

ELEMENTOS DE ECONOMETRIA

JAN KMENTA

Profesor de Economía
Michigan State University

Traducción de:

MANUEL PEÑALVER
Licenciado en Ciencias Económicas

Adaptación de:

LUIS BARBE DURAN
Catedrático de Teoría Económica
de la Universidad Autónoma de Barcelona

 **VICENS universidad**

330.

k _ _ _ e

A Joan

Dirección de Edición: Albert Vicens

Primera edición, 1977

Título Original:
Elements of Econometrics

© THE MACMILLAN COMPANY, New York, 1971

© JAN KMENTA, 1971

© MANUEL PEÑALVER, 1977
Sobre la versión española

Prohibida la reproducción total o parcial de la presente obra sin permiso expreso, por escrito, del Editor. Se exceptúan de esta prohibición las citas breves en artículos de periódicos o revistas destinadas a reseñar esta obra.

—Depósito Legal: B. 45 059-1977

—ISBN: 84-316-1448 X

1.º de Orden V. V.: A-715

—IMPRESO EN ESPAÑA

—PRINTED IN SPAIN

editado por Editorial VICENS VIVES A.ª de C.ª y S.ª de C.ª
impreso por Gráficas INSTAR S.ª de C.ª de C.ª y S.ª de C.ª

Prefacio

Este libro está dirigido a los economistas, es decir, a quienes deben comprender las técnicas econométricas modernas, y, quizá, quienes utilizarlas y no a los futuros econométricos que piensen participar en el desarrollo y perfeccionamiento de su disciplina. Su contenido está planificado para abarcar un curso anual de introducción a la econometría, que muchas Universidades ofrecen a los estudiantes de economía en el primer año de postgraduado. Sin embargo, también puede servir como texto para un curso de estadística económica o empresarial para estudiantes no graduados de nivel avanzado, o como libro de referencia para economistas de cualquier especialidad que quieran estar al corriente de los trabajos actuales de investigación económica aplicada. Tanto los estudiantes como los profesores de economía necesitan, cada vez más, estar familiarizados con los métodos cuantitativos para entender la literatura económica actual y ser efectivos en sus propias tareas de investigación. Mi intención ha sido abarcar la mayor parte de los métodos estadísticos y econométricos que los economistas modernos pueden requerir. En mi opinión, se necesita un libro que proporcione al economista tanta econometría como sea posible sin que ello requiera una gran habilidad matemática por su parte. A partir de mi experiencia con estudiantes postgraduados en la Universidad de Wisconsin y en la Universidad del Estado de Michigan, he descubierto que es totalmente posible partir de un nivel de conocimientos estadísticos muy reducido y pasar con gran rapidez a sofisticados temas econométricos. Naturalmente, ello requiere un serio esfuerzo, tanto de los estudiantes como del profesor. Por mi parte, he sido lo bastante afortunado como para encontrarme con estudiantes que tenían un gran interés, y que encontraban verdadera satisfacción en ser capaces de resolver problemas que, sólo dos semanas antes, parecían estar totalmente fuera de su alcance. En cuanto al trabajo del profesor, creo que su obligación es subrayar en todo momento la necesidad de alcanzar una comprensión total, en lugar de un aprendizaje del tipo de libro de cocina, y no ahorrar ningún esfuerzo para hacerlo todo tan claro como sea posible. Esta es la intención con que se ha escrito este libro.

Como el libro ya supera en buena medida la extensión que originalmente se pretendía, no se han incluido algunos temas. Uno de ellos es el análisis espectral, que aunque proporciona una nueva técnica de interés, aún no se ha integrado estrechamente con la teoría económica, y resulta bastante difícil de manejar por un no especialista. Asimismo, un análisis del enfoque bayesiano de la estadística constituiría otro tema interesante (algunos de mis amigos dirían el tema interesante), pero para hacerle justicia se requeriría un espacio considerable. Un tercer tema que se ha dejado de lado, sintiéndolo mucho, es el de la simulación de modelos econométricos. El economista interesado en estos temas puede consultar textos más especializados. El objetivo de este libro es dotarle de las herramientas básicas que faciliten y hagan más eficiente esta consulta.

En relación con este libro, debo hacer constar muchos agradecimientos, algunos directos y otros indirectos. A Arthur Goldberger, cuya excelente enseñanza y devoción a la investigación —así como claridad en ambos campos— han inspirado gran parte del espíritu de mi obra, le debo un agradecimiento constante. Varios de los temas de los que me ocupé en el libro han sido desarrollados por Arnold Zellner, hacia quien me siento agradecido no sólo por este motivo, sino porque siempre se ha mostrado generoso discutiendo problemas de econometría y prestándome la ayuda de su excepcional visión. También he recibido ayuda editorial y aliento de Lawrence Klein, quien amablemente leyó el manuscrito en su totalidad y me ofreció gran cantidad de comentarios y sugerencias de gran utilidad. Desgraciadamente, no he sido capaz de incorporarlas todas debido a limitaciones de tiempo y espacio, pero el libro sin duda se hubiera visto beneficiado de haberse seguido tales sugerencias en su totalidad. He recibido la ayuda indirecta, pero importante, de Robert Basman y Karl Brunner, quienes me hicieron penosamente consciente de muchas deficiencias metodológicas, y de James Ramsey, cuyos conocimientos técnicos e imaginativa visión siempre han estado a mi disposición. Debo agradecimientos específicos a Roy Gilbert por su ayuda en la calculadora, a Jeffrey Roth que verificó los ejemplos, y a William Ruble por su ayuda en los programas de ecuaciones simultáneas. La Sra. Elisabeth Belfer, de The Macmillan Company, merece especial mención por su concienzuda ayuda en la lectura de pruebas. Naturalmente, el mayor agradecimiento lo debo a mi esposa, que ha sido el lector de pruebas más paciente además de encargarse de todos los asuntos familiares que, de otro modo, se hubiese inmiscuido en mi tarea.

Por último, quiero expresar mi aprecio y agradecimiento a Arnold Zellner y Martin Geisel por su autorización para citar trabajos no publicados, y a J. Zubrzycki por su permiso para utilizar los datos de los ejemplos en la sección 9.2. También agradezco al albacea literario de Sir Ronald Fisher, F.R.S., y a Oliver & Boyd Ltd. de Edimburgo, su autorización para reproducir la tabla D-2 de su libro *Statistical Methods for Research Workers*, al profesor E. S. Pearson y a los administradores de *Biometrika* su autorización para reproducir

el material de las tablas D-3, D-4 y D-5; y a la Rand Corporation su permiso para reproducir el material de la tabla D-6, de su libro *A Million Random Digits and One Hundred Thousand Deviates*. La publicación del material de la tabla D-5 también fue autorizada por el profesor J. Durbin y el profesor J. S. Watson.

East Lansing

J. K.

Indice

Prólogo	VII
-------------------	-----

PRIMERA PARTE

Teoría Estadística Básica

Capítulo 1. Introducción a la Inferencia Estadística	3
1.1. Conceptos básicos de la inferencia estadística	3
1.2. La naturaleza de la inferencia estadística	8
1.3. Distribuciones muestrales	9
1.4. Propiedades de las distribuciones muestrales	12
1.5. Obtención de distribuciones muestrales	16
Capítulo 2. Obtención experimental de distribuciones muestrales	21
2.1. Distribución muestral de la proporción de aciertos en la muestra	23
2.2. Distribución muestral de la media de la muestra	27
Capítulo 3. Probabilidad y distribuciones de probabilidad	35
3.1. Conjuntos y espacios muestrales	37
3.2. Permutaciones y combinaciones	40
3.3. Teoremas básicos de la teoría de la probabilidad	47
3.4. Funciones de probabilidad y variables aleatorias discretas	56
3.5. Variables aleatorias continuas y funciones de probabilidad	62
3.6. Esperanza matemática	66
Capítulo 4. Obtención teórica de distribuciones muestrales	81
4.1. Distribución muestral de la proporción de aciertos de la muestra: La distribución binominal	82
4.2. La distribución normal como caso límite de la distribución binominal	91
4.3. Distribución muestral de la media de la muestra	104

Capítulo 5. Contrastación de hipótesis	129
5.1. Establecimiento y valoración de pruebas para la contrastación	129
5.2. Distribución de algunos estadísticos de prueba	157
Capítulo 6. Estimación	179
6.1. Propiedades de los estimadores	180
6.2. Métodos de estimación	200
6.3. Intervalos de confianza	219

SEGUNDA PARTE

Teoría Econométrica Básica

Capítulo 7. Regresión simple	231
7.1. Relaciones entre variables	231
7.2. El modelo de regresión	236
7.3. Estimación de los parámetros de la regresión	241
7.4. Otros resultados de la inferencia estadística	255
Capítulo 8. Violaciones de los supuestos básicos	293
8.1. Heteroscedasticidad	295
8.2. Perturbaciones autorregresivas	319
8.3. Variable explicativa estocástica	354
Capítulo 9. Estimación con datos deficientes	367
9.1. Errores de medición	367
9.2. Estimación a partir de datos agrupados	385
9.3. Estimación faltando algunas observaciones	401
Capítulo 10. Regresión múltiple	413
10.1. Estimación de los parámetros de la regresión	413
10.2. Otros resultados de la inferencia estadística	425
10.3. Multicolinealidad	433
10.4. Errores de especificación	447
Capítulo 11. Formulación y estimación de modelos especiales	459
11.1. Modelos con variables binarias	459
11.2. Modelos con restricciones en los coeficientes	514
11.3. Modelos no lineales	539
11.4. Modelos con retardos distribuidos	544

Capítulo 12. El modelo generalizado de regresión lineal y sus aplicaciones	599
12.1. El modelo generalizado de regresión lineal	599
12.2. Consideración conjunta de datos de series temporales y de sección transversal	610
12.3. Regresiones aparentemente no relacionadas	621
Capítulo 13. Sistemas de ecuaciones simultáneas	639
13.1. Descripción de los sistemas de ecuaciones simultáneas	640
13.2. El problema de la identificación	648
13.3. Métodos de estimación de ecuaciones aisladas	661
13.4. Métodos de estimación de sistemas	689
13.5. Comparación de los métodos alternativos de estimación y problemas especiales	698
13.6. Análisis de los modelos econométricos dinámicos	708
Apéndice	721
A. Álgebra de Sumatorios	721
B. Elementos de álgebra matricial	721
C. Diseño del método de cálculo para la estimación minimocuadrática	724
Lecturas adicionales sugeridas	745

PRIMERA PARTE: TEORIA ESTADISTICA BASICA

Capítulo

1

Introducción a la Inferencia Estadística

Hasta principios del siglo XIX, se consideró que la estadística se ocupaba de las características del Estado, particularmente de las relacionadas con sus instituciones políticas y militares. La descripción de tales características fue al principio de carácter literario, pero con el tiempo se hicieron cada vez más numéricas. La adquisición de un carácter cada vez más cuantitativo se vio acompañada por la extensión a otros campos distintos de los relacionados con los asuntos del Estado. Los posteriores adelantos en la teoría de la probabilidad permitieron el desarrollo de la teoría estadística, que hace posible la generalización científica a partir de información incompleta —en otras palabras, la inferencia estadística—.

Debido a este proceso histórico, la disciplina que se conoce con el nombre de «estadística» consta de dos partes: la estadística descriptiva y la inferencia estadística. La estadística descriptiva se ocupa de la recogida, organización y presentación de los datos, mientras que la inferencia estadística realiza generalizaciones, de una parte a un todo. La inferencia estadística, como cualquier otra ciencia, se ocupa del desarrollo de sus métodos (teoría estadística) y de su utilización (aplicación de la estadística).

En la econometría, básicamente nos interesa la inferencia estadística. La estadística descriptiva sólo nos importa hasta el punto en que las medidas obtenidas por los estadísticos descriptivos con el fin de resumir distintas características de los datos —promedios, medidas de dispersión, etc.— se utilizan también en la inferencia estadística. La diferencia básica radica en el hecho de que, en el ámbito de la estadística descriptiva, dichas medidas representan fines en sí mismas, mientras que en la inferencia estadística son sólo instrumentos en el proceso de investigación.

1.1 Conceptos básicos de la inferencia estadística

Antes de explicar de forma detallada la naturaleza de la inferencia estadística es preciso introducir algunos conceptos básicos. Los más importantes son el de población y el de muestra.

Una *población* puede definirse como el conjunto de todas las observaciones posibles de medidas o resultados. Ejemplos de ello son los ingresos de los habitantes de un país en un periodo de tiempo determinado, la renta nacional de un país a lo largo de un cierto número de periodos de tiempo, y todos los resultados obtenidos al lanzar una moneda al aire repetidamente. Las poblaciones pueden ser finitas o infinitas. Una *población finita* es aquella cuyo número de observaciones posibles es menor que infinito. Sin embargo, la distinción entre poblaciones finitas e infinitas es más sutil de lo que pueda parecer a primera vista. Por ejemplo, una serie de las cifras de la renta nacional de los Estados Unidos durante un cierto número de años, por ejemplo, 1948-1967, representa un conjunto finito de veinte observaciones y por tanto, puede parecer una población finita. Pero esto representaría una interpretación muy limitada de los hechos históricos, ya que implicaría que las veinte mediciones de la renta nacional eran las únicas posibles, es decir, que la historia sólo podía haber seguido un determinado curso. Naturalmente, muy pocas personas aceptarían una concepción del mundo tan extremadamente fatalista; la mayoría admitiría la posibilidad de que se hubiesen producido otros valores, aunque sólo fuesen ligeramente distintos, de la renta nacional. Este último punto de vista es el subyacente en prácticamente toda la investigación económica y econométrica relacionada con la política económica y será el utilizado a lo largo de todo este libro. Así, pues, una población de rentas nacionales en un intervalo de tiempo determinado incluye no sólo la historia real, representada por los valores que se observaron, sino también la historia potencial, que consiste en todos los valores que podían haber ocurrido pero no ocurrieron. La población definida de este modo es obviamente una población infinita. Del mismo modo, la población de todos los resultados posibles al lanzar una moneda al aire es también infinita, puesto que el proceso puede generar un número infinito de resultados, en este caso «caras» y «cruces». La mayor parte de las poblaciones de las que se ocupa la econometría son infinitas.

En relación con el concepto de población, aparece el concepto de *muestra*, que es un conjunto de medidas o resultados seleccionado a partir de la población. La selección la puede efectuar el investigador, en cuyo caso hablamos de un experimento muestral, o puede ocurrir de forma independiente ya sea intencionadamente o por su propia naturaleza. En el último caso, el investigador es un mero observador y esta situación es particularmente frecuente en econometría. Aunque las muestras que proceden de poblaciones infinitas pueden ser también infinitas, la importancia de este tipo de muestras es, como máximo, de carácter teórico. En la práctica, sólo nos encontramos con muestras finitas y, desgraciadamente, muy a menudo con muestras muy pequeñas. Como las muestras se obtienen mediante una selección a partir de una población determinada, el principio de selección desempeña claramente un importante papel en la determinación de la composición de la muestra. En econometría limitamos nuestra atención a las muestras obtenidas de acuerdo con algún me-

mo casual determinado. Estas muestras reciben el nombre de *muestras probabilísticas*. Un tipo importante de muestras probabilísticas son las *muestras aleatorias*. En las poblaciones finitas, el principio para seleccionar una muestra aleatoria consiste en conceder a cada individuo de la población la misma posibilidad de ser elegido. En el caso de poblaciones infinitas, una muestra es aleatoria si cada observación (de una medición o un resultado) es independiente de todas las demás observaciones. El significado de *independencia* se definirá de forma rigurosa más adelante, y de momento basta con señalar que dos sucesos (ya sean medidos o contados) son independientes si el hecho de que ocurra uno de ellos no ejerce ninguna influencia sobre la ocurrencia del otro.

Tanto las poblaciones como las muestras se pueden describir a través de sus características. Las características numéricas de una población se denominan *parámetros*; las características de una muestra, expresadas en forma de alguna medida resumen, reciben el nombre de *estadísticos* (plural de «estadístico»). Estas características pueden ser, por ejemplo, la tendencia central de las mediciones (la media o la moda), su dispersión (por ejemplo, la desviación standard), o, en caso de tratarse de fenómenos cualitativos, la proporción de observaciones de un tipo determinado. Naturalmente, los parámetros de una población infinita no se pueden observar nunca; en cambio, los parámetros de una población finita teóricamente se podrían observar aunque en la práctica puede resultar imposible.

A partir de lo que llevamos dicho hasta aquí, debería quedar claro que la estadística se ocupa de los fenómenos que puedan ser medidos o contados. Cuando nos referimos a un fenómeno que se puede medir, hablamos de una *variable* para indicar una cantidad homogénea que puede tomar valores distintos en distintas observaciones. Si el fenómeno sólo se puede contar pero no medir (cada observación representa una unidad de cuenta), hacemos referencia a un *atributo*. Por tanto, un atributo es la presencia o la ausencia de una característica determinada. El resultado de un suceso tal como, por ejemplo, el nacimiento de un niño origina una observación respecto a un atributo de sexo (es decir, «varón» o «no varón»); el resultado de lanzar un dado se puede clasificar atendiendo a la presencia o ausencia de un «1», un «2», etc... En cierto sentido, el concepto de atributo resulta redundante porque podemos asignar simplemente el valor 1 a la presencia de una característica determinada y el valor 0 a su ausencia. En este caso, equiparamos el concepto de «atributo» al concepto de *variable binaria* o *cualitativa*. El mismo concepto recibe también el nombre de «variable ficticia».

La definición de *variable*, y su propio nombre, subraya la posibilidad de variación en los distintos puntos de observación. Por otra parte, las cantidades que no pueden variar de una observación a otra reciben el nombre de *constantes* (Obsérvese que los parámetros son también constantes, puesto que no varían de una observación a otra.) Si la cantidad en cuestión no es una cons-

tante sino una variable, nos puede interesar conocer la causa general de la variación. En especial, es importante distinguir entre las variaciones que se pueden manipular y controlar, y las que no se pueden controlar totalmente. En el caso de estas últimas, su existencia se debe al azar. Un ejemplo obvio de variación incontrolada serían los resultados de lanzar una moneda al aire (sin trucos, naturalmente), pero existen muchos otros ejemplos que son menos obvios. En realidad, tal como se analizará detalladamente a lo largo del libro, la mayor parte de las variables económicas están determinadas siempre, parcialmente, por el azar. Las variables cuyos valores no se pueden determinar antes de ser observados, ni controlar totalmente, se denominan *variables aleatorias* o *estocásticas*; su característica principal consiste en el hecho de que toman distintos valores (o están en distintos intervalos), con probabilidades que no son la unidad. Por el contrario, una variable *no aleatoria*, o *no estocástica*, o *fija*, es aquella que es totalmente controlable o que, al menos, se puede predecir por completo. Las constantes se pueden considerar casos especiales de variables fijas.

Otra clasificación importante de las variables es la que distingue entre variables continuas y variables discretas. Una *variable continua* es una variable que puede tomar cualquier valor sobre el eje numérico o una parte de él. Los ejemplos típicos son el tiempo y la temperatura, pero la renta, el gasto, y otras variables similares se pueden considerar variables continuas. En realidad, la mayor parte de las variables económicas son continuas o casi continuas. Esta última posibilidad se añade para responder a las posibles objeciones señalando, por ejemplo, que en la práctica no se pueden observar valores monetarios inferiores a un dólar (o, posiblemente, un centavo). Contrariamente a las variables continuas, las *variables discretas* sólo pueden tomar algunos valores determinados sobre la escala numérica. Estos valores están separados normalmente (aunque no siempre) por intervalos de la misma longitud. Ejemplos de este tipo de variables son el número de hijos de una familia, el número de puntos obtenidos al lanzar un dado, y todas las variables binarias.

Cuadro 1-1

Número de hijos (= Variable)	Número de familias (= Frecuencia absoluta)	Proporción de familias (= Frecuencia relativa)
0	4	0.0625
1	12	0.1875
2	20	0.3125
3	16	0.2500
4	8	0.1250
5 o más	4	0.0625
Totales	64	1.0000

En las tablas que siguen se ha utilizado el sistema de puntuación anterior.

El último concepto que debemos introducir ahora es el de *distribución*. En el caso de una muestra tenemos una distribución de frecuencias y cuando se trata de una población hacemos referencia a una distribución de probabilidades. Una *distribución de frecuencias* representa una ordenación de los datos de tal forma que nos indique el número de observaciones correspondiente a cada valor de la variable (en el caso de variables discretas) o a cada intervalo de los valores de la variable (si se trata de variables continuas). El número de observaciones en cada clase (un punto si se trata de una variable discreta o un intervalo si es una variable continua) se denomina *frecuencia absoluta*. Esta última debe distinguirse de la *frecuencia relativa* que nos indica la proporción de observaciones en cada clase, y no su número. Por ejemplo, consideremos una muestra de 64 familias, observando el número de hijos de cada una de ellas. Los resultados podrían ser los que aparecen en el cuadro 1-1. Otro ejemplo, relativo en esta ocasión a una variable continua, viene dado por la distribución de la renta familiar en los Estados Unidos, en 1962 (cuadro 1-2). En este caso no aparecen las frecuencias absolutas, y las frecuencias relativas se presentan en forma de porcentajes en lugar de proporciones. Los datos de una muestra en forma de serie temporal, como, por ejemplo, las cifras de la renta nacional correspondientes a un cierto número de años, se podrían presentar también en forma de distribución de frecuencias, pero, normalmente, no se suele hacer así. El hecho de que las distintas observaciones se hayan efectuado en distintos períodos de tiempo sólo tiene importancia en la medida en que la población de la que se ha extraído la muestra puede haber experimentado cambios a lo largo del tiempo.

En una población, el concepto correspondiente a la distribución de frecuencias de una muestra se conoce con el nombre de *distribución de probabilidades*. Considérese, por ejemplo, la población de las familias de los Estados

Cuadro 1-2

Renta deducidos los impuestos (= Variable)	Porcentaje de unidades de consumo * (= Frecuencia relativa)
Menos de \$2,000	12.7
\$2,000 a \$4,999	31.9
\$5,000 a \$9,999	39.7
\$10,000 o más	15.7
Total	100.0

Fuente: *Statistical Abstract of the United States, 1965*, página 341.

* Incluye familias e individuos que viven solos (número total = 57 890 000).

Unidos clasificadas según la renta obtenida en 1962, tal como aparecen en el cuadro 1-2. Resulta evidente que el hecho de decir que el 12,7 % de las familias obtuvieron una renta inferior a 2000 dólares equivale a afirmar que la probabilidad de seleccionar (al azar) una familia cuya renta sea inferior a 2000 dólares es 0,127. Si la población es infinita, las probabilidades se pueden considerar como los *límites* de las frecuencias relativas (esto se explicará de forma más rigurosa en el capítulo 3). Si consideramos la distribución de probabilidades de una variable como la contrapartida de la distribución de frecuencias de una muestra, podemos ver que es posible analizar más de una variable a la vez. Por ejemplo, una distribución que indique la probabilidad de fallecimiento a distintas edades hace referencia a una sola variable —es una *distribución univariante*—. Sin embargo, si obtenemos estas probabilidades para cada sexo por separado, consideramos dos variables y obtenemos una *distribución bivalente*. Otra clasificación en la que se tuviesen en cuenta mayor número de características produciría una *distribución multivalente*.

1.2 La naturaleza de la inferencia estadística

Tras haber introducido, aunque sea brevemente, algunos de los conceptos más importantes de la teoría estadística, podemos pasar a describir la naturaleza de la inferencia estadística. Como se ha señalado anteriormente, esta última se dedica a efectuar generalizaciones respecto a la población, en base a la información proporcionada por la muestra. Naturalmente, este procedimiento es corriente en la vida diaria: efectuamos generalizaciones acerca de la temperatura de nuestra sopa sobre la base de la primera cucharada, o respecto a la esperanza de vida de un par de zapatillas de tenis, basándonos en la experiencia pasada. Esto es precisamente lo que se hace en la inferencia estadística, salvo que, en este último caso, se hace de forma más científica. Lo que convierte la aplicación de la inferencia estadística en un proceso científico es el hecho de que tomamos en consideración la forma de seleccionar la muestra, y de que expresamos las generalizaciones en términos de probabilidades. Por ejemplo, en lugar de decir que las zapatillas de tenis duran cinco años, especificamos un intervalo e indicamos el nivel de probabilidades correspondiente.

En resumen, utilizamos las muestras para establecer juicios respecto a las poblaciones de las que provienen las muestras. Si la población es infinita, no se podrá observar nunca por completo, y los juicios respecto a ella sólo pueden proceder de una muestra. Pero incluso en el caso de que la población sea finita, puede haber buenas razones para observar únicamente una muestra, puesto que las observaciones pueden ser destructivas (como en el caso de probar la sopa o de medir la duración de las bombillas), o, en el mejor de los casos, caras. En general, no nos interesa conocerlo todo respecto a una población, sino que estamos interesados solamente en *algunas* de sus características, a las que, como se recordará, denominamos *parámetros*. El objetivo del muestreo, y de todo lo relacionado con la inferencia estadística, es efectuar juicios acerca de los parámetros de la población, basados en los estadísticos de la muestra. En realidad, estos juicios son pronósticos dotados de un cierto grado de confianza, y pueden ser de dos tipos según hagan referencia a la estimación de un parámetro, o a la contrastación de alguna hipótesis respecto a un parámetro. La estimación se realiza mediante los *estimadores* que son fórmulas que describen un procedimiento para efectuar conjeturas acerca del valor de un parámetro determinado de la población; el valor específico de un estimador recibe el nombre de *estimación*. Los juicios que toman la forma de *contrastación de hipótesis* implican un supuesto previo respecto al valor de un parámetro. Si la información procedente de la muestra proporciona evidencia contraria a la hipótesis, rechazamos esta última; en caso contrario, la seguimos manteniendo. Para la contrastación de hipótesis, la evidencia proporcionada por las observaciones que constituyen la muestra se resume en un *estadístico de prueba*; a través de este último se llega a una decisión relativa a la hipótesis.

Las muestras proporcionan información acerca de las poblaciones de las que proceden. Cuando se trata de un problema de estimación, esta información se resume en forma de un estimador, y si se trata de la contrastación de una hipótesis, la información se resume en un estadístico de prueba. En ambos casos utilizamos alguna fórmula, aplicándola a los valores observados en la muestra. Los valores del estimador y del estadístico de prueba obtenidos por este procedimiento están estrechamente relacionados, lo cual resulta lógico si consideramos que ambos proceden de la misma fuente de información, es decir, la muestra. En ambos casos, el valor de un estimador o de un estadístico de prueba representa un pronóstico acerca del valor del correspondiente parámetro de la población. Ahora bien, es evidente que diferentes muestras darán como resultado pronósticos diferentes, algunos de ellos más próximos a la verdad (al valor real del parámetro) que los otros. Naturalmente, en la práctica sólo disponemos en general de una muestra y, por tanto, de un solo pronóstico, pero resulta muy importante saber qué otros pronósticos podríamos haber obtenido si hubiésemos dispuesto de distintas muestras. Si todas las muestras posibles conducen a pronósticos que en todo caso están próximos a la verdad, cualquiera de ellos resultará de fiar. Por otra parte, si todas las muestras posibles producen pronósticos que difieren ampliamente, sólo algunos de ellos estarán cerca de la verdad y no podremos confiar mucho en ninguno por separado. El tercero de los casos extremos que se pueden presentar ocurre cuando todos los pronósticos posibles son similares pero están alejados del valor real del parámetro.

1.3 Distribuciones muestrales

El párrafo anterior sugiere que la forma de saber en qué medida se puede confiar en un pronóstico consiste en conocer la conducta de todos los que se

pueden obtener a partir de todas las muestras posibles. Podemos imaginar que se van obteniendo muestras, una detrás de otra, y que para cada una de ellas se calcula el valor de la suposición (por ejemplo, la estimación de un parámetro de la población), ordenándolas en forma de una distribución. Si tuviésemos un número infinito de muestras, la distribución resultante se denominaría *distribución muestral*. Considérese, por ejemplo, el problema de estimar la renta familiar media en los Estados Unidos durante un año determinado, a partir de una muestra de 100 familias. Una de las posibilidades consiste en calcular la renta familiar media de la muestra y usarla como estimación de la media de la población. Naturalmente, podríamos utilizar como estimador la moda, la mediana, y cualquier otra medida. Supongamos que estimamos la media de la población utilizando la media de la muestra y queremos saber en qué medida podemos confiar en este estimador. Una forma de saberlo consistiría en obtener un número infinito de muestras de este tipo, calculando el valor de la media de cada una de ellas y ordenando estos valores en forma de distribución. Obsérvese que, pese a que la población de todas las familias de los Estados Unidos es una población finita, el número de muestras que se pueden extraer de esta población es infinito si cada familia tiene la posibilidad de aparecer en cada una de las muestras. Este procedimiento de muestreo se denomina *muestreo con reposición*. Estudiando la distribución muestral resultante, conoceríamos por completo la conducta posible de nuestro pronóstico. Si en realidad conociéramos de antemano las características de la población, este ejercicio serviría para ampliar nuestro conocimiento respecto a la relación entre la media de la muestra y la de la población, y este conocimiento podría ser utilizado en otras ocasiones en las que nos viésemos limitados a disponer de una sola muestra.

Si cada familia que aparece en la muestra está seleccionada al azar, no podemos saber de antemano cuál es su renta y, por tanto, en este caso, la renta familiar es una variable aleatoria. Además, la renta media obtenida en una muestra también es una variable aleatoria. Ello significa que la distribución muestral de la media de la muestra (basada en un número infinito de muestras) es en realidad una distribución de probabilidad. Esta distribución puede ser discreta o continua según la variable de la población sea una variable discreta o continua; en nuestro ejemplo, la distribución muestral es continua porque la renta es una variable continua. Naturalmente, esta idea es completamente general: *una distribución muestral es una distribución de probabilidad de un estimador o de un estadístico de prueba*.

Resulta evidente que las muestras de distinto tamaño proporcionan distinta cuantía de información respecto a la población de la que proceden. Por tanto, los estimadores que utilizan toda la información contenida en una muestra, y se basan en muestras de distinto tamaño, poseerán un grado distinto de fiabilidad. A fin de evitar los efectos de los distintos tamaños muestrales sobre la calidad de un estimador, todas las distribuciones muestrales se refieren

siempre a muestras del mismo tamaño. Los efectos de las variaciones en el tamaño de la muestra se estudian a través de la comparación de distintas distribuciones muestrales.

Supongamos que estamos estudiando la población de todos los valores posibles de una variable X (por ejemplo, las rentas familiares en un determinado año) y que nos interesa estimar un parámetro θ (por ejemplo, la renta familiar media). Para estimar este parámetro, utilizamos un estadístico muestral $\hat{\theta}$. En la terminología corriente, $\hat{\theta}$ es un estimador de θ , y un valor concreto de $\hat{\theta}$ (obtenido a partir de una muestra determinada) es una estimación de θ . (Una convención corriente consiste en utilizar letras griegas para describir los parámetros de la población, y esas mismas letras griegas con acento circunflejo, tildes, etc., para referirse a los estimadores respectivos.) Si X es una variable continua (como en el caso de la renta familiar), la distribución muestral de $\hat{\theta}$ tendrá una apariencia similar a la de la distribución de la figura 1-1. Como se ha señalado anteriormente, esto es en realidad una distribución de probabilidad, y como este tipo de distribuciones no se definirá hasta el capítulo 3, no resulta posible en este momento estudiar de forma rigurosa las distribuciones muestrales. Sin embargo, se puede entender mejor el concepto de distribución muestral si, de momento, lo consideramos como una distribución de frecuencias relativas calculada a partir de un número infinito de observaciones (en este caso, de muestras). En la figura 1-1, las frecuencias relativas de $\hat{\theta}$ se miden a lo largo del eje $f(\hat{\theta})$.

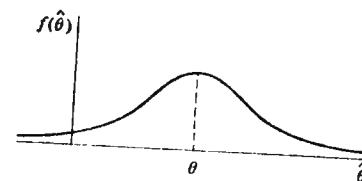


Figura 1-1

Un procedimiento razonable, y generalmente útil, para juzgar la calidad de un pronóstico consiste en valorar la calidad del procedimiento empleado para obtenerlo. Supongamos, por ejemplo, que un hombre que ha dedicado una cantidad de tiempo considerable al estudio de las carreras de caballos acude a un hipódromo con su esposa, que es totalmente ignorante en este tipo de materias, y supongamos que en la primera carrera la esposa acierta en su apuesta mientras que su marido la pierde. A partir de este hecho, no obtenemos necesariamente la conclusión de que los pronósticos de la esposa son mejores que los de su marido, y de que en la próxima carrera tenemos que apostar por el mismo caballo que ella. Tal conclusión sólo se puede alcanzar después de obtener suficiente información (es decir, mayor número de observaciones) y de que el

«sistema» de la esposa haya producido mayores beneficios (o menores pérdidas) que el de su marido. Lógicamente, si sólo se permite una apuesta —la de la primera carrera—, el pronóstico de la esposa es mejor sin discusión, ya que produjo un beneficio mientras que el marido obtuvo una pérdida, pero esto es un caso especial de muy poco interés. En general, el resultado de cualquier pronóstico concreto se considera de poco interés; lo realmente importante es con qué frecuencia un procedimiento determinado produce malos resultados, y con qué frecuencia los produce buenos. En otras palabras, debemos conocer los resultados de una gran cantidad de pronósticos, todos ellos basados en el mismo procedimiento. Esta es precisamente la información que proporcionan las distribuciones muestrales. Para comparar la calidad de los pronósticos, se comparan los resultados de los procedimientos por medio de los cuales se obtuvieron aquéllos, lo cual quiere decir que se comparan sus distribuciones muestrales.

1.4 Propiedades de las distribuciones muestrales

La discusión anterior nos condujo a definir qué es un buen procedimiento para efectuar pronósticos y, por lo tanto, un buen estimador. En otros términos, queremos saber cuáles son los rasgos específicos de una distribución muestral que nos permiten juzgar un determinado estimador. En el capítulo 6 nos ocuparemos de este problema de manera detallada, pero ahora se pueden efectuar algunas observaciones. Podemos empezar con el caso más simple, que es el correspondiente a un estimador perfecto. Un estimador perfecto es el que no tiene posibilidad de error, es decir, aquel cuya distribución muestral está concentrada totalmente en un solo punto, y este punto coincide con el valor real del parámetro que se quiere estimar. No es preciso decir que los estimadores perfectos resultan muy infrecuentes. Un caso en el que nos podemos encontrar con un estimador perfecto es aquel en el cual no se produce variación en la población. En el ejemplo correspondiente al muestreo de la temperatura de un plato de sopa a partir de una cucharada, nuestra conjetura sería perfecta si la temperatura fuese la misma en todas las partes del plato, y usáramos como estimador la temperatura de la cucharada. Normalmente, esto se podría conseguir batiendo cuidadosamente la sopa antes de probarla. En el ejemplo de la renta familiar, tendríamos un estimador perfecto si todas las familias tuviesen la misma renta y usáramos la media de una muestra como estimación de la media de la población. Otro caso que puede producir estimación perfecta es aquel en que la muestra es de tamaño infinito. Pero debe señalarse que la existencia de cualquiera de estos dos casos no garantiza por sí solo una estimación perfecta; si se ignora una parte de la información muestral o su totalidad, o se usa erróneamente, podemos equivocarnos. Por ejemplo, si nuestro estimador de la renta familiar media no fuese la media de la muestra sino esta última aumentada en un diez por ciento (quizá debido a la creencia

de que la gente tiende a subestimar sus rentas), sin que hubiese ninguna necesidad de hacer este aumento (si en realidad la gente indicaba sus rentas correctamente), el estimador no sería perfecto a pesar de la inexistencia de variación en la población, e independientemente del tamaño de la muestra.

Casi siempre, los estimadores no son perfectos, sino que son de tal tipo que únicamente una pequeña proporción de sus valores coincide con el verdadero valor del parámetro o se aproxima mucho a él. Ello implica que deberíamos contentarnos con unos resultados inferiores, y estos últimos se pueden resumir enunciando algunas propiedades de los estimadores que normalmente se consideran deseables. En este punto sólo mencionaremos la idea básica subyacente en tres de tales propiedades y en otro capítulo se efectuará un análisis amplio y elaborado de este tema. Quizá la más conocida de las propiedades que resultan deseables en un estimador es la de que sea *insesgado*. Un estimador insesgado es aquel cuya distribución muestral tiene una media igual al parámetro que se pretende estimar. Un estimador perfecto proporciona un pronóstico perfecto en cada ocasión; los estimadores insesgados sólo proporcionan resultados perfectos en promedio. Los estimadores insesgados producirán estimaciones en algunos casos superiores y en otros inferiores al verdadero valor del parámetro, pero los excesos y los defectos «se compensan» al efectuar una cantidad infinita de estimaciones. Si la distribución muestral es simétrica, el hecho de que un estimador sea insesgado implica que la mitad de las estimaciones posibles es mayor que el valor del parámetro, y la otra mitad es inferior a él. En la figura 1-1 se representa una situación de este tipo (en el caso de un estimador continuo). Debe hacerse hincapié en el hecho de que la insesgadez no nos indica nada respecto a las distancias entre las estimaciones y el valor del parámetro, aparte del hecho de que estas distancias (positivas y negativas) suman cero en total. Es perfectamente posible que un estimador insesgado no origine nunca una estimación igual al valor del parámetro. Considérese el ejemplo de la estimación de la renta familiar media a través de la media de una muestra. Aceptemos la proposición —que demostraremos más adelante— de que la media de la muestra es un estimador insesgado de la media de la población, y supongamos que la renta familiar media de la población es 6573.46 dólares; obviamente, la probabilidad de que una media muestral proporcione exactamente esta cifra es prácticamente nula.

Otra propiedad importante que resulta deseable en los estimadores es la *eficiencia*, y hace referencia a la distancia entre los valores del estimador y el valor del parámetro. Desgraciadamente, no existe ninguna definición de eficiencia generalmente aceptada que abarque todos los casos. Sin embargo, si consideramos únicamente los estimadores insesgados, si existe una definición generalmente aceptada; en este caso, un estimador eficiente es aquel que posee la menor dispersión, es decir, aquel cuya distribución muestral tiene la varianza más pequeña. En la figura 1-2 se representan dos estimadores, θ_1 cuya distribución muestral es $f_1(\hat{\theta})$ y θ_2 cuya distribución muestral es $f_2(\hat{\theta})$. Ambos

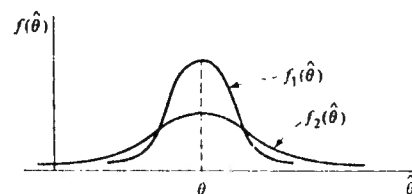


Figura 1-2

estimadores son insesgado, pero $\hat{\theta}_2$ tiene mayor dispersión que $\hat{\theta}_1$ y es, por tanto, menos eficiente. Si no fuese posible obtener otro estimador insesgado cuya variancia fuese menor que la de $\hat{\theta}_1$, este último sería un estimador eficiente entre la familia de todos los estimadores insesgados. Si no queremos limitarnos únicamente a los estimadores insesgados, el concepto de eficiencia se hace ambiguo, puesto que en tal caso debemos considerar la relación entre sesgo y variancia. La figura 1-3 representa un caso de este tipo: el estimador $\hat{\theta}_1$ es

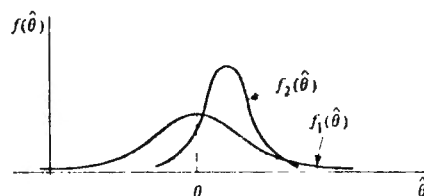


Figura 1-3

insesgado pero tiene una gran variancia, mientras que $\hat{\theta}_2$ es sesgado pero su variancia es más pequeña. En este caso no podemos decir cuál de los dos estimadores es más eficiente, a menos que asignemos pesos relativos (es decir, precio) al sesgo y a la variancia. Sin embargo, hay que señalar que la variancia mínima no es, en sí misma, una propiedad deseable; en caso contrario, podríamos, simplemente, utilizar alguna constante (cuya variancia es, por definición, cero) independientemente de la evidencia proporcionada por la muestra. La distribución muestral de un «estimador» de este tipo se concentraría totalmente en un punto y, sin embargo, el estimador sería totalmente inútil.

Otra propiedad deseable es la *consistencia*. Esta propiedad hace referencia a los cambios en la distribución muestral a medida que aumenta el tamaño de la muestra. Se dice que un estimador es consistente si su distribución muestral tiende a concentrarse alrededor del verdadero valor del parámetro cuando el tamaño de la muestra tiende a infinito. La figura 1-4 presenta las distribuciones muestrales de un estimador consistente, correspondientes a distintos tamaños

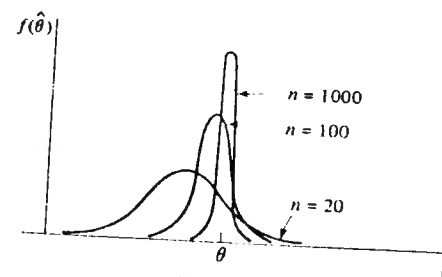


Figura 1-4

de la muestra. A medida que nos movemos desde una muestra de pequeño tamaño hacia una de menor tamaño, suceden dos cosas: a) el sesgo se hace más pequeño, y b) las estimaciones están menos dispersas. La consistencia es una propiedad importante porque nos garantiza que las estimaciones mejoran al ampliar el tamaño de la muestra. En este caso, siempre que sea posible aumentar el tamaño de la muestra, podremos adquirir un mayor grado de fiabilidad gastando más en muestreo. Incluso las observaciones de la contabilidad nacional pueden hacerse más numerosas si se calculan los datos para periodos de tiempo más cortos.

La estimación de un parámetro a partir de una muestra se puede comparar al tiro al blanco con fusil. En este paralelismo, el centro de la diana representa el verdadero valor del parámetro. Cada disparo representa una estimación (muestra) concreta, el fusil es el estimador (es decir, la fórmula de estimación), y la distancia hasta el blanco representa el tamaño de la muestra. En la práctica, normalmente sólo se dispone de una muestra y, por tanto, se puede efectuar una sola estimación; es decir —en este paralelismo— únicamente se nos permite efectuar un disparo. Sin embargo, a partir de la calidad del fusil, se puede juzgar la calidad de cualquier disparo antes de efectuarlo. El fusil se puede juzgar según sus resultados reales, es decir, efectuando un gran número de disparos con él, o bien examinando su estructura, el tipo de material utilizado, etcétera. El primer método equivale a la obtención empírica de las propiedades de un estimador, mientras que el segundo corresponde a su obtención teórica. Un fusil insesgado es aquel que realiza disparos cuyos impactos se distribuyen de forma aleatoria alrededor del centro de la diana. Si comparamos todos los fusiles insesgados, podemos considerar eficiente aquel cuyos impactos se hallan más concentrados alrededor del centro del blanco. Por último, podemos considerar que un fusil es consistente si la probabilidad de que el impacto del disparo esté a una cierta distancia (pequeña) del centro del blanco aumenta a medida que disminuye la distancia entre el tirador y el blanco. Obsérvese que la calidad de un fusil se evalúa a partir de su comportamiento repetido (real o esperado) y no por un simple disparo. Dando un único disparo, puede ocu-

rrir que un fusil de poca precisión efectúe un impacto directo en el centro del blanco, mientras que un fusil claramente superior y de alta precisión puede no lograrlo. Claramente, esto no alterará nuestras conclusiones acerca de las calidades respectivas de los dos fusiles a menos que tienda a ocurrir repetidamente.

Hasta aquí, no hemos considerado el problema de la construcción de los estimadores. Como se ha señalado anteriormente, un estimador es una fórmula para generar estimaciones. Considérese, por ejemplo, la media de la muestra como estimador de la media de la población; en este caso, la fórmula requiere que tomemos todos los valores de la variable observados en la muestra, los sumemos y los dividamos por el número de observaciones. Una muestra concreta dará como resultado una estimación concreta. En esta ocasión, se ha elegido la media de la muestra como estimador de la media de la población, más o menos debido a que ello parece intuitivamente plausible, y éste es, en realidad, uno de los medios para la obtención de estimadores —es decir, pensar en una idea plausible y probar con ella—, lo cual, en este contexto, significa obtener las propiedades de estos estimadores. Otro procedimiento consiste en diseñar un estimador, desarrollando una fórmula que satisfaga ciertas condiciones, y asegurando así que se cumplan al menos algunas de las propiedades que resultan deseables. Por último, se pueden utilizar algunos principios que, si bien no garantizan directamente la obtención de las propiedades deseadas, resultan sin embargo atractivos por otros motivos. En la sección 6.2 se analizarán detalladamente.

1.5 Obtención de distribuciones muestrales

El fin principal del análisis anterior consiste en explicar la crucial importancia de las distribuciones muestrales en la inferencia estadística. El problema que aparece a continuación es el de obtener las distribuciones muestrales de ciertos estimadores, lo cual, en general, puede hacerse experimental o teóricamente. La *obtención experimental* se basa en la simulación: se crea una población determinada (que, por tanto, se conoce totalmente) y se extrae de ella un gran número de muestras aleatorias. Estas muestras nos permiten construir una distribución muestral aproximada del estimador que queremos examinar. El resultado que obtenemos es un resultado específico, en el sentido de que sólo tiene aplicación respecto a la población concreta (caracterizada por unos parámetros determinados) que estamos examinando, pero normalmente confiamos en que estos resultados se pueden generalizar, por lo menos en cierta medida y estas generalizaciones se pueden contrastar a través de nuevos experimentos con poblaciones cuyos parámetros tengan valores distintos. La *obtención teórica* de distribuciones muestrales utiliza la teoría de la probabilidad; en lugar de extraer físicamente un gran número de muestras, como se hace en el enfoque experimental, es posible conocer lo que ocurriría sin tener que hacerlo practi-

camente y, por tanto, la teoría nos permite ahorrar una gran cantidad de trabajo, lo cual es una gran ventaja. Otra ventaja de la teoría es que es más precisa: en los experimentos no podemos producir nunca un número infinito de muestras, pero en la teoría se puede manejar el concepto de infinito con gran facilidad. Y, por último, pero no menos importante, contrariamente a lo que ocurre con el método experimental, los resultados de la obtención teórica son completamente generales, por lo menos dentro de unos límites bien definidos. Por tanto, la obtención teórica de distribuciones muestrales es claramente superior a la obtención experimental; el único problema es que no siempre somos capaces de trabajar con ella. En algunas ocasiones los problemas son tan complicados que nuestro conocimiento resulta simplemente insuficiente. Tal ha sido el caso, en especial, con los problemas de estimación en la econometría moderna, como se demostrará al estudiar las relaciones económicas simultáneas. En el próximo capítulo utilizaremos el enfoque experimental para obtener las distribuciones muestrales de dos estimadores de uso corriente. En este caso, el motivo para hacerlo así no es la dificultad para poner en práctica las leyes de la probabilidad, sino que nuestra intención es familiarizar al lector en la mayor medida posible con el concepto de distribución muestral. Confiamos en que proporcionando un contenido físico a la teoría, a través de los experimentos, la idea quedará mucho más clara. También demostraremos el enfoque teórico del mismo problema, comparándolo con el enfoque experimental; sin embargo, antes de intentar la obtención teórica, se deben introducir los instrumentos básicos proporcionados por la teoría de la probabilidad.

EJERCICIOS:

En la exposición de los principios básicos de la inferencia estadística hemos supuesto que el lector posee un conocimiento adecuado del álgebra elemental. Las preguntas cortas de los ejercicios 1-1 a 1-8 están destinadas a comprobar la validez de tal supuesto.

1-1. Representar gráficamente las funciones siguientes:

$$a. y = x + 1. \quad b. y = \sqrt{x}. \quad c. \log y = \frac{1}{2} \log x.$$

1-2. Hallar x e y :

$$ax + by = b^2 \quad y \quad x + y^2 = a.$$

1-3. Simplificar las expresiones siguientes:

$$a. \frac{x^5}{x^2}, \quad b. \sqrt[3]{x^6}, \quad c. \frac{(x/y) - (y/x)}{(x-y)/(x+y)}.$$

1-4. Hallar $\sum_{i=1}^n x_i$ en los casos siguientes:

$$a. x_i = i. \quad b. x_i = 1 + 2i. \quad c. x_i = 2^{i-1}.$$

1-5. Resolver la ecuación: $x^2 - 5x + 6 = 0$.

1-6. Hallar el valor máximo de y sabiendo que:

$$y = 1 + 2x - x^2.$$

1-7. Obtener la primera y segunda derivadas de y en las expresiones siguientes:

$$\text{a. } y = (x + a)^2. \quad \text{b. } y = \sqrt{x}. \quad \text{c. } y = \frac{a - x}{b - x}$$

1-8. Suponiendo que $x > 0$ e $y > 0$, probar que

$$(x/y + y/x) \geq 2.$$

A lo largo de la exposición se ha supuesto también que el lector está familiarizado con los conceptos básicos de la estadística descriptiva. En los ejercicios 1-9 a 1-12 se deberán poner en práctica estos conocimientos. (Recuérdese que en el Apéndice A se proporcionan las nociones básicas del álgebra de sumatorios.)

1-9. La suma de diez números, $x_1, x_2, \dots, x_{10}$ es 60, y la suma de sus cuadrados es 396. Hallar los valores siguientes:

a. Su media aritmética, $\bar{x}$.

b. La desviación standard (s) de x (utilizando como denominador 10).

$$\text{c. } \sum_{i=1}^{10} (x_i - \bar{x})/s. \quad \text{d. } \sum_{i=1}^{10} 2(x_i - 5).$$

e. La media aritmética de z_i sabiendo que $z_i = x_i - (x_i - \bar{x})/s$.

1-10. Trazar un diagrama correspondiente a cada una de las siguientes distribuciones de frecuencias:

a. La frecuencia es la misma para todos los valores de las variables.

b. La variable toma el mismo valor en cada observación.

c. La distancia entre la primera cuartila y la mediana es el doble de la distancia entre esta última y la tercera cuartila.

d. El valor de la desviación standard es cero.

e. El valor de la media aritmética es inferior al de la moda.

1-11. Supóngase que queremos determinar la edad media de una población de individuos determinada. Supóngase, además, que una cuarta parte de los individuos indican en el censo un año menos de su verdadera edad, la mitad de los individuos proporcionan su edad correcta, y una cuarta parte añaden dos años a su verdadera edad. Determinar la relación entre la edad media correcta y la edad media obtenida a partir del censo.

1-12. Las distribuciones de frecuencias en términos percentuales de los salarios mensuales de los obreros industriales en la Unión Soviética durante los años 1929 y 1934, es decir antes y después de la reforma salarial de 1931-1933, se presentan en el cuadro 1-3¹. A este respecto, se ha afirmado que

«los salarios correspondientes a los grupos de menores ingresos experimentaron los mayores aumentos, y los menores aumentos correspondieron a los grupos de ingresos medios. A pesar de que se redujo un poco la diferencia entre los salarios más bajos y los salarios medios, aumentó la diferencia entre los salarios de los trabajadores de ingresos medios y los de ingresos elevados.»¹

Compruébese si los datos apoyan esta afirmación.

Cuadro 1-3

Salarios mensuales en rublos	Marzo 1929	Octubre 1934
Menos de 30	2.3	0.0
30 a 49	15.2	0.8
50 a 69	23.6	2.6
70 a 89	21.1	7.9
90 a 109	15.0	12.7
110 a 129	9.4	14.3
130 a 149	5.8	13.3
150 a 169	1.8	10.6
170 a 199	4.1	11.9
200 a 249	1.3	12.0
250 a 299	0.4	6.6
300 y más	0.0	7.3
	100.0	100.0

1. A. Aganbegian, «Methods of Analysing and Calculating the Distribution of Workers and Employees by the Amount of Wages», *Problems of Economics*, vol. 3, octubre 1960, p. 24.

Obtención experimental de distribuciones muestrales

La distribución muestral de un estimador se puede considerar como una distribución de frecuencias relativas de los valores del estimador, obtenidas a partir de un número infinito de muestras aleatorias, todas del mismo tamaño, y procedentes de la misma población. Experimentalmente, lo podemos hacer de la forma siguiente: en primer lugar, debemos crear la población dotándola de unas características determinadas (parámetros). A continuación, elegimos el parámetro que queremos estimar y la fórmula para estimarlo a partir de la información proporcionada por la muestra. Después se extraen un gran número de muestras aleatorias de igual tamaño, calculando el valor del estimador para cada muestra. Por último, se ordenan los resultados en forma de una distribución de frecuencias relativas. Esto nos dará una aproximación de la distribución muestral del estimador elegido, y decimos «aproximación» porque sólo tenemos un número finito de muestras (aunque sea grande), mientras que una distribución muestral auténtica está basada en un número infinito de muestras.

El problema de crear o simular una población es generalmente muy sencillo. Por otra parte, algunas poblaciones no deben existir por fuerza físicamente para que podamos extraer muestras de ellas. Un ejemplo de este caso lo constituye la población de los posibles resultados en cualquier mecanismo de azar, como por ejemplo la población de los resultados al lanzar una moneda al aire. Supongamos que queremos estimar la probabilidad de obtener cara, y usamos como estimador la proporción de caras obtenidas en 30 lanzamientos. En este caso, podemos tomar una moneda con una probabilidad dada de obtener cara (por ejemplo, una moneda perfecta, en cuyo caso esta probabilidad es una mitad), lanzarla treinta veces y anotar el resultado. Si repetimos esta operación un gran número de veces, podremos construir una aproximación razonable de la distribución muestral de la proporción de caras obtenidas en muestras de tamaño 30. Es posible que para otras poblaciones sea necesario que exista una representación material de cada unidad antes de que podamos extraer muestras. Por ejemplo, la población de las familias de los Estados Unidos se puede representar mediante fichas que contengan la información adecuada, con

una ficha para cada familia. En este caso, una muestra aleatoria de familias vendría dada por una muestra aleatoria de las fichas.

En este capítulo se estudia la obtención experimental de distribuciones muestrales en dos casos simples. En el caso A consideramos las muestras de atributos. En particular, estamos interesados en estimar la proporción de individuos (o resultados, objetos, etc.) que poseen un atributo determinado, usando como estimador la proporción obtenida en la muestra. En tal caso, nuestra tarea consiste en obtener la distribución muestral de este estimador. Para que se vea más claro, podemos imaginar el problema del estimar la proporción de individuos que toman café en una población determinada. En el caso B nos ocuparemos del muestreo correspondiente a una variable (no binaria) y nos interesará obtener la distribución muestral de la media de la muestra como estimador de la media de la población. A modo de ilustración, podemos pensar en una variable que indique el número de visitas al dentista por adulto y año, considerando el problema de estimar el número medio de visitas al dentista por persona, correspondiente a la población.

Los aspectos mecánicos del experimento de muestreo que debemos realizar son los mismos en ambos casos. Tenemos una urna que contiene una gran cantidad de bolas con anotaciones distintas, que representan las unidades de la población. Se agita la urna y se escogen al azar un número de bolas igual al tamaño deseable de la muestra. Se anota el resultado y se devuelven las bolas a la urna, repitiéndose la operación gran número de veces. Como después de cada extracción las bolas se reintegran a la urna y ésta se agita antes de extraer otra muestra, cada muestra tiene la misma probabilidad de ser elegida. Sin embargo, existe una pequeña complicación porque nuestra población, aunque sea grande, es finita y, además, las bolas no se devuelven a la urna hasta después de formar totalmente la muestra. Esto significa que la segunda bola tiene una posibilidad de ser elegida ligeramente mayor que la primera, la tercera una posibilidad mayor que la segunda, etc., ya que, al ir disminuyendo el número de bolas en la urna, aumenta la posibilidad de ser elegida para cada una de las que quedan. Pero si la población es grande, el cambio en las probabilidades es tan pequeño que puede dejarse de lado. Debe señalarse que sólo se puede simular una población, representando las unidades a través de bolas o cualquier otro objeto, cuando la variable no es continua. Si la población estuviese formada por todos los valores posibles pertenecientes a un intervalo determinado, no se podría representarla por un conjunto de objetos de carácter discreto y, en tal caso, tendríamos que usar algún otro medio de simulación. Un artefacto que daría buen resultado, y permitiría conseguir una simulación razonable de las variables cuyos valores estén en un intervalo finito, sería un disco con una aguja girando libremente en torno a su centro. Existen también otros métodos, y cabe destacar los basados en calculadoras electrónicas, pero en los casos que se estudian, aquí no se plantea este problema ya que no utilizamos variables continuas.

2.1 Distribución muestral de la proporción de aciertos en la muestra

Consideremos la proporción de aciertos (por ejemplo, la proporción de bebedores de café) de la muestra como un estimador de la proporción de aciertos en la población. Queremos obtener la distribución muestral de este estimador mediante sucesivos muestreos. La población es una urna conteniendo gran número de bolas idénticas en todos los aspectos excepto el color, y en esta población concreta el 70 % de las bolas son rojas mientras que el 30 % son blancas. Podemos considerar que las bolas rojas representan los aciertos (los bebedores de café) y las bolas blancas representan los fallos. Realicemos dos experimentos con esta población: en el primero de ellos extraemos 100 muestras de tamaño 4, y en el segundo, 100 muestras de tamaño 16. Obsérvese que en cada caso utilizamos 100 muestras como aproximación de un número infinito de ellas, y, por lo tanto, los resultados contendrán errores de aproximación. Si lo descáramos, podríamos reducir estos errores ampliando el número de muestras, pero, para nuestros objetivos, el grado de aproximación proporcionado por las 100 muestras resulta suficiente. Para describir los resultados utilizaremos la notación siguiente:

- π = proporción de aciertos en la población;
- $\hat{\pi}$ = proporción de aciertos en la muestra (un estimador de π);
- f = frecuencia relativa;
- n = tamaño de la muestra.

En los experimentos, $\hat{\pi}$ tomará 100 valores, es decir, uno para cada muestra. Obsérvese que en la población, $\pi = 0,7$.

EXPERIMENTO A.1: 100 muestras de tamaño 4. En este caso, $n = 4$, de modo que cada muestra sólo puede dar uno de entre 5 resultados posibles (ningún acierto, un acierto, etc.). Estos resultados aparecen en el cuadro 2-1 y la figura 2-1. La

Cuadro 2-1

Número	Aciertos Proporción: $\hat{\pi}$	Frecuencia	
		Absoluta *	Relativa: f
0	0.00	1	0.01
1	0.25	6	0.06
2	0.50	28	0.28
3	0.75	42	0.42
4	1.00	23	0.23
		100	1.00

última columna del cuadro 2-1 indica las frecuencias que proporcionan una aproximación a la distribución muestral que pretendemos hallar. Las características principales de esta distribución muestral son

$$\text{Media} = \sum_{i=0}^4 f_i \hat{\pi}_i = 0.01 \times 0 + 0.06 \times 0.25 + 0.28 \times 0.50 + 0.42 \times 0.75 + 0.23 \times 1.00 = 0.700.$$

$$\text{Desviación standard} = \sqrt{\sum_{i=0}^4 f_i (\hat{\pi}_i - 0.7)^2} = 0.2233.$$

(Obrévese que, en este caso, $\hat{\pi}_0 = 0$, $\hat{\pi}_1 = 0.25$, $\hat{\pi}_2 = 0.50$, etc., y f_0 , f_1 , f_2 , etc., son las frecuencias relativas correspondientes.) El examen de la distribución muestral obtenida indica que si hubiéramos utilizado la proporción de aciertos de la muestra de tamaño 4 como estimación de la proporción de aciertos de la población hubiésemos subestimado seriamente esta última en el 35 % de los casos ($0.01 + 0.06 + 0.28 = 0.35$), la hubiésemos sobreestimado en el 23 % de los casos, y hubiésemos estado muy cerca del verdadero valor un 42 % de las veces. Estos porcentajes son, naturalmente, aproximados porque nuestra experiencia se reduce a 100 muestras.

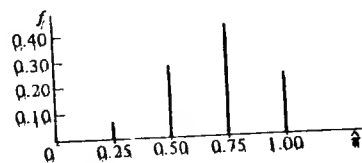


Figura 2-1

EXPERIMENTO A.2: 100 muestras de tamaño 16. En este caso, $n = 16$ y, por tanto, existen 17 resultados posibles. Estos resultados aparecen en el cuadro 2-2 y la figura 2-2. Las características principales de la distribución muestral son

$$\text{Media} = \sum_{i=0}^{16} f_i \hat{\pi}_i = 0.7006.$$

$$\text{Desviación standard} = \sqrt{\sum_{i=0}^{16} f_i (\hat{\pi}_i - 0.7006)^2} = 0.1191.$$

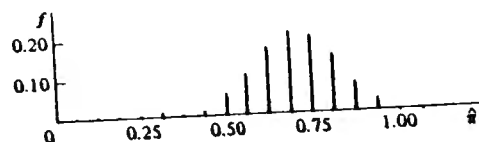


Figura 2-2

(Obrévese que, en este caso, $\hat{\pi}_0 = 0$, $\hat{\pi}_1 = 1/16$, $\hat{\pi}_2 = 2/16$, etc., y f_0 , f_1 , f_2 , etc., son las frecuencias relativas correspondientes.) La distribución muestral obtenida indica una elevada concentración de las estimaciones alrededor del valor correcto: el 95 % de las estimaciones se hallan en el intervalo comprendido entre 0.5 y 0.9, y un alto porcentaje está muy próximo al parámetro de la población. Contrariamente a lo que ocurría en el experimento anterior, la muestra no sugiere en ningún caso que la población está compuesta totalmente de aciertos o de errores.

Los principales resultados de estos experimentos se pueden resumir en los siguientes puntos: I) en ambos casos, la media de la distribución muestral resulta ser prácticamente igual al valor del parámetro de la población; II) la dispersión de la distribución muestral correspondiente a las muestras de tamaño 16 es inferior a la de la distribución correspondiente a las muestras de tamaño 4, y la desviación standard de la primera es aproximadamente la mitad de la de la segunda; y III) la distribución muestral correspondiente a las muestras de tamaño 16 es considerablemente más simétrica que la de las muestras de tamaño 4.

Cuadro 2-2

Número	Aciertos Proporción: $\hat{\pi}$	Frecuencia	
		Absoluta	Relativa: f
0	0.0000	0	0.00
1	0.0625	0	0.00
2	0.1250	0	0.00
3	0.1875	0	0.00
4	0.2500	0	0.00
5	0.3125	1	0.01
6	0.3750	0	0.00
7	0.4375	1	0.01
8	0.5000	5	0.05
9	0.5625	10	0.10
10	0.6250	17	0.17
11	0.6875	21	0.21
12	0.7500	20	0.20
13	0.8125	15	0.15
14	0.8750	7	0.07
15	0.9375	3	0.03
16	1.0000	0	0.00
		100	1.00

Estos resultados se han obtenido a partir de repetidas muestras procedentes de una población dicotómica (es decir, que sólo contiene dos tipos de individuos) cuya proporción de aciertos es 0,7. Únicamente se han obtenido dos distribuciones muestrales, correspondientes a las muestras de tamaño 4 y de tamaño 16, pero, a pesar del carácter específico de los experimentos, los resultados permiten claramente realizar algunas generalizaciones acerca de la distribución muestral de la proporción de aciertos observados en una muestra ($\hat{\pi}$) como estimador de la proporción de aciertos de la población (π). Estas generalizaciones son las siguientes:

1. $\hat{\pi}$ es un *estimador insesgado* de π (es decir, la media de la distribución muestral de $\hat{\pi}$ es igual al parámetro de la población π).
2. A medida que aumenta el tamaño de la muestra, la distribución muestral de $\hat{\pi}$ se concentra cada vez más alrededor de π . Ello implica que $\hat{\pi}$ es un *estimador consistente* de π .
3. Las distribuciones muestrales de $\hat{\pi}$ basadas en muestras grandes tienden a ser más simétricas que las basadas en muestras pequeñas.

La cuarta generalización resulta menos obvia y presenta un mayor grado de incertidumbre que las otras tres, estando relacionada con la generalización 2; como la dispersión (o, su contraria, la concentración) se mide a través de la desviación standard, es de esperar que exista cierta relación entre la variación en el tamaño de la muestra y la variación de la desviación standard. Si observamos que, al aumentar cuatro veces el tamaño de la muestra, la desviación standard se ha reducido a la mitad, podemos pensar que se cumple también la proposición siguiente:

4. La desviación standard de la distribución muestral de $\hat{\pi}$ es inversamente proporcional a la raíz cuadrada del tamaño de la muestra.

Esta última generalización resulta más arriesgada que las otras tres. Por una parte, la desviación standard correspondiente a las muestras de tamaño 16 sólo es aproximadamente igual a la mitad de la de las muestras de tamaño 4; por otra, sólo podemos generalizar a partir de dos tamaños muestrales (es decir, de una sola proporción). La primera dificultad se puede eliminar utilizando más de 100 muestras y aumentando así la exactitud, y la segunda realizando un mayor número de experimentos.

Quizá se haya observado que nuestro experimento muestral no nos permite efectuar ninguna generalización acerca de la eficiencia de $\hat{\pi}$. Aunque, desde luego, sería interesante ver si existe algún otro estimador insesgado cuya variancia sea menor que la de $\hat{\pi}$, nuestros experimentos no pueden dar respuesta a esta pregunta, ya que sólo se considera un estimador y, por tanto, no se pueden efectuar comparaciones; pero, aunque se considerasen varios estima-

dores alternativos en lugar de uno solo, en el mejor de los casos los experimentos muestrales resolverían únicamente el problema de la eficiencia en relación con la pequeña clase de los estimadores utilizados, y no respecto a todos los estimadores (insesgados) como sería de desear.

Tenemos que confiar en que las generalizaciones anteriores se cumplen para todos los valores del parámetro de la población π , y para todos los tamaños muestrales n . En el capítulo 4 descubriremos si ello ocurre o no, al obtener la distribución muestral mediante un procedimiento teórico.

2.2 Distribución muestral de la media de la muestra

Consideremos ahora el problema de obtener la distribución muestral de la media de la muestra como estimador de la media de la población. La población que estamos simulando es, por ejemplo, la de todos los individuos adultos de una cierta región geográfica, y cada uno de ellos se caracteriza por el número de visitas al dentista en un año determinado. La variable X representa, pues, el número de visitas al dentista por individuo adulto en un año. A fin de simplificar la construcción de la población, supondremos que X sólo puede tomar los valores 0, 1, 2, ..., 9, y que cada uno de estos valores se observa con igual probabilidad. Una distribución de este tipo recibe el nombre de *distribución uniforme discreta*. La población se simula mediante una urna que contiene un gran número de bolas idénticas en todos los sentidos excepto por el número que llevan impresas. Estos números son 0, 1, 2, ..., 9, y la urna contiene una cantidad igual de bolas con cada uno de estos números. En el cuadro 2-3 se proporciona una descripción más elaborada de esta población.

Cuadro 2-3

Valores de X	Frecuencia relativa en la población
0	0.1
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
	1.0

Media de la población:
 $\mu = 4.5$
 Desviación standard de la población:
 $\sigma = 2.8723$.

Al igual que en el caso A, realizaremos de nuevo dos experimentos muestrales. En el primero de ellos, extraeremos 100 muestras de tamaño 5 y obtendremos la distribución muestral de la media de la muestra. Para obtener alguna idea respecto a su eficiencia, obtendremos también la distribución muestral de un estimador alternativo, en este caso la mediana de la muestra. Por tanto, cada muestra se utilizará para producir dos estimaciones de la media de la población.

En el segundo experimento, se extraerán 100 muestras de tamaño 10, obteniendo una distribución de la media de la muestra. Al igual que en los experimentos del caso A, nos daremos por muy satisfechos con la aproximación a infinito proporcionada por las 100 muestras. Para describir los resultados se utilizará la siguiente notación:

- μ = media de la población;
- σ = desviación standard de la población;
- $\bar{X}$ = media de la muestra;
- $\tilde{X}$ = mediana de la muestra;
- f = frecuencia relativa;
- n = tamaño de la muestra.

EXPERIMENTO B.1: 100 muestras de tamaño 5. Consideremos en primer lugar la distribución muestral de la media de la muestra, que constituye nuestro máximo interés. Lógicamente, aunque la variable (y , por tanto, las observaciones individuales) solo puede tomar valores enteros de 0 a 9, las medias de las muestras no serán generalmente enteros y, por tanto, la distribución muestral será una dis-

Cuadro 2-4

Valores de la media de la muestra: $\bar{X}$		Frecuencia	
Intervalo	Punto medio	Absoluta	Relativa: f
0,5 a 1,499	1	1	0,01
1,5 a 2,499	2	5	0,05
2,5 a 3,499	3	12	0,12
3,5 a 4,499	4	31	0,31
4,5 a 5,499	5	28	0,28
5,5 a 6,499	6	15	0,15
6,5 a 7,499	7	5	0,05
7,5 a 8,499	8	3	0,03
8,5 a 9,499	9	0	0,00
		100	1,00

tribución de frecuencias cuyas clases estarán definidas por intervalos y no por puntos. Naturalmente, para representar cada clase podemos elegir un valor, como por ejemplo el centro de cada intervalo. La distribución obtenida como resultado del experimento aparece en el cuadro 2-4 y se ilustra mediante la figura 2-3.

Las principales características de esta distribución son

$$\text{Media} = \sum_{i=1}^9 f_i \bar{x}_i = 4.60.$$

$$\text{Desviación standard} = \sqrt{\sum_{i=1}^9 f_i (\bar{x}_i - 4.60)^2} = 1.3638.$$

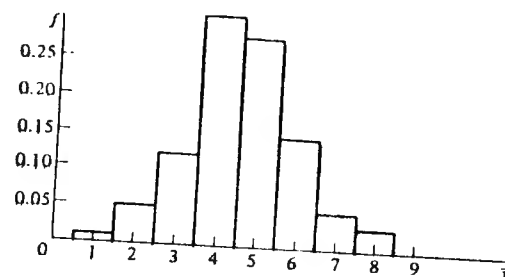


Figura 2-3

Los resultados indican que un 59 % de los valores estimados ($0.31 + 0.28 = 0.59$) están comprendidos en el intervalo ± 1 alrededor del valor real 4,5, y que el 86 % de las estimaciones están comprendidas en el intervalo ± 2 alrededor de dicho valor.

A continuación se presenta la distribución muestral de la mediana de la muestra. Como el tamaño de la muestra es un número impar, la mediana será siempre un entero. La distribución aparece en el cuadro 2-5. Las características principales de esta distribución son

$$\text{Media} = \sum_{i=0}^9 f_i \bar{x}_i = 4.68.$$

$$\text{Desviación standard} = \sqrt{\sum_{i=0}^9 f_i (\bar{x}_i - 4.68)^2} = 1.8758.$$

Cuadro 2-5

Valores de la mediana de la muestra: $\tilde{x}$	Frecuencia	
	Absoluta	Relativa: f
0	1	0.01
1	4	0.04
2	4	0.04
3	19	0.19
4	23	0.23
5	14	0.14
6	17	0.17
7	11	0.11
8	5	0.05
9	2	0.02
	100	1.00

La distribución se representa gráficamente en la figura 2-4. A fin de facilitar la comparación con la distribución muestral de la media de la muestra, en esta figura se reproduce la distribución de la figura 2-3 mediante la línea de puntos.

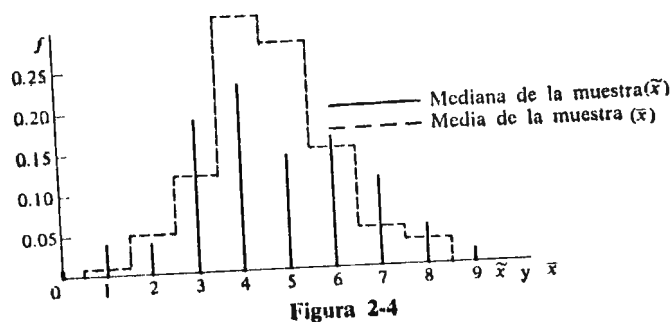


Figura 2-4

A primera vista resulta evidente que las dos distribuciones son muy distintas, y que la de $\bar{x}$ es mucho menos regular y considerablemente más dispersa que la de $\tilde{x}$.

EXPERIMENTO B.2: 100 muestras de tamaño 10. Los resultados de este experimento se resumen en el cuadro 2-6 y la figura 2-5. Las principales características de esta distribución son

$$\text{Media} = \sum_{i=1}^7 f_i \bar{x}_i = 4.57,$$

$$\text{Desviación standard} = \sqrt{\sum_{i=1}^7 f_i (\bar{x}_i - 4.57)^2} = 1.0416.$$

Cuadro 2-6

Valores de la media de la muestra: $\bar{x}$		Frecuencia	
Intervalo	Punto medio	Absoluta	Relativa: f
0.5 a 1.499	1	0	0.00
1.5 a 2.499	2	1	0.01
2.5 a 3.499	3	14	0.14
3.5 a 4.499	4	34	0.34
4.5 a 5.499	5	32	0.32
5.5 a 6.499	6	16	0.16
6.5 a 7.499	7	3	0.03
7.5 a 8.499	8	0	0.00
8.5 a 9.499	9	0	0.00
		100	1.00

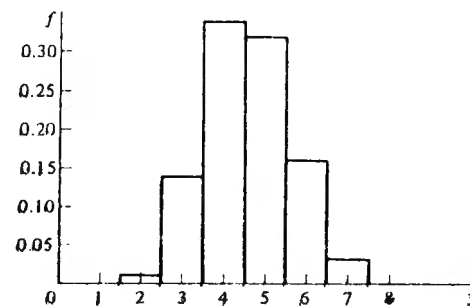


Figura 2-5

En el caso B, los resultados de nuestros experimentos muestrales se pueden resumir de la forma siguiente: I) la media de la distribución muestral de la media de la muestra es aproximadamente igual a la media de la población para los dos tamaños muestrales considerados; II) la dispersión de la distribución muestral de la media de la muestra de tamaño 10 es inferior a la de las muestras de tamaño 5,

y la variancia de esta última ($1,3638^2$) es casi el doble de la de la primera ($1,0416^2$); y III) la media de la distribución muestral de la mediana de la muestra también es aproximadamente igual a la media de la población, pero para muestras del mismo tamaño su variación es mayor.

Estos resultados, aunque se basan en un caso específico, permiten claramente efectuar algunas generalizaciones:

1. La media de la muestra, $\bar{X}$, es un *estimador insesgado* de la media de la población, μ .
2. A medida que aumenta el tamaño de la muestra, la distribución muestral de $\bar{X}$ se va concentrando cada vez más en torno a la media de la población, y, por tanto, $\bar{X}$ es un *estimador consistente* de μ .
3. La mediana de la muestra $\bar{X}$ es un estimador insesgado de μ , pero su variancia es mayor que la de $\bar{X}$ y, por tanto, $\bar{X}$ es un estimador ineficiente de μ .

Estas generalizaciones se derivan de forma inmediata de los resultados principales de nuestros experimentos, pero la generalización que se efectúa a continuación resulta menos obvia y algo más arriesgada.

4. La desviación standard de la distribución muestral de $\bar{X}$ es inversamente proporcional a la raíz cuadrada del tamaño de la muestra.

La generalización 4 sólo se ve aproximadamente confirmada por nuestro experimento, ya que para ello debemos considerar que $1,3638/1,0416 = 1,3093$ es «aproximadamente igual» a $\sqrt{2} = 1,4142$.

Cuando estudiemos la obtención teórica de las distribuciones muestrales, consideraremos nuevamente estas cuatro generalizaciones y, de momento, las mantendremos de forma provisional, confiando en poder confirmarlas o rechazarlas más adelante, pero antes debemos dominar los instrumentos básicos de la teoría de la probabilidad y éste es el tema del capítulo 3.

EJERCICIOS:

2-1. Utilizando 100 muestras, construir distribuciones muestrales experimentales de la proporción de caras obtenidas en muestras de 4 y 16 lanzamientos de una moneda, respectivamente. Calcular la media y la desviación standard de ambas distribuciones y comparar los resultados con los obtenidos en el caso A de este capítulo.

2-2. Con una población consistente en los números 1, 2 y 3, en iguales proporciones,

- a. Construir todas las muestras posibles de tamaño 3, observando que la frecuencia relativa de cada muestra es la misma. Escribir la distribución resultante de la media de la muestra y la de la mediana. Calcular las medias y las desviaciones standard de las dos distribuciones. (NOTA: de esta población se pueden extraer 27 muestras distintas de tamaño 3.)
 - b. Efectuar las mismas operaciones para las muestras de tamaño 4. Comparar la razón entre la desviación standard de la media y la de la mediana con la de la media y la mediana obtenidas en el apartado A.
- 2-3.** Considérese que, si al lanzar tres monedas iguales se obtienen dos caras y una cruz, se obtiene un «acierto». Efectuar 10 lanzamientos de este tipo y anotar el número de aciertos. Repetir esta misma operación 100 veces, efectuando 10 lanzamientos cada vez y anotando el número de aciertos.
- a. Presentar la distribución muestral resultante.
 - b. Calcular la media y la desviación standard de la proporción de aciertos.
- 2-4.** Utilizando la tabla de desviaciones normales aleatorias del Apéndice D, construir una distribución muestral empírica de la media de la muestra para muestras de tamaño 5 extraídas de una población normal.

Probabilidad y distribuciones de probabilidad

Pese a que el concepto de probabilidad se utiliza muy a menudo en muchas de las ramas de la ciencia y en la vida cotidiana, se trata de un concepto muy difícil de definir, y acerca del cual existe gran controversia. Vamos a hacer mención de los puntos de vista más importantes y a indicar la naturaleza de dichas dificultades. Según el enfoque denominado *clásico*, la probabilidad de un resultado favorable viene dada por el cociente f/n , en el que n es el número de resultados posibles, mutuamente excluyentes e igualmente posibles, mientras que f es el número de resultados que se consideran favorables. Dos resultados son mutuamente excluyentes si la aparición de uno de ellos elimina la posibilidad de que ocurra el otro; por ejemplo, si al lanzar una moneda al aire se obtiene una cara, no se puede obtener una cruz en el mismo lanzamiento. Además, los resultados son «igualmente posibles» si se espera de antemano que, a largo plazo, cada resultado ocurrirá con la misma frecuencia. Por tanto, al lanzar un dado de seis caras (sin trazar), la probabilidad de obtener un 3 es $1/6$ ya que existen seis resultados posibles y, suponiendo que el dado no esté trucado, todos ellos son igualmente probables. Para ver otro ejemplo, considérese la probabilidad de obtener dos caras al lanzar dos monedas insesgadas. Los resultados posibles son

1.ª moneda	2.ª moneda
Cara	Cara
Cara	Cruz
Cruz	Cara
Cruz	Cruz

Por tanto, existen cuatro (no tres) resultados posibles e igualmente probables, y la probabilidad de obtener dos caras es, pues, $1/4$. Obsérvese que esta definición de probabilidad se puede adaptar también muy fácilmente a los casos continuos. Por ejemplo, considérese la esfera de un reloj (sin el mecanismo) con una de las agujas, que puede girar libremente. En este caso, si hacemos girar la aguja, ésta puede pararse en un número infinito de puntos. Como los

puntos no tienen ninguna dimensión y, por tanto, no tienen longitud, la probabilidad de que la aguja se detenga en un punto concreto es cero. Sin embargo, contrariamente a los puntos, cualquier intervalo posee una longitud distinta de cero y, en consecuencia, la probabilidad de que la aguja se detenga en el interior de cualquier intervalo es positiva y se puede calcular. La definición de probabilidad mencionada anteriormente se puede reformular en términos de longitudes (o volúmenes, etc.) en lugar de números; por ejemplo, en el caso que acabamos de mencionar, la probabilidad de que la aguja se detenga entre el minuto 10 y el minuto 12 es $1/30$.

El uso de la definición clásica de probabilidad plantea dos importantes dificultades. La primera de ellas tiene su origen en el hecho de que la definición depende de modo crucial del supuesto de que todos los resultados son igualmente probables. Si se nos preguntara cuál es la probabilidad de obtener una cara cuando la moneda está sesgada de un modo no especificado, no seríamos capaces de dar una respuesta. La segunda dificultad es de tipo similar. Existen algunos sucesos para los cuales resulta imposible —a partir del conocimiento actual— obtener las probabilidades a priori, como por ejemplo, las tablas de mortalidad, las tasas de participación en la fuerza de trabajo, las variaciones de la renta, y muchas otras. Ambas dificultades se deben al hecho de que la definición clásica se basa en el análisis a priori y las probabilidades obtenidas a través de la definición clásica se llaman a veces «probabilidades a priori» para resaltar su naturaleza teórica.

Es posible evitar las dificultades asociadas al enfoque clásico si utilizamos el concepto *objetivo* o de *frecuencia* de la probabilidad. Este enfoque representa un desarrollo más reciente en la teoría de la probabilidad y define a las probabilidades como los límites de las frecuencias relativas, cuando el número de observaciones tiende a infinito. Como aproximación de las probabilidades se pueden usar las frecuencias relativas correspondientes a un gran número de pruebas, y si lanzamos muchas veces una moneda insesgada observaremos que la proporción de caras (es decir, la frecuencia relativa) tiende a ser estable y próxima a $1/2$. En este enfoque objetivo, las probabilidades se consideran determinadas empíricamente y por ello a veces se denominan «probabilidades positivas». La dificultad de este enfoque es su dependencia de las observaciones: como no se puede observar un número infinito, las probabilidades determinadas empíricamente son necesariamente aproximaciones de los valores límite. Otra dificultad existente es que, en ciertas ocasiones, la frecuencia relativa puede no tender hacia un valor límite.

Un tercer enfoque para definir las probabilidades es el que las considera *grados de creencia racional*. Esta definición abarca afirmaciones de carácter probabilístico hechas a menudo, a veces de modo implícito, que no se pueden justificar por medio de las definiciones clásica o de frecuencia de probabilidad. Nos referimos a los casos en que no es posible contar (o medir) los resultados favorables ni los posibles, es decir, las afirmaciones como «es casi seguro que

suspenderé el examen de mañana» o «es muy probable que no sepamos nunca toda la verdad respecto a la muerte del presidente Kennedy». En ninguno de estos casos se puede utilizar una teoría bien definida para obtener probabilidades a priori ni pensar en repeticiones naturales a fin de conseguir probabilidades a posteriori.

La diversidad de enfoques respecto a la probabilidad puede dejarnos perplejos, pero, afortunadamente, origina relativamente pocas dificultades en la práctica. En parte, ello es debido a que la probabilidad concreta asignada a un suceso por cada uno de los enfoques es a menudo la misma. Por ejemplo, si una persona respetable nos enseña una moneda que parece normal, se considerará que la probabilidad de obtener cara es $1/2$, sea cual sea la definición de probabilidad que se utilice. En los casos en que existe desacuerdo (o en los que la definición clásica no nos da una respuesta) se puede utilizar la interpretación de la probabilidad basada en la frecuencia, y esto es lo que haremos cuando resulte necesario.

3.1 Conjuntos y espacios muestrales

Los principios de la teoría de la probabilidad se pueden desarrollar muy fácilmente por medio de la simple *teoría de conjuntos*. Ello parece también deseable debido a que la teoría de conjuntos ha adquirido gran popularidad en la estadística moderna y, en cierta medida, también en la economía. Un *conjunto* es una colección de objetos (o miembros, o elementos) definidos y claramente diferenciados. Por ejemplo, un conjunto puede estar formado por los números 1, 2, y 3, y lo representaremos de la forma siguiente

$$S = \{1, 2, 3\}.$$

Nótese, sin embargo, que un conjunto de los cinco números 1, 2, 2, 3, 3, se representa también

$$S = \{1, 2, 3\}.$$

ya que sólo tres de los cinco números están diferenciados. Obsérvese también que el orden en que se escriben los elementos no tiene ninguna importancia. Un conjunto se puede especificar ya sea enumerando todos sus elementos, o indicando una regla que nos permita decidir si un objeto determinado pertenece o no a él. Por tanto, podemos imaginar una regla que defina el conjunto de las familias de los Estados Unidos en un momento de tiempo determinado, en lugar de tener una lista completa. Un *conjunto vacío* o *nulo* es el que no contiene ningún elemento y lo expresamos como

$$S = \emptyset.$$

Si el elemento a pertenece al conjunto S , escribimos

$$a \in S;$$

y si no pertenece a él, escribimos
y si no pertenece a él, escribimos

$$a \notin S.$$

Si todos los elementos de S_1 lo son también de S , S_1 es un subconjunto de S . Esto se expresa de la forma siguiente

$$S_1 \subseteq S.$$

Por ejemplo, si $S_1 = \{1\}$ y $S = \{1, 2, 3\}$, $S_1 \subseteq S$. S_1 es un subconjunto estricto de S si éste contiene por lo menos un elemento que no aparece en S_1 ; esto se escribe de la forma siguiente

$$S_1 \subset S.$$

Si $S_1 = \{1\}$ y $S = \{1, 2, 3\}$, S_1 es un subconjunto estricto de S . Como ejemplo lógico y de construcción de conjuntos, considérese la afirmación de que

$$\emptyset \subseteq S.$$

Esto debe ser cierto porque, de no serlo, $\emptyset$ tendría al menos un elemento que no estaría contenido en S , pero $\emptyset$ no tiene ningún elemento y, por tanto, la afirmación debe cumplirse. Además, si $S \neq \emptyset$, S contendrá al menos un elemento no contenido en $\emptyset$, de modo que $\emptyset$ debe ser un subconjunto estricto de S , es decir, $\emptyset \subset S$.

Definamos ahora dos conceptos de particular importancia para la teoría de la probabilidad. El primero de ellos es la *unión de conjuntos*. La unión de los conjuntos S_1 y S_2 se define como el conjunto de elementos que pertenecen a S_1 , o a S_2 , o a ambos. Si llamamos S a la unión de los conjuntos S_1 y S_2 , podemos escribir

$$S = S_1 \cup S_2.$$

Por ejemplo, si $S_1 = \{a, b, c, 2\}$, y $S_2 = \{1, 2, 3\}$, $S = S_1 \cup S_2 = \{a, b, c, 1, 2, 3\}$. En la figura 3-1 se efectúa una representación diagramática de este concepto. El otro concepto importante para la teoría de la probabilidad

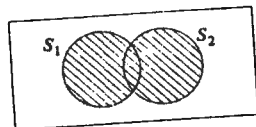


Figura 3-1

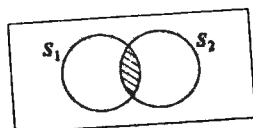


Figura 3-2

es el de *intersección de conjuntos*. La intersección de los conjuntos S_1 y S_2 es el conjunto de los elementos que pertenecen a ambos conjuntos S_1 y S_2 . Si llamamos S a este conjunto tenemos

$$S = S_1 \cap S_2.$$

Por ejemplo, si $S_1 = \{a, b, c, 2\}$, y $S_2 = \{1, 2, 3\}$, $S = S_1 \cap S_2 = \{2\}$. En la figura 3-2 se representa gráficamente este concepto.

El álgebra de conjuntos está basada en un pequeño número de postulados básicos y de leyes. Estos postulados incluyen, entre otras, las llamadas *leyes asociativa y conmutativa*. Considérense los conjuntos $S_1, S_2, S_3, \dots$, todos los cuales son subconjuntos de un cierto conjunto S . En este caso, la propiedad conmutativa señala que:

$$S_1 \cup S_2 = S_2 \cup S_1.$$

$$S_1 \cap S_2 = S_2 \cap S_1.$$

y la propiedad asociativa indica que:

$$(S_1 \cup S_2) \cup S_3 = S_1 \cup (S_2 \cup S_3),$$

$$(S_1 \cap S_2) \cap S_3 = S_1 \cap (S_2 \cap S_3).$$

Estas propiedades nos permiten extender las definiciones de unión e intersección de conjuntos al caso en que tenemos más de dos de ellos.

El conjunto más importante en la teoría de la probabilidad y en la teoría del muestreo es el llamado *espacio muestral*, y se trata del conjunto cuyos elementos representan todos los resultados posibles y diferenciados de un experimento (experimento que se puede haber realizado a propósito o puede haber ocurrido por causas naturales). Por tanto, el espacio muestral correspondiente al experimento de lanzar una moneda al aire consiste en los dos elementos cara, cruz, el espacio muestral correspondiente al experimento de lanzar un dado consta de seis elementos, 1, 2, 3, 4, 5, 6, etc. Los espacios muestrales que constan de un número finito de elementos (o de un número infinito, pero cuyos elementos se pueden contar) se denominan *espacios muestrales discretos* y todos los demás se llaman *continuos*. Debe observarse que el espacio muestral correspondiente a un experimento no es necesariamente único, es decir que dos o más espacios muestrales distintos pueden referirse a un mismo experimento. Por ejemplo, supóngase que el experimento consiste en lanzar dos monedas al aire. Un espacio muestral correspondiente a este experimento es

{ninguna cara, una cara, dos caras}.

y otro espacio muestral es

{cara en ambas monedas, cruz en ambas monedas, cara en la primera y cruz en la segunda, cruz en la primera y cara en la segunda}.

La diferencia entre el primer espacio muestral y el segundo es que un elemento del primero (una cara) se divide, en el segundo conjunto, en dos elementos. En general, resulta deseable utilizar espacios muestrales cuyos elementos no se pueden subdividir. Por último, en relación con los espacios muestrales discretos, obsérvese que un suceso es simplemente un subconjunto del espacio muestral.

3.2 Permutaciones y combinaciones

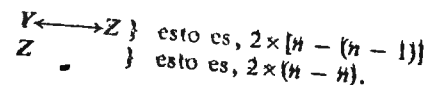
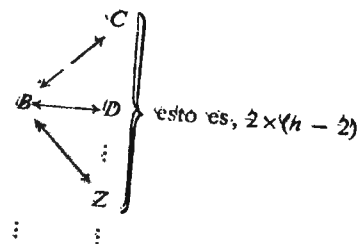
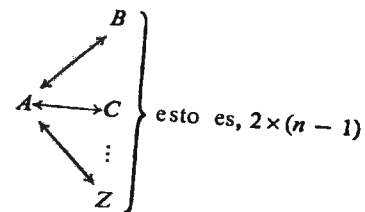
Los espacios muestrales discretos se pueden definir confeccionando una lista completa, y sus elementos pueden contarse de la misma forma, pero también es posible desarrollar fórmulas que simplifiquen esta tarea, en especial cuando se trata de un gran número de elementos. Estas fórmulas hacen referencia al número de *permutaciones* y *combinaciones* de los distintos resultados, y serán especialmente útiles para la obtención teórica de distribuciones muestrales cuando nos ocupemos de los atributos.

Consideremos en primer lugar las permutaciones. Entendemos por *permutación* una agrupación de objetos en un orden determinado, y nos interesa obtener el número de permutaciones que se pueden formar con los elementos de un conjunto determinado. Por ejemplo, considérese el conjunto A, B, C . En este conjunto existen tres tipos de permutaciones posibles, las que tienen un elemento, las de dos elementos, y las de tres. Su enumeración completa es la siguiente:

1. Permutaciones posibles de un elemento: A, B, C (es decir, tres).
2. Permutaciones posibles de dos elementos: AB, BA, AC, CA, BC, CB (es decir, seis).
3. Permutaciones posibles de tres elementos: $ABC, BAC, ACB, CAB, BCA, CBA$ (es decir, seis).

Consideremos ahora el caso general de un conjunto formado por n elementos $\{A, B, \dots, Z\}$.

1. Permutaciones posibles de un elemento: $A, B, \dots, Z$. Existen obviamente n permutaciones de este tipo.
2. Permutaciones posibles de dos elementos: $AB, BA, AC, CA, \dots, AZ, ZA, BC, CB, \dots, YZ, ZY$. La construcción de estas permutaciones de dos elementos se puede ver explícitamente de la forma siguiente:



(Las flechas con dos puntas indican que las asociaciones ocurren en ambos sentidos, es decir, $A \leftrightarrow B$ representa AB y también BA .) La suma total de estas permutaciones es

$$2[(n - 1) + (n - 2) + \dots + 2 + 1] = 2 \left[\left(\frac{n - 1}{2} \right) n \right] = n(n - 1).$$

Este resultado puede obtenerse también si observamos que en el caso de las permutaciones de dos elementos cada uno de los n elementos se asocia con los $(n - 1)$ restantes. Como en total existen n elementos, el número total de permutaciones debe ser $n(n - 1)$.

3. Permutaciones posibles de tres elementos: $ABC, BAC, ADC, DAC, \dots, AZX, ZAX, BCA, CBA, \dots, YZX, ZYX$. Es decir, cada una de las $n(n - 1)$ permutaciones obtenidas en 2 se asocia con los $(n - 2)$ elementos restantes del conjunto. Por tanto, la suma total es $n(n - 1)(n - 2)$.

Podríamos seguir con este proceso, pero las respuestas ya deben resultar evidentes y las podemos resumir en el cuadro 3-1. Si llamamos P , al número de

permutaciones de r elementos distintos obtenidos de un conjunto de n elementos, a partir de la fórmula para las permutaciones de r elementos del cuadro 3-1 tenemos

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1).$$

Cuadro 3-1

Número de elementos	Número de permutaciones
1	n
2	$n(n-1)$
3	$n(n-1)(n-2)$
$\vdots$	$\vdots$
r	$n(n-1)(n-2)\cdots(n-r+1)$
$\vdots$	$\vdots$
$(n-1)$	$n(n-1)(n-2)\cdots 2$
n	$n(n-1)(n-2)\cdots 2 \times 1$

Podemos simplificar esta expresión por medio de la llamada notación factorial. El *factorial* de un número n se indica mediante el símbolo $n!$ y se define como

$$n! = n(n-1)(n-2)\cdots 3 \times 2 \times 1,$$

siendo n cualquier número entero positivo. Podemos definir también

$$0! = 1.$$

Utilizando estos símbolos, la fórmula correspondiente a las permutaciones se convierte en

$${}_nP_r = \frac{n!}{(n-r)!} \quad (3.1)$$

Resulta interesante señalar lo siguiente:

1. ${}_nP_0 = 1$, es decir, un conjunto vacío sólo se puede ordenar de una forma.
2. Supongamos que tenemos n objetos y que k de ellos son exactamente el mismo objeto. En este caso

$${}_nP_k^{(k)} = \frac{n!}{k!} \quad (3.2)$$

Por ejemplo, consideremos el número de permutaciones de las cuatro letras de la palabra POOH:

POOH HPOO OPHO
POHO HOPO OPOH
PHOO HOOP OHPO
OHOP
OOHP
OOPH
Total = 12.

$${}_4P_4^{(2)} = \frac{4!}{2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12.$$

Este resultado se puede extender al caso de n objetos, k de los cuales son iguales entre sí, l son también iguales entre sí y distintos de los k anteriores, etc... En este caso

$${}_nP_n^{(k, l, \dots)} = \frac{n!}{k! l! \dots} \quad (3.3)$$

3. Si un conjunto de objetos se puede ordenar en m_1 formas distintas y otro conjunto de objetos se puede ordenar en m_2 formas, el número total de permutaciones es $m_1 \times m_2$. Ello se conoce con el nombre de *principio multiplicativo*.

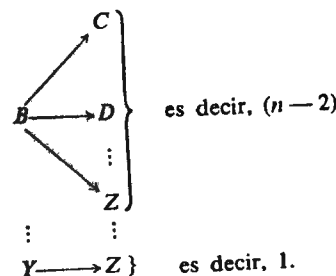
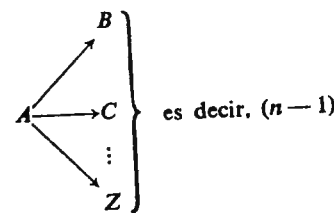
Como ejemplo de este principio, consideremos el número de permutaciones correspondientes a los resultados del experimento de lanzar dos monedas. En este caso, m_1 es el número de permutaciones correspondiente a los resultados posibles al lanzar la primera moneda (cara y cruz) y es igual a 2, mientras que m_2 indica los resultados de lanzar la segunda moneda y es también igual a 2; por tanto, el número total de permutaciones es $2 \times 2 = 4$. Esto se puede comprobar muy fácilmente: los resultados posibles al lanzar dos monedas son cara cara, cara cruz, cruz cara, cruz cruz, es decir, cuatro en total. Como otro ejemplo, consideremos el número de permutaciones en los resultados obtenidos al lanzar dos dados de seis caras. En este caso, $m_1 = 6$ y $m_2 = 6$, o sea que el número total es 36.

Todas las permutaciones con los mismos elementos representan una combinación determinada. Más exactamente, una *combinación* es un subconjunto formado por r elementos seleccionados de un conjunto de n elementos distintos, y sin fijarnos en su orden. Suponemos que $n \geq r$. Consideremos el conjunto $\{A, B, C\}$. Con los elementos de este conjunto podemos formar combinaciones de uno, dos, o tres elementos. Estas combinaciones son las siguientes.

1. Combinaciones de *un* elemento: $\{A\}$, $\{B\}$, $\{C\}$ (es decir, tres).
2. Combinaciones de *dos* elementos: $\{A, B\}$, $\{A, C\}$, $\{B, C\}$ (es decir, tres).
3. Combinaciones de *tres* elementos: $\{A, B, C\}$ (es decir, una).

Consideremos ahora el caso general de n elementos $\{A, B, C, \dots, Z\}$.

1. Combinaciones de *un* elemento: $\{A\}, \{B\}, \dots, \{Z\}$. El número de ellas es n .
2. Combinaciones de dos elementos: $\{A, B\}, \{A, C\}, \dots, \{Z, X\}$. Pueden representarse de la forma siguiente:



El total es

$$(n-1) + (n-2) + \dots + 1 = \frac{1}{2}n(n-1) = \frac{1}{2}n^2P_2.$$

Este resultado puede obtenerse también si nos fijamos en que cada combinación de dos letras permite obtener dos permutaciones. Por tanto, para obtener el número total de combinaciones de dos letras basta con dividir el número total de permutaciones de dos letras por dos.

3. Combinaciones de *tres* elementos: $\{A, B, C\}, \{A, B, D\}, \dots, \{X, Y, Z\}$. Puesto que cada combinación de tres letras permite obtener $3! = 6$ permutaciones, el total de combinaciones de tres letras es

$$\frac{1}{6}n^3P_3 = \frac{n(n-1)(n-2)}{3!}.$$

Podríamos seguir de esta forma para las combinaciones de cualquier número de elementos; los resultados se resumen en el cuadro 3-2. El número

Cuadro 3-2

Número de elementos	Número de combinaciones
1	n
2	$\frac{n(n-1)}{2}$
3	$\frac{n(n-1)(n-2)}{3!}$
$\vdots$	$\vdots$
r	$\frac{n(n-1)\dots(n-r+1)}{r!}$
$\vdots$	$\vdots$
$(n-1)$	n
n	1

de combinaciones de r elementos distintos extraídos de un conjunto de n elementos, sin tener en cuenta el orden, se representa generalmente por la notación $\binom{n}{r}$. A partir del cuadro 3-2 podemos ver que

$$(3.4) \quad \binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

Obsérvese que

$$\binom{n}{0} = 1,$$

$$\binom{n}{n} = 1,$$

$$\binom{n}{r} = \binom{n}{n-r}.$$

Uno de los usos más conocidos de la fórmula correspondiente al número de combinaciones es la obtención de los *coeficientes binomiales*. Considérense las expansiones algebraicas siguientes:

$$(a+b)^0 = 1,$$

$$(a+b)^1 = a+b,$$

$$(a+b)^2 = a^2 + 2ab + b^2,$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4,$$

etcétera.

Los coeficientes numéricos de estos desarrollos reciben el nombre de coeficientes binomiales. Obsérvese que vienen dados por la fórmula correspondiente al número de combinaciones, ya que se podrían haber escrito también en la forma

$$(a + b)^0 = 1,$$

$$(a + b)^1 = \binom{1}{0}a + \binom{1}{1}b,$$

$$(a + b)^2 = \binom{2}{0}a^2 + \binom{2}{1}ab + \binom{2}{2}b^2,$$

$$(a + b)^3 = \binom{3}{0}a^3 + \binom{3}{1}a^2b + \binom{3}{2}ab^2 + \binom{3}{3}b^3,$$

$$(a + b)^4 = \binom{4}{0}a^4 + \binom{4}{1}a^3b + \binom{4}{2}a^2b^2 + \binom{4}{3}ab^3 + \binom{4}{4}b^4,$$

etcétera.

Y en general,

$$(3.5) \quad (a + b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \dots + \binom{n}{n}b^n$$

$$= \sum_{r=0}^n \binom{n}{r} a^{n-r} b^r.$$

Un procedimiento sencillo para calcular los coeficientes binomiales consiste en utilizar lo que se conoce con el nombre de *triángulo de Pascal* (figura 3-3).

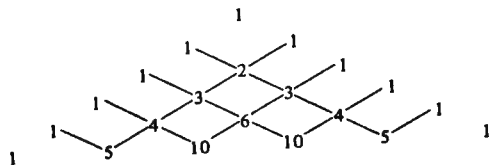


Figura 3-3

Cada fila empieza con un 1 y termina con otro 1; todos los demás números se obtienen sumando los dos números más próximos en la fila anterior.

La fórmula binomial (3.5) facilita la obtención del total de subconjuntos que pueden formarse a partir de un conjunto de n elementos. El primer subconjunto es el conjunto vacío, a continuación tenemos n subconjuntos de un elemento cada uno, $n(n-1)/2$ subconjuntos de dos elementos, y así sucesivamente. El total viene dado por la suma

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n}.$$

Estos sumandos son los coeficientes binomiales correspondientes a la expansión cuando $a = 1$ y $b = 1$, es decir

$$(1 + 1)^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n}.$$

Pero, obviamente,

$$(1 + 1)^n = 2^n.$$

Por tanto, obtenemos el resultado de que *un conjunto de n elementos tiene 2^n subconjuntos*.

3.3 Teoremas básicos de la teoría de la probabilidad

Consideremos ahora los teoremas básicos de la teoría de la probabilidad. Sean $A, B, C, \dots$ los sucesos representados por los subconjuntos de un espacio muestral discreto S e indiquemos mediante $P(A), P(B), P(C), \dots$ sus respectivas probabilidades. Además, postulamos que

$$0 \leq P(A) \leq 1 \quad \text{para cada subconjunto } A \text{ de } S, \quad \text{y } P(S) = 1.$$

Teorema 1. Si A es el suceso «no A », $P(\bar{A}) = 1 - P(A)$.

Confiamos en que esto no requiere mayor explicación. Si la probabilidad de que llueva es 0,40, resulta obvio que la probabilidad de que no llueva es 0,60. El teorema 1 se representa en la figura 3-4.



Figura 3-4

Antes de seguir adelante, debemos efectuar una digresión con el fin de aclarar el sentido de las afirmaciones condicionales del tipo «si..., entonces...», y otras parecidas, que se utilizan a menudo en el lenguaje de los teoremas y las pruebas. Para efectuar esta aclaración, debemos distinguir diversos tipos de condiciones; en especial, debemos distinguir entre una condición «necesaria», una condición «suficiente», y una condición «necesaria y suficiente». Ilustraremos esta distinción por medio de dos proposiciones, P_1 y P_2 , no idénticas. Consideremos, en primer lugar, una condición *suficiente*: si P_1 es cierta, entonces P_2 también es cierta. En ocasiones, esto se expresa señalando que « P_1 implica P_2 ». En este caso, P_1 es una condición suficiente de P_2 . Por ejemplo, P_1 puede ser «ser madre» y P_2 , «ser mujer», ya que la maternidad es, obviamente, condición suficiente de pertenencia al sexo femenino. Otro ejemplo puede ser que P_1 sea la afirmación «hoy es martes» y P_2 la afirmación «mañana es miércoles». Ahora, consideremos una condición *necesaria*: si P_1 no es cierta, P_2 tampoco es cierta. O, alternativamente, P_2 es cierta sólo si P_1 es cierta. Aquí, P_1 es una condición necesaria para P_2 . Por ejemplo, «ser mujer» es condición necesaria para «ser madre». Obsérvese que una condición suficiente puede ser o no condición necesaria, y al revés, una condición necesaria puede o no ser condición suficiente. Un ejemplo de condición suficiente pero no necesaria es «ser madre» como condición para «ser mujer», ya que es posible ser mujer sin ser madre. Si invertimos el orden de estas proposiciones, estableciendo «ser mujer» como condición para «ser madre», obtenemos una condición necesaria pero no suficiente, ya que ser mujer no basta para ser madre. Este último ejemplo ilustra una propiedad universal de la relación de suficiencia: si P_1 es condición suficiente para P_2 , P_2 es condición necesaria para P_1 . Por último, tenemos el caso de las condiciones necesarias y suficientes: si ocurre P_1 , ocurrirá P_2 , y si ocurre P_2 ocurrirá P_1 . Este tipo de condiciones se expresan mediante la afirmación «si y solamente si». Por ejemplo, «si y solamente si hoy es martes, mañana será miércoles»; es decir, el hecho de que «hoy es martes» no es sólo suficiente sino también necesario para que sea cierto que «mañana es miércoles».

A lo largo de nuestra exposición, la afirmación condicional utilizada más a menudo será «si P_1 , entonces P_2 ». Ello significa que P_1 es una condición suficiente para P_2 , pero puede ser o no condición necesaria. El hecho de que esta afirmación no incluya la cláusula «solamente si» no debe interpretarse como indicación de que P_1 no es condición necesaria para P_2 , y solamente indica que, para nuestros fines, no importa que P_1 sea o no condición necesaria para P_2 . Sólo se introducirá dicha cláusula cuando la condición de necesidad sea relevante para la discusión posterior. Con esto, damos por terminada la digresión, volviendo al análisis de los teoremas básicos de la probabilidad.

Teorema 2. (Teorema aditivo) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Este teorema indica que la probabilidad de A , o B , o ambos es igual a la

probabilidad de A más la probabilidad de B menos la probabilidad de que A y B ocurran a la vez, y se ilustra en la figura 3-5. A la suma $P(A) + P(B)$, hay que restarle el término $P(A \cap B)$ porque A y B se superponen y, en caso contrario, contaríamos dos veces dicho término.

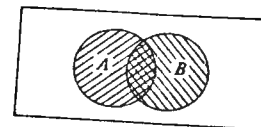


Figura 3-5

EJEMPLO 1. Consideremos la probabilidad de obtener una espada o una figura al extraer una carta de una baraja de 52 cartas. Sea A = espada, y B = figura. En este caso

$$P(A) = \frac{13}{52} \text{ (ya que hay 13 espadas en la baraja),}$$

$$P(B) = \frac{12}{52} \text{ (puesto que la baraja contiene 12 figuras),}$$

$$P(A \cap B) = \frac{3}{52} \text{ (ya que 3 de las figuras son espadas).}$$

Y por tanto

$$P(A \cup B) = \frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52}.$$

EJEMPLO 2. ¿Cuál es la probabilidad de obtener un «1» o un «2» al lanzar un dado?

$$\begin{aligned} P("1" \cup "2") &= P("1") + P("2") - P("1" \cap "2") \\ &= \frac{1}{6} + \frac{1}{6} - 0 \\ &= \frac{1}{3}. \end{aligned}$$

El segundo ejemplo nos conduce al teorema que se ocupa de la probabilidad de los sucesos mutuamente excluyentes, caso que ya mencionamos al tratar de la definición clásica de probabilidad. En el lenguaje de la teoría de conjuntos, los sucesos (es decir, subconjuntos de un espacio muestral) son mutuamente excluyentes si y solamente si su intersección es un conjunto vacío. El teorema

Teorema 3: Si A y B son mutuamente excluyentes, $P(A \cup B) = P(A) + P(B)$.

Este teorema se ilustra en la figura 3-6, y se puede extender a cualquier número de sucesos mutuamente excluyentes. En particular, si $A, B, C, \dots, Z$ se excluyen mutuamente,

$$P(A \cup B \cup C \cup \dots \cup Z) = P(A) + P(B) + P(C) + \dots + P(Z).$$

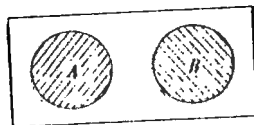


Figura 3-6

Examinemos ahora los sucesos que no se excluyen entre sí, y consideremos, por ejemplo, el caso en que elegimos al azar a una persona de una población y nos fijamos en dos características, si es fumador o no (F o $\bar{F}$), y cuál es su sexo (V o H). Es evidente que estas dos características no son mutuamente excluyentes, por lo menos en una sociedad moderna. El espacio muestral correspondiente a este experimento es $\{(VF), (HF), (V\bar{F}), (H\bar{F})\}$, en el que VF representa «varón fumador», etc. Si se trata de una población finita, la distribución se puede expresar a través de las frecuencias absolutas, de la forma siguiente:

		Hábito de fumar		Totales
		F	$\bar{F}$	
Sexo	V	a	b	$a + b$
	H	c	d	$c + d$
Totales		$a + c$	$b + d$	N

en la que $N = a + b + c + d$ representa el total de la población. En forma de probabilidades, la distribución sería

		Hábito de fumar		
		F	$\bar{F}$	
Sexo	V	$P(V \cap F)$	$P(V \cap \bar{F})$	$P(V)$
	H	$P(H \cap F)$	$P(H \cap \bar{F})$	$P(H)$
		$P(F)$	$P(\bar{F})$	1

Las probabilidades del centro de la tabla, representadas por medio de intersecciones de conjuntos, reciben el nombre de *probabilidades conjuntas*. Por ejemplo, $P(V \cap F)$ es la probabilidad de que un individuo elegido al azar sea varón y fumador, es decir, que posea las dos características conjuntamente. Las probabilidades que aparecen en la última fila y la última columna de la tabla reciben el nombre de *probabilidades marginales*, y así, por ejemplo, $P(V)$ es la probabilidad de que el individuo elegido sea varón, sin considerar si es fumador o no, $P(F)$ es la probabilidad de que sea fumador, sin tener en cuenta el sexo, etc. Veamos ahora cómo funciona en este caso el teorema aditivo que se ha definido anteriormente

$$\begin{aligned} P(V \cup F) &= P(V) + P(F) - P(V \cap F) = \frac{a+b}{N} + \frac{a+c}{N} - \frac{a}{N} = \\ &= \frac{a+b+c}{N} = 1 - \frac{d}{N} = 1 - P(H \cap \bar{F}). \end{aligned}$$

Es decir, la probabilidad de que el individuo elegido sea varón, o sea fumador, o ambas cosas a la vez, es simplemente igual a 1 menos la probabilidad de que sea mujer y no fumador (la única categoría que no está incluida en $V \cup F$). Del mismo modo,

$$\begin{aligned} P(V \cup \bar{F}) &= 1 - P(H \cap F), \\ P(H \cup F) &= 1 - P(V \cap \bar{F}), \\ P(H \cup \bar{F}) &= 1 - P(V \cap F). \end{aligned}$$

Supongamos ahora que queremos saber la probabilidad de que un individuo de un determinado sexo sea fumador (o no fumador), o de que un individuo que es fumador sea varón (o mujer). Estas probabilidades se conocen con el nombre de *probabilidades condicionadas*, y se representan por medio de la expresión $P(F|V)$, leyéndose «probabilidad de F condicionada a V », etc. Por ejemplo, $P(F|V)$ significa que se trata de un varón y queremos saber la probabilidad de que sea fumador. En una población finita, esta probabilidad es, lógicamente, el total de varones fumadores dividido por el número total de varones, y, por tanto,

$$P(F|V) = \frac{a}{a+b}$$

Del mismo modo,

$$P(\bar{F}|V) = \frac{b}{a+b}.$$

$$P(F|H) = \frac{c}{c+d},$$

$$P(\bar{F}|H) = \frac{d}{c+d},$$

$$P(V|F) = \frac{a}{a+c},$$

$$P(H|F) = \frac{c}{a+c},$$

$$P(V|\bar{F}) = \frac{b}{b+d},$$

$$P(H|\bar{F}) = \frac{d}{b+d}.$$

Obsérvese que

$$\begin{aligned} P(F|V) + P(\bar{F}|V) &= 1, \\ P(F|H) + P(\bar{F}|H) &= 1, \text{ etcétera.} \end{aligned}$$

Si lo representamos en forma de probabilidades, podemos escribir

$$P(F|V) = \frac{P(F \cap V)}{P(V)},$$

$$P(\bar{F}|V) = \frac{P(\bar{F} \cap V)}{P(V)}, \text{ etcétera.}$$

Estas expresiones son válidas tanto para una población finita como infinita. Este análisis nos conduce a un importante teorema:

Teorema 4. (Probabilidad condicionada). Si A y B son subconjuntos de un espacio muestral discreto, y $P(B) \neq 0$, $P(A|B) = P(A \cap B)/P(B)$.

Es decir, la probabilidad de A , condicionada a B , es igual a la probabilidad conjunta de A y B , dividida por la probabilidad marginal de B (distinta de cero).

En relación con este teorema de la probabilidad condicionada, deben tenerse en cuenta los puntos siguientes:

1. $A|B$ no es un conjunto.

2. $P(A|B)$ y $P(B|A)$ no son necesariamente iguales. En realidad, estas dos probabilidades son iguales si y solamente si $P(A) = P(B) \neq 0$.
3. Cuando escribimos $P(A|B)$ no indicamos con ello ninguna ordenación temporal entre A y B , y no importa que B ocurra antes, después, o en el mismo momento que A .
4. El teorema de la probabilidad condicionada se puede escribir también como $P(A \cap B) = P(A|B)[P(B)]$ y, en consecuencia, a veces recibe el nombre de *teorema multiplicativo*.

EJEMPLO 1.

$$P(\text{figura}|\text{espada}) = \frac{3/52}{13/52} = \frac{3}{13}$$

$$\text{y} \quad P(\text{espada}|\text{figura}) = \frac{3/52}{12/52} = \frac{3}{12}$$

EJEMPLO 2. Supongamos que lanzamos por dos veces una moneda insesgada. ¿Cuál es la probabilidad de que el resultado del segundo lanzamiento sea una cara, si el resultado del primero fue cara? Representemos por C_i el resultado «obtener cara en el i -ésimo lanzamiento». En este caso

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Consideremos este segundo ejemplo más detalladamente. El resultado muestra que la probabilidad de obtener cara en el segundo lanzamiento, si se ha obtenido cara en el primero, es $1/2$. Pero ésta es precisamente la probabilidad de obtener cara en *cualquier* lanzamiento, independientemente de lo que haya ocurrido en el primero. En realidad, sería muy sorprendente obtener un resultado distinto puesto que, obviamente, las monedas no tienen memoria y, por tanto, lo ocurrido en el primer lanzamiento resulta irrelevante para determinar lo que va a ocurrir en el segundo. El tipo de sucesos en los cuales el hecho de que ocurra uno de ellos no afecta a la probabilidad del otro se denominan *sucesos independientes*. Por tanto, si A es independiente de B , debemos tener

(3.6)

$$P(A|B) = P(A).$$

Es decir, la probabilidad de A , condicionada a B , es igual a la probabilidad marginal de A . Si utilizamos la expresión de $P(A|B)$ que aparece en el teorema 4 (con el supuesto de que $P(B) \neq 0$), obtenemos

(3.7)

$$\frac{P(A \cap B)}{P(B)} = P(A).$$

Si suponemos, asimismo, que $P(A) \neq 0$, podemos escribir de nuevo la ecuación (3.7) en la forma

$$(3.8) \quad \frac{P(A \cap B)}{P(A)} = P(B).$$

Pero la parte izquierda de la ecuación (3.8) es precisamente $P(B|A)$, o sea que

$$(3.9) \quad P(B|A) = P(B).$$

Es decir, si A es independiente de B , B es independiente de A . Las ecuaciones (3.7) y (3.8) nos conducen al siguiente teorema, de gran importancia:

Teorema 5. (Independencia) Si $P(A) \neq 0$ y $P(B) \neq 0$, A y B son independientes si y solamente si $P(A \cap B) = P(A) \times P(B)$.

En otras palabras, A y B son independientes si y solamente si su probabilidad conjunta es igual al producto de sus probabilidades marginales. El teorema se puede extender a fin de abarcar cualquier número de sucesos. En particular, $A, B, C, \dots, Z$ (todos ellos con probabilidad distinta de cero) son independientes si y solamente si

$$P(A \cap B \cap C \dots \cap Z) = P(A) \times P(B) \times P(C) \times \dots \times P(Z).$$

EJEMPLO 1. Supongamos que lanzamos dos dados. ¿Cuál es la probabilidad de obtener un número par en uno de ellos y un número impar en el otro? Llamemos E al hecho de obtener un número par, y Q al de obtener un número impar. En este caso, $P(E) = 1/2$, $P(Q) = 1/2$, $P(E \cap Q) = 1/2 \times 1/2 = 1/4$.

EJEMPLO 2. ¿Cuál es la probabilidad de obtener tres caras al lanzar tres veces una moneda?

$$P(C_1 \cap C_2 \cap C_3) = P(C_1) \times P(C_2) \times P(C_3) = 1/2 \times 1/2 \times 1/2 = 1/8.$$

EJEMPLO 3. Supongamos que el hecho de fumar, o, no, es independiente del sexo, es decir, $P(V \cap F) = P(V) \times P(F)$. Significa esto necesariamente que $P(H \cap \bar{F}) = P(H) \times P(\bar{F})$?

	F	$\bar{F}$	
V	$P(V \cap F)$	$P(V \cap \bar{F})$	$P(V)$
H	$P(H \cap F)$	$P(H \cap \bar{F})$	$P(H)$
	$P(F)$	$P(\bar{F})$	1

Ahora bien,

$$P(H \cap F) = P(F) - P(V \cap F) = P(F)[1 - P(V)] = P(F)P(H)$$

$$\begin{aligned} P(H \cap \bar{F}) &= P(H) - P(H \cap F) = P(H) - P(H)P(F) \\ &= P(H)[1 - P(F)] = P(H)P(\bar{F}). \end{aligned}$$

Por tanto, la respuesta es afirmativa.

Quienes no conocen a fondo la teoría de la probabilidad, tienden a confundir con frecuencia la aplicabilidad del teorema aditivo, que hace referencia a los sucesos mutuamente excluyentes, con la del teorema multiplicativo que se refiere a los sucesos independientes. Tal confusión se debe a que no distinguen entre $P(A \cup B)$ (es decir, la probabilidad de que ocurra A , o B , o ambos a la vez) y $P(A \cap B)$ (es decir, la probabilidad de que A y B ocurran a la



Figura 3-7

vez) (ver figura 3-7): Ahora bien, si A y B son mutuamente excluyentes, sabemos que

$$(3.10) \quad P(A \cap B) = 0,$$

lo cual se deriva del teorema 3 y se ilustra en la figura 3-6. Para que A y B sean independientes, es preciso que

$$(3.11) \quad P(A \cap B) = P(A) \times P(B)$$

Pero las ecuaciones (3.10) y (3.11) sólo pueden ser ciertas simultáneamente si $P(A)$ o $P(B)$ (o ambas) son cero. Por otra parte, el teorema 5 elimina esta posibilidad y, por tanto, los sucesos mutuamente excluyentes no pueden ser, a la vez, independientes. El hecho resulta obvio incluso mediante un razonamiento mucho menos formal. Si A y B son mutuamente excluyentes, el hecho de que ocurra uno de ellos elimina la posibilidad de que ocurra el otro, es decir, al ocurrir A , la probabilidad de B se hace cero. Sin embargo, se han definido como sucesos independientes aquellos en los que el hecho de producirse uno de ellos no afecta en forma alguna a la probabilidad del otro, lo cual, ciertamente, no es el caso si se trata de dos sucesos que se excluyen mutuamente. Por el mismo razonamiento, también es evidente que dos sucesos independientes no pueden ser mutuamente excluyentes.

3.4 Funciones de probabilidad y variables aleatorias discretas

Supongamos que realizamos el experimento consistente en lanzar dos monedas al aire. El espacio muestral asociado a este experimento se puede describir como $S = \{CC, CZ, ZC, ZZ\}$. Cada elemento de este espacio muestral corresponde a un determinado número de caras (o cruces). Por tanto, tenemos que

Elementos del espacio muestral	Número de caras
CC	2
CZ	1
ZC	1
ZZ	0

De forma alternativa, podemos presentar esta relación de la forma siguiente:

Número de caras	Elementos del espacio muestral
0	ZZ
1	ZC, CZ
2	CC

Resulta bien obvio que el principio de asociar los elementos de un espacio muestral a una característica numérica se puede aplicar de forma general a cualquier espacio muestral. Esta característica numérica recibe el nombre de *variable aleatoria discreta* (o *estocástica*). Por tanto, una variable aleatoria discreta es una variable cuyos valores están asociados a los elementos de un espacio muestral. Una forma comúnmente utilizada para representar las variables aleatorias consiste en emplear mayúsculas (por ejemplo, X), utilizando letras minúsculas (por ejemplo, x) para representar sus valores específicos; si estos valores aparecen en un cierto orden (por ejemplo, una secuencia de observaciones), este orden se indica mediante un subíndice (x_1, x_2 , etc.). En el ejemplo anterior, hemos asociado el número de caras con los elementos del espacio muestral, pero también habiésemos podido elegir como característica numérica la proporción de caras en lugar del número absoluto. Obviamente, tanto una como otra son variables aleatorias.

Los espacios muestrales están formados por elementos relacionados con los resultados de un experimento. x , por tanto, se puede asociar una cierta probabilidad a cada uno de los elementos. Además, como cada valor de una variable aleatoria discreta está asociado a uno o varios elementos del espacio muestral, ello significa que es posible asociar una probabilidad a cada uno de estos valores. Es decir, podemos describir una variable aleatoria discreta como una

variable que toma distintos valores, cada uno de ellos con una probabilidad determinada. En el ejemplo anterior, consistente en lanzar dos monedas, cada elemento del espacio muestral consta de dos sucesos independientes. Si las monedas son insesgadas, tenemos

$$P(Z \cap Z) = P(Z)P(Z) = 1/4$$

$$P(Z \cap C) = P(Z)P(C) = 1/4$$

$$P(C \cap Z) = P(C)P(Z) = 1/4$$

$$P(C \cap C) = P(C)P(C) = 1/4$$

Representemos por x el número de caras y por $f(x)$ la probabilidad de obtener este número. En este caso podemos escribir

Elementos del espacio muestral	Número de caras: x	Probabilidad de x : $f(x)$
ZZ	0	1/4
ZC	1	1/4
CZ	1	1/4
CC	2	1/4
		1

Alternativamente, podemos presentar la información anterior en la forma

x	$f(x)$
0	1/4
1	1/2
2	1/4
	1

La distribución anterior recibe el nombre de función de probabilidad. Podemos generalizar fácilmente esta idea con el fin de presentar la definición siguiente:

Si X es una variable aleatoria discreta que toma los valores $x_1, x_2, \dots, x_n$, cuyas probabilidades son $f(x_1), f(x_2), \dots, f(x_n)$, el conjunto de pares

$$x_1 \quad f(x_1)$$

$$x_2 \quad f(x_2)$$

$$\vdots$$

$$x_n \quad f(x_n)$$

es una función de probabilidad (o distribución de probabilidad) de X .

Como ejemplo, consideremos el experimento consistente en lanzar dos dados. En este caso, tomemos como variable aleatoria el número de puntos obtenidos, y sus valores serán 2, 3, ..., 12. Podemos considerar que el espacio muestral correspondiente a este experimento está formado por todas las permutaciones posibles de los dos conjuntos de números, del 1 al 6. Según el principio multiplicativo, estas permutaciones son $6 \times 6 = 36$, y cada una de ellas tiene la misma probabilidad (suponiendo que los dados no estén trucados). La distribución de probabilidad resultante, correspondiente al número de puntos obtenidos, se presenta en el cuadro 3-3.

Cuadro 3-3

x	Elementos del espacio muestral	$f(x)$
2	11	1/36
3	12, 21	2/36
4	13, 31, 22	3/36
5	14, 41, 23, 32	4/36
6	15, 51, 24, 42, 33	5/36
7	16, 61, 25, 52, 34, 43	6/36
8	26, 62, 35, 53, 44	5/36
9	36, 63, 45, 54	4/36
10	46, 64, 55	3/36
11	56, 65	2/36
12	66	1/36
		1

Cuando X toma un valor, por ejemplo x_i , esto representa un suceso determinado, y $f(x_i)$ representa la probabilidad de este suceso. Como X sólo puede tomar un valor a la vez, la probabilidad de que el valor de X sea x_i o x_j ($x_i \neq x_j$) es la probabilidad de dos sucesos que se excluyen mutuamente y, por tanto, según el teorema 3, es igual a $f(x_i) + f(x_j)$. Por ejemplo, la probabilidad de que el número de puntos obtenidos al lanzar dos dados sea 4 u 8 es igual a $3/36 + 5/36 = 2/9$. Y, naturalmente, la probabilidad de que X tome cualquier valor distinto de x_i es $1 - f(x_i)$ como indica el teorema 1.

En algunos casos, se quiere hallar la probabilidad de que X tome un valor menor o igual que un número determinado. Estas probabilidades reciben el nombre de *probabilidades acumuladas* y se representan normalmente por $F(x)$. Si $x_1, x_2, \dots, x_n$ son los valores de X en orden de magnitud creciente, es decir, si $x_1 < x_2 < \dots < x_n$, la probabilidad acumulada de x_k viene dada por la expresión

$$(3.12) \quad F(x_k) = f(x_1) + f(x_2) + \dots + f(x_k) = \sum_{i=1}^k f(x_i).$$

Los valores no comprendidos en el campo de variación de X (es decir, los valores inferiores a x_1 o superiores a x_n) tienen una probabilidad igual a cero, y por tanto también podemos escribir

$$(3.13) \quad F(x_k) = \sum_{i=-\infty}^k f(x_i).$$

La distribución que proporciona las probabilidades acumuladas $F(x)$ correspondientes a cada valor de X se denomina *distribución acumulada* de X . La distribución acumulativa correspondiente al experimento de lanzar dos dados aparece en el cuadro 3-4. La función de probabilidad y la distribución acumulativa de este experimento se representan gráficamente en la figura 3-8. Es interesante señalar varios puntos en relación con las probabilidades acumuladas.

1. Si x_n es el mayor valor de X , $F(x_n) = 1$.
2. $F(-\infty) = 0$, $F(\infty) = 1$.
3. $F(x_i) - F(x_{i-1}) = f(x_i)$ ($x_i > x_{i-1} > x_{i-2} > \dots$).

Cuadro 3-4

x	$F(x)$
2	1/36
3	3/36
4	6/36
5	10/36
6	15/36
7	21/36
8	26/36
9	30/36
10	33/36
11	35/36
12	1

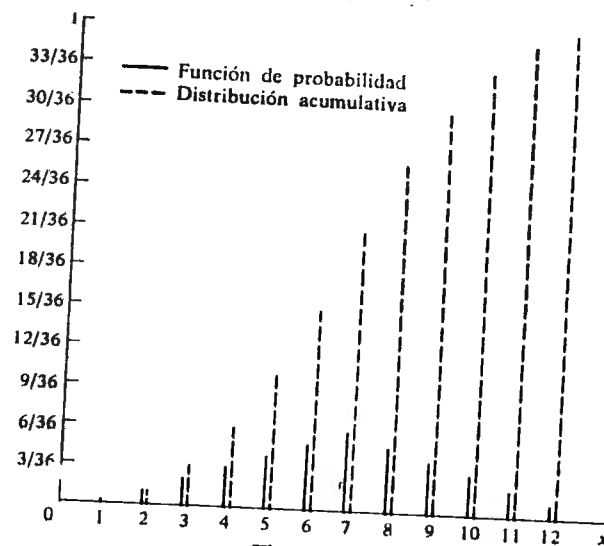


Figura 3-8

Hasta este momento nos hemos ocupado únicamente de una variable aleatoria y su distribución de probabilidad. Este tipo de distribuciones se pueden denominar *distribuciones univariantes*. Para ampliar nuestro campo de análisis podemos considerar un espacio muestral que contiene más de una varia-

ble aleatoria a la vez, y la distribución de probabilidad correspondiente se llamará distribución *multivariante*; el caso especial consistente en dos variables aleatorias recibe el nombre de distribución *bivariante*. Como ejemplo de ello, consideremos el experimento consistente en extraer una carta de una baraja de 52 cartas. Cada una de estas cartas se distingue por dos características, el número y el palo. Representemos el número por X , que toma los valores 1 (para el as), 2, 3, ..., 10, 11(J), 12(Q), y 13(K), y el palo por Y , con valores 1 (corazones), 2 (diamantes), 3 (tréboles), y 4. La distribución de probabilidad *bivariante* correspondiente a este experimento es la siguiente:

	Valores de X							$f(y)$
	1	2	3	...	12	13		
Valores de Y	1	1/52	1/52	1/52	...	1/52	1/52	13/52
	2	1/52	1/52	1/52	...	1/52	1/52	13/52
	3	1/52	1/52	1/52	...	1/52	1/52	13/52
	4	1/52	1/52	1/52	...	1/52	1/52	13/52
$f(x)$		4/52	4/52	4/52	...	4/52	4/52	1

La probabilidad de que X tome el valor x e Y tome el valor y se denomina *probabilidad conjunta* de x e y , y se representa por $f(x, y)$. Por ejemplo, la probabilidad de obtener una Q de corazones es $f(12, 1) = 1/52$. La probabilidad de que X tome el valor x , *independientemente del valor de Y* , se denomina *probabilidad marginal* de x ; la distribución de estas probabilidades se llama *distribución marginal de X* y, en el ejemplo anterior, aparece en la última fila. Del mismo modo, la *distribución marginal de Y* está formada por las probabilidades de los distintos valores de Y , independientemente del valor que tome X , y en el ejemplo aparece en la última columna a la derecha. De este modo, la probabilidad de obtener una Q (de cualquier palo) es $4/52$, y la probabilidad de obtener un corazón (cualquiera de ellos) es $13/52$. Las distribuciones de probabilidades marginales son, en realidad, distribuciones univariantes y se representan como éstas, es decir, $f(x)$, $f(y)$, etc. Obsérvese que las probabilidades de la distribución marginal de una variable se obtienen sumando las correspondientes probabilidades asociadas a cada uno de los valores de la otra variable. Por tanto, en el caso de la distribución *bivariante* podemos escribir

$$(3.14) \quad f(x_i) = \sum_{j=1}^{\infty} f(x_i, y_j) = f(x_i, y_1) + f(x_i, y_2) + \dots$$

y

$$(3.15) \quad f(y_j) = \sum_{i=1}^{\infty} f(x_i, y_j) = f(x_1, y_j) + f(x_2, y_j) + \dots$$

Por último, la probabilidad de que X sea igual a x , cuando Y es igual a y , se denomina *probabilidad de x condicionada a y* , y se representa por la expresión $f(x|y)$. Del mismo modo, la *probabilidad de y condicionada a x* se representa por $f(y|x)$. Aplicando el teorema 4, obtenemos

$$(3.16) \quad f(x|y) = \frac{f(x, y)}{f(y)}$$

En el ejemplo consistente en extraer una carta de una baraja, la probabilidad de obtener, por ejemplo, una Q, si la carta es un corazón, es

$$f(12|1) = \frac{1/52}{13/52} = \frac{1}{13}$$

y la probabilidad de que sea corazón si es una Q es

$$f(1|12) = \frac{1/52}{4/52} = \frac{1}{4}$$

Al considerar las distribuciones multivariantes, la cuestión de la dependencia o independencia se hace importante. Anteriormente se ha definido la independencia como la condición según la cual la probabilidad condicionada de un suceso es igual a su probabilidad marginal. Por tanto, X e Y son independientes si y solamente si, para todos los valores de X e Y , se cumple que

$$f(x|y) = f(x) \quad [f(y) \neq 0]$$

y

$$f(y|x) = f(y) \quad [f(x) \neq 0].$$

Ello significa que la distribución marginal y la condicionada son iguales para cada variable. Otra implicación del teorema 5 es que X e Y son independientes si y solamente si

$$(3.17) \quad f(x, y) = f(x)f(y)$$

para todos los valores de X e Y . Esto se puede generalizar para cualquier número de variables aleatorias. En particular, consideraremos que las variables aleatorias discretas $X, Y, Z, \dots$ son independientes si, y solamente si,

$$(3.18) \quad f(x, y, z, \dots) = f(x)f(y)f(z) \dots$$

para todos los valores de $X, Y, Z, \dots$

Por ejemplo, consideremos el experimento de lanzar dos veces, dos dados, y llamemos X a la variable cuyos valores representan el número de puntos del

primer lanzamiento, e Y a la variable cuyos valores indican los puntos del segundo lanzamiento. Obviamente, las probabilidades de los distintos resultados del segundo lanzamiento no se ven afectadas en absoluto por el resultado del primero, y viceversa, de modo que X e Y son independientes. Obsérvese que la distribución (marginal) de X es exactamente la misma que la de Y y la podemos ver en el cuadro 3-3, donde se puede apreciar que la probabilidad de obtener dos veces 7 puntos es $(6/36) \times (6/36) = 1/36$, la de obtener 4 puntos y luego 8 es $(3/36) \times (5/36) = 5/432$, etcétera.

Aparte de los distintos ejemplos, el análisis de las funciones de probabilidad realizado en esta sección ha sido completamente general. Es evidente que las distintas situaciones experimentales pueden dar lugar a distintas distribuciones de probabilidad, y al describir estas distribuciones no es siempre necesario escribir la totalidad de la distribución como se ha hecho en los ejemplos; en muchos casos se puede obtener una fórmula algebraica de $f(x)$ (e incluso para las distribuciones multivariantes) que proporciona una descripción completa de la distribución que nos interesa. Algunas distribuciones son muy corrientes, en el sentido de que describen las probabilidades de muchas de las situaciones experimentales con las que nos encontramos en la práctica. Una de las distribuciones de probabilidad más corrientes, y la más sencilla, es la llamada *distribución uniforme*. En esta distribución, la probabilidad de que X tome cualquier valor, de entre un número de ellos, es la misma, o sea que $f(x)$ es constante para todos los valores de X . Esta distribución describe las probabilidades de los distintos resultados obtenidos al lanzar un dado, de extraer de una baraja una carta de determinado palo, de ganar un premio en la lotería, y muchas otras. Otra distribución extraordinariamente frecuente es la llamada *distribución binomial*, que tiene gran importancia en la inferencia estadística, y de la que nos ocuparemos en detalle en la sección 4.1.

3.5 Variables aleatorias continuas y funciones de probabilidad

En el caso discreto, los elementos del espacio muestral se hallan representados por puntos separados por distancias finitas, y podemos asignar un valor numérico a cada punto y una probabilidad a cada valor. Sin embargo, existen muchos experimentos para los cuales el espacio muestral no consiste en un número de puntos, sino que abarca un intervalo completo (o una cantidad de intervalos). Las variables aleatorias asociadas a los resultados de tales experimentos se denominan *variables aleatorias continuas*. Un ejemplo de este tipo de experimentos consiste en observar una aguja que puede girar libremente en una esfera de reloj. En este caso, la variable aleatoria puede ser el tiempo (por ejemplo, en horas), indicado por la aguja, al parar. Evidentemente, la aguja se puede parar en un número infinito de puntos entre 0 y 12; y, por tanto, la probabilidad de que se pare en cualquier punto concreto es cero. Por otra parte, la probabilidad de que se pare en un intervalo alrededor de cualquier punto concreto, es distinta de cero, y se puede calcular.

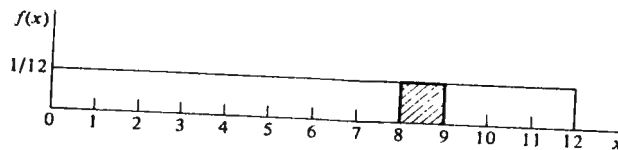


Figura 3-9

Las probabilidades correspondientes a cada intervalo, en el experimento anterior, se representan gráficamente en la figura 3-9. La probabilidad de que el valor de X esté comprendido en un intervalo determinado viene dada por el área debajo de la curva (que, en este caso, es una línea recta). Por ejemplo, el área rayada de la figura 3-9 representa la probabilidad de que x esté entre 8 y 9 y es $1/12$. Esta idea se puede generalizar a fin de aplicarla a otros experimentos relacionados con variables continuas. La *distribución de probabilidad de una variable continua X (denominada función de densidad) se representa mediante una curva, y la probabilidad de que X tome un valor comprendido en el intervalo a, b ($a < b$) viene dada por el área comprendida entre esta curva y el eje de abscisas, entre los puntos a y b* . La mayor parte de las distribuciones de probabilidad con las que trataremos son distribuciones continuas.

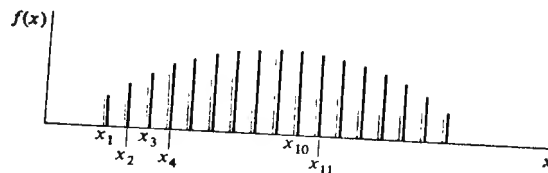


Figura 3-10

Para analizar con más detalle la función de densidad, la podemos comparar con la función de probabilidad de una variable discreta. Supongamos que tenemos una variable aleatoria discreta que puede tomar los valores $x_1, x_2, \dots, x_n$, que se hallan ordenados en orden creciente. Supongamos que la función de probabilidad de esta variable es la representada en la figura 3-10. La probabilidad de que X tome un valor superior a x_3 pero menor o igual que x_{10} es la siguiente:

$$P(x_3 < x \leq x_{10}) = f(x_4) + f(x_5) + \dots + f(x_{10}) = \sum_{i=4}^{10} f(x_i).$$

De modo alternativo, podríamos expresarla como

$$P(x_3 < x \leq x_{10}) = F(x_{10}) - F(x_3),$$

en la que $F(x)$ representa la función de distribución acumulativa. Supongamos ahora que tenemos una variable aleatoria continua cuya densidad de probabilidad es la representada en la figura 3-11. En este caso, la probabilidad de que X tome un valor comprendido en el intervalo x_3, x_{10} viene dada por el área rayada debajo de la curva. Pero las áreas por debajo de una curva se pueden calcular por medio de integrales, de modo que si sabemos que la fórmula al-

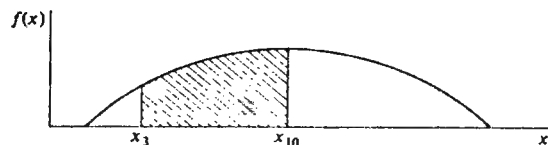


Figura 3-11

gebraica que describe la función de densidad de la figura 3-11 es $f(x)$, el área por debajo de la curva, entre los puntos x_3 y x_{10} , viene dada por la integral correspondiente. Por tanto,

$$P(x_3 < x < x_{10}) = \int_{x_3}^{x_{10}} f(x)dx.$$

(La probabilidad de que $x = x_{10}$ es cero y, por tanto, $P(x_3 < x < x_{10})$ y $P(x_3 < x \leq x_{10})$ son equivalentes.) La integración entre x_3 y x_{10} en el caso de la variable continua es análoga a la suma de las probabilidades efectuadas en el caso de variables discretas. En general, si X es una variable aleatoria continua, la probabilidad de que tome un valor comprendido en el intervalo a, b viene dada por la expresión:

$$(3.19) \quad P(a < x < b) = \int_a^b f(x)dx,$$

en la que $f(x)$ es la correspondiente función de densidad. En realidad, en muy pocos casos será necesario operar con integrales explícitas, pero, en cambio, nos proporcionan un marco conceptual muy útil para estudiar las densidades de probabilidad.

La probabilidad de que X tome cualquier valor es 1 (es decir, existe certeza) y, por tanto,

$$(3.20) \quad P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Además, la probabilidad de que X tome cualquier valor menor o igual que un valor x es

$$(3.21) \quad F(x) = \int_{-\infty}^x f(x)dx.$$

Al igual que en el caso discreto, $F(x)$ recibe el nombre de *probabilidad acumulada de x* . Obsérvese que

1. $F(-\infty) = 0$, y $F(+\infty) = 1$.
2. Si $a < b$,

$$F(b) - F(a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^b f(x)dx = P(a < x < b).$$

(En otras palabras, la diferencia entre dos probabilidades acumuladas es igual a la probabilidad simple.)

La probabilidad acumulada se representa gráficamente en la figura 3-12.

Al igual que en el caso de las variables discretas, los espacios muestrales continuos pueden incluir más de una variable a la vez. La distribución de probabilidad correspondiente a este caso recibe el nombre de *densidad de probabilidad multivariante*. Dicha distribución indica la *probabilidad conjunta* de

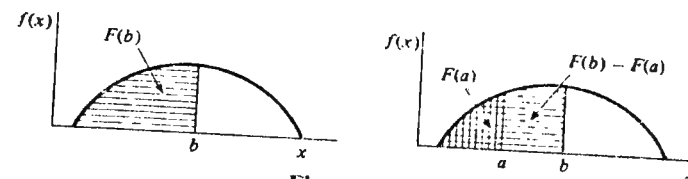


Figura 3-12

que cada una de las variables consideradas se halle comprendida en un intervalo concreto; si, por ejemplo, las variables son X, Y, Z , la *densidad de probabilidad conjunta* será $f(x, y, z)$. Análogamente a lo expuesto para el caso discreto, podemos definir la *densidad marginal* de X como la densidad de probabilidad de X sean cuales sean los valores de las demás variables. Obsérvese que la densidad marginal de una variable se obtiene integrando la distribución conjunta para todos los valores de la otra variable, es decir,

$$(3.22) \quad f(x) = \int_{-\infty}^{+\infty} f(x, y)dy \quad \text{and} \quad f(y) = \int_{-\infty}^{+\infty} f(x, y)dx.$$

Asimismo, al igual que en el caso discreto, las variables aleatorias continuas $X, Y, Z, \dots$ son *independientes* si, y solamente si,

$$f(x, y, z, \dots) = f(x)f(y)f(z) \dots$$

Algunas distribuciones continuas resultan especialmente importantes debido a que aparecen a menudo en la práctica. Una de tales distribuciones es la *distribución uniforme continua*, que hemos utilizado en el ejemplo de la aguja en el capítulo 2. Se halla representada en la figura 3-9. Otro ejemplo lo constituye

la llamada *distribución normal*, que es extraordinariamente corriente y que estudiaremos en detalle en la sección 4.2. También introduciremos otras distribuciones cuando surja la ocasión.

3.6 Esperanza matemática

Las distribuciones de probabilidad, al igual que las distribuciones de frecuencias corrientes, presentan varias características, las más conocidas de las cuales son la media y la variancia, y estas características se definen en términos de sus *valores esperados* o *esperanzas matemáticas*. A fin de aclarar estos términos, podemos efectuar una analogía con una distribución de frecuencias corriente. Supongamos, por ejemplo, que tenemos una muestra de 64 familias clasificadas según el número de hijos, tal como aparece en el cuadro 3-5.

Cuadro 3-5

Número de hijos: x	Número de familias: n	Proporción de familias: f
0	4	0.06250
1	12	0.18750
2	20	0.31250
3	16	0.25000
4	8	0.12500
5	2	0.03125
6	2	0.03125
	64	1.00000

En este cuadro, n representa las frecuencias absolutas y f las frecuencias relativas. Existen siete valores distintos de X , $x_1 = 0$, $x_2 = 1$, ..., $x_7 = 6$, asociados a las distintas frecuencias. Consideremos ahora el número medio de hijos por familia. Aunque existen varias clases de promedios, en este caso, casi todo el mundo elegiría probablemente la media aritmética —el número total de hijos en la muestra, dividido por el número de familias—. Esta viene dada por

$$\bar{x} = \frac{\sum_{i=1}^7 n_i x_i}{\sum_{i=1}^7 n_i} = \frac{1}{64}(4 \times 0 + 12 \times 1 + \dots + 2 \times 6) = 2.40625.$$

Este mismo resultado se puede obtener también utilizando las frecuencias relativas:

$$\begin{aligned}\bar{x} &= \sum_{i=1}^7 f_i x_i = (0.06250 \times 0 + 0.18750 \times 1 + \dots + 0.03125 \times 6) \\ &= 2.40625.\end{aligned}$$

Es decir, la media aritmética de una distribución de frecuencias es, en realidad, una media ponderada de los distintos valores de la variable, con unas ponderaciones que son las correspondientes frecuencias relativas. Consideremos ahora una *variable aleatoria discreta* X con la siguiente función de probabilidad:

x	$f(x)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_n	$f(x_n)$

en la que n es algún número finito. En este caso, el *valor esperado* o la *esperanza matemática* de X es

$$(3.23) \quad E(X) = \sum_{i=1}^n x_i f(x_i).$$

Podemos ver que $E(X)$ no es más que un promedio ponderado de los distintos valores de X con unas ponderaciones determinadas por las probabilidades respectivas. Por esta razón, $E(X)$ se identifica con la media de la población μ ; es decir,

$$(3.24) \quad \mu = E(X).$$

A partir de esta exposición, cabe esperar que la analogía entre la media de una distribución de frecuencias y el valor esperado de una variable aleatoria discreta X resulte suficientemente clara.

Se utiliza la expresión *valor esperado* para subrayar la relación entre la media de la población y la expectativa respecto al resultado de un experimento. Supongamos, por ejemplo, que nos dicen que al lanzar un dado ganaremos un número de dólares igual al número de puntos obtenidos. El problema es saber qué cantidad esperamos ganar, antes de lanzar realmente el dado. Si formamos nuestras expectativas según la regla dada por la ecuación (3.23), la respuesta es

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = 3.5.$$

Aunque, en realidad, no es posible (y, por lo tanto, no «esperamos», en el sentido coloquial de la expresión) ganar 3,50 dólares, esta cifra representa un resumen de todos los resultados posibles. Si lanzáramos el dado un número infinito de veces, la ganancia media de cada lanzamiento sería precisamente 3,50 dólares.

El *valor esperado de una variable continua* se define de forma muy parecida. Como ya mencionamos anteriormente, las variables continuas se pueden considerar como casos límites de variables discretas, en las que los valores que puede tomar la variable se van haciendo cada vez más numerosos y más próximos entre sí. Por tanto, si X es una variable aleatoria continua, con densidad de probabilidad $f(x)$, su valor esperado (o esperanza matemática) es

$$(3.25) \quad E(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

Se integra entre $-\infty$ y $+\infty$ para asegurarnos de que consideramos todos los valores posibles de X .

Hay que señalar que existen algunas distribuciones de probabilidad para las cuales *no existe un valor esperado de la variable*; es decir, $E(X)$ puede ser igual a infinito. Un ejemplo clásico es el constituido por la función de probabilidad:

x	$f(x)$
2	1/2
4	1/4
8	1/8
16	1/16
$\vdots$	$\vdots$

Esta función de probabilidad es totalmente válida, ya que $f(x) \geq 0$ para cada x , y

$$\sum_{i=1}^{\infty} f(x_i) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1.$$

En realidad, podemos interpretar $f(x)$ como la probabilidad de obtener una cara al lanzar una vez una moneda, 2 caras al efectuar 2 lanzamientos, 3 caras en 3 lanzamientos, y así sucesivamente en orden ascendente. El valor esperado de X es¹

$$E(X) = 2 \times \frac{1}{2} + 4 \times \frac{1}{4} + 8 \times \frac{1}{8} + \dots = 1 + 1 + 1 + \dots = \infty.$$

1. Este caso se conoce con el nombre de *la paradoja de San Petersburgo*. Supongamos que nos proponen que lancemos una moneda al aire y nos ofrecen dos dólares si obtenemos cara en el primer lanzamiento pero no en el segundo; cuatro dólares si obtenemos cara en el primero y el segundo, pero no en el tercero; ocho dólares si obtenemos cara en los tres primeros pero no en el cuarto, y así sucesivamente. ¿Cuál es para nosotros el valor de este juego? Según los cálculos, el valor esperado (es decir, el beneficio esperado en este caso) es infinito, pero no es probable que nadie esté dispuesto a pagar esta suma (ni siquiera, simplemente un millón de dólares) por el privilegio de participar en él. Este es el aspecto paradójico de la situación.

Afortunadamente, este tipo de distribuciones no es muy frecuente. De ahora en adelante supondremos que nos ocupamos de distribuciones cuyos valores esperados existen. Cuando éste no sea el caso, lo señalaremos explícitamente.

El concepto de esperanza matemática se puede ampliar fácilmente con el fin de aplicarlo a otros problemas, además de la obtención del valor medio de X . En particular, si X es una variable aleatoria y $g(X)$ es una función de esta variable, se cumple que,

$$(3.26) \quad Eg(X) = \begin{cases} \sum_{i=1}^{\infty} g(x_i)f(x_i) & (X \text{ discreta}) \\ \int_{-\infty}^{+\infty} g(x)f(x)dx & (X \text{ continua}). \end{cases}$$

Como ejemplo, demostraremos el teorema siguiente.

Teorema 6. Si X es una variable aleatoria, y a y b son constantes,

$$E(aX + b) = aE(X) + b.$$

Demostración: a) X discreta:

$$\begin{aligned} E(aX + b) &= \sum_{i=1}^n (ax_i + b)f(x_i) \\ &= \sum_i ax_i f(x_i) + \sum_i bf(x_i) \\ &= a \sum_i x_i f(x_i) + b \sum_i f(x_i) \\ &= aE(X) + b. \end{aligned}$$

b) X continua:

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{+\infty} (ax + b)f(x)dx \\ &= \int axf(x)dx + \int bf(x)dx \\ &= aE(X) + b. \end{aligned}$$

1. Este caso se conoce con el nombre de *la paradoja de San Petersburgo*. Supongamos que nos proponen que lancemos una moneda al aire y nos ofrecen dos dólares si obtenemos cara en el primer lanzamiento pero no en el segundo; cuatro dólares si obtenemos cara en el primero y el segundo, pero no en el tercero; ocho dólares si obtenemos cara en los tres primeros pero no en el cuarto, y así sucesivamente. ¿Cuál es para nosotros el valor de este juego? Según los cálculos, el valor esperado (es decir, el beneficio esperado en este caso) es infinito, pero no es probable que nadie esté dispuesto a pagar esta suma (ni siquiera, simplemente un millón de dólares) por el privilegio de participar en él. Este es el aspecto paradójico de la situación.

Una aplicación directa de (3.26) nos permite obtener ciertas características especiales de las distribuciones de probabilidad llamadas *momentos*, los cuales constituyen una familia de parámetros que caracterizan a la distribución. Podemos distinguir dos clases de momentos: momentos respecto al origen (es decir, cero) y momentos respecto a la media (es decir, μ). Los *momentos respecto al origen* se definen como

$$(3.27) \quad \mu'_r = E(X^r) = \begin{cases} \sum_i x_i^r f(x_i) & (X \text{ discreta}) \\ \int_{-\infty}^{+\infty} x^r f(x) dx & (X \text{ continua}) \end{cases}$$

para $r = 0, 1, 2, \dots$. De modo que, al sustituir los distintos valores de r , obtenemos los momentos de distinto orden. Por ejemplo,

$$\mu'_0 = 1,$$

$$\mu'_1 = E(X) = \mu.$$

Y, por tanto, la media es el momento de primer orden respecto al origen. Los momentos de orden superior respecto al origen se utilizan menos a menudo y no poseen ningún nombre especial. Los *momentos respecto a la media* se definen como

$$(3.28) \quad \mu_r = E[(X - \mu)^r] = \begin{cases} \sum_i (x_i - \mu)^r f(x_i) & (X \text{ discreta}) \\ \int_{-\infty}^{+\infty} (x - \mu)^r f(x) dx & (X \text{ continua}) \end{cases}$$

para $r = 0, 1, 2, \dots$. Obsérvese que

$$\mu_0 = 1,$$

$$\mu_1 = E(X - \mu) = E(X) - \mu = 0,$$

$$\mu_2 = E(X - \mu)^2 = \text{Var}(X).$$

Por tanto, la variancia es el momento de segundo orden respecto a la media, y su raíz cuadrada recibe el nombre de *desviación standard* de la distribución.

A continuación, consideramos las características más importantes de una distribución.

1. La *media* $\mu = E(X)$ es una medida de la tendencia central de una distribución. Su rasgo característico es el hecho de que, en la población, la suma

de todas las desviaciones de los valores de X respecto de μ es cero. Esto es cierto tanto si se trata de una variable discreta como de una variable continua. En particular, si la distribución es discreta tenemos que:

$$\sum_i (x_i - \mu) f(x_i) = \sum_i x_i f(x_i) - \mu = E(X) - \mu = 0.$$

Si la distribución es simétrica, la media está situada en el centro de la distribución.

2. La *variancia* σ^2 o $\text{Var}(X)$, que a menudo se representa como

$$(3.29) \quad \text{Var}(X) = E[X - E(X)]^2,$$

es una medida de la extensión o dispersión de la distribución. En realidad, se trata de la media de los cuadrados de las desviaciones de X respecto de μ , puesto que a partir de la definición de esperanza matemática,

$$\sigma^2 = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i) & (X \text{ discreta}) \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx & (X \text{ continua}) \end{cases}$$

Naturalmente, la variancia no puede ser nunca negativa. Si los valores de X están muy concentrados, el punto de concentración tiene que ser la media (o al menos, estar muy próximo a ella) y la variancia será muy pequeña. En el caso extremo en el que todos los valores de X son iguales (es decir, si X es una constante), la variancia será igual a cero. Obsérvese que un procedimiento alternativo para obtener la variancia es el siguiente:

$$(3.30) \quad \begin{aligned} \sigma^2 &= E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2, \end{aligned}$$

puesto que μ es una constante. Además, podemos obtener el teorema siguiente:

Teorema 7. Si X es una variable aleatoria y a y b son constantes,

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Demostración:

$$\text{Var}(aX + b) = E[(aX + b) - E(aX + b)]^2$$

Según la definición de variancia,

$$E(aX + b) = aE(X) + b$$

Ahora bien, según el teorema 6 en la que por definición $E(X) = \mu$. En consecuencia,

$$\begin{aligned}\text{Var}(aX + b) &= E[aX + b - (a\mu + b)]^2 = E[aX - a\mu]^2 \\ &= E[a^2(X - \mu)^2] = a^2E[(X - \mu)^2] = a^2\text{Var}(X).\end{aligned}$$

En ocasiones, resulta más conveniente utilizar la *desviación standard* de X que es simplemente

$$\sigma = \sqrt{\text{Var}(X)}.$$

3. Una *medida de asimetría* viene dada por la expresión

$$(3.31) \quad \mu_3 = E(X - \mu)^3.$$

Si la distribución es simétrica, μ_3 será igual a cero. Si la distribución es asimétrica hacia la izquierda (es decir, si el lado izquierdo es más alargado), μ_3 será negativa, mientras que si la distribución es asimétrica hacia la izquierda, la medida será positiva. En la figura 3-13, la primera distribución, $f_1(x)$, es asimétrica hacia la izquierda, la segunda, $f_2(x)$, es simétrica, y la tercera, $f_3(x)$, es

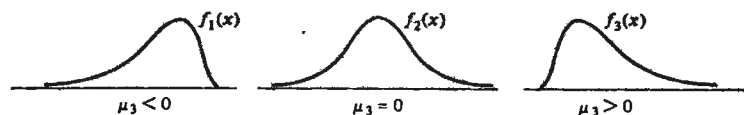


Figura 3-13

asimétrica hacia la derecha. En muchos casos es preferible utilizar una medida de asimetría algo distinta, llamada α_3 y que se define como

$$(3.32) \quad \alpha_3 = \frac{\mu_3}{\sigma^3}.$$

El denominador de esta expresión sirve para que en todos los casos la medida se exprese en una misma escala, con el fin de que las comparaciones entre la asimetría de distintas distribuciones no se vean alteradas por las diferencias en

dispersion. Como $\sigma^3 = [\text{Var}(X)]^{3/2}$ será siempre positiva, el signo de α_3 siempre viene dado por el signo de μ_3 .

Hasta este punto, hemos considerado únicamente distribuciones de probabilidad univariantes, pero no existe ninguna dificultad para extender el concepto de esperanza matemática a las distribuciones que tienen más de una variable aleatoria. Supongamos que tenemos dos variables aleatorias X e Y con una distribución conjunta $f(x, y)$ y queremos determinar el valor esperado de cualquier función de X e Y , por ejemplo, $g(X, Y)$. En este caso,

$$E[g(X, Y)] = \begin{cases} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) f(x_i, y_j) & (X \text{ and } Y \text{ discreta}) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy & (X \text{ and } Y \text{ continua}). \end{cases} \quad (3.33)$$

Esto puede generalizarse para cualquier número de variables aleatorias. Algunas funciones de las variables aleatorias tienen un interés especial y las estudiamos en los teoremas que aparecen a continuación, que son muy importantes para la teoría econométrica. Los demostraremos para el caso de dos variables discretas, pero su extensión a los demás casos es inmediata.

Teorema 8. El valor esperado de una suma de variables aleatorias es igual a la suma de sus valores esperados, es decir, $E(X + Y + Z + \dots) = E(X) + E(Y) + E(Z) + \dots$

Demostración:

$$\begin{aligned}E(X + Y) &= \sum_i \sum_j (x_i + y_j) f(x_i, y_j) = \sum_i \sum_j x_i f(x_i, y_j) + \sum_i \sum_j y_j f(x_i, y_j) \\ &= \sum_i x_i \sum_j f(x_i, y_j) + \sum_j y_j \sum_i f(x_i, y_j) = \sum_i x_i f(x_i) + \sum_j y_j f(y_j) \\ &= E(X) + E(Y).\end{aligned}$$

Teorema 9. El valor esperado de una combinación lineal de variables aleatorias es igual a la combinación lineal de sus valores esperados, es decir,

$$E(aX + bY + cZ + \dots) = aE(X) + bE(Y) + cE(Z) + \dots,$$

donde $a, b, c, \dots$ son cualesquiera constantes.

$$\begin{aligned}
 E(aX + bY) &= \sum_i \sum_j (ax_i + by_j)f(x_i, y_j) \\
 &= \sum_i \sum_j ax_i f(x_i, y_j) + \sum_i \sum_j by_j f(x_i, y_j) \\
 &= a \sum_i x_i \sum_j f(x_i, y_j) + b \sum_j y_j \sum_i f(x_i, y_j) \\
 &= aE(X) + bE(Y).
 \end{aligned}$$

Teorema 10. Si X e Y son dos variables aleatorias independientes, el valor esperado de su producto es igual al producto de sus valores esperados, es decir, $E(XY) = E(X)E(Y)$.

Demostración:

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j).$$

Pero si X e Y son independientes,

$$f(x_i, y_j) = f(x_i)f(y_j).$$

Y por tanto,

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i)f(y_j) = \sum_i x_i f(x_i) \sum_j y_j f(y_j) = E(X)E(Y).$$

Es importante señalar la diferencia entre el valor esperado de una suma y el de un producto. El valor esperado de una suma es *siempre* igual a la suma de los valores esperados, mientras que el valor esperado de un producto sólo es igual al producto de los valores esperados *si* las variables son independientes.

El último teorema que queremos presentar relativo a los valores esperados requiere la definición de *covariancia* entre dos variables aleatorias, X e Y , que se representa generalmente como $\text{Cov}(X, Y)$ (y, a veces, σ_{XY}) y se define como

$$(3.34) \quad \text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)].$$

El signo de la covariancia depende de cuál sea la dirección de la asociación entre X e Y . Si se trata de una asociación positiva —es decir, si los valores pequeños de X tienden a estar asociados a los valores pequeños de Y , y los valores grandes de X a los valores grandes de Y —, la covariancia será *positiva*. Por otra parte, si existe asociación negativa —es decir, los valores pequeños de X asociados a los valores grandes de Y y viceversa—, la covariancia será *negativa*. (Se consideran valores «pequeños» los que son inferiores a la media y valores «grandes» los superiores a la media.) Es decir, si existe asociación

positiva $[x - E(X)]$ y $[y - E(Y)]$ tendrán generalmente el mismo signo y, por tanto, su producto tenderá a ser positivo, originando así una covariancia positiva. Pero si la asociación es negativa, $[x - E(X)]$ y $[y - E(Y)]$ tenderán a tomar signos opuestos y, por tanto, su producto será negativo, reflejándose esto en el signo de la covariancia. En la figura 3-14 se ilustran estos casos.

$E(Y)$	$x - E(X) < 0$	$x - E(X) > 0$
	$y - E(Y) > 0$	$y - E(Y) > 0$
0	$x - E(X) < 0$	$x - E(X) > 0$
	$y - E(Y) < 0$	$y - E(Y) < 0$
	$E(X)$	

Figura 3-14

Efectuada esta introducción, podemos presentar el teorema siguiente:

Teorema 11. Si X e Y son dos variables aleatorias independientes,

$$\text{Cov}(X, Y) = 0.$$

Demostración:

$$\begin{aligned}
 \text{Cov}(X, Y) &= E[X - E(X)][Y - E(Y)] \\
 &= E[XY - YE(X) - XE(Y) + E(X)E(Y)] \\
 &= E(XY) - E(X)E(Y).
 \end{aligned}$$

Pero si X e Y son independientes, según el teorema 10,

$$E(XY) = E(X)E(Y);$$

y por tanto,

$$\text{Cov}(X, Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

Debe señalarse que a pesar de que la independencia implica necesariamente covariancia cero, lo contrario no es cierto. Resulta fácil encontrar casos en los que $\text{Cov}(X, Y) = 0$ y, sin embargo, X e Y no son independientes. Considérese, por ejemplo, la distribución siguiente:

		Valores de Y		$f(x)$
		0	1	
Valores de X	1	0	1/3	1/3
	2	1/3	0	1/3
	3	0	1/3	1/3
$f(y)$		1/3	2/3	1

evidentemente, X e Y no son independientes, ya que la distribución condicional y la marginal son completamente distintas. Ahora bien,

$$\text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y).$$

$$\text{Pero } E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = 1 \times 1 \times \frac{1}{3} + 2 \times 0 \times \frac{1}{3} + 3 \times 1 \times \frac{1}{3} = \frac{4}{3}.$$

$$\text{Además, } E(X) = 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2$$

$$\text{y } E(Y) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}.$$

$$\text{Por tanto, } \text{Cov}(X, Y) = \frac{4}{3} - 2 \times \frac{2}{3} = 0.$$

Las variables cuya covarianza es cero se llaman variables *no correlacionadas*. La exposición que antecede ha tenido por objeto demostrar que la independencia es una condición suficiente pero no necesaria para que se produzca una covarianza cero. En un sentido amplio, la diferencia entre independencia y no correlación radica en el hecho de que la independencia elimina la posibilidad de cualquier tipo de relación entre las variables, mientras que la no correlación sólo elimina las relaciones lineales.

Como último punto, obsérvese que, análogamente al concepto de probabilidad condicionada, también existe el de *esperanza condicionada* de Y para una X dada, que se define de la forma siguiente:

$$(3.35) \quad E(Y|X) = \begin{cases} \sum_i y_i f(y_i|x_i) & (Y \text{ discreta}) \\ \int_{-\infty}^{+\infty} y f(y|x) dy & (Y \text{ continua}). \end{cases}$$

Es decir, la esperanza de Y condicionada a X es igual a la media de la distribución de Y condicionada a X . Utilizando los valores del ejemplo anterior, la esperanza de Y condicionada a X cuando esta última toma el valor 1 es

$$E(Y|x=1) = 0 \times \frac{0}{1/3} + 1 \times \frac{1/3}{1/3} = 1.$$

Del mismo modo

$$E(Y|x=2) = 0 \times \frac{1/3}{1/3} + 1 \times \frac{0}{1/3} = 0$$

y

$$E(Y|x=3) = 0 \times \frac{0}{1/3} + 1 \times \frac{1/3}{1/3} = 1,$$

mientras que la media de Y no condicionada es

$$E(Y) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}.$$

El concepto de esperanza condicionada se puede generalizar aplicándolo a cualquier función (de valor único) de Y , por ejemplo, $g(Y)$. En este caso, tenemos que

$$(3.36) \quad E[g(Y)|X] = \begin{cases} \sum_i g(y_i) f(y_i|x_i) & (Y \text{ discreta}) \\ \int_{-\infty}^{+\infty} g(y) f(y|x) dy & (Y \text{ continua}). \end{cases}$$

La esperanza condicionada y la no condicionada tienen una diferencia importante: $E[g(Y)]$ siempre es una constante, pero $E[g(Y)|X]$ no es necesariamente constante y puede ser función de X .

En el resto del libro se efectuarán muy a menudo operaciones relacionadas con valores esperados. Ello se debe a que estamos interesados en la obtención de las propiedades de las distribuciones muestrales y, como se ha señalado anteriormente, éstas son precisamente distribuciones de probabilidad de estimadores o estadísticos de prueba, y los valores esperados son un medio muy conveniente para describir las características de estas distribuciones.

EJERCICIOS:

- 3.1. Enumerar los elementos de los espacios muestrales correspondientes a los experimentos siguientes:
 - a. Extraer (con reposición) dos bolas de una urna que contiene 7 bolas blancas y 3 rojas.
 - b. Lanzar una moneda al aire y, si se obtiene cara, lanzar un dado (si sale cruz se interrumpe el experimento).
 - c. Extraer cartas de una baraja de 52 naipes hasta que aparezca una figura.
- 3.2. ¿De cuántas formas distintas pueden sentarse alrededor de una mesa 3 hombres y 3 mujeres suponiendo que a) no se impone ninguna restricción sobre

el orden en que se distribuyen; b) cada mujer tiene que estar sentada entre dos hombres; c) igual que en el caso b), pero la dueña de la casa insiste en sentarse en el sitio más cercano a la cocina?

- 3-3. ¿Cuántas muestras distintas de tamaño 20 se pueden extraer (con reposición) de una población de 30 objetos?
- 3-4. Las preguntas siguientes hacen referencia a las probabilidades de los resultados de lanzar un dado. Calcular la probabilidad de a) obtener un número par o un número mayor que 3; b) obtener un 4 y, a continuación, un 5; c) no obtener ningún 4 al lanzarlo dos veces; d) obtener al menos un 4 al lanzarlo dos veces; e) no obtener dos 7 al lanzarlo dos veces.
- 3-5. Considérese el siguiente conjunto de resultados igualmente probables:

$$S = \{1, 2, \dots, n\}$$

en el que n es un múltiplo de seis. Definir

$$S_1 = \{1, 3, \dots, (n-1)\}.$$

$$S_2 = \{2, 4, \dots, n\}.$$

$$S_3 = \{3, 6, \dots, n\}.$$

Obtener las probabilidades siguientes:

- a. $P(n)$, $P(S_1)$, $P(S_2)$, y $P(S_3)$. e. $P(S_1 \cup S_3)$.
- b. $P(S_1 \cap S_3)$. f. $P(S_2 \cup S_3)$.
- c. $P(S_2 \cap S_3)$. g. $P[S_1(S_2 \cap S_3)]$.
- d. $P(S_1 \cup S_2)$.

- 3-6. Representemos por S_1 y S_2 dos subconjuntos del espacio muestral S . Dadas $P(S_1)$, $P(S_2)$, y $P(S_1 \cap S_2)$, calcular

- a. $P(\bar{S}_1 \cup \bar{S}_2)$. c. $P(\bar{S}_1 \cap S_2)$.
- b. $P(\bar{S}_1 \cap \bar{S}_2)$. d. $P(\bar{S}_1 \cup S_2)$.

- 3-7. Supongamos que se pide a tres niños que adivinen cuál de tres jarras contiene una galleta. Si el primero acierta, se queda con la galleta. Si no, se elimina la jarra vacía y el segundo niño tiene que adivinar dónde está la galleta. Si acierta se queda con ella y si falla se la queda el tercer niño. ¿Hay alguna ventaja en ser el primero, el segundo o el tercero? Dar las razones.

- 3-8. En Michigan, las placas de matrícula de los automóviles están formadas por dos letras seguidas de cuatro números. ¿Cuántas matrículas distintas pueden formarse de este modo?

- 3-9. Demostrar que $\sum_{i=0}^n \binom{a}{i} \binom{b}{n-i} = \binom{a+b}{n}$, $a, b > n$.

- 3-10. En los datos correspondientes a una gran ciudad, observamos la siguiente distribución de solicitudes de empleo no cualificado, clasificadas según el período de paro del solicitante:

Duración del paro (semanas)	0	1	2	3	4	5	6	7	8	9-12
Proporción de solicitantes	0.25	0.20	0.15	0.10	0.10	0.05	0.04	0.03	0.02	0.06

- a. ¿Cuál es la duración esperada de paro de un solicitante?
- b. Calcular el valor de la desviación standard y de la medida de asimetría α_1 de esta distribución.
- 3-11. Si X es una variable aleatoria con media μ y variancia σ^2 , y definimos $Z = X - (X - \mu)/\sigma^2$, calcular $E(Z)$ y $\text{Var}(Z)$.
- 3-12. Siendo X una variable aleatoria con media μ , y b una constante distinta de μ , demostrar que $E(X - \mu)^2 < E(X - b)^2$.

- 3-13. Sean X e Y dos variables aleatorias, y a y b dos constantes. Demostrar que

- a. $\text{Var}(X + a) = \text{Var}(X)$.
- b. $\text{Var}(Y + b) = \text{Var}(Y)$.
- c. $\text{Cov}[(X + a), (Y + b)] = \text{Cov}(X, Y)$.

- 3-14. Considérese una muestra de variables $X_1, X_2, \dots, X_n$, distribuidas de forma idéntica e independiente, cada una de ellas con media μ y variancia σ^2 . Calcular (para $i = 1, 2, \dots, n$ e $i \neq j$):

- a. $E(X_i \bar{X})$. f. $E[X_i(X_j - \bar{X})]$.
- b. $E(X_i^2)$. g. $\text{Var}(X_i - \bar{X})$.
- c. $E(\bar{X}^2)$. h. $\text{Cov}(X_i, \bar{X})$.
- d. $E[X_i(X_i - \bar{X})]$. i. $\text{Cov}[(X_i - \bar{X}), \bar{X}]$.
- e. $E[\bar{X}(X_i - \bar{X})]$. j. $\text{Cov}[(X_i - \bar{X}), (X_j - \bar{X})]$.

- 3-15. Considérense tres variables aleatorias X , Y y Z . Demostrar que a) el hecho de que X no esté correlacionada con Y , e Y no esté correlacionada con Z , implica que X no esté correlacionada con Z ; b) el hecho de que X sea independiente de Y , e Y sea independiente de Z , no implica que X sea independiente de Z .

3-16. Demostrar que

$$\mu_r = \mu_r' + (-1)^1 \binom{r}{1} \mu_{r-1}' + \dots + (-1)^{r-1} \binom{r}{r-1} \mu_1'$$

para $r = 1, 2, 3, \dots$ (Si la demostración para el caso general resulta demasiado difícil, demostrarlo para $r = 1, 2$, y 3 .)3-17. Considérese la siguiente distribución conjunta de X e Y :

		X				
		- 2	- 1	0	1	2
Y	40	0,09	0,15	0,27	0,25	0,04
	20	0,01	0,05	0,08	0,05	0,01

- Calcular las distribuciones marginales de X e Y .
- Calcular la distribución de X condicionada a Y cuando $y = 20$.
- Indicar si existe correlación entre X e Y .
- Indicar si X e Y son independientes.

- 3-18. Una empresa tiene que elegir entre dos proyectos de inversión. El proyecto A producirá un beneficio de 20 000 dólares si tiene éxito, o una pérdida de 2 000 dólares si fracasa, mientras que el proyecto B producirá 25 000 dólares si tiene éxito o una pérdida de 5 000 si fracasa. La probabilidad de que el proyecto tenga éxito es la misma en ambos casos. Hallar esta probabilidad sabiendo que, en base al beneficio esperado, los dos proyectos se consideran equivalentes.

Obtención teórica de distribuciones muestrales

El motivo para dedicar el capítulo 3 al estudio de la probabilidad y las distribuciones de probabilidad ha sido la necesidad de disponer de instrumentos que permitan la obtención teórica de las distribuciones muestrales de los estimadores y los estadísticos de prueba, así como facilitar la comprensión del proceso y los resultados de la estimación y la contrastación de hipótesis. En este capítulo vamos a *utilizar* la teoría de la probabilidad para obtener varias distribuciones muestrales. Como ya se explicó en la sección 1.5, las distribuciones muestrales pueden obtenerse experimental o teóricamente. En el enfoque experimental, se construye la población y se extraen de ella un gran número de muestras, cada una de las cuales nos proporciona un valor del estimador o estadístico de prueba que nos interesa. La distribución de frecuencias de estos valores constituye una aproximación a la distribución de probabilidad que hubiésemos obtenido si el número de muestras hubiese sido infinito en lugar de grande. Este enfoque experimental presenta dos graves inconvenientes. El primero de ellos es que debemos contentarnos con aproximaciones, en lugar de obtener la forma exacta de las distribuciones muestrales, y el segundo es que, en sentido estricto, los resultados obtenidos son aplicables solamente a la población concreta en la que se ha basado el experimento. Las distribuciones muestrales obtenidas de forma teórica no presentan ninguna de estas dos dificultades y, por tanto, son claramente superiores a las obtenidas de forma experimental. El único problema existente para utilizar el enfoque teórico es que éste puede no funcionar en todos los casos —en el sentido de que nuestra habilidad y conocimientos matemáticos pueden ser insuficientes para obtener los resultados deseados—. Esto sucede de forma especialmente frecuente en relación a los estimadores de los sistemas de relaciones económicas, como veremos en el capítulo 13.

En este capítulo nos limitaremos a la obtención de las distribuciones muestrales de la media y la proporción de aciertos de la muestra. Estas dos distribuciones ya se han obtenido de forma experimental, para poblaciones específicas, en el capítulo 2 y se pueden obtener matemáticamente para cualquier población sin gran dificultad. Más adelante, cuando surja la ocasión, estudiaremos otras distribuciones muestrales.

4.1 Distribución muestral de la proporción de aciertos de la muestra:

4.1.1 La distribución binomial. Consideremos una población en la que cada unidad se puede clasificar de acuerdo con el hecho de que posea o no un atributo determinado. Tal población puede ser finita o infinita. Si no es infinita, supondremos que, al extraer una muestra, cada unidad extraída se devuelve antes de extraer una nueva unidad (es decir, efectuamos muestreo con reposición), lo cual implica que no existe ningún límite respecto al tamaño de la muestra. Las unidades que posean el atributo considerado recibirán el nombre de *aciertos*, y las que no lo posean se denominarán *fallos*. Como ejemplo concreto, podemos pensar que consideramos la población adulta de los Estados Unidos en un momento de tiempo determinado, y que el atributo que nos interesa es el hecho de que un individuo tome café o no lo tome. La extracción de un individuo que toma café se considera un acierto, y la de un individuo que no toma café se considera un fallo. Al igual que en la sección 2, usaremos la notación siguiente:

π = proporción (o límite de la frecuencia relativa) de aciertos en la población;

ρ = proporción (o límite de la frecuencia relativa) de fallos en la población;

n = tamaño de la muestra;

$\hat{\pi}$ = proporción de aciertos en la muestra;

X = número de aciertos en la muestra.

Obsérvese que, al efectuar una extracción aleatoria, la probabilidad de obtener un acierto es π y la de obtener un fallo es ρ . Indiquemos el «acierto» por S y el «fallo» por F . Por tanto, tenemos que

$$P(S) = \pi,$$

$$y \quad P(F) = \rho, \quad \text{y} \quad P(F) = 1 - \pi,$$

ya que F es «no S ».

Obtengamos ahora las distribuciones muestrales de X y $\hat{\pi}$ correspondientes a muestras de distinto tamaño. Al existir una correspondencia, de uno a uno entre X y $\hat{\pi}$ debida al hecho de que

$$\hat{\pi} = \frac{X}{n}, \quad \text{y} \quad X = n\hat{\pi},$$

si conocemos la distribución de X conoceremos totalmente la distribución de $\hat{\pi}$, y viceversa. Efectuaremos una descripción completa de las distribuciones muestrales correspondientes a cada tamaño de la muestra utilizando a una variable aleatoria; no obstante, calcularemos las principales características

la distribución para X y también para $\hat{\pi}$. Al presentar cada distribución, se calcularán los valores de X , su probabilidad, y los valores de $xf(x)$ y $x^2f(x)$, necesarios para calcular la media y la variancia.

Distribución muestral con muestras de tamaño 1

Número de aciertos: x	Probabilidad: $f(x)$	$xf(x)$	$x^2f(x)$
0	$P(F) = \rho$	0	0
1	$P(S) = \pi$	π	π
Suma	$\rho + \pi = 1$	π	π

Media y variancia de X :

$$E(X) = \sum_i x_i f(x_i) = \pi,$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 = \sum_i x_i^2 f(x_i) - \left[\sum_i x_i f(x_i) \right]^2 \\ &= \pi - \pi^2 = \pi(1 - \pi) = \pi\rho, \end{aligned}$$

Media y variancia de $\hat{\pi}$:

$$E(\hat{\pi}) = E\left(\frac{X}{n}\right) = E(X) = \pi,$$

$$\text{Var}(\hat{\pi}) = \text{Var}\left(\frac{X}{n}\right) = \text{Var}(X) = \pi\rho.$$

Distribución muestral con muestras de tamaño 2

Número de aciertos: x	Probabilidad: $f(x)$	$xf(x)$	$x^2f(x)$
0	$P(F)P(F) = \rho^2$	0	0
1	$P(S)P(F) + P(F)P(S) = 2\pi\rho$	$2\pi\rho$	$2\pi\rho$
2	$P(S)P(S) = \pi^2$	$2\pi^2$	$4\pi^2$
Suma	$(\rho + \pi)^2 = 1$	$2\pi(\rho + \pi)$	$2\pi(\rho + 2\pi)$

Obsérvese que sólo se puede obtener «ningún acierto» si tanto la primera observación como la segunda son fallos. Como cada unidad extraída es aleatoria, los resultados de la primera y segunda observación son independientes y, por

lo tanto, la probabilidad de que aparezcan dos fallos es igual al producto $P(F)P(F)$. El mismo razonamiento se aplica a la probabilidad de que aparezcan dos aciertos. Sin embargo, «un acierto y un fallo» se pueden obtener de dos formas distintas, ya sea un acierto seguido de un fallo o un fallo en primer lugar y un acierto a continuación. Los dos resultados se excluyen mutuamente, de modo que la probabilidad de un acierto y un fallo es $P(S)P(F) + P(F)P(S)$ tal como se ha indicado.

Media y variancia de X :

$$E(X) = 2\pi(\rho + \pi) = 2\pi,$$

$$\text{Var}(X) = 2\pi(\rho + 2\pi) - [2\pi]^2 = 2\pi\rho + 4\pi^2 - 4\pi^2 = 2\pi\rho.$$

Media y variancia de $\hat{\pi}$:

$$E(\hat{\pi}) = E\left(\frac{X}{2}\right) = \pi,$$

$$\begin{aligned}\text{Var}(\hat{\pi}) &= \text{Var}\left(\frac{X}{2}\right) = \frac{1}{4} \text{Var}(X) \quad (\text{según el Teorema 7}) \\ &= \frac{\pi\rho}{2}.\end{aligned}$$

Distribución muestral con muestras de tamaño 3

Número de aciertos:

Probabilidad:

0	$f(x)$	$xf(x)$	$x^2f(x)$
0	$P(S)P(F)P(F) + P(F)P(S)P(F) + P(F)P(F)P(S) = 3\pi\rho^2$	$3\rho^2\pi$	$3\rho^2\pi$
1	$P(S)P(S)P(F) + P(S)P(F)P(S) + P(F)P(S)P(S) = 3\pi^2\rho$	$6\rho\pi^2$	$12\rho\pi^2$
2	$P(F)P(F)P(F) = \pi^3$	$3\pi^3$	$9\pi^3$
Suma $(\rho + \pi)^3 = 1$			

Media y variancia de X :

$$E(X) = 3\rho^2\pi + 6\rho\pi^2 + 3\pi^3 = 3\pi(\rho^2 + 2\rho\pi + \pi^2) = 3\pi(\rho + \pi)^2 = 3\pi.$$

$$\begin{aligned}\text{Var}(X) &= 3\rho^2\pi + 12\rho\pi^2 + 9\pi^3 - (3\pi)^2 = 3\rho^2\pi + 12\rho\pi^2 + 9\pi^3 - 9\pi^2 \\ &= 3\rho^2\pi + 12\rho\pi^2 - 9\pi^2 = 3\rho\pi(\rho + 4\pi - 3\pi) = 3\rho\pi.\end{aligned}$$

Media y variancia de $\hat{\pi}$:

$$E(\hat{\pi}) = E\left(\frac{X}{3}\right) = \pi.$$

$$\text{Var}(\hat{\pi}) = \text{Var}\left(\frac{X}{3}\right) = \frac{1}{9} \text{Var}(X) = \frac{\pi\rho}{3}.$$

Podríamos seguir con este proceso utilizando muestras cada vez mayores, pero los resultados ya son evidentes: las probabilidades de obtener 0, 1, 2, ..., n aciertos en una muestra de tamaño n vienen dadas por los términos $(\rho + \pi)^n$ res-

Cuadro 4-1

x	$f(x)$				
	Tamaño 1	Tamaño 2	Tamaño 3	Tamaño 4...	Tamaño n
0	ρ	ρ^2	ρ^3	ρ^4	ρ^n
1	π	$2\rho\pi$	$3\rho^2\pi$	$4\rho^3\pi$	$\binom{n}{1}\rho^{n-1}\pi$
2		π^2	$3\rho\pi^2$	$6\rho^2\pi^2$	$\binom{n}{2}\rho^{n-2}\pi^2$
3			π^3	$4\rho\pi^3$	$\binom{n}{3}\rho^{n-3}\pi^3$
4				π^4	$\binom{n}{4}\rho^{n-4}\pi^4$
$\vdots$					$\vdots$
$n-1$					$\binom{n}{n-1}\rho\pi^{n-1}$
n					π^n

pectivos, es decir la llamada *expansión binomial*. El cuadro 4-1 presenta en forma resumida las distribuciones de X correspondientes a varios tamaños muestrales. Obsérvese que el primer elemento de la última columna se podría escribir también en la forma

$$\rho^n = \binom{n}{0}\rho^n\pi^0,$$

y el último elemento de esta misma columna, en la forma

$$\pi^n = \binom{n}{n}\rho^0\pi^n.$$

Por tanto, la probabilidad de obtener x aciertos en una muestra de tamaño n es

$$f(x) = \binom{n}{x}\rho^{n-x}\pi^x.$$

Del mismo modo, la probabilidad de obtener una proporción de aciertos π en una muestra de tamaño n es

$$(4.2) \quad f(\hat{\pi}) = \binom{n}{n\hat{\pi}} p^{n(1-\hat{\pi})} \pi^{n\hat{\pi}}$$

Estas probabilidades se pueden interpretar como los límites de las frecuencias relativas, es decir, las frecuencias que se obtendrían si tomásemos un número infinito de muestras de cada tamaño. Las distribuciones definidas por dichas probabilidades son, pues, las verdaderas (y exactas) distribuciones muestrales: (4.1) define la distribución muestral del número de aciertos, y (4.2) la distribución muestral de la proporción de aciertos.

El cálculo de las probabilidades individuales de la distribución binomial resulta muy laborioso a menos que el tamaño de la muestra sea muy pequeño. Si se utiliza el *triángulo de Pascal* (figura 3-3), que proporciona los valores de

$$\binom{n}{1}, \binom{n}{2}, \binom{n}{3}, \text{ etcétera.}$$

para cada n , nos ahorramos parte del trabajo, pero es posible eliminar todos los cálculos si utilizamos las tablas de *probabilidades binomiales* que nos dan las probabilidades correspondientes a los distintos valores de n y π . Para los valores grandes de n , es posible obtener una aproximación razonablemente precisa de la distribución binomial, por medio de la llamada *distribución normal* (salvo en el caso en que π es muy pequeña y $n\pi$ se mantiene constante cuando n tiende a infinito). En la sección 4.2 estudiaremos esta aproximación de forma detallada.

Obtenemos ahora las características básicas de la distribución muestral de X y de π . En especial, nos interesa la media, la variancia y la asimetría de dichas distribuciones. En el cuadro 4-2 se resumen la media y la variancia co-

11. Dichas tablas se encuentran en algunos manuales de estadística, entre ellos, John E. Freund, *Mathematical Statistics*. Englewood Cliffs, N. J., Prentice-Hall, 1962.

Cuadro 4-2

Tamaño muestral	Media		Variancia	
	Número de aciertos: x	Proporción de aciertos: $\hat{\pi}$	Número de aciertos: x	Proporción de aciertos: $\hat{\pi}$
1	π	π	πp	πp
2	2π	π	$2\pi p$	$\pi p/2$
3	3π	π	$3\pi p$	$\pi p/3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$n\pi$	π	$n\pi p$	$\pi p/n$

respondientes a las muestras de tamaño 1, 2 y 3. La generalización para las muestras de tamaño n resulta inmediata. Sería posible probar la validez de esta generalización sin gran dificultad, pero la demostración acarrearía un elevado número de fastidiosas operaciones algebraicas y, por tanto, la omitimos. Para la medida de asimetría, podemos utilizar α_3 , que se define como

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}},$$

en la que μ_2 es la variancia y μ_3 es el momento de tercer orden respecto de la media. En este caso, se puede demostrar que la expresión correspondiente a α_3 , tanto para la distribución de X como para la de π , es

(4.3)

$$\alpha_3 = \frac{2p - 1}{\sqrt{n\pi p}}.$$

En el cuadro 4-3 se resumen las características básicas de las dos distribuciones muestrales.

Cuadro 4-3

Característica	Distribución de X	Distribución de π
Media	$n\pi$	π
Variancia	$n\pi p$	$\pi p/n$
α_3 (asimetría)	$(2p - 1)/\sqrt{n\pi p}$	$(2p - 1)/\sqrt{n\pi p}$

En este momento disponemos de un conocimiento completo de la distribución muestral de $\hat{\pi}$ para cualquier parámetro específico π de la población y cualquier muestra de tamaño n , y, por tanto, podemos obtener conclusiones concretas respecto a las *propiedades* de $\hat{\pi}$ (la proporción de aciertos en la muestra) como estimador de π (la proporción de aciertos o la probabilidad de obtener un acierto, en la población). Las principales conclusiones son las siguientes

1. La media de la distribución muestral de $\hat{\pi}$ es igual al parámetro de la población π . Es decir, $\hat{\pi}$ es un *estimador insesgado* de π .
2. La desviación standard de la distribución muestral de $\hat{\pi}$ es

(4.4)

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi p}{n}},$$

y, por tanto, la distribución se hace cada vez más concentrada a medida que aumenta el tamaño de la muestra. Este hecho, unido a la conclusión anterior, implica que $\hat{\pi}$ es un estimador consistente de π .

3. La fórmula correspondiente a $\sigma_{\hat{\pi}}$ indica que la dispersión de la distribución muestral de $\hat{\pi}$ (medida por su desviación standard) es inversamente proporcional a la raíz cuadrada del tamaño de la muestra. Podemos considerar este hecho de la forma siguiente: tomemos dos muestras, la primera de ellas de tamaño n , y la segunda de tamaño $2n$. Si utilizamos la fórmula adecuada, que aparece en el cuadro 4-3, tenemos que

$$\frac{\sigma_{\hat{\pi}}(\text{segunda muestra})}{\sigma_{\hat{\pi}}(\text{primera muestra})} = \frac{\sqrt{\pi\rho/2n}}{\sqrt{\pi\rho/n}} = \frac{1}{\sqrt{2}};$$

$$\sigma_{\hat{\pi}}(\text{segunda muestra}) = \frac{\sigma_{\hat{\pi}}(\text{primera muestra})}{\sqrt{2}}$$

En otras palabras, al multiplicar por k el tamaño de la muestra, la desviación standard de la distribución muestral queda dividida por $\sqrt{k}$.

4. Dado el tamaño de la muestra, la distribución muestral de $\hat{\pi}$ tiene una dispersión máxima cuando el parámetro de la población π es igual a $1/2$ y una dispersión mínima cuando π es 0 ó 1. Esto se deriva del hecho de que $0 \leq \pi \leq 1$ y, por tanto,

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi\rho}{n}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

es máxima cuando $\pi = 1/2$, y mínima cuando $\pi = 0$ ó 1. En realidad, el valor máximo que $\sigma_{\hat{\pi}}$ puede tomar es $1/(2\sqrt{n})$, y su valor mínimo es 0.

5. La asimetría de la distribución muestral de $\hat{\pi}$ es inversamente proporcional a la raíz cuadrada del tamaño de la muestra. Ello se deriva, lógicamente, de la ecuación (4.3).
6. Para un tamaño muestral dado, la asimetría de la distribución muestral de $\hat{\pi}$ es mínima cuando π es igual a $1/2$, y es máxima cuando π toma el valor 0 ó 1. De nuevo, esto puede verse a partir de la ecuación (4.3): α_3 es cero cuando $\pi = 1/2$ y toma un valor máximo cuando π es cero o uno.

Estas conclusiones se han obtenido mediante la ayuda de la teoría de la probabilidad y su aplicación es totalmente universal. En la sección 2.1 vamos de obtener la distribución muestral de $\hat{\pi}$ a partir de un experimento

lacionado con una población con $\pi = 0,70$. El experimento consistió en formar 100 muestras de tamaño 4 y 100 muestras de tamaño 16, las cuales nos permitieron construir distribuciones de frecuencias (una para cada tamaño muestral) que supusimos se aproximaban a las verdaderas distribuciones muestrales. Los resultados nos permitieron efectuar generalizaciones acerca de la distribución muestral de π para cualquier π y cualquier tamaño muestral. En este momento podemos, en primer lugar, comprobar aquellos resultados con el fin de estudiar en qué medida las distribuciones obtenidas de forma experimental se ajustaban a las distribuciones teóricas y, en segundo lugar, verificar o refutar la validez de las generalizaciones.

Comparemos en primer lugar las distribuciones experimentales y teóricas, y sus características.

Distribución muestral con muestras de tamaño 4 (cuadro 4-4). Los resultados correspondientes a las probabilidades obtenidas de forma experimental

Cuadro 4-4

Proporción de aciertos: $\hat{\pi}$	$f(\hat{\pi})$ en los experimentos	$f(\hat{\pi})$ teórica
0.00	0.01	0.01
0.25	0.06	0.07
0.50	0.28	0.26
0.75	0.42	0.41
1.00	0.23	0.24
Media	0.7	0.7
Desviación Standard	0.2233	0.2291

son las frecuencias relativas del cuadro 2-1. Las probabilidades teóricas se obtuvieron a partir del libro de Freund,² tomando sólo dos decimales, y pueden calcularse de la forma siguiente:

$$f(0) = \left(\frac{4!}{4!0!}\right)(0.3)^4(0.7)^0 = (0.3)^4 = 0.0081 \approx 0.01,$$

$$f\left(\frac{1}{4}\right) = \left(\frac{4!}{3!1!}\right)(0.3)^3(0.7)^1 = 4 \times (0.3)^3 \times (0.7) = 0.0756 \approx 0.07,$$

etcétera.

² Ibid. tabla 1.

Distribución muestral con muestras de tamaño 16 (cuadro 4-5). Los resultados correspondientes a las probabilidades obtenidas de forma experimental se han tomado del cuadro 2-2. Las probabilidades teóricas tienen el mismo origen que las del cuadro 4-4, tomando nuevamente dos decimales.

Es evidente que los resultados experimentales describen con gran precisión las distribuciones muestrales; los errores son muy pequeños, y es muy poco probable que tengan importancia. Por tanto, en este caso, las distribuciones basadas en 100 muestras se aproximan mucho a las que se obtendrían a partir de un número infinito de muestras. Además, las generalizaciones efectuadas en base a los resultados experimentales —es decir, insesgadez, consistencia, variación relativa de la desviación standard, y disminución de la asimetría— resultan ser correctas todas ellas. Sin embargo, si los comparamos con los resultados teóricos, los resultados experimentales están en desventaja en dos sentidos. En primer lugar, no nos proporcionan ninguna fórmula para el cálculo de la variancia, la medida de asimetría, ni, naturalmente, las probabilidades individuales. En segundo lugar, y ello es mucho más importante, no tenemos ninguna garantía de que las generalizaciones efectuadas a partir de las distribuciones experimentales sean válidas en realidad, ya que los resultados de experimentos aislados no pueden probar conclusión alguna, y pueden sólo sugerirlas.

Cuadro 4-5

Proporción de aciertos: π	$f(\pi)$ en los experimentos	$f(\pi)$ teórica
0	0	0
1/16	0	0.00
2/16	0	0.00
3/16	0	0.00
4/16	0	0.00
5/16	0.01	0.00
6/16	0	0.01
7/16	0.01	0.02
8/16	0.05	0.05
9/16	0.10	0.10
10/16	0.17	0.17
11/16	0.21	0.21
12/16	0.20	0.20
13/16	0.15	0.15
14/16	0.07	0.07
15/16	0.03	0.02
1	0	0.00
Media	0.7006	0.7
Desviación standard	0.1191	0.1146

4.2 La distribución normal como caso limite de la distribución binomial

Uno de los resultados de la sección 4.1 fue el descubrimiento de que la distribución binomial tiende a ser cada vez más simétrica a medida que aumenta n (el tamaño de la muestra), independientemente del valor de π . Incluso las distribuciones en las que π toma un valor próximo a cero (o a la unidad), que son muy asimétricas cuando n es pequeña, tienden a hacerse simétricas cuando n es mayor. La figura 4-1, en la que se representan las probabilidades binomiales correspondientes a $\pi = 0,10$ para distintos valores de n , demuestra este resultado. Obsérvese también que, al aumentar n , los puntos son cada vez más numerosos y la línea que los une se hace cada vez más redondeada.

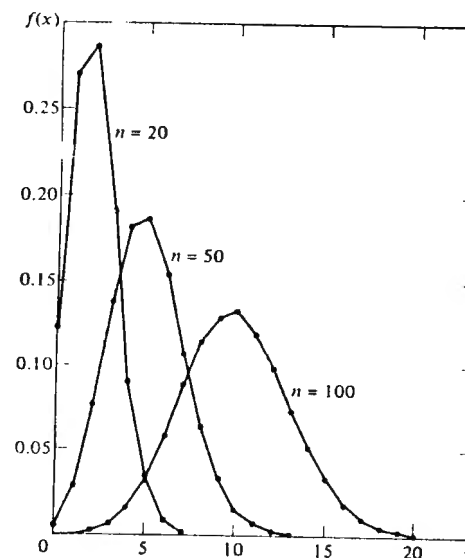


Figura 4-1

En este apartado ampliaremos este concepto afirmando que, a medida que n tiende a infinito, la distribución binomial se aproxima a la llamada *distribución normal*.³ La distribución normal es una distribución continua cuya densidad de probabilidad viene dada por la fórmula

$$(4.5) \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}[(x-\mu)/\sigma]^2},$$

3. Esta demostración se encuentra en muchos libros de texto de estadística. Una demostración relativamente simple es la que aparece en G. Udny Yule y M. G. Kendall, *An Introduction to the Theory of Statistics*. Londres, Charles Griffin, 1950, pp. 177-81.

en la que σ = desviación standard de X ,

μ = media de X ,

$\pi = 3,14159...$ (no debe confundirse con el parámetro de la población π),

$e = 2,71828...$

En la figura 4-2 se representa gráficamente una distribución normal, y se puede ver que la distribución es simétrica alrededor de la media μ y se extiende desde $-\infty$ a $+\infty$. Algo más adelante se estudian las demás propiedades.

El hecho de que cuando n tiende a infinito la distribución binomial se acerca a la normal significa que, en realidad, cuando n es grande podemos utilizar la distribución normal como aproximación a la binomial y, en este caso, si llamamos X al número de aciertos en la muestra, podemos escribir

$$(4.6) \quad f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n\pi p}} e^{-\frac{1}{2}[(x - n\pi)/\sqrt{n\pi p}]^2}$$

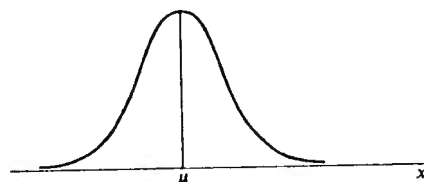


Figura 4-2

ya que la media de X es $n\pi$ y la variancia es $n\pi p$. Del mismo modo, si n es grande, la función de probabilidad de la proporción de aciertos en la muestra ($\hat{\pi}$) se puede representar mediante la ecuación

$$(4.7) \quad f(\hat{\pi}) = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{\pi p}} e^{-\frac{1}{2}[(\hat{\pi} - \pi)/\sqrt{\pi p/n}]^2},$$

ya que la media de $\hat{\pi}$ es π y su variancia es $\pi p/n$. La exactitud de estas aproximaciones depende de n y π . Si π no difiere mucho de $1/2$, la correspondencia entre la curva normal y la binomial resulta muy grande incluso para valores pequeños de n . En general, en la mayor parte de las ocasiones prácticas se puede utilizar la distribución normal como aproximación aceptable de la binomial sin dudarle mucho siempre que $n \geq 30$.

Al utilizar la fórmula de la distribución normal como aproximación de una binomial debemos tener en cuenta el hecho de que estamos tratando de obtener una aproximación de una distribución *discreta* mediante una distribución

continua. Para ello podemos representar los valores de la variable discreta a través de los intervalos en torno a cada uno de dichos puntos. Por ejemplo, si $n = 20$, los puntos y sus correspondientes intervalos serían los que aparecen

Cuadro 4-6

x (Discreta)	x (Aproximación continua)	$\hat{\pi}$ (Discreta)	$\hat{\pi}$ (Aproximación continua)
0	$-\frac{1}{2}$ a $\frac{1}{2}$	0	$-1/40$ a $1/40$
1	$\frac{1}{2}$ a $1\frac{1}{2}$	$1/20$	$1/40$ a $3/40$
2	$1\frac{1}{2}$ a $2\frac{1}{2}$	$2/20$	$3/40$ a $5/40$
3	$2\frac{1}{2}$ a $3\frac{1}{2}$	$3/20$	$5/40$ a $7/40$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
20	$19\frac{1}{2}$ a $20\frac{1}{2}$	1	$39/40$ a $41/40$

en el cuadro 4-6. La distribución normal se extiende, en realidad, desde $-\infty$ hasta $+\infty$ y, por tanto, se puede tomar el primer intervalo a partir de $-\infty$ y el último puede acabar en $+\infty$ en lugar de tomar los límites inferior y superior que aparecen en el cuadro.

La reconciliación entre los valores discretos de la distribución binomial y los continuos de la normal es más fácil a medida que aumenta el tamaño de la muestra. Esto se puede ver de forma especialmente clara en el caso de la distribución de $\hat{\pi}$. Al aumentar n , los intervalos correspondientes a cada valor de $\hat{\pi}$ se van haciendo más pequeños. Como ejemplo, consideremos los intervalos correspondientes al punto $\hat{\pi} = 0,1$ para varios tamaños muestrales:

n	Intervalo
10	0,05 a 0,15
20	0,075 a 0,125
50	0,090 a 0,110
100	0,095 a 0,105
1000	0,0995 a 0,1005
	etc.

Obsérvese que, por ejemplo, cuando $n = 1000$, sustituimos la probabilidad de que $\hat{\pi}$ sea igual a $0,1$ por la probabilidad de que $\hat{\pi}$ esté comprendida en el intervalo $0,1 \pm 0,0005$, lo cual, evidentemente, no es una mala aproximación. Si representamos las probabilidades binomiales de $\hat{\pi}$ por medio de rectángulos cuya base sea el intervalo apropiado y la altura igual a $n f(\hat{\pi})$ (para que el área total sea igual a la unidad), podemos ver que la línea quebrada se va

haciendo cada vez más suave cuando n se va haciendo mayor. En la figura 4-3 podemos ver este proceso para un valor de π igual a 0,1.

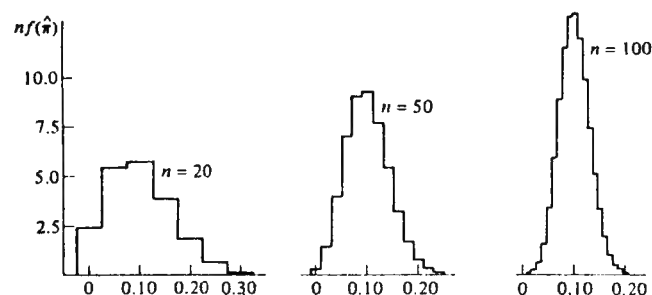


Figura 4-3

Al afirmar que la distribución binomial converge hacia la normal cuando n tiende a infinito, damos por supuesto que el parámetro de la población π (y por tanto p) es constante y distinto de 0 y de la unidad. Cuando π no es constante, sino que disminuye a medida que $n \rightarrow \infty$ (de modo que $n\pi$ permanece constante), en el límite, la distribución binomial no toma la forma de la normal, sino que se convierte en lo que se conoce con el nombre de *distribución de Poisson*. Esta distribución se hizo famosa debido a que se ajustaba extraordinariamente bien a la frecuencia de muertes debidas a la cox de un caballo en el ejército prusiano a finales del siglo XIX. La distribución de Poisson no tiene gran importancia para nuestros fines ya que limitamos nuestra atención a las poblaciones cuya proporción de aciertos (probabilidad) es fija. En los casos en que π es igual a cero o a la unidad (es decir, cuando la población está formada únicamente de aciertos o de fallos), la proporción de aciertos (π) es la misma en todas las muestras e igual a π . Ello significa que las distribuciones muestrales se hallan concentradas en un solo punto, independientemente del tamaño de la muestra y, por tanto, no se plantea la cuestión de la convergencia hacia la distribución normal.

En econometría, la distribución normal es muy importante, no solamente debido al hecho de que representa la forma límite de la distribución binomial, sino también porque se aplica a muchas más situaciones. En la sección siguiente y en otros puntos a lo largo del libro se verá claramente este hecho, y por ello resulta conveniente estudiar la distribución normal con mayor detalle. Definamos, en primer lugar, sus rasgos principales.

1. La distribución es *continua* y *simétrica* en torno a su media μ . Esto tiene las implicaciones siguientes: a) la media, la mediana y la moda

- son iguales; y b) la media divide al área comprendida entre la curva normal y el eje de abscisas en dos mitades exactamente iguales.
- El recorrido de la distribución se extiende desde $-\infty$ hasta $+\infty$, es decir, la distribución *no está limitada*.
- La altura máxima de la curva normal es la correspondiente al punto $x = \mu$ y los puntos de inflexión (es decir, los puntos en que la distribución empieza a hacerse cada vez más plana) son los correspondientes a $x = \mu \pm \sigma$. Ello significa que la desviación standard mide la distancia comprendida entre el centro de la distribución y un punto de inflexión como puede verse en la figura 4-4.

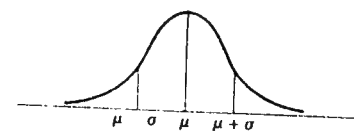


Figura 4-4

4. La distribución normal se define por completo mediante dos parámetros, la media y la variancia. Esto significa que, si conocemos μ y σ^2 correspondientes a una distribución normal, sabemos todo lo que hay que saber respecto a ella. Obsérvese que la distribución binomial también se especifica totalmente mediante dos parámetros, π y n . La figura 4-5 presenta varias comparaciones de dos distribuciones normales. El caso a) representa dos distribuciones normales con distintas medias, pero con la misma variancia; en el caso b) las medias son iguales, pero las variancias son distintas, y en c) tanto las medias como las variancias son distintas.

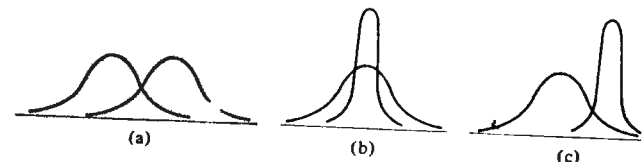


Figura 4-5

5. La última característica que mencionamos resultará especialmente útil más adelante, y la presentamos en forma de teorema:

Teorema 12. Si $X, Y, Z, \dots$, son variables aleatorias independientes, y se dis-

tribuyen normalmente, y $a, b, c, \dots$, son constantes, la combinación lineal $aX + bY + cZ \dots$ también se distribuye normalmente.

La demostración de este teorema se puede hallar en muchos libros de texto de estadística matemática⁴.

Una vez enumeradas las principales propiedades de la distribución normal, aún tenemos que enfrentarnos con el problema de cómo operar con ella. En especial, estamos interesados en calcular las distintas probabilidades de una variable que se distribuye normalmente. En el caso de la distribución binomial, esto no presentaba, en principio, ningún problema, ya que los términos de la expansión binomial se determinaban mediante una simple fórmula. Sin embargo, en el caso de la distribución normal, la fórmula de la densidad de probabilidad es muy complicada y, afortunadamente, no tenemos necesidad de utilizarla; las probabilidades, representadas por las áreas correspondientes entre la curva y el eje de abscisas, se pueden obtener a partir de tablas. Naturalmente, a partir de distribuciones normales distintas se obtendrán probabilidades distintas, pero estas diferencias sólo pueden ser debidas a las diferencias entre las medias y las variancias, y, por tanto, no presentan ninguna dificultad. Si conocemos las áreas correspondientes a una curva normal determinada, podemos obtener las correspondientes a cualquier otra normal considerando simplemente la diferencia entre las medias y las variancias. La distribución concreta cuyas áreas (correspondientes a intervalos relativamente pequeños) se han tabulado es una distribución normal con media $\mu = 0$ y variancia $\sigma^2 = 1$, y recibe el nombre de *distribución normal standarizada o tipificada* (también llamada a veces *distribución normal unitaria*).

El problema de calcular las probabilidades correspondientes a una variable X que se distribuye normalmente se puede enunciar de la forma siguiente: ¿cómo podemos calcular la probabilidad de que x esté comprendida en el intervalo limitado por los puntos x_1 y x_2 , si conocemos a) μ y σ^2 de X , y b) las áreas comprendidas entre la curva normal tipificada y el eje de abscisas?

Para obtener la solución, introduciremos la notación siguiente:

Z = variable que se distribuye normalmente con media cero y variancia igual a la unidad (es decir, «variable normal tipificada»);

$P(x_1 < x < x_2)$ = probabilidad de que X esté comprendida entre x_1 y x_2 ($x_1 < x_2$);

$P(z_1 < z < z_2)$ = probabilidad de que Z esté comprendida entre z_1 y z_2 ($z_1 < z_2$).

Para ello procederemos en dos etapas: en primer lugar, obtendremos la rela-

4. Véase, por ejemplo, R. L. Anderson y T. A. Bancroft, *Statistical Theory in Economics*, Nueva York, McGraw-Hill, 1952, pp. 53-64.

ción entre X y Z , y a continuación examinaremos la relación entre las correspondientes áreas limitadas por las curvas.

Sabemos que X se distribuye normalmente y, por tanto, las funciones lineales de X serán también normales (ver teorema 12). En general, podemos representar estas funciones lineales en la forma

$$aX + b,$$

en la que a y b son constantes. Si obtenemos unos valores de a y b que hagan que la media de $(aX + b)$ sea cero y su variancia sea la unidad, habremos obtenido una variable normal tipificada. Es decir, pretendemos que

$$E(aX + b) = 0 \quad \text{y} \quad \text{Var}(aX + b) = 1.$$

Podemos escribir estas condiciones en la forma

$$a\mu + b = 0 \quad (\text{según el teorema 6})$$

$$a^2\sigma^2 = 1 \quad (\text{según el teorema 7}).$$

Despejando a y b obtenemos

$$a = \frac{1}{\sigma} \quad \text{y} \quad b = -\frac{\mu}{\sigma}.$$

Y, por tanto,

$$aX + b = \frac{X - \mu}{\sigma}.$$

La variable $(X - \mu)/\sigma$ tiene una media cero y su variancia es igual a la unidad, y, por tanto, se trata de una variable normal tipificada, o sea que

$$(4.8) \quad \frac{X - \mu}{\sigma} = Z.$$

Por tanto, podemos transformar cualquier variable normal con media μ y variancia σ^2 en una variable normal tipificada, si la expresamos en términos de desviaciones respecto de su media, dividiendo cada desviación por σ .

Consideremos ahora $P(x_1 < x < x_2)$, con $x_1 < x_2$, o sea, una afirmación en términos de probabilidad acerca de X , e intentemos obtener otra afirmación acerca de Z que sea exactamente equivalente. Evidentemente,

$$\frac{X - \mu}{\sigma} = Z \quad \text{implica} \quad X = \sigma Z + \mu.$$

Por tanto, podemos escribir

$$x_1 = \sigma z_1 + \mu \quad \text{y} \quad x_2 = \sigma z_2 + \mu.$$

Y sustituyendo, vemos que

$$P(x_1 < x < x_2) = P(\sigma z_1 + \mu < \sigma z + \mu < \sigma z_2 + \mu).$$

Simplificando la expresión anterior obtenemos

$$(4.9) \quad P(x_1 < x < x_2) = P(z_1 < z < z_2),$$

en la que
$$z_1 = \frac{x_1 - \mu}{\sigma} \quad \text{y} \quad z_2 = \frac{x_2 - \mu}{\sigma}.$$

Es decir, podemos ver que la probabilidad de que x esté comprendida entre x_1 y x_2 es igual a la probabilidad de que una variable normal tipificada esté comprendida entre los puntos $(x_1 - \mu)/\sigma$ y $(x_2 - \mu)/\sigma$.

A modo de ejemplo, consideremos una variable X que se distribuye normalmente con media 5 y desviación standard igual a 2. El problema consiste en obtener la probabilidad $P(2 < x < 3)$. Para ello debemos obtener una afirmación equivalente en términos de probabilidad, correspondiente a la variable normal tipificada Z . En este caso, el límite inferior x_1 es 2 y $x_2 = 3$. Sabiendo que

$$z_1 = \frac{x_1 - \mu}{\sigma},$$

tenemos que

$$z_1 = \frac{2 - 5}{2} = -\frac{3}{2}.$$

Y del mismo modo,
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{3 - 5}{2} = -1.$$

Por tanto, $P(2 < x < 3) = P(-3/2 < z < -1)$. Este resultado se representa gráficamente en la figura 4-6. Las dos áreas en negro debajo de las curvas son exactamente iguales.

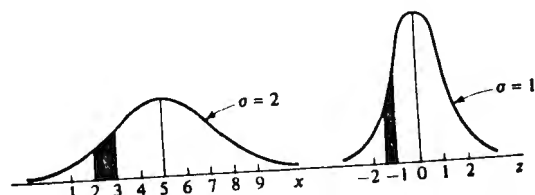


Figura 4-6

Tras estos prolegómenos, lo único que queda por hacer es aprender a utilizar las tablas que indican las áreas correspondientes a la distribución normal tipificada. Esta tabla se puede hallar prácticamente en todos los libros de texto de introducción a la estadística. La forma de presentación más corriente consiste en indicar las probabilidades de que Z tome un valor comprendido entre 0 y un número positivo z_0 (tomando dos decimales) tal como se repre-

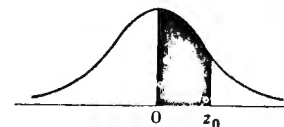
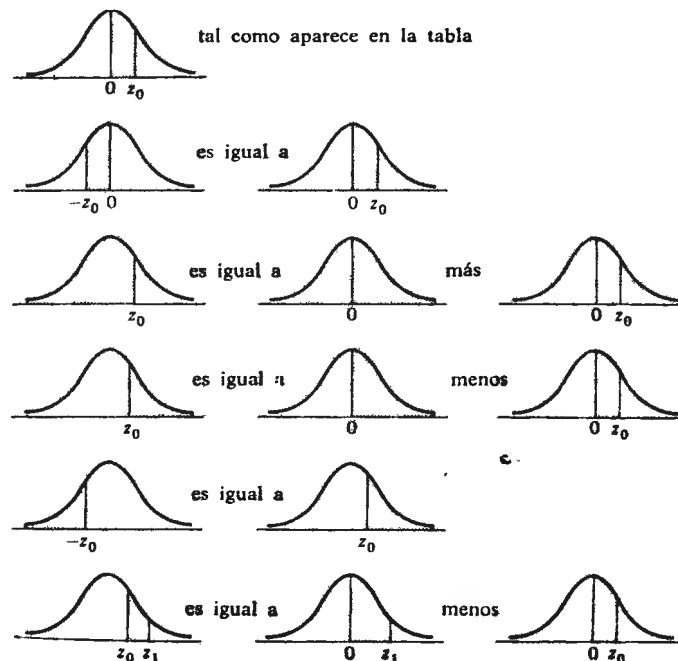


Figura 4-7

senta gráficamente en la figura 4-7. La tabla incluida en el Apéndice D de este libro es de este tipo. Las probabilidades indicadas sólo hacen referencia a una mitad de la distribución, pero como ésta es perfectamente simétrica, ello resulta totalmente suficiente. En este caso, la probabilidad (área) mayor será 0,5



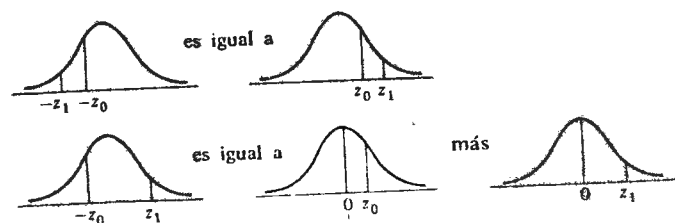


Figura 4-8

correspondiendo al punto en que $z = +\infty$. Sin embargo, como la cola de la distribución disminuye al principio con rapidez, la probabilidad de que z esté comprendida entre 0 y 3 ya es casi 0,5 (en realidad, es 0,4987) y, debido a ello, la mayor parte de las tablas acaban en este punto. (Algunas tablas proporcionan la probabilidad de que z esté comprendida entre $-\infty$ y un número positivo; en este tipo de tablas la probabilidad menor que aparece es 0,5 y la mayor, es próxima a 1.) En la figura 4-8 se indica la forma de utilizar las tablas para la resolución de distintos problemas. Obsérvese que $z_0 > 0$ y $z_1 > z_0$.

Veamos ahora dos ejemplos del uso de las tablas que indican el área que se halla debajo de la curva normal tipificada. El primer ejemplo nos da información que será útil más adelante, y el segundo hace referencia a la relación entre la distribución normal y la binomial en un caso específico que también tiene interés para nuestro análisis.

EJEMPLO 1. Siendo X una variable aleatoria normal con media μ y variancia σ^2 , calcular:

- $P(\mu - \sigma < x < \mu + \sigma)$,
- $P(\mu - 2\sigma < x < \mu + 2\sigma)$,
- $P(\mu - 3\sigma < x < \mu + 3\sigma)$.

- los dos valores de X que delimitan la parte central del área situada debajo de la curva, abarcando un 95 % de ella,
- los dos valores de X que delimitan la parte central del área situada debajo de la curva, abarcando un 99 % de ella.

Las respuestas son las siguientes.



Figura 4-9

a) Queremos obtener el área representada en la figura 4-9. Los límites inferior y superior, expresados en términos de la variable normal tipificada, son los siguientes

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{(\mu - \sigma) - \mu}{\sigma} = -1$$

y

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{(\mu + \sigma) - \mu}{\sigma} = +1.$$

Por tanto,

$$P(-1 < z < +1) = P(0 < z < +1) + P(0 < z < +1) = 0.3413 + 0.3413 = 0.6826.$$

Es decir, la probabilidad de que x tome un valor comprendido en el intervalo limitado por una desviación standard en cada sentido, a partir de la media, es de 68,26 %.

b)

$$\begin{aligned} P(\mu - 2\sigma < x < \mu + 2\sigma) &= P\left(\left[\frac{(\mu - 2\sigma) - \mu}{\sigma}\right] < z < \left[\frac{(\mu + 2\sigma) - \mu}{\sigma}\right]\right) \\ &= P(-2 < z < +2) = 0.4772 + 0.4772 = 0.9544. \end{aligned}$$

Es decir, la probabilidad de que x esté comprendida en el intervalo formado por dos desviaciones standard en cada sentido, a partir de la media, es de 95,44 %.

c)

$$\begin{aligned} P(\mu - 3\sigma < x < \mu + 3\sigma) &= P\left(\left[\frac{(\mu - 3\sigma) - \mu}{\sigma}\right] < z < \left[\frac{(\mu + 3\sigma) - \mu}{\sigma}\right]\right) \\ &= P(-3 < z < +3) = 0.4987 + 0.4987 = 0.9974. \end{aligned}$$

Es decir, la probabilidad de que x esté comprendida en el intervalo formado por tres desviaciones standard en cada sentido, a partir de la media, es de 99,74 %. En otros términos, prácticamente la totalidad de los valores de X se hallan limitados a un intervalo cuya longitud es igual a seis desviaciones standard; naturalmente, el punto central de este intervalo es la media.

Todos los problemas anteriores han sido casos en los que conocíamos los valores de X y hemos obtenido las áreas situadas debajo de la curva y limitadas por dichos valores. Los dos problemas siguientes son del tipo contrario: en esta ocasión conocemos el área y queremos hallar los valores de X que la limitan.

d) En este caso sabemos que $P(x_1 < x < x_2) = 0.95$, y que el intervalo entre x_1 y x_2 se halla centrado en torno a la media. El problema consiste en hallar x_1 y x_2 . Para resolverlo, calculamos en primer lugar los límites correspondientes en el caso de la distribución normal tipificada. Dada la afirmación en términos de probabilidad correspondiente a X , la correspondiente a la variable normal tipificada Z es

$P(z_1 < z < z_2) = 0,95$, con el intervalo entre z_1 y z_2 centrado en torno al valor 0 (figura 4-10). Al estar centrado en torno a cero, tenemos

$$P(0 < z < z_2) = P(z_1 < z < 0) = \frac{0,95}{2} = 0,475.$$

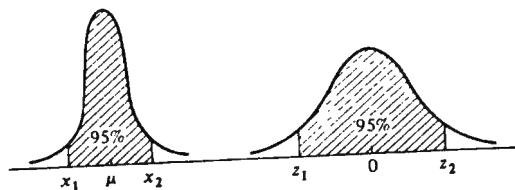


Figura 4-10

Buscando en las columnas del centro de la tabla que indica las áreas de la distribución normal tipificada, vemos que el valor de Z correspondiente al área 0,475 es 1,96. Así, pues,

$$z_2 = 1,96,$$

$$z_1 = -1,96.$$

Y en consecuencia,

$$\frac{x_2 - \mu}{\sigma} = 1,96,$$

$$\frac{x_1 - \mu}{\sigma} = -1,96.$$

De ello se deriva que

$$x_2 = \mu + 1,96 \sigma$$

y

$$x_1 = \mu - 1,96 \sigma.$$

Es decir, el intervalo $\mu \pm 1,96\sigma$ contiene el 95 % central de los valores de X .

e)

$$P(x_1 < x < x_2) = 0,99,$$

$$P(z_1 < z < z_2) = 0,99,$$

$$P(0 < z < z_2) = 0,495.$$

El valor de Z correspondiente al área 0,495 es 2,57. Así pues,

$$z_2 = 2,57,$$

$$z_1 = -2,57,$$

y, en consecuencia,

$$x_2 = \mu + 2,57 \sigma$$

y

$$x_1 = \mu - 2,57 \sigma.$$

Es decir, el intervalo $\mu \pm 2,57\sigma$ contiene el 99 % central de los valores de X .

EJEMPLO 2. Consideremos la distribución de la proporción muestral $\hat{\pi}$ correspondiente a muestras de tamaño 16, extraídas de una población con una proporción de aciertos $\pi = 0,7$. Esta distribución es la distribución binomial que obtuvimos anteriormente y aparece en la última columna del cuadro 4-5 (reproduciéndose también en la última columna del cuadro 4-7). Se supone que debemos hallar una aproximación normal a esta distribución. La solución es la siguiente: en primer lugar, debemos reemplazar los puntos $\hat{\pi}$ de la distribución binomial discreta por intervalos pertenecientes a la correspondiente distribución normal. Este proceso se ha explicado al principio de esta sección. El paso siguiente consiste en obtener las probabilidades normales correspondientes a cada intervalo. Como en este caso la distribución binomial de $\hat{\pi}$ tiene una media de 0,7 y una desviación standard igual a 0,11464, la distribución normal que buscamos debe tener como características unos parámetros que tomen dichos valores. Para obtener las probabilidades, debemos efectuar una transformación y pasar a la variable normal tipificada, que viene dada por

$$z = \frac{\hat{\pi} - 0,7}{0,11464} = 8,72872\hat{\pi} - 6,11010.$$

Ello nos permitirá establecer los intervalos en términos de z en lugar de $\hat{\pi}$, obteniendo las probabilidades correspondientes a partir de la tabla de las áreas de la distribución normal tipificada. Los resultados aparecen en el cuadro 4-7, y resulta obvio que la aproximación normal a la distribución binomial se halla muy cercana, a pesar del hecho de que el tamaño de la muestra es únicamente 16.

Al estudiar la distribución normal, comenzamos afirmando que se trata del caso límite de la distribución binomial. Debe señalarse que ésta es sólo una de las formas por las que se puede llegar a la citada distribución, y existen también otras líneas de razonamiento que nos permiten llegar hasta ella⁵. La más importante de estas vías alternativas es la que obtiene la distribución normal basándose en el comportamiento de los errores aleatorios. Por ejemplo, considérese el problema de medir la longitud de un objeto. En circunstancias normales, cada una de las mediciones puede estar sujeta a un cierto error y supongamos que cada error es el resultado de un número indefinidamente grande de causas, cada una de las cuales produce una pequeña desviación. Si suponemos que todas estas pequeñas desviaciones son iguales y que las de signo positivo y las de signo negativo son igualmente probables, se puede demostrar que los errores se distribuyen normalmente alrededor de cero, es decir, que las mediciones se distribuyen normalmente alrededor del valor «verdadero». Na-

5. Para una exposición panorámica de este tema, ver C. R. Rao, *Linear Statistical Inference and Its Applications*. Nueva York, Wiley, 1965, pp. 126 y ss.

Cuadro 4-7

π (Valores discretos)	π (Intervalos)	z (Intervalos)	$F(\pi)$ (Aproxi- mación normal) *	$f(\pi)$ (Aproxi- mación normal) †	$f(\pi)$ (Distri- bución bino- mial)
0	-1/32 a 1/32	$-\infty$ a -5.84	0.0000	0.0000 (0.00)	0.00
1/16	1/32 a 3/32	-5.84 a -5.29	0.0000	0.0000 (0.00)	0.00
2/16	3/32 a 5/32	-5.29 a -4.75	0.0000	0.0000 (0.00)	0.00
3/16	5/32 a 7/32	-4.75 a -4.20	0.0000	0.0000 (0.00)	0.00
4/16	7/32 a 9/32	-4.20 a -3.66	0.0000	0.0000 (0.00)	0.00
5/16	9/32 a 11/32	-3.66 a -3.11	0.0009	0.0009 (0.00)	0.00
6/16	11/32 a 13/32	-3.11 a -2.57	0.0050	0.0041 (0.00)	0.01
7/16	13/32 a 15/32	-2.57 a -2.02	0.0217	0.0167 (0.02)	0.02
8/16	15/32 a 17/32	-2.02 a -1.48	0.0694	0.0477 (0.05)	0.05
9/16	17/32 a 19/32	-1.48 a -0.93	0.1762	0.1068 (0.11)	0.10
10/16	19/32 a 21/32	-0.93 a -0.38	0.3520	0.1758 (0.17)	0.17
11/16	21/32 a 23/32	-0.38 a 0.16	0.5636	0.2116 (0.21)	0.21
12/16	23/32 a 25/32	0.16 a 0.71	0.7611	0.1975 (0.20)	0.20
13/16	25/32 a 27/32	0.71 a 1.25	0.8944	0.1333 (0.13)	0.15
14/16	27/32 a 29/32	1.25 a 1.80	0.9641	0.0697 (0.07)	0.07
15/16	29/32 a 31/32	1.80 a 2.34	0.9904	0.0263 (0.03)	0.02
1	31/32 a 33/32	2.34 a $+\infty$	1.0000	0.0096 (0.01)	0.00

* $F(\pi)$ corresponde al área entre $-\infty$ y el límite superior de cada intervalo.

† $f(\pi)$ es el área correspondiente a cada intervalo. Las cifras entre paréntesis son las probabilidades tomando sólo dos decimales.

turalmente, la base de este procedimiento se puede interpretar de forma completamente general. En particular, se puede considerar que el «error» representa cualquier clase de desviación respecto de cierta conducta sistemática, y ésta es la interpretación subyacente en la mayor parte de los desarrollos teóricos de la econometría moderna. Ello se verá claramente cuando iniciemos el estudio de los modelos de regresión. En este punto debemos mencionar que el uso extensivo de la distribución normal ha dado lugar a la siguiente abreviatura:

$X \sim N(\mu, \sigma^2)$ significa que X es una variable aleatoria que se distribuye normalmente con media μ y variancia σ^2 . En consecuencia, $X \sim N(0,1)$ representa la variable normal tipificada. En adelante, utilizaremos dicha notación.

4.3 Distribución muestral de la media de la muestra

En el caso de la distribución muestral de la proporción de aciertos en la muestra, estudiado en la sección 4.1, nos ocupamos del muestreo a partir de una población de carácter dicotómico. Cada observación se clasifica como

acierto o fallo y consideramos la proporción de aciertos en la muestra como un estimador de la probabilidad de aciertos en la población. Como ya se mencionó anteriormente, en lugar de llamar aciertos o fallos a las observaciones, podemos asignarles valores numéricos, por ejemplo 0 en lugar de fallo y 1 en lugar de acierto. De este modo, en vez de tratar con atributos, nos ocuparemos de una *variable binaria*. Supongamos que llamamos Y a esta variable y que, en una muestra de seis observaciones, observamos los valores siguientes:

$$y_1 = 0,$$

$$y_2 = 1,$$

$$y_3 = 0,$$

$$y_4 = 0,$$

$$y_5 = 1,$$

$$y_6 = 0.$$

En este caso, la proporción de aciertos observada es

$$\hat{\pi} = \frac{0 + 1 + 0 + 0 + 1 + 0}{6} = \frac{2}{6} = \frac{1}{3} \sum_{i=1}^6 y_i = \bar{y}.$$

Es decir, la proporción de aciertos en la muestra no es más que la media muestral de Y . Además, sabemos que la probabilidad de obtener un acierto en la población es el límite de la frecuencia relativa de los aciertos cuando el número de observaciones tiende a infinito. Como el límite de la frecuencia relativa de y es su probabilidad —a la que hemos llamado $f(y)$ — y la observación de un acierto significa que $y = 1$, resulta que

$$\pi = P(\text{acierto}) = P(y = 1) = f(1),$$

o

$$\pi = 1 \times f(1).$$

Esto puede escribirse también en la forma

$$\pi = 0 \times f(0) + 1 \times f(1),$$

ya que cualquier número multiplicado por cero es cero. Pero al escribir el valor de π de esta forma, vemos que en realidad se trata de la media ponderada de los distintos valores de Y (es decir, 0 y 1) y que las ponderaciones vie-

nen dadas por las probabilidades respectivas. Esta es precisamente la definición de la esperanza matemática de Y , tal como la establecimos en la ecuación (3.23). Por lo tanto,

$$\pi = E(Y) = \mu_Y.$$

En otras palabras, la probabilidad de obtener un acierto en la población es la media de la población de Y y, por tanto, la distribución muestral de la proporción de aciertos en la muestra (como estimador de la probabilidad de obtener un acierto) se puede considerar como una distribución muestral de la media de la muestra (como estimador de la media de la población) cuando la variable es binaria.

Llegado a este punto, estamos interesados en la obtención de la distribución muestral de la media de la muestra en los casos en que la variable puede tomar más de dos valores y tiene una distribución cualquiera. En general, cabe esperar que el hecho de que la variable de la población (es decir, las distribuciones «originarias») se distribuya de forma distinta dará lugar a distribuciones muestrales de la media de la muestra también distintas. Por ejemplo, no existe ningún motivo aparente por el cual la distribución de la media de la muestra correspondiente a muestras extraídas de una población muy asimétrica deba tener la misma forma que la correspondiente a muestras extraídas de una población simétrica. Sin embargo, ni queremos ni podemos estudiar, por separado, todas las formas imaginables de distribuciones originarias. Para nuestros fines basta con que nos limitemos al estudio detallado de dos distribuciones —la distribución uniforme discreta y la distribución normal— considerando las restantes distribuciones en términos generales. La distribución uniforme discreta tiene especial interés porque la utilizamos como base del experimento muestral de la sección 2.2; además, proporciona una base adecuada para ilustrar uno de los métodos de obtención teórica de distribuciones muestrales. Estudiaremos también la distribución normal, porque muchas de las inferencias que se realizan en la econometría dependen en gran medida del supuesto de normalidad de la población originaria.

En primer lugar, nos enfrentamos con el problema de obtener la distribución muestral de la media de la muestra cuando la variable que nos ocupa (por ejemplo, X) tiene una *distribución uniforme discreta*. Esto significa que X puede tomar un número finito de valores distintos, cada uno de ellos con la misma probabilidad. Las variables binarias representan un caso especial, y ya sabemos que dan lugar a la distribución binomial de la media de la muestra (ver sección 4.1). En este apartado nos ocupamos de un problema más general. Empezaremos especificando la distribución originaria en la misma forma que se utilizó en el experimento muestral de la sección 2.2. Como se recordará, X toma los valores 0, 1, 2, ..., 9, y la probabilidad de cada uno de ellos es $1/10$. Tal como

e calculó anteriormente, las principales características de esta distribución son:

Media:

$$E(X) = 4.5$$

Variancia:

$$\text{Var}(X) = 2.8723^2 = 8.25.$$

Más adelante generalizaremos este caso para hacerlo aplicable a cualquier número de valores distintos de X .

Obtengamos ahora la distribución muestral de X como estimador de $\mu = 4.5$. Para ello utilizaremos el procedimiento de la sección 4.1. «formando por partes» las probabilidades de los valores individuales del estimador y calculando las características principales de la distribución resultante. Al igual que entonces, efectuaremos este proceso para muestras de distinto tamaño, e intentaremos descubrir la pauta que nos permita desarrollar una fórmula general aplicable a muestras de cualquier tamaño.

Distribución muestral de X con muestras de tamaño 1. En este caso, X puede tomar 10 valores distintos: 0, 1, 2, ..., 9, al igual que la variable X . La probabilidad de cada uno de estos valores es $1/10$, tal como indican el cuadro 4-8 y la figura 4-11.

$\bar{x}$	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$\bar{x}^2f(\bar{x})$
0	$P(0) = 1/10$	0	0
1	$P(1) = 1/10$	1/10	1/10
2	$P(2) = 1/10$	2/10	4/10
3	$P(3) = 1/10$	3/10	9/10
4	$P(4) = 1/10$	4/10	16/10
5	$P(5) = 1/10$	5/10	25/10
6	$P(6) = 1/10$	6/10	36/10
7	$P(7) = 1/10$	7/10	49/10
8	$P(8) = 1/10$	8/10	64/10
9	$P(9) = 1/10$	9/10	81/10
Suma	1	45/10	285/10

Cuadro 4-8

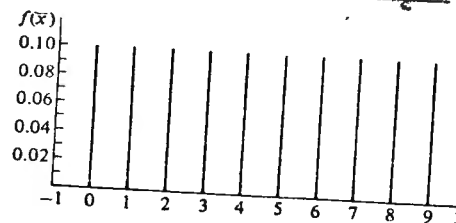


Figura 4-11

Media y variancia de $\bar{X}$:

$$E(\bar{X}) = \sum_i \bar{x}_i f(\bar{x}_i) = 4.5,$$

$$\text{Var}(\bar{X}) = E(\bar{X}^2) - [E(\bar{X})]^2 = \sum_i \bar{x}_i^2 f(\bar{x}_i) - \left[\sum_i \bar{x}_i f(\bar{x}_i) \right]^2 = 28.5 - 4.5^2 = 8.25.$$

Naturalmente, en este caso la media y la variancia de X son las mismas que las de la variable X en la población.

Distribución muestral de X con muestras de tamaño 2. En este caso, los valores posibles de X son 0, 1/2, 2/2, ..., 18/2. Es decir, X puede tomar 19 valores distintos y la distribución de probabilidades se presenta en el cuadro 4-9. Podemos sustituir la enumeración completa de las distintas probabilidades del cuadro 4-9 por las fórmulas combinatorias que se obtuvieron en la sección 3.2. Concretamente, podemos escribir:

$$P(0, 0) = \binom{1}{1} / 10^2$$

$$P(0, 1) + P(1, 0) = \binom{2}{1} / 10^2$$

$$P(0, 2) + P(2, 0) + P(1, 1) = \binom{3}{1} / 10^2$$

$$P(0, 3) + P(3, 0) + P(1, 2) + P(2, 1) = \binom{4}{1} / 10^2$$

$$P(0, 9) + P(9, 0) + P(8, 1) + P(1, 8) + P(2, 7) + P(7, 2) + P(3, 6) + P(6, 3) + P(4, 5) + P(5, 4) = \binom{10}{1} / 10^2$$

$$P(1, 9) + P(9, 1) + P(2, 8) + P(8, 2) + P(3, 7) + P(7, 3) + P(4, 6) + P(6, 4) + P(5, 5) = \left[\binom{11}{1} - 2 \right] / 10^2$$

$$P(8, 9) + P(9, 8) = \left[\binom{18}{1} - 16 \right] / 10^2$$

$$P(9, 9) = \left[\binom{19}{1} - 18 \right] / 10^2.$$

La distribución resultante tiene una forma «triangular» tal como puede verse en la figura 4-12.

(Cuadro 4-9)

$\bar{x}$	$f(\bar{x})$		$\bar{x}f(\bar{x})$	$\bar{x}^2f(\bar{x})$
0	$P(0,0)$	$= 1/10^2$	0.00	0.000
1/2	$P(0,1) + P(1,0)$	$= 2/10^2$	0.01	0.005
2/2	$P(0,2) + P(2,0) + P(1,1)$	$= 3/10^2$	0.03	0.030
3/2	$P(0,3) + P(3,0) + P(1,2) + P(2,1)$	$= 4/10^2$	0.06	0.090
4/2	...	$= 5/10^2$	0.10	0.200
5/2	...	$= 6/10^2$	0.15	0.375
6/2	...	$= 7/10^2$	0.21	0.630
7/2	...	$= 8/10^2$	0.28	0.980
8/2	...	$= 9/10^2$	0.36	1.440
9/2	$P(0,9) + P(9,0) + P(1,8) + P(8,1)$ $+ P(2,7) + P(7,2) + P(3,6) + P(6,3)$ $+ P(4,5) + P(5,4)$	$= 10/10^2$	0.45	2.025
10/2	$P(1,9) + P(9,1) + P(2,8) + P(8,2)$ $+ P(3,7) + P(7,3) + P(4,6) + P(6,4)$ $+ P(5,5)$	$= 9/10^2$	0.45	2.250
⋮	⋮	⋮	⋮	⋮
17/2	$P(8,9) + P(9,8)$	$= 2/10^2$	0.17	1.445
18/2	$P(9,9)$	$= 1/10^2$	0.09	0.810
	Suma	1	4.50	24.375

Media y variancia de $\bar{X}$:

$$E(\bar{X}) = 4.5,$$

$$\text{Var}(\bar{X}) = 24.375 - 4.5^2 = 24.375 - 20.25 = 4.125.$$

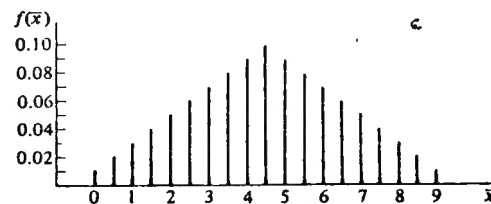


Figura 4-12

Distribución muestral de $\bar{X}$ con muestras de tamaño 3. En este caso, los valores posibles de $\bar{X}$ son 0, 1/3, 2/3, ..., 27/3. Las probabilidades correspondientes están calculadas en el cuadro 4-10. La distribución es perfectamente simétrica en torno al punto 4.5 y su representación gráfica aparece en la figura 4-13.

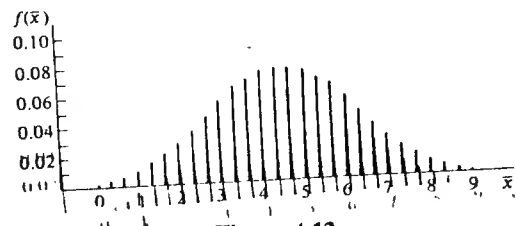


Figura 4-13

Figura 4-13

Media y variancia de $\bar{X}$:

Media y variancia de $\bar{X}$

$$E(\bar{X}) = 4.5,$$

$$\text{Var}(\bar{X}) = 23.00 - 4.5^2 = 2.75.$$

Al llegar aquí ya debe verse con claridad de qué forma se pueden seguir construyendo las distribuciones muestrales y no es preciso que sigamos haciéndolo de manera detallada. Por el contrario, presentaremos en forma resumida las distribuciones correspondientes a distintos tamaños muestrales. Confiamos en que ello inspire una generalización que haga posible determinar, a partir de una fórmula la probabilidad de cualquier valor de $\bar{X}$ correspondiente a una muestra de cualquier tamaño.

El cuadro 4-11 se ha construido de forma que la presentación sea lo más concisa posible, sin perder claridad. Como los valores de $\bar{X}$ siempre tienen la forma $0/n, 1/n, 2/n, \dots, 9n/n$, cada uno de ellos depende del tamaño de la muestra, n . Para no tener que presentar una columna distinta de valores de $\bar{X}$ para cada tamaño muestral, utilizamos solamente los numeradores, 0, 1, 2, ... Estos números son independientes del tamaño de la muestra, en todos los sentidos menos por lo que se refiere a la determinación del último de ellos. Naturalmente, sólo existe un número $n\bar{x}$ de ellos, como indica el título de la primera columna. Al escribir los números de dicha columna, nos limitamos únicamente a los situados alrededor de los puntos «de ruptura», indicados por los múltiplos de 10. Las mismas probabilidades individuales son fracciones con denominador 10ⁿ. Para no repetir cada vez este denominador, sólo escribimos los numeradores y por ello el título de la columna es «10ⁿf(x)» en lugar del valor f(x) corriente. Por lo demás, el cuadro es totalmente convencional

Cuadro 4-10

$\bar{x}$	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$\bar{x}^2f(\bar{x})$
0	$P(0,0,0) = \left(\frac{2}{2}\right)/10^3 = 1/10^3$	0.000	0.000 000
1/3	$P(0,0,1) + P(0,1,0) + P(1,0,0) = \left(\frac{3}{2}\right)/10^3 = 3/10^3$	0.001	0.000 333
2/3	$\dots = \left(\frac{4}{2}\right)/10^3 = 6/10^3$	0.004	0.002 667
3/3	$\dots = \left(\frac{5}{2}\right)/10^3 = 10/10^3$	0.010	0.010 000
4/3	$\dots = \left(\frac{6}{2}\right)/10^3 = 15/10^3$	0.020	0.026 667
5/3	$\dots = \left(\frac{7}{2}\right)/10^3 = 21/10^3$	0.035	0.058 333
6/3	$\dots = \left(\frac{8}{2}\right)/10^3 = 28/10^3$	0.056	0.112 000
7/3	$\dots = \left(\frac{9}{2}\right)/10^3 = 36/10^3$	0.084	0.196 000
8/3	$\dots = \left(\frac{10}{2}\right)/10^3 = 45/10^3$	0.120	0.320 000
9/3	$\dots = \left(\frac{11}{2}\right)/10^3 = 55/10^3$	0.165	0.495 000
10/3	$\dots = \left[\left(\frac{12}{2}\right) - 3\right]/10^3 = 63/10^3$	0.210	0.700 000
11/3	$\dots = \left[\left(\frac{13}{2}\right) - 9\right]/10^3 = 69/10^3$	0.253	0.927 667
12/3	$\dots = \left[\left(\frac{14}{2}\right) - 18\right]/10^3 = 73/10^3$	0.292	1.168 000
13/3	$\dots = \left[\left(\frac{15}{2}\right) - 30\right]/10^3 = 75/10^3$	0.325	1.408 333
14/3	$\dots = \left[\left(\frac{16}{2}\right) - 45\right]/10^3 = 75/10^3$	0.350	1.633 333
15/3	$\dots = \left[\left(\frac{17}{2}\right) - 63\right]/10^3 = 73/10^3$	0.365	1.825 000
:	:	:	:
26/3	$\dots = \left[\left(\frac{28}{2}\right) - 375\right]/10^3 = 3/10^3$	0.026	0.225 333
27/3	$\dots = \left[\left(\frac{29}{2}\right) - 405\right]/10^3 = 1/10^3$	0.009	0.081 000
Suma		1	4.500 23.000 000

$n\bar{x}$	$10^n f(\bar{x})$									
	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$
0	1	$\binom{1}{1} = 1$	$\binom{2}{2} = 1$	$\binom{3}{3} = 1$	$\binom{4}{4} = 1$	$\binom{5}{5} = 1$	$\binom{6}{6} = 1$	$\binom{7}{7} = 1$	$\binom{8}{8} = 1$	$\binom{9}{9} = 1$
1	1	$\binom{2}{1} = 2$	$\binom{3}{2} = 3$	$\binom{4}{3} = 4$	$\binom{5}{4} = 5$	$\binom{6}{5} = 6$	$\binom{7}{6} = 7$	$\binom{8}{7} = 8$	$\binom{9}{8} = 9$	$\binom{10}{9} = 10$
2	1	$\binom{3}{1} = 3$	$\binom{4}{2} = 6$	$\binom{5}{3} = 10$	$\binom{6}{4} = 15$	$\binom{7}{5} = 21$	$\binom{8}{6} = 28$	$\binom{9}{7} = 36$	$\binom{10}{8} = 45$	$\binom{11}{9} = 55$
3	1	$\binom{4}{1} = 4$	$\binom{5}{2} = 10$	$\binom{6}{3} = 20$	$\binom{7}{4} = 35$	$\binom{8}{5} = 56$	$\binom{9}{6} = 84$	$\binom{10}{7} = 120$	$\binom{11}{8} = 165$	$\binom{12}{9} = 220$
4	1	$\binom{5}{1} = 5$	$\binom{6}{2} = 15$	$\binom{7}{3} = 35$	$\binom{8}{4} = 70$	$\binom{9}{5} = 126$	$\binom{10}{6} = 210$	$\binom{11}{7} = 330$	$\binom{12}{8} = 495$	$\binom{13}{9} = 715$
5	1	$\binom{6}{1} = 6$	$\binom{7}{2} = 21$	$\binom{8}{3} = 56$	$\binom{9}{4} = 126$	$\binom{10}{5} = 252$	$\binom{11}{6} = 462$	$\binom{12}{7} = 792$	$\binom{13}{8} = 1287$	$\binom{14}{9} = 2002$
6	1	$\binom{7}{1} = 7$	$\binom{8}{2} = 28$	$\binom{9}{3} = 84$	$\binom{10}{4} = 210$	$\binom{11}{5} = 462$	$\binom{12}{6} = 924$	$\binom{13}{7} = 1716$	$\binom{14}{8} = 3003$	$\binom{15}{9} = 4862$
7	1	$\binom{8}{1} = 8$	$\binom{9}{2} = 36$	$\binom{10}{3} = 120$	$\binom{11}{4} = 330$	$\binom{12}{5} = 792$	$\binom{13}{6} = 1716$	$\binom{14}{7} = 3432$	$\binom{15}{8} = 6435$	$\binom{16}{9} = 11440$
8	1	$\binom{9}{1} = 9$	$\binom{10}{2} = 45$	$\binom{11}{3} = 165$	$\binom{12}{4} = 495$	$\binom{13}{5} = 1287$	$\binom{14}{6} = 3003$	$\binom{15}{7} = 6435$	$\binom{16}{8} = 12870$	$\binom{17}{9} = 24310$
9	1	$\binom{10}{1} = 10$	$\binom{11}{2} = 55$	$\binom{12}{3} = 220$	$\binom{13}{4} = 715$	$\binom{14}{5} = 2002$	$\binom{15}{6} = 5005$	$\binom{16}{7} = 11440$	$\binom{17}{8} = 24310$	$\binom{18}{9} = 48620$
10		$\binom{11}{1} - 2 = 9$	$\binom{12}{2} - 3 = 63$	$\binom{13}{3} - 4 = 282$	$\binom{14}{4} - 5 = 996$	$\binom{15}{5} - 6 = 3003$	$\binom{16}{6} - 7 = 7920$	$\binom{17}{7} - 8 = 17160$	$\binom{18}{8} - 9 = 35430$	$\binom{19}{9} - 10 = 71520$
11		$\binom{12}{1} - 4 = 8$	$\binom{13}{2} - 9 = 69$	$\binom{14}{3} - 16 = 348$	$\binom{15}{4} - 25 = 1,340$	$\binom{16}{5} - 36 = 3003$	$\binom{17}{6} - 49 = 7920$	$\binom{18}{7} - 64 = 17160$	$\binom{19}{8} - 81 = 35430$	$\binom{20}{9} - 100 = 71520$
12		$\binom{13}{1} - 6 = 7$	$\binom{14}{2} - 18 = 73$	$\binom{15}{3} - 40 = 415$	$\binom{16}{4} - 75 = 1,745$	$\binom{17}{5} - 100 = 3003$	$\binom{18}{6} - 147 = 7920$	$\binom{19}{7} - 200 = 17160$	$\binom{20}{8} - 271 = 35430$	$\binom{21}{9} - 361 = 71520$
13		$\binom{14}{1} - 8 = 6$	$\binom{15}{2} - 21 = 74$	$\binom{16}{3} - 64 = 448$	$\binom{17}{4} - 100 = 2,000$	$\binom{18}{5} - 147 = 3003$	$\binom{19}{6} - 216 = 7920$	$\binom{20}{7} - 300 = 17160$	$\binom{21}{8} - 400 = 35430$	$\binom{22}{9} - 529 = 71520$
14		$\binom{15}{1} - 10 = 5$	$\binom{16}{2} - 25 = 79$	$\binom{17}{3} - 80 = 477$	$\binom{18}{4} - 126 = 2,178$	$\binom{19}{5} - 200 = 3003$	$\binom{20}{6} - 294 = 7920$	$\binom{21}{7} - 400 = 17160$	$\binom{22}{8} - 529 = 35430$	$\binom{23}{9} - 729 = 71520$
15		$\binom{16}{1} - 12 = 4$	$\binom{17}{2} - 30 = 84$	$\binom{18}{3} - 100 = 504$	$\binom{19}{4} - 165 = 2,364$	$\binom{20}{5} - 252 = 3003$	$\binom{21}{6} - 378 = 7920$	$\binom{22}{7} - 500 = 17160$	$\binom{23}{8} - 649 = 35430$	$\binom{24}{9} - 900 = 71520$
16		$\binom{17}{1} - 14 = 3$	$\binom{18}{2} - 36 = 89$	$\binom{19}{3} - 120 = 531$	$\binom{20}{4} - 210 = 2,590$	$\binom{21}{5} - 330 = 3003$	$\binom{22}{6} - 528 = 7920$	$\binom{23}{7} - 676 = 17160$	$\binom{24}{8} - 841 = 35430$	$\binom{25}{9} - 1100 = 71520$
17		$\binom{18}{1} - 16 = 2$	$\binom{19}{2} - 40 = 94$	$\binom{20}{3} - 140 = 560$	$\binom{21}{4} - 252 = 2,848$	$\binom{22}{5} - 360 = 3003$	$\binom{23}{6} - 672 = 7920$	$\binom{24}{7} - 840 = 17160$	$\binom{25}{8} - 1024 = 35430$	$\binom{26}{9} - 1296 = 71520$
18		$\binom{19}{1} - 18 = 1$	$\binom{20}{2} - 45 = 95$	$\binom{21}{3} - 160 = 574$	$\binom{22}{4} - 280 = 3,068$	$\binom{23}{5} - 462 = 3003$	$\binom{24}{6} - 800 = 7920$	$\binom{25}{7} - 1000 = 17160$	$\binom{26}{8} - 1216 = 35430$	$\binom{27}{9} - 1600 = 71520$
19		$\binom{20}{1} - 20 = 0$	$\binom{21}{2} - 50 = 100$	$\binom{22}{3} - 180 = 600$	$\binom{23}{4} - 300 = 3,300$	$\binom{24}{5} - 504 = 3003$	$\binom{25}{6} - 900 = 7920$	$\binom{26}{7} - 1200 = 17160$	$\binom{27}{8} - 1440 = 35430$	$\binom{28}{9} - 1960 = 71520$
20		$\binom{21}{1} - 21 = 0$	$\binom{22}{2} - 55 = 104$	$\binom{23}{3} - 200 = 630$	$\binom{24}{4} - 350 = 3,590$	$\binom{25}{5} - 600 = 3003$	$\binom{26}{6} - 1000 = 7920$	$\binom{27}{7} - 1400 = 17160$	$\binom{28}{8} - 1760 = 35430$	$\binom{29}{9} - 2400 = 71520$
21		$\binom{22}{1} - 22 = 0$	$\binom{23}{2} - 60 = 109$	$\binom{24}{3} - 220 = 660$	$\binom{25}{4} - 400 = 3,900$	$\binom{26}{5} - 720 = 3003$	$\binom{27}{6} - 1200 = 7920$	$\binom{28}{7} - 1760 = 17160$	$\binom{29}{8} - 2200 = 35430$	$\binom{30}{9} - 3000 = 71520$
22		$\binom{23}{1} - 23 = 0$	$\binom{24}{2} - 66 = 114$	$\binom{25}{3} - 240 = 690$	$\binom{26}{4} - 450 = 4,250$	$\binom{27}{5} - 840 = 3003$	$\binom{28}{6} - 1400 = 7920$	$\binom{29}{7} - 2000 = 17160$	$\binom{30}{8} - 2500 = 35430$	$\binom{31}{9} - 3600 = 71520$
23		$\binom{24}{1} - 24 = 0$	$\binom{25}{2} - 70 = 119$	$\binom{26}{3} - 260 = 715$	$\binom{27}{4} - 500 = 4,650$	$\binom{28}{5} - 1000 = 3003$	$\binom{29}{6} - 1760 = 7920$	$\binom{30}{7} - 2500 = 17160$	$\binom{31}{8} - 3000 = 35430$	$\binom{32}{9} - 4000 = 71520$
24		$\binom{25}{1} - 25 = 0$	$\binom{26}{2} - 75 = 124$	$\binom{27}{3} - 280 = 740$	$\binom{28}{4} - 560 = 5,040$	$\binom{29}{5} - 1320 = 3003$	$\binom{30}{6} - 2000 = 7920$	$\binom{31}{7} - 3000 = 17160$	$\binom{32}{8} - 3600 = 35430$	$\binom{33}{9} - 5000 = 71520$
25		$\binom{26}{1} - 26 = 0$	$\binom{27}{2} - 80 = 129$	$\binom{28}{3} - 300 = 765$	$\binom{29}{4} - 600 = 5,440$	$\binom{30}{5} - 1500 = 3003$	$\binom{31}{6} - 2200 = 7920$	$\binom{32}{7} - 3500 = 17160$	$\binom{33}{8} - 4200 = 35430$	$\binom{34}{9} - 6000 = 71520$
26		$\binom{27}{1} - 27 = 0$	$\binom{28}{2} - 84 = 134$	$\binom{29}{3} - 320 = 790$	$\binom{30}{4} - 660 = 5,840$	$\binom{31}{5} - 1680 = 3003$	$\binom{32}{6} - 2500 = 7920$	$\binom{33}{7} - 4000 = 17160$	$\binom{34}{8} - 5000 = 35430$	$\binom{35}{9} - 7000 = 71520$
27		$\binom{28}{1} - 28 = 0$	$\binom{29}{2} - 89 = 139$	$\binom{30}{3} - 340 = 815$	$\binom{31}{4} - 720 = 6,240$	$\binom{32}{5} - 1920 = 3003$	$\binom{33}{6} - 2800 = 7920$	$\binom{34}{7} - 4800 = 17160$	$\binom{35}{8} - 5700 = 35430$	$\binom{36}{9} - 8000 = 71520$
28		$\binom{29}{1} - 29 = 0$	$\binom{30}{2} - 90 = 144$	$\binom{31}{3} - 360 = 840$	$\binom{32}{4} - 800 = 6,640$	$\binom{33}{5} - 2100 = 3003$	$\binom{34}{6} - 3000 = 7920$	$\binom{35}{7} - 5400 = 17160$	$\binom{36}{8} - 6600 = 35430$	$\binom{37}{9} - 9000 = 71520$
29		$\binom{30}{1} - 30 = 0$	$\binom{31}{2} - 95 = 149$	$\binom{32}{3} - 380 = 865$	$\binom{33}{4} - 880 = 7,040$	$\binom{34}{5} - 2310 = 3003$	$\binom{35}{6} - 3300 = 7920$	$\binom{36}{7} - 6000 = 17160$	$\binom{37}{8} - 7500 = 35430$	$\binom{38}{9} - 10000 = 71520$
30		$\binom{31}{1} - 31 = 0$	$\binom{32}{2} - 100 = 154$	$\binom{33}{3} - 400 = 890$	$\binom{34}{4} - 960 = 7,440$	$\binom{35}{5} - 2520 = 3003$	$\binom{36}{6} - 3600 = 7920$	$\binom{37}{7} - 6700 = 17160$	$\binom{38}{8} - 8400 = 35430$	$\binom{39}{9} - 11000 = 71520$
31		$\binom{32}{1} - 32 = 0$	$\binom{33}{2} - 105 = 159$	$\binom{34}{3} - 420 = 915$	$\binom{35}{4} - 1040 = 7,840$	$\binom{36}{5} - 2730 = 3003$	$\binom{37}{6} - 3900 = 7920$	$\binom{38}{7} - 7200 = 17160$	$\binom{39}{8} - 9300 = 35430$	$\binom{40}{9} - 12000 = 71520$
32		$\binom{33}{1} - 33 = 0$	$\binom{34}{2} - 110 = 164$	$\binom{35}{3} - 440 = 940$	$\binom{36}{4} - 1120 = 8,240$	$\binom{37}{5} - 2940 = 3003$	$\binom{38}{6} - 4200 = 7920$	$\binom{39}{7} - 7700 = 17160$	$\binom{40}{8} - 10400 = 35430$	$\binom{41}{9} - 13000 = 71520$
33		$\binom{34}{1} - 34 = 0$	$\binom{35}{2} - 115 = 169$	$\binom{36}{3} - 460 = 965$	$\binom{37}{4} - 1200 = 8,640$	$\binom{38}{5} - 3150 = 3003$	$\binom{39}{6} - 4600 = 7920$	$\binom{40}{7} - 8200 = 17160$	$\binom{41}{8} - 11500 = 35430$	$\binom{42}{9} - 14000 = 71520$
34		$\binom{35}{1} - 35 = 0$	$\binom{36}{2} - 120 = 174$	$\binom{37}{3} - 480 = 990$	$\binom{38}{4} - 1280 = 9,040$	$\binom{39}{5} - 3360 = 3003$	$\binom{40}{6} - 5000 = 7920$	$\binom{41}{7} - 8800 = 17160$	$\binom{42}{8} - 12600 = 35430$	$\binom{43}{9} - 15000 = 71520$
35		$\binom{36}{1} - 36 = 0$	$\binom{37}{2} - 125 = 179$	$\binom{38}{3} - 500 = 1,015$	$\binom{39}{4} - 1360 = 9,440$	$\binom{40}{5} - 3570 = 3003$	$\binom{41}{6} - 5400 = 7920$	$\binom{42}{7} - 9300 = 17160$	$\binom{43}{8} - 13700 = 35430$	$\binom{44}{9} - 16000 = 71520$
36		$\binom{37}{1} - 37 = 0$	$\binom{38}{2} - 130 = 184$	$\binom{39}{3} - 520 = 1,040$	$\binom{40}{4} - 1440 = 9,840$	$\binom{41}{5} - 3780 = 3003$	$\binom{42}{6} - 5800 = 7920$	$\binom{43}{7} - 9800 = 17160$	$\binom{44}{8} - 14800 = 35430$	$\binom{45}{9} - 17000 = 71520$
37		$\binom{38}{1} - 38 = 0$	$\binom{39}{2} - 135 = 189$	$\binom{40}{3} - 540 = 1,065$	$\binom{41}{4} - 1520 = 10,240$	$\binom{42}{5} - 3990 = 3003$	$\binom{43}{6} - 6200 = 7920$	$\binom{44}{7} - 10300 = 17160$	$\binom{45}{8} - 15900 = 35430$	$\binom{46}{9} - 18000 = 71520$
38		$\binom{39}{1} - 39 = 0$	$\binom{40}{2} - 140 = 194$	$\binom{41}{3} - 560 = 1,090$	$\binom{42}{4} - 1600 = 10,640$	$\binom{43}{5} - 4200 = 3003$	$\binom{44}{6} - 6600 = 7920$	$\binom{45}{7} - 10800 = 17160$	$\binom{46}{8} - 17000 = 35430$	$\binom{47}{9} - 19000 = 71520$
39		$\binom{40}{1} - 40 = 0$	$\binom{41}{2} - 145 = 199$	$\binom{42}{3} - 580 = 1,115$	$\binom{43}{4} - 1680 = 11,040$	$\binom{44}{5} - 4410 = 3003$	$\binom{45}{6} - 7000 = 7920$	$\binom{46}{7} - 11300 = 17160$	$\binom{47}{8} - 18100 = 35430$	$\binom{48}{9} - 20000 = 71520$
40		$\binom{41}{1} - 41 = 0$	$\binom{42}{2} - 150 = 204$	$\binom{43}{3} - 600 = 1,140$	$\binom{44}{4} - 1760 = 11,440$	$\binom{45}{5} - 4620 = 3003$	$\binom{46}{6} - 7400 = 7920$	$\binom{47}{7} - 11800 = 17160$	$\binom{48}{8} - 19200 = 35430$	$\binom{49}{9} - 21000 = 71520$
41		$\binom{42}{1} - 42 = 0$	$\binom{43}{2} - 155 = 209$	$\binom{44}{3} - 620 = 1,165$	$\binom{45}{4} - 1840 = 11,840$	$\binom{46}{5} - 4830 = 3003$	$\binom{47}{6} - 7800 = 7920$	$\binom{48}{7} - 12300 = 17160$	$\binom{49}{8} - 20300 = 35430$	$\binom{50}{9} - 22000 = 71520$
42		$\binom{43}{1} - 43 = 0$	$\binom{44}{2} - 160 = 214$	$\binom{45}{3} - 640 = 1,190$	$\binom{46}{4} - 1920 = 12,240$	$\binom{47}{5} - 5040 = 3003$	$\binom{48}{6} - 8200 = 7920$	$\binom{49}{7} - 12800 = 17160$	$\binom{50}{8} - 21400 = 35430$	$\binom{51}{9} - 23000 = 71520$
43		$\binom{44}{1} - 44 = 0$	$\binom{45}{2} - 165 = 219$	$\binom{46}{3} - 660 = 1,215$	$\binom{47}{4} - 2000 = 12,640$	$\binom{48}{5} - 5250 = 3003$	$\binom{49}{6} - 8600 = 7920$	$\binom{50}{7} - 13300 = 17160$	$\binom{51}{8} - 22500 = 35430$	$\binom{52}{9} - 24000 = 71520$
44		$\binom{45}{1} - 45 = 0$	$\binom{46}{2} - 170 = 224$	$\binom{47}{3} - 680 = 1,240$	$\binom{48}{4} - 2080 = 13,040$	$\binom{49}{5} - 5460 = 3003$	$\binom{50}{6} - 9000 = 7920$	$\binom{51}{7} - 13800 = 17160$	$\binom{52}{8} - 23600 = 35430$	$\binom{53}{9} - 25000 = 71520$
45		$\binom{46}{1} - 46 = 0$	$\binom{47}{2} - 175 = 229$	$\binom{48}{3} - 700 = 1,265$	$\binom{49}{4} - 2160 = 13,440$	$\binom{50}{5} - 5670 = 3003$	$\binom{51}{6} - 9400 = 7920$	$\binom{52}{7} - 14300 = 17160$	$\binom{53}{8} - 24700 = 35430$	$\binom{54}{9} - 26000 = 71520$
46		$\binom{47}{1} - 47 = 0$	$\binom{48}{2} - 180 = 234$	$\binom{49}{3} - 720 = 1,290$	$\binom{50}{4} - 2240 = 13,840$	$\binom{51}{5} - 5880 = 3003$	$\binom{52}{6} - 9800 = 7920$	$\binom{53}{7} - 14800 = 171$		

En este momento puede verse que estas secuencias son simplemente los coeficientes de probabilidad dados diez lugares antes. Es decir,

$$c_1 = n \binom{n\bar{x} - 10 + n - 1}{n - 1},$$

y, por tanto, las probabilidades de los distintos valores de $\bar{x}$ para $n\bar{x}$ entre 10 y 19 son

$$f(\bar{x}) = \left[\binom{n\bar{x} + n - 1}{n - 1} - n \binom{n\bar{x} - 10 + n - 1}{n - 1} \right] / 10^n.$$

Por ejemplo, la probabilidad de que $\bar{x}$ de la muestra de tamaño 4 sea igual a 4.5 es

$$\left[\binom{18 + 4 - 1}{4 - 1} - 4 \binom{18 - 10 + 4 - 1}{4 - 1} \right] / 10^4 = \left[\binom{21}{3} - 4 \binom{11}{3} \right] / 10^4 = 0.0670.$$

Si aplicamos la fórmula obtenida para el segundo grupo de coeficientes a la obtención de los coeficientes del tercer grupo, vemos que es necesario efectuar una nueva corrección. Por ejemplo, consideremos el coeficiente correspondiente a $n\bar{x} = 20$ en la muestra de tamaño 3. El coeficiente es igual a

$$\binom{22}{2} - 195 = 36.$$

Pero si aplicamos la fórmula del párrafo anterior obtenemos

$$\binom{22}{2} - 3 \binom{12}{2} = 33 \neq 36.$$

Por tanto, necesitamos un nuevo factor de corrección al que llamaremos c_2 y para $n\bar{x}$ entre 20 y 29 tendremos

$$f(\bar{x}) = \left[\binom{n\bar{x} + n - 1}{n - 1} - n \binom{n\bar{x} - 10 + n - 1}{n - 1} + c_2 \right] / 10^n.$$

Los valores de c_2 son:

$n\bar{x}$	$n = 3$	$n = 4$	$n = 5$
20	3	6	10
21	9	24	50
22	18	60	150
23	30	120	350
⋮	⋮	⋮	⋮

Obsérvese que todos los valores de la columna « $n = 3$ » son múltiplos de 3, los de la columna « $n = 4$ » son múltiplos de 6, y los de la columna « $n = 5$ » son múltiplos de 10. La relación entre estos factores comunes y los tamaños muestrales respectivos puede identificarse, después de reflexionar un poco, de la forma siguiente:

$$3 = \binom{3}{2}, 6 = \binom{4}{2}, 10 = \binom{5}{2}.$$

En general, el factor común es igual a $\binom{n}{2}$. Si dividimos cada columna por su factor común respectivo, obtenemos

$n\bar{x}$	$n = 3$	$n = 4$	$n = 5$
20	1	1	1
21	3	4	5
22	6	10	15
23	10	20	35
⋮	⋮	⋮	⋮

Estas secuencias son simplemente los coeficientes de probabilidad obtenidos veinte lugares antes y, por tanto, la fórmula correspondiente a la probabilidad de los distintos valores de $\bar{x}$ cuando $n\bar{x}$ se halla entre 20 y 29 es

$$f(\bar{x}) = \left[\binom{n\bar{x} + n - 1}{n - 1} - n \binom{n\bar{x} - 10 + n - 1}{n - 1} + \binom{n}{2} \binom{n\bar{x} - 20 + n - 1}{n - 1} \right] / 10^n.$$

Como ejemplo, consideremos la probabilidad de que $\bar{x}$ de la muestra de tamaño 4 sea igual a 6. Esta probabilidad es

$$\left[\binom{24 + 4 - 1}{4 - 1} - 4 \binom{24 - 10 + 4 - 1}{4 - 1} + 6 \binom{24 - 20 + 4 - 1}{4 - 1} \right] / 10^4 = 0.0415.$$

Podríamos continuar este proceso, pero la pauta ya resulta evidente. Para calcular la probabilidad de cualquier valor de $\bar{x}$, correspondiente a cualquier tamaño muestral n , basta con extender la fórmula anterior para $f(\bar{x})$, añadiendo factores de corrección adicionales del mismo tipo. En general, la distribución de probabilidad de $\bar{X}$ será

$$(4.10) \quad f(\bar{x}) = \frac{1}{10^n} \left[\binom{n\bar{x} + n - 1}{n - 1} - \binom{n}{1} \binom{n\bar{x} - 10 + n - 1}{n - 1} + \binom{n}{2} \binom{n\bar{x} - 20 + n - 1}{n - 1} - \binom{n}{3} \binom{n\bar{x} - 30 + n - 1}{n - 1} + \dots \right].$$

La suma de la ecuación (4.10) debe proseguir mientras los términos $(n\bar{x} - 10)$, $(n\bar{x} - 20)$,... sean positivos o cero, es decir, hasta que el sustraendo correspondiente, 10, 20, 30,... sea igual o mayor que $n\bar{x}$.

La ecuación (4.10) nos proporciona la distribución muestral de $\bar{X}$ para muestras extraídas de una población uniforme discreta en la que X puede tomar los valores 0, 1, 2,..., 9. La fórmula se puede modificar fácilmente con el fin de aplicarla a una situación en la que X puede tomar, con igual probabilidad, cualesquiera valores distintos k , separados por distancias iguales. Para efectuar esta generalización, basta con sustituir los valores 10, 20, 30,... de la fórmula (4.10) por los valores k , $2k$, $3k$,... obteniendo

$$(4.11) \quad f(\bar{x}) = \frac{1}{k^n} \left[\binom{n\bar{x} + n - 1}{n-1} - \binom{n}{1} \binom{n\bar{x} - k + n - 1}{n-1} + \binom{n}{2} \binom{n\bar{x} - 2k + n - 1}{n-1} - \binom{n}{3} \binom{n\bar{x} - 3k + n - 1}{n-1} + \dots \right].$$

Obtenemos ahora las características básicas de las distribuciones muestrales expresadas en las ecuaciones (4.10) y (4.11). Para la primera, los valores de las medias y las variancias que ya se obtuvieron previamente en el caso de las muestras de tamaño 1, 2, 3, 4 y 5, se resumen en el cuadro 4-12. (Las

Cuadro 4-12

Tamaño muestral	Media	Variancia
1	4.5	8.25
2	4.5	4.125
3	4.5	2.75
4	4.5	2.0625
5	4.5	1.65

variancias correspondientes a las muestras de tamaño 1, 2 y 3 se obtuvieron previamente; las de las muestras de tamaño 4 y 5 se pueden obtener a partir de la última línea del cuadro 4-11, restando 4.5^2 en cada caso.) El rasgo más sobresaliente del cuadro 4-12 es el hecho de que el valor de la media es igual a la media de la población, en todos los tamaños muestrales examinados. Por lo que se refiere a la variancia, obsérvese que el valor correspondiente a la muestra de tamaño 1 es exactamente igual a la variancia de la población y que los demás valores disminuyen proporcionalmente respecto al tamaño de la muestra, es decir que

$$4.125 = \frac{8.25}{2}.$$

$$2.75 = \frac{8.25}{3},$$

$$2.0625 = \frac{8.25}{4},$$

$$1.65 = \frac{8.25}{5}.$$

Por último, como vimos que todas las distribuciones muestrales son exactamente simétricas, el momento de tercer orden μ_3 (y, en consecuencia, también α_3) debe ser igual a cero. Estos resultados se pueden generalizar fácilmente con el fin de aplicarlos a cualquier tamaño muestral n . En este caso tenemos que en las que

$$E(\bar{X}) = \mu,$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n},$$

$$\alpha_3 = 0,$$

donde

$$\sigma_{\bar{X}}^2 = \text{Var}(\bar{X}) \quad \text{y} \quad \sigma_X^2 = \text{Var}(X).$$

Se puede demostrar que estas generalizaciones son perfectamente válidas y, además, no sólo se aplican a la distribución de la ecuación (4.10), sino también a su versión generalizada, presentada en la ecuación (4.11). (Las pruebas relativas a la media y a la variancia se presentarán hacia el final de esta sección, y la correspondiente a α_3 se deja para el lector.)

Al llegar a este punto podemos efectuar afirmaciones precisas acerca de las propiedades de la media de la muestra como estimador de la media de la población cuando las muestras provienen de una *población uniforme discreta* con valores de X separados por espacios iguales. Podemos extraer las conclusiones siguientes:

1. $\bar{X}$ es un estimador insesgado de μ .
2. La variancia de $\bar{X}$ es igual a la variancia de X dividida por el tamaño de la muestra. Por lo tanto, al aumentar el tamaño de la muestra, la distribución de $\bar{X}$ se hace cada vez más concentrada.
3. Las conclusiones 1 y 2 tomadas conjuntamente, implican que $\bar{X}$ es un estimador consistente de μ .
4. La distribución de $\bar{X}$ es perfectamente simétrica.

Estas propiedades de $\bar{X}$ se han deducido a partir de la distribución muestral obtenida teóricamente. En la sección 2.2 se intentó obtener estas propiedades efectuando un experimento muestral con una población uniforme discreta en la que X tomaba los valores 0, 1, 2,..., 9. El experimento consistió en extraer 100

muestras de tamaño 5 y 100 muestras de tamaño 10. A continuación, se destruyeron las distribuciones de frecuencias de la media y la mediana muestrales correspondientes a las muestras de tamaño 5 y la de la media muestral de las muestras de tamaño 10, tomando las distribuciones de frecuencias así obtenidas como aproximaciones de las verdaderas distribuciones muestrales de dichos estimadores. Los resultados nos permitieron obtener algunas conclusiones provisionales relativas a las propiedades de los estimadores. Ahora podemos comprobar los resultados experimentales correspondientes a la media de la muestra comparándolos con los obtenidos teóricamente.

En primer lugar, fijémonos en la distribución muestral de la media de la muestra para el tamaño 5. Los resultados experimentales de este caso son los presentados en el cuadro 4-13 y se reproducen en el cuadro 4-14 juntamente con los resultados teóricos obtenidos a partir del cuadro 4-11, todos ellos presentados de forma que permita la comparación.

A continuación, se comparan los resultados experimentales y teóricos de las muestras de tamaño 10. Los resultados experimentales son los que aparecen en el cuadro 2-6, y los teóricos se han calculado por medio de la fórmula (4.10) y las expresiones que se han obtenido para $E(\bar{X})$ y $\sigma_{\bar{X}}$. En el cuadro 4-14 se presentan ambas distribuciones de forma que permita su comparación. Naturalmente, la distribución teórica es perfectamente simétrica, aunque ello no es evidente de forma inmediata a partir de los datos del cuadro 4-14 debido a la forma en que se han espaciado los límites de los intervalos.

Los cuadros 4-13 y 4-14 indican que las distribuciones de frecuencias obtenidas experimentalmente proporcionan una aproximación razonable de las distribuciones muestrales verdaderas. Las conclusiones provisionales acerca de $\bar{X}$

Cuadro 4-13

Intervalo: $\bar{x}$	Tamaño de la muestra 5	
	$f(\bar{x})$ en los experimentos	$f(\bar{x})$ teórica
0.5 a 1.499	0.01	0.01
1.5 a 2.499	0.05	0.05
2.5 a 3.499	0.12	0.16
3.5 a 4.499	0.31	0.28
4.5 a 5.499	0.28	0.28
5.5 a 6.499	0.15	0.16
6.5 a 7.499	0.05	0.05
7.5 a 8.499	0.03	0.01
8.5 a 9.499	0.00	0.00
Media	4.60	4.5
Desviación standard	1.3638	1.2845

Cuadro 4-14

Intervalo: $\bar{x}$	Tamaño de la muestra 10	
	$f(\bar{x})$ en los experimentos	$f(\bar{x})$ teórica
0.5 a 1.499	0.00	0.00
1.5 a 2.499	0.01	0.01
2.5 a 3.499	0.14	0.11
3.5 a 4.499	0.34	0.35
4.5 a 5.499	0.32	0.37
5.5 a 6.499	0.16	0.14
6.5 a 7.499	0.03	0.02
7.5 a 8.499	0.00	0.00
8.5 a 9.499	0.00	0.00
Media	4.57	4.5
Desviación standard	1.0416	0.9829

como estimador de la media de la población —es decir, insesgadez, consistencia y disminución de la variancia en proporción al tamaño de la muestra— resultan ser correctas todas ellas. Sin embargo, los resultados experimentales no nos proporcionaron ninguna fórmula para obtener la variancia y no fueron suficientemente claros para sugerir la simetría. Los resultados teóricos no solamente son más precisos, sino que son también considerablemente más explícitos y generales.

Hasta este momento, hemos analizado el muestreo efectuado a partir de una población uniforme discreta. Pasemos ahora al muestreo a partir de poblaciones normales. En particular, nuestro objetivo consiste en obtener la distribución muestral de la media de la muestra sabiendo que la variable X , de cuyos valores está compuesta la muestra, se distribuye normalmente con media μ y variancia σ^2 , es decir, sabiendo que $X \sim N(\mu, \sigma^2)$. Obsérvese que aquí efectuamos un pequeño cambio de notación: para describir la variancia de la población utilizamos σ^2 en lugar de σ_x^2 . En este caso sabemos que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} (X_1 + X_2 + \cdots + X_n) = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \cdots + \frac{1}{n} X_n.$$

Podemos considerar las n observaciones $X_1, X_2, \dots, X_n$ como n variables, cada una de las cuales posee la misma distribución, la misma media y la misma variancia que X . Es decir, X_1 indica todos los valores posibles de X que se pueden obtener al efectuar la primera observación, X_2 los valores posibles de X que se pueden obtener al efectuar la segunda, y así sucesivamente. Esta forma de con-

siderar la muestra —que, naturalmente, es perfectamente legítima— hace que el análisis posterior sea más sencillo y más claro. Para un tamaño muestral determinado, n (y, por tanto, también $1/n$) es una constante y, en consecuencia, podemos considerar a $\bar{X}$ como una combinación lineal de n variables normales independientes, $X_1, X_2, \dots, X_n$. Pero como ya sabemos por el teorema 12 de este capítulo, una combinación lineal de variables aleatorias independientes que se distribuyen normalmente también se distribuirá normalmente. Ello nos evita todo el trabajo correspondiente a la obtención de la forma de la distribución muestral de $\bar{X}$, ya que el teorema implica claramente que si X (y por tanto, $X_1, X_2, \dots, X_n$) es una normal, $\bar{X}$ es también normal, independientemente del tamaño de la muestra.

Tras haber obtenido este resultado, nuestra tarea se limita a la obtención de la media y la variancia de la distribución muestral de $\bar{X}$. Tal como se señaló en la sección 4.2, el conocimiento de la media y la variancia de una distribución normal basta para definirla por completo y, afortunadamente, en este caso la obtención de estos dos parámetros resulta ser relativamente simple. Tomemos la media en primer lugar:

$$E(\bar{X}) = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right]$$

Como $1/n$ es una constante, según el teorema 6 podemos escribir esta expresión en la forma

$$E(\bar{X}) = \frac{1}{n} E \sum_{i=1}^n X_i.$$

Pero, según el teorema 8, la esperanza de una suma es igual a la suma de las esperanzas y, por tanto,

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)].$$

Ahora bien, cada uno de los términos $X_1, X_2, \dots, X_n$ tiene la misma media que X , es decir μ ; o sea, que

$$E(\bar{X}) = \frac{1}{n} [\mu + \mu + \dots + \mu] = \frac{n\mu}{n} = \mu.$$

Es decir, la media de la distribución muestral de $\bar{X}$ es igual a la media de la población μ . Obtengamos a continuación la variancia de $\bar{X}$, a la que representamos por $\sigma_{\bar{X}}^2$. Sabemos que

$$\sigma_{\bar{X}}^2 = \text{Var}(\bar{X}) = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] \quad (\text{según el teorema 2})$$

Según el supuesto de muestreo aleatorio, los términos $X_1, X_2, \dots, X_n$ son independientes entre sí y, por tanto, podemos utilizar el teorema siguiente.

Teorema 13. *La variancia de una suma de variables aleatorias independientes es igual a la suma de sus variancias.*

Demostración: Sean $X_1, X_2, \dots, X_n$ un conjunto de variables aleatorias independientes. En este caso,

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n X_i\right] &= E[(X_1 + X_2 + \dots + X_n) - E(X_1 + X_2 + \dots + X_n)]^2 \\ &= E\{[X_1 - E(X_1)] + [X_2 - E(X_2)] + \dots + [X_n - E(X_n)]\}^2 \\ &= \sum_{i=1}^n E[X_i - E(X_i)]^2 + 2 \sum_{i < j} E[X_i - E(X_i)][X_j - E(X_j)]. \end{aligned}$$

Pero, como las variables son independientes entre sí, todas sus covariancias son iguales a cero y, por tanto,

$$\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}(X_i).$$

Y podemos escribir la variancia de $\bar{X}$ en la forma

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2} [\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)].$$

Ahora bien, todos los términos $X_1, X_2, \dots, X_n$ tienen la misma variancia que X , σ^2 , y, por tanto,

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2} [\sigma^2 + \sigma^2 + \dots + \sigma^2] = \frac{\sigma^2}{n}.$$

O sea, que la variancia de la distribución muestral de $\bar{X}$ es igual a la variancia de la población dividida por el tamaño de la muestra. Podemos resumir los resultados anteriores en la forma siguiente:

$$\text{Si } X \sim N(\mu, \sigma^2), \bar{X} \sim N(\mu, \sigma^2/n).$$

Este resultado tiene gran importancia y lo utilizaremos muy a menudo a lo largo del resto de esta obra.

Los resultados anteriores implican ciertas propiedades de $\bar{X}$ como estimador de la media (μ) de una población normal.

1. Como $E(\bar{X}) = \mu$, $\bar{X}$ es un estimador insesgado de μ .
2. Como $\sigma_{\bar{X}}^2 = \sigma^2/n$, la distribución de $\bar{X}$ se hace cada vez más concentrada a medida que aumenta el tamaño de la muestra.
3. Las dos propiedades anteriores implican que $\bar{X}$ es un estimador consistente de μ .
4. La distribución de $\bar{X}$ es normal y, por tanto, es perfectamente simétrica,

Estas propiedades son las mismas que las de $\bar{X}$ como estimador de la media de una población uniforme discreta, a pesar de que las distribuciones muestrales son distintas entre sí.

Otra implicación interesante de estos resultados hace referencia a la relación entre la variancia de la distribución muestral de $\bar{X}$ y la variancia de X en la población. Obsérvese que la dispersión de $\bar{X}$ depende sólo de dos factores, la dispersión de la variable X en la población y el tamaño de la muestra. Dado un tamaño muestral, cuanto menor sea la variancia de la población, nuestras conjeturas respecto a la media de la población serán menos dispersas (es decir, más fiables). Decimos «sea» para poner de relieve que el tamaño de la variancia de la población (contrariamente al tamaño de la muestra) no está bajo nuestro control —ni siquiera en teoría—. Por tanto, en caso de que quisiéramos estimar la media de cualquier variable normalmente distribuida, como por ejemplo los gastos familiares anuales de familias con rentas de nivel similar, obtendríamos una estimación más fiable si se tratara de una población menos conformista. (Como hemos visto anteriormente, la relación entre la variancia de $\bar{X}$ y la de X es la misma tanto si la población originaria es una población uniforme discreta, como si es normal. Sin embargo, el primer caso es prácticamente menos interesante y, por tanto, no hacemos ningún comentario acerca de él, por el momento.)

Con lo anterior, se da por finalizada la obtención detallada de las distribuciones muestrales de la media de la muestra para dos poblaciones originarias —la población uniforme discreta y la población normal—. Esta última es de gran importancia práctica, pero, aún así, en el trabajo econométrico se presentan a menudo casos en que el muestreo se efectúa respecto a poblaciones no normales (y no uniformes). Sin embargo, no es en absoluto posible estudiar en detalle todas las demás distribuciones, ni disponemos de criterio alguno que resulte útil para destacar algunas de ellas con preferencia a las demás, y, por lo tanto, nos ocuparemos de las distribuciones restantes en forma resumida si bien intentaremos obtener el mayor número posible de conclusiones que tengan aplicabilidad general. Afortunadamente, es posible avanzar en este sentido, estableciendo resultados que son independientes de la forma de la distribución originaria de X .

En el caso del muestreo a partir de una población normal, ya se ha obtenido un rasgo de la distribución muestral de $\bar{X}$ que no depende en absoluto de la distribución de X en la población. Ello hace referencia a la media y la varian-

de $\bar{X}$. Supongamos que la variable X posee cualquier distribución, con media μ y variancia σ^2 , y que consideramos a $X_1, X_2, \dots, X_n$ como n variables cada una de las cuales tiene exactamente la misma media y variancia que X . En este caso, la media de $\bar{X}$ viene dada por

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \mu,$$

y la variancia de $\bar{X}$ es

$$\text{Var}(\bar{X}) = \text{Var}\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right] = \frac{\sigma^2}{n}.$$

Resumimos estos dos resultados en el teorema siguiente.

Teorema 14. Si X es una variable con media μ y variancia σ^2 , cualquiera que sea la distribución de X , la distribución muestral de $\bar{X}$ posee la misma media μ y una variancia igual a σ^2/n .

Otro rasgo de la distribución muestral de $\bar{X}$, que es independiente de la distribución de X , es el presentado en el teorema que se expone a continuación. Conocido generalmente con el nombre de *teorema central del límite*, es una de las proposiciones más importantes de la teoría estadística. Su demostración puede verse en la mayor parte de los libros de texto de estadística matemática y no la efectuaremos aquí.⁶

Teorema 15. (Teorema central del límite.) Si X posee cualquier distribución con media μ y variancia σ^2 , al aumentar el tamaño de la muestra, la distribución de $\bar{X}$ se aproxima a una distribución normal con media μ y variancia σ^2/n .

Este resultado es realmente notable. Su implicación es que, cuando las muestras son grandes, podemos efectuar una aproximación a la distribución de $\bar{X}$ mediante la distribución normal, sea cual sea la distribución originaria de X . Esta implicación es la que proporciona importancia práctica al teorema central del límite.

Como aplicación de este teorema, tratemos de utilizar la distribución normal para obtener una aproximación de la distribución muestral de $\bar{X}$ basada en

6. La formulación exacta del teorema central del límite varía de un texto a otro. Aquí seguiremos la de R. L. Anderson y T. A. Bancroft, *Statistical Theory in Research*, Nueva York, McGraw-Hill, 1952, p. 71, que permite entender el teorema con mayor claridad que otras presentaciones. Sin embargo, debe ponerse de relieve que la validez de este teorema está limitada por el supuesto de que las variables X_i son independientes y proceden de una población cuyos parámetros son constantes.

muestras que provienen de una población uniforme discreta. En particular, consideremos una población originaria de X tal que ésta pueda tomar únicamente los valores 0, 1, 2, ..., 9. Esta es la misma población que nos sirvió de base para el experimento muestral expuesto en la sección 2.2. En este experimento, se construyeron las distribuciones de $\bar{X}$ para 100 muestras de tamaño 5 y 100 muestras de tamaño 10. En la primera parte de esta sección obtuvimos la fórmula exacta de la distribución de $\bar{X}$, de modo que conocemos con exactitud cuál es su distribución real. Tratemos ahora de utilizar una distribución normal (con media $\mu = 4.5$ y variancia $\sigma^2/n = 8.25/n$) como aproximación de la verdadera distribución. Aunque se supone que la aproximación sólo es válida para muestras grandes, y evidentemente las muestras de tamaño 5 ó 10 no pueden ser consideradas grandes, es interesante ver cómo funciona con muestras pequeñas.

En primer lugar, utilicemos la distribución muestral de $\bar{X}$ con muestras de tamaño 5. En este caso, la media y la desviación standard de la distribución son

$$E(\bar{X}) = 4.5$$

$$\sigma_{\bar{x}} = \sqrt{\frac{8.25}{5}} = 1.2845.$$

Por tanto, la variable normal tipificada Z —necesaria para utilizar la tabla de probabilidades normales— es

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 4.5}{1.2845} = 0.7785\bar{x} - 3.5033.$$

Cuadro 4-15

Tamaño de la muestra 5							
Intervalo: $\bar{x}$	Intervalo: z	$f(z)$ (Normal acumulada)	$f(\bar{x})$ (Aprox. normal)	$f(\bar{x})$ (teórica)	$f(\bar{x})$ (en los experimentos)		
0.5 a 1.499	$-\infty$ a -2.33	0.010	0.01	0.01	0.01		
1.5 a 2.499	-2.33 a -1.56	0.059	0.05	0.05	0.05		
2.5 a 3.499	-1.56 a -0.78	0.218	0.16	0.16	0.12		
3.5 a 4.499	-0.78 a 0.00	0.500	0.28	0.28	0.31		
4.5 a 5.499	0.00 a 0.78	0.782	0.28	0.28	0.28		
5.5 a 6.499	0.78 a 1.56	0.941	0.16	0.16	0.15		
6.5 a 7.499	1.56 a 2.33	0.990	0.05	0.05	0.05		
7.5 a 8.499	2.33 a 3.11	0.999	0.01	0.01	0.01		
8.5 a 9.499	3.11 a ∞	1.000	0.00	0.00	0.00		

Para construir la distribución normal, seguiremos el procedimiento de la sección 4.2 ilustrado en el cuadro 4-7. En el cuadro 4-15, se presentan los resultados junto a las probabilidades experimentales y teóricas, que se reproducen del cuadro 4-13. Los resultados obtenidos a partir de la aproximación normal son muy buenos. En realidad, resulta que la distribución normal que ajustamos es exactamente igual —por lo menos tomando sólo los dos primeros decimales— a la distribución exacta de $\bar{X}$ obtenida teóricamente.

A continuación, consideremos la distribución muestral con muestras de tamaño 10. La media y la desviación standard de esta distribución son

$$E(\bar{X}) = 4.5.$$

$$\sigma_{\bar{x}} = \sqrt{\frac{8.25}{10}} = 0.9083.$$

La correspondiente variable normal tipificada es

$$z = \frac{\bar{x} - 4.5}{0.9083} = 1.1010\bar{x} - 4.9549.$$

Los cálculos se presentan en el cuadro 4-16; las columnas que indican las probabilidades experimentales y teóricas están tomadas del cuadro 4-14. Los resultados indican nuevamente que la distribución normal proporciona una aproximación muy buena de la distribución muestral de $\bar{X}$. En este caso, el ajuste parece un poco peor que en el cuadro 4-15, porque, cuando el tamaño de la muestra viene dado por un número par, la distribución teórica sólo tiene un

Cuadro 4-16

Tamaño de la muestra 10					
Intervalo: $\bar{x}$	Intervalo: z	$f(z)$ (Normal acumulada)	$f(\bar{x})$ (Aprox. normal)	$f(\bar{x})$ (teórica)	$f(\bar{x})$ (en los experimentos)
0.5 a 1.499	$-\infty$ a -3.30	0.001	0.00	0.00	0.00
1.5 a 2.499	-3.30 a -2.20	0.014	0.02	0.01	0.01
2.5 a 3.499	-2.20 a -1.10	0.136	0.12	0.11	0.14
3.5 a 4.499	-1.10 a 0.00	0.500	0.36	0.35	0.34
4.5 a 5.499	0.00 a 1.10	0.864	0.36	0.37	0.32
5.5 a 6.499	1.10 a 2.20	0.986	0.12	0.14	0.16
6.5 a 7.499	2.20 a 3.30	0.999	0.02	0.02	0.03
7.5 a 8.499	3.30 a 4.40	1.000	0.00	0.00	0.00
8.5 a 9.499	4.40 a ∞	1.000	0.00	0.00	0.00

valor máximo en el punto 4.5, mientras que, cuando el tamaño de la muestra es un número impar, la distribución teórica tiene dos valores máximos iguales, uno de ellos inmediatamente antes, y el otro inmediatamente después, del punto 4.5. Tal como se han especificado los límites de los intervalos, las distribuciones basadas en una muestra par presentan una apariencia engañosa de asimetría.

Los resultados anteriores indican que la distribución normal puede proporcionar una aproximación aceptable de la distribución muestral de $\bar{X}$ incluso aunque el tamaño de la muestra sea pequeño, pero esto no es cierto en general. El rasgo específico del caso anterior que hizo que la aproximación normal diese tan buenos resultados es la simetría de la distribución originaria y, en consecuencia, la de la distribución de $\bar{X}$ para cualquier tamaño muestral. En los casos en que la distribución originaria es muy asimétrica, haría falta una muestra mayor para poder obtener resultados satisfactorios utilizando una distribución normal para $\bar{X}$, y esto es, de hecho, lo que puede esperarse legítimamente del teorema central del límite.

Para finalizar nuestro estudio de la distribución muestral de la media de la muestra podemos poner de relieve los resultados más importantes que se han conseguido. Se empezó obteniendo la distribución muestral de $\bar{X}$ con muestras procedentes de una *población uniforme discreta*. La distribución de $\bar{X}$ obtenida resultó ser perfectamente simétrica y susceptible de aproximación por medio de una distribución normal, cualquiera que fuese el tamaño de la muestra. A continuación, se descubrió que la distribución de $\bar{X}$ para muestras procedentes de una *población normal* es exactamente normal para todo tamaño muestral. Por último, a través del teorema central del límite, se llegó a la conclusión de que la distribución de $\bar{X}$ para una muestra grande será aproximadamente normal sea cual sea la población originaria. Todos estos resultados señalan la importancia de la distribución normal en la inferencia estadística. Cuando se utilizan muestras grandes, se puede confiar *siempre* en la distribución normal para describir la distribución muestral de $\bar{X}$. (Estrictamente hablando, el teorema central del límite no es válido en el caso en que la variancia de la población no es finita y su presentación en el texto se basa en el supuesto de que estos casos se han excluido de nuestro análisis. Obsérvense además las restricciones mencionadas en la nota 6 de este capítulo.) Si la población originaria no es normal, esta descripción será aproximada, y si es normal será exacta. Cuando se utilizan muestras pequeñas, la distribución normal proporcionará una descripción perfecta de la distribución de $\bar{X}$ si la población originaria es normal, y una descripción aproximada si dicha población no es normal pero es simétrica. Sólo sería desaconsejable utilizar una distribución normal como aproximación de la distribución de $\bar{X}$ si las muestras fuesen pequeñas y la población originaria es asimétrica. Por último, respecto a las propiedades de la media de la muestra como estimador de la media de la población, se halló que, sea cual sea la población originaria, la distribución de $\bar{X}$ tiene una

media igual a la media de la población y una variancia igual a la variancia de la población dividida por el tamaño de la muestra. Por consiguiente, $\bar{X}$ es un estimador insesgado y consistente de la media de la población.

EJERCICIOS:

- 4-1. En una cierta ciudad, un 20 % de los consumidores utilizan jabón de la marca D. ¿Cuáles son las probabilidades de que en un ascensor en el que se hallan diez personas, 0, 1, 2, ..., 10 de ellas utilicen dicho jabón?
- 4-2. ¿Cuántas veces tendríamos que lanzar una moneda al aire para que la probabilidad de que la proporción de caras se halle entre 0.40 y 0.60 sea por lo menos 0.95?
- 4-3. Sabiendo que el número medio de aciertos en n pruebas es 8 y que la desviación standard es 2, calcular n y π .
- 4-4. Considérese una población uniforme discreta con $X = 0, 1, \dots, 5$. Obtener la distribución muestral del recorrido (es decir, la diferencia entre la observación mayor y la menor) para muestras de tamaño 4.
- 4-5. Considérese una población uniforme continua con $0 < x < 1$. ¿Cuál es la probabilidad de que la media de 20 observaciones esté comprendida entre 0.4 y 0.6?
- 4-6. Una población está compuesta por 10 bolas, 5 de las cuales son blancas y las otras 5 rojas. Construir la distribución muestral de la proporción de bolas blancas en una muestra de 5 bolas extraídas *sin reposición*.
- 4-7. Una empresa que opera en una cierta ciudad ha sido acusada de efectuar excesivos beneficios. Como prueba de ello se ha dicho que la tasa de beneficio de la empresa en el año anterior fue del 22 %, mientras que el promedio nacional en esta industria fue del 16 %. En defensa de la empresa, los directivos afirman que en dicha industria los beneficios son muy variables y que es frecuente encontrar desviaciones importantes respecto de la media, señalando también que, aunque la tasa media de beneficios de la industria fue del 16 %, la desviación standard ha sido del 4 %. Suponiendo que los beneficios se distribuyen normalmente, ¿cuál es la probabilidad de obtener un beneficio del 22 % o mayor por razones puramente aleatorias?
- 4-8. Sabiendo que $X \sim N(8, 16)$, calcular las probabilidades siguientes:
 - a. $P(6 < X < 10)$. b. $P(10 < X < 12)$. c. $P(X < 0)$. d. $P(X > 20)$.
 - e. Calcular también los dos valores de X que delimitan el 50 % central del área situada debajo de la curva.
- 4-9. Definamos $Y =$ renta y $X = \log$ de Y . Se ha afirmado en muchas ocasiones que la distribución de X es aproximadamente normal. Comprobar esta

afirmación ajustando una curva normal a la distribución de X obtenida a partir de los datos del cuadro 4-17.

Cuadro 4-17 Renta de las familias y los individuos que viven solos (E.E.U.U., 1962)

Renta	Frecuencia (en porcentaje)
Menos de \$2,000	12.0
\$2,000 a \$2,999	8.5
\$3,000 a \$3,999	9.8
\$4,000 a \$4,999	10.2
\$5,000 a \$5,999	10.2
\$6,000 a \$7,499	14.1
\$7,500 a \$9,999	15.7
\$10,000 a \$14,999	12.3
\$15,000 o más	7.2

Fuente: *Statistical Abstract of the United States, 1965*. Ministerio de Comercio de los E.E.U.U., p. 341.

Contrastación de hipótesis

Al comienzo de esta obra se indicó que existen básicamente dos tipos de inferencias estadísticas: la estimación y la contrastación de hipótesis. En ambos casos se trata de efectuar juicios respecto a algún aspecto desconocido de una población determinada, en base a la información proporcionada por una muestra. El aspecto investigado puede ser el valor de uno o más parámetros de la población o, menos frecuentemente, la forma funcional de la población. Según el tipo de pregunta que se plantee, se tratará de un problema de estimación o de contrastación de hipótesis. En el caso de la estimación, nos preguntamos cuál es el valor de un parámetro determinado, mientras que en la contrastación de hipótesis partimos de una afirmación que hace referencia a la población y la pregunta es, en este caso, si dicha afirmación es correcta o falsa. En otros sentidos, los dos casos son bastante similares, ya que en ambas ocasiones la respuesta se obtiene combinando los conocimientos previos y los supuestos relativos a la población con la evidencia proporcionada por la muestra. También en los dos casos se efectúa un uso considerable del concepto de distribución muestral que se ha desarrollado en secciones anteriores. Por último, cualquiera que sea el tipo de pregunta, la respuesta es siempre provisional. Sin embargo, existen ciertas diferencias de enfoque que hacen aconsejable un estudio separado y, en consecuencia, dedicaremos este capítulo al problema de la contrastación de hipótesis y el siguiente a la estimación.

5.1 Establecimiento y valoración de pruebas para la contrastación

Se define como hipótesis un supuesto que hace referencia a la población. Generalmente, se efectúa más de un supuesto de este tipo, pero no todos ellos se contrastan. Los supuestos que no se pretenden contrastar se denominan *hipótesis sostenidas*, y consisten en todos los supuestos que estamos dispuestos a efectuar y a creer válidos. Naturalmente, no podemos estar nunca completamente seguros de que estos supuestos son válidos, ya que, de lo contrario, dejarían de ser supuestos para convertirse en hechos. En este sentido, la situación habitual consiste en el hecho de que estamos convencidos de que los supues-

tos se cumplen al menos de forma aproximada, de modo que las hipótesis sostenidas son casi correctas. Los supuestos restantes, sujetos a verificación, se denominan *hipótesis contrastables* y consisten normalmente en afirmaciones relativas al valor de algún parámetro de la población. En la teoría estadística, este tipo de hipótesis reciben el nombre de *hipótesis nulas*, ya que implican que no existe diferencia entre el verdadero valor del parámetro de la población y el que se supone en la hipótesis.

A modo de ejemplo, consideremos la afirmación de que «los economistas y los psicólogos gastan, por término medio, en propinas, durante sus congresos anuales, una cantidad igual». Esto se puede interpretar como una hipótesis contrastable que señala que la media de la población es la misma en los dos grupos profesionales. Normalmente, contrastaríamos esta hipótesis obteniendo una muestra aleatoria de economistas y otra de psicólogos, y comparando las medias muestrales respectivas (en la forma que estudiaremos a continuación). En este caso, la hipótesis sostenida puede estar formada por los siguientes supuestos.

1. Si existen diferencias en el comportamiento medio con respecto a las propinas, éstas se deben por completo a la profesión del individuo, y los demás factores, como pueden ser la renta, el sexo o la edad, no tienen ninguna importancia.
2. La cantidad media que se gasta en propinas en cada grupo se distribuye normalmente con la misma variancia.
3. No existe ninguna presunción a priori de que alguna de las dos medias poblacionales sea mayor que la otra.

El primero de estos supuestos implica que al efectuar la contrastación no es necesario tener en cuenta ningún otro factor aparte de la profesión. El segundo supuesto es necesario para determinar la distribución muestral del estadístico de prueba, y el último supuesto es el que determina la hipótesis alternativa a la hipótesis nula que consideramos. En este caso, esta hipótesis alternativa es la de que las medias *no* son iguales, y al plantear la contrastación es preciso especificarla.

La idea de una hipótesis alternativa es de gran importancia y conviene considerarla con más detalle. La hipótesis nula es una proposición contrastable y, por lo tanto, es necesario que exista una proposición contraria, ya que en otro caso no habría ninguna necesidad de efectuar la contrastación. Esta proposición contraria recibe el nombre de *hipótesis alternativa*. Supongamos que la hipótesis nula señala que la media de la población, μ , tiene un valor μ_0 . Normalmente, llamamos H_0 a la hipótesis nula y H_A a la alternativa. En este caso, la hipótesis alternativa puede ser, por ejemplo, la proposición de que μ es igual a otro valor, μ_A . Es decir, tendremos que

$$H_0: \mu = \mu_0.$$

$$H_A: \mu = \mu_A.$$

Si el caso es éste, la implicación es que μ puede ser igual a μ_0 o a μ_A , pero no puede tener ningún otro valor. Evidentemente, esta situación es muy poco corriente, ya que significa que tenemos realmente mucha información a priori respecto a la media de la población y que de lo único que no estamos seguros es de cuál de los dos valores es el correcto. Frecuentemente, el conocimiento previo relativo a la media de la población (o a cualquier otro parámetro de la población) es mucho más reducido. Si no sabemos absolutamente nada respecto a μ , la hipótesis alternativa será la de que μ *no* es igual a μ_0 y la representaremos de la forma siguiente

$$H_0: \mu = \mu_0,$$

$$H_A: \mu \neq \mu_0.$$

En algunos casos estamos bastante seguros de que μ es mayor (o menor) que μ_0 y en estas ocasiones tenemos

$$H_0: \mu = \mu_0,$$

$$H_A: \mu > \mu_0 \quad (\text{o } H_A: \mu < \mu_0).$$

Las hipótesis que tienen una forma general como $\mu \neq \mu_0$, $\mu > \mu_0$ o $\mu < \mu_0$ se denominan *hipótesis compuestas*, y las que presentan una afirmación específica como $\mu = \mu_0$ se llaman *hipótesis simples*.

Las afirmaciones de carácter específico son más fáciles de refutar que las que tienen un carácter vago o general y, por tanto, resulta deseable —y así ha ocurrido en la práctica común— que los problemas de contrastación de hipótesis se formulen con una hipótesis nula tan específica como sea posible. Por tanto, si tenemos dos hipótesis rivales —como suele ocurrir a menudo— y una de ellas es simple y la otra compuesta, elegimos como hipótesis nula a contrastar la simple. Ello significa que en muchas ocasiones introducimos como hipótesis nula la proposición que realmente deseamos refutar. Un buen ejemplo de ello es el correspondiente a las pruebas de un nuevo medicamento. Existe una obvia presunción de que, en términos de porcentaje medio de curaciones, el nuevo medicamento dará mejores resultados que la terapia o medicamento anteriores, ya que, en caso contrario, no tendría ningún sentido efectuar dicha contrastación. Sin embargo, la hipótesis nula es en este caso la proposición de que el nuevo medicamento produce el mismo porcentaje medio de curaciones que el viejo y la hipótesis alternativa será que este porcentaje medio es mayor en el caso del nuevo medicamento.

La decisión de cuál de las dos hipótesis rivales debe tomarse como hipótesis nula tiene algunas implicaciones que es preciso no dejar de lado. Según la

metodología establecida, una hipótesis nula es una proposición que será considerada válida a menos que la evidencia nos haga dudar seriamente de ello. En este sentido, una contrastación estadística se parece a un juicio ante los tribunales. En un juicio, el acusado es considerado inocente a menos que la evidencia sugiera que es culpable, *más allá de toda duda razonable*. Del mismo modo, la hipótesis nula se considera válida a menos que la evidencia sugiera —también más allá de toda duda razonable— que no lo es. (Sin embargo, mientras en los tribunales la definición de «duda razonable» es la misma en todos los casos, en las pruebas estadísticas puede variar según cuál sea el coste de efectuar un veredicto incorrecto.) Además, al igual que en el tribunal es el fiscal quien tiene que probar la culpabilidad del acusado, en las pruebas estadísticas es el estadístico quien tiene que probar que la hipótesis nula es incorrecta. Naturalmente, la palabra «probar» no debe tomarse en sentido absoluto en ninguno de los dos casos, puesto que siempre existe una «sombra» de duda; sólo Dios sabe si un hombre es realmente culpable o una hipótesis nula realmente incorrecta. Por último, también el procedimiento resulta similar; en el tribunal se presenta toda la evidencia y demás información que resulta relevante para el caso, y se pondera de acuerdo con las reglas establecidas por la ley, llegándose a un veredicto de «culpable» o «no culpable». Del mismo modo, cuando se efectúa una contrastación estadística, se utiliza toda la evidencia y la información previa de acuerdo con unas reglas determinadas y se llega a la conclusión de «rechazar» o «no rechazar» la hipótesis nula. Cabe señalar que, igual que el tribunal dicta su veredicto en la forma de «no culpable» en lugar de «inocente», la conclusión de una contrastación estadística consiste en «no rechazar» en lugar de «aceptar».

El paralelismo entre los juicios y las contrastaciones estadísticas finaliza de repente cuando consideramos la aplicación de la Quinta Enmienda de la Constitución de los Estados Unidos. Contrariamente a lo que ocurre con un acusado, que no debe «verse sujeto a grave riesgo de su vida o a perjuicio grave en dos ocasiones por una misma falta», las hipótesis nulas están *siempre* sujetas a la posibilidad de contrastación. En realidad, aunque las hipótesis nulas se consideran válidas a menos que tengamos evidencia clara en sentido contrario, esta opinión siempre tiene carácter provisional. En este sentido, una hipótesis nula se parece a un campeón de cualquier deporte, que en cualquier momento puede verse superado por otro deportista. En realidad, es posible imaginar el progreso de la ciencia como un proceso en el que se establecen hipótesis y a continuación se efectúa un gran esfuerzo para obtener nuevos datos que permitan refutar dichas hipótesis. Únicamente pueden sobrevivir a estos ataques continuados las hipótesis más robustas haciéndose merecedoras de nuestra confianza, al menos hasta que se produzca un nuevo ataque.

Hasta aquí no se ha mencionado en absoluto la cuestión relativa al origen de las hipótesis. En principio, este problema tiene una respuesta simple: Las hipótesis económicas se obtienen a partir de la teoría económica; sin embargo,

en la práctica, el asunto es mucho más complejo. En primer lugar, la teoría económica es en muy pocas ocasiones lo suficientemente precisa y detallada como para producir hipótesis a las que se puedan aplicar pruebas estadísticas. Por ejemplo, sería una tarea difícil tratar de encontrar en la literatura económica un análisis teórico que dé lugar a una proposición en la que se especifique un valor concreto del multiplicador del gasto público, detallando además todos los supuestos implicados en la hipótesis sostenida. Normalmente, la teoría económica sólo especifica las interrelaciones entre las distintas variables económicas describiéndolas normalmente en términos muy generales. Por tanto, el econométra debe especificar la forma matemática de estas relaciones y presentar la hipótesis sostenida de forma más completa y con mayor detalle. La hipótesis nula señala generalmente que la relación postulada *no* existe, lo cual normalmente significa que el valor de uno o más parámetros es igual a cero, mientras que la hipótesis alternativa afirma que la relación *sí* existe. La segunda dificultad presentada por la teoría económica como origen de las hipótesis es que en muchas ocasiones las variables a las que hace referencia son difíciles de definir, o de medir, o tienen ambos problemas a la vez. Por desgracia, esto no sólo ocurre en conceptos cuya dificultad es tan notoria como en el caso del capital, sino que también se aplica a conceptos claramente menos problemáticos como pueden ser la renta o el consumo. Por último, existen gran cantidad de problemas, cuya teoría económica está muy poco desarrollada, ofreciendo por tanto muy poca ayuda en la obtención de hipótesis relevantes. En estos casos, los investigadores recurren generalmente a teorizaciones *ad hoc*, es decir, establecen hipótesis sostenidas y contrastables utilizando su sentido común y toda la ayuda que puedan obtener de la teoría. La mayor parte del trabajo econométrico aplicado que se ha efectuado hasta la actualidad es de este tipo, o por lo menos puede decirse que predomina este tipo de trabajo. Todas estas dificultades para establecer hipótesis económicas explican en cierta medida el mayor grado de concentración en los problemas de estimación y la menor dedicación a los de contrastación de hipótesis, existente en la investigación econométrica.

El establecimiento de la hipótesis nula y de su alternativa representa el primer paso cuando nos ocupamos de un problema relativo a la contrastación de hipótesis. El paso siguiente consiste en diseñar un criterio que nos permita decidir, basándonos en la evidencia disponible, si vamos a rechazar o no la hipótesis nula. Esta regla o criterio es, en principio, la misma sea cual sea el problema: se define un estadístico de prueba y unos límites que dividen al espacio muestral en una región en la que se rechaza la hipótesis y otra región en la que no se rechaza. El estadístico de prueba es simplemente una fórmula que nos indica la forma de confrontar la hipótesis nula con la evidencia, y es una variable aleatoria cuyo valor varía de una muestra a otra. La *región en la que se rechaza la hipótesis* se denomina a veces *región crítica* y es un subconjunto del espacio muestral tal que si el valor del estadístico de prueba perte-

nece a él, se rechaza la hipótesis nula. Del mismo modo, la región en la que no se rechaza la hipótesis, llamada normalmente *región de aceptación*, es un subconjunto o espacio muestral tal que si el valor del estadístico de prueba pertenece a él, no se rechaza la hipótesis nula. El límite entre las regiones de aceptación y de rechazo se determina a partir de la información previa relativa a la distribución del estadístico de prueba, mediante la especificación de la hipótesis alternativa, y a partir de consideraciones relativas a los costes de la obtención de una conclusión incorrecta. Un rasgo importante de este límite es su independencia respecto a la información de la muestra; en realidad, su determinación es anterior, en sentido lógico, a la obtención de la muestra.

Criterio de contrastación

El procedimiento para establecer un criterio que permita rechazar la hipótesis nula se puede explicar haciendo referencia al caso corriente, de libro de texto, de contrastación de hipótesis. Consideremos una hipótesis nula que afirma que la media de una variable X es igual a μ_0 . Supongamos que la hipótesis sostenida está formada por los siguientes supuestos: 1) X se distribuye normalmente; 2) la variancia de X es σ^2 y la conocemos. La hipótesis sostenida no nos dice nada respecto a μ (por ejemplo, si μ puede tomar solamente valores positivos, u otra información de este tipo) y, por tanto, la hipótesis alternativa es simplemente que μ no es igual a μ_0 . Por tanto, podemos escribir

$$H_0: \mu = \mu_0.$$

$$H_A: \mu \neq \mu_0.$$

Desarrollemos ahora un estadístico de prueba adecuado a este caso. La información que obtendremos de la muestra obviamente nos dirá algo respecto a la media de la población; el problema consiste en cómo utilizar esta información. El análisis de la sección 4.3 indicó que la media de la muestra posee ciertas propiedades que resultan deseables para un estimador de la media de la población. Ello sugiere que para resumir la evidencia muestral relativa a la media de la población podemos utilizar la media de la muestra. En este caso, un criterio obvio para rechazar o no la hipótesis nula puede ser el siguiente: si el valor de $\bar{X}$ es muy distinto de μ_0 , rechazar la hipótesis; si no es muy distinto, no rechazarla.

El criterio anterior resulta claramente inútil para definir las regiones críticas y de aceptación a menos que establezcamos con exactitud qué valores de $\bar{X}$ vamos a considerar «muy distintos» de μ_0 y cuáles no. Para tomar una decisión acerca de ello es preciso considerar la distribución muestral de $\bar{X}$. Si X es normal y tiene, en realidad, una media μ_0 y una variancia σ^2 , $\bar{X}$ será una normal con media μ_0 y variancia σ^2/n (en la que n es el tamaño de la muestra). Naturalmente, como la distribución normal se extiende entre $-\infty$ y $+\infty$,

es posible observar cualquier valor de $\bar{X}$ sea cual sea la media de la población. Sin embargo, si la media correcta es μ_0 , los valores de $\bar{X}$ correspondientes a los intervalos próximos a μ_0 tendrán una mayor probabilidad que los de los intervalos (de la misma longitud) situados muy lejos de μ_0 . Es natural, pues, que consideremos «muy distintos de μ_0 » aquellos valores de $\bar{X}$ que ocurrirían por casualidad —si la verdadera media fuese μ_0 — en muy pocas ocasiones. Al decir «en muy pocas ocasiones» queremos decir, por lo menos de momento, «con probabilidad 0.01».

Al llegar a este punto, es necesario introducir en la discusión la hipótesis alternativa que afirma que $\mu \neq \mu_0$. En realidad, esto significa que, si la hipótesis nula no se cumple, la verdadera media puede estar en cualquier lado de μ_0 . Por lo tanto, los valores de $\bar{X}$ muy superiores a μ_0 así como los valores mucho más pequeños constituirían evidencia en contra de la hipótesis nula. (Si la hipótesis alternativa fuese, por ejemplo, $\mu > \mu_0$, únicamente los valores de $\bar{X}$ mucho mayores que μ_0 representarían evidencia en contra de la hipótesis nula.) Es decir, los límites entre la región crítica y la región de aceptación deben ser tales que tengamos que rechazar la hipótesis nula si el valor de $\bar{X}$ resulta ser demasiado pequeño o demasiado grande comparado con μ_0 de forma que su aparición por casualidad sea muy poco probable. Como hemos decidido llamar muy poco probable (o muy raro) a un suceso si su probabilidad es únicamente 0.01, dicha probabilidad debe estar «repartida» por igual entre los valores demasiado pequeños de $\bar{X}$ y los demasiado grandes. En otras palabras, los límites deben establecerse de tal forma que la probabilidad de que $\bar{X}$ sea demasiado pequeña es 0.005 y la probabilidad de que $\bar{X}$ sea demasiado grande es también 0.005. Por tanto, según el teorema aditivo (teorema 2), la probabilidad de que $\bar{X}$ sea demasiado grande o demasiado pequeña, en comparación con μ_0 , será $0.005 + 0.005 = 0.01$ tal como deseamos.

Llamemos μ_L al valor por debajo del cual consideraremos que $\bar{X}$ es demasiado pequeña y μ_H al valor por encima del cual consideraremos que $\bar{X}$ es demasiado grande. En la figura 5-1 se han señalado estos puntos correspondientes a la distribución muestral de $\bar{X}$. En este caso

$$P(\bar{X} < \mu_L) = 0.005,$$

$$P(\bar{X} > \mu_H) = 0.005;$$

y por lo tanto,

$$P(\mu_L \leq \bar{X} \leq \mu_H) = 0.99.$$

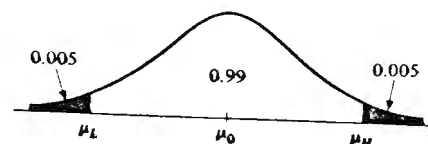


Figura 5-1

En realidad, podríamos considerar a X como un estadístico de prueba adecuado a este caso, y tomar como región de aceptación el intervalo entre μ_L y μ_H , pero esto no resulta muy práctico porque no sabemos cuál es la posición de μ_L ni la de μ_H . Sin embargo, podemos obtener con facilidad la posición de los puntos de una distribución normal tipificada, que constituye la contrapartida de los previamente mencionados, ya que en este caso disponemos de tablas para las probabilidades. El procedimiento a utilizar para ello se explicó con detalle en la sección 4.2.

Dado que $X \sim N(\mu_0, \sigma^2/n)$, la variable normal tipificada correspondiente será

$$(5.1) \quad Z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}.$$

Y éste será el estadístico de prueba.

Ahora queremos hallar los valores de z que corresponden a μ_L y μ_H en la escala de x . Estos valores de Z , a los que llamaremos z_L y z_H , se pueden obtener considerando que deben delimitar a cada lado 0,005 del área situada de-

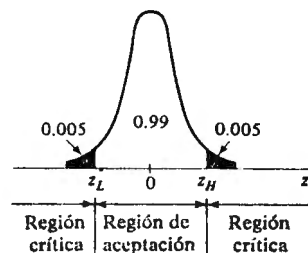


Figura 5-2

bajo de la curva, como indica la figura 5-2. Es decir, a partir de las tablas de la probabilidad normal debemos obtener el valor z_L tal que

$$P(z < z_L) = 0.005,$$

o, lo que es lo mismo,

$$P(z \geq z_L) = 0.995.$$

Del mismo modo, tenemos que obtener z_H tal que

$$P(z > z_H) = 0.005,$$

o, del mismo modo,

$$P(z \leq z_H) = 0.995.$$

Consultando las tablas de la probabilidad normal, obtenemos

$$z_L = -2.575$$

$$z_H = +2.575$$

Ello completa nuestra tarea. El intervalo comprendido entre -2.575 y $+2.575$ representa la región de aceptación, y los intervalos entre $-\infty$ y -2.575 , y entre $+2.575$ y $+\infty$ constituyen la región crítica. (La probabilidad de obtener un valor de Z igual a z_L o a z_H es cero y, por tanto, el hecho de que la región de aceptación incluya o no los valores límites no tiene importancia práctica.) El criterio para rechazar o no la hipótesis nula es, pues, el siguiente:

$$\text{rechazarla si } \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} < -2.575 \text{ o si } \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > +2.575;$$

$$\text{no rechazarla si } -2.575 \leq \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \leq +2.575.$$

Esta división entre la región crítica y la región de aceptación depende de nuestra decisión anterior en la que sólo hemos considerado como «muy distintos» de μ_0 los valores de X cuya probabilidad de ocurrir por casualidad es 0,01. En otras palabras, si extraemos un número infinito de muestras de una población cuya media es μ_0 , sólo en un 1% de los casos obtendríamos un valor de X que nos llevaría a rechazar, de forma incorrecta, la hipótesis nula. Esta probabilidad se conoce con el nombre de *nivel de significación* de la prueba. Naturalmente, no hay nada sagrado en esta cifra del 1% que hemos elegido; en la ausencia de toda información, hubiésemos podido elegir cualquier otra cifra. Si hubiéramos elegido un porcentaje mayor, la región de aceptación hubiese sido más estrecha, y la región crítica más amplia. En realidad, el 1% es uno de los niveles de significación «populares», pero esto es lo único que puede decirse a su favor en este momento. Antes de finalizar esta sección, estudiaremos la posibilidad de efectuar una elección más racional del nivel de significación, pero consideremos primero un ejemplo numérico del procedimiento de contrastación que acabamos de examinar.

111 MPLO. Los estudios psicológicos indican que en la población en general, la inteligencia —medida por el cociente intelectual (IQ)— se distribuye normalmen-

te con media 100 y desviación standard 16. Supongamos que queremos contrastar si una subpoblación determinada —por ejemplo, los zurdos— tienen una media distinta. En la hipótesis sostenida suponemos que entre los zurdos la inteligencia se distribuye normalmente y con una desviación standard igual a la de la población global, es decir, 16. Llamemos μ al IQ medio de los zurdos. La hipótesis nula y la alternativa serán

$$H_0: \mu = 100,$$

$$H_A: \mu \neq 100.$$

El estadístico de prueba será

$$\frac{(\bar{X} - 100)\sqrt{n}}{16}.$$

Como nivel de significación, elegimos el 5% que es otro de los niveles «populares»; es decir, si el valor de X es tan distinto de la media 100 considerada en la hipótesis nula, que sólo ocurriría por casualidad en un 5% de los casos, consideraremos que dicha hipótesis nula es improbable y, por tanto, la rechazaremos. Para encontrar los límites de la región de aceptación tenemos que localizar los valores de la variable normal tipificada que dejan en cada extremo de la distribución un 2,5% del área total, tal como se indica en la figura 5-3. A partir de las tablas de la probabilidad normal, obtenemos los puntos

$$z_L = -1.96,$$

$$z_H = +1.96.$$

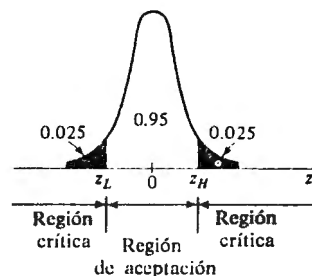


Figura 5-3

Y, en consecuencia, el criterio es:

$$\text{no rechazar } H_0 \text{ si } -1.96 \leq \frac{(\bar{x} - 100)\sqrt{n}}{16} \leq +1.96;$$

rechazarla en los demás casos.

Al llegar aquí, podemos extraer una muestra y calcular los valores de X y del estadístico de prueba. Supongamos que la muestra consiste en 400 observaciones correspondientes a personas zurdas y que el IQ medio es 99. En este caso, el estadístico de prueba será

$$\frac{(99 - 100)\sqrt{400}}{16} = -1.25.$$

Es evidente que este resultado pertenece a la región de aceptación, de modo que la muestra no proporciona ninguna evidencia en contra de la hipótesis nula. En otras palabras, no existe ninguna evidencia de que el IQ medio de las personas zurdas sea distinto del de la población en general. Con esto finaliza la respuesta al problema.

En lo expuesto hasta ahora no hemos efectuado supuesto alguno que nos sirva de ayuda en la formulación de la hipótesis alternativa y por ello dicha hipótesis tenía que ser de la forma $\mu \neq \mu_0$. A consecuencia de ello, la región en la que se rechazaba la hipótesis nula comprendía los dos extremos de la distribución del estadístico de prueba. Las contrastaciones con este tipo de región crítica reciben el nombre de *contrastaciones de dos extremos*. Sin embargo, en ciertas ocasiones es posible efectuar algunos supuestos que permiten una especificación menos general de la hipótesis alternativa. Concretamente, en algunas ocasiones tenemos casi la certeza de que la única alternativa posible a la afirmación de que $\mu = \mu_0$ es la afirmación de que μ es mayor (o menor) que μ_0 . Por ejemplo, se ha afirmado que la propensión marginal al ahorro (PMA) de los «perceptores de beneficios» es mayor que la de los trabajadores. En este caso, la hipótesis nula sería la de que ambas propensiones son iguales y la hipótesis alternativa sería la afirmación de que la PMA de los perceptores de beneficios es mayor. En estas ocasiones, los valores de X (y, por lo tanto, los del estadístico de prueba) que se considerarían como prueba en contra de la hipótesis nula se concentrarían únicamente en uno de los extremos de la distribución. Las contrastaciones de este tipo reciben el nombre de *contrastaciones de un extremo*.

Con el fin de ilustrar este punto, consideremos de nuevo una variable $X \sim N(\mu, \sigma^2)$ cuya variancia es conocida. Supongamos que las hipótesis nula y alternativa se pueden representar en la forma siguiente:

$$H_0: \mu = \mu_0,$$

$$H_A: \mu > \mu_0.$$

Como estadístico de prueba utilizaremos

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma},$$

como indica la ecuación (5.1). Supongamos ahora que deseamos efectuar la contrastación a un nivel de significación del 1 %, es decir, que rechazaremos la hipótesis nula si el valor de X es tan superior al de μ_0 que la probabilidad de que ocurra de forma casual es solamente 0.01. En este caso, la región

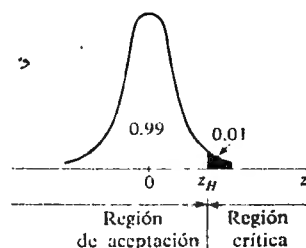


Figura 5-4'

crítica y la de aceptación son las que indica la figura 5-4'. El valor de z_H que se puede obtener en las tablas de probabilidad normal es 2.327 y, por tanto, en este caso, el criterio para rechazar H_0 es

$$\text{rechazarla si } \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > 2.327;$$

no rechazarla en los demás casos.

Si hubiésemos elegido un nivel de significación del 5 % en lugar del 1 %, el criterio sería

$$\text{rechazarla si } \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > 1.645;$$

no rechazarla en los demás casos.

Tipos de error

El criterio para rechazar o no la hipótesis nula, basándonos en la evidencia proporcionada por la muestra, no constituye una garantía de que lleguemos a la conclusión correcta. Con idénticos de forma detallada las clases de error que podemos realizar. Supongamos que tenemos que contrastar una hipótesis respecto a la media de la población, igual que en el caso anterior. La solución de este problema está compuesta esencialmente de dos partes: establecimiento de los límites entre la región crítica y la región de aceptación, y obtención del valor muestral del estadístico de prueba. Los resultados posibles son dos: el valor del estadístico de prueba puede estar incluido en la región de aceptación

o no estarlo. Tomemos en primer lugar la segunda posibilidad. En este caso, el estadístico de prueba tiene un valor tal que, si la hipótesis nula fuese en realidad cierta, la probabilidad de que dicho valor se hubiese obtenido por casualidad sería muy pequeña, por ejemplo, el 5 % o el 1 %. Ello significa que, si repitiésemos esta contrastación un número infinito de veces y la hipótesis nula fuese en realidad cierta, *rechazaríamos de forma incorrecta* esta hipótesis nula en un 5 % o un 1 % (o cualquiera que sea el nivel de significación) de las ocasiones. Este error recibe el nombre de *error del tipo I*. Con anterioridad, hemos comparado las contrastaciones estadísticas con los juicios ante los tribunales, en los que la inocencia del acusado (la hipótesis nula) se ve desafiada por la afirmación de culpabilidad realizada por el fiscal (la hipótesis alternativa). Siguiendo con este paralelismo, el error del tipo I estaría representado por el error cometido al condenar a un inocente. En la contrastación estadística, la probabilidad de cometer este error está determinada exactamente por el nivel de significación elegido. Consideremos ahora el otro posible resultado de la contrastación, o sea, el caso en que el valor del estadístico de prueba está comprendido en la región de aceptación. En este caso, no rechazamos la hipótesis nula, sino que seguimos creyendo que dicha hipótesis es cierta. Sin embargo, no podemos evitar la posibilidad de que esta conclusión sea errónea, y que la hipótesis nula sea en realidad falsa. Esta clase de error recibe el nombre de *error del tipo II*. En términos del paralelismo con un juicio, el error del tipo II equivaldría a absolver a un culpable. Normalmente, en la contrastación estadística se desconoce la probabilidad exacta de este tipo de error.

Podemos ilustrar fácilmente la idea general subyacente en los dos tipos de error en el caso —desgraciadamente poco frecuente— en que se contrasta una hipótesis nula simple con un alternativa también *simple*. Supongamos que las hipótesis son

$$H_0: \mu = \mu_0,$$

$$H_A: \mu = \mu_A,$$

en las que μ_0 y μ_A son unas cantidades determinadas y, además, $\mu_A > \mu_0$. Al igual que en el caso anterior, suponemos que X es una normal y que conocemos su variancia σ^2 . Por tanto, podemos identificar a las dos hipótesis con dos poblaciones alternativas, las dos normales y con una misma variancia σ^2 pero con medias distintas. Para una muestra de tamaño n , cada población genera su propia distribución muestral de X . Para efectuar la contrastación debemos establecer los límites entre la región crítica y la de aceptación. Como sabemos, estos límites dependerán del nivel de significación que elijamos y de la hipótesis alternativa. El nivel de significación puede definirse, a priori, tomando por ejemplo el 5 %. Como la hipótesis alternativa es que $\mu = \mu_A$ y, además, $\mu_A > \mu_0$, los únicos valores de X que constituirán evidencia en contra de H_0

son los valores muy superiores a μ_0 . Es decir, en este caso la contrastación apropiada es la que sólo tiene en cuenta un extremo, y la región crítica se concentrará en el extremo derecho de la distribución. Teniendo presente estas consideraciones, podemos determinar el límite entre la región crítica y la región de aceptación correspondiente al estadístico de prueba

$$Z = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$$

Para un nivel de significación del 5% cuando la región crítica se concentra en el extremo derecho de la distribución, el valor límite z_H es igual a 1,645 y, por tanto, la región de aceptación será

$$z \leq 1.645.$$

Si la verdadera media es μ_0 , la probabilidad de que un valor de X se encuentre en la región de aceptación es 0,95. Para obtener la probabilidad del error del tipo II, tenemos que calcular la probabilidad de que un valor de X se encuentre en la región de aceptación cuando la verdadera media no es μ_0 sino μ_A . Podemos calcular esta probabilidad del modo siguiente: Observemos en primer lugar que

$$P(z > 1.645) = 0.05$$

se puede escribir en la forma

$$P\left[\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} > 1.645\right] = 0.05.$$

Consideremos la desigualdad del interior del corchete. Si multiplicamos ambos lados de dicha desigualdad por $\sigma/\sqrt{n}$, obtenemos

$$(\bar{x} - \mu_0) > 1.645 \left[\frac{\sigma}{\sqrt{n}} \right].$$

Sumando μ_0 a ambos lados obtenemos

$$\bar{x} > \mu_0 + 1.645 \left[\frac{\sigma}{\sqrt{n}} \right].$$

Y, por tanto, podemos escribir

$$P\left[\bar{x} > \mu_0 + 1.645 \left(\frac{\sigma}{\sqrt{n}} \right)\right] = 0.05,$$

que es una afirmación probabilística que hace referencia a X y no a Z . Por tanto, el límite entre la región crítica y la región de aceptación en el eje de las x viene dado por $[\mu_0 + 1.645(\sigma/\sqrt{n})]$. Este límite tendrá un valor conocido porque sabemos cuáles son los valores de μ_0 , σ y n . En la figura 5-5 se representan las dos distribuciones y los correspondientes puntos límites.

Consideremos ahora la distribución muestral de X para muestras de la población cuya media es μ_A . Esta distribución será una normal con media igual a la media de la población (es decir, μ_A) y variancia σ^2/n . Por tanto, la única

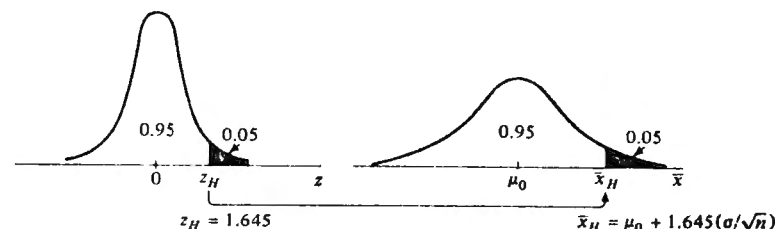


Figura 5-5

diferencia entre esta distribución y la que procede de la población cuya media es μ_0 hace referencia a las medias de estas dos distribuciones. En la figura 5-6 se comparan las dos distribuciones de X y se pueden ver claramente las probabilidades de los dos tipos de error que aparecen en la contrastación de hipótesis. Recuérdese que el error del tipo I es el que se comete al rechazar H_0 cuando ésta es en realidad la hipótesis correcta. Ello ocurre cuando $\bar{x}$ se encuentra a la derecha del punto límite $\bar{x}_H$. Obsérvese que si la hipótesis correcta es H_0 , $\bar{X}$ tiene la distribución 1 de la figura 5-6 y, por tanto, la probabilidad del error del tipo I viene dada por el nivel de significación elegido (0,05) y corresponde al área en negro. El error del tipo II lo efectuamos cuando no rechazamos H_0 y ésta es, en realidad, falsa. Ello ocurre siempre que x se encuentra a la izquierda de $\bar{x}_H$. En este caso, si H_0 fuese falsa, es decir, si la verdadera media no fuese μ_0 , la única posibilidad alternativa es que la verdadera media sea μ_A , pero siendo así, la media muestral $\bar{X}$ tendría la distribución 2 y, por tanto, la probabilidad de cometer el error del tipo II viene dada por el

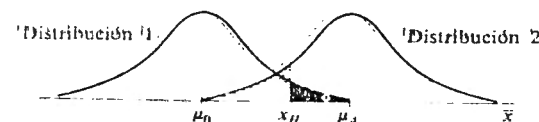


Figura 5-6

área rayada de la figura 5-6. En realidad, no sabemos cuál es la verdadera media y, por lo tanto, no podemos saber cuál es la distribución correcta de $\bar{X}$. Si la verdadera distribución de $\bar{X}$ es «1», la contrastación dará como resultado una conclusión incorrecta en el 5 % de los casos (o cualquier otro nivel de significación que elijamos). Si la verdadera distribución de $\bar{X}$ es «2», la contrastación dará resultados incorrectos con una probabilidad correspondiente al área rayada a la izquierda de $\bar{x}_H$.

Aunque la discusión anterior se limita al caso de una hipótesis alternativa simple, pone de relieve una importante faceta de la contrastación de hipótesis que es igualmente relevante en los casos en los que existen alternativas compuestas. Se trata del hecho de que *al reducir la probabilidad de un tipo de error aumentamos la probabilidad del otro*. Por ejemplo, podemos hacer que la probabilidad del error del tipo I (rechazar H_0 cuando en realidad es cierta) sea tan pequeña como queramos si establecemos un nivel de significación muy pequeño. En relación a la figura 5-6 esto implica desplazar el punto límite $\bar{x}_H$ hacia la derecha. Pero cuando hacemos esto estamos aumentando el área rayada, que representa la probabilidad del error del tipo II (no rechazar H_0 cuando en realidad es falsa). Del mismo modo, podemos reducir la probabilidad del error del tipo II si aumentamos el nivel de significación, es decir, si desplazamos $\bar{x}_H$ hacia la izquierda, pero entonces aumentaríamos la probabilidad del error del tipo I. Refiriéndonos de nuevo a la comparación entre la contrastación estadística y los juicios, es evidente que podríamos disminuir la probabilidad de condenar a un inocente absolviendo a casi todos los acusados, pero entonces aumentaríamos la probabilidad de no condenar a los culpables. Del mismo modo, podríamos reducir la probabilidad de absolver a un culpable exigiendo una evidencia menos rigurosa para dictar una condena, pero ello aumentaría la probabilidad de condenar a los inocentes. En la contrastación estadística, la única forma para reducir a la vez las probabilidades de los dos tipos de error consiste en aumentar el tamaño de la muestra (suponiendo que el estadístico de prueba que utilizamos es el mejor que podemos emplear).

EJEMPLO. Un fabricante produce dos clases de neumáticos, uno de los cuales tiene una esperanza de vida de 25 000 millas, mientras que la esperanza de vida del otro es de 30 000 millas. La variación de la duración en torno a la esperanza de vida es la misma para las dos clases, siendo la desviación standard de 3 000 millas. Puede suponerse que la distribución es normal. Las dos clases de neumático pueden distinguirse solamente por una contraseña impresa en ellos. Al efectuar el inventario se descubre que por error se han dejado 100 neumáticos sin contraseña. El encargado de efectuar el control cree que los neumáticos son de la clase de menor duración, pero recomienda efectuar una prueba. Esta prueba se va a efectuar con una muestra de 4 neumáticos. ¿Cuál debe ser el criterio de contrastación adecuado?

En primer lugar, especifiquemos la hipótesis nula y la alternativa del modo siguiente:

$$H_0: \mu = 25,000,$$

$$H_A: \mu = 30,000.$$

A continuación, establezcamos el límite entre la región de aceptación y la región crítica. Este dependerá del nivel de significación que escojamos. Vamos a considerar los niveles de significación del 1 %, 5 % y 10 %, determinando las probabilidades de error del tipo II correspondientes a cada uno de estos niveles de significación. Como $\mu_0 (= 25\,000) < \mu_A (= 30\,000)$, utilizaremos la región crítica en el extremo superior en todos los casos.

1. Si el nivel de significación es el 1 %, el punto límite en la escala de las z (en la distribución normal tipificada) será el punto que limita al 1 % superior del área que se halla debajo de la curva. A partir de las tablas de probabilidad normal podemos ver que este punto es el

$$z_H = 2.327.$$

Para obtener la probabilidad del error del tipo II tenemos que hallar el punto correspondiente en la escala de las $\bar{x}$, es decir $\bar{x}_H$. Ya sabemos, por el análisis anterior, que

$$\bar{x}_H = \mu_0 + 2.327 \frac{\sigma}{\sqrt{n}}.$$

Sustituyendo, obtenemos

$$\bar{x}_H = 25,000 + 2.327 \left[\frac{3000}{\sqrt{4}} \right] = 28,490.5.$$

Ahora tenemos que calcular la probabilidad de que $\bar{x} < 28\,490.5$ sabiendo que la media de $\bar{X}$, es decir, $E(\bar{X})$ es 30 000. Para ello tenemos que efectuar la transformación correspondiente y pasar a la variable normal tipificada. Podemos escribir que

$$\begin{aligned} P[\bar{x} < 28,490.5 | E(\bar{X}) = 30,000] &= P\left[z < \frac{(28,490.5 - 30,000)\sqrt{4}}{3000}\right] \\ &= P[z < -1.0063]. \end{aligned}$$

En las tablas de probabilidad normal hallamos

$$P[z < -1.0063] = 0.1571.$$

Esta es la probabilidad de efectuar un error del tipo II.

2. Si el nivel de significación elegido es el 5 % tenemos

$$z_H = 1.645,$$

$$y \quad \bar{x}_H = 25,000 + 1.645 \left[\frac{3000}{\sqrt{4}} \right] = 27,467.5.$$

La probabilidad de que $\bar{x} < 27,467.5$ cuando la media de $\bar{X}$ es 30 000 es

$$P[\bar{x} < 27,467.5 \mid E(\bar{X}) = 30,000] = P\left[z < \frac{(27,467.5 - 30,000)\sqrt{4}}{3000}\right] \\ = P[z < -1.6883] = 0.0457.$$

3. Por último, si el nivel de significación es el 10 %

$$z_H = 1.280,$$

$$y \quad \bar{x}_H = 25,000 + 1.280 \left[\frac{3000}{\sqrt{4}} \right] = 26,920.$$

Por tanto,

$$P[\bar{x} < 26,920 \mid E(\bar{X}) = 30,000] = P\left[z < \frac{(26,920 - 30,000)\sqrt{4}}{3000}\right] \\ = P[z < -2.0533] = 0.0200.$$

Resumiendo, los resultados son los siguientes

z	Límite $\bar{x}$	Probabilidad de	
		Error del tipo I	Error del tipo II
2,327	28 490,5	0,01	0,1571
1,645	27 467,5	0,05	0,0457
1,280	26 920,0	0,10	0,0200

Estos resultados indican que, como cabía esperar, las dos probabilidades están inversamente relacionadas. El problema es, pues, cuál de estos pares de probabilidades debemos tomar como elección óptima. La respuesta depende del coste de efectuar cada uno de estos dos tipos de error. Si el error consistente en rechazar la hipótesis nula cuando en realidad es cierta (error del tipo I) es caro en relación al error consistente en no rechazar la hipótesis nula cuando en realidad es falsa (error del tipo II), será racional que establezcamos una probabilidad pequeña del primer tipo de error. Si, por el contrario, el coste de cometer un error del tipo I es pequeño en comparación con el coste del error del tipo II, resultará deseable que la probabilidad del primer tipo de error sea grande (reduciendo, por lo tanto, la probabilidad del segundo tipo).

Se puede ilustrar este punto ampliando la información del ejemplo considerado. Supongamos que el fabricante vende los neumáticos de menor duración a 10 dólares, mientras que el precio de los más duraderos es de 12 dólares. Supongamos

además, que los neumáticos de menor duración están garantizados con «devolución del dinero» si deben reemplazarse antes de las 25 000 millas, sin ninguna garantía para la otra clase de neumáticos. Con esta información podemos estimar el coste de cada tipo de error.

Coste del error del tipo I. Si rechazamos la hipótesis de que $\mu = 25 000$, los neumáticos se venderán al precio mayor. Esto representa un beneficio de $110 \times (12 - 10) = 200$ dólares. Por otra parte, si la verdadera media es en realidad 25 000 millas, la mitad de los neumáticos (en la población) tendrán una duración inferior. Por tanto, en una partida de 100 neumáticos el número esperado de devoluciones será 50 y el desembolso correspondiente se elevará a $50 \times 12 = 600$ dólares. El coste total de cometer el error del tipo I es, pues, $600 - 200 = 400$ dólares.

Coste del error del tipo II. Si no se rechaza la hipótesis de que la media es 25 000 millas cuando en realidad esta media es de 30 000 millas, los neumáticos se venderán por un total de 1 000 dólares en lugar de 1 200. Sin embargo, esta última cifra debe ajustarse a la baja porque la garantía representa en realidad una reducción de precio. Si la media es de 30 000 millas, un 4,78 % de los neumáticos durarán menos de 25 000 millas. Esto se debe a que

$$P\left[z < \frac{25,000 - 30,000}{3000}\right] = 0.0478.$$

Por tanto, en una partida de 100 neumáticos de la clase más duradera, el número esperado de devoluciones será 4,78. En este caso, el coste de la garantía es $4,78 \times 12 = 57,36$ dólares y el ingreso esperado por la venta de estos 100 neumáticos es $1 200 - 57,36 = 1 142,64$. El coste total neto de cometer el error del tipo II es, pues, $1 142,64 - 1 000 = 142,64$ dólares.

Al disponer de la información adicional respecto a este problema descubrimos que el error del tipo I resulta más costoso que el error del tipo II y, por tanto, el comportamiento racional será establecer una probabilidad menor del error del tipo I respecto a la del otro tipo. La medida en que la primera probabilidad debe ser menor la podemos obtener comparando las pérdidas esperadas correspondientes a cada nivel de significación. Se define como «pérdida esperada» el importe de la pérdida multiplicado por la probabilidad de que se produzca. Por ejemplo, si la pérdida es de 400 dólares y la probabilidad de que dicha pérdida ocurra es 0,01, la pérdida esperada será 4 dólares. Los valores de las pérdidas esperadas en este ejemplo son las siguientes

Error del tipo I		Error del tipo II	
Probabilidad	Pérdida esperada	Probabilidad	Pérdida esperada
0,01	4,00 dólares	0,1571	22,41 dólares
0,05	20,00 »	0,0457	6,52 »
0,10	40,00 »	0,0200	2,85 »

Si suponemos que el fabricante no obtiene ninguna utilidad del hecho de correr un riesgo, la elección racional será la que produzca igual pérdida esperada para cada tipo de error. Las cifras anteriores indican que esto ocurrirá en algún punto

entre los niveles de significación del 1 % y el 5 %. Si efectuamos los cálculos correspondientes a los niveles de este intervalo, obtenemos los resultados siguientes:

Error del tipo I		Error del tipo II	
Probabilidad	Pérdida esperada	Probabilidad	Pérdida esperada
0,02	8,00 dólares	0,1006	14,35 dólares
0,03	12,00 »	0,0732	10,44 »
0,04	16,00 »	0,0567	8,09 »

Resulta, pues, que el nivel de significación óptimo está entre el 2 % y 3 %. Si lo fijamos en el 2,75 % la pérdida esperada del error del tipo I será aproximadamente igual a la del error del tipo II (unos 11,00 dólares aproximadamente). Este será, pues, el nivel de significación óptimo en este problema.

El ejemplo anterior pone de relieve la importancia de considerar las implicaciones de las distintas decisiones efectuadas al establecer contrastaciones estadísticas en términos de coste. La teoría estadística moderna pone un énfasis especial en este punto y proporciona un marco formal para la incorporación de las funciones de pérdidas y de riesgo en la determinación del criterio de contrastación adecuado. Desgraciadamente, sólo una parte muy pequeña de ello tiene interés o utilidad para la investigación econométrica ya que las ideas previas (o incluso posteriores) acerca de las pérdidas debidas a conclusiones incorrectas son o bien del todo inexistentes o tan generales que no ofrecen ninguna guía. Por ejemplo, consideremos el problema de si los activos líquidos afectan o no al gasto en consumo. El econométra especificará la hipótesis sostenida y la contrastable, y calculará el valor del estadístico de prueba adecuado, pero no tiene ninguna idea en absoluto acerca de cuál es el coste —para él, para la profesión, para la sociedad— de obtener una conclusión incorrecta. En consecuencia, en la econometría se ha convertido en práctica corriente la utilización del enfoque tradicional de la estadística clásica, es decir, fijar el nivel de significación al 1 % o al 5 % y utilizar un estadístico de prueba que haga lo más pequeña posible la probabilidad del error de tipo II. Si el estadístico de prueba toma un valor comprendido en la región crítica delimitada por el nivel de significación del 5 %, se dice que es *significativo*; si el valor del estadístico de prueba está comprendido en la región crítica correspondiente al nivel de significación del 1 % se dice que es *altamente significativo*. Sin embargo, estos dos niveles de significación no tienen otra ventaja especial aparte del hecho de que son ampliamente utilizados y es sólo esta popularidad la que elimina la competencia por parte de otros niveles de significación distintos.

Potencia de la contrastación

Al estudiar la clasificación de los errores en los dos tipos posibles se ha considerado el caso en que la hipótesis nula y la hipótesis alternativa eran sim-

ples. La idea se puede generalizar con facilidad al caso más corriente en que se contrasta una hipótesis nula con respecto a una alternativa compuesta. Supongamos que tenemos

$$H_0: \mu = \mu_0,$$

$$H_A: \mu \neq \mu_0.$$

Igual que en el caso anterior, la probabilidad del error del tipo I (rechazar H_0 siendo cierta) viene dada por el nivel de significación. Sin embargo, la probabilidad del error del tipo II (no rechazar H_0 siendo falsa) ya no es una cantidad determinada para cada nivel de significación, sino que depende del valor de μ . La probabilidad del error del tipo II para valores de μ próximos a μ_0 es grande en comparación con la probabilidad de este mismo error para valores de μ que están alejados de μ_0 . Por ejemplo, si la hipótesis nula afirma que la media es igual a 10, la probabilidad de no rechazar H_0 es claramente mayor si la verdadera media es 15 que si es 20. Ello se ilustra en la figura 5-7 en la que la probabilidad del error del tipo II está indicada por el área rayada.

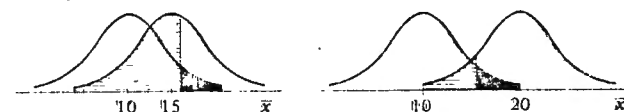


Figura 5-7

Como es lógico, podemos calcular la probabilidad del error del tipo II para cualquier valor de μ . Cuanto menor sea esta probabilidad, mejor será la contrastación para discriminar entre hipótesis ciertas y falsas. En el caso representado en la figura 5-7, la contrastación discriminará con mayor claridad entre la hipótesis nula y la alternativa si esta última es 20 que en el caso de que sea 15. En la terminología estadística corriente, diremos que cuanto menor sea la probabilidad de no rechazar H_0 cuando ésta sea falsa, más *potente* es la contrastación. Es decir, la potencia de una contrastación se mide por la *probabilidad de rechazar H_0 cuando sea falsa*. La probabilidad del error del tipo II es la probabilidad de no rechazar H_0 cuando ésta es falsa y, por tanto, la potencia de una contrastación es igual a

$$1 - P(\text{Error del tipo II}).$$

Además, como en el caso de alternativas compuestas la probabilidad del error del tipo II depende del valor de μ , la potencia de la contrastación también depende de μ . Si representamos gráficamente las probabilidades de rechazar H_0 cuando es falsa en el eje de ordenadas y los valores de μ en el eje de abscisas, obtenemos la llamada *función de potencia* de una contrastación.

EJEMPLO. Consideremos las siguientes hipótesis correspondientes a la media de una variable $X \sim N(\mu, 81)$

$$H_0: \mu = 10,$$

$$H_A: \mu \neq 10.$$

Supongamos que el estadístico de prueba se basa en una muestra de tamaño 9 y que el nivel de significación elegido es el 5%. La contrastación adecuada es la de dos extremos y la región de aceptación viene dada por

$$-1.96 \leq \frac{(\bar{x} - 10)\sqrt{9}}{\sqrt{81}} \leq 1.96$$

$$\text{o} \quad -1.96 \leq \frac{\bar{x} - 10}{3} \leq 1.96.$$

En la escala de las x , la región de aceptación equivalente es:

$$4.12 \leq \bar{x} \leq 15.88.$$

Para obtener la función de potencia tenemos que calcular $[1 - P(\text{Error del tipo II})]$ para la región de aceptación y para los distintos valores de μ . Empecemos con algún valor muy pequeño de μ , por ejemplo -10 . En este caso

$$\begin{aligned} P(\text{Error del tipo II}) &= P(4.12 \leq \bar{x} \leq 15.88 | \mu = -10) \\ &= P\left[\frac{4.12 - (-10)}{3} \leq z \leq \frac{15.88 - (-10)}{3}\right] \\ &= P(4.707 \leq z \leq 8.627) = 0.0000, \end{aligned}$$

en la que z es, igual que en el caso anterior, el valor de la variable normal tipificada. El valor de la función de potencia es, pues, $1 - 0.0000 = 1.0000$. Si efectuamos los cálculos correspondientes a otros valores de μ obtenemos los resultados siguientes:

μ	Potencia	μ	Potencia
-10	1.0000	10	0.0500
-5	0.9988	12	0.1022
0	0.9152	14	0.2659
2	0.7602	16	0.5159
4	0.5159	18	0.7602
6	0.2659	20	0.9152
8	0.1022	25	0.9988

Esta función de potencia se representa gráficamente en la figura 5-8. El gráfico confirma nuestra afirmación de que la potencia de la contrastación aumenta a medida que μ se aleja de μ_0 en la dirección (o direcciones) especificada por la hipótesis

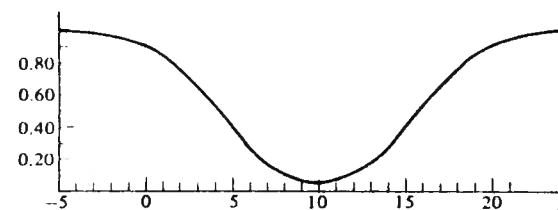


Figura 5-8

alternativa. En este caso, vemos que si el verdadero valor de μ fuese -5 o $+25$, rechazaríamos correctamente la hipótesis falsa prácticamente en todos los casos. Si el verdadero valor de μ fuese 0 o 20, rechazaríamos correctamente la hipótesis falsa en un 91,5 % de los casos. Obsérvese que si la verdadera media fuese casi igual a 10, sólo rechazaríamos correctamente la hipótesis falsa en el 5 % de los casos aproximadamente.

La función de potencia representada en la figura 5-8 tiene una forma que es típica de las contrastaciones simétricas, de dos extremos, relacionadas con la media de una población normal. Las contrastaciones de un extremo dan lugar a una curva completamente distinta que no presentaremos aquí. El lector que esté interesado en ello puede efectuar los cálculos y representarla gráficamente por sí mismo, o puede consultar casi cualquiera de los libros de texto corrientes sobre inferencia estadística. Unicamente queremos señalar que la función de potencia es otra de las formas de demostrar algunas de las debilidades de la contrastación estadística. La función de potencia ideal —que no hace falta decir que no existe— sería la que proporcionase un valor de 1 para todos los valores de μ (o cualquiera que fuese el parámetro relevante) especificados por la hipótesis alternativa, y un valor 0 para el valor de μ especificado por la hipótesis nula.

Calidad de la contrastación

¿Cuál es la mejor contrastación que podemos diseñar dados un tamaño muestral y una información previa determinada? Podemos responder a esta pregunta analizando las elecciones que debemos efectuar al establecer el criterio de contrastación. En general, las decisiones relevantes abarcan tres aspectos distintos:

1. Elección del nivel de significación.

2. Elección del tipo de región de aceptación.
3. Elección del estadístico de prueba.

Ya hemos visto que en el primer apartado podemos efectuar una elección racional *siempre y cuando* tengamos la posibilidad de valorar las pérdidas debidas a una conclusión incorrecta. De lo contrario —y esto es lo que ocurre generalmente en la investigación econométrica—, no disponemos de ninguna base firme para elegir el nivel de significación, como no sea la tradición.

La segunda área de elección hace referencia a la posición de la región de aceptación sobre el eje de las z o de las x . Supongamos que el problema incluye las hipótesis

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0$$

y que el nivel de significación elegido es el 5%. Anteriormente, en un problema de este tipo, elegimos una contrastación de dos extremos y colocamos la región de aceptación de forma simétrica en torno al punto cero del eje de las z , o, de forma equivalente, en torno al punto μ_0 en el eje de las $\bar{x}$. La región de aceptación cubría entonces el 95% central del área y dejaba un 2.5% del área en cada extremo para constituir la región crítica. Naturalmente, podríamos haber tomado el mismo nivel de significación del 5% —y en consecuencia, la misma probabilidad de error del tipo I—, pero colocando la región de aceptación en una posición distinta. Por ejemplo, podríamos haber situado la región de aceptación de forma que delimitase un 1% del área en el extremo inferior y un 4% en el extremo superior. El número de posibilidades es obviamente infinito. Anteriormente, se justificó la elección de una región de aceptación simétrica basándonos en gran medida en la intuición; en este momento, podemos efectuar una defensa más argumentada en favor de dicha elección. Todas las regiones de aceptación que conservan el mismo nivel de significación mantienen también, automáticamente, la misma probabilidad de error del tipo I y, por tanto, el argumento en favor de una región simétrica sólo puede discurrir en términos de las probabilidades del error del tipo II. Pero como el problema que nos ocupa tiene una hipótesis alternativa compuesta, los distintos valores de μ correspondientes a la hipótesis alternativa estarán asociados a distintas probabilidades de error del tipo II. Así, pues, tenemos que comparar las *funciones de potencia* de las contrastaciones basadas en regiones de aceptación con distinta delimitación, y no simplemente comparar probabilidades individuales. Supongamos que comparamos la función de potencia de una contrastación con una región de aceptación simétrica, con la de una contrastación cuya región de aceptación es asimétrica, como, por ejemplo, la que delimita un 1% del área en el extremo inferior y un 4% en el extremo superior. En este caso,

La contrastación simétrica resulta ser más potente para los valores de μ inferiores a μ_0 como puede comprobarse fácilmente efectuando los cálculos correspondientes. La función de potencia indica la capacidad de la contrastación para discriminar entre una hipótesis verdadera y una hipótesis falsa y, por lo tanto, la comparación muestra que la contrastación simétrica cumple este cometido mejor que la asimétrica (que tiene un extremo inferior más pequeño y un extremo inferior mayor) cuando $\mu < \mu_0$, pero cuando $\mu > \mu_0$ ocurre lo contrario. Si no existe ninguna razón para que estemos interesados en discriminar de forma más efectiva entre las dos hipótesis cuando μ está a un lado de μ_0 que cuando está en el otro lado, la contrastación simétrica es, evidentemente, más adecuada que la asimétrica. Si existe alguna razón, debe estar incluida en la información previa relativa a la contrastación y, en este caso, la región de aceptación se delimitará del modo que corresponde. Normalmente, (cuando H_A es $\mu \neq \mu_0$) la elección más razonable que se puede efectuar es la de una región de aceptación simétrica. Es interesante señalar que esta región de aceptación es la más corta de entre todas las basadas en el mismo estadístico de prueba y para un mismo nivel de significación. Mediante un razonamiento similar se puede establecer también que, para los problemas cuya hipótesis alternativa es del tipo $\mu > \mu_0$ o $\mu < \mu_0$, la contrastación más adecuada es, como ya se ha mencionado anteriormente, la correspondiente a un extremo.

Sólo queda por considerar el tercero de los campos de elección, que corresponde al estadístico de prueba. Para el problema consistente en contrastar una hipótesis relativa a la media de una población normal —el único tipo de problema que se estudia de forma específica en esta sección— se utiliza como estadístico de prueba la variable normal tipificada Z construida de la forma siguiente

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$$

en la que μ_0 es la media de la población postulada por la hipótesis nula y $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. La utilización de $\bar{X}$ se justificó basándonos en que se trata de una estimación insesgada y consistente de μ , y, por tanto, representa un resumen adecuado de la información que la muestra proporciona acerca de la media de la población. Para una contrastación de dos extremos con un nivel de significación del 5%, la región de aceptación para el estadístico Z es

$$-1.96 \leq z \leq 1.96;$$

y para $\bar{X}$,

$$\mu_0 - 1.96\sigma_{\bar{x}} \leq \bar{x} \leq \mu_0 + 1.96\sigma_{\bar{x}}.$$

Estudiemos ahora la razón de que las propiedades que resultan deseables en un estimador (insesgaredad, etc.) son también deseables cuando se trata de contrastar hipótesis. Consideremos un estadístico de prueba similar a Z pero

en el que, en lugar de utilizar X , tomamos un estimador no especificado de μ al que llamaremos $\hat{\mu}$. La variancia de este estimador, que tampoco está especificada, es $\sigma_{\hat{\mu}}$. La única restricción que imponemos a $\hat{\mu}$ es que se distribuya normalmente (al menos de forma aproximada) a fin de poder efectuar la transformación a la variable normal tipificada. En este caso, esta última vendrá dada por

$$Z^* = \frac{\hat{\mu} - E(\hat{\mu})}{\sigma_{\hat{\mu}}}$$

Expresada de esta forma, Z^* no incluye a μ_0 y, por lo tanto, no cumple la función básica que debe cumplir un estadístico de prueba, es decir, confrontar la hipótesis nula con la evidencia de la muestra. Sin embargo, podemos introducir fácilmente μ_0 en la fórmula, escribiendo

$$Z^* = \frac{\hat{\mu} - E(\hat{\mu})}{\sigma_{\hat{\mu}}} + \frac{\mu_0}{\sigma_{\hat{\mu}}} - \frac{\mu_0}{\sigma_{\hat{\mu}}} = \frac{(\hat{\mu} - \mu_0) - [E(\hat{\mu}) - \mu_0]}{\sigma_{\hat{\mu}}}$$

Si la hipótesis nula es válida y, por tanto, μ_0 es la media correcta, $[E(\hat{\mu}) - \mu_0]$ representa el sesgo y sólo será igual a cero si $\hat{\mu}$ es un estimador insesgado de μ . Para una contrastación con dos extremos al nivel de significación del 5%, la región de aceptación será

$$-1.96 \leq \frac{(\hat{\mu} - \mu_0) - [E(\hat{\mu}) - \mu_0]}{\sigma_{\hat{\mu}}} \leq 1.96.$$

La región de aceptación equivalente en el eje de $\hat{\mu}$ será

$$\mu_0 - 1.96\sigma_{\hat{\mu}} + [E(\hat{\mu}) - \mu_0] \leq \hat{\mu} \leq \mu_0 + 1.96\sigma_{\hat{\mu}} + [E(\hat{\mu}) - \mu_0].$$

La primera propiedad de $\hat{\mu}$ que vamos a considerar es la *insesgadez*. Supongamos que $\hat{\mu}$ es un estimador sesgado de μ . La consecuencia obvia de este hecho consistirá en que la región de aceptación anterior no será simétrica en torno a μ_0 . Por ejemplo, si $\mu_0 = 10$, $E(\hat{\mu}) = 5$, y $\sigma_{\hat{\mu}} = 3$, la región de aceptación será

$$10 - 1.96 \times 3 + (5 - 10) \leq \hat{\mu} \leq 10 + 1.96 \times 3 + (5 - 10),$$

es decir,

$$-0.88 \leq \hat{\mu} \leq 10.88,$$

que evidentemente no es simétrica en torno a 10. Si utilizamos una región de aceptación asimétrica como ésta, la función de potencia de la contrastación estará desplazada y su valor mínimo corresponderá al punto $E(\hat{\mu})$ en lugar de μ_0 . A menos que tengamos alguna razón especial para querer ser capaces de

discriminar entre hipótesis verdaderas y falsas de forma tan desigual, esta contrastación no podrá resultar adecuada. Naturalmente, si no sabemos cuál es el sesgo de $\hat{\mu}$, no podremos delimitar los límites entre la región de aceptación y la región crítica, como hicimos anteriormente, y, por tanto, no podremos efectuar la contrastación. Pero aún es peor el caso en que el estimador es sesgado pero no lo sabemos. En este caso, estaríamos utilizando una región de aceptación incorrecta para el nivel de significación elegido y las conclusiones serían, pues, erróneas. Todo ello se aplica por igual a las contrastaciones de un extremo y señala el interés de utilizar un estimador insesgado de μ al construir el estadístico de prueba.

Consideremos ahora la *consistencia*. Anteriormente se ha definido como consistente un estimador cuando su distribución tiende a concentrarse en torno al verdadero valor del parámetro al aumentar el tamaño de la muestra. Salvo en algunos casos especiales (en los estimadores que no tienen media o variancia finitas, o ninguna de las dos) esto significa que al aumentar el tamaño de la muestra disminuirán el sesgo (si lo hay) y la variancia de la distribución, y que, en el límite, ambos tenderán a cero. Por lo tanto, cuanto mayor es el tamaño muestral, más estrecha es la región de aceptación en la escala de las $\hat{\mu}$ y más

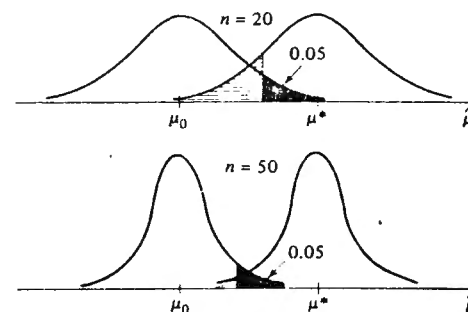


Figura 5-9

potente es la contrastación. En la figura 5-8 se ilustra este caso. El valor de la media postulado por la hipótesis nula es μ_0 , y μ^* es uno de los valores de la media postulados por la hipótesis alternativa. El área en negro representa la probabilidad del error del tipo I, y el área rayada la probabilidad del error del tipo II. El gráfico indica cómo disminuye la probabilidad del error del tipo II o, de forma equivalente, aumenta la potencia de la contrastación al aumentar el tamaño de la muestra, dado el nivel de significación. Si al aumentar el tamaño de la muestra la variancia de la distribución no disminuyese —es decir, si el estimador no fuese consistente— este aumento de la muestra no reduciría la

probabilidad del error del tipo II. Esto no resulta deseable, ya que significa que el coste adicional, debido al aumento en el tamaño de la muestra, no produce mayor información respecto a la media de la población.

La última propiedad a la que haremos referencia en este contexto es la *eficiencia*. Si tenemos dos estimadores de la media de la población, y los dos son insesgados y se distribuyen normalmente pero tienen variancias distintas, el estimador con la variancia mayor no es eficiente. La región de aceptación de este estimador —medida en el eje de las $\hat{\mu}$ — será mayor que la del otro y la potencia de la contrastación correspondiente será menor. En la figura 5-10 se presenta un ejemplo de este caso. En el gráfico, $\hat{\mu}_I$ es más eficiente que $\hat{\mu}_{II}$. Si comparásemos las funciones de potencia de las contrastaciones correspondientes a estos dos estimadores, veríamos que los valores de la función de μ_I son siempre mayores que los de la función μ_{II} . Por tanto, la eficiencia es claramente un rasgo muy deseable al establecer un estadístico.

El análisis del diseño óptimo de una contrastación que acabamos de efectuar indica que, en la mayor parte de los problemas econométricos, el único procedimiento para obtener la mejor contrastación consiste en utilizar el estimador óptimo al construir el estadístico de prueba adecuado. Como normalmente no se dispone de la información que permitiría efectuar una elección racional del nivel de significación o de la localización de la región de aceptación, dada la información previa y el tamaño muestral, la única posibilidad de avance es la relativa al estimador óptimo. Como veremos, en algunos casos concretos la investigación ya se ha completado, es decir, ya se ha obtenido el estimador óptimo¹. Desgraciadamente esto sólo ocurre en problemas simples como, por

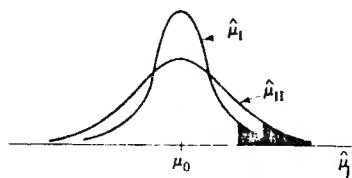


Figura 5-10

ejemplo, la estimación de la media de una población normal, pero en muchos problemas econométricos la búsqueda sigue en pie. Esta es, al menos en parte, la razón por la que la investigación econométrica pura se ha concentrado fuertemente en los problemas de estimación.

1. Sin embargo, también existe el problema de si estamos utilizando *toda* la información previa de la forma más efectiva posible. Normalmente, parte de los conocimientos previos (por ejemplo, los resultados numéricos de las estimaciones anteriores) no se utilizan en absoluto, a menos que se adopte el enfoque bayesiano de la inferencia estadística.

Comentarios finales

Con lo dicho hasta aquí, llegamos al final de la sección dedicada a la obtención y valoración de las contrastaciones estadísticas. Como el lector recordará, se inició la discusión dividiendo las hipótesis en hipótesis sostenidas por un lado e hipótesis contrastables por otro. Esta división presenta un problema que no se ha planteado y que es el relativo a qué hipótesis deberían tomarse como hipótesis sostenidas, es decir fuera de discusión. Parece que sería más razonable contrastar todos los supuestos utilizados en lugar de sólo unos cuantos. Desde luego, esto sería más seguro que confiar en nuestras creencias previas que son las que constituyen la hipótesis sostenida, pero la dificultad de ello radica en el hecho de que nuestro conocimiento *real* de la población es muy escaso y, por lo tanto, si no efectuásemos algunos supuestos, la única fuente de información sería la muestra. A consecuencia de ello, la posibilidad de cometer errores sería mayor y la potencia de la prueba disminuiría. Tomemos por ejemplo el problema de contrastar una hipótesis acerca de la media de una población normal cuya variancia es conocida, es decir, el problema considerado a lo largo de esta sección. Una parte de la hipótesis sostenida es el supuesto de que conocemos la variancia de la población. Si no efectuamos dicho supuesto no conoceremos la variancia de la distribución de la media de la muestra y no podremos construir el estadístico de prueba de la ecuación (5.1). Naturalmente, podemos estimar la variancia de la población a partir de los datos de la muestra, pero al usar la estimación en lugar del valor real de la variancia aumentamos el grado de incertidumbre respecto a la población y, como es de esperar, obtenemos una región de aceptación más amplia. Aún podemos elegir el nivel de significación que deseemos, pero, con una región de aceptación más amplia, la potencia de la contrastación será menor (en la sección 5.2 se explica con detalle esta contrastación). Por tanto, al sustituir un supuesto por una estimación a partir de la muestra, perdemos cierta información y debemos pagar por ello a través de una contrastación menos potente. Claro está que, si no tenemos ninguna idea respecto al valor de la variancia de la población, no hay elección posible. Por otra parte, si confiamos en los supuestos contenidos en la hipótesis sostenida, obtenemos unos resultados estrictamente condicionados a dichos supuestos, que no tienen ninguna validez al margen de ellos. Este hecho elemental parece olvidarse con frecuencia en la investigación econométrica aplicada.

5.2 Distribución de algunos estadísticos de prueba

En la sección 5.1 hemos ilustrado las ideas subyacentes en la contrastación de hipótesis estadísticas considerando el problema de contrastar una hipótesis respecto a la media de una población normal con una variancia conocida σ^2 . Como estadístico de prueba se propuso

$$Z = \frac{(\bar{X} - \mu_0)}{\sigma_{\bar{X}}},$$

en la que μ_0 es la media postulada por la hipótesis nula, y σ_x es la desviación standard de la distribución de $\bar{X}$, y es igual a $\sigma/\sqrt{n}$. Si $\bar{X}$ es el mejor estimador que podemos hallar de la media de la población, Z es el mejor estadístico de prueba que podemos construir. Como veremos en el capítulo 6, $\bar{X}$ es (salvo en circunstancias especiales) el estimador óptimo de la media de la población de modo que, en general, no es posible mejorar Z .

Contrastación de la diferencia entre dos medias

Cuando el problema comprende *dos muestras independientes* procedentes de dos poblaciones normales con variancias conocidas, se puede utilizar un procedimiento de contrastación similar al que acabamos de describir. Se puede plantear el problema de si existe alguna diferencia entre las medias de las dos poblaciones. Supongamos que la media de la primera es μ_1 y la media de la segunda es μ_2 . En este caso, la hipótesis nula es

$$H_0: \mu_1 = \mu_2,$$

y la hipótesis alternativa es

$$H_A: \mu_1 \neq \mu_2 \quad (\text{o } \mu_1 > \mu_2).$$

Como resumen muestral de la evidencia disponible acerca de las medias de las poblaciones podemos utilizar las medias muestrales respectivas, $\bar{X}_1$ y $\bar{X}_2$. Si la hipótesis nula es correcta ($\mu_1 - \mu_2 = 0$) y, por tanto, el valor de $(\bar{X}_1 - \bar{X}_2)$ será en muy pocas ocasiones muy distinto de cero. En consecuencia, los valores muestrales de $(\bar{X}_1 - \bar{X}_2)$ que sean muy distintos de cero se podrán considerar como evidencia en contra de la hipótesis nula. Si obtenemos la distribución de $(\bar{X}_1 - \bar{X}_2)$, podremos especificar las regiones crítica y de aceptación adecuadas. Como tanto $\bar{X}_1$ como $\bar{X}_2$ proceden de poblaciones normales, su diferencia también se distribuirá normalmente. La media de esta distribución es

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2,$$

y, si la hipótesis nula es cierta, esta media es igual a cero. Por tanto, lo único que debemos obtener es la variancia. Como X_1 y X_2 son las medias de dos muestras independientes, su covariancia es cero y, por tanto,

$$\text{Var}(\bar{X}_1 - \bar{X}_2) = \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

en la que σ_1^2 y σ_2^2 son las variancias de las dos poblaciones normales de donde proceden las muestras, y n_1 y n_2 son los tamaños muestrales respectivos. Cuan-

do las dos medias de las poblaciones son iguales, como afirma la hipótesis nula,

$$(\bar{X}_1 - \bar{X}_2) \sim N\left[0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right].$$

En este caso, la variable normal tipificada correspondiente es

$$(5.2) \quad Z_{(\bar{X}_1 - \bar{X}_2)} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{(\bar{X}_1 - \bar{X}_2)}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}.$$

Este es el estadístico de prueba apropiado y con la ayuda de las tablas de probabilidad normal podemos definir la región crítica y de aceptación correspondientes al mismo.

Estimación de σ^2

En las dos contrastaciones consideradas anteriormente hemos supuesto que se conocía la variancia de la población. Normalmente no sucede así, y esta variancia debe estimarse a partir de la muestra, pero al utilizar una *estimación* de la variancia de la población en lugar de su valor real, es preciso modificar las contrastaciones relativas a la media de la población. Antes de efectuar las modificaciones necesarias, estudiaremos el problema de estimar la variancia de la población de $X \sim N(\mu, \sigma^2)$. Como posible candidato, podemos considerar la variancia de la muestra, a la que llamaremos $\hat{\sigma}^2$ y que se define de la forma siguiente:

$$(5.3) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Naturalmente, cada muestra distinta daría como resultado un valor distinto de $\hat{\sigma}^2$, y nos interesa saber si este estimador posee algunas de las propiedades que resultan deseables y cuál es su distribución.

Podemos empezar examinando el sesgo y para ello tomaremos la esperanza matemática de $\hat{\sigma}^2$ y comprobaremos si es igual o no a σ^2 . Si lo es, diremos que $\hat{\sigma}^2$ es insesgada. Sabemos que

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n} \sum_{i=1}^n E(X_i - \bar{X})^2.$$

El valor de $E(X_i - \bar{X})^2$ no lo conocemos, pero sabemos que

$$E(X_i - \mu)^2 = \text{Var}(X_i) = \sigma^2,$$

ya que X_i tiene exactamente la misma distribución que X y, por lo tanto, la misma variancia. Además, también sabemos que

$$E(\bar{X} - \mu)^2 = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

según el teorema 13 (sección 4.3). De ahí que podamos escribir de nuevo la expresión de $E(\hat{\sigma}^2)$ sumando y restando simultáneamente μ :

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n} \sum_i E[(X_i - \mu) - (\bar{X} - \mu)]^2 \\ &= \frac{1}{n} \sum_i [E(X_i - \mu)^2 + E(\bar{X} - \mu)^2 - 2E(X_i - \mu)(\bar{X} - \mu)] \\ &= \frac{1}{n} \sum_i E(X_i - \mu)^2 + \frac{1}{n} \sum_i E(\bar{X} - \mu)^2 - \frac{2}{n} \sum_i E(X_i - \mu)(\bar{X} - \mu) \\ &= \frac{1}{n} \sum_i \sigma^2 + \frac{1}{n} \sum_i \frac{\sigma^2}{n} - 2E(\bar{X} - \mu) \frac{1}{n} \sum_i (X_i - \mu) \\ &= \sigma^2 + \frac{\sigma^2}{n} - 2E(\bar{X} - \mu)^2 = \sigma^2 + \frac{\sigma^2}{n} - 2 \frac{\sigma^2}{n}. \end{aligned}$$

Es decir,

$$(5.4) \quad E(\hat{\sigma}^2) = \left[\frac{n-1}{n} \right] \sigma^2,$$

que no es igual a σ^2 . Ello significa que $\hat{\sigma}^2$ es un estimador *sesgado* de σ^2 .

Este resultado, pese a ser negativo, resulta de utilidad, ya que sugiere un fácil procedimiento para obtener un estimador insesgado de $\hat{\sigma}^2$. Si multiplicamos ambos lados de la ecuación (5.4) por $n/(n-1)$ obtenemos

$$\left[\frac{n}{n-1} \right] E(\hat{\sigma}^2) = \sigma^2,$$

que puede escribirse en la forma

$$E \left[\frac{n}{n-1} \right] \hat{\sigma}^2 = \sigma^2,$$

o también

$$E \left[\frac{n}{n-1} \right] \frac{1}{n} \sum_i (X_i - \bar{X})^2 = \sigma^2,$$

es decir,

$$(5.5) \quad E \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 = \sigma^2,$$

o sea que un estimador de σ^2 , al que llamaremos s^2 y que se define como

$$(5.6) \quad s^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2,$$

es un estimador *insesgado* de σ^2 .

Debido a su insesgadez, utilizaremos s^2 como estimador preferido de σ^2 . Su distribución, que no obtendremos ni presentaremos aquí, es un caso especial de la llamada distribución gamma². La forma exacta de la distribución de s^2 depende de dos parámetros, la variancia de la población σ^2 y el tamaño muestral n . La distribución es siempre asimétrica hacia la derecha para tamaños muestrales pequeños y se va haciendo cada vez más simétrica al aumentar el tamaño muestral. Como se ha indicado, la media de la distribución es igual a σ^2 y su variancia viene dada³ por

$$(5.7) \quad \text{Var}(s^2) = \frac{2\sigma^4}{n-1}.$$

Como s^2 es insesgado y su variancia tiende a cero cuando $n \rightarrow \infty$, s^2 es un estimador *consistente* de σ^2 . Otro de los rasgos de s^2 que es interesante señalar es el hecho de que s^2 es independiente de $\bar{X}$, en el sentido de que la distribución conjunta de s^2 y $\bar{X}$ es igual al producto de sus distribuciones marginales respectivas. Es decir,

$$f(s^2, \bar{x}) = f(s^2)f(\bar{x}).$$

La distribución X^2

Para obtener la probabilidad de que s^2 se halle en un intervalo determinado, tendríamos que conocer las áreas situadas debajo de la curva gamma correspondiente. Esta curva sería distinta para cada combinación de los valores de σ^2 y n , y calcular el área apropiada resultaría muy complicado. Ya nos hemos encontrado con un problema similar al estudiar la distribución normal, que también es distinta para cada combinación de los valores de los parámetros

2. Véase, por ejemplo, John E. Freund, *Mathematical Statistics*. Englewood Cliffs, N. J., Prentice-Hall, 1962, pp. 127-28. Para el caso especial de s^2 tenemos, siguiendo la notación de Freund, $\alpha = (n-1)/2$ y $\beta = 2\sigma^2/(n-1)$.

3. *Ibid.*, p. 147. Obsérvese que la media de s^2 es igual a σ^2 sea cual sea la distribución originaria de X , pero la fórmula de la variancia (5.8) se basa en el supuesto de que X se distribuye normalmente.

—en este caso, μ y σ^2 — y cuya fórmula matemática es, asimismo, muy complicada. En el caso de la distribución normal resolvimos el problema transformando la variable normal que nos interesaba en otra con media 0 y variancia 1 (la «variable normal tipificada») cuyas probabilidades se hallan tabuladas. En el caso de s^2 , la solución es similar. La transformación en este caso se hace pasando a una variable llamada *ji-cuadrado* (χ^2) que se define como

$$(5.8) \quad \chi^2 = \frac{(n-1)s^2}{\sigma^2},$$

cuyas probabilidades están calculadas. Sin embargo, contrariamente a lo que ocurre con la distribución normal tipificada, la forma de la distribución de χ^2 varía al variar el tamaño de la muestra. Cuando la muestra es pequeña, la distribución es simétrica hacia la derecha, pero, a medida que el tamaño de la muestra aumenta, se va haciendo cada vez más simétrica. Naturalmente, χ^2 no tiene ningún valor negativo. Debido a esta dependencia de la distribución ji-cuadrado respecto al tamaño de la muestra, normalmente se utiliza un subíndice para identificarla. Por ejemplo, la ecuación (5.8) se escribiría normalmente en la forma

$$(5.8a) \quad \chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}.$$

Una forma equivalente de escribir esta ecuación es

$$(5.8b) \quad \chi_{n-1}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}.$$

El subíndice de χ^2 , que es igual al tamaño de la muestra menos uno, se llama el número de *grados de libertad* y frecuentemente se representa por v en lugar de $(n-1)$. La expresión «grados de libertad» hace referencia al número de cuadrados independientes en el numerador del estadístico ji-cuadrado, es decir, en

$$\sum_{i=1}^n (X_i - \bar{X})^2.$$

El número total de cuadrados en esta expresión es n , pero sólo hay $(n-1)$ cuadrados independientes porque una vez calculados los $n-1$ primeros, el valor del último queda determinado automáticamente. La razón de ello es la presencia de $\bar{X}$ en la expresión y el hecho de que una de las características bien conocidas de $\bar{X}$ es que

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

representa una restricción que debe cumplirse. Por ejemplo, supongamos que tenemos tres cuadrados y que los valores de los dos primeros son

$$(x_1 - \bar{x})^2 = 2^2,$$

$$(x_2 - \bar{x})^2 = 4^2.$$

Si el tercer cuadrado fuese independiente de los otros dos, podría tener cualquier valor, por ejemplo, 5^2 . Pero esto no puede suceder, porque en este caso tendríamos

$$\sum_{i=1}^3 (x_i - \bar{x}) = 2 + 4 + 5,$$

cuya suma es 11 y no cero como tendría que ser. En realidad, si los dos primeros cuadrados son 2^2 y 4^2 , el tercero tiene que ser necesariamente $(-6)^2 = 36$, porque sólo en este caso obtenemos $2 + 4 - 6 = 0$. Esta explicación de la obtención del número de grados de libertad resulta bastante simple en este contexto, pero en otras ocasiones el tema es mucho más complicado. Por este motivo, la mayor parte de los libros de texto de estadística elemental no profundizan demasiado en la cuestión y sugieren que el lector puede pensar en los grados de libertad simplemente como un nombre que se da a un parámetro.

A partir del conocimiento de la distribución de probabilidad de χ^2 podemos obtener las probabilidades de la distribución muestral de s^2 . Podemos verlo de la forma siguiente. Si a y b son dos constantes cualquiera tales que $0 \leq a \leq b$, tenemos

$$\begin{aligned} P(a \leq \chi_{n-1}^2 \leq b) &= P\left[a \leq \frac{(n-1)s^2}{\sigma^2} \leq b\right] \\ &= P\left[a\left(\frac{\sigma^2}{n-1}\right) \leq s^2 \leq b\left(\frac{\sigma^2}{n-1}\right)\right], \end{aligned}$$

o sea, que una afirmación de tipo probabilístico respecto a χ^2 se puede transformar fácilmente en otra equivalente respecto a s^2 . Por último, podemos señalar que la media y la variancia de la distribución de ji-cuadrado son

$$E(\chi_{n-1}^2) = n - 1$$

y

$$\text{Var}(\chi_{n-1}^2) = 2(n-1),$$

mientras que la moda es $(n-3)$. En la figura 5-11 se representa gráficamente la distribución de ji-cuadrado para distintos grados de libertad, v .

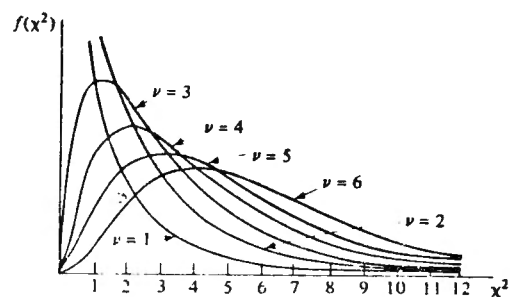


Figura 5-11

Podemos ver con mayor claridad la naturaleza de la distribución ji-cuadrado comparándola con la distribución normal. Esta última se utiliza en estadística por dos razones fundamentales. En primer lugar, se trata de una distribución importante en sí misma, ya que proporciona una descripción de gran cantidad de poblaciones. No es muy difícil encontrar variables relativas a fenómenos sociales o naturales que se distribuyen —o podemos suponer que se distribuyen— normalmente. Ejemplos de ello son la inteligencia, la estatura de los individuos, los errores de medición o de comportamiento, el número de aciertos en muchas observaciones y muchos otros. La segunda razón para usar la distribución normal se debe al hecho de que esta distribución describe algunas distribuciones muestrales muy importantes y, en particular, la de la media de la muestra. Por lo tanto, la distribución normal puede utilizarse para describir una población originaria o una distribución muestral. Por el contrario, la distribución ji-cuadrado sólo sirve para describir algunas distribuciones muestrales, una de las cuales es la del estadístico de la ecuación (5.8)⁴. No hay ninguna población originaria importante cuya distribución corresponda a la ji-cuadrado.

La distribución χ^2

Una vez solucionado el problema de la estimación de la variancia de la población, podemos volver al problema original que consistía en contrastar una

4. Además de describir la distribución de la expresión (5.8), la distribución ji-cuadrado también corresponde a un estadístico que se utiliza para contrastar la independencia en las tablas de contingencia y la bondad del ajuste de las distribuciones de frecuencia. La contrastación de la independencia en las tablas de contingencia la estudiaremos en el capítulo 11, en relación con las variables binarias; la contrastación de la bondad del ajuste se estudia al final de esta sección.

hipótesis respecto a la media de la población de una variable normalmente distribuida cuya variancia no se conocía. Recapitulando, sabemos que si

$$X \sim N(\mu, \sigma^2),$$

se cumple que

$$\bar{X} \sim N\left[\mu, \frac{\sigma^2}{n}\right].$$

Además, si la hipótesis nula es de la forma

$$H_0: \mu = \mu_0,$$

el estadístico de prueba correspondiente es

$$\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0, 1),$$

suponiendo que conocemos σ . Cuando, como en este caso, no conocemos el valor de σ , lo podemos sustituir por una estimación obtenida mediante la fórmula

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

que procede directamente de la ecuación (5.6). En este caso, el estadístico de prueba sería

$$\frac{(\bar{X} - \mu_0)\sqrt{n}}{s}.$$

El problema consiste en obtener su distribución. Obsérvese que si dividimos el numerador y el denominador por σ y reordenamos el resultado, obtenemos

$$\frac{(\bar{X} - \mu_0)\sqrt{n}/\sigma}{\sqrt{(n-1)s^2/((n-1)\sigma^2)}}.$$

El numerador es la variable normal tipificada Z y en el denominador vemos que $(n-1)s^2/\sigma^2 = \chi_{n-1}^2$ como indica la ecuación (5.8a). O sea, que podemos escribir

$$\frac{Z}{\sqrt{\chi_{n-1}^2/(n-1)}}.$$

Esta variable tiene una distribución que se conoce con el nombre de *distribución t*. Debido a la presencia de la variable ji-cuadrado en el denominador, la forma de esta distribución depende del tamaño de la muestra y, por tanto, normalmente se utiliza un subíndice para identificarla. En este caso escribiremos

$$(5.9) \quad \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t_{n-1}.$$

En esta expresión, $(n - 1)$ es el número de grados de libertad. Naturalmente, estos grados de libertad proceden de la variable ji-cuadrado que aparece en el denominador. Pero contrariamente a lo que ocurre con la distribución ji-cuadrado, la distribución t es siempre simétrica; su media es cero y su variancia es $(n - 1)/(n - 3)$ y cuando n es grande, esta última se aproxima a la unidad. Al aumentar el tamaño de la muestra, la distribución t se aproxima a la distribución normal tipificada. En el Apéndice D se reproduce una tabla que presenta las probabilidades correspondientes a distintos tamaños muestrales.

Para contrastar la hipótesis de que la media de la población tiene un valor determinado, tenemos que especificar los límites entre la región crítica y la región de aceptación para el estadístico de prueba de la ecuación (5.9). Estos dependerán de la forma que tome la hipótesis alternativa, del nivel de significación que se desee y del número de grados de libertad. Supongamos que queremos contrastar

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0.$$

El nivel de significación deseado es un número α . En este caso, la región de aceptación se define como

$$-t_{n-1, \alpha/2} \leq \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \leq t_{n-1, \alpha/2}.$$

En esta expresión, $t_{n-1, \alpha/2}$ representa el valor del estadístico t con $n - 1$ grados de libertad, que delimita un porcentaje $\alpha/2$ del área de la distribución en cada extremo. Este valor se puede obtener a partir de la tabla de probabilidades de la distribución t . Dicha tabla se presenta de forma que cada fila corresponde a un número distinto de grados de libertad, y cada columna a un área distinta en el extremo de la distribución. Por ejemplo, en la fila correspondiente a «14» y la columna «0.05», hallamos el valor 1.761. Esto significa que la probabilidad de que un valor del estadístico t (obtenido a partir de una muestra de tamaño 15) sea superior a 1.761 (o menor que -1.761) es 0.05. Es decir, los valores del estadístico de prueba se presentan en la parte central de la tabla y las probabilidades en el margen. Esta ordenación es distinta de la de las

tablas de probabilidad normal, ya que en ellas el orden es exactamente el contrario. La tabla de los valores de la t presenta en realidad muchas distribuciones distintas (tantas como filas) y, por tanto, es mucho menos detallada que la de la probabilidad normal.

Consideremos ahora la contrastación con dos extremos que hemos mencionado anteriormente. Si el nivel de significación que deseamos es el 5 %, el valor $\alpha/2$ es 0.025, y la columna que nos interesa es la correspondiente a este número. La fila que debemos tomar depende del tamaño de la muestra. Algunos de los valores que podemos obtener son los siguientes

Tamaño de la muestra (n)	Grados de libertad (v)	Valor de $t_{n-1, 0.025}$
5	4	2.776
10	9	2.262
20	19	2.093
30	29	2.045
∞		1.960

Obsérvese que al aumentar n , los valores de t son cada vez más pequeños, y se aproximan a 1.960, que es el valor que correspondería a una variable normal tipificada con el mismo nivel de significación. Si comparamos con las contrastaciones en las que se conoce la variancia de la población, y por tanto no es preciso estimarla, la región de aceptación basada en la distribución t es siempre más amplia. Ello puede verse si nos fijamos en que en cada columna de la tabla de la t todos los valores son mayores que el valor que aparece en la última fila, que corresponde a la variable normal tipificada. Como consecuencia de esta mayor región de aceptación, la probabilidad del error del tipo II es mayor, y por tanto, la contrastación es menos potente.

La contrastación que hemos analizado corresponde al caso de dos extremos, pero el procedimiento a utilizar para las contrastaciones de un extremo es muy similar. En lugar de efectuar una descripción de tipo general, lo ilustraremos con un ejemplo numérico.

EJEMPLO. Un trabajador que efectúa una determinada operación en la fabricación de un producto emplea 7 minutos, por término medio, para realizarla. Un analista sugiere una forma distinta de efectuar la operación y decide tomar una muestra para descubrir si se produce algún ahorro de tiempo. Las hipótesis nula y alternativa son las siguientes

$$H_0: \mu = 7 \text{ min. (o 420 seg.)}$$

$$H_A: \mu < 7 \text{ min.}$$

El nivel de significación será el 5 % tradicional y el tamaño de la muestra es 16. Suponiendo que la población originaria es normal, el estadístico de prueba apropiado es la t de la ecuación (5.9) con 15 grados de libertad. El punto límite de la distribución t se puede buscar en la tabla, en la fila «15» y la columna «0,05». El criterio de contrastación que obtenemos es

$$\begin{aligned} &\text{no rechazar la hipótesis si } \frac{(\bar{X} - 420)\sqrt{16}}{s} \geq -1.753; \\ &\text{rechazarla en caso contrario.} \end{aligned}$$

Obsérvese que $\bar{x}$ y s se miden en segundos. Las observaciones obtenidas son

6 min. 26 seg.	6 min. 0 seg.	7 min. 30 seg.	6 min. 4 seg.
6 min. 38 seg.	7 min. 0 seg.	7 min. 8 seg.	6 min. 42 seg.
6 min. 48 seg.	7 min. 12 seg.	7 min. 20 seg.	7 min. 6 seg.
6 min. 58 seg.	6 min. 46 seg.	6 min. 22 seg.	6 min. 48 seg.

Los resultados obtenidos son los siguientes:

$$\bar{x} = 6,528/16 = 408 \text{ (o sea, 6 min. 48 seg.)}$$

$$s = \sqrt{9872/15} = 25.65.$$

Por tanto, el valor del estadístico de prueba es

$$\frac{(408 - 420)\sqrt{16}}{25.65} = -1.871.$$

Este punto se halla en el exterior de la región de aceptación y, por tanto, la hipótesis «7 minutos» debe rechazarse al nivel de significación del 5 %.

Contrastaciones respecto a la media de una población no normal

El análisis anterior se ocupa de las contrastaciones relativas a las poblaciones de una variable distribuida normalmente cuya variancia se desconoce. Para resolver el problema correspondiente a la diferencia entre las medias de dos poblaciones normales, a partir de dos muestras independientes, podemos establecer un procedimiento similar, pero no lo vamos a estudiar aquí porque se puede hallar en cualquier texto introductorio de estadística matemática. Lo que sí vamos a considerar es la correspondiente al valor de la media de una variable que *no se distribuye normalmente* (pero tiene una variancia finita). En este caso, si la muestra es grande podemos utilizar el teorema central del límite (teorema 15) que señala que, sea cual sea la distribución de X en las muestras grandes, la distribución de $\bar{X}$ será aproximadamente normal. Teniendo

en cuenta esto y el hecho de que cuando n es grande la distribución t es aproximadamente normal, podemos ver que

$$\frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \quad (n \geq 30)$$

tiene una distribución que se aproxima a la distribución normal tipificada y, por tanto, en este caso podemos utilizar la contrastación que se ha estudiado en la sección 5.1. Si no conocemos la distribución originaria y el tamaño de la muestra es pequeño, podemos acudir a las contrastaciones llamadas *sin distribución* o *no paramétricas* que se estudian en muchos libros de texto.

Contrastaciones de σ^2

Consideremos ahora algunas contrastaciones de la variancia de la población. Supongamos que queremos contrastar la hipótesis nula de que la variancia de una población normal tiene un valor determinado contra la hipótesis alternativa de que su valor es otro. Es decir,

$$H_0: \sigma^2 = \sigma_0^2,$$

$$H_A: \sigma^2 \neq \sigma_0^2.$$

Podemos contrastar esta hipótesis tomando el estimador muestral de σ^2 definido en la ecuación (5.6), es decir,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

y estableciendo las regiones crítica y de aceptación para la distribución de s^2 . Como estadístico de prueba utilizamos una transformación de s^2 que viene dada por la expresión

$$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

en la que $(n-1)$ es el número de *grados de libertad*. Las propiedades de esta distribución ya las hemos estudiado antes. Si elegimos un nivel de significación α , en una contrastación de dos extremos la región de aceptación es

$$\chi_{n-1, 1-\alpha/2}^2 \leq \frac{(n-1)s^2}{\sigma_0^2} \leq \chi_{n-1, \alpha/2}^2,$$

en la que los subíndices $1-\alpha/2$ y $\alpha/2$ se refieren al área situada a la *derecha* del correspondiente punto límite de χ^2 . La tabla de las probabilidades de la distribución ji-cuadrado tienen la misma forma que la de la t . Las filas corres-

ponden a los grados de libertad, las columnas a las probabilidades y las cifras en el centro de la tabla son los valores correspondientes de χ^2 . Por ejemplo, el valor que aparece en la fila «4» y la columna «0,975» es 0,484. La fila indica que se trata de la distribución ji-cuadrado para las muestras de tamaño 5, y el valor 0,484 indica que la probabilidad de que la variable ji-cuadrado tome un valor superior al mismo es 0,975. En otras palabras, 0,484 es el límite inferior de un intervalo que se extiende hasta $+\infty$, y 0,975 es la probabilidad de que un valor de χ^2_4 se halle en este intervalo. A continuación, indicamos algunos de los valores de χ^2 para un nivel de significación de 5% en una contrastación de dos extremos.

Tamaño de la muestra (n)	Grados de libertad (v)	Valores de	
		$\chi^2_{n-1, 0.975}$	$\chi^2_{n-1, 0.025}$
5	4	0,484	11,143
10	9	2,700	19,023
20	19	8,907	32,852
30	29	16,047	45,722
1001	1000	914	1090

Cuando n es grande, podemos obtener las probabilidades de la ji-cuadrado a partir del hecho de que $[\sqrt{2\chi^2} - \sqrt{2(n-1)}]$ tiene una distribución que se aproxima a la distribución normal tipificada. Por ejemplo, cuando $(n-1) = 1000$ vemos que

$$\sqrt{2 \times 914} - \sqrt{2 \times 1000} = -1.966$$

y
$$\sqrt{2 \times 1090} - \sqrt{2 \times 1000} = +1.969$$

que son casi los valores $-1,960$ y $+1,960$ correspondientes a una variable normal tipificada. Aunque este análisis se ha aplicado a una contrastación de dos extremos, puede adaptarse fácilmente a las contrastaciones de un extremo en las que $H_A: \sigma^2 > \sigma_0^2$ o $\sigma^2 < \sigma_0^2$, y dejamos esta tarea para el lector.

EJEMPLO. Tomemos nuevamente el ejemplo anterior relativo al tiempo empleado por un trabajador al efectuar una tarea. Supongamos que se afirma que la variancia de esta variable (que se supone se distribuye normalmente) es 30^2 segundos. El analista afirma que el nuevo método de efectuar la operación que él propone también alterará la variancia del tiempo utilizado. En este caso, la hipótesis nula y la alternativa serán

$$H_0: \sigma^2 = 900,$$

$$H_A: \sigma^2 \neq 900.$$

Vamos a contrastar la hipótesis nula a un nivel de significación del 5%, y utilizaremos las 16 observaciones obtenidas en el ejemplo anterior. La región de aceptación se define como

$$6.262 \leq \frac{(n-1)s^2}{900} \leq 27.488,$$

en la que los valores límites se han obtenido en la tabla de la distribución ji-cuadrado. El valor del estadístico de prueba, calculado a partir de las observaciones es

$$\frac{9872}{900} = 10.97,$$

y, por tanto, se halla en el interior de la región de aceptación. En este caso, la muestra no proporciona, pues, evidencia en contra de la hipótesis nula.

La distribución F

Otra contrastación que se utiliza muy a menudo, relativa a la variancia de la población, es la correspondiente a la hipótesis de que dos poblaciones normales tienen la misma variancia. La hipótesis alternativa más corriente es la de que una de las variancias es mayor que la otra⁵. Es decir,

$$H_0: \sigma_1^2 = \sigma_2^2,$$

$$H_A: \sigma_1^2 > \sigma_2^2,$$

en la que tomamos como *primera* población a la que, según H_A , tiene la variancia mayor. Se puede contrastar la hipótesis nula obteniendo una muestra de cada una de las dos poblaciones y calculando las estimaciones s_1^2 y s_2^2 de las variancias respectivas. Se supone que las muestras se obtienen independientemente y que su tamaño es n_1 y n_2 , respectivamente. Como estadístico de prueba podemos tomar el cociente s_1^2/s_2^2 . Si la hipótesis nula es cierta, este cociente sólo será distinto de la unidad porque las estimaciones muestrales no son exactamente iguales a los respectivos parámetros. En cualquier caso, es de esperar que, a medida que las dos muestras aumentan de tamaño, el cociente se acercará a la unidad a menos que la hipótesis nula sea falsa.

Para efectuar la contrastación tenemos que establecer el punto límite entre la región de aceptación y la región crítica, y para ello tenemos que conocer la distribución muestral de s_1^2/s_2^2 . Dividamos el numerador por σ_1^2 y el denomina-

⁵ Naturalmente, también podríamos considerar una hipótesis alternativa con dos extremos, del tipo $\sigma_1^2 \neq \sigma_2^2$, pero es mucho menos corriente y no la vamos a estudiar.

dor por σ_2^2 ; si la hipótesis nula es cierta, el cociente no variará. Por tanto, podemos escribir

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2},$$

que equivale a

$$\frac{(n_1 - 1)s_1^2/(n_1 - 1)\sigma_1^2}{(n_2 - 1)s_2^2/(n_2 - 1)\sigma_2^2} = \frac{\chi_{n_1-1}^2/(n_1 - 1)}{\chi_{n_2-1}^2/(n_2 - 1)}$$

según la ecuación (5.8a). Como las dos muestras son independientes entre sí, el numerador y el denominador de esta expresión también son independientes. Mencionamos este punto porque la distribución de un cociente de dos variables ji-cuadrado independientes, cada una de las cuales está dividida por su correspondiente número de grados de libertad, es la conocida *distribución F*. Esta distribución es asimétrica y depende de dos parámetros, el número de grados de libertad del numerador y el del denominador. Estos dos números se indican generalmente en forma de subíndices de *F* para asegurar una identificación adecuada. Por tanto, podemos escribir

$$(5.10) \quad \frac{s_1^2}{s_2^2} \sim F_{n_1-1, n_2-1}.$$

Los valores de la distribución *F* se pueden obtener a partir de las tablas correspondientes. Normalmente, se presentan dos tablas, una para el nivel de significación del 5 % y la otra para el de 1 %. Cada tabla indica el valor límite de *F* para la contrastación de un extremo. Las filas de cada tabla corresponden al número de grados de libertad del denominador, y las columnas a los grados de libertad del numerador. Por ejemplo, si buscamos en la tabla del nivel de significación del 5 % en la fila «10» y la columna «15» obtenemos el valor 2,85. Ello significa que si tenemos dos muestras independientes, una de las cuales es de tamaño 16 y la otra de tamaño 11, la probabilidad de que el cociente (s_1^2/s_2^2) sea mayor que 2,85 es 0,05. Es decir, el valor 2,85 corresponde al límite inferior de un intervalo que se extiende hasta $+\infty$, y la probabilidad de que el valor de (s_1^2/s_2^2) se halle en el interior de este intervalo es 0,05.

Estas contrastaciones de las variaciones de la población sólo son exactamente ciertas si las poblaciones originarias son normales. Sin embargo, existen algunas indicaciones de que los resultados se pueden aplicar también en gran medida a otras clases de poblaciones originarias siempre y cuando no sean muy distintas de una población normal⁶. Pero en caso de que existan buenas razones para sospechar que la población originaria es muy asimétrica o tiene forma de U, no es posible confiar en estas contrastaciones.

6. Para un análisis de este tema, véase, por ejemplo, G. Udny Yule y M. G. Kendall. *An Introduction to the Theory of Statistics*. Londres, Griffin, 1950, p. 486.

Contrastación de la bondad del ajuste

La contrastación de la *bondad del ajuste* se aplica a los problemas en que se trata de decidir si la distribución de frecuencias de una muestra es compatible con cierta distribución teórica. Por ejemplo, se utilizaría para contrastar el supuesto de que una variable se distribuye normalmente. En general, la hipótesis nula consiste en la proposición de que cierta variable tiene una distribución de probabilidad determinada, mientras que la hipótesis alternativa señala que dicha proposición no es cierta. Para contrastar la hipótesis nula utilizamos la distribución de frecuencias obtenida en la muestra como evidencia relativa a la forma de la distribución de la población. En este caso, el estadístico de prueba utilizado corrientemente es

$$\sum_{i=1}^m \frac{(f_i - e_i)^2}{e_i},$$

en la que f_i es la frecuencia muestral en el i -ésimo intervalo, e_i es la frecuencia esperada en la distribución teórica (que se postula) y m es el número de intervalos. Se puede demostrar que este estadístico de prueba posee una distribución que para *muestras grandes* se aproxima a la *distribución ji-cuadrado*. En particular, si la muestra es grande, se cumple que

$$(5.11) \quad \sum_{i=1}^m \frac{(f_i - e_i)^2}{e_i} \sim \chi_{m-k-1}^2,$$

en la que el subíndice $(m - k - 1)$ se refiere al número de grados de libertad. Las frecuencias muestrales f_i se pueden observar, y las frecuencias teóricas e_i pueden calcularse utilizando la fórmula y la distribución especificada en la hipótesis nula. Esta fórmula incluirá probablemente algunos parámetros desconocidos que deberán ser reemplazados por sus correspondientes estimaciones muestrales. Por ejemplo, si la hipótesis nula señala que la distribución de la población es normal, será necesario estimar la media y la variancia de esta distribución a partir de la muestra. (En realidad, para que se cumpla la expresión (5.11) las estimaciones deben ser de una forma determinada, concretamente del tipo «máximoverosímil» —expresión que se explicará en la sección 6.2—. En este momento basta con señalar que X es una estimación máximoverosímil y que s^2 lo es aproximadamente en las muestras grandes.) El número de grados de libertad se obtiene de la forma siguiente:

m = número de intervalos;

k = número de parámetros que se han tenido que reemplazar por estimaciones muestrales.

Para que la contrastación sea razonablemente satisfactoria es preciso que $m \geq 5$ y $e_i \geq 5$ para cada i .

Si la hipótesis nula es cierta, f_i se puede tomar como una estimación muestral de e_i y la expresión (5.11) sólo será distinta de cero porque estamos observando una muestra en lugar de la población total. Por lo tanto, si observamos una muestra en la que el valor del estadístico de prueba (5.11) es grande, consideraremos este hecho como evidencia contraria a la hipótesis nula. Para efectuar la contrastación tenemos que obtener el límite entre la región crítica y la de aceptación, y éste depende del número de grados de libertad y del nivel de significación elegido, pudiendo obtenerse en la tabla de la ji-cuadrado. Obsérvese que el estadístico (5.11) no puede ser negativo y, por tanto, la evidencia contraria a la hipótesis nula sólo puede tomar la forma de valores muy grandes (pero no muy pequeños); a consecuencia de ello, la contrastación que resulta adecuada es la de *un extremo*.

EJEMPLO. Los economistas se interesan a menudo en la distribución de la renta personal. Consideremos la hipótesis de que la renta familiar se distribuye normalmente. Para contrastar esta hipótesis podemos utilizar los datos del cuadro 5-1. Estos datos se pueden tomar como una muestra de una población que incluye

Cuadro 5-1

Intervalo	Punto medio *	Porcentaje de familias †
Menos de \$2,000	1,130	12.7
\$2,000 a \$2,999	2,560	9.4
\$3,000 a \$3,999	3,490	10.8
\$4,000 a \$4,999	4,510	11.7
\$5,000 a \$5,999	5,480	11.4
\$6,000 a \$7,499	6,690	14.4
\$7,500 a \$9,999	8,570	13.9
\$10,000 a \$14,999	11,960	10.5
\$15,000 o más	22,780	5.2
Total		100.0
	Número total	57,890,000

* Los puntos medios se han calculado dividiendo la renta total (después de deducidos los impuestos) de cada grupo de renta por el número de familias del grupo.

† Incluye personas que viven solas.

Fuente: *Statistical Abstract of the United States, 1965*, Ministerio de Comercio de EE.UU., cuadro 467.

estas las rentas posibles que se hubiesen podido obtener en el año 1962 en los EE.UU. El estadístico que se utilizará en la contrastación es

$$\sum_{i=1}^g \frac{(f_i - e_i)^2}{e_i} = \left[\frac{\sum f_i}{100} \right] \sum \frac{(p_i - \pi_i)^2}{\pi_i},$$

en el que p_i = frecuencias observadas, en porcentaje, y π_i = frecuencias esperadas, también en porcentaje. Las frecuencias esperadas tienen que calcularse ajustando una distribución normal a los datos observados. Para ello, tenemos que estimar dos parámetros (la media y la variancia) a partir de la muestra. Utilizaremos como estimadores la media muestral y la variancia muestral, cuyos valores son

$$\bar{x} = 6507 \quad \text{y} \quad \hat{\sigma}^2 = 4920^2.$$

A fin de obtener las frecuencias de la distribución normal con la media y variancia citadas, seguiremos el procedimiento expuesto en la sección 4-2 e ilustrado en el cuadro 4-7. En primer lugar, formamos la variable normal tipificada

$$Z = \frac{X - 6507}{4920}$$

(en la que X = renta) y calcularemos de nuevo los límites de cada intervalo en términos de esta variable. A continuación, obtendremos las probabilidades normales correspondientes a cada grupo de renta, a partir de la tabla de las áreas situadas debajo de la curva normal. Los resultados son los presentados en el cuadro 5-2. Utilizando estos resultados vemos que

$$\sum_{i=1}^g \frac{(f_i - e_i)^2}{e_i} = 9,454,950.$$

En la tabla de la ji-cuadrado, el valor correspondiente a $9 - 2 - 1 = 6$ grados de libertad al nivel de significación del 1 % es 16,812. Los valores inferiores se hallan en la región de aceptación y los mayores en la región crítica. En el caso que nos ocupa el valor del estadístico de prueba es muy superior al valor límite de 16,812 y, por tanto, se rechazará la hipótesis nula. Es decir, los datos no resultan ser consistentes con la proposición de que las rentas familiares se distribuyen normalmente.

Conclusión

Con esto llegamos al final de esta sección, en la que se han presentado algunas contrastaciones básicas. Los objetivos buscados han sido dos. En primer lugar, se ha querido ilustrar el desarrollo de los procedimientos de contrastación en general, de modo que el lector pudiese ver en términos concretos los problemas que se presentan y la forma de resolverlos. En realidad, los problemas concretos que aparecen en esta sección y sus correspondientes contras-

taciones no se presentan muy a menudo en econometría. Ello se debe a que los modelos estadísticos que hemos utilizado son demasiado simples para satisfacer los requerimientos normales de la teoría económica. Concretamente, la concentración en una sola variable y la exclusión de todos los demás factores

Cuadro 5-2

Intervalos		Probabilidades normales acumuladas	Probabilidades normales	
x	z		$f(z)$	Porcentaje
Menos de 2,000	$-\infty$ a -0.92	0.1788	0.1788	17.9
2,000 a 2,999	-0.92 a -0.71	0.2388	0.0600	6.0
3,000 a 3,999	-0.71 a -0.51	0.3050	0.0662	6.6
4,000 a 4,999	-0.51 a -0.31	0.3783	0.0733	7.3
5,000 a 5,999	-0.31 a -0.11	0.4562	0.0779	7.8
6,000 a 6,999	-0.11 a 0.20	0.5793	0.1231	12.3
7,500 a 9,999	0.20 a 0.71	0.7612	0.1819	18.2
10,000 a 14,999	0.71 a 1.73	0.9582	0.1970	19.7
15,000 o más	1.73 a $+\infty$	1.0000	0.0418	4.2
			1.0000	100.0

no puede hacer justicia a la complejidad de las relaciones económicas. Sin embargo, las sencillas contrastaciones estudiadas en esta sección tienen un rasgo en común con las que se aplican en situaciones más complejas. Esta característica es el uso de las distribuciones que se han analizado en las páginas anteriores: la distribución normal, la ji-cuadrado, la t y la F . Este ha sido el segundo y más importante de los objetivos de esta sección. El estudio de estas sencillas contrastaciones ha permitido introducir de forma natural dichas distribuciones y nos ha dado la oportunidad de subrayar sus principales características y de relacionarlas entre sí. He aquí, pues, la razón por la que esta sección resulta verdaderamente indispensable para una completa comprensión de los métodos econométricos.

EJERCICIOS:

5-1. Sea $X \sim N(\mu, 81)$. La hipótesis nula y la alternativa son

$$H_0: \mu = 10,$$

$$H_A: \mu > 10.$$

Representar gráficamente la función de potencia de esta contrastación considerando que el estadístico de prueba se basa en una muestra de tamaño 9 y que el nivel de significación elegido es el 5%.

5-2. Al efectuar un análisis de mercado de los precios de los comestibles, se obtuvieron dos muestras del precio de un determinado producto. La muestra I se obtuvo en un área ciudadana de elevada densidad de población y la muestra II en los suburbios. Los resultados son los siguientes:

	Muestra I	Muestra II
n	14	18
$\sum_{i=1}^n p_i$	12.60	14.96
$\sum_{i=1}^n p_i^2$	1.68	1.96

en la que p_i = precio obtenido en la tienda i . Contrastar la hipótesis de que no existe diferencia entre el precio medio de este determinado producto en una y otra área. (Utilícese el nivel de significación del 5%.)

- 5-3. Al asignar subsidios federales entre distintas localidades, una determinada ciudad se clasificó como área de renta elevada. Las autoridades locales se mostraron disconformes y presentaron pruebas, basadas en una muestra aleatoria de 25 familias, de que la renta familiar media de la ciudad era 7 145 dólares. Un estudio a escala nacional de ciudades de tamaño similar indicó que la renta familiar media era de 6 500 dólares y la desviación standard 920 dólares. ¿Son consistentes los resultados muestrales presentados por las autoridades locales con su afirmación de que la ciudad en cuestión no es más próspera que las demás ciudades similares del país?
- 5-4. Sea Y = renta y $X = \log Y$. Contrastar la hipótesis de que X se distribuye normalmente, utilizando los datos correspondientes a la distribución de la renta en los Estados Unidos, presentados en el cuadro 5-1.
- 5-5. Plantear y realizar un experimento con la distribución ji-cuadrado para contrastar que las últimas cifras de los números que aparecen en la tabla de desviaciones normales del Apéndice D son aleatorias.
- 5-6. El propietario de una pizzería quiere establecer un servicio de reparto a domicilio para el que necesita una furgoneta. La elección se limita básicamente a dos marcas, A y B. La furgoneta A cuesta 300 dólares más que la B, pero se afirma que su consumo medio es de 24 millas por galón y, por tanto, mejor que las 20 millas por galón correspondientes a B. El propietario de la pizzería puede utilizar cada furgoneta durante 4 días antes de tomar una decisión. ¿Cuál debería ser el criterio de contrastación adecuado, teniendo

taciones no se presentan muy a menudo en econometría. Ello se debe a que los modelos estadísticos que hemos utilizado son demasiado simples para satisfacer los requerimientos normales de la teoría económica. Concretamente, la concentración en una sola variable y la exclusión de todos los demás factores

Cuadro 5-2

Intervalos		Probabilidades normales acumuladas	Probabilidades normales	
x	z		$f(z)$	Porcentaje
Menos de 2,000	$-\infty$ a -0.92	0.1788	0.1788	17.9
2,000 a 2,999	-0.92 a -0.71	0.2388	0.0600	6.0
3,000 a 3,999	-0.71 a -0.51	0.3050	0.0662	6.6
4,000 a 4,999	-0.51 a -0.31	0.3783	0.0733	7.3
5,000 a 5,999	-0.31 a -0.11	0.4562	0.0779	7.8
6,000 a 7,499	-0.11 a 0.20	0.5793	0.1231	12.3
7,500 a 9,999	0.20 a 0.71	0.7612	0.1819	18.2
10,000 a 14,999	0.71 a 1.73	0.9582	0.1970	19.7
15,000 o más	1.73 a $+\infty$	1.0000	0.0418	4.2
			1.0000	100.0

no puede hacer justicia a la complejidad de las relaciones económicas. Sin embargo, las sencillas contrastaciones estudiadas en esta sección tienen un rasgo en común con las que se aplican en situaciones más complejas. Esta característica es el uso de las distribuciones que se han analizado en las páginas anteriores: la distribución normal, la ji-cuadrado, la t y la F . Este ha sido el segundo y más importante de los objetivos de esta sección. El estudio de estas sencillas contrastaciones ha permitido introducir de forma natural dichas distribuciones y nos ha dado la oportunidad de subrayar sus principales características y de relacionarlas entre sí. He aquí, pues, la razón por la que esta sección resulta verdaderamente indispensable para una completa comprensión de los métodos econométricos.

EJERCICIOS:

5-1. Sea $X \sim N(\mu, 81)$. La hipótesis nula y la alternativa son

$$H_0: \mu = 10,$$

$$H_A: \mu > 10.$$

Representar gráficamente la función de potencia de esta contrastación considerando que el estadístico de prueba se basa en una muestra de tamaño 9 y que el nivel de significación elegido es el 5%.

5-2. Al efectuar un análisis de mercado de los precios de los comestibles, se obtuvieron dos muestras del precio de un determinado producto. La muestra I se obtuvo en un área ciudadana de elevada densidad de población y la muestra II en los suburbios. Los resultados son los siguientes:

	Muestra I	Muestra II
n	14	18
$\sum_{i=1}^n p_i$	12.60	14.96
$\sum_{i=1}^n p_i^2$	1.68	1.96

en la que p_i = precio obtenido en la tienda i . Contrastar la hipótesis de que no existe diferencia entre el precio medio de este determinado producto en una y otra área. (Utilícese el nivel de significación del 5%.)

5-3. Al asignar subsidios federales entre distintas localidades, una determinada ciudad se clasificó como área de renta elevada. Las autoridades locales se mostraron disconformes y presentaron pruebas, basadas en una muestra aleatoria de 25 familias, de que la renta familiar media de la ciudad era 7 145 dólares. Un estudio a escala nacional de ciudades de tamaño similar indicó que la renta familiar media era de 6 500 dólares y la desviación standard 920 dólares. ¿Son consistentes los resultados muestrales presentados por las autoridades locales con su afirmación de que la ciudad en cuestión no es más próspera que las demás ciudades similares del país?

5-4. Sea Y = renta y $X = \log Y$. Contrastar la hipótesis de que X se distribuye normalmente, utilizando los datos correspondientes a la distribución de la renta en los Estados Unidos, presentados en el cuadro 5-1.

5-5. Plantear y realizar un experimento con la distribución ji-cuadrado para contrastar que las últimas cifras de los números que aparecen en la tabla de desviaciones normales del Apéndice D son aleatorias.

5-6. El propietario de una pizzería quiere establecer un servicio de reparto a domicilio para el que necesita una furgoneta. La elección se limita básicamente a dos marcas, A y B. La furgoneta A cuesta 300 dólares más que la B, pero se afirma que su consumo medio es de 24 millas por galón y, por tanto, mejor que las 20 millas por galón correspondientes a B. El propietario de la pizzería puede utilizar cada furgoneta durante 4 días antes de tomar una decisión. ¿Cuál debería ser el criterio de contrastación adecuado, teniendo

en cuenta que la vida útil esperada de cada furgoneta es 50 000 millas y que el coste de la gasolina es de 40 centavos por galón?

- 5-7. Se ha afirmado que las fluctuaciones cíclicas en la demanda de trabajadores de edades inferiores a 20 años son mayores que las de la demanda correspondiente a hombres casados. Contrastar esta afirmación utilizando los datos del cuadro 5-3.

Cuadro 5-3 Desempleo en los EE.UU.

	Tasa de desempleo, %	
	Hombres casados	Jóvenes 14-19 años
1949	3.4	12.2
1950	4.6	11.3
1951	1.5	7.7
1952	1.4	8.0
1953	1.7	7.1
1954	4.0	11.4
1955	2.6	10.2
1956	2.3	10.4
1957	2.8	10.8
1958	5.1	14.4
1959	3.6	13.2
1960	3.7	13.6
1961	4.6	15.2

Fuente: *Economic Report of the President, 1965*, Washington, D.C., U.S. Government Printing Office, 1965, p. 217.

Estimación

En la introducción ya comentamos la división tradicional de los problemas de la inferencia estadística en problemas de contrastación de hipótesis y de estimación. La similitud entre ambos tipos radica en el hecho de que tanto uno como otro se ocupan de problemas relativos al valor de algún parámetro o parámetros desconocidos de la población. La diferencia consiste en que en la estimación, contrariamente a lo que sucede en la contrastación de la hipótesis, no efectuamos ninguna afirmación previa cuya credibilidad se vaya a discutir. En la contrastación de hipótesis, los ingredientes iniciales son la información previa (en forma de una hipótesis sostenida) y una afirmación relativa al valor del parámetro en cuestión. En la estimación también empezamos a partir de la información previa (en forma de un modelo), pero no tenemos ninguna idea previa respecto al valor del parámetro. (La especificación a priori del posible recorrido del parámetro —por ejemplo, la especificación de que debe ser positivo— se considera parte del modelo.) Como ya hemos mencionado anteriormente, los problemas de estimación han recibido mayor atención que los de contrastación de hipótesis por parte de los econométricos y, por tanto, la teoría de la estimación para los distintos tipos de modelos económicos está bastante avanzada aunque, naturalmente, siguen existiendo dificultades.

Podemos dividir la teoría de la estimación en dos partes, la estimación por puntos y la estimación por intervalos. En la *estimación por puntos*, el objetivo consiste en calcular, utilizando la información a priori y la proporcionada por la muestra, un valor que sea, en algún sentido, el mejor pronóstico acerca del valor real del parámetro que nos interesa. En la *estimación por intervalos*, se utiliza la misma información con el fin de obtener un intervalo que tenga un nivel determinado de probabilidad de abarcar el verdadero valor del parámetro. Como un intervalo queda completamente definido por sus límites, la estimación del intervalo equivale a estimar sus límites. Normalmente, estos intervalos reciben el nombre de *intervalos de confianza* y pueden considerarse también como una posible medida de la precisión de un estimador por puntos. Al final de este capítulo los consideraremos en este sentido en la sección 6.3.

El problema de la estimación por puntos consiste en obtener una estimación que represente nuestro mejor pronóstico del valor del parámetro. Para resolver este problema tenemos que hacer dos cosas: primeramente hemos de concretar el significado del «mejor pronóstico» y, en segundo lugar, tenemos que diseñar estimadores que satisfagan este criterio o al menos se aproximen a ello. En otras palabras, tenemos que especificar lo que queremos y establecer una fórmula que nos diga cómo hacerlo o al menos cómo conseguir una buena aproximación. La primera parte del problema —la definición del mejor pronóstico— equivale a especificar las propiedades que se pueden considerar deseables en un estimador. Como un estimador es una variable aleatoria cuyo valor varía de una muestra a otra, sus propiedades son en realidad las de su distribución muestral. Estas propiedades se estudiarán en la sección 6.1. La segunda parte del problema corresponde al establecimiento de unos estimadores que tengan al menos algunas de las propiedades que se desean y será el tema de la sección 6.2. La última sección se dedicará al estudio de los intervalos de confianza, como ya hemos mencionado.

6.1 Propiedades de los estimadores

Consideremos una variable aleatoria X cuya distribución se caracteriza, entre otras cosas, por un parámetro θ que queremos estimar. La población originaria está formada, pues, por todos los valores posibles de X , y θ es una de las características paramétricas de esta población. La variable X puede ser discreta o continua, o puede ser incluso un atributo (es decir, una variable binaria). Un ejemplo podría ser la renta familiar —que es una variable continua— y su media. Esta especificación constituye un problema de estimación relativamente simple; un problema más complicado podría ser la estimación conjunta de varios parámetros correspondientes a distintas variables. En realidad, el problema de estimación simple es más que suficiente para describir las distintas propiedades de los estimadores y esbozar lo métodos básicos de estimación. En los capítulos siguientes consideraremos los problemas de estimación más complejos, que son típicos de la econometría.

Para estimar un parámetro de la población combinamos la información a priori que podamos tener con la información que nos proporciona la muestra. La información a priori no es más que lo que, en el contexto de la contrastación de hipótesis, denominados «hipótesis sostenida». Cuando se trata de estimación, normalmente utilizamos la expresión *modelo* pero *hipótesis sostenida* resulta también una expresión totalmente aceptable. Esta información previa hace referencia a la población de X , y puede consistir en supuestos acerca de la forma de las distribuciones, el valor de ciertos parámetros distintos de θ , o algún tipo de especificación (por ejemplo, el recorrido) de la propia θ . La información proporcionada por la muestra es la correspondiente a las observaciones muestrales $X_1, \dots, X_2, X_n$. La forma de utilizar la información a fin de

obtener una estimación de θ viene indicada por la fórmula de estimación a la que llamamos *estimador*. Es posible que exista más de una de estas fórmulas entre las que elegir, y normalmente esto es lo que ocurrirá. En la mayor parte de este capítulo consideraremos únicamente los problemas en los que la información a priori es mínima, y en el resto del libro se estudiarán problemas relativos a modelos más elaborados.

Llamaremos $\hat{\theta}$ al estimador del parámetro θ , que es una de las características de la distribución de X . Como $\hat{\theta}$ se obtiene sustituyendo en una fórmula las observaciones muestrales de X , podemos escribir

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n),$$

que se lee « $\hat{\theta}$ es función de $x_1, x_2, \dots, x_n$ ». (Si no se utilizan *todas* las observaciones muestrales, la expresión se modifica en la forma que corresponda.) Esta función puede tener cualquier forma *salvo* por la restricción de que no debe incluir ningún parámetro desconocido ni, naturalmente, la propia θ . Las características básicas de la distribución de $\hat{\theta}$ son su media $E(\hat{\theta})$ y su variancia,

$$\text{Var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 = E(\hat{\theta}^2) - [E(\hat{\theta})]^2.$$

La desviación standard de $\hat{\theta}$, que se define como $\sqrt{\text{Var}(\hat{\theta})}$, recibe el nombre de *error standard* de $\hat{\theta}$. Los conceptos siguientes son también muy importantes:

$$\text{Error muestral} = \hat{\theta} - \theta,$$

$$\text{Sesgo} = E(\hat{\theta}) - \theta,$$

$$\text{Media del error al cuadrado} = E(\hat{\theta} - \theta)^2.$$

El *error muestral* es simplemente la diferencia entre el valor del estimador y el verdadero valor del parámetro que se estima. Naturalmente, este error muestral varía de una muestra a otra. El *sesgo* es la diferencia entre la media de la distribución muestral de un estimador determinado y el verdadero valor del parámetro. Para cada estimador, este valor es una cantidad fija que puede ser o no igual a cero. Por último, la *media del error al cuadrado* es un concepto relacionado con la dispersión de la distribución del estimador y, por tanto, es similar al concepto de variancia. La diferencia entre la variancia de un estimador y la media del error al cuadrado es que la primera mide la dispersión de la distribución *en torno a la media*, mientras que la última mide la dispersión *en torno al verdadero valor del parámetro*. Si la media de la distribución

coincide con dicho valor, la variancia y la media del error al cuadrado son iguales, pero, de lo contrario, son distintos.

La relación entre la media del error al cuadrado (MEC) y la variancia puede mostrarse explícitamente de la forma siguiente

$$\begin{aligned}\text{MEC } (\hat{\theta}) &= E(\hat{\theta} - \theta)^2 = E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= E\{[\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta]\}^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta].\end{aligned}$$

Consideremos el último término,

$$2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = 2\{[E(\hat{\theta})]^2 - [E(\hat{\theta})]^2 - \theta E(\hat{\theta}) + \theta E(\hat{\theta})\} = 0.$$

Teniendo en cuenta esto, y observando que el valor esperado de una constante es simplemente esta misma constante, podemos escribir

$$\text{MEC}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 = \text{variancia más sesgo al cuadrado}.$$

Es decir, el valor de la media del error al cuadrado no puede ser nunca inferior al de la variancia, y la diferencia entre los dos es precisamente igual al sesgo al cuadrado.

Pasemos ahora a considerar algunas de las propiedades que normalmente se conceptúan como deseables en los estimadores. Podemos dividir estas propiedades en dos grupos, según el tamaño de la muestra. Las *propiedades de las muestras finitas o pequeñas* hacen referencia a las propiedades de la distribución muestral de los estimadores basados en muestras de cualquier tamaño fijo. Las propiedades de las muestras finitas pueden caracterizar a las estimaciones obtenidas a partir de cualquier número de observaciones, y normalmente se llaman propiedades de las muestras pequeñas porque se cumplen *incluso* aunque las muestras sean pequeñas. Por otra parte, las *propiedades de las muestras grandes o asintóticas* se limitan solamente a las distribuciones muestrales que proceden de muestras cuyo tamaño tiende a infinito. Se supone que estas propiedades, cuando son de aplicación, sólo se cumplen aproximadamente cuando el tamaño de la muestra es grande, y que posiblemente no se cumplen en absoluto cuando las muestras son pequeñas. Cuando consideremos las distintas propiedades asintóticas estudiaremos esto con mayor precisión, pero, en primer lugar, nos ocuparemos de las muestras finitas.

Propiedades de las muestras pequeñas

La primera propiedad que vamos a mencionar es la *inesgadez*. Esta propiedad de los estimadores es la más ampliamente conocida entre los investigadores que se dedican al trabajo empírico. En la sección 1.4 ya hemos explicado el significado de la inesgadez y hemos señalado que un estimador inesgado es aquel cuya media es igual al valor del parámetro de la población que vamos a estimar. En esta ocasión, habiendo estudiado ya el significado y la terminología de la esperanza matemática, podemos definir la inesgadez de forma a la vez técnica y precisa del modo siguiente:

$$(6.1) \quad \hat{\theta} \text{ es un estimador inesgado de } \theta \text{ si } E(\hat{\theta}) = \theta.$$

En la figura 6.1 se representa gráficamente un estimador inesgado; como

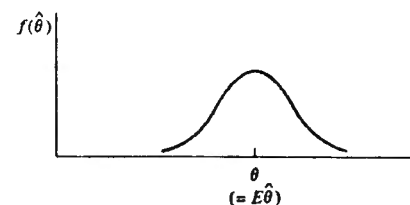


Figura 6-1

la distribución considerada es simétrica, la media se halla en el centro de la distribución y coincide con el valor del parámetro. Como ejemplo de estimador inesgado podemos tomar la media de la muestra como estimador de la media de la población, ya que

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu.$$

(Obsérvese que $E(X_i)$ es la media de X_i , que es μ .)

Es preciso señalar que la inesgadez no es, en sí misma, una propiedad muy satisfactoria, ya que no implica absolutamente nada respecto a la *dispersión* de la distribución del estimador. Un estimador que sea inesgado, pero que tenga variancia muy grande, producirá a menudo estimaciones muy alejadas

del objetivo. Por otra parte, un estimador cuya variancia es muy pequeña pero es sesgado —y no sabemos cuál es el sesgo— es todavía menos útil. Podemos ver este punto tomando el caso extremo de un estimador cuya variancia es cero. No es difícil construir un estimador de este tipo, ya que todas las constantes, cuya variancia es, por definición, cero, sirven para ello. Por tanto, si decidimos que nuestro estimador de θ será siempre el número 5, la distribución muestral de este «estimador» estará totalmente concentrada en el punto $\hat{\theta} = 5$. Pero un estimador de este tipo tiene, obviamente, muy poco sentido, puesto que no presta ninguna atención a la evidencia proporcionada por la muestra y, por tanto, prescinde de toda la información que tenga este origen. En vista de este argumento, parece que lo deseable sería un estimador que minimizase la media del error al cuadrado. Efectivamente, puede demostrarse que este estimador sería óptimo en el caso en que la pérdida debida al uso de un estimador en lugar del verdadero valor del parámetro aumenta cuando la distancia entre $\hat{\theta}$ y θ también aumenta. Desgraciadamente, en la práctica, la fórmula de un estimador que proporcione un valor mínimo de la media del error al cuadrado incluye la mayor parte de las veces al verdadero valor del parámetro que vamos a estimar. Obviamente, esto hace que la fórmula sea totalmente inútil; sería como si una receta de cocina para confeccionar un pastel empezase diciendo «tómese un pastel...». Nuestra definición de estimador eliminó de forma específica este tipo de fórmulas no considerándolas merecedoras del título de «estimador».

El análisis anterior proporciona una base para introducir el concepto de *eficiencia* en la estimación. Como ya mencionamos en la sección 1.4, no existe un acuerdo general acerca de la definición más apropiada de este concepto en la literatura estadística. Algunos autores equiparan eficiencia con valor mínimo de la media del error al cuadrado, pese a la dificultad que acabamos de mencionar; otros autores sólo definen la eficiencia en el contexto de las propiedades asintóticas, pero no en el de las muestras finitas, y otros consideran que un estimador es eficiente si (y únicamente si) es insesgado y al mismo tiempo tiene una variancia mínima. Este último concepto de eficiencia se está haciendo muy corriente entre los económetras y será el que adoptaremos aquí. Por consiguiente, definiremos formalmente la eficiencia en la forma siguiente:

- (6.2) $\hat{\theta}$ es un estimador eficiente de θ si satisface las condiciones siguientes:
- $\hat{\theta}$ es insesgado;
 - $\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta})$, en la que $\tilde{\theta}$ es cualquier otro estimador insesgado de θ .

Este tipo de estimador eficiente que acabamos de definir se denomina también a veces «estimador insesgado de variancia mínima» o «estimador insesgado óptimo». Obsérvese, que, según esta definición, un estimador que sea ligeramente sesgado no podrá llamarse eficiente, aunque su variancia sea muy pe-

queña. En la figura 6.2 se representa gráficamente este concepto de eficiencia. En esta figura se presentan las distribuciones de tres estimadores de θ , $\hat{\theta}_a$, $\hat{\theta}_b$ y $\hat{\theta}_c$. El que tiene la menor variancia es $\hat{\theta}_a$, pero no es eficiente porque es sesgado. Por otra parte, $\hat{\theta}_b$ y $\hat{\theta}_c$ son insesgados, pero este último tiene una variancia mayor que $\hat{\theta}_b$, o sea que tampoco es eficiente. Por tanto, sólo queda $\hat{\theta}_b$, que es eficiente *siempre y cuando* no exista otro estimador insesgado cuya variancia sea más pequeña.

Esta última frase nos lleva a considerar un aspecto práctico de la eficiencia, que es el problema de decidir si un determinado estimador es eficiente o no. Este tipo de problema no nos ha causado ninguna preocupación en el caso de la insesgadez, ya que, en dicho caso, se trata de un problema que, al menos en principio, es trivial. Lo único que tenemos que hacer para comprobar si un estimador es insesgado o no es calcular su esperanza matemática, es decir la media de su distribución muestral. Pero cuando se trata de la eficiencia, y en

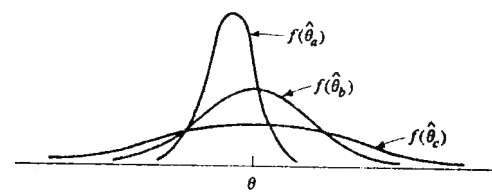


Figura 6-2

especial, de la condición de variancia mínima, el problema es potencialmente muy complejo. Como tenemos que efectuar una afirmación acerca de la variancia de todos los estimadores insesgados, que pueden ser infinitos, puede resultar muy difícil afirmar que un estimador concreto es eficiente. Una forma de evitar esta dificultad consiste en reducir nuestras exigencias y, en lugar de decir que un determinado estimador insesgado es mejor —en el sentido de que su variancia es más pequeña— o al menos es tan bueno como cualquier otro estimador insesgado, nos podemos dar por satisfechos señalando que el estimador en cuestión es mejor que algún otro estimador insesgado. Por tanto, al comparar dos estimadores insesgados, nos podemos ocupar simplemente de su *eficiencia relativa* y declarar que el estimador que tiene la variancia más pequeña es más eficiente que el otro. Un ejemplo de ello es la comparación entre la media y la mediana de la muestra como estimadores de la media de la población. Como ya vimos en la sección 2.2, ambos estimadores son insesgados, pero la variancia de la media de la muestra es más pequeña que la de

la mediana. En este caso, podemos decir que la media de la muestra es un estimador más eficiente de la media de la población que la mediana de la muestra.

Afortunadamente, en algunos casos no tenemos necesidad de limitarnos a comparar las variancias de un pequeño número de estimadores, ya que podemos efectuar una afirmación clara respecto a la eficiencia en sentido absoluto. La razón de ello es la existencia del teorema que presentamos a continuación, y que se conoce con el nombre de *desigualdad de Cramer-Rao*.

Teorema 16. Sea X una variable aleatoria con una distribución de probabilidad $f(x)$ que se caracteriza por los parámetros $\theta_1, \theta_2, \dots, \theta_k$, y sea $\hat{\theta}_i$ cualquier estimador insesgado de θ_i obtenido a partir de una muestra $X_1, X_2, \dots, X_n$. Definamos $L = \log f(x_1, x_2, \dots, x_n)$; la función L se conoce con el nombre de *función logarítmica de verosimilitud* de una muestra determinada. Consideremos la siguiente matriz:

$$\begin{bmatrix} -E\left[\frac{\partial^2 L}{\partial \theta_1^2}\right] & -E\left[\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}\right] & \cdots & -E\left[\frac{\partial^2 L}{\partial \theta_1 \partial \theta_k}\right] \\ -E\left[\frac{\partial^2 L}{\partial \theta_2 \partial \theta_1}\right] & -E\left[\frac{\partial^2 L}{\partial \theta_2^2}\right] & \cdots & -E\left[\frac{\partial^2 L}{\partial \theta_2 \partial \theta_k}\right] \\ \vdots & \vdots & \ddots & \vdots \\ -E\left[\frac{\partial^2 L}{\partial \theta_k \partial \theta_1}\right] & -E\left[\frac{\partial^2 L}{\partial \theta_k \partial \theta_2}\right] & \cdots & -E\left[\frac{\partial^2 L}{\partial \theta_k^2}\right] \end{bmatrix}$$

Esta matriz se conoce con el nombre de *matriz de información*. Tomemos ahora la inversa de la matriz de información y llamemos I^{ii} al elemento de la i -ésima fila y la i -ésima columna de esta matriz inversa. En este caso, la desigualdad de Cramer-Rao señala que

$$\text{Var}(\hat{\theta}_i) \geq I^{ii}.$$

Este teorema nos permite construir un límite inferior (mayor que cero) para la variancia de cualquier estimador insesgado, suponiendo que podamos especificar la forma funcional en la distribución originaria.¹ El límite inferior especificado en este teorema se denomina *límite inferior de Cramer-Rao*. Si podemos obtener un estimador insesgado cuya variancia sea igual al límite inferior de Cramer-Rao, sabremos que no existe ningún otro estimador insesgado que tenga

1. Esta desigualdad se cumple bajo condiciones muy generales. Los detalles completos se presentan, por ejemplo, en C. R. Rao, *Linear Statistical Inference and Its Applications*. Nueva York, Wiley, 1965, pp. 265 y ss.

una variancia más pequeña, y el estimador considerado será eficiente. Por ejemplo, sabemos que la variancia de la media muestral $\bar{X}$ es igual a (σ^2/n) . Ahora bien, se puede demostrar que, si la población originaria es una normal, (σ^2/n) es en realidad el límite inferior de Cramer-Rao para los estimadores insesgados de la media de la población normal. Debe señalarse, sin embargo, que el uso de la desigualdad de Cramer-Rao no nos soluciona siempre el problema, ya que el límite inferior puede no ser *alcanzable* por ningún estimador insesgado. Por ejemplo, cuando se trata de estimar la variancia de una población normal, el límite inferior de Cramer-Rao para un estimador insesgado es $(2\sigma^4/n)$, pero no existe ningún estimador insesgado de σ^2 que tenga una variancia tan pequeña como ésta.²

El análisis anterior indica que no siempre es fácil obtener la eficiencia. En el mejor de los casos, tenemos que poder especificar la forma de la distribución originaria y confiar en la existencia de un estimador insesgado cuya variancia sea igual al límite inferior de Cramer-Rao. Si no sabemos la forma de la distribución originaria, o si no se puede alcanzar el límite inferior de Cramer-Rao, existen pocas esperanzas de que podamos establecer si un estimador determinado es eficiente o no. Por este motivo, podemos estar dispuestos a abandonar la idea de buscar un estimador cuya variancia sea la mínima de entre todos los estimadores insesgados, y podemos limitar nuestra atención a una clase más reducida de estimadores insesgados. En realidad, resulta que el problema de hallar un estimador insesgado cuya variancia sea mínima es relativamente sencillo si nos limitamos a los estimadores que sean *funciones lineales* de las observaciones muestrales. Este hecho ha dado lugar a la definición de un concepto de eficiencia más especializado, que presentamos a continuación:

(6.3) $\hat{\theta}$ es un estimador lineal insesgado óptimo (ELIO) de θ si satisface las tres condiciones siguientes:

- $\hat{\theta}$ es una función lineal de las observaciones muestrales;
- $\hat{\theta}$ es insesgado;
- $\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta})$, en la que $\tilde{\theta}$ es cualquier otro estimador lineal insesgado de θ .

La condición de linealidad significa que para una muestra $X_1, X_2, \dots, X_n$, el estimador debe tener la forma $a_1X_1 + a_2X_2 + \dots + a_nX_n$ en la que $a_1, a_2, \dots, a_n$ son constantes. Por ejemplo, $\bar{X}$ es un estimador lineal ya que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n.$$

Debemos señalar que el hecho de limitarnos a la clase de los estimadores lineales

2. *Ibid.*, p. 268.

les puede no ser demasiado restrictivo, ya que gran cantidad de funciones se pueden aproximar razonablemente bien mediante una función lineal para un amplio campo de valores. En general, nos podemos encontrar con alguna de las situaciones siguientes:

1. El estimador eficiente es lineal en las observaciones muestrales. En este caso, el estimador lineal insesgado óptimo y el estimador eficiente son idénticos. Por ejemplo, como $\bar{X}$ es un estimador eficiente de la media de una población normal, μ , y además $\bar{X}$ es lineal en las observaciones neutrales, esta última será también el estimador lineal insesgado óptimo de μ .
2. El estimador eficiente es *aproximadamente* lineal. En este caso, el estimador lineal insesgado óptimo no es eficiente, pero es probable que su variancia sea muy parecida a la del estimador eficiente.
3. El estimador eficiente es altamente no lineal. En este caso, la variancia del estimador eficiente puede ser mucho más pequeña que la del estimador lineal insesgado óptimo. Cabe esperar que a menudo esta situación se presente muy poco.

Otra propiedad de las muestras finitas de un estimador que se menciona a veces es la *suficiencia*. Se dice que un estimador es suficiente si utiliza toda la información relativa al parámetro contenida en la muestra. Como el valor de cada observación nos dice algo acerca de la población, para que un estimador sea suficiente debe basarse en los valores de todas las observaciones muestrales. Por ejemplo, la mediana de la muestra no es un estimador suficiente, ya que utiliza solamente el orden, pero no los valores de las observaciones muestrales. Obsérvese que la suficiencia no es en sí misma deseable; obviamente, no nos preocupamos de si una fórmula de estimación utiliza o no todas las observaciones muestrales si nos proporciona buenas estimaciones del parámetro en cuestión. La importancia real de la suficiencia radica en el hecho de que esta última es una condición necesaria para la eficiencia.⁵ Es decir, un estimador no puede ser eficiente —según la definición efectuada en (6.2)— si no utiliza toda la información proporcionada por la muestra. Este es el aspecto importante de la suficiencia y el motivo por el cual la mencionamos.

Estas tres propiedades —insesgadez, eficiencia y ser lineal insesgado y óptimo— son todas las correspondientes a las muestras pequeñas cuya importancia hace que se mencionen generalmente en las obras de econometría. Todas ellas se definen en términos de medias y variancia, y, por tanto, no se pueden obtener para los estimadores cuyas medias o variancias no existen. Conside-

remos, por ejemplo, la insesgadez. Si $\hat{\theta}$ es un estimador de θ cuya distribución es continua y puede representarse por $f(\hat{\theta})$, para que $\hat{\theta}$ sea insesgado es necesario que

$$E(\hat{\theta}) = \theta.$$

Pero, por definición,

$$E(\hat{\theta}) = \int_{-\infty}^{+\infty} \hat{\theta} f(\hat{\theta}) d\hat{\theta}.$$

Ahora bien, esta integral representa el área que se halla debajo de la curva $\hat{\theta} f(\hat{\theta})$, entre $-\infty$ y $+\infty$, y no podemos excluir la posibilidad de que esta área sea infinita. Si esto es así, diremos que la integral es *divergente* y que la media de esta distribución *no existe*. Mencionamos este punto porque nos encontraremos con estimadores cuya media o variancia puede no existir, y el lector debe entender claramente qué es lo que ello significa.

Propiedades asintóticas

Pasemos ahora a considerar las propiedades asintóticas de los estimadores. Como ya dijimos anteriormente, estas propiedades se refieren a la distribución de un estimador en el caso en que el tamaño de la muestra es grande y tiende a infinito. En general, la distribución de un estimador, basada en una muestra de un tamaño determinado, es distinta de la distribución del mismo estimador cuando procede de una muestra de otro tamaño. Estas distribuciones pueden ser distintas no sólo respecto a la media o a la variancia, sino incluso en relación con su forma matemática. Por ejemplo, tomemos la distribución de la media de las muestras de una población uniforme discreta, ejemplo que ya utilizamos al empezar la sección 4.3. En aquella ocasión vimos que, cuando las muestras eran de tamaño 1, la distribución era uniforme, es decir, de forma rectangular; cuando las muestras eran de tamaño 2, la distribución era triangular (ver figura 4-12), y cuando las muestras eran de tamaño 5 y 10, la distribución era parecida a la normal. El proceso por el que la distribución de la media muestral correspondiente a *cualquier* tipo de población varía se describe en el teorema central del límite (es decir, el teorema 15 de la sección 4.3). Básicamente, este teorema afirma que, a medida que aumenta el tamaño de la muestra, la distribución de la media muestral se acerca a la distribución muestral. En este caso, decimos que la distribución normal es la *distribución asintótica* (o *límite*) de la media de la muestra. En general, cuando la distribución de un estimador tiende cada vez más a tomar la forma de una determinada distribución al aumentar el tamaño de la muestra, esta última distribución recibe el nombre de *distribución asintótica* de dicho estimador.

El hecho de que usemos la palabra «asintótica» no debe hacer pensar al lector que esta distribución es necesariamente la forma final tomada por la distribución de un estimador cuando el tamaño de la muestra tiende a infi-

5. Véase, por ejemplo, A. M. Mood y F. A. Graybill, *Introduction to the Theory of Statistics*. Nueva York, McGraw-Hill, 1963, p. 176.

nito. En realidad, lo que ocurre típicamente en este caso es que la distribución del estimador se reduce a un solo punto —y cabe esperar que este punto sea el verdadero valor del parámetro—. (Cuando una distribución se concentra por completo en un solo punto, se llama *degenerada*.) Como ejemplo, tomemos de nuevo la distribución de la media muestral. Ya sabemos (por el teorema 14 de la sección 4.3) que para *todos* los tamaños muestrales, la media de esta distribución es igual a la media de la población y su variancia es igual a (σ^2/n) , en la que σ^2 es la variancia de la población, y n es el tamaño de la muestra. Ahora bien, como puede verse claramente, cuando el tamaño de la muestra tiende a infinito, (σ^2/n) tiende a cero, y la distribución se concentrará en la media de la población. Una representación gráfica de esta distribución nos mostraría una línea vertical de altura igual a uno. Es evidente que esto *no* es la distribución normal que, como sabemos por el teorema central del límite, representa la distribución asintótica de la media muestral. Cuando hablamos de la distribución asintótica no queremos decir la forma final de la distribución, que puede ser una forma degenerada, sino la forma que esta distribución tiende a tomar en la última parte del proceso por el cual tiende hacia el colapso final (de producirse éste). Con respecto a la distribución de la media muestral, al aumentar el tamaño de la muestra la distribución tendrá una variancia cada vez más pequeña, pero también se parecerá cada vez más a una distribución normal. Un momento antes de que la distribución se concentre en un solo punto, no podrá distinguirse de una distribución normal, aunque se tratará de un caso en que la variancia será extraordinariamente pequeña.

Una vez aclarado el significado del término «distribución asintótica», podemos pasar ahora al problema de cómo determinar su existencia y su forma. En muchos casos, éste es un problema relativamente sencillo. En primer lugar, algunos estimadores tienen una distribución cuya forma no varía, sea cual sea el tamaño de la muestra, y esta forma es conocida. En este caso, los estimadores tendrán también esta forma cuando el tamaño de la muestra es grande y tiende a infinito. Por tanto, la distribución asintótica de estos estimadores es igual a la distribución de las muestras finitas. Un ejemplo de este caso es la media de la muestra como estimador de la media de una población normal. La distribución de esta media muestral es una normal sea cual sea el tamaño de la muestra y su media es igual a la media de la población mientras que su variancia es (σ^2/n) . Por tanto, la distribución asintótica de la media de la muestra es también una normal de media μ y variancia (σ^2/n) . En segundo lugar, algunos estimadores tienen una distribución conocida para todos los tamaños muestrales, aunque no necesariamente con la misma forma para todos ellos. La distribución asintótica de estos estimadores es la distribución basada en una muestra cuyo tamaño tiende a infinito. Un ejemplo de este caso es la distribución de la proporción de aciertos en una muestra. Como ya vimos en la sección 4.2, esta distribución es binomial, pero tiende a una nor-

mal cuando n tiende a infinito. Por tanto, podemos decir que la distribución asintótica de la proporción de aciertos de una muestra es una normal. En tercer lugar, en algunos casos no conocemos la distribución del estimador para todos los tamaños muestrales, pero sí la correspondiente al caso en que $n \rightarrow \infty$. Un ejemplo de este tipo de estimadores es la media de la muestra como estimador de la media de una población que no es normal. Por el teorema central del límite sabemos que esta distribución tiende a hacerse normal cuando $n \rightarrow \infty$. Las tres clases de estimadores que acabamos de mencionar abarcan la mayor parte de los casos con que nos encontramos en los problemas de estimación en economía. Además, en la práctica, muchas distribuciones asintóticas resultan ser normales, lo cual resulta muy conveniente teniendo en cuenta que la normal es una distribución muy conocida.

Las distribuciones asintóticas pueden caracterizarse, al igual que las demás distribuciones, por sus momentos. Los más importantes son la media, llamada *media asintótica*, y la variancia, llamada también *variancia asintótica*. La media asintótica se puede obtener simplemente calculando el límite (cuando $n \rightarrow \infty$) de la media de la muestra finita. Consideremos un estimador $\hat{\theta}$. La media de este estimador es, por definición, igual a su esperanza matemática, es decir, $E(\hat{\theta})$. La *media asintótica* es igual a la *esperanza asintótica*, y viene dada por el $\lim_{n \rightarrow \infty} E(\hat{\theta})$. Sin embargo, la *variancia asintótica* no es igual a $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta})$. La razón de ello es que, en el caso de los estimadores cuya variancia disminuye al aumentar n , esta variancia tenderá a cero cuando $n \rightarrow \infty$. Esto es lo que ocurre cuando la distribución se concentra en un solo punto, pero, como ya hemos señalado, *no* es lo mismo la distribución asintótica que la distribución concentrada (degenerada), y la variancia de aquella *no* es cero. Tomemos, por ejemplo, la distribución de la media de la muestra. Su distribución asintótica es una normal y su variancia es (σ^2/n) . Pero $\lim_{n \rightarrow \infty} (\sigma^2/n) = 0$, y esto no es la variancia de una distribución normal. Así pues, la expresión «variancia asintótica» es un tanto engañosa; estrictamente hablando, se trata de la abreviación de la expresión «variancia de la distribución asintótica». La fórmula que defina automáticamente la variancia asintótica es⁴

$$\text{Var asintótica}(\hat{\theta}) = \frac{1}{n} \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\theta} - \lim_{n \rightarrow \infty} E\hat{\theta})]^2.$$

Vamos a describir ahora tres de las llamadas propiedades asintóticas que se consideran deseables en un estimador: insesgadez asintótica, consistencia, y eficiencia asintótica. Dos de ellas, la insesgadez asintótica y la eficiencia asintótica, se definen en términos de las características de la distribución asintótica del estimador, tal como las acabamos de definir, mientras que la tercera, es

4 Ver A. S. Goldberger, *Econometric Theory*, Nueva York, Wiley, 1964, p. 116.

decir, la consistencia, se define como una característica de la distribución «concentrada» (es decir, degenerada) cuando $n \rightarrow \infty$.

Consideremos en primer lugar la *inesegadez asintótica*.

(6.4) $\hat{\theta}$ es un estimador asintóticamente inesgado de θ si $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$.

Esta definición afirma simplemente que un estimador es inesgado asintóticamente si se convierte en inesgado cuando el tamaño de la muestra tiende a infinito. Obsérvese que, si un estimador es inesgado, también lo será asintóticamente, pero lo contrario no es necesariamente cierto. La inesgadez implica la inesgadez asintótica porque, cuando un estimador es inesgado, su esperanza es igual al verdadero valor del parámetro para *todos* los tamaños muestrales, incluyendo los próximos a infinito. Un ejemplo corriente de un estimador sesgado pero asintóticamente inesgado es la variancia de la muestra:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

y, según la expresión (5.4),

$$\lim_{n \rightarrow \infty} E(\hat{\sigma}^2) = \lim_{n \rightarrow \infty} \left[\frac{n-1}{n} \right] \sigma^2 = \sigma^2,$$

ya que, cuando $n \rightarrow \infty$, $(n-1)/n$ tiende a la unidad.

La propiedad que vamos a analizar a continuación es la *consistencia*. Como ya hemos dicho, esta propiedad se define en términos de la distribución «concentrada» de un estimador cuando $n \rightarrow \infty$. El punto en el que se concentra la distribución de un estimador se llama *límite probabilístico de $\hat{\theta}$* , y se representa a menudo, abreviado, como $\lim \text{prob } \hat{\theta}$. Podemos definirlo más formalmente del modo siguiente: sea θ^* un punto que puede ser o no igual a θ . En este caso, la afirmación

$$\lim \text{prob } \hat{\theta} = \theta^*$$

equivale a la afirmación

$$\lim_{n \rightarrow \infty} P(\theta^* - \varepsilon \leq \hat{\theta} \leq \theta^* + \varepsilon) = 1$$

en la que ε es cualquier número positivo arbitrariamente pequeño. Ahora bien, se considera que un estimador es consistente si se concentra en el punto correspondiente al verdadero valor del parámetro. Concretamente,

(6.5) $\hat{\theta}$ es un estimador consistente de θ si $\lim \text{prob } \hat{\theta} = \theta$.

Un procedimiento para descubrir si un estimador es consistente estriba en

observar la conducta del sesgo y la variancia del estimador cuando el tamaño de la muestra tiende a infinito. Si al aumentar el tamaño de la muestra se produce a la vez una disminución en el sesgo (en caso de que lo haya) y en la variancia, y si este proceso sigue hasta que tanto el primero como la segunda tienden a cero cuando $n \rightarrow \infty$, el estimador en cuestión es consistente. Esto se representó gráficamente en la figura 1-4. Como la suma de la variancia y el sesgo al cuadrado es igual a la media del error al cuadrado, la desaparición de los dos primeros cuando $n \rightarrow \infty$ equivale a la desaparición de la media del error al cuadrado. En consecuencia, podemos decir lo siguiente:

(6.6) Si $\hat{\theta}$ es un estimador de θ y $\lim_{n \rightarrow \infty} \text{MEC}(\hat{\theta}) = 0$, $\hat{\theta}$ es un estimador consistente de θ .

En general, la condición mencionada en (6.6) es una condición suficiente, pero no necesaria para la consistencia. Es decir, es posible obtener estimadores cuya media del error al cuadrado no tiende a cero cuando $n \rightarrow \infty$, y que, sin embargo, son consistentes. Puede presentarse una situación de este tipo en el caso en que la distribución asintótica del estimador es de una forma tal que su media o su variancia no existen, y esta posibilidad hace más complicado el problema de determinar si un estimador es consistente o no. Afortunadamente, los estimadores con media o variancia asintótica no existentes no son muy frecuentes.

EJEMPLO. A continuación, presentamos un ejemplo⁵ de un estimador consistente cuya media del error al cuadrado no tiende a cero cuando $n \rightarrow \infty$. Sea $\hat{\alpha}$ un estimador de α , con la siguiente distribución de probabilidad

$\hat{\alpha}$	$f(\hat{\alpha})$
α	$1 - \frac{1}{n}$
n	$\frac{1}{n}$

Es decir, $\hat{\alpha}$ sólo puede tomar los dos valores α y n . Se puede ver fácilmente que $\hat{\alpha}$ es consistente, ya que, cuando $n \rightarrow \infty$, la probabilidad de que sea igual a α tiende a la unidad. Pero

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{MEC}(\hat{\alpha}) &= \lim_{n \rightarrow \infty} E(\hat{\alpha} - \alpha)^2 \\ &= \lim_{n \rightarrow \infty} \left[(\alpha - \alpha)^2 \left(1 - \frac{1}{n}\right) + (n - \alpha)^2 \left(\frac{1}{n}\right) \right] = \infty. \end{aligned}$$

Si excluimos explícitamente de nuestra consideración este tipo de estima-

5. Este ejemplo me lo sugirió el profesor Phoebus Dhrymes.

dores y nos limitamos a los estimadores con medias y variancias asintóticas finitas, la condición (6.6) representará una condición necesaria y suficiente para la consistencia. Algunos autores llaman a este concepto, un poco más limitado, *consistencia del error al cuadrado*.⁶ Como el hecho de que la media del error al cuadrado de un estimador tiende a cero cuando $n \rightarrow \infty$ implica que el sesgo también tiende a cero, los estimadores que sean consistentes por la regla del error al cuadrado son también forzosamente insesgados asintóticamente. Lo contrario no es cierto, porque la insesgadez asintótica, por sí sola, no es suficiente para implicar consistencia del error al cuadrado.

Una característica importante de los estimadores consistentes es el hecho de que toda función continua de un estimador consistente es también un estimador consistente. El teorema que presentamos a continuación establece dicha propiedad:

Teorema 17 (teorema de Slutsky). Si $\lim \text{prob } \hat{\theta} = \theta$, y $g(\hat{\theta})$ es una función continua de $\hat{\theta}$, $\lim \text{prob } g(\hat{\theta}) = g(\theta)$.

No vamos a presentar aquí la prueba de este teorema.⁷ Esta propiedad de los estimadores consistentes, que se describe a veces diciendo que «la consistencia persiste», resulta muy conveniente y, más adelante, la utilizaremos en gran medida. Por ejemplo, esta propiedad significa que si $\hat{\theta}$ es un estimador consistente de θ , $(1/\hat{\theta})$ es un estimador consistente de $(1/\theta)$, $\log \hat{\theta}$ es un estimador consistente de $\log \theta$, etc. Téngase bien presente que, en general, esto no se aplica a la insesgadez, es decir, que esta última, al revés de la consistencia, no «persiste». Concretamente, el hecho de que $\hat{\theta}$ sea un estimador insesgado de θ no implica que $(1/\hat{\theta})$ sea un estimador insesgado de $(1/\theta)$, o que $\log \hat{\theta}$ sea un estimador insesgado de $\log \theta$, y así sucesivamente.

La última de las propiedades deseables que queremos mencionar es la *eficiencia asintótica*, que hace referencia a la dispersión de la distribución asintótica de un estimador. La eficiencia asintótica sólo está definida para los estimadores cuya media y variancia asintóticas existen (es decir, son un número finito). En realidad, esta propiedad nos proporciona un criterio de elección dentro de la familia de estimadores que satisfacen la propiedad de consistencia del error al cuadrado (y son, por tanto, asintóticamente insesgados). Como la distribución de los estimadores consistentes se concentra en el verdadero valor del parámetro cuando $n \rightarrow \infty$, debería darse prioridad a los estimadores que se aproximan a este punto de la forma más rápida posible. Estos estimadores serán aquellos cuyas distribuciones asintóticas tengan una variancia menor, ya que estas últimas representan la etapa final antes de que la dis-

tribución se concentre por completo, y los estimadores que tienen la variancia más pequeña están más cerca de este punto que los demás estimadores consistentes.

Las consideraciones anteriores nos conducen a efectuar la siguiente definición de eficiencia asintótica:

(6.7) $\hat{\theta}$ es un estimador asintóticamente eficiente de θ si cumple las condiciones siguientes:

- $\hat{\theta}$ tiene una distribución asintótica con media y variancia finitas,
- $\hat{\theta}$ es consistente.
- No existe ningún otro estimador consistente de θ cuya variancia asintótica sea inferior a la de $\hat{\theta}$.

Las dos primeras condiciones, tomadas conjuntamente, señalan que el estimador debe cumplir la consistencia del error al cuadrado para poder aspirar a ser asintóticamente eficiente. Para establecer si esta condición se satisface o no, bastará con obtener el límite (cuando $n \rightarrow \infty$) de la media del error al cuadrado. El estimador cumplirá o no dicha condición según que este límite de la media del error al cuadrado sea o no igual a cero. Establecer si un estimador consistente satisface la tercera condición de eficiencia asintótica resulta más difícil, igual que sucede cuando se trata de establecer la eficiencia en el caso de las muestras finitas. Como en aquella ocasión, el problema de la menor variancia asintótica sólo puede resolverse para los estimadores en los que conocemos la forma de la distribución de la población originaria. En el caso de estos estimadores podemos establecer la eficiencia asintótica comparando su variancia asintótica con el límite inferior de Cramer-Rao (que se definió en el teorema 16); si los dos son iguales, el estimador es asintóticamente eficiente. Así pues, tanto la eficiencia como la eficiencia asintótica se establecen en base al límite inferior de Cramer-Rao —en el caso de la eficiencia lo comparamos con la variancia ordinaria (de la muestra finita), y en el de la eficiencia asintótica, con la variancia asintótica—. Si un estimador es eficiente, lo será para *todos* los tamaños muestrales, por grandes que sean, y, por tanto, la eficiencia implica la eficiencia asintótica, pero, no obstante, lo contrario no es cierto.

Comentarios finales

Con esto damos por finalizado el estudio de las propiedades que resultan deseables en los estimadores. Podemos enumerar estas propiedades de la forma siguiente:

6. Véase, por ejemplo, Mood y Graybill. *op. cit.*, p. 175.

7. Véase S. S. Wilks, *Mathematical Statistics*. Nueva York, Wiley, 1962, pp. 102-103.

Propiedades de las muestras
finitas (pequeñas)

Insesgadez
Eficiencia
ELIO

Propiedades de las muestras
asintóticas (grandes)

Insesgadez asintótica
Consistencia
Eficiencia asintótica

Al considerar por primera vez la estimación, en la sección 1.4, comparamos la estimación de un parámetro con el tiro al blanco con fusil. El centro de la diana puede considerarse como el verdadero valor del parámetro, el fusil es el estimador, y cada disparo representa una estimación concreta (calculada a partir de una muestra determinada). La distancia hasta el blanco se relaciona inversamente con el tamaño de la muestra. En este paralelismo, las propiedades deseables de un estimador se describen en términos de las características del fusil. Un fusil insesgado es el que realiza unos disparos cuyos impactos se distribuyen de forma aleatoria alrededor del centro de la diana, mientras que los impactos de los disparos de un fusil sesgado tienen como centro algún punto distinto. Si comparamos todos los fusiles insesgados, podemos considerar como eficiente aquel cuyos impactos se hallan, en conjunto, más próximos al centro del blanco. Si sabemos la clase de proyectil que se está utilizando, podemos calcular la mínima dispersión posible de los impactos; este valor corresponde al límite inferior de Cramer-Rao. Respecto a la propiedad ELIO, la comparación se limita a los fusiles de un tipo concreto y relativamente simple. Un fusil ELIO es insesgado y produce unos impactos que se hallan más próximos al centro del blanco que los de cualquier otro fusil insesgado del mismo tipo. Si pasamos ahora a las propiedades asintóticas, debemos considerar lo que sucede cuando reducimos la distancia hasta el blanco. La insesgadez asintótica significa que los impactos tienden a distribuirse alrededor del centro del blanco, a medida que disminuye la distancia hasta él. La consistencia significa que, según va disminuyendo esta distancia, aumenta la probabilidad de efectuar un impacto en el centro del blanco o en un punto muy próximo a él. La consistencia del error al cuadrado (que implica consistencia) se puede considerar como la tendencia de los impactos a tener como centro el del blanco, y hallarse cada vez más concentrados a medida que disminuye la distancia. Por último, se puede considerar que el fusil es asintóticamente eficiente si satisface la condición de consistencia del error al cuadrado y si sus impactos están más próximos al centro del blanco que los de todos los demás fusiles consistentes, cuando la distancia hasta el blanco es muy pequeña.

El análisis de las propiedades deseables de los estimadores que hemos efectuado en esta sección es fundamental porque proporciona la base de gran parte del trabajo econométrico. Acabaremos la sección con un ejemplo acerca de cómo obtener las propiedades de tres estimadores distintos de la media de una población normal.

EJEMPLO. Sea X una variable que se distribuye normalmente con media μ y variancia σ^2 . Consideremos el problema de estimar μ a partir de una muestra aleatoria de observaciones $X_1, X_2, \dots, X_n$. Se proponen tres estimadores:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

$$\bar{\mu} = \frac{1}{n+1} \sum_{i=1}^n X_i,$$

$$\hat{\mu} = \frac{1}{2} X_1 + \frac{1}{2n} \sum_{i=2}^n X_i.$$

¿Qué propiedades deseables poseen cada uno de estos estimadores?

1. Insesgadez

$$E(\bar{X}) = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu;$$

es decir, $\bar{X}$ es insesgada.

$$E(\bar{\mu}) = E\left[\frac{1}{n+1} \sum_{i=1}^n X_i\right] = \left[\frac{n}{n+1}\right] \mu;$$

y, por tanto, $\bar{\mu}$ es sesgada.

$$\begin{aligned} E(\hat{\mu}) &= E\left[\frac{1}{2} X_1 + \frac{1}{2n} \sum_{i=2}^n X_i\right] = \frac{1}{2} E(X_1) + \frac{1}{2n} \sum_{i=2}^n E(X_i) \\ &= \frac{1}{2} \mu + \left[\frac{n-1}{2n}\right] \mu = \left[\frac{2n-1}{2n}\right] \mu; \end{aligned}$$

luego $\hat{\mu}$ es sesgada.

2. *Eficiencia.* Como $\bar{\mu}$ y $\hat{\mu}$ son sesgadas, el único estimador que tiene la posibilidad de ser eficiente es $\bar{X}$. Sabemos que la variancia de $\bar{X}$ es (σ^2/n) y que la muestra procede de una población normal. Por tanto, podemos obtener el límite inferior de Cramer-Rao correspondiente a la variancia de los estimadores insesgados de μ , y se puede demostrar que este límite inferior es también (σ^2/n) , por lo cual podemos decir que $\bar{X}$ es un estimador eficiente de μ .

3. *ELIO*. En este caso, el único estimador que puede satisfacer esta propiedad también es $\bar{X}$, ya que los otros dos son sesgados. Asimismo, $\bar{X}$ es lineal, ya que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \cdots + \frac{1}{n} X_n.$$

Por último, como $\bar{X}$ es eficiente, posee la variancia más pequeña de entre *todos* los estimadores insesgados y, por tanto, dicha variancia también tiene que ser la menor de las de los estimadores insesgados que sean lineales respecto a las observaciones. Por consiguiente, $\bar{X}$ es un estimador lineal insesgado óptimo. (En la sección 6.2 se realiza la obtención concreta de la propiedad ELIO de $\bar{X}$ independientemente de la forma de la distribución originaria.)

4. Insesgadez asintótica

$$\lim_{n \rightarrow \infty} E(\bar{X}) = \lim_{n \rightarrow \infty} \mu = \mu;$$

es decir, $\bar{X}$ es asintóticamente insesgado.

$$\lim_{n \rightarrow \infty} E(\hat{\mu}) = \lim_{n \rightarrow \infty} \left[\frac{n}{n+1} \right] \mu = \mu;$$

y, por tanto, $\hat{\mu}$ es asintóticamente insesgado.

$$\lim_{n \rightarrow \infty} E(\hat{\mu}) = \lim_{n \rightarrow \infty} \left[\frac{2n-1}{2n} \right] \mu = \mu;$$

luego, $\hat{\mu}$ es asintóticamente insesgado.

5. Consistencia

$$\text{MEC}(\bar{X}) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n},$$

$$\lim_{n \rightarrow \infty} \text{MEC}(\bar{X}) = 0;$$

es decir, $\bar{X}$ es un estimador consistente.

$$\begin{aligned} \text{MEC}(\hat{\mu}) &= \text{Var}(\hat{\mu}) + (\text{Sesgo de } \hat{\mu})^2 = \text{Var} \left[\frac{1}{n+1} \sum_{i=1}^n X_i \right] + \left[\left(\frac{n}{n+1} \right) \mu - \mu \right]^2 \\ &= \left[\frac{1}{n+1} \right]^2 \sum_{i=1}^n \text{Var}(X_i) + \left[\frac{-1}{n+1} \right]^2 \mu^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}, \end{aligned}$$

$$\lim_{n \rightarrow \infty} \text{MEC}(\hat{\mu}) = 0;$$

y, por tanto, $\hat{\mu}$ es un estimador consistente.

$$\begin{aligned} \text{MEC}(\hat{\mu}) &= \text{Var}(\hat{\mu}) + (\text{Sesgo de } \hat{\mu})^2 \\ &= \text{Var} \left[\frac{1}{2} X_1 + \frac{1}{2n} \sum_{i=2}^n X_i \right] + \left[\left(\frac{2n-1}{2n} \right) \mu - \mu \right]^2 \\ &= \frac{1}{4} \text{Var}(X_1) + \left[\frac{1}{2n} \right]^2 \sum_{i=2}^n \text{Var}(X_i) + \left[\frac{-1}{2n} \right]^2 \mu^2 \\ &= \frac{(n^2 + n - 1)\sigma^2 + \mu^2}{4n^2}, \end{aligned}$$

$$\lim_{n \rightarrow \infty} \text{MEC}(\hat{\mu}) = \frac{\sigma^2}{4}.$$

Como $\lim_{n \rightarrow \infty} \text{MEC}(\hat{\mu})$ no es igual a cero, el estimador $\hat{\mu}$ no cumple la condición de consistencia del error al cuadrado. En realidad, puede demostrarse que $\hat{\mu}$ tampoco es consistente en el sentido general, pero la prueba de ello escapa del ámbito de este libro.

6. *Eficiencia asintótica*. Los estimadores $\bar{X}$ y $\hat{\mu}$ satisfacen la condición de consistencia del error al cuadrado y, por tanto, tienen la posibilidad de ser asintóticamente eficientes, pero no en cambio $\hat{\mu}$. Como $\bar{X}$ es eficiente para *todos* los tamaños muestrales, también lo será cuando el tamaño de la muestra tienda a infinito y, por tanto, también es un estimador asintóticamente eficiente. Obsérvese que $\text{Var}(\bar{X}) = (\sigma^2/n)$, sea cual sea el tamaño de la muestra, y por ello esta última expresión es también la variancia asintótica de $\bar{X}$. Por lo que respecta a $\hat{\mu}$ podemos ver lo siguiente

$$\text{Var}(\hat{\mu}) = \frac{n\sigma^2}{(n+1)^2} = \left[\frac{n}{n+1} \right]^2 \frac{\sigma^2}{n}.$$

Cuando las muestras son grandes, el término $n(n+1)$ tendrá un valor próximo a la unidad y, por tanto, la variancia asintótica de $\hat{\mu}$ será (σ^2/n) . Como este valor es también el de la variancia asintótica de $\bar{X}$, y esta última es asintóticamente eficiente, $\hat{\mu}$ será, asimismo, asintóticamente eficiente.

En el cuadro 6-1 presentamos los anteriores resultados en forma resumida. La conclusión general es que $\bar{X}$ es el mejor estimador en las muestras pequeñas, pero, cuando las muestras son grandes, $\bar{X}$ y $\hat{\mu}$ son igualmente correctos. El tercer esti-

mador, $\bar{x}$, sólo posee una de las propiedades que resultan deseables, la insesgader asintótica.

Cuadro 6-1

Propiedades	Estimador		
	$\bar{X}$	β	μ
Muestras finitas			
Insesgadez	Sí	No	No
Eficiencia	Sí	No	No
ELIO	Sí	No	No
Propiedades asintóticas			
Insesgadez	Sí	Sí	Sí
Consistencia	Sí	Sí	No
Eficiencia	Sí	Sí	No

6.2 Métodos de estimación

Una vez definidas las propiedades que deseamos posean los estimadores, se nos presenta el problema de idear unas fórmulas de estimación que generen unas estimaciones capaces de satisfacer todas estas propiedades o algunas de ellas. En la primera parte de la sección 6.1 se definió un estimador como una fórmula de estimación que no incluye ningún parámetro desconocido. Es posible crear estimadores de muchas formas distintas. Un procedimiento consiste en acudir a algún principio que resulte más o menos plausible intuitivamente, y obtener una fórmula basada en dicho principio, comprobando a continuación si el estimador resultante posee alguna de las propiedades antes mencionadas. Los estimadores reciben nombres que indican la naturaleza del principio utilizado para obtener su fórmula. El *método de los momentos*, el *método de los mínimos cuadrados* y el *método de la máxima verosimilitud*, proporcionan todos ellos estimadores de este tipo. Otro procedimiento para obtener estimadores consiste en construir una fórmula de estimación de tal forma que incorpore en ella las propiedades que deseamos en los estimadores. El llamado *método ELIO*, que produce estimadores lineales insesgados óptimos, es el más significativo de este grupo.

Método de los momentos

El *método de los momentos* es, probablemente, el procedimiento de estimación más antiguo que se conoce en estadística. Este método está basado en el

simple principio de que los momentos de la distribución de la población deben estimarse a través de los correspondientes momentos de la muestra. De esta forma, la media de la población debe estimarse a través de la media de la muestra, la variancia de la población a través de la variancia de la muestra, etcétera. Por lo que respecta a las propiedades de los estimadores obtenidos por este procedimiento, puede demostrarse que, en condiciones muy generales, cumplen la propiedad de consistencia del error al cuadrado (y, por tanto, son generalmente consistentes) y son asintóticamente normales.⁸ Naturalmente, pueden cumplir también alguna de las demás propiedades, pero no necesariamente. Por ejemplo, la media de la muestra como estimador de la media de la población posee otras propiedades además de la de consistencia. Cuando los momentos de la población no existen, este método no es aplicable, y cuando se trata de problemas de estimación más complicados, puede ser difícil de aplicar.

Estimación por mínimos cuadrados

Otro método utilizado desde hace mucho tiempo es el *método de los mínimos cuadrados* que se puede emplear para estimar los momentos respecto al origen de la distribución de una población. El principio subyacente a este método es un poco más complejo que el correspondiente al método de los momentos. Consideremos una variable aleatoria X y su momento de orden r respecto al origen:

$$E(X^r) = \mu_r',$$

en la que $r = 0, 1, 2, \dots$. La muestra a utilizar se representa por $X_1, X_2, \dots, X_n$. Para obtener el estimador por mínimos cuadrados de μ_r' , formamos la suma siguiente:

$$\sum_{i=1}^n [X_i^r - \mu_r']^2.$$

Seleccionamos como estimador por mínimos cuadrados o *estimador minimocuadrático* el valor de μ_r' que hace lo más pequeña posible esta suma anterior. Por ejemplo, para obtener el estimador minimocuadrático de la media de la población μ ($=\mu_1'$), obtenemos el valor de μ que minimiza la suma

$$\sum_{i=1}^n [X_i - \mu]^2.$$

Obsérvese que X_i^r es la i -ésima observación de X^r y $E(X_i^r) = \mu_r'$ es la media

⁸ Véase, por ejemplo, Mood y Graybill, *op. cit.*, p. 187.

de X_i' , y, por tanto, la expresión que debemos minimizar es la suma de los cuadrados de las desviaciones de los valores observados respecto de su media. El estimador por mínimos cuadrados de esta media es el valor que hace a esta suma lo más pequeña posible.

Para obtener la fórmula de la estimación por mínimos cuadrados tenemos que resolver el problema de minimizar una función respecto a una «variable» determinada —en este caso, respecto a μ' —. Como es bien sabido, el cálculo infinitesimal elemental nos dice que la condición necesaria para la existencia de un mínimo (o un máximo) es que la primera derivada de la función que queremos minimizar sea igual a cero. Es decir, tenemos que derivar la suma de los cuadrados de los errores, respecto a μ' , igualar a cero esta derivada y despejar μ' en la ecuación que obtenemos. La solución de esta ecuación satisface la condición necesaria para que exista un mínimo o un máximo; sin embargo, puede demostrarse fácilmente que lo que obtenemos es, en realidad, un mínimo y no un máximo, y que la solución representa, pues, el estimador minimocuadrático de μ' .

Podemos ilustrar la obtención de la fórmula de estimación por mínimos cuadrados tomando el problema de la estimación de la media de la población. En este caso, la suma que debemos minimizar es

$$\sum_{i=1}^n (X_i - \mu)^2.$$

Derivando respecto a μ , obtenemos

$$\frac{d \sum_i (X_i - \mu)^2}{d\mu} = \sum_i \left[\frac{d(X_i - \mu)^2}{d\mu} \right] = \sum_i 2(X_i - \mu)(-1) = -2 \sum_i (X_i - \mu).$$

Igualando a cero esta derivada y poniendo un acento circunflejo sobre μ , para indicar que es el estimador de μ (y no su verdadero valor) el que satisface esta ecuación, obtenemos

$$-2 \sum_i (X_i - \hat{\mu}) = 0,$$

$$\left(\sum_i X_i \right) - n\hat{\mu} = 0,$$

$$\hat{\mu} = \frac{1}{n} \sum_i X_i = \bar{X}.$$

Veremos, pues, que el estimador minimocuadrático de la media de la población es la media de la muestra, al igual que el obtenido por el método de los momentos. Como otro ejemplo, tomemos el estimador por mínimos cuadrados de $E(X') = \mu'_2$. En este caso, minimizamos la suma

$$\sum_{i=1}^n (X_i^2 - \mu'_2)^2.$$

Derivando respecto a μ'_2 , tenemos

$$\frac{d \sum_i (X_i^2 - \mu'_2)^2}{d(\mu'_2)} = -2 \sum_i (X_i^2 - \mu'_2).$$

Igualando a cero esta derivada, tenemos

$$-2 \sum_i (X_i^2 - \mu'_2) = 0,$$

$$\left(\sum_i X_i^2 \right) - n\mu'_2 = 0,$$

$$\mu'_2 = \frac{1}{n} \sum_i X_i^2,$$

que es el mismo estimador de μ'_2 que obtuvimos por el método de los momentos.

Las propiedades de los estimadores minimocuadráticos tienen que establecerse en cada caso. En los dos ejemplos del párrafo anterior, estos estimadores coincidían con los estimadores por momentos y, por tanto, podemos afirmar con justificación que son consistentes. Pero esto no es necesariamente así en los modelos más complicados y, en realidad, gran parte de la econometría moderna debe su existencia al descubrimiento de que, en muchos modelos económicos, los estimadores minimocuadráticos son inconsistentes.

Estimación de la máxima verosimilitud

El tercer método de estimación es el de la *máxima verosimilitud*, que está basado en la idea, relativamente simple, de que las poblaciones distintas generan muestras distintas, y que es más probable que una muestra deter-

minada proceda de algunas poblaciones que de otras. A fin de ilustrar esta idea, consideremos el caso de las poblaciones normales y una muestra determinada de n observaciones. Las observaciones muestrales son puntos sobre el eje numérico, distribuidos alrededor de la media. Supongamos que la media muestral observada es 5. La pregunta es: ¿A qué población es más verosímil que pertenezca la muestra? En general, todas las poblaciones normales son posibles. Como las poblaciones normales quedan totalmente definidas por la media y la variancia, la única diferencia entre ellas es la correspondiente a estos dos parámetros. De momento, consideremos poblaciones que tienen la misma variancia. De entre éstas, la que tenga una media 5 generará, naturalmente, muestras con una media igual o próxima a 5 con mayor frecuencia que una población cuya media sea 6, y esta última generará dicho tipo de muestras más a menudo que una población de media 7, etc. Si consideramos las poblaciones cuya media es menor que 5 podremos efectuar afirmaciones similares.

En la figura 6-3 se representa gráficamente el argumento anterior. Los puntos $x_1, x_2, \dots, x_{10}$ representan 10 observaciones muestrales determinadas. En sentido estricto, estas observaciones pueden proceder de cualquier población normal, ya que el recorrido de las poblaciones normales se extiende desde $-\infty$ hasta $+\infty$ (en el gráfico aparecen tres de estas poblaciones). Sin embargo, si la verdadera población es A o C, la probabilidad de que las

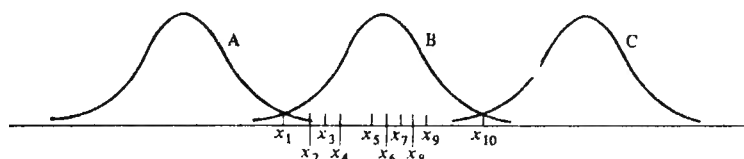


Figura 6-3

observaciones muestrales aparezcan en el intervalo indicado (es decir, entre x_1 y x_{10}) es muy pequeña. Por otra parte, si la verdadera población es B, la probabilidad de que obtengamos observaciones en este intervalo es muy grande. Por tanto, podemos llegar a la conclusión de que es más verosímil que esta muestra concreta proceda de la población B que de las poblaciones A o C.

En el ejemplo que acabamos de considerar, no hemos tenido en cuenta las poblaciones que tiene distinta variancia además de medias distintas. Esta ampliación complica un poco la explicación del principio de la verosimilitud, pero conduce a la misma conclusión. Una muestra determinada puede proceder de una población con cualquier media y cualquier variancia, pero algunas po-

drán generar una muestra de este tipo más a menudo que otras. Del mismo modo que es más verosímil que una muestra de media 5 proceda de una población de media 5 que de una población con la misma variancia, pero cuya media es 6 o 7, también una muestra con una variancia grande es más verosímil que proceda de una población cuya variancia es grande que de otra con una variancia pequeña. Lo único que se requiere es que consideremos las combinaciones de la media y la variancia específicas de la población en relación con las de la media.

Teniendo presentes estos comentarios de tipo introductorio, podemos definir ahora los *estimadores de la máxima verosimilitud*:

Si una variable aleatoria X tiene una distribución de probabilidad $f(x)$ que se caracteriza por unos parámetros $\theta_1, \theta_2, \dots, \theta_k$, y observamos una muestra cuyos valores son $x_1, x_2, \dots, x_n$, los estimadores maximoverosímiles de $\theta_1, \theta_2, \dots, \theta_k$, son los valores de estos parámetros que generarían con mayor frecuencia la muestra observada.

En otras palabras, los estimadores maximoverosímiles de $\theta_1, \theta_2, \dots, \theta_k$, son aquellos valores para los cuales la probabilidad (o densidad de probabilidad) del conjunto particular de valores muestrales es máxima. Es decir, que para obtener los estimadores maximoverosímiles de $\theta_1, \theta_2, \dots, \theta_k$, tenemos que hallar los valores que maximizan $f(x_1, x_2, \dots, x_n)$.

Tomemos un ejemplo sencillo con el fin de ilustrar el concepto de estimador maximoverosímil (EMV). Supongamos que X es una variable binaria que toma el valor 1 con una probabilidad π , y el valor 0 con una probabilidad $(1-\pi)$. Es decir,

$$f(0) = 1 - \pi,$$

$$f(1) = \pi.$$

Esto significa que la distribución de X se puede definir mediante el parámetro π , que se puede considerar como la proporción de aciertos (o la probabilidad de acertar) en la población. Supongamos una muestra aleatoria —obtenida mediante muestreo con reposición— compuesta por las tres observaciones siguientes:

$$\{1, 1, 0\}.$$

El problema consiste en obtener el EMV de π . A partir de la descripción de la población, resulta evidente que π no puede ser menor que 0 ni mayor que 1. Con

el fin de obtener la población que generaría con mayor frecuencia la muestra $\{1, 1, 0\}$, podemos considerar simplemente varios valores de π comprendidos entre 0 y 1, y calcular la probabilidad de obtener la muestra en cuestión correspondiente a estos valores. Empecemos con el valor $\pi = 0$. En este caso, en la población no hay «aciertos» y, por tanto, sería imposible que se observase dos veces el valor 1. Así pues, la probabilidad de obtener esta muestra cuando $\pi = 0$ es exactamente 0. Consideremos ahora el valor $\pi = 1/10$. En este caso, la probabilidad de obtener un 1 es $1/10$, y la de obtener un 0 es $9/10$. La probabilidad de obtener la muestra en este caso será, pues,

$$f(1, 1, 0) = f(1)f(1)f(0) = \frac{1}{10} \times \frac{1}{10} \times \frac{9}{10} = \frac{9}{1000},$$

ya que las observaciones muestrales son independientes. (Como se recordará, en el caso de que sean independientes, la probabilidad conjunta es igual al producto de las probabilidades —es decir, marginales—.) Por tanto, si $\pi = 1/10$, la probabilidad de observar esa muestra concreta es $9/1000$. De forma similar, podemos calcular la probabilidad de obtener la muestra $\{1, 1, 0\}$, correspondiente a otros valores de π . Los resultados son los siguientes:

π	$f(1, 1, 0)$
0	0
1/10	0.009
2/10	0.032
3/10	0.063
4/10	0.096
5/10	0.125
6/10	0.144
7/10	0.147
8/10	0.128
9/10	0.081
1	0.000

La función $f(1, 1, 0)$ es la *función de verosimilitud* de la muestra $\{1, 1, 0\}$. Para efectuar los cálculos, hemos seleccionado valores de π que varían en intervalos de $1/10$. Naturalmente, podríamos haber elegido intervalos más pequeños, ya que la función de verosimilitud es continua. La figura 6.4, basada en los cálculos anteriores, indica que la función de verosimilitud de esta muestra

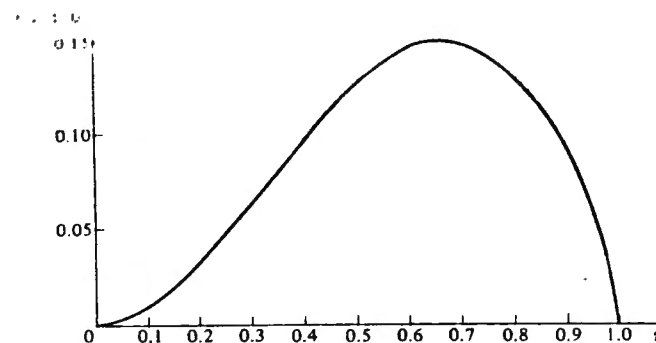


Figura 6-4

se maximiza cuando π es aproximadamente 0.7. Es decir, una población de $\pi = 0.7$ generaría muestras de la forma $\{1, 1, 0\}$ con mayor frecuencia que cualquier otra población. Por tanto, el EMV de π es 0.7 (véase también el ejemplo 1 que presentamos un poco más adelante).

El concepto de *función de verosimilitud* es básico para obtener las estimaciones maximoverosímiles y, por tanto, merece una explicación más detallada. La función de verosimilitud, representada generalmente por ℓ , es el nombre que se da a la fórmula de la distribución de probabilidad conjunta de la muestra. Como se recordará, en las secciones 3.4 y 3.5 se estudió la distribución de probabilidad conjunta de las variables aleatorias $X, Y, Z, \dots$ que se expresó como $f(x, y, z, \dots)$. En aquella ocasión se señaló que si X, Y, Z , eran independientes, podíamos escribir

$$f(x, y, z, \dots) = f(x)f(y)f(z) \dots$$

Consideremos ahora una variable aleatoria X , cuya distribución de probabilidad es $f(x)$ y se caracteriza por los parámetros $\theta_1, \theta_2, \dots, \theta_k$. Una muestra aleatoria $X_1, X_2, \dots, X_n$ representa un conjunto de n variables aleatorias independientes, cada una de las cuales tiene exactamente la misma distribución de probabilidad que X . En este caso, la función de verosimilitud ℓ se define mediante la fórmula de la distribución de probabilidad conjunta de la muestra, es decir,

$$(6.8a) \quad \ell = f(x_1, x_2, \dots, x_n).$$

Las observaciones muestrales son independientes y, por tanto, también podemos escribir

$$(6.8b) \quad \ell = f(x_1)f(x_2) \dots f(x_n).$$

Aunque la fórmula de la distribución de probabilidad conjunta de la muestra es exactamente igual que la de la función de verosimilitud, su interpretación es distinta. En el caso de la distribución de probabilidad conjunta, los parámetros $\theta_1, \theta_2, \dots, \theta_k$ se consideran fijos, y los valores de X (que representan las observaciones muestrales) son variables. En el caso de la función de verosimilitud, consideramos que los valores de los parámetros pueden variar, pero los valores de X son unos números determinados observados en una muestra concreta. Las estimaciones maximoverosímiles se obtienen maximizando la función de verosimilitud respecto a los parámetros.

La obtención de los estimadores maximoverosímiles implica, pues, especificar la función de verosimilitud y descubrir los valores de los parámetros que hacen que esta función tome su valor máximo. Como ya dijimos en relación con el método de los mínimos cuadrados, la condición necesaria para que una función tome un valor máximo (o mínimo) es que la primera derivada de la función en este punto sea igual a cero. Si la función de verosimilitud únicamente tiene un parámetro desconocido, sólo habrá una primera derivada en la que se deba cumplir esta condición, pero, sin embargo, el número de parámetros desconocidos en la función en general es mayor que uno y, en consecuencia, se ha de acudir a las derivadas parciales. En este caso, es preciso que las derivadas parciales de ℓ respecto a cada uno de los parámetros desconocidos sean igual a cero, o sea que si los parámetros desconocidos son $\theta_1, \theta_2, \dots, \theta_k$, las ecuaciones que expresan las condiciones necesarias para la existencia de un máximo (o un mínimo) son:

$$(6.9a) \quad \frac{\partial \ell}{\partial \theta_1} = 0, \quad \frac{\partial \ell}{\partial \theta_2} = 0, \quad \dots, \quad \frac{\partial \ell}{\partial \theta_k} = 0.$$

Tenemos que resolver, pues, k ecuaciones para obtener los valores de los k parámetros desconocidos. Estas ecuaciones se llaman a veces condiciones de primer grado para la existencia de un máximo (o un mínimo) y nos garantizan que para los valores de $\theta_1, \theta_2, \dots, \theta_k$ obtenidos al resolver dichas ecuaciones, encontraremos un valor máximo o un valor mínimo de ℓ . Para que podamos estar seguros de que la solución de las ecuaciones (6.9a) nos proporciona en realidad un valor máximo de ℓ , deben cumplirse determinadas condiciones de segundo grado. El estudio de estas condiciones sale del ámbito de nuestro análisis, pero no es muy difícil demostrar que se cumplen en todos los casos de los que nos vamos a ocupar.⁹ No obstante, un sencillo procedimiento para asegurarnos de

que no hemos obtenido un mínimo (en lugar de un máximo) consiste en calcular el valor de ℓ correspondiente a la solución de las ecuaciones (6.9a) y, a continuación, calcular el valor de ℓ correspondiente a unos valores ligeramente distintos de $\theta_1, \theta_2, \dots, \theta_k$. Si este segundo resultado es una cantidad menor que la primera, aquella no puede haber sido un mínimo.

Es preciso efectuar un último comentario respecto al procedimiento de maximización, relativo a la forma de las condiciones de primer grado. En la práctica, estas condiciones se suelen presentar con una forma algo distinta a como se ha hecho en las ecuaciones (6.9a). Esta formulación alternativa se basa en el hecho de que el logaritmo de ℓ es una «transformación monótonica» de ℓ . Esto significa que siempre que ℓ aumenta, su logaritmo también aumenta, y cuando ℓ disminuye, su logaritmo también disminuye. Por tanto, el punto correspondiente al valor máximo de ℓ es también el punto que corresponde al máximo valor del logaritmo de ℓ . Como ℓ es una fórmula correspondiente a una distribución de probabilidad conjunta, no puede ser negativa en ningún caso y, por tanto, no existe ningún problema para obtener su logaritmo. En la figura 6-5 se representa gráficamente la monotonización de la transformación logarítmica en el caso en que sólo existe un

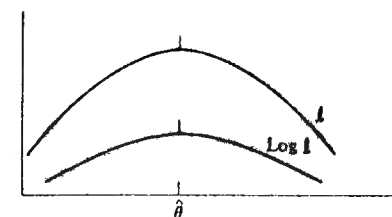


Figura 6-5

parámetro desconocido (θ). El punto $\hat{\theta}$ corresponde claramente a un máximo tanto en la función de ℓ como en la de $\log \ell$. Así pues, no tiene importancia que maximicemos una u otra de estas funciones y, como prácticamente es más fácil resolver las condiciones de primer grado cuando se trabaja con $\log \ell$, escribiremos

$$L = \log_e \ell,$$

y escribiremos de nuevo las condiciones de primer grado en la forma

$$(6.9b) \quad \frac{\partial L}{\partial \theta_1} = 0, \quad \frac{\partial L}{\partial \theta_2} = 0, \quad \dots, \quad \frac{\partial L}{\partial \theta_k} = 0.$$

Vamos a ilustrar la obtención de los EMV' con dos ejemplos. El primero de ellos se refiere al problema de estimar la proporción de «aciertos» en la población y suministra una generalización teórica de la solución gráfica presentada en la figura 6-4. El segundo ejemplo se ocupa del sencillo problema de la obtención del EMV de la media de una población normal.

9. Una descripción simple de las condiciones de segundo grado se puede encontrar, por ejemplo, en T. Yamane, *Mathematics for Economists*. Englewood Cliffs, N. J., Prentice-Hall, 1962, pp. 344-45.

EJEMPLO 1. Supongamos que X es una variable binaria que toma el valor 1 con una probabilidad π , y el valor 0 con la probabilidad $(1-\pi)$. Consideraremos que la observación de un «1» es un «acierto», y si observamos un «0» es un «fallo». Tenemos pues

$$f(0) = 1 - \pi,$$

$$f(1) = \pi.$$

Obsérvese que podemos describir la distribución de probabilidad de X como

$$f(x) = (1 - \pi)^{1-x} \pi^x.$$

Podemos comprobar que esta fórmula es correcta sustituyendo x por sus valores 0 y 1. De esta forma, obtenemos

$$f(0) = (1 - \pi)^{1-0} \pi^0 = (1 - \pi)^1 = (1 - \pi)$$

$$\text{y} \quad f(1) = (1 - \pi)^{1-1} \pi^1 = \pi^1 = \pi,$$

lo cual coincide con la especificación anterior. Nótese también que

$$E(X) = 0 \times (1 - \pi) + 1 \times \pi = \pi,$$

o sea que podemos considerar a π como la media de X . Supongamos ahora que extraemos una muestra aleatoria de n valores $\{x_1, x_2, \dots, x_n\}$. El problema consiste en hallar el EMV de π y para ello formamos la función de verosimilitud:

$$\begin{aligned} \ell &= f(x_1)f(x_2)\dots f(x_n) \\ &= [(1 - \pi)^{1-x_1}\pi^{x_1}][(1 - \pi)^{1-x_2}\pi^{x_2}]\dots [(1 - \pi)^{1-x_n}\pi^{x_n}] \\ &= (1 - \pi)^{(1-x_1)+(1-x_2)+\dots+(1-x_n)} \pi^{x_1+x_2+\dots+x_n} \\ &= (1 - \pi)^{n-\sum x_i} \pi^{\sum x_i}. \end{aligned}$$

El logaritmo de esta función es

$$L = (n - \sum x_i) \log(1 - \pi) + (\sum x_i) \log \pi.$$

El único parámetro desconocido que aparece en esta fórmula es π . Si derivamos L con respecto a π , obtendremos

$$\frac{dL}{d\pi} = (n - \sum x_i) \left[\frac{1}{1 - \pi} \right] (-1) + (\sum x_i) \left[\frac{1}{\pi} \right].$$

haciendo a cero esta derivada y colocando un «triángulo» encima de π para indicar que estamos resolviendo la ecuación con respecto a un estimador de π , obtenemos

$$\frac{(\sum x_i) - n}{1 - \hat{\pi}} + \frac{\sum x_i}{\hat{\pi}} = 0.$$

Para todos los valores de $\hat{\pi}$ distintos de 0 y 1, podemos multiplicar ambos lados de la ecuación por $\hat{\pi}(1 - \hat{\pi})$ obteniendo

$$[(\sum x_i) - n]\hat{\pi} + (\sum x_i)(1 - \hat{\pi}) = 0,$$

y despejando,

$$\hat{\pi} = \frac{1}{n} \sum x_i.$$

como $\sum x_i$ es el número de «aciertos» y n el número de observaciones de la muestra, el EMV de π es simplemente la proporción de aciertos que se ha obtenido en la muestra. En el caso específico que hemos considerado anteriormente, cuando la muestra era $\{1, 1, 0\}$, el EMV de π es

$$\hat{\pi} = \frac{2}{3} \approx 0.7.$$

EJEMPLO 2. Consideremos una variable aleatoria X que se distribuye normalmente con media μ y variancia σ^2 , es decir, $X \sim N(\mu, \sigma^2)$. Hemos observado una muestra aleatoria $\{x_1, x_2, \dots, x_n\}$. Hallar el EMV de μ . La función de densidad normal se define como

$$f(x) = (2\pi\sigma^2)^{-1/2} e^{-1/2[(x-\mu)/\sigma]^2}$$

en la que $\pi = 3.14159$. Su logaritmo es

$$\log f(x) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left[\frac{x - \mu}{\sigma} \right]^2$$

ya que $\log_e e = 1$. La función de verosimilitud es

$$\ell = f(x_1)f(x_2)\dots f(x_n).$$

y su logaritmo es

$$L = \sum_{i=1}^n \log f(x_i).$$

Sustituyendo $\log f(x_i)$ por su valor tenemos

$$L = \sum_i \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2.$$

En L aparecen ahora dos parámetros desconocidos: μ y σ^2 . Si derivamos respecto a cada uno de ellos obtenemos

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= -\frac{1}{2\sigma^2} \sum_i 2(x_i - \mu)(-1), \\ \frac{\partial L}{\partial(\sigma^2)} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (x_i - \mu)^2. \end{aligned}$$

E igualando a cero estas derivadas, se obtienen

$$(6.10a) \quad \frac{1}{\sigma^2} \sum_i (x_i - \hat{\mu}) = 0,$$

$$(6.10b) \quad -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (x_i - \hat{\mu})^2 = 0.$$

Para los valores de $\hat{\sigma}^2$ distintos de cero, la primera ecuación se reduce a

$$\sum_i (x_i - \hat{\mu}) = 0,$$

o sea,

$$\hat{\mu} = \frac{1}{n} \sum_i x_i.$$

Por tanto, el estimador maximoverosímil de la media de una población normal es igual a la media de la muestra. Debe señalarse que las dos ecuaciones (6.10a) y (6.10b) se pueden resolver también respecto a $\hat{\sigma}^2$, que es el EMV de σ^2 . Si multiplicamos la segunda ecuación por $2\hat{\sigma}^4$ obtenemos

$$-n\hat{\sigma}^2 + \sum_i (x_i - \hat{\mu})^2 = 0.$$

Y sustituyendo $\hat{\mu}$ por $\bar{x}$, y despejando $\hat{\sigma}^2$ obtendremos

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2.$$

1, decir, el estimador maximoverosímil de la variancia de una población normal es simplemente igual a la variancia de la muestra. Como ya se indicó anteriormente, esta variancia de la muestra es un estimador *sesgado* de la variancia de la población, y ya vimos en la ecuación (5.4) que

$$E(\hat{\sigma}^2) = \left[\frac{n-1}{n} \right] \sigma^2.$$

Mencionamos este hecho para señalar que el estimador maximoverosímil no es necesariamente insesgado en todos los casos.

Si conocemos la función de verosimilitud, no sólo podemos obtener los EMV de μ y σ^2 , sino también los límites inferiores de Cramer-Rao para las variancias de los estimadores insesgados de estos dos parámetros. Estos límites inferiores se pueden obtener mediante la fórmula de la matriz de información presentada en el teorema 16 (sección 6.1):

$$\begin{bmatrix} -E \frac{\partial^2 L}{\partial \mu^2} & -E \frac{\partial^2 L}{\partial \mu \partial(\sigma^2)} \\ -E \frac{\partial^2 L}{\partial \mu \partial(\sigma^2)} & -E \frac{\partial^2 L}{\partial(\sigma^2)^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

El límite inferior de Cramer-Rao para un estimador insesgado de la media de una población normal viene dado por el elemento del ángulo superior izquierdo de esta última matriz, y por tanto, es igual a (σ^2/n) .

El principio de la máxima verosimilitud está basado en una idea que resulta intuitivamente atractiva y que consiste en elegir los parámetros de los que es más probable que proceda la muestra que se ha obtenido en realidad. Sin embargo, esta atracción tiene poco valor en sí misma, a menos que los estimadores resultantes posean algunas de las propiedades que deseamos. Pero, en realidad, esto es lo que ocurre, y se puede demostrar¹⁰ que, en condiciones generales, los estimadores maximoverosímiles son

1. Consistentes respecto al error al cuadrado y, además, consistentes en sentido general.
2. Asintóticamente eficientes.

Otro de los rasgos interesantes de los EMV es que su distribución asintótica es normal, y puede obtenerse fácilmente una fórmula para calcular sus variancias asintóticas. Concretamente, las variancias asintóticas de los EMV son los

10. Ver, por ejemplo, Wilks, *op. cit.*, pp. 358-65. Las condiciones generales en las que los EMV cumplen las propiedades establecidas no son muy restrictivas. La única condición que puede causar problemas en algunos casos, y que se olvida muy a menudo, es la necesidad de que el número de parámetros en la función de verosimilitud sea finito; véase J. Malinvaud, *Statistical Methods of Econometrics* (Chicago, Rand McNally, 1966, p. 359).

elementos de la diagonal principal de la inversa de la matriz de información y, por tanto, coinciden con los límites inferiores de Cramer-Rao. En los casos de muestras finitas, tomamos como estimaciones de las variancias asintóticas los elementos de la diagonal principal de la matriz

$$\begin{bmatrix} -\frac{\partial^2 L}{\partial \theta_1^2} & -\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} & \cdots & -\frac{\partial^2 L}{\partial \theta_1 \partial \theta_k} \\ -\frac{\partial^2 L}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 L}{\partial \theta_2^2} & \cdots & -\frac{\partial^2 L}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 L}{\partial \theta_k \partial \theta_1} & -\frac{\partial^2 L}{\partial \theta_k \partial \theta_2} & \cdots & -\frac{\partial^2 L}{\partial \theta_k^2} \end{bmatrix}^{-1}$$

para unos valores de $\theta_i = \text{EMV}$ de θ_i ($i = 1, 2, \dots, k$). Todas estas propiedades son sólo asintóticas, pero a menudo esto es lo único que podemos esperar.

Estimación lineal insesgada óptima

El último método de estimación que vamos a considerar en esta parte es el *método de estimación lineal insesgada óptima*. Contrariamente a lo que ocurría en los casos anteriores, a través de este método se obtiene una fórmula de estimación que nos garantiza, por definición, determinadas propiedades que deseamos. Sea $\hat{\theta}$ un estimador lineal insesgado óptimo (ELIO) de un parámetro θ . En este caso, la fórmula de $\hat{\theta}$ debe cumplir las condiciones siguientes:

1. $\hat{\theta}$ debe ser una función lineal de las observaciones muestrales;
2. $E(\hat{\theta}) = \theta$;
3. $\text{Var}(\hat{\theta}) \leq \text{Var}(\theta^*)$, en la que θ^* es cualquier otro estimador lineal insesgado de θ .

Además, la fórmula de $\hat{\theta}$ no debe incluir a θ ni a ningún otro parámetro desconocido, ya que, de lo contrario, $\hat{\theta}$ no sería un estimador. Para construir un estimador lineal insesgado óptimo tenemos que hallar la función lineal de las observaciones muestrales que satisface las condiciones 2 y 3. Ilustraremos el procedimiento a utilizar para hallar esta función, obteniendo el ELIO de la media de la población.

Supongamos que una variable aleatoria X procede de una población con media μ y variancia σ^2 . Las observaciones muestrales son $X_1, X_2, \dots, X_n$ y queremos obtener el ELIO de μ ($\hat{\mu}$). Para ello, tomaremos cada una de las tres condiciones en forma consecutiva.

1. *Linealidad*. Como $\hat{\mu}$ debe ser una combinación lineal de las observaciones muestrales, podemos escribir

$$(6.11) \quad \hat{\mu} = \sum_{i=1}^n a_i X_i,$$

en la que $a_1, a_2, \dots, a_n$ son unas constantes a determinar. El problema de hallar el ELIO de μ se reduce, pues, al problema de especificar las constantes $a_1, a_2, \dots, a_n$ de forma tal que se satisfagan las condiciones 2 y 3.

2. *Insensgadez*. Para que $\hat{\mu}$ sea insesgada, es preciso que

$$E(\hat{\mu}) = \mu.$$

Ahora bien,

$$E(\hat{\mu}) = E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E(X_i) = \sum_{i=1}^n a_i \mu = \mu \sum_{i=1}^n a_i.$$

Es decir, para que $E(\hat{\mu})$ sea igual a μ , es necesario que $\sum a_i = 1$. La condición es, pues, que la suma de las constantes $a_1, a_2, \dots, a_n$ sea la unidad.

3. *Variancia mínima*. Por último, queremos que la variancia de $\hat{\mu}$ sea más pequeña que la de todos los demás estimadores de μ que satisfacen las condiciones anteriores. Sabemos que

$$\begin{aligned} \text{Var}(\hat{\mu}) &= \text{Var}\left[\sum_{i=1}^n a_i X_i\right] = E\left[\sum_{i=1}^n a_i X_i - E\left(\sum_{i=1}^n a_i X_i\right)\right]^2 \\ &= E\left[\sum_{i=1}^n a_i X_i - \mu \sum_{i=1}^n a_i\right]^2 = E\left[\sum_{i=1}^n a_i (X_i - \mu)\right]^2 \\ &= E\left[\sum_{i=1}^n a_i^2 (X_i - \mu)^2\right] + E\left[\sum_{i=1}^n \sum_{j=1}^n a_i a_j (X_i - \mu)(X_j - \mu)\right] \\ &= \sum_{i=1}^n a_i^2 E(X_i - \mu)^2 + \sum_{i=1}^n \sum_{j=1}^n a_i a_j E(X_i - \mu)(X_j - \mu) \quad (i \neq j). \end{aligned}$$

Pero

$$E(X_i - \mu)^2 = \text{Var}(X_i) = \sigma^2,$$

y como X_i y X_j son independientes,

$$E(X_i - \mu)(X_j - \mu) = \text{Cov}(X_i, X_j) = 0.$$

Y, por tanto,

$$(6.12) \quad \text{Var}(\hat{\mu}) = \sum_{i=1}^n a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^n a_i^2.$$

Esto significa que tenemos que hallar unos valores de $a_1, a_2, \dots, a_n$ tales que $\sum a_i = 1$ (según la condición 2) y que, a la vez, la expresión $\sigma^2 \sum a_i^2$ sea la más pequeña posible. En otras palabras, el problema consiste en minimizar la expresión $\sigma^2 \sum a_i^2$ sujeta a la condición de que $\sum a_i = 1$. Se trata, pues, de la minimización condicionada de una función, que puede efectuarse con la ayuda del método de los multiplicadores de Lagrange.¹¹

En pocas palabras, el método de los multiplicadores de Lagrange funciona de la forma siguiente. Supongamos que queremos hallar los valores de $z_1, z_2, \dots, z_m$ que minimizan (o maximizan) el valor de la función $F(z_1, z_2, \dots, z_m)$ sujeta a la condición de que $G(z_1, z_2, \dots, z_m) = 0$. La función G es la restricción, expresada de forma que todos sus términos se encuentren en el lado izquierdo de la ecuación. A continuación, se forma una nueva función, H , que se define de la forma siguiente

$$(6.13) \quad H = F(z_1, z_2, \dots, z_m) - \lambda G(z_1, z_2, \dots, z_m).$$

λ es el multiplicador de Lagrange y su valor se obtiene conjuntamente con los valores de $z_1, z_2, \dots, z_m$ que minimizan (o maximizan) F condicionada a G . Para obtener la solución deseada, se deriva H respecto de $z_1, z_2, \dots, z_m$ y λ , igualando a cero todas estas derivadas. Esto da como resultado un sistema de $(m+1)$ ecuaciones con $(m+1)$ incógnitas. Su solución proporciona las condiciones (necesarias) de primer grado; las condiciones de segundo grado, que nos dicen si la solución es un máximo o un mínimo, no las vamos a exponer aquí.¹² Una aplicación muy conocida del método de los multiplicadores de Lagrange a la economía es la correspondiente al problema de la maximización de la utilidad, condicionada a la restricción de balance.¹³

Pasemos ahora al problema concreto de minimizar $\sigma^2 \sum a_i^2$, condicionada a $\sum a_i = 1$. En este problema, la función F que queremos minimizar es

$$F(a_1, a_2, \dots, a_n) = \sigma^2 \sum a_i^2,$$

y la restricción G es

$$G(a_1, a_2, \dots, a_n) = \sum a_i - 1.$$

Seguendo el método de los multiplicadores de Lagrange, formaremos la ecuación

$$H = \sigma^2 \sum a_i^2 - \lambda \left[\sum a_i - 1 \right].$$

Las condiciones de primer grado son:

$$(6.14a) \quad \frac{\partial H}{\partial a_1} = 0, \quad \frac{\partial H}{\partial a_2} = 0, \dots, \quad \frac{\partial H}{\partial a_n} = 0, \quad \frac{\partial H}{\partial \lambda} = 0$$

o, en forma explícita,

$$(6.14b) \quad \begin{aligned} 2a_1\sigma^2 - \lambda &= 0, \\ 2a_2\sigma^2 - \lambda &= 0, \\ &\vdots \\ 2a_n\sigma^2 - \lambda &= 0, \\ -\left[\sum a_i - 1\right] &= 0. \end{aligned}$$

Tenemos pues $(n+1)$ ecuaciones con las incógnitas $a_1, a_2, \dots, a_n$ y λ . A partir de las primeras n ecuaciones obtenemos

$$a_1 = \frac{\lambda}{2\sigma^2}, \quad a_2 = \frac{\lambda}{2\sigma^2}, \quad \dots, \quad a_n = \frac{\lambda}{2\sigma^2}.$$

Y sustituyendo en la última ecuación tenemos

$$-\left[\frac{n\lambda}{2\sigma^2} - 1\right] = 0,$$

o

$$\lambda = \frac{2\sigma^2}{n}.$$

Y, por tanto,

$$a_1 = \frac{1}{n}, \quad a_2 = \frac{1}{n}, \quad \dots, \quad a_n = \frac{1}{n}.$$

11. Véase, por ejemplo, Yamane, *op. cit.*, pp. 116-20.

12. Véase, por ejemplo, *ibid.*, pp. 345-48.

13. Véase, por ejemplo, M. Friedman, *Price Theory*. Chicago, Aldine, 1962, p. 40.

Estas son, pues, las constantes que harán que $\hat{\mu}$ sea insesgada y su variancia sea mínima. Sustituyendo estos valores de $a_1, a_2, \dots, a_n$ en la fórmula de $\hat{\mu}$ de la ecuación (6.11) obtendremos

$$\hat{\mu} = \sum_i \frac{1}{n} X_i = \frac{1}{n} \sum_i X_i = \bar{X}.$$

En otras palabras, el ELIO de la media de la población es la media de la muestra. Además, si sustituimos los valores de a_i en la fórmula de la variancia de $\hat{\mu}$ que aparece en la expresión (6.12) obtenemos

$$\text{Var}(\hat{\mu}) = \sigma^2 \sum_i \left(\frac{1}{n}\right)^2 = \sigma^2 n \left(\frac{1}{n}\right)^2 = \frac{\sigma^2}{n},$$

que es la conocida fórmula de la variancia de la media de la muestra.

Los estimadores lineales insesgados óptimos pueden considerarse un caso especial de los *estimadores lineales ponderados que minimizan el error al cuadrado* (ELPMEC) a los que llamaremos θ^* y que se obtienen minimizando la expresión

$$[w \text{Var}(\theta^*) + (1 - w)(\text{Sesgo de } \theta^*)^2],$$

en la que w es un factor de ponderación predeterminado con valores $0 \leq w \leq 1$. Si hacemos $w \rightarrow 0$ obtenemos el ELIO. Muchos valores de w quedan «prohibidos» porque dan lugar a fórmulas que incluyen parámetros desconocidos. A modo de ejemplo, consideremos un ELPMEC de la media de la población cuando $w = 1/2$. Este estimador es

$$\mu^* = \sum_i a_i X_i,$$

en la que las constantes $a_1, a_2, \dots, a_n$ deben tomar unos valores que minimicen la expresión

$$\frac{1}{2} [\text{Var}(\mu^*) + (\text{Sesgo de } \mu^*)^2].$$

Si establecemos $w = 1/2$, ello implica que deseamos minimizar la media del error al cuadrado de μ^* . La fórmula resultante es

$$\mu^* = \left[\frac{n\mu^2}{n\mu^2 + \sigma^2} \right] \bar{X},$$

lo cual indica claramente que μ^* no puede considerarse como un estimador, ya que μ^2 y σ^2 no se conocen. Sin embargo, como para todos los valores de μ^2 y σ^2 distintos de cero,

$$\frac{n\mu^2}{n\mu^2 + \sigma^2} < 1,$$

podemos ver que el valor del estimador lineal de μ que minimiza la media del error al cuadrado es menor que $\bar{X}$, aunque, naturalmente, no podemos saber en qué medida.

Conclusión

Con esto damos por finalizada la consideración de los métodos de estimación a este nivel. Hemos dedicado nuestra atención a cuatro métodos básicos y en capítulos posteriores desarrollaremos métodos adicionales, pero, sin embargo, los métodos que hemos estudiado proporcionan la base de la mayor parte (o la totalidad) de los métodos restantes. Hemos ilustrado la utilidad de estos cuatro métodos aplicándolos al problema de la estimación de la media de la población, y todos ellos nos han proporcionado el mismo estimador, es decir, la media de la muestra. Este resultado es realmente tranquilizador, puesto que ya sabemos por la sección 6.1 que la media de la muestra posee todas las propiedades óptimas como estimador de la media de la población.

6.3 Intervalos de confianza

Vamos a considerar ahora el problema de la precisión de un estimador. Supongamos que estamos interesados en un parámetro θ de la población, para el que poseemos un estimador $\hat{\theta}$. Supongamos también que $\hat{\theta}$ posee todas las propiedades óptimas de un estimador e incorpora todo nuestro conocimiento respecto a la población en cuestión. Como este conocimiento no es completo, al utilizar $\hat{\theta}$ en lugar de θ estaremos cometiendo un cierto error. El problema que se plantea es cuál es el tamaño de este error y, en este contexto, hablaremos de la precisión de un estimador. Es decir, tras haber obtenido el mejor estimador que se puede construir, dado nuestro limitado conocimiento, podemos estar interesados en saber lo bien que este estimador va a funcionar.

La respuesta está obviamente relacionada con la dispersión de la distribución muestral del estimador. Si esta dispersión es pequeña, una proporción grande de las estimaciones se encontrará en un pequeño intervalo en torno al verdadero valor del parámetro; si la dispersión es grande, esta misma proporción de estimaciones se encontrará en un intervalo mayor. Por tanto, el grado de precisión de un estimador podría medirse por la desviación standard de su distribución muestral, es decir, por su error standard. En la mayor parte de los casos prácticos, no se conoce este valor, pero es posible estimarlo del error standard calculado, junto al valor del estimador. Este error standard en econometría se ha convertido en una práctica casi común la presentación del error standard calculado, junto al valor del estimador. Este error standard aparece generalmente entre paréntesis, debajo del valor del estimador, es decir,

$$\begin{array}{c} \bar{X} \\ (s_{\bar{X}}) \end{array}$$

Cuando conocemos la forma de la distribución muestral del estimador, existe un método más sistemático y explícito para indicar la precisión de dicho estimador. En este caso, es posible construir los llamados *intervalos de confianza* para el parámetro de la población. Podemos explicar con mayor claridad este concepto de intervalos de confianza haciendo referencia al análisis de la contrastación de hipótesis efectuada en la sección 5.1, y tomando como ilustración el problema de estimar la media de una población normal. En este caso, usamos como estimador la media de la muestra $\bar{X}$, que posee todas las propiedades óptimas. Sabemos que, si la población normal en cuestión tiene una media μ y una variancia σ^2 , la distribución de la media de la muestra será también una normal con media μ y variancia (σ^2/n) , es decir, $\bar{X} \sim N(\mu, \sigma^2/n)$. Por tanto,

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1),$$

en la que $N(0, 1)$ es la distribución normal tipificada, cuyas áreas se han calculado y tabulado (ver Apéndice D). Con este conocimiento, podemos efectuar ciertas afirmaciones de tipo probabilístico, que, a su vez, nos permitirán construir los intervalos de confianza para μ .

El lector recordará que, cuando la variable tiene una distribución normal tipificada, el 95 % de sus valores se halla comprendido en el intervalo -1.96 y $+1.96$. Es decir, podemos escribir

$$(6.15) \quad P(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq +1.96) = 0.95.$$

Esta afirmación implica que el 95 % de las muestras extraídas de una población normal con media μ y variancia σ^2 tendrán una media $\bar{X}$ tal que se cumplirá que

$$(6.16) \quad -1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq +1.96$$

Si multiplicamos todos los miembros de esta desigualdad por $\sqrt{\sigma^2/n}$, obtenemos

$$-1.96\sqrt{\sigma^2/n} \leq (\bar{X} - \mu) \leq +1.96\sqrt{\sigma^2/n}.$$

Restando $\bar{X}$ a cada uno los términos, obtenemos

$$-1.96\sqrt{\sigma^2/n} - \bar{X} \leq (-\mu) \leq +1.96\sqrt{\sigma^2/n} - \bar{X}.$$

Y, por último, multiplicando por -1 e invirtiendo los signos, se obtiene

$$(6.17) \quad \bar{X} - 1.96\sqrt{\sigma^2/n} \leq \mu \leq \bar{X} + 1.96\sqrt{\sigma^2/n}.$$

La expresión (6.17) recibe el nombre de *intervalo de confianza del 95 %*, para la media de la población μ . La probabilidad de que este intervalo comprenda a la verdadera media es igual a 0.95. Ello significa que, si se obtuviese un número infinito de muestras de la población que estamos considerando, y se calculase el intervalo de confianza para cada muestra, según la expresión (6.17), el 95 % de estos intervalos comprendería a la verdadera media μ . Esta medida del «95 %» representa nuestro grado de confianza de que el intervalo —construido en base a una muestra determinada— incluya a la verdadera media de la población. Obsérvese que no podemos decir que «la probabilidad de que μ esté comprendida en el intervalo considerado es 0.95», ya que μ es un valor fijo y no una variable aleatoria. La única afirmación de tipo probabilístico que se puede hacer respecto a μ es que μ tomará su verdadero valor con una probabilidad 1, y todos los demás valores con probabilidad 0. Sin embargo, los puntos extremos del intervalo —y, por tanto, el propio intervalo— son aleatorios.

Naturalmente, al establecer un intervalo de confianza, podemos elegir el nivel de confianza que deseemos; sin embargo, debemos constatar que cuando mayor sea este último, más amplio será el intervalo correspondiente y, por tanto, la información acerca de la precisión del estimador será de menor utilidad. Podemos ver este hecho considerando el caso extremo en que el nivel

de confianza es el del 100 %. En este caso, el intervalo de confianza correspondiente, obtenido a partir de un estimador que se distribuye normalmente, se extiende entre $-\infty$ y $+\infty$, lo cual, obviamente, no nos proporciona información alguna acerca de la precisión del estimador. Por otra parte, los intervalos de confianza más estrechos irán asociados a niveles de confianza más pequeños. El problema que hemos planteado aquí es muy parecido al correspondiente al nivel de significación en el contexto de la contrastación de hipótesis. En ambos casos, una solución común consiste en utilizar los niveles utilizados más a menudo por otros investigadores. En el caso de la contrastación de hipótesis se señaló la existencia de dos niveles de significación utilizados corrientemente, el 5 % y el 1 %. De forma similar, para los intervalos de confianza, los dos niveles corrientes son el 95 % y el 99 %. El intervalo de confianza del 95 % para la media de una población normal —cuando se utiliza $\bar{X}$ como estimador— se indicó en la expresión (6.17), y el intervalo de confianza del 99 % es

$$(6.18) \quad \bar{X} - 2.57\sqrt{\sigma^2/n} \leq \mu \leq \bar{X} + 2.57\sqrt{\sigma^2/n}.$$

Para construir este intervalo hemos utilizado el hecho de que el 99 % de los valores de la distribución normal tipificada se hallan entre -2.57 y $+2.57$.

El intervalo de confianza de la expresión (6.17) se ha obtenido a partir de una afirmación de carácter probabilístico acerca de la distribución normal tipificada. Concretamente, hemos utilizado los puntos límites -1.96 y $+1.96$, que contienen el 95 % del área total. Estos puntos límite no son los únicos que podemos hallar, ya que otros límites contienen también un 95 % del área, como, por ejemplo, -2.10 y $+1.85$ o -2.20 y $+1.80$, entre otros. La diferencia es que el intervalo comprendido entre -1.96 y $+1.96$ contiene la porción *central* del área, ya que delimita un 2.5 % del área en cada extremo de la distribución, mientras que los demás intervalos son asimétricos. El hecho de que el intervalo comprendido entre -1.96 y $+1.96$ sea simétrico implica que es el intervalo *más corto* de entre todos los intervalos que contienen un 95 % del área. Esto, a su vez, significa que el intervalo de confianza del 95 % resultante es menor que todos los demás intervalos con el mismo nivel de confianza. Se puede obtener la misma conclusión respecto al intervalo de confianza del 99 % obtenido en la expresión (6.18). Naturalmente, para cada nivel de confianza, es más deseable un intervalo pequeño que otro mayor.

EJEMPLO. Como ejemplo numérico, consideremos el problema siguiente. Supongamos que queremos construir un intervalo de confianza del 95 % para la media

de $X \sim N(\mu, 16)$ y que hemos tomado una muestra de 400 observaciones, obteniendo $\bar{x} = 99$. En este caso, el intervalo de confianza del 95 % es

$$99 - 1.96\sqrt{16/400} \leq \mu \leq 99 + 1.96\sqrt{16/400}$$

$$98.61 \leq \mu \leq 99.39.$$

Es decir, la probabilidad de que el intervalo comprendido entre 98.61 y 99.39 contenga a la media verdadera es 0.95. Obsérvese que, en este caso, el intervalo de confianza del 99 % es

$$99 - 2.57\sqrt{16/400} \leq \mu \leq 99 + 2.57\sqrt{16/400}$$

$$98.36 \leq \mu \leq 99.64.$$

Es posible que el lector haya observado la similitud entre los intervalos de confianza para la media de una población, y las regiones de aceptación para una contrastación respecto a la media de una población. Consideremos estas similitudes en forma explícita. Supongamos, por ejemplo, que estamos interesados en una variable $X \sim N(\mu, \sigma^2)$ y que queremos contrastar la hipótesis nula

$$H_0: \mu = \mu_0$$

contra la alternativa,

$$H_A: \mu \neq \mu_0.$$

La región de aceptación correspondiente al nivel de significación del 5 % es

$$-1.96 \leq \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}} \leq +1.96.$$

Que puede escribirse también en la forma

$$\bar{X} - 1.96\sqrt{\sigma^2/n} \leq \mu_0 \leq \bar{X} + 1.96\sqrt{\sigma^2/n}.$$

Comparemos ahora esta expresión con el intervalo de confianza del 95 % para una población de media μ , obtenido en la expresión (6.17):

$$\bar{X} - 1.96\sqrt{\sigma^2/n} \leq \mu \leq \bar{X} + 1.96\sqrt{\sigma^2/n}.$$

La conclusión es que el intervalo de confianza del 95 % es simplemente el intervalo que contiene todas las hipótesis acerca de la media de la población (es decir, las μ_0) que serían aceptadas en una contrastación de dos extremos con un nivel de significación del 5 %. Puede obtenerse un resultado similar en relación con el intervalo de confianza del 99 % y la contrastación de dos extremos con un nivel de significación del 1 %. La diferencia entre regiones de aceptación e intervalos de confianza corresponde a la diferencia entre contrastación de hipótesis y estimación: en un caso, efectuamos una afirmación relativa a la población y comprobamos si la evidencia proporcionada por la muestra la contradice o no; en el otro, la población se toma como un interrogante que tenemos que resolver a través de la muestra.

Hasta aquí hemos supuesto que, al construir los intervalos de confianza, utilizamos estimadores que poseen las propiedades óptimas. Al estudiar los intervalos de confianza para la media de una población normal, se tomó como estimador la media de la muestra, que satisface dicha condición. Vamos a ocuparnos ahora de la deseabilidad de estas propiedades óptimas para la construcción de intervalos de confianza. En primer lugar, si el estimador en cuestión estuviese sesgado y no se conociera la magnitud de este sesgo, el nivel de confianza establecido sería incorrecto. Esto se puede demostrar fácilmente sustituyendo $\bar{X}$ en las expresiones (6.17) o (6.18), por $(\bar{X} + B)$, siendo B el sesgo cuyo valor no conocemos. Es evidente que el intervalo obtenido utilizando $(\bar{X} + B)$ corresponde a una afirmación de carácter probabilístico distinta de la del intervalo en el que sólo aparece $\bar{X}$ (salvo en el caso de que $B = 0$) y, por tanto, estos dos intervalos se caracterizan por unos niveles de confianza distintos. En segundo lugar, si el estimador es insesgado, pero no es eficiente, el intervalo de confianza es más amplio que en caso de que sea eficiente. Esto deriva del hecho de que la variancia de un estimador ineficiente es mayor que la de un estimador eficiente, lo cual «empuja» hacia afuera los puntos límites del intervalo de confianza. Por último, la consistencia del error al cuadrado garantiza que, al aumentar el tamaño de la muestra, el intervalo de confianza se reduce cada vez más y que, en el límite, se concentra por completo en el punto correspondiente al verdadero valor del parámetro.

En este análisis de los intervalos de confianza, hemos utilizado a modo de ilustración el problema de construir los intervalos de confianza para la media de una población normal con variancia conocida. En la práctica, son muy pocos los casos en los que se conoce esta variancia de la población y, por tanto, es preciso estimarla a partir de la muestra. Ya hemos obtenido anteriormente un estimador insesgado de σ^2 , cuya fórmula indicada en la expresión (5.6) es

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Además, sabemos también por la expresión (5.9) que si

$$\frac{(\bar{X} - \mu)}{\sqrt{\sigma^2/n}} \sim N(0, 1),$$

se cumplirá que

$$\frac{(\bar{X} - \mu)}{\sqrt{s^2/n}} \sim t_{n-1},$$

en la que t_{n-1} representa la distribución t con $(n-1)$ grados de libertad. Esto nos permite efectuar la siguiente afirmación:

$$(6.19) \quad P\left[-t_{n-1, \alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \leq +t_{n-1, \alpha/2}\right] = 1 - \alpha,$$

en la que $t_{n-1, \alpha/2}$ indica el valor del estadístico t con $(n-1)$ grados de libertad que delimita una parte $\alpha/2$ del área de la distribución t en cada extremo. El término $(1 - \alpha)$ representa el área comprendida entre los puntos $-t_{n-1, \alpha/2}$ y $+t_{n-1, \alpha/2}$. A partir de la expresión (6.19), podemos construir un intervalo de confianza para μ al nivel de confianza que deseemos. Por ejemplo, el intervalo de confianza del 95 % es

$$(6.20) \quad \bar{X} - t_{n-1, 0.025} \sqrt{s^2/n} \leq \mu \leq \bar{X} + t_{n-1, 0.025} \sqrt{s^2/n}.$$

EJEMPLO. Como ejemplo numérico, consideremos el problema de construir el intervalo de confianza del 95 % para la media μ de $X \sim N(\mu, \sigma^2)$ sabiendo que $\bar{X} = 20$, $s^2 = 100$ y $n = 25$. En este caso, el valor de $t_{24, 0.025}$ es 2.064 y, por tanto, el intervalo de confianza del 95 % para μ es

$$20 - 2.064 \sqrt{100/25} \leq \mu \leq 20 + 2.064 \sqrt{100/25}$$

$$15.872 \leq \mu \leq 24.128.$$

De forma similar, podemos construir los intervalos correspondientes al nivel de confianza del 99 % o a cualquier otro nivel que podamos desear.

Esta idea de intervalo de confianza, que hemos desarrollado respecto a la media de una población normal, se puede utilizar de forma general en relación a cualquier parámetro para el que dispongamos de un estimador con dis-

tribución muestral conocida. Por ejemplo, podemos construir un intervalo de confianza para la variancia de una población normal, ya que sabemos que $[(n-1)s^2/\sigma^2]$ tiene una distribución ji-cuadrado con $(n-1)$ grados de libertad. En nuestra exposición hemos considerado los intervalos de confianza como un procedimiento para medir formalmente la precisión de un estimador, pero, sin embargo, el punto de vista alternativo y más tradicional consiste en considerar los intervalos de confianza como un tema más o menos separado que se estudia bajo el título de «estimación por intervalos», diferenciándolo de la «estimación por puntos» que es el tema de las secciones 6.1 y 6.2. No hemos seguido este enfoque tradicional, porque la relación entre la «estimación por puntos» y la «estimación por intervalos» es tan estrecha que la separación resulta artificial.

EJERCICIOS:

- 6-1. Sea $X \sim N(\mu, \sigma^2)$. Considérense dos muestras aleatorias independientes de observaciones de X . Los tamaños de las muestras son n_1 y n_2 , y las medias $\bar{X}_1$ y $\bar{X}_2$, respectivamente. Se proponen dos estimadores de la media de la población:

$$\hat{\mu} = \frac{1}{2} (\bar{X}_1 + \bar{X}_2).$$

$$\bar{\mu} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}.$$

Comparar las propiedades de estos estimadores.

- 6-2. Sea $X \sim N(\mu, \sigma^2)$. Considérese el siguiente estimador de la media de la población, obtenido a partir de una muestra aleatoria de n observaciones de X :

$$\bar{\mu} = \bar{X} + \frac{a}{n},$$

en la que a es una constante finita.

- ¿Cuál es la media asintótica y la variancia asintótica de $\bar{\mu}$?
- Demostrar que $\bar{\mu}$ es consistente y asintóticamente eficiente.

- 6-3. Sea X el número de aciertos en una muestra de tamaño n . Suponemos que las observaciones son independientes. Dos estimadores de la proporción de aciertos en la muestra, π , son

$$\hat{\pi} = \frac{X}{n},$$

$$\bar{\pi} = \frac{(X+1)}{(n+1)}.$$

Estudiar las propiedades de estos estimadores.

- 6-4. Considérese un dado con k caras, numeradas 1, 2, ..., k . Se lanza el dado y el número más alto que aparece es 9. En base a esta observación, obtener la estimación maximoverosimil de k , y la función de verosimilitud.
- 6-5. Sea X una variable aleatoria con media μ y variancia σ^2 . Obtener un estimador lineal de μ , $\hat{\mu}$, tal que

$$\frac{\text{Var}(\hat{\mu})}{\sigma^2} + \frac{(\text{Sesgo de } \hat{\mu})^2}{\mu^2}$$

sea mínima.

- 6-6. Sea $X \sim N(\mu, \sigma^2)$. Considérense los dos estimadores siguientes de σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Comprobar que

- $\text{Var}(\hat{\sigma}^2) < \text{Var}(s^2)$;
- $\text{MEC}(\hat{\sigma}^2) < \text{MEC}(s^2)$;
- ambos estimadores son consistentes.

(Téngase en cuenta que: $\text{Var}(\hat{\sigma}^2) = 2 \sigma^4(n-1)/n^2$.)

- 6-7. A partir de una población normal se han obtenido las 20 observaciones siguientes:

0.464	0.137	2.455	-0.323
0.060	-2.526	-0.531	-0.194
1.486	-0.354	-0.634	0.697
1.022	-0.472	1.279	3.521
1.394	-0.555	0.046	0.321

- Obtener el intervalo de confianza del 95 % para la media de la población.
- Representar gráficamente la función de verosimilitud de la media de la población, y la estimación MV de μ .

(NOTA: La estimación MV de μ no depende de σ^2 y, por tanto, para la representación gráfica podemos tomar un valor de σ^2 que sea conveniente, por ejemplo, la unidad.)

6-8. La empresa Air Conditioning Maintenance and Repair Co. considera la posibilidad de establecer una sucursal en Belair Hills. Supóngase que el lector se ha comprometido a efectuar un estudio a través de una muestra para estimar el porcentaje de familias de esta comunidad que utilizan aire acondicionado. Las condiciones del contrato entre el investigador y A.C.M.R. son las siguientes:

1. Los límites de confianza requeridos para la proporción de la población son del 95 %.
2. La muestra utilizada como base para la estimación debe ser estrictamente aleatoria y contener como mínimo 30 familias.
3. El investigador recibirá 500 dólares, menos 10 dólares por cada unidad (en términos porcentuales) del intervalo estimado de la proporción de la población; por ejemplo, si se obtiene que, con el 95 % de confianza, la proporción de la población está comprendida entre 0,4 y 0,6, se recibirá $500 - 10(60 - 40) = 300$ dólares.

En la investigación preliminar se ha descubierto que el coste de obtener información muestral es de 2 dólares por familia incluida en la muestra. Los resultados de las entrevistas efectuadas con las familias incluidas en la muestra se pueden obtener a través del «censo» de Belair Hills que se presenta en el cuadro 6-2. Hay 250 familias, cada una de las cuales ocupa una casa; las que tienen aire acondicionado son las que tienen una señal al lado. (NOTA: el intervalo de confianza del 95 % para la proporción de aciertos en una población finita, cuando la muestra es grande, es

$$\hat{\pi} \pm 1.96 \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} \sqrt{\frac{N - n}{N - 1}},$$

en la que n = tamaño de la muestra, y N = tamaño de la población.)

- a. Efectuar un informe completo dirigido a A.C.M.R., describiendo la técnica utilizada para seleccionar la muestra, y el método de estimación, presentando también los resultados.
- b. Presentar la factura para A.C.M.R.
- c. Calcular la remuneración *net*a obtenida por el trabajo realizado.
- d. Presentar un resumen de una carta dirigida a un amigo con el que se pretende formar sociedad, explicándole la forma ingeniosa con que se ha aplicado el principio de maximización del beneficio en el trabajo para la empresa A.C.M.R.

Cuadro 6-2 Censo completo de las familias residentes en Belair Hills

Fam- lia N.º	Fam- lia N.º	Fam- lia N.º	Fam- lia N.º	Fam- lia N.º	Fam- lia N.º	Fam- lia N.º	Fam- lia N.º
1 ✓	33	65	96	127	158	189	220
2	34 ✓	66 ✓	97 ✓	128 ✓	159 ✓	190	221
3 ✓	35	67	98	129 ✓	160 ✓	191	222
4 ✓	36 ✓	68	99	130 ✓	161	192 ✓	223
5	37 ✓	69 ✓	100	131	162	193	224
6 ✓	38	70 ✓	101	132	163	194	225 ✓
7	39 ✓	71 ✓	102 ✓	133	164 ✓	195 ✓	226
8	40 ✓	72	103	134	165 ✓	196	227 ✓
9 ✓	41 ✓	73	104	135	166 ✓	197 ✓	228 ✓
10	42	74 ✓	105	136	167 ✓	198	229 ✓
11 ✓	43	75	106	137 ✓	168 ✓	199 ✓	230
12 ✓	44	76	107 ✓	138	169 ✓	200	231 ✓
13 ✓	45	77 ✓	108	139	170 ✓	201	232 ✓
14	46 ✓	78 ✓	109 ✓	140	171 ✓	202	233 ✓
15 ✓	47	79 ✓	110	141	172	203	234
16	48 ✓	80	111 ✓	142	173	204 ✓	235
17	49	81	112 ✓	143	174 ✓	205	236
18	50	82	113	144	175	206 ✓	237 ✓
19 ✓	51	83 ✓	114 ✓	145 ✓	176	207 ✓	238
20 ✓	52	84	115	146	177	208	239
21	53	85 ✓	116 ✓	147 ✓	178 ✓	209 ✓	240
22	54	86 ✓	117 ✓	148 ✓	179	210	241
23 ✓	55	87	118	149	180	211	242 ✓
24 ✓	56 ✓	88	119 ✓	150 ✓	181	212	243
25 ✓	57	89	120	151	182	213	244
26 ✓	58 ✓	90	121 ✓	152	183 ✓	214	245
27 ✓	59 ✓	91	122 ✓	153 ✓	184	215 ✓	246 ✓
28 ✓	60	92 ✓	123 ✓	154	185	216 ✓	247
29 ✓	61 ✓	93 ✓	124	155 ✓	186	217 ✓	248
30 ✓	62	94 ✓	125	156	187	218	249 ✓
31	63	95 ✓	126	157	188	219	250 ✓
32 ✓	64 ✓						

SEGUNDA PARTE: TEORIA ECONOMETRICA BASICA

Capítulo

7

Regresión simple

La teoría económica se ocupa fundamentalmente de las relaciones entre variables. Las relaciones de oferta y demanda, las funciones de costes, las de producción y muchas otras son familiares a todo estudiante de economía. En realidad, todo el cuerpo de la teoría económica se puede considerar como una colección de relaciones entre variables.¹ Como ya se señaló en el capítulo 1, la econometría se ocupa de la contrastación de las proposiciones teóricas incorporadas en estas relaciones y de la estimación de los parámetros que aparecen en ellas. En esta segunda parte se estudiarán los distintos métodos que se pueden utilizar para realizar esta tarea, y los problemas que se plantean al hacerlo. En este capítulo estudiaremos el caso más sencillo, correspondiente a una relación lineal entre dos variables susceptibles de medición, y los capítulos posteriores se ocuparán de los casos más complejos.

7.1 Relaciones entre variables

Una forma apropiada para iniciar el análisis consiste en definir los nuevos conceptos con los que vamos a operar. Definimos una *relación* entre las variables X e Y como el conjunto de todos los valores de X e Y caracterizados por una ecuación determinada. Por ejemplo, si la ecuación que la caracteriza es

$$y = \alpha + \beta x,$$

en la que α y β son constantes, la relación entre X e Y es el conjunto $\{x, y\}$ formado por todos los valores posibles de X e Y que satisfacen la ecuación. Generalmente, la forma de la ecuación que la caracteriza da el nombre

¹ Ver, por ejemplo, Paul A. Samuelson, *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press, 1947.

a la relación correspondiente: una ecuación lineal describe una relación lineal, una ecuación exponencial describe una relación exponencial, etc. El concepto de relación está estrechamente unido a los de dominio y recorrido. Si una relación entre X e Y se caracteriza por la ecuación $y = f(x)$, el *dominio* de esta relación es el conjunto de todos los valores posibles de X , y el *recorrido* es el conjunto de los correspondientes valores de Y . En la práctica, las relaciones se describen generalmente por medio de la ecuación que las caracteriza, y el dominio y el recorrido no se indican explícitamente, aunque están implícitos en ella.

Las relaciones pueden ser estocásticas o determinísticas. Una relación entre X e Y es *determinística* si a cada elemento del dominio le corresponde *exactamente* un elemento del recorrido. Es decir, una relación entre X e Y caracterizada por la ecuación $y = f(x)$ es determinística si para cada valor de X sólo existe un correspondiente valor de Y . Sin embargo, las *variables* X e Y pueden ser ambas no estocásticas (es decir, pueden tomar valores totalmente controlables o susceptibles de predicción), o ambas estocásticas. Ello significa que una relación puede ser determinística (no estocástica), aunque las dos variables que aparecen en ella sean estocásticas; sin embargo, si las dos variables son estocásticas y la relación es determinística, la distribución de Y condicionada a X es degenerada (ver pág. 190). Por otra parte, se dice que una relación entre X e Y es estocástica si para cada valor de X existe una distribución de probabilidad de valores de Y . Por tanto, para cada valor determinado de X , la variable Y puede tomar, en este caso, algún valor determinado —o hallarse en el interior de cierto intervalo— con una probabilidad mayor que cero y menor que uno.

Con el fin de ilustrar esta distinción entre relaciones estocásticas y determinísticas, supongamos que llevamos a cabo una serie de experimentos para determinar la demanda de manzanas Mackintosh a distintos precios. Sea $q_t =$ cantidad de manzanas vendidas en el momento t , y $p_t =$ precio en centavos. Las manzanas se pusieron a la venta a un precio determinado en cada ocasión en que se reunió la clase durante el trimestre. Al final del trimestre, los resultados pueden ser los siguientes:

p_t	q_t
25	1
20	3
15	5
10	7
5	9
0	11

Los resultados se pueden resumir en forma de una «ecuación de demanda» del modo siguiente

$$q_t = 11 - 0.4p_t.$$

En este caso, la relación entre el precio y la cantidad es tal que *en todo momento* en que las manzanas se ofrecieron a 25 centavos cada una, sólo se vendió una de ellas. En todos los casos en que el precio fue de 20 centavos, se vendieron 3, etc. Esta relación es determinística, puesto que a cada precio le corresponde siempre una sola cantidad de manzanas vendidas. Consideremos ahora otro conjunto de resultados (cuadro 7-1). La «ecuación de demanda» debe escribirse ahora en la forma

$$q_t = 11 - 0.4p_t + \varepsilon_t,$$

Cuadro 7-1

p_t	q_t
25	0 manzanas 25 % de las veces
	1 manzana 50 % de las veces
	2 manzanas 25 % de las veces
20	2 manzanas 25 % de las veces
	3 manzanas 50 % de las veces
	4 manzanas 25 % de las veces
⋮	⋮
0	10 manzanas 25 % de las veces
	11 manzanas 50 % de las veces
	12 manzanas 25 % de las veces

en la que ε_t es una variable aleatoria con la siguiente distribución de probabilidad, sea cual sea el precio especificado:

ε_t	$f(\varepsilon_t)$
-1	0.25
0	0.50
+1	0.25
	1.00

Generalmente, esta variable recibe el nombre de *perturbación aleatoria*, debido a que «perturba» una relación que, de lo contrario, sería determinística (una expresión alternativa para ε_i es «término de error aleatorio»). Esta última relación es estocástica, ya que, debido a la presencia del término de perturbación, hay varias cantidades demandadas distintas correspondientes a cada precio, y cada cantidad tiene una probabilidad determinada. En la figura 7-1 se muestra una representación gráfica de estas dos relaciones.

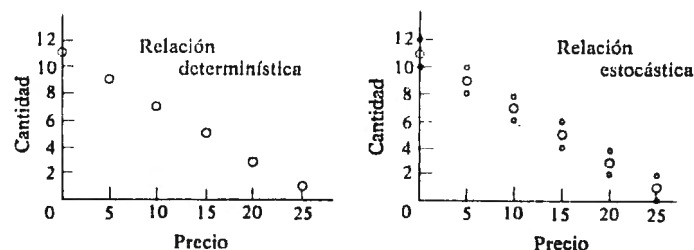


Figura 7-1

Estudiemos ahora la dependencia entre las dos variables que aparecen en una relación. En primer lugar, tomemos una relación determinística caracterizada por

$$y = f(x).$$

En este caso, si $f(x)$ no es constante para todos los valores de X (es decir, si $f(x)$ no es constante para todos los elementos del dominio), diremos que Y depende funcionalmente de X . En otras palabras, se considera que Y depende de X si un cambio en X implica un cambio en Y , al menos para algunos valores de X . Si en un diagrama de dos dimensiones medimos los valores de Y a lo largo del eje vertical, y los de X en el eje horizontal, Y dependerá de X siempre que todos los puntos no se hallen situados a lo largo de una recta horizontal. En el caso de la *relación estocástica*, decimos que Y depende funcionalmente de X si la distribución de probabilidad de Y no es la misma para todos los valores de X . Un caso típico de dependencia de Y respecto de X aparece cuando la media de Y varía al tomar X distintos valores. Sin embargo, se considera que Y depende de X , aunque la media de Y permanezca constante para todos los valores de X , siempre que varíe alguna otra caracte-

ra de la distribución de Y al variar X . Por ejemplo, si al aumentar X aumenta la variancia de Y , esto por sí solo haría que Y fuese dependiendo de X según nuestra definición.

Es interesante señalar que, en el ejemplo numérico de la curva de demanda estocástica que hemos utilizado anteriormente, la media de la cantidad demandada varía al variar el precio, mientras que la variancia permanece inalterada. Concretamente, tenemos

p_i	$E(q_i)$	$\text{Var}(q_i)$
25	1	0.5
20	3	0.5
15	5	0.5
10	7	0.5
5	9	0.5
0	11	0.5

En un caso de dependencia más general, tanto la media como la variancia de Y pueden variar en respuesta a las variaciones de X .

En la teoría económica todas las relaciones se presentan, por regla general, en forma determinística. Esto no se debe a que los economistas crean que el azar se halla totalmente ausente de las relaciones económicas, sino a que consideran que las perturbaciones estocásticas tienen una importancia menor que las influencias sistemáticas. La introducción de las perturbaciones estocásticas en las relaciones económicas complicaría en gran medida las tareas de los teóricos. Sin embargo, el énfasis en la necesidad de *contrastación* de las teorías económicas, que se encuentra a menudo en la literatura económica, implica la creencia de que existen factores estocásticos. Si las relaciones teóricas fuesen en realidad determinísticas, no se plantearía el problema de la contrastación estadística y, para obtener los valores de los parámetros desconocidos, lo único que tendríamos que hacer sería efectuar mediciones precisas y no contrastaciones. Para ilustrar esto podemos considerar una teoría que nos diga que Y depende linealmente de X . Si la relación entre Y y X fuese realmente determinística, bastaría con medir dos pares de valores de X e Y , y si la recta que une a estos dos puntos fuese horizontal, rechazaríamos la teoría mientras en todos los demás casos la teoría quedaría verificada. La ordenada en el origen y la pendiente podrían obtenerse directamente a partir del gráfico. Sin embargo, si la relación entre X e Y fuese estocástica, nuestras observaciones de los valores de las dos variables tendrían que considerarse como una muestra y ésta se utilizaría para contrastar una proposición acerca de la población, mientras que la pendiente y la ordenada en el origen tendrían que obtenerse a través de una estimación.

7.2 El modelo de regresión

La econometría se ocupa exclusivamente de las relaciones estocásticas. La forma más simple de relación estocástica entre las variables X e Y recibe el nombre de *modelo de regresión lineal simple* y se describe formalmente del modo siguiente

$$(7.1) \quad Y_i = \alpha + \beta X_i + \epsilon_i,$$

en la que Y se denomina «variable dependiente», X es la «variable explicativa», y ϵ la «perturbación estocástica», mientras que α y β son los «parámetros de regresión», desconocidos. El subíndice i indica que se trata de la observación i -ésima. Los valores de las variables X e Y son observables, pero no así los de ϵ . Las observaciones de X e Y pueden efectuarse a lo largo del tiempo, en cuyo caso decimos que se trata de «datos de series temporales», o para distintos individuos, grupos, objetos, o áreas geográficas, en cuyo caso decimos que se trata de «datos de sección o corte transversal». Por tanto, el subíndice i puede indicar el período o momento de tiempo i , o el individuo, objeto... i . Naturalmente, pueden combinarse datos de ambos tipos obteniéndose entonces «datos agrupados de series temporales y sección transversal»; por ejemplo, podemos tener datos sobre los gastos de consumo y la renta de N economías domésticas individuales a lo largo de T períodos de tiempo. En este caso, sería conveniente utilizar dos subíndices, pero los datos de este tipo no son muy frecuentes. Normalmente, las relaciones agregadas, como, por ejemplo, las funciones de consumo agregadas, las relaciones de demanda de mercado, o las funciones de producción agregadas, se estiman a partir de datos de series temporales, mientras que las microrrelaciones, como, por ejemplo, las funciones del gasto de las economías domésticas o las funciones de producción de las empresas, se estiman a partir de datos de sección transversal obtenidos mediante encuestas muestrales. El origen de los datos no se tiene en cuenta de modo explícito en la obtención de los estimadores de los parámetros de la regresión, pero, como veremos, las propiedades de estos estimadores dependen de ciertos supuestos relativos a las observaciones, y es más probable que algunos de estos supuestos no sean válidos cuando los datos son de un cierto tipo y no de otro. En este sentido, el tipo de datos utilizados en la estimación resulta importante.

La naturaleza estocástica del modelo de regresión implica que para cada valor de X existe una distribución de probabilidad completa de valores de Y . Esto significa que el valor de Y no se puede predecir en ningún caso con toda exactitud. La incertidumbre respecto a Y aparece como consecuencia del término de perturbación estocástico ϵ , el cual, al ser aleatorio, hace que Y también lo sea. Considérese, por ejemplo, la función de producción de una empresa. Supongamos que la cantidad producida depende de la cantidad de trabajo

de alguna forma específica y según el plan técnico de producción. Esta función de producción puede aplicarse a corto plazo, cuando las cantidades de los demás inputs pueden considerarse fijas, pero, en general, la misma cantidad de trabajo dará lugar a distintas cantidades producidas, debido a las variaciones en el clima, la conducta humana, la frecuencia de averías en las máquinas y otros muchos factores. La cantidad producida, que es la variable dependiente en este caso, dependerá no sólo de la cantidad de trabajo, que es la variable explicativa, sino también de una gran cantidad de causas aleatorias, que resumimos en la forma de la perturbación estocástica. La distribución de probabilidad de Y y sus características quedan pues determinados por los valores de X y la distribución de probabilidad de ϵ . Si la relación «técnica» entre trabajo y cantidad producida estuviese especificada totalmente y de forma correcta, podríamos medir el valor de ϵ a partir de las observaciones de X e Y después de cada serie de producción. En realidad, esto no sucede casi nunca y nos podemos considerar afortunados si al menos conocemos la forma matemática de la relación sin conocer sus parámetros. Normalmente, la forma matemática de la relación tiene que suponerse y los valores de los parámetros se estiman a partir de las observaciones de X e Y . Utilizando los valores estimados de los parámetros, podemos entonces «estimar» los valores de la perturbación estocástica para cada par de valores de X e Y .

Puede verse, pues, que la especificación completa del modelo de regresión no incluye solamente la forma de la ecuación de regresión tal como aparece en la expresión (7.1), sino también la especificación de la distribución de probabilidad de la perturbación, y una indicación de cómo se han obtenido los valores de la variable explicativa. Esta información viene dada por lo que llamaremos *supuestos básicos*. Estos supuestos, que se consideran de aplicación para todas las observaciones, son los siguientes:

(7.2) *Normalidad*: ϵ_i se distribuye normalmente.

(7.3) *Media cero*: $E(\epsilon_i) = 0$.

(7.4) *Homoscedasticidad*: $E(\epsilon_i^2) = \sigma^2$.

(7.5) *No autorregresión*: $E(\epsilon_i \epsilon_j) = 0$ ($i \neq j$).

(7.6) *X no estocástica*: X_i es una variable no estocástica con valores fijos en muestras repetidas y, tal que, para todo tamaño muestral,

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

es un número finito distinto de cero.

La especificación completa del modelo de regresión lineal simple consta, pues, de la ecuación de regresión (7.1) y los cinco supuestos básicos (7.2) a (7.6).² Esto representa el llamado «modelo clásico de regresión lineal normal» que constituye el punto de partida de la mayor parte del trabajo realizado en la teoría econométrica.

Examinemos ahora el significado de los distintos supuestos. Los dos primeros señalan que, para cada valor de X_i , el término de perturbación se distribuye normalmente alrededor de cero. Las implicaciones son que ϵ_i es continuo y sus valores se extienden entre $-\infty$ y $+\infty$, se distribuye simétricamente en torno a su media, y su distribución queda totalmente determinada por dos parámetros, la media y la variancia. La racionalización del supuesto de normalidad se basa en el mismo argumento que se aplica al comportamiento de los errores de medición aleatorios y que se mencionó hacia el final de la sección 4.2. Concretamente, podemos considerar cada valor de la perturbación estocástica como el resultado de un gran número de pequeñas causas, cada una de las cuales produce una pequeña desviación de la variable dependiente respecto del valor que tomaría si la relación fuese determinística. En estas circunstancias, la analogía con la conducta de los errores de medición puede ser válida y los supuestos de normalidad y media cero son apropiados. El tercer supuesto, que hace referencia a la homoscedasticidad, significa que todas las perturbaciones tienen una misma variancia σ^2 cuyo valor es desconocido. Este supuesto elimina la posibilidad de que la dispersión de las perturbaciones sea, por ejemplo, mayor para los valores grandes de X_i que para los pequeños. En términos del ejemplo de las funciones de producción, el supuesto de homoscedasticidad implica que la variación en la cantidad producida es la misma cuando la cantidad de trabajo es 20, 100, o cualquier otro número de unidades. El cuarto supuesto requiere que las perturbaciones no sean autorregresivas. Bajo este supuesto, el hecho de que la producción sea hoy mayor que la esperada no debe conducir mañana a una producción mayor (o menor) que la esperada. Obsérvese que los supuestos (7.3) y (7.5) tomados conjuntamente implican que las perturbaciones no están correlacionadas, y los supuestos (7.2), (7.3) y (7.5) tomados conjuntamente implican que las perturbaciones son independientes en sentido probabilístico.³ El último supuesto, que señala que la variable explicativa no es estocástica, resulta claro. Este supuesto nos limita

a ciertos tipos de las situaciones en que los valores de X_i son controlables o totalmente susceptibles de predicción. La afirmación adicional de que los valores de X_i es el mismo en todas las muestras. Por último, el requisito de que $(1/n) \sum (X_i - \bar{X})^2$ sea un número finito distinto de cero significa que los valores de X_i en la muestra no deben ser todos iguales y que no pueden aumentar el tamaño de la muestra.

Los supuestos subyacentes en el modelo clásico de regresión normal lineal se utilizan para la obtención de estimadores de los parámetros de regresión. Como se supone que el término de perturbación se distribuye normalmente con media cero, lo único que no se conoce acerca de su distribución es su variancia σ^2 . Por tanto, el modelo descrito por las expresiones (7.1) a (7.6) incluye en total tres parámetros desconocidos, los parámetros de regresión α y β y la variancia de la perturbación σ^2 . Sin embargo, debe ponerse de relieve que no se ignora la posibilidad de que no se cumplan uno o más de los supuestos básicos. En realidad, el capítulo 8 se dedica precisamente a este problema y en él examinaremos lo que ocurre con las propiedades de los estimadores obtenidos en este capítulo cuando dejan de cumplirse varios supuestos. También intentaremos obtener estimadores alternativos que sean apropiados para las situaciones que se presentan, cuando sea necesario.

Una vez efectuada la especificación completa del modelo de regresión tal como aparece descrito por la ecuación de regresión y los cinco supuestos básicos, podemos examinar con mayor detalle algunas de sus características básicas. Concretamente, fijémonos en la distribución de probabilidad de la variable dependiente Y_i . En primer lugar, podemos obtener la media de Y_i calculando la esperanza matemática de la ecuación (7.1), y obtenemos

$$(7.7) \quad E(Y_i) = E(\alpha + \beta X_i + \epsilon_i) = \alpha + \beta X_i.$$

Esto se deriva del hecho de que se ha especificado que α y β son parámetros, X_i es no estocástica (es decir, es un número determinado), y la media de ϵ_i es 0 según la expresión (7.3). Además, la variancia de Y_i es

$$(7.8) \quad \begin{aligned} \text{Var}(Y_i) &= E\{Y_i - E(Y_i)\}^2 = E[(\alpha + \beta X_i + \epsilon_i) - (\alpha + \beta X_i)]^2 \\ &= E(\epsilon_i^2) = \sigma^2. \end{aligned}$$

Para obtener este resultado hemos utilizado en primer lugar la definición general de variancia, sustituyendo a continuación en ella los valores de Y_i y $E(Y_i)$ procedentes de las ecuaciones (7.1) y (7.7), respectivamente, y, por último, hemos utilizado el supuesto de homoscedasticidad indicado en la expresión (7.4). Por lo que respecta a la distribución de Y_i , a partir de la ecuación (7.1) podemos ver que Y_i es simplemente una función lineal de ϵ_i . Como ésta se distribuye normalmente, Y_i se distribuirá normalmente según el teorema 12 y,

2. Estrictamente hablando, existe un supuesto adicional que se efectúa sólo de forma implícita, y es el de que no existe ningún otro modelo de regresión cuya perturbación esté correlacionada con la ϵ_i del modelo de la ecuación (7.1). Cuando nos ocupemos de las «regresiones aparentemente no relacionadas» se verá claramente la importancia de este supuesto.

3. En el caso de las variables aleatorias que se distribuyen normalmente, la no existencia de correlación implica independencia. Ver, por ejemplo, A. S. Goldberger, *Econometric Theory*, Nueva York, Wiley, 1964, pp. 107-108.

por tanto, podemos afirmar que Y_i es una variable que se distribuye normalmente con media $(\alpha + \beta X_i)$ y variancia σ^2 , es decir, que $Y_i \sim N(\alpha + \beta X_i, \sigma^2)$. En la figura 7-2 se representa gráficamente esta distribución. Obsérvese que las medias de las distribuciones se hallan todas ellas sobre una recta, y que cada distribución tiene exactamente la misma variancia.

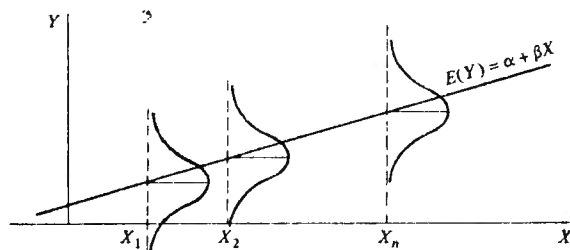


Figura 7-2

La ecuación (7.7) que da el valor medio de Y correspondiente a cada valor de X se conoce con el nombre de *recta de regresión de la población*. Su ordenada en el origen, α , indica el valor medio de Y cuando X toma el valor cero, y su pendiente, β , indica la variación en el valor medio de Y correspondiente a un cambio de una unidad en el valor de X . Si, por ejemplo, Y representa el consumo agregado y X la renta agregada, α indica el nivel de consumo de subsistencia y β la propensión marginal al consumo. Como los valores de estos parámetros son desconocidos, la recta de regresión de la población también es desconocida. Al estimar los valores de α y β obtenemos la *recta de regresión de la muestra* que sirve de estimación de la de la población. Si α y β se estiman a través de $\hat{\alpha}$ y $\hat{\beta}$, respectivamente, la recta de regresión de la muestra viene dada por

$$(7.9) \quad \hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i,$$

en la que $\hat{Y}_i$ es el valor ajustado de Y_i . La mayor parte de los valores observados de Y_i , a todos ellos, no se hallarán exactamente sobre la recta de regresión de la muestra y, por tanto, los valores de $\hat{Y}_i$ diferirán de los de Y_i . Estas diferencias reciben el nombre de *residuos*, y se representan por e_i . Por tanto, tenemos que distinguir entre las dos expresiones siguientes:

$$Y_i = \alpha + \beta X_i + e_i \quad (\text{población});$$

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i \quad (\text{muestra}).$$

Obsérvese que, en general, e_i es distinta de ϵ_i , porque $\hat{\alpha}$ y $\hat{\beta}$ difieren de los verdaderos valores de α y β . En realidad, los residuos e_i pueden considerarse como «estimaciones» de las perturbaciones ϵ_i . (De forma alternativa, podemos decir que se utiliza la distribución de e_i para estimar los parámetros de la distribución de ϵ_i .) Esto se representa gráficamente en la figura 7-3. En la sección 7.3 estudiaremos un procedimiento para estimar los parámetros de la regresión y, por tanto, la recta de regresión de la población.

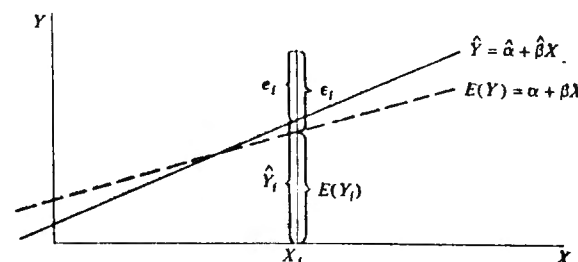


Figura 7-3

7.3 Estimación de los parámetros de la regresión

Podemos considerar el problema de la estimación de los parámetros del modelo de regresión como si se tratase de estimar los parámetros de la distribución de probabilidad de la variable dependiente Y_i . Como ya hemos indicado, bajo los supuestos efectuados en el modelo, la variable Y_i se distribuye normalmente con media $E(Y_i) = \alpha + \beta X_i$ y variancia $\text{Var}(Y_i) = \sigma^2$. El problema de estimar los parámetros de la regresión α y β equivale, pues, a la estimación de la media de Y_i , y puede resolverse a través de varios métodos distintos de estimación como se indicó en la sección 6.2. Vamos a utilizar tres de dichos métodos —mínimos cuadrados, estimación lineal insesgada óptima, y máxima verosimilitud— comparando los estimadores resultantes y sus propiedades. El objetivo consiste en obtener un estimador que tenga el mayor número posible de las propiedades que deseamos en los estimadores, con el fin de poderlo utilizar para la contrastación de hipótesis acerca del modelo de regresión y para efectuar predicciones.

Estimación por mínimos cuadrados

Vamos a empezar por la obtención de los *estimadores minimocuadráticos* de α y β . El principio de la estimación por mínimos cuadrados consiste en

minimizar la suma de los cuadrados de las desviaciones de los valores observados respecto de su media. Es decir, tenemos que hallar el valor de la media que haga lo más pequeña posible la suma mencionada. En este caso, hemos de minimizar la suma S que viene dada por

$$S = \sum_{i=1}^n [Y_i - E(Y_i)]^2,$$

$$o \quad S = \sum_{i=1}^n [Y_i - \alpha - \beta X_i]^2.$$

Para obtener los valores de α y β que minimizan esta suma tenemos que derivar S respecto de α y β . Es decir,

$$\begin{aligned} \frac{\partial S}{\partial \alpha} &= \sum_i \frac{\partial (Y_i - \alpha - \beta X_i)^2}{\partial \alpha} = \sum_i 2(Y_i - \alpha - \beta X_i)(-1) \\ &= -2 \sum_i (Y_i - \alpha - \beta X_i), \end{aligned}$$

y

$$\begin{aligned} \frac{\partial S}{\partial \beta} &= \sum_i \frac{\partial (Y_i - \alpha - \beta X_i)^2}{\partial \beta} = \sum_i 2(Y_i - \alpha - \beta X_i)(-X_i) \\ &= -2 \sum_i X_i(Y_i - \alpha - \beta X_i). \end{aligned}$$

Igualando a cero estas derivadas, y poniendo un acento circunflejo sobre α y β para indicar que las ecuaciones resultantes las cumplen los estimadores minimocuadráticos de α y β , y no sus valores verdaderos, obtenemos

$$-2 \sum_i (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0,$$

$$-2 \sum_i X_i(Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0;$$

o, de modo equivalente,

$$(7.10) \quad \sum Y_i = \hat{\alpha}n + \hat{\beta}(\sum X_i),$$

$$(7.11) \quad \sum X_i Y_i = \hat{\alpha}(\sum X_i) + \hat{\beta}(\sum X_i^2).$$

Las ecuaciones se conocen normalmente con el nombre de ecuaciones normales minimocuadráticas.⁴ Como podemos escribir

$$Y_i = \alpha + \beta X_i + e_i,$$

en la que e_i representa los «residuos minimocuadráticos», las ecuaciones normales minimocuadráticas se pueden expresar más simplemente en la forma

$$(7.10a) \quad \sum e_i = 0,$$

$$(7.11a) \quad \sum X_i e_i = 0.$$

Podemos despejar $\hat{\alpha}$ y $\hat{\beta}$ en las ecuaciones (7.10) y (7.11), obteniendo para $\hat{\beta}$ la solución

$$(7.12) \quad \hat{\beta} = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{n(\sum X_i^2) - (\sum X_i)^2}.$$

Podemos escribir esta expresión de forma algo distinta. Obsérvese que

$$\begin{aligned} (7.13) \quad n \sum (X_i - \bar{X})(Y_i - \bar{Y}) &= n(\sum X_i Y_i) - n\bar{X}(\sum Y_i) - n\bar{Y}(\sum X_i) + n^2 \bar{X} \bar{Y} \\ &= n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i) - (\sum X_i)(\sum Y_i) + (\sum X_i)(\sum Y_i) \\ &= n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i), \end{aligned}$$

que es el numerador de la expresión correspondiente a $\hat{\beta}$. Además,

$$\begin{aligned} (7.14) \quad n \sum (X_i - \bar{X})^2 &= n(\sum X_i^2) - 2n\bar{X}(\sum X_i) + n^2 \bar{X}^2 \\ &= n(\sum X_i^2) - 2(\sum X_i)^2 + (\sum X_i)^2 \\ &= n(\sum X_i^2) - (\sum X_i)^2, \end{aligned}$$

4. Para recordarlo mejor, quizá sea útil considerar la regla siguiente. En la ecuación de regresión, $Y_i = \alpha + \beta X_i + e_i$, el multiplicador de α es 1 y el multiplicador de β es X_i . La primera ecuación normal minimocuadrática se obtiene multiplicando los dos lados de la ecuación de regresión por 1, sumando las observaciones, y prescindiendo del último término correspondiente a las e_i . La segunda ecuación se obtiene multiplicando por X_i los dos lados de la ecuación de regresión, sumando todas las observaciones, y prescindiendo del último término correspondiente a las e_i . Esta regla se puede ampliar a las ecuaciones de regresión con cualquier número de variables explicativas.

que es el denominador de la expresión correspondiente a $\hat{\beta}$. Por tanto, podemos escribir

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2},$$

ya que la n del numerador se compensa con la del denominador. Podemos simplificar aún más introduciendo una nueva notación para representar las desviaciones de X_i e Y_i en cuanto a sus medias respectivas. Concretamente, podemos definir

$$x'_i = X_i - \bar{X} \quad \text{y} \quad y'_i = Y_i - \bar{Y}.$$

Naturalmente, ello implica que $\sum x'_i = 0$ y $\sum y'_i = 0$. En este caso, la expresión del estimador minimocuadrático de $\hat{\beta}$ se reduce a

$$(7.12a) \quad \hat{\beta} = \frac{\sum x'_i y'_i}{\sum x'^2_i}.$$

Una vez obtenida $\hat{\beta}$, se puede conseguir fácilmente el valor de $\hat{\alpha}$ a partir de la ecuación (7.10). Es decir,

$$(7.15) \quad \hat{\alpha} = \frac{1}{n} (\sum Y_i) - \hat{\beta} \frac{1}{n} (\sum X_i) = \bar{Y} - \hat{\beta} \bar{X},$$

lo cual significa que la recta de regresión de la muestra

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

pasa a través del punto $(\bar{X}, \bar{Y})$. El valor de $\hat{\alpha}$ indica la ordenada en el origen, y el de $\hat{\beta}$ la pendiente de la recta de regresión de la muestra. Al final de esta sección estudiaremos las propiedades muestrales de estos estimadores.

En el ejemplo numérico, consideremos los datos correspondientes a los precios y cantidades de naranjas vendidas en un supermercado durante doce

Cuadro 7-2

Precio, cent. por libra	Cantidad, libras
100	55
90	70
80	90
70	100
70	90
70	105
70	80
65	110
60	125
60	115
55	130
50	130

días consecutivos, que se presentan en el cuadro 7-2. Sea X_i el precio e Y_i la cantidad vendida en el i -ésimo día. Consideremos, asimismo, que la función de demanda es de la forma

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

y que, además, se satisfacen los supuestos básicos del modelo clásico de regresión normal. Queremos obtener la estimación minimocuadrática de α y β . Efectuando las operaciones apropiadas obtenemos

$$\bar{X} = 70, \quad \sum x'_i y'_i = -3550,$$

$$\bar{Y} = 100, \quad \sum x'^2_i = 2250.$$

(Obsérvese también que $\sum y'^2_i = 6300$, dato que utilizaremos más adelante.) Las estimaciones minimocuadráticas son

$$\hat{\beta} = \frac{-3550}{2250} = -1.578,$$

$$\hat{\alpha} = 100 - (-1.578) \times 70 = 210.460,$$

de modo que la recta de regresión estimada de la muestra es

$$\hat{Y}_i = 210.460 - 1.578 X_i.$$

Esta es la curva de demanda estimada. Como la función es lineal, la elasticidad-precio de la demanda es distinta para los distintos precios. En el punto correspondiente al precio medio (es decir, cuando $X_i = 70$), la elasticidad-precio estimada es

$$\hat{\eta}_{X_i=\bar{x}} = (-1.578) \times \frac{70}{100} = -1.105,$$

lo que indica que se estima que la demanda en este punto es ligeramente elástica.

Estimación lineal insesgada óptima

Pasemos ahora a la obtención de los *estimadores lineales insesgados óptimos* (ELIO) de α y β . Este método de estimación requiere que el estimador sea una combinación lineal de las observaciones muestrales, que sea insesgado, y que su variancia sea menor que la de todos los demás estimadores lineales insesgados. Utilizaremos este método para obtener, en primer lugar, el ELIO de β , al que llamaremos $\hat{\beta}$. Según la condición de linealidad, tenemos que

$$(7.16) \quad \hat{\beta} = \sum_i a_i Y_i,$$

en la que a_i ($i = 1, 2, \dots, n$) son ciertas constantes que debemos obtener. Ahora bien,

$$E(\hat{\beta}) = E\left(\sum a_i Y_i\right) = \sum a_i E(Y_i) = \sum a_i (\alpha + \beta X_i) = \alpha \left(\sum a_i\right) + \beta \left(\sum a_i X_i\right).$$

Ello significa que para que $\hat{\beta}$ sea insesgado es preciso que

$$\sum a_i = 0 \quad \text{y} \quad \sum a_i X_i = 1.$$

Por último, se requiere que la variancia de $\hat{\beta}$ sea menor que la de todos los demás estimadores que satisfacen las condiciones anteriores. La variancia de $\hat{\beta}$ viene dada por la expresión

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}\left(\sum a_i Y_i\right) = E\left[\left(\sum a_i Y_i\right) - E\left(\sum a_i Y_i\right)\right]^2 \\ &= E\left[\sum a_i (Y - EY)\right]^2 = E\left[\sum a_i \varepsilon_i\right]^2. \end{aligned}$$

La expresión de esta suma de cuadrados se puede desarrollar más. Consideremos, por ejemplo, el cuadrado de la suma de tres números a los que llamaremos c_1, c_2 y c_3 . Tenemos que

$$(c_1 + c_2 + c_3)^2 = c_1^2 + c_2^2 + c_3^2 + 2c_1c_2 + 2c_1c_3 + 2c_2c_3.$$

Es decir, el cuadrado de una suma es igual a la suma de los cuadrados más dos veces la suma de los productos cruzados, siendo estos últimos todos los pares posibles en los que el primer subíndice es más pequeño que el segundo. En realidad, podemos escribir

$$\left(\sum_i c_i\right)^2 = \sum_i c_i^2 + 2 \sum_{i < j} c_i c_j.$$

Este resultado, que es de interés general, se puede ampliar para cualquier número de términos. Por tanto, obtenemos

$$\begin{aligned} (7.17) \quad \text{Var}(\hat{\beta}) &= E \sum_i (a_i \varepsilon_i)^2 + 2E \sum_{i < j} (a_i \varepsilon_i)(a_j \varepsilon_j) \\ &= \sum_i a_i^2 E(\varepsilon_i^2) + 2 \sum_{i < j} a_i a_j E(\varepsilon_i \varepsilon_j) = \sigma^2 \sum_i a_i^2 \end{aligned}$$

utilizando los supuestos de homoscedasticidad y no autorregresión.

El problema consiste ahora en obtener unos valores de $a_1, a_2, \dots, a_n$ tales que $\sum a_i = 0$, $\sum a_i X_i = 1$, y, a la vez, $\sigma^2 \sum a_i^2$ sea lo más pequeña posible. Es decir, tenemos que minimizar

$$\sigma^2 \sum a_i^2$$

sujeta a

$$\sum a_i = 0 \quad \text{y} \quad \left(\sum a_i X_i\right) - 1 = 0.$$

Se trata de un problema de minimización condicionada que puede resolverse a través del *método de los multiplicadores de Lagrange*. Este método ya se describió anteriormente, para el caso en que sólo existe una restricción, al estudiar la obtención de los estimadores lineales insesgados óptimos (sección 6.2). La extensión al caso en que existen dos o más restricciones puede

hacerse por simple analogía.⁵ En el problema que nos ocupa en este caso, formamos una nueva función

$$H = \sigma^2 \sum_i a_i^2 - \lambda_1 \left(\sum_i a_i \right) - \lambda_2 \left[\left(\sum_i a_i X_i \right) - 1 \right],$$

consistente en la función a minimizar, las dos restricciones que se establecen, y los dos multiplicadores de Lagrange, λ_1 y λ_2 . Para obtener la solución que buscamos, debemos derivar H respecto de $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 , o igualar a cero cada una de estas derivadas. Es decir,

$$(7.18a) \quad \frac{\partial H}{\partial a_1} = 0, \quad \frac{\partial H}{\partial a_2} = 0, \dots, \frac{\partial H}{\partial a_n} = 0, \quad \frac{\partial H}{\partial \lambda_1} = 0, \quad \frac{\partial H}{\partial \lambda_2} = 0,$$

o, de forma explícita,

$$(7.18b) \quad \begin{aligned} 2a_1\sigma^2 - \lambda_1 - \lambda_2 X_1 &= 0 \\ 2a_2\sigma^2 - \lambda_1 - \lambda_2 X_2 &= 0 \\ &\vdots \\ 2a_n\sigma^2 - \lambda_1 - \lambda_2 X_n &= 0 \\ - \sum_i a_i &= 0 \\ - \left(\sum_i a_i X_i \right) + 1 &= 0. \end{aligned}$$

Esto da como resultado $(n+2)$ ecuaciones, a partir de las cuales deben obtenerse las incógnitas $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 . Las primeras n ecuaciones se pueden escribir nuevamente en la forma

$$(7.19) \quad \begin{aligned} a_1 &= \frac{1}{2\sigma^2} (\lambda_1 + \lambda_2 X_1) \\ a_2 &= \frac{1}{2\sigma^2} (\lambda_1 + \lambda_2 X_2) \\ &\vdots \\ a_n &= \frac{1}{2\sigma^2} (\lambda_1 + \lambda_2 X_n). \end{aligned}$$

Resumiendo estas ecuaciones, obtenemos

$$(7.20) \quad \sum_i a_i = \frac{1}{2\sigma^2} (\lambda_1 n + \lambda_2 \sum_i X_i).$$

Además, si multiplicamos la primera ecuación de (7.19) por X_1 , la segunda por X_2 , la tercera por X_3 , y así sucesivamente, y a continuación sumamos las n ecuaciones obtendremos

$$(7.21) \quad \sum_i a_i X_i = \frac{1}{2\sigma^2} \left[\lambda_1 \left(\sum_i X_i \right) + \lambda_2 \left(\sum_i X_i^2 \right) \right].$$

Sustituyendo ahora estos valores de $\sum a_i$ y $\sum a_i X_i$ según las ecuaciones (7.20) y (7.21) en las dos últimas ecuaciones de (7.18b) obtenemos

$$\begin{aligned} - \frac{1}{2\sigma^2} (\lambda_1 n + \lambda_2 \sum_i X_i) &= 0, \\ - \frac{1}{2\sigma^2} \left[\lambda_1 \left(\sum_i X_i \right) + \lambda_2 \left(\sum_i X_i^2 \right) \right] &= -1. \end{aligned}$$

Por este procedimiento hemos obtenido dos ecuaciones con dos incógnitas, λ_1 y λ_2 . El lector puede comprobar fácilmente que la solución es

$$\begin{aligned} \lambda_1 &= \frac{-2\sigma^2 \sum_i X_i}{n(\sum_i X_i^2) - (\sum_i X_i)^2}, \\ \lambda_2 &= \frac{2n\sigma^2}{n(\sum_i X_i^2) - (\sum_i X_i)^2}. \end{aligned}$$

Estas expresiones para λ_1 y λ_2 se pueden sustituir en las ecuaciones (7.19) obteniendo las soluciones correspondientes a $a_1, a_2, \dots, a_n$. Es decir,

$$(7.22) \quad a_i = \frac{-(\sum_i X_i) + n X_i}{n(\sum_i X_i^2) - (\sum_i X_i)^2} \quad (i = 1, 2, \dots, n).$$

Estas son, pues, las constantes que hacen que $\hat{\beta}$ sea un estimador insesgado y su variancia sea mínima.⁶ Sustituyendo estos valores de a_i en la fórmula de $\hat{\beta}$ de la expresión (7.16), obtenemos

$$\begin{aligned} \hat{\beta} &= \sum_i \left[\frac{-(\sum_i X_i) + n X_i}{n(\sum_i X_i^2) - (\sum_i X_i)^2} \right] Y_i = \sum_i \left[\frac{-Y_i(\sum_i X_i) + n X_i Y_i}{n(\sum_i X_i^2) - (\sum_i X_i)^2} \right] \\ &= \frac{-(\sum_i Y_i)(\sum_i X_i) + n(\sum_i X_i Y_i)}{n(\sum_i X_i^2) - (\sum_i X_i)^2}. \end{aligned}$$

5. Ver, por ejemplo, T. Yamane, *Mathematics for Economists*. Englewood Cliffs, N. J.: Prentice-Hall, 1962, pp. 119-20.

6. Puede demostrarse que también se cumplen las condiciones de segundo grado para la existencia de un mínimo. Para una ampliación, ver, por ejemplo, Yamane, *op. cit.*, pp. 345-48.

Este es exactamente el mismo resultado que hablamos obtenido para el estimador minimocuadrático de β . Según la expresión (7.12) tenemos

$$\hat{\beta} = \beta = \frac{\sum x_i y_i'}{\sum x_i'^2},$$

utilizando la notación abreviada de la expresión (7.12a).

La aplicación del principio del ELIO no sólo nos lleva a la obtención de la fórmula para el estimador en cuestión, sino que nos indica su variancia. La fórmula de la variancia de $\hat{\beta}$ indicada en la ecuación (7.17) es

$$\text{Var}(\hat{\beta}) = \sigma^2 \sum a_i^2.$$

Con el fin de eliminar el término $\sum a_i^2$, utilizamos el resultado de la ecuación (7.22). Multiplicando ambos lados de dicha ecuación por a_i y sumando todas las observaciones obtenemos

$$\sum a_i^2 = \frac{-(\sum X_i)(\sum a_i) + n(\sum a_i X_i)}{n(\sum X_i^2) - (\sum X_i)^2}.$$

Pero, a partir de las dos últimas ecuaciones de (7.18b), sabemos que

$$\sum a_i = 0 \quad \text{y} \quad \sum a_i X_i = 1.$$

Esto significa que, en realidad,

$$\sum a_i^2 = \frac{n}{n(\sum X_i^2) - (\sum X_i)^2},$$

o sea, que

$$(7.23) \quad \text{Var}(\hat{\beta}) = \frac{n\sigma^2}{n(\sum X_i^2) - (\sum X_i)^2} = \frac{\sigma^2}{\sum x_i'^2},$$

utilizando el resultado de la ecuación (7.14). Esta es, pues, la variancia del ELIO (y, también, del estimador minimocuadrático) de $\hat{\beta}$.⁷

7. Debe señalarse que, en general, $\hat{\beta}$ no es el estimador lineal de β que hace mínima la media de los errores al cuadrado. La fórmula de la combinación lineal de observaciones muestrales que proporcionan una MEC mínima es

$$\left(\frac{\beta^2}{(\sigma^2/\sum x_i'^2) + \beta^2} \right) \hat{\beta}.$$

Esta expresión no puede considerarse como un estimador (porque incluye parámetros desconocidos), pero muestra que —cuando β y σ^2 son distintos de cero— el valor del estimador lineal de β que hace mínima la MEC es menor que $\hat{\beta}$ en una cuantía desconocida ($\hat{\beta} > 0$).

Una vez obtenido el ELIO de β nos queda la tarea de obtener el de α , es decir, la ordenada en el origen de la recta de regresión. La obtención de este estimador —al que llamaremos $\hat{\alpha}$ — sigue exactamente el mismo proceso que la de $\hat{\beta}$ y, por tanto, no la vamos a presentar aquí. Como en este último caso, la obtención del ELIO de α no sólo da como resultado la fórmula del estimador, sino también la de su variancia. Los resultados son los siguientes:

$$(7.24) \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X},$$

$$(7.25) \quad \text{Var}(\hat{\alpha}) = \frac{\sigma^2(\sum X_i^2)}{n(\sum x_i'^2)} = \frac{\sigma^2(\sum x_i'^2 + n\bar{X}^2)}{n(\sum x_i'^2)} = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i'^2} \right].$$

Observando que $\hat{\beta}$ y $\hat{\beta}$ son iguales, y comparando la ecuación (7.24) con la (7.15), podemos ver que el ELIO de α es igual al estimador minimocuadrático de dicho parámetro. En la sección siguiente estudiaremos la interpretación de la fórmula de su variancia.

Estimación maximoverosímil

El último método que vamos a aplicar es el método de la máxima verosimilitud. Como ya se explicó en la sección 6.2, se consideran estimadores maximoverosímiles (EMV) de los parámetros de una población determinada aquellos valores de los parámetros que generarían con mayor frecuencia la muestra observada. Con el fin de hallar estos estimadores, tenemos que obtener la función de verosimilitud de las observaciones de la muestra y maximizarlas con respecto a los parámetros desconocidos. En el caso de nuestro modelo de regresión, la muestra está compuesta de las observaciones de las n variables $Y_1, Y_2, \dots, Y_n$. Estas variables se distribuyen normalmente con medias $(\alpha + \beta X_1), (\alpha + \beta X_2), \dots, (\alpha + \beta X_n)$ y una variancia común igual a σ^2 . Representemos estas observaciones por $y_1, y_2, \dots, y_n$. La función de verosimilitud es

$$\ell = f(y_1, y_2, \dots, y_n).$$

Vamos a probar ahora que las observaciones son independientes entre sí (en sentido probabilístico), de modo que su distribución de probabilidad conjunta se puede expresar como producto de sus distribuciones (marginales) individuales.

Podemos empezar señalando que, si las perturbaciones del modelo de regresión se distribuyen normalmente, no son autorregresivas, y tienen una media cero, tienen que ser independientes. Por tanto, el único problema consiste en mostrar que la independencia de las perturbaciones implica la independencia de las Y . Esto puede probarse utilizando el teorema siguiente.

Teorema 18. (Cambio de variable.) Si una variable aleatoria X tiene una densidad de probabilidad $f(x)$, y la variable Z es una función de X tal que entre X y Z existe una correspondencia de una a una, la densidad de probabilidad de Z es $f(z) = |dx/dz| f(x)$, siendo $dx/dz \neq 0$.

En esta expresión, $|dx/dz|$ representa la derivada de x respecto de z en valor absoluto. La demostración de este teorema puede verse en otras obras.⁸ Su importancia radica en el hecho de que, en condiciones generales, nos permite obtener la distribución de una variable a partir del conocimiento de la distribución de otra variable que esté relacionada con la primera. En el contexto del modelo de regresión que estamos considerando, la distribución que conocemos es la de ε_i , y la que queremos obtener es la de Y_i . Como sabemos que

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

existe obviamente una correspondencia de una a una entre Y_i y ε_i . Por tanto, podemos escribir

$$f(y_i) = \left| \frac{d\varepsilon_i}{dY_i} \right| f(\varepsilon_i).$$

Pero

$$\varepsilon_i = Y_i - \alpha - \beta X_i,$$

o sea que

$$\frac{d\varepsilon_i}{dY_i} = 1.$$

Y , en consecuencia,

$$f(y_i) = f(\varepsilon_i).$$

Por tanto, si las ε son independientes, también deben serlo las Y .

Tras haber obtenido este resultado, podemos presentar la función de verosimilitud en la forma

$$\ell = f(y_1)f(y_2) \dots f(y_n).$$

Como los valores de los parámetros que maximizan ℓ son los mismos que los que maximizan su logaritmo, podemos operar con $L = \log \ell$ en lugar de operar con la propia ℓ . Por tanto, queremos maximizar

$$L = \sum_{i=1}^n \log f(y_i).$$

8. Ver, por ejemplo, John E. Freund, *Mathematical Statistics*. Englewood Cliffs, N. J., Prentice-Hall, 1962, pp. 132-33.

Como cada ε_i se distribuye normalmente con media $(\alpha + \beta X_i)$ y varianza σ^2 , a partir de la fórmula de la distribución normal tenemos

$$\log f(y_i) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left[\frac{Y_i - \alpha - \beta X_i}{\sigma} \right]^2,$$

en la que $\pi = 3,14159 \dots$ (Al escribir la fórmula, utilizamos mayúsculas para representar los valores de las variables Y_i y X_i con el fin de homogeneizar la notación con la que se utiliza corrientemente para el modelo de regresión simple en otros libros de texto.) Por tanto,

$$(7.26) \quad L = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i (Y_i - \alpha - \beta X_i)^2.$$

En la función L aparecen tres parámetros desconocidos, α , β y σ^2 . Derivando la función respecto a cada uno de ellos, obtenemos

$$\frac{\partial L}{\partial \alpha} = -\frac{1}{2\sigma^2} \sum_i 2(Y_i - \alpha - \beta X_i)(-1),$$

$$\frac{\partial L}{\partial \beta} = -\frac{1}{2\sigma^2} \sum_i 2(Y_i - \alpha - \beta X_i)(-X_i),$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (Y_i - \alpha - \beta X_i)^2.$$

Igualando a cero estas derivadas y colocando un «triángulo» sobre los parámetros que queremos estimar, obtenemos

$$\frac{1}{2\hat{\sigma}^2} \sum_i (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0,$$

$$\frac{1}{2\hat{\sigma}^2} \sum_i X_i (Y_i - \hat{\alpha} - \hat{\beta} X_i) = 0,$$

$$-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_i (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 = 0.$$

Una simple manipulación de las dos primeras ecuaciones da como resultado

$$\sum Y_i = \hat{\alpha}n + \hat{\beta}(\sum X_i),$$

$$\sum X_i Y_i = \hat{\alpha}(\sum X_i) + \hat{\beta}(\sum X_i^2).$$

Estas ecuaciones son exactamente iguales que las ecuaciones normales minimocuadráticas indicadas en las ecuaciones (7.10) y (7.11). Esto significa que los *estimadores maximoverosímiles de α y β son iguales a los estimadores minimocuadráticos*. La tercera ecuación nos proporciona el estimador maximoverosímil de σ^2 , que es igual a

$$(7.27) \quad \hat{\sigma}^2 = \frac{1}{n} \sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2;$$

o, como $\hat{\alpha}$ y $\hat{\beta}$ coinciden con los estimadores minimocuadráticos,

$$(7.27a) \quad \hat{\sigma}^2 = \frac{1}{n} \sum e_i^2,$$

en la que, según la notación ya definida anteriormente, los términos e_i representan los residuos minimocuadráticos. Como $\sum e_i = 0$, a partir de la ecuación (7.27a) se deduce que el EMV de la variancia de la perturbación es igual a la variancia muestral de los residuos minimocuadráticos.

Conclusión

En resumen, vemos, pues, que los tres métodos de estimación conducen a las mismas estimaciones de los parámetros de regresión. En otras palabras, bajo los supuestos del modelo clásico de regresión normal lineal, los estimadores minimocuadráticos de los parámetros de regresión equivalen a los estimadores lineales insesgados óptimos y a los estimadores maximoverosímiles. Sin embargo, mientras que el método de los mínimos cuadrados sólo no proporciona las fórmulas de los estimadores de α y β , el método de estimación LIO nos proporciona también las fórmulas de sus variancias y el método de estimación de la máxima verosimilitud nos proporciona una fórmula para el estimador de σ^2 , resultados adicionales que son de gran utilidad.

Consideremos ahora las propiedades de los estimadores minimocuadráticos de α y β . Empezando por las propiedades de las muestras finitas, vemos inmediatamente que los estimadores minimocuadráticos son *insesgados* porque son ELIO. Podemos mostrar también que son *eficientes*. Los límites inferiores de

los errores cuadráticos dependientes a los estimadores insesgados de α y β vienen dados por los dos primeros elementos de la diagonal principal de la siguiente matriz:

$$\begin{bmatrix} -E\left(\frac{\partial^2 L}{\partial \alpha^2}\right) & -E\left(\frac{\partial^2 L}{\partial \alpha \partial \beta}\right) & -E\left(\frac{\partial^2 L}{\partial \alpha \partial \sigma^2}\right) \\ -E\left(\frac{\partial^2 L}{\partial \beta \partial \alpha}\right) & -E\left(\frac{\partial^2 L}{\partial \beta^2}\right) & -E\left(\frac{\partial^2 L}{\partial \beta \partial \sigma^2}\right) \\ -E\left(\frac{\partial^2 L}{\partial \sigma^2 \partial \alpha}\right) & -E\left(\frac{\partial^2 L}{\partial \sigma^2 \partial \beta}\right) & -E\left(\frac{\partial^2 L}{\partial (\sigma^2)^2}\right) \end{bmatrix}^{-1} = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{\sum X_i}{\sigma^2} & 0 \\ \frac{\sum X_i}{\sigma^2} & \frac{\sum X_i^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{\sigma^2 \sum X_i^2}{n(\sum X_i^2)} & \frac{-\bar{X}\sigma^2}{\sum X_i^2} & 0 \\ \frac{-\bar{X}\sigma^2}{\sum X_i^2} & \frac{\sigma^2}{\sum X_i^2} & 0 \\ 0 & 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

Si efectuamos una comparación de estos dos primeros elementos de la diagonal principal con las fórmulas (7.25) y (7.23) del texto, vemos que, en efecto, coinciden con las variancias de los parámetros de regresión.

Por último, los estimadores minimocuadráticos poseen todas las propiedades asintóticas deseables, ya que coinciden con los estimadores maximoverosímiles y, como es sabido, estos últimos son *asintóticamente insesgados, consistentes y asintóticamente eficientes*. Por tanto, los estimadores minimocuadráticos de los parámetros de regresión del modelo clásico de regresión normal lineal poseen todas las propiedades deseables, asintóticas y de las muestras finitas. En la sección 7.4 nos ocuparemos de otras características de estos estimadores y estudiaremos otros problemas de inferencia estadística en el contexto del modelo de regresión.

7.4 Otros resultados de la inferencia estadística

En la sección 7.3 se obtuvieron los estimadores minimocuadráticos de los parámetros de regresión, estableciéndose sus propiedades deseables. Ahora vamos a considerar otras características de estos estimadores y a mostrar cómo puede utilizarse el modelo de regresión para contrastar hipótesis acerca de los parámetros de regresión y para efectuar predicciones.

Distribución de $\hat{\alpha}$ y $\hat{\beta}$

A partir de los resultados que se han obtenido anteriormente resulta fácil deducir la distribución de los estimadores minimocuadráticos $\hat{\alpha}$ y $\hat{\beta}$. En primer

lugar, como dichos estimadores son insesgados, sus medias son iguales a los verdaderos valores de α y β , respectivamente. En segundo lugar, conocemos sus variancias a partir de la obtención de las propiedades ELIO. Por último, como $\hat{\alpha}$ y $\hat{\beta}$ son combinaciones lineales de las variables normales independientes $Y_1, Y_2, \dots, Y_n$, deben también distribuirse normalmente (ver teorema 12). Es decir, podemos escribir

$$(7.28) \quad \hat{\alpha} \sim N\left[\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)\right]$$

$$\hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{\sum x_i^2}\right],$$

Consideremos ahora con más detenimiento las variancias de $\hat{\alpha}$ y $\hat{\beta}$. Si nos fijamos en las fórmulas podemos ver lo siguiente:

1. Cuanto mayor sea la variancia de la perturbación (σ^2), mayores serán las variancias de $\hat{\alpha}$ y $\hat{\beta}$.
2. Cuanto mayor sea la dispersión de los valores de la variable explicativa X , menores serán las variancias de $\hat{\alpha}$ y $\hat{\beta}$.
3. Si todos los valores de X fuesen iguales, es decir, si $X_1 = X_2 = \dots = X_n$, ambas variancias serían infinitamente grandes.
4. La variancia de $\hat{\alpha}$ toma su valor mínimo cuando $\bar{X} = 0$ (siendo $\sum x_i^2 \neq 0$).

El primer punto se obvio; significa que cuanto mayor sea la dispersión de las perturbaciones en torno a la recta de regresión de la población, mayor será la dispersión de nuestras «conjeturas» acerca del valor de los parámetros de la regresión. Si todas las perturbaciones estuviesen concentradas totalmente en las medias —es decir, si todas las perturbaciones fuesen iguales a cero—, nuestras «conjeturas» respecto a los valores de los parámetros de regresión serían siempre perfectas (si observásemos al menos dos valores distintos de la variable dependiente, claro está). El segundo punto se basa en el hecho de que, cuanto mayor sea la dispersión de las X , mayor será el valor $\sum x_i^2$. En realidad, si pudiésemos elegir libremente un número determinado de valores de X en el interior de cierto intervalo —por ejemplo, entre a y b (siendo $0 < a < b$)—, la elección óptima consistiría en elegir una mitad de las X igual a a y la otra mitad igual a b . Si lo hiciésemos así maximizaríamos $\sum x_i^2$. El tercer punto se deriva del hecho de que, si todos los valores de la variables explicativa fuesen iguales, el valor de $\sum x_i^2$ sería igual a cero, y cualquier número finito dividido por cero es igual a infinito. Otra forma de explicar este mismo punto es señalar que, si todos los valores observados de Y se hallasen sobre una vertical (que es lo que ocurriría si todos ellos correspondiesen al mis-

mo valor de Y), no podríamos efectuar inferencia alguna respecto a la pendiente o a la ordenada en el origen de la recta de regresión. El último punto tiene menor importancia práctica, ya que se refiere solamente a la variancia de $\hat{\alpha}$. Si el recorrido de X incluye valores negativos además de positivos, $\text{Var}(\hat{\alpha})$ sería lo más pequeña posible si se seleccionasen los valores de forma que $\bar{X}$ fuese igual a cero. En este caso, la $\text{Var}(\hat{\alpha})$ sería igual a σ^2/n , que es el valor mínimo que puede alcanzar.

EJEMPLO. Con el fin de ilustrar el aumento de eficiencia que puede conseguirse mediante una elección adecuada de los valores de la variable explicativa, utilizaremos el ejemplo de la sección 7.3 correspondiente a la estimación de la demanda de naranjas. Los valores de X (=precio de las naranjas) eran los siguientes: 100, 90, 80, 70, 70, 70, 65, 60, 60, 55 y 50. Para estos doce valores vimos que $\bar{X} = 70$ y $\sum x_i^2 = 2250$. En este caso, las variancias de los estimadores minimocuadráticos son

$$\text{Var}(\hat{\alpha}) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right] = \sigma^2\left[\frac{1}{12} + \frac{70^2}{2250}\right] = 2.261111\sigma^2.$$

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2}{2250} = 0.000444\sigma^2.$$

Supongamos ahora que en lugar de los valores anteriores hubiésemos tenido $X_1 = X_2 = \dots = X_6 = 100$ y $X_7 = X_8 = \dots = X_{12} = 50$. En este caso, tendríamos $\bar{X} = 75$ y $\sum x_i^2 = 7500$. Las variancias resultantes serían, ahora,

$$\text{Var}(\hat{\alpha}) = \sigma^2\left[\frac{1}{12} + \frac{75^2}{7500}\right] = 0.833333\sigma^2.$$

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{7500} = 0.000133.$$

Comparando las variancias correspondientes a los dos casos obtenemos

$$\frac{\text{Var}(\hat{\alpha}) \text{ caso I}}{\text{Var}(\hat{\alpha}) \text{ caso II}} = \frac{2.261111 \sigma^2}{0.833333 \sigma^2} = 2.713.$$

$$\frac{\text{Var}(\hat{\beta}) \text{ caso I}}{\text{Var}(\hat{\beta}) \text{ caso II}} = \frac{0.000444 \sigma^2}{0.000133 \sigma^2} = 3.338.$$

Es decir, la variancia de $\hat{\alpha}$ en el primer caso es más de dos veces y media mayor que la del segundo caso, y la de $\hat{\beta}$ más de 3 1/3 veces mayor.

Es evidente que el aumento de eficiencia a consecuencia de la elección óptima de los valores de X resulta considerable. Naturalmente, en la práctica

la dificultad consiste en que, normalmente, el economista no dispone de la posibilidad de elección en este aspecto, ya que el muestreo lo ha efectuado una tercera persona y él recibe solamente los resultados muestrales obtenidos.⁹

Covariancia de $\hat{\alpha}$ y $\hat{\beta}$

Un problema de cierto interés es el correspondiente a la relación entre $\hat{\alpha}$ y $\hat{\beta}$. Al utilizar $\hat{\alpha}$ en lugar de α y $\hat{\beta}$ en lugar de β , estamos cometiendo errores muestrales, y es de cierta importancia saber si estos dos errores muestrales tendrán el mismo signo o no. Es decir, queremos conocer el signo de

$$E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta),$$

que, por definición, es la covariancia entre $\hat{\alpha}$ y $\hat{\beta}$. Ahora bien, según la ecuación (7.15) tenemos

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

Como se ha señalado anteriormente, el modelo de regresión es

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Sumando todas las observaciones muestrales y dividiendo por n , obtenemos

$$(7.29) \quad \bar{Y} = \alpha + \beta\bar{X} + \bar{\varepsilon} \quad \text{or} \quad \alpha = \bar{Y} - \beta\bar{X} - \bar{\varepsilon},$$

o sea, que

$$(7.30) \quad \hat{\alpha} - \alpha = (\bar{Y} - \hat{\beta}\bar{X}) - (\bar{Y} - \beta\bar{X} - \bar{\varepsilon}) = -(\hat{\beta} - \beta)\bar{X} + \bar{\varepsilon}.$$

Además, según la ecuación (7.12a) tenemos

$$\hat{\beta} = \frac{\sum x'_i y'_i}{\sum x_i'^2},$$

pero, restando la ecuación (7.29) de la (7.1), obtenemos

$$(7.31) \quad (Y_i - \bar{Y}) = \beta(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon}),$$

9. Sin embargo, debe señalarse que la optimalidad del procedimiento de muestreo que «acumula» los valores de X en los dos extremos del recorrido depende de forma crucial de la linealidad del modelo. Este diseño de las muestras sería muy poco adecuado para los modelos en los que, en lugar de efectuar el supuesto de linealidad, se intentase contrastarla.

En el texto se utiliza la notación abreviada para indicar desviaciones respecto de las medias muestrales.

$$(7.11a) \quad y_i = \beta x_i + \varepsilon_i.$$

Sustituyendo este valor en la fórmula de $\hat{\beta}$ obtenemos

$$\hat{\beta} = \frac{\sum x'_i(\beta x'_i + \varepsilon'_i)}{\sum x_i'^2} = \beta + \frac{\sum x'_i \varepsilon'_i}{\sum x_i'^2},$$

o sea, que

$$(7.32) \quad \begin{aligned} \hat{\beta} - \beta &= \frac{\sum x'_i \varepsilon'_i}{\sum x_i'^2} = \frac{\sum x'_i(\varepsilon_i - \bar{\varepsilon})}{\sum x_i'^2} \\ &= \frac{\sum x'_i \varepsilon_i - \bar{\varepsilon} \sum x'_i}{\sum x_i'^2} = \frac{\sum x'_i \varepsilon_i}{\sum x_i'^2}, \end{aligned}$$

que es el error muestral de $\hat{\beta}$. Así pues, combinando las ecuaciones (7.30) y (7.32) obtenemos

$$\begin{aligned} E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta) &= E[-(\hat{\beta} - \beta)\bar{X} + \bar{\varepsilon}](\hat{\beta} - \beta) \\ &= -\bar{X}E(\hat{\beta} - \beta)^2 + E\bar{\varepsilon} \left[\frac{\sum x'_i \varepsilon_i}{\sum x_i'^2} \right]. \end{aligned}$$

Consideremos el último término de esta expresión:

$$\begin{aligned} E\bar{\varepsilon} \left[\frac{\sum x'_i \varepsilon_i}{\sum x_i'^2} \right] &= E \left[\frac{\sum x'_i \varepsilon_i \bar{\varepsilon}}{\sum x_i'^2} \right] = \frac{\sum x'_i E\varepsilon_i (1/n)(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i + \dots + \varepsilon_n)}{\sum x_i'^2} \\ &= \frac{(1/n) \sum x'_i (E\varepsilon_i \varepsilon_1 + E\varepsilon_i \varepsilon_2 + \dots + E\varepsilon_i^2 + \dots + E\varepsilon_i \varepsilon_n)}{\sum x_i'^2} \\ &= \frac{(1/n) \sum x'_i (0 + 0 + \dots + \sigma^2 + \dots + 0)}{\sum x_i'^2} = \frac{(1/n)\sigma^2 \sum x'_i}{\sum x_i'^2} = 0 \end{aligned}$$

ya que $\sum x'_i = 0$.

$$(7.33) \quad E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta) = -\bar{X}E(\hat{\beta} - \beta)^2 = -\bar{X}\text{Var}(\hat{\beta}) = -\bar{X} \left[\frac{\sigma^2}{\sum x_i'^2} \right],$$

según la ecuación (7.23). Esta es, pues, la covariancia entre $\hat{\alpha}$ y $\hat{\beta}$. A partir de este resultado podemos ver que, *siempre que $\bar{X}$ sea positiva*, puede esperarse que los errores muestrales de $\hat{\alpha}$ y $\hat{\beta}$ tengan signos opuestos. En este caso, podemos esperar que una estimación por exceso del verdadero valor de α vaya acompañada de una estimación por defecto del verdadero valor de β , y viceversa.

Método de las semi-medias

Es interesante comparar la recta de regresión muestral ajustada por el método de los mínimos cuadrados con la ajustada por algún otro método. Uno de los métodos que se ha utilizado en aplicaciones prácticas es el llamado *método de las semi-medias*. Este método consiste en disponer las observaciones de forma que los valores de X se hallen colocados en orden creciente, dividiéndolos en dos partes iguales (o aproximadamente iguales), y calculando los valores medios de X e Y en cada parte por separado. Esto nos proporciona dos puntos, uno para cada parte, y la recta de regresión estimada es la que pasa por estos dos puntos. Para mayor simplicidad, supongamos que el número de observaciones es par y definamos $\bar{X}_A$ e $\bar{Y}_A$ como las medias muestrales de los primeros $n/2$ valores de X e Y , y $\bar{X}_B$ e $\bar{Y}_B$ las medias muestrales de los últimos $n/2$ valores de X e Y . Además, sean a y b los estimadores «semi-medios» de α y β , respectivamente. Los valores de a y b pueden calcularse a partir de

$$\bar{Y}_A = a + b\bar{X}_A$$

$$\bar{Y}_B = a + b\bar{X}_B,$$

ya que se requiere que la recta de regresión estimada pase por los puntos $(\bar{X}_A, \bar{Y}_A)$ y $(\bar{X}_B, \bar{Y}_B)$. La solución es

$$b = \frac{\bar{Y}_A - \bar{Y}_B}{\bar{X}_A - \bar{X}_B} \quad \text{y} \quad a = \bar{Y}_A - b\bar{X}_A.$$

Calculemos la media y la variancia de b . Como

$$Y_i = \alpha + \beta X_i + e_i,$$

tenemos

$$\bar{Y}_A = \alpha + \beta \bar{X}_A + \bar{e}_A,$$

$$\bar{Y}_B = \alpha + \beta \bar{X}_B + \bar{e}_B,$$

en las que $\bar{e}_A$ y $\bar{e}_B$ son las medias muestrales (no observadas) de las perturbaciones de la regresión, correspondientes a la primera y segunda mitad de las observaciones, respectivamente. Por tanto,

$$\bar{Y}_A - \bar{Y}_B = \beta(\bar{X}_A - \bar{X}_B) + (\bar{e}_A - \bar{e}_B).$$

En este caso, obtenemos

$$E(b) = E\left[\frac{\bar{Y}_A - \bar{Y}_B}{\bar{X}_A - \bar{X}_B}\right] = \frac{E[\beta(\bar{X}_A - \bar{X}_B) + (\bar{e}_A - \bar{e}_B)]}{\bar{X}_A - \bar{X}_B} = \frac{\beta(\bar{X}_A - \bar{X}_B)}{\bar{X}_A - \bar{X}_B} = \beta,$$

... que es un estimador insesgado de β . Además,

$$E(b^2) = E\left[\left(\frac{\bar{Y}_A - \bar{Y}_B}{\bar{X}_A - \bar{X}_B} - \beta\right)^2\right] = E\left[\frac{\beta(\bar{X}_A - \bar{X}_B) + (\bar{e}_A - \bar{e}_B)}{\bar{X}_A - \bar{X}_B} - \beta\right]^2 \\ = E\left[\frac{\bar{e}_A - \bar{e}_B}{\bar{X}_A - \bar{X}_B}\right]^2 = \frac{E(\bar{e}_A)^2 + E(\bar{e}_B)^2 - 2E(\bar{e}_A\bar{e}_B)}{(\bar{X}_A - \bar{X}_B)^2}.$$

Pero, según los supuestos básicos acerca de e_i , tenemos

$$E(\bar{e}_A)^2 = E\left[\frac{1}{n/2}(e_1 + e_2 + \dots + e_{n/2})\right]^2 \\ = \left[\frac{1}{n/2}\right]^2 [E(e_1^2) + E(e_2^2) + \dots + E(e_{n/2}^2)] \\ = \left[\frac{1}{n/2}\right]^2 (\sigma^2 + \sigma^2 + \dots + \sigma^2) = \left[\frac{1}{n/2}\right]^2 (n/2)\sigma^2 = \frac{\sigma^2}{n/2},$$

y, de forma similar,

$$E(\bar{e}_B)^2 = \frac{\sigma^2}{n/2}.$$

Por último,

$$E(\bar{e}_A\bar{e}_B) = E\left[\frac{1}{n/2}(e_1 + e_2 + \dots + e_{n/2})\right] \left[\frac{1}{n/2}(e_{n/2+1} + e_{n/2+2} + \dots + e_n)\right] = 0,$$

según el supuesto de no autorregresión. Por lo tanto, obtenemos

$$\text{Var}(b) = \frac{\sigma^2/(n/2) + \sigma^2/(n/2)}{(\bar{X}_A - \bar{X}_B)^2} = \frac{4\sigma^2}{n(\bar{X}_A - \bar{X}_B)^2}.$$

Naturalmente, tenemos

$$\text{Var}(b) \geq \text{Var}(\hat{\beta}),$$

en la que $\hat{\beta}$ es el estimador minimocuadrático de β .

EJEMPLO. Demostraremos las ventajas del método de los mínimos cuadrados sobre el método de las semi-medias a través del ejemplo de la demanda de naranjas. Tenemos 12 observaciones que se pueden dividir en dos grupos de 6 observaciones cada una. Los resultados son los siguientes

$$\begin{array}{ll} \bar{X}_A = 60 & \bar{Y}_A = 115 \\ \bar{X}_B = 80 & \bar{Y}_B = 85 \end{array}$$

Esto da como resultado

$$b = \frac{115 - 85}{60 - 80} = -1.5$$

$$a = 115 - (-1.5) \times 60 = 205.$$

La recta de regresión estimada por el método de las semi-medias es, pues,

$$\hat{Y}_i = 205 - 1.5X_i.$$

La variancia de b es

$$\text{Var}(b) = \frac{4\sigma^2}{12(60 - 80)^2} = \frac{\sigma^2}{1200}$$

Comparando $\text{Var}(b)$ con $\text{Var}(\hat{\beta})$ obtenemos

$$\frac{\text{Var}(b)}{\text{Var}(\hat{\beta})} = \frac{\sigma^2/1200}{\sigma^2/2250} = 1.875.$$

Vemos, pues, que el método de las semi-medias produce unas estimaciones de la pendiente de la recta de regresión cuya variancia es un 87,5 % mayor que la variancia de la correspondiente estimación minimocuadrática. En otras palabras, si formásemos nuestras «conjeturas» sobre la pendiente de la recta de regresión de la población a través del método de las semimedias, dichas conjeturas se hallarían considerablemente más dispersas alrededor del verdadero valor que las basadas en la fórmula de los mínimos cuadrados.

Estimación de $\text{Var}(\hat{\alpha})$ y $\text{Var}(\hat{\beta})$

Bajo los supuestos del modelo clásico de regresión normal lineal, los estimadores minimocuadráticos de α y β poseen todas las propiedades deseables de los estimadores, pero su utilidad dependerá del tamaño de sus variancias. Si estas últimas son muy grandes, el hecho de que ningún otro estimador insesgado pueda tener una variancia menor servirá de escaso consuelo. Si las variancias son grandes, es probable que las conjeturas acerca de los verdaderos valores de los parámetros sean muy poco corrientes. En la sección 7.3 se obtuvieron las fórmulas de las variancias de $\hat{\alpha}$ y $\hat{\beta}$, pero dichas fórmulas incluyen el parámetro desconocido σ^2 , o sea que su evaluación resulta imposible. Sin embargo, podemos estimar σ^2 , y en realidad ya hemos obtenido una fórmula de estimación al estudiar los estimadores maximoverosímiles en la sección 7.3.

El estimador $\hat{\sigma}^2$ es un estimador de la obtención de los estimadores de α y β . La fórmula, dada en la ecuación 7.27, es

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2.$$

Como se trata de un estimador maximoverosímil de σ^2 , poseerá todas las propiedades asintóticas deseables, pero tienen que obtenerse sus propiedades de muestras pequeñas. Concretamente, podemos estar interesados en comprobar si $\hat{\sigma}^2$ es un estimador insesgado de σ^2 o no. Para ello, escribiremos de nuevo la expresión de $\hat{\sigma}^2$ en forma algo distinta. En primer lugar, sustituyendo Y_i por su valor se obtiene

$$\hat{\sigma}^2 = \frac{1}{n} \sum (\alpha + \beta X_i + \varepsilon_i - \hat{\alpha} - \hat{\beta}X_i)^2 = \frac{1}{n} \sum [-(\hat{\alpha} - \alpha) - (\hat{\beta} - \beta)X_i + \varepsilon_i]^2.$$

A continuación, sustituyendo $(\hat{\alpha} - \alpha)$ por su valor según la ecuación (7.30), obtendremos

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum [(\hat{\beta} - \beta)\bar{X} - \bar{\varepsilon} - (\hat{\beta} - \beta)X_i + \varepsilon_i]^2 \\ &= \frac{1}{n} \sum [-(\hat{\beta} - \beta)x_i' + \varepsilon_i']^2 \\ &= \frac{1}{n} \sum [(\hat{\beta} - \beta)^2 x_i'^2 + \varepsilon_i'^2 - 2(\hat{\beta} - \beta)\varepsilon_i'x_i'] \\ &= \frac{1}{n}(\hat{\beta} - \beta)^2 \sum x_i'^2 + \frac{1}{n} \sum \varepsilon_i'^2 - \frac{2}{n}(\hat{\beta} - \beta) \sum \varepsilon_i'x_i'. \end{aligned}$$

Pero a partir de (7.32) tenemos

$$\sum x_i' \varepsilon_i' = (\hat{\beta} - \beta) \sum x_i'^2,$$

o sea, que podemos escribir

$$\hat{\sigma}^2 = -\frac{1}{n}(\hat{\beta} - \beta)^2 \sum x_i'^2 + \frac{1}{n} \sum \varepsilon_i'^2.$$

Tomando esperanzas matemáticas en ambos lados, obtenemos

$$E(\hat{\sigma}^2) = -\frac{1}{n} \sum x_i'^2 E(\hat{\beta} - \beta)^2 + \frac{1}{n} \sum E(\varepsilon_i'^2).$$

Ahora bien,
$$E(\hat{\beta} - \beta)^2 = \text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2},$$

y
$$\begin{aligned} E(e_i^2) &= E(e_i - \bar{e})^2 = E(e_i^2) + E(\bar{e}^2) - 2E(e_i\bar{e}) \\ &= \sigma^2 + \frac{\sigma^2}{n} - 2E(e_i(e_1 + e_2 + \dots + e_i + \dots + e_n))/n \\ &= \sigma^2 + \frac{\sigma^2}{n} - 2[E(e_ie_1) + E(e_ie_2) + \dots + E(e_i^2) + \dots + E(e_ie_n)]/n \\ &= \sigma^2 + \frac{\sigma^2}{n} - 2(0 + 0 + \dots + \sigma^2 + \dots + 0)/n \\ &= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} = \left[\frac{n-1}{n}\right]\sigma^2. \end{aligned}$$

Utilizando estos resultados obtenemos

$$\begin{aligned} (7.34) \quad E(\hat{\sigma}^2) &= -\frac{1}{n} \left(\sum x_i^2 \right) \left[\frac{\sigma^2}{\sum x_i^2} \right] + \frac{1}{n} \sum \left[\frac{n-1}{n} \right] \sigma^2 \\ &= -\frac{\sigma^2}{n} + \left[\frac{n-1}{n} \right] \sigma^2 = \left[\frac{n-2}{n} \right] \sigma^2. \end{aligned}$$

Es decir, $\hat{\sigma}^2$ es un estimador *sesgado* de σ^2 . Sin embargo, dado el resultado que hemos obtenido en la ecuación (7.34), resulta fácil construir un estimador insesgado de σ^2 . Si multiplicamos los dos lados de la ecuación (7.34) por $n/(n-2)$ obtenemos

$$\left[\frac{n}{n-2} \right] E(\hat{\sigma}^2) = \sigma^2,$$

o
$$E \left[\frac{n}{n-2} \right] \frac{1}{n} \sum (Y_i - \hat{\alpha} - \beta X_i)^2 = \sigma^2,$$

lo cual se reduce a

$$E \left[\frac{1}{n-2} \right] \sum (Y_i - \hat{\alpha} - \beta X_i)^2 = \sigma^2.$$

El término en el numerador es la suma de los cuadrados de las desviaciones de Y_i respecto a la recta de regresión, al que podemos llamar s^2 , viene dado por

$$(7.35) \quad s^2 = \frac{1}{n-2} \sum (Y_i - \hat{\alpha} - \beta X_i)^2 = \frac{1}{n-2} \sum e_i^2.$$

Como asintóticamente no existe diferencia alguna entre $1/(n-2)$ y $1/n$, s^2 es asintóticamente igual a $\hat{\sigma}^2$, y posee, por tanto, las mismas propiedades asintóticas óptimas.

Para calcular el valor de s^2 , se puede simplificar la fórmula (7.35) con el fin de evitar la necesidad de calcular los valores individuales de e_i , o sea las desviaciones de los valores observados respecto a la recta de regresión de la muestra. Sustituyendo $\hat{\alpha}$ por su valor obtenemos

$$\begin{aligned} s^2 &= \frac{1}{n-2} \sum [Y_i - (\bar{Y} - \beta \bar{X}) - \beta X_i]^2 = \frac{1}{n-2} \sum (y_i' - \beta x_i')^2 \\ &= \frac{1}{n-2} \left[\sum y_i'^2 + \beta^2 \sum x_i'^2 - 2\beta \sum x_i' y_i' \right]; \end{aligned}$$

pero, a partir de (7.12a), tenemos

$$\beta \sum x_i'^2 = \sum x_i' y_i',$$

o sea, que

$$\beta^2 \sum x_i'^2 = \beta \sum x_i' y_i'.$$

Utilizando este resultado llegamos a

$$(7.36) \quad s^2 = \frac{1}{n-2} \left[\sum y_i'^2 - \beta \sum x_i' y_i' \right],$$

que resulta mucho más fácil de calcular que el resultado indicado en la ecuación (7.35).

Si utilizamos s^2 como estimador de σ^2 podemos obtener los estimadores de $\text{Var}(\hat{\alpha})$ y $\text{Var}(\hat{\beta})$; dichos estimadores serán insesgados y tendrán las propiedades asintóticas óptimas. Siguiendo la notación corriente, llamados $s_{\hat{\alpha}}^2$ al estimador de $\text{Var}(\hat{\alpha})$ y $s_{\hat{\beta}}^2$ al de $\text{Var}(\hat{\beta})$. Las fórmulas correspondientes son

$$(7.37) \quad s_{\hat{\alpha}}^2 = s^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i'^2} \right]$$

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum x_i'^2}.$$

La raíz cuadrada de estos estimadores, $s_{\hat{\alpha}}$ y $s_{\hat{\beta}}$, respectivamente, representan los errores standard estimados de $\hat{\alpha}$ y $\hat{\beta}$ y se utilizan muy a menudo como medidas de la precisión de $\hat{\alpha}$ y $\hat{\beta}$. (En los trabajos de investigación se emplea frecuentemente la expresión «errores standard» para referirse a $s_{\hat{\alpha}}$ y $s_{\hat{\beta}}$, en lugar de «errores standard estimados», pero como los verdaderos errores standard no se pueden conocer casi nunca, la omisión de la palabra «estimados» no crea generalmente ninguna confusión.)

Intervalos de confianza para α y β

Es posible obtener una indicación más formal de la precisión de $\hat{\alpha}$ y $\hat{\beta}$ construyendo los intervalos de confianza correspondientes a α y β . Empecemos con el de β . Como

$$\hat{\beta} \sim N(\beta, \sigma_{\hat{\beta}}^2),$$

en la que $\sigma_{\hat{\beta}}^2 = \text{Var}(\hat{\beta})$, de ello se sigue que

$$\frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}} \sim N(0, 1).$$

Además, a partir de (5.8b) sabemos que

$$\frac{\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2}{\sigma^2} \sim \chi_{n-2}^2.$$

En este caso, el número de grados de libertad de la distribución ji-cuadrado es $(n-2)$, ya que se han «utilizado» dos grados de libertad para calcular $\hat{\alpha}$ y $\hat{\beta}$. Obsérvese que podemos escribir

$$\frac{\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2}{\sigma^2} = \frac{(n-2)s^2}{\sigma^2} = \frac{(n-2)s^2/(\sum x_i'^2)}{\sigma^2/(\sum x_i'^2)} = \frac{(n-2)s_{\hat{\beta}}^2}{\sigma_{\hat{\beta}}^2}.$$

Por tanto, tenemos

$$\frac{(n-2)s_{\hat{\beta}}^2}{\sigma_{\hat{\beta}}^2} \sim \chi_{n-2}^2.$$

y, en consecuencia,

$$\frac{(\hat{\beta} - \beta)/\sigma_{\hat{\beta}}}{\sqrt{(n-2)s_{\hat{\beta}}^2/(n-2)\sigma_{\hat{\beta}}^2}} = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}$$

es un cociente en el que el numerador es una variable normal tipificante, y el denominador una variable $[\chi_{n-2}^2/(n-2)]^{1/2}$ independiente. Como se explicó

en la sección 7.2, este cociente tiene una distribución t con $(n-2)$ grados de libertad. Es decir,

$$(7.38) \quad \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \sim t_{n-2}.$$

Mediante una deducción similar, podemos obtener también

$$(7.39) \quad \frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}} \sim t_{n-2}.$$

Estos resultados nos permiten efectuar las siguientes afirmaciones de tipo probabilístico:

$$P\left[-t_{n-2, \lambda/2} \leq \frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}} \leq +t_{n-2, \lambda/2}\right] = 1 - \lambda,$$

$$P\left[-t_{n-2, \lambda/2} \leq \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \leq +t_{n-2, \lambda/2}\right] = 1 - \lambda,$$

en las que $t_{n-2, \lambda/2}$ indica el valor del estadístico t con $(n-2)$ grados de libertad que delimita $\lambda/2$ del área de la distribución t en cada extremo. Este valor se puede obtener a partir de las tablas de la distribución t para cualquier valor de λ que deseemos. El término $(1 - \lambda)$ indica el área de la distribución entre los puntos $-t_{n-2, \lambda/2}$ y $+t_{n-2, \lambda/2}$. A partir de estas afirmaciones probabilísticas, podemos construir los intervalos de confianza para α y β en la forma

$$(7.40) \quad \hat{\alpha} - t_{n-2, \lambda/2}s_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{n-2, \lambda/2}s_{\hat{\alpha}}$$

$$\hat{\beta} - t_{n-2, \lambda/2}s_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{n-2, \lambda/2}s_{\hat{\beta}}.$$

La probabilidad de que el intervalo de confianza así especificado incluya el verdadero valor del parámetro de regresión es $(1 - \lambda)$, y recibe el nombre de «nivel de confianza». Como ya se mencionó en la sección 6.3, los niveles utilizados más corrientemente son el 95 % y el 99 %.

EJEMPLO. Como ejemplo numérico, construyamos los intervalos de confianza del 95 % para α y β en la función que representa la demanda de naranjas a partir de los datos presentados en la sección 7.3. Para ello tenemos que calcular las estimaciones de los errores standard de $\hat{\alpha}$ y $\hat{\beta}$ mediante las fórmulas (7.36) y (7.37). A partir de los datos de este ejemplo, ya se calculó con anterioridad

$$\bar{X} = 70,$$

$$\bar{Y} = 100,$$

$$\sum x_i' y_i' = -3550,$$

$$\sum x_i'^2 = 2250,$$

lo cual da como resultado

$$\hat{\alpha} = 210.460,$$

$$\hat{\beta} = -1.578.$$

Además, tenemos

$$\sum y_i^2 = 6300.$$

Sustituyendo estos valores en la ecuación (7.36) obtenemos

$$s^2 = \frac{1}{12-2} [6300 - (-1.578)(-3550)] = 69.8.$$

Sustituyendo ahora en la ecuación (7.37) se obtiene

$$s_{\hat{\alpha}}^2 = 69.8 \left[\frac{1}{12} + \frac{70^2}{2250} \right] = 157.825555$$

$$y \quad s_{\hat{\beta}}^2 = \frac{69.8}{2250} = 0.031022.$$

Las estimaciones resultantes de los errores standard de $\hat{\alpha}$ y $\hat{\beta}$ son

$$s_{\hat{\alpha}} = 12.563,$$

$$s_{\hat{\beta}} = 0.176.$$

La última información que necesitamos para la construcción de los intervalos de confianza es el valor apropiado de t . Como poseemos 12 observaciones, tenemos 10 grados de libertad. Además, como el nivel de confianza deseado es el 95 %, queremos hallar el valor de t que delimita 0,025 del área en cada extremo de la distribución. Por tanto, buscaremos en la fila «10» y la columna «0,025» de la tabla de la t . El valor correspondiente es 2,228 y, por tanto, los intervalos de confianza del 95 % para α y β son

$$210.460 - 2.228 \times 12.563 \leq \alpha \leq 210.460 + 2.228 \times 12.563$$

$$182.470 \leq \alpha \leq 238.450,$$

$$y \quad -1.578 - 2.228 \times 0.176 \leq \beta \leq -1.578 + 2.228 \times 0.176$$

$$-1.970 \leq \beta \leq -1.186.$$

Intervalo de confianza para $E(Y_i)$

Podemos ampliar el uso de los intervalos de confianza y considerar la precisión con que toda la recta de regresión de la muestra representa a la recta de regresión de la población. Esta última viene dada por $E(Y_i) = \alpha + \beta X_i$ y

representa los valores de Y en un cierto recorrido. Su estimador es la recta de regresión de la muestra $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$. Como $E(Y_i) = \alpha + \beta X_i$, $\hat{Y}_i$ es un estimador sesgado de $E(Y_i)$. Se pueden establecer también otras propiedades deseables de $\hat{Y}_i$, como estimador de $E(Y_i)$; podemos argumentar que, como $\hat{\alpha}$ y $\hat{\beta}$ son los mejores estimadores que podemos construir de α y β , $(\hat{\alpha} + \hat{\beta} X_i)$ es el mejor estimador de $(\alpha + \beta X_i)$. Para obtener el intervalo de confianza para cualquier punto determinado sobre la recta de regresión de la población $E(Y_i)$ tenemos que obtener la variancia de su estimador $\hat{Y}_i$, a la que llamaremos $\sigma_{\hat{Y}_i}^2$. Esta puede obtenerse de la forma siguiente:

$$\begin{aligned} \sigma_{\hat{Y}_i}^2 &= E[\hat{Y}_i - E(\hat{Y}_i)]^2 = E[(\hat{\alpha} + \hat{\beta} X_i) - (\alpha + \beta X_i)]^2 \\ &= E[(\hat{\alpha} - \alpha) + (\hat{\beta} - \beta) X_i]^2 \\ &= E(\hat{\alpha} - \alpha)^2 + E(\hat{\beta} - \beta)^2 X_i^2 + 2E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta) X_i \\ &= \text{Var}(\hat{\alpha}) + X_i^2 \text{Var}(\hat{\beta}) + 2X_i \text{Cov}(\hat{\alpha}, \hat{\beta}) \\ &= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right] + X_i^2 \left[\frac{\sigma^2}{\sum x_i^2} \right] - 2X_i \bar{X} \left[\frac{\sigma^2}{\sum x_i^2} \right]. \end{aligned}$$

El último resultado se ha obtenido sustituyendo los valores de las ecuaciones (7.23), (7.25) y (7.33). Efectuando algunas manipulaciones adicionales se obtiene

$$\begin{aligned} (7.41) \quad \sigma_{\hat{Y}_i}^2 &= \frac{\sigma^2}{\sum x_i^2} \left[\frac{\sum x_i^2}{n} + \bar{X}^2 + X_i^2 - 2X_i \bar{X} \right] \\ &= \frac{\sigma^2}{\sum x_i^2} \left[\frac{\sum x_i^2}{n} + (X_i - \bar{X})^2 \right] = \sigma^2 \left[\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x_i^2} \right]. \end{aligned}$$

Tras haber obtenido la media y la variancia de $\hat{Y}_i$, tenemos que considerar su distribución. Esto resulta muy sencillo: como $(\hat{\alpha} + \hat{\beta} X_i)$ es una combinación lineal de las variables aleatorias independientes $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ que se distribuyen normalmente, $\hat{Y}_i$ se distribuye también normalmente. Por tanto, tenemos

$$\hat{Y}_i \sim N \left[\alpha + \beta X_i, \sigma^2 \left(\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x_i^2} \right) \right];$$

y, en consecuencia,

$$\frac{\hat{Y}_i - (\alpha + \beta X_i)}{\sigma_{\hat{Y}_i}} \sim N(0, 1).$$

En general, la expresión correspondiente a $\sigma^2 \hat{Y}_i$ indicada por la ecuación (7.41), no se puede calcular, ya que incluye el parámetro desconocido σ^2 , pero, sin embargo, podemos sustituir dicho parámetro por su estimador insesgado, consistente, y asintóticamente eficiente s^2 expresado en la ecuación (7.35). Esto da como resultado un estimador de $\sigma^2 \hat{Y}_i$, al que podemos llamar $s^2 \hat{Y}_i$, que posee estas mismas propiedades, y que se define de la forma siguiente

$$(7.42) \quad s^2_{\hat{Y}_i} = s^2 \left[\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum X_i^2} \right].$$

Entonces tenemos, pues,

$$\frac{\hat{Y}_i - (\alpha + \beta X_i)}{s_{\hat{Y}_i}} \sim t_{n-2},$$

y el intervalo de confianza para $(\alpha + \beta X_i)$ será

$$\hat{Y}_i - t_{n-2, \lambda/2} s_{\hat{Y}_i} \leq (\alpha + \beta X_i) \leq \hat{Y}_i + t_{n-2, \lambda/2} s_{\hat{Y}_i},$$

en la que $(1 - \lambda)$ es el nivel de confianza elegido. Como este intervalo de confianza se puede calcular para cualquier valor de X_i dentro del dominio correspondiente, podemos construir intervalos de confianza para *todos* los puntos de la recta de regresión de la población. Esto nos permite construir una *zona de confianza* para la recta de regresión de la población en su conjunto.

EJEMPLO. A modo de ilustración, podemos utilizar el ejemplo de la sección 7.3 correspondiente a la demanda de naranjas y construir la zona de confianza del 95 % para la curva de demanda de la población. Como se recordará, la recta de regresión estimada a partir de la muestra era

$$\hat{Y}_i = 210.460 - 1.578 X_i.$$

La estimación de la variancia de $\hat{Y}_i$ es

$$s^2_{\hat{Y}_i} = 69.8 \left[\frac{1}{12} + \frac{(X_i - 70)^2}{2250} \right],$$

y el valor del estadístico t es 2.228 como anteriormente. Estos son todos los ingredientes que necesitamos para el cálculo de la zona de confianza. El cuadro 7-3 presenta los resultados de dichos cálculos. Las últimas dos columnas representan

los límites superior e inferior de la zona de confianza para la recta de regresión de la población que el intervalo más estrecho es el correspondiente a $X_i = 70$, es

Cuadro 7-3

X_i	$\hat{Y}_i$	$s_{\hat{Y}_i}$	$2.228 s_{\hat{Y}_i}$	Intervalo de confianza 95 %	
				Límite inferior	Límite superior
0	210.46	12.57	28.01	182.45	238.47
10	194.66	10.84	24.15	170.51	218.81
20	178.88	9.14	20.36	158.52	199.24
30	163.10	7.45	16.60	146.50	179.70
40	147.32	5.81	12.94	134.38	160.26
50	131.54	4.27	9.51	122.03	141.05
60	115.76	2.98	6.64	109.12	122.40
70	99.98	2.41	5.37	94.61	105.35
80	84.20	2.98	6.64	77.56	90.84
90	68.42	4.27	9.51	58.91	77.93
100	52.64	5.81	12.94	39.70	65.58
110	36.86	7.45	16.60	20.26	53.46
120	21.08	9.14	20.36	0.72	41.44

decir, a $\bar{X}$. Los intervalos se van haciendo más amplios a medida que nos alejamos de $\bar{X}$, y si unimos los puntos adecuados, obtenemos los límites superior e inferior de la zona de confianza para la recta de regresión de la población que aparecen representados en la figura 7-4.

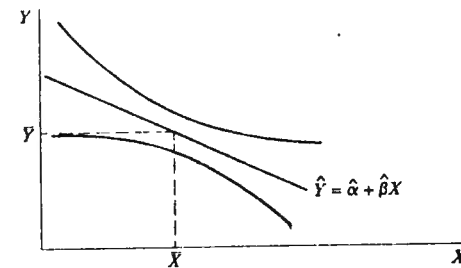


Figura 7-4

Descomposición de la variación muestral de Y

Para complementar los resultados de la estimación que hemos obtenido, pueden utilizarse ciertos conceptos relacionados con la descomposición de la

variación muestral de los valores de la variable dependiente.¹⁰ A modo de ilustración, considérese la variación de Y que aparece en la figura 7-5. En esta última, se han representado los valores de Y obtenidos en una muestra determinada junto a los valores correspondientes de X . Esta representación gráfica se conoce generalmente con el nombre de «diagrama de dispersión» (o nube

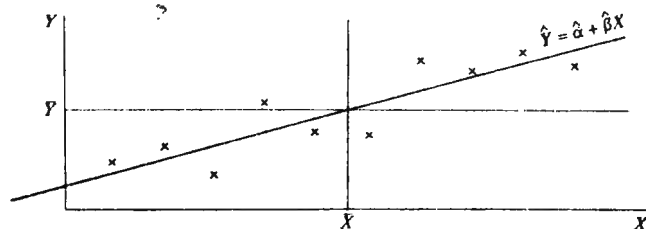


Figura 7-5

de puntos). En la figura 7-5 se presentan 10 observaciones de Y correspondientes a 10 valores distintos de X . Podemos preguntarnos cuál es la razón por la que los valores de Y difieren de una observación a otra. La respuesta, según el modelo de regresión que se postula, es que la variación de Y se debe en parte a los cambios de X —que dan lugar a cambios en el valor esperado de Y — y en parte a los efectos de la perturbación aleatoria. La cuestión siguiente es, pues, qué parte de la variación observada de Y puede atribuirse a la variación de X , y qué parte al efecto aleatorio de la perturbación. Se puede responder a esta pregunta con la ayuda de ciertas medidas que se estudian a continuación. En primer lugar, definamos la expresión muestral de Y . Si no hubiese ninguna variación, al representar los valores de Y respecto a los de X , todos los puntos se hallarían situados sobre una recta horizontal, y puesto que si todos los valores de Y fuesen iguales serían también iguales a la media de la muestra, la recta horizontal sería la correspondiente al punto $\bar{Y}$ en la figura 7-5. Ahora bien, en la práctica, los valores observados de Y se hallan distribuidos alrededor de dicha recta, de forma que la variación de $\bar{Y}$ podría

10. Al hablar de «variación de la variable dependiente» hacemos referencia a los cambios de Y , de una observación muestral a otra. Esto debe distinguirse de la «variancia de Y », que se refiere a la dispersión de los valores de Y_i correspondientes a un valor fijo de X , por ejemplo X_i . En el caso de la variación muestral de Y , los valores de X pueden variar de una observación a otra.

de los valores observados de Y respecto de $\bar{Y}$. Una forma de resumir estas distancias es la suma de los cuadrados de los valores, que generalmente recibe el nombre de «suma total de los cuadrados» y que puede expresarse en forma abreviada por STC . Es decir, de-

$$STC = \sum_i (Y_i - \bar{Y})^2 = \sum_i Y_i^2.$$

Nuestro objetivo consiste en descomponer esta suma de cuadrados en dos partes, una de las cuales debe representar las variaciones de Y que pueden atribuirse a las variaciones de X , y la otra debe considerarse como representación de las variaciones de Y atribuibles a causas aleatorias.

Volvamos ahora a la figura 7-5 y a las observaciones muestrales que aparecen en ella. Supongamos que se ha obtenido una recta de regresión muestral por el método de los mínimos cuadrados y que se ha trazado sobre el diagrama de dispersión tal como allí aparece. Tal como indica el mismo nombre del método de estimación, esta línea es tal que la suma de los cuadrados de las desviaciones respecto de la misma es mínima y, en consecuencia, a veces se denomina la «recta mejor ajustada». Tomemos ahora una observación específica, Y_i , correspondiente al valor X_i de la variable X para estudiar la distancia vertical entre (X_i, Y_i)

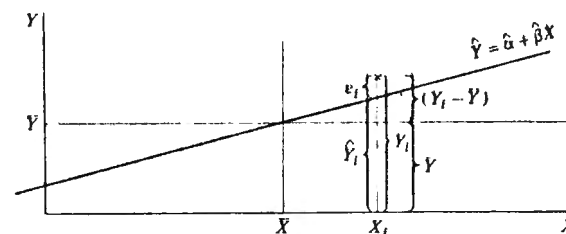


Figura 7-6

y el valor $\bar{Y}$. En la figura 7-6 podemos ver que dicha distancia puede dividirse en dos partes, una de las cuales representa la distancia entre el punto observado y la recta de regresión de la muestra, y la otra la distancia entre dicha recta de regresión y el valor $\bar{Y}$. Es decir, tenemos

$$Y_i = \hat{Y}_i + e_i,$$

en la que $\hat{Y}_i$ es el punto de la recta de regresión muestral correspondiente al valor X_i . Restando $\bar{Y}$ de ambos lados de esta expresión obtenemos

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + e_i.$$

Distancia total des- de $\bar{Y}$	Distancia entre la recta de regresión e $\bar{Y}$	Residuo
---	---	---------

Este análisis se aplica a una sola observación, pero como estamos interesados en la obtención de una medida resumen para *todas* las observaciones muestrales, elevamos al cuadrado ambos lados de esta igualdad y sumamos respecto a todas las observaciones muestrales. Esto da como resultado

$$\begin{aligned}\sum_i (Y_i - \bar{Y})^2 &= \sum_i [(\hat{Y}_i - \bar{Y}) + e_i]^2 \\ &= \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i e_i^2 + 2 \sum_i (\hat{Y}_i - \bar{Y}) e_i.\end{aligned}$$

Consideremos el último término del lado derecho de esta igualdad. Si sustituimos $\hat{Y}_i$ por su valor obtenemos

$$\begin{aligned}2 \sum_i (\hat{Y}_i - \bar{Y}) e_i &= 2 \sum_i (\hat{\alpha} + \beta X_i - \bar{Y}) e_i \\ &= \hat{\alpha} \sum_i e_i + \beta \sum_i X_i e_i - \bar{Y} \sum_i e_i.\end{aligned}$$

Pero, según las ecuaciones (7.10a) y (7.11a), sabemos que $\sum_i e_i = 0$ y $\sum_i X_i e_i = 0$, y, por tanto, podemos llegar a la conclusión de que

$$2 \sum_i (\hat{Y}_i - \bar{Y}) e_i = 0.$$

En consecuencia,

$$(7.43) \quad \sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i e_i^2$$

Suma to- tal de cuadrados (STC)	Suma de cuadrados de la re- gresión (SCR)	Suma de cuadrados de los errores (SCE)
--	---	--

La ecuación (7.43) puede desarrollarse de la forma siguiente:

$$\begin{aligned}(7.44) \quad \text{SCR} &= \sum_i (\hat{Y}_i - \bar{Y})^2 = \sum_i (\hat{\alpha} + \beta X_i - \bar{Y})^2 \\ &= \sum_i [(\bar{Y} - \beta \bar{X}) + \beta X_i - \bar{Y}]^2 \\ &= \sum_i [-\beta(X_i - \bar{X})]^2 \\ &= \beta^2 \sum_i (X_i - \bar{X})^2 = \beta^2 \sum_i x_i'^2.\end{aligned}$$

Hemos visto, pues, que la variación muestral de Y (STC) puede descomponerse en dos partes, una de las cuales describe la variación de los valores ajustados de Y , mientras que la otra describe la variación de los residuos de la regresión. Es decir, SRC representa el efecto estimado de X sobre la variación de Y , y SCE el efecto estimado de la perturbación aleatoria.

La descomposición de la variación muestral de Y da lugar a una medida de la «bondad del ajuste» que se conoce con el nombre de *coeficiente de determinación*, y se representa por R^2 . Dicho coeficiente es, simplemente, la proporción de la variación de Y que puede atribuirse a la variación de X . Como

$$\text{STC} = \text{SCR} + \text{SCE},$$

dividiendo ambos lados por STC obtenemos

$$1 = \frac{\text{SCR}}{\text{STC}} + \frac{\text{SCE}}{\text{STC}}.$$

El coeficiente de determinación se define como

$$(7.45) \quad R^2 = \frac{\text{SCR}}{\text{STC}} = \frac{\hat{\beta}^2 \sum x_i'^2}{\sum y_i'^2},$$

o bien

$$(7.45a) \quad R^2 = 1 - \frac{\text{SCE}}{\text{STC}} = 1 - \frac{\sum e_i^2}{\sum y_i'^2}.$$

R^2 es una medida que se utiliza corrientemente para describir hasta qué punto la recta de regresión de la muestra se ajusta a los datos observados. Nótese que R^2 no puede ser negativo ni mayor que la unidad, es decir,

$$0 \leq R^2 \leq 1.$$

Un valor de R^2 igual a cero indica el peor ajuste que se puede obtener, mientras que un valor unitario indica el mejor ajuste posible.

Una condición necesaria pero no suficiente para que R^2 sea igual a cero es que la recta de regresión de la muestra sea horizontal —es decir, que $\hat{\beta}$ sea igual a cero—. Obsérvese que la recta de regresión de la muestra puede ser horizontal por varias razones distintas. Esto se ilustra gráficamente en la figura 7-7. En el caso a), las observaciones se hallan dispersas de forma aleatoria

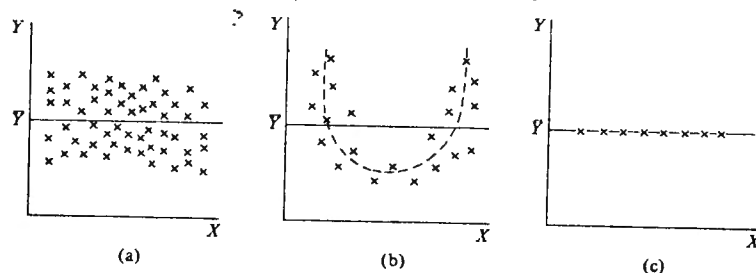


Figura 7-7

en torno al valor $\bar{Y}$. En el caso b), las observaciones se hallan dispersas en torno a una curva tal que la recta mejor ajustada es una horizontal. En este caso, existe una relación entre X e Y , pero dicha relación es extremadamente no lineal, de modo que una recta proporciona un ajuste muy deficiente. Por último, en el caso c) todos los valores de Y son iguales, independientemente del valor de X . Este es un caso excepcional. Cuando todos los valores de Y son constantes, no hay ninguna variación que deba explicarse y, por tanto, la cuestión de la descomposición de la variación es irrelevante. El valor de R^2 en esta situación es indeterminado.

Deben señalarse dos extremos más relativos a la descomposición de la variación muestral de Y . En primer lugar, no se ha aludido a los problemas de la inferencia estadística en ningún punto de la discusión. Esta omisión ha sido deliberada, ya que el objetivo ha consistido simplemente en proporcionar cierta información relativa a la muestra. En segundo lugar, la descomposición de STC desarrollada anteriormente depende crucialmente del uso del método de los mínimos cuadrados en la obtención de la recta de regresión de la muestra. Si hubiésemos usado un método de estimación que produjera una recta de regresión muestral distinta, la descomposición de STC en SCR y SCE no hubiese sido posible. Sin embargo, por lo que respecta a R^2 , podemos generalizar la fórmula (7.45a) con el fin de hacerla aplicable a cualquier método de estimación definiendo R^2 en la forma siguiente

$$(7.45b) \quad R^2 = 1 - \frac{\text{Suma de cuadrados de los residuos}}{\text{STC}}$$

En esta expresión los residuos están representados por las desviaciones respecto de la recta de regresión de la muestra, independientemente del método de estimación utilizado. Pero si la recta de regresión muestral es distinta de la que se hubiese obtenido por el método de los mínimos cuadrados, no podemos seguir interpretando R^2 como una medida de la proporción de la variación de Y atribuible a la regresión muestral. En este caso, R^2 se utilizaría únicamente como medida de la bondad del ajuste.

Si consideramos a R^2 como un estadístico de tipo descriptivo, podemos preguntarnos cuál es el valor de la información que proporciona. Concretamente, supongamos que en una muestra determinada obtenemos un valor muy pequeño de R^2 . Esto significa que la recta de regresión de la muestra se ajusta muy poco a las observaciones. Una explicación posible es que X es una variable explicativa deficiente en el sentido de que su variación deja inafectada a la variable Y . Esto es una proposición acerca de la recta de regresión de la población —que afirma que la recta de regresión de la población es horizontal— y, en consecuencia, se puede probar con referencia a la muestra. Vamos a ocuparnos ahora de cómo puede hacerse. La validez de la contrastación depende de la validez de la hipótesis sostenida —es decir, de la especificación correcta de la ecuación de regresión y de la validez de los supuestos básicos—. En particular, la especificación correcta de la ecuación de regresión implica que en el modelo no existe ninguna otra variable explicativa y que el efecto de X_i sobre $E(Y_i)$ es de tipo lineal. Si no rechazamos la hipótesis de que la recta de regresión de la población es horizontal, en realidad estamos afirmando que Y está influenciada únicamente por la perturbación aleatoria ϵ . Otra explicación posible de un valor pequeño de R^2 es que, aunque X es la variable explicativa relevante, su influencia sobre Y es débil en comparación con la influencia de la perturbación aleatoria. En realidad, éste parece ser el caso en las relaciones que describen la conducta del consumidor a partir de datos de corte transversal. Por ejemplo, un valor corriente de R^2 correspondiente a varias funciones de la conducta del consumidor según los datos del Survey Research Centre de la Universidad de Michigan (con unas 3000 observaciones) es alrededor de 0,20. Ello indicaría que el 80 % de la variación del comportamiento de un consumidor a otro en la muestra puede corresponder a factores distintos de la variable o variables explicativas. Una tercera explicación posible para un valor pequeño de R^2 es que la ecuación de regresión esté mal especificada. En la práctica, ésta es a menudo la conclusión a que el investigador llega en dicho caso. El valor de R^2 tiende a ser considerado como un indicador de la «corrección» de la especificación del modelo. Obviamente, se trata de un criterio puramente operacional, sin fundamento alguno en la inferencia estadística. Cuando consideremos los errores de especificación nos ocuparemos nuevamente de este punto. En cualquier caso, al presentar los resultados de la regresión se acostumbra a señalar el valor de R^2 junto con los resultados del procedimiento de estimación.

EJEMPLO. La descomposición de STC y el cálculo de R^2 pueden ilustrarse a través del ejemplo de la demanda de naranjas introducido en la sección 7.3. A partir de cálculos anteriores tenemos:

$$\sum y_i^2 = 6300,$$

$$\sum x_i^2 = 2250,$$

$$\beta = -1.578.$$

Por tanto,

$$STC = \sum y_i^2 = 6300,$$

$$SCR = \sum x_i^2 = 5602,$$

$$SCE = STC - SCR = 698.$$

Es decir,

$$6300 = 5602 + 698.$$

$$(STC) = (SCR) + (SCE)$$

El coeficiente de determinación es

$$R^2 = \frac{SCR}{STC} = \frac{5602}{6300} = 0.889.$$

Esto significa que el 88,9 % de la variación muestral de Y se puede atribuir a la variación de los valores ajustados de Y , es decir, a $\hat{Y}$. El valor de R^2 indica que la recta de regresión de la muestra se ajusta francamente bien a las observaciones. Esto se representa gráficamente en la figura 7-8.

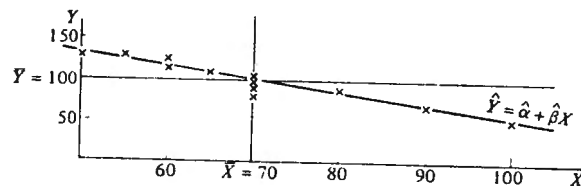


Figura 7-8

Contrastación de hipótesis

Pasemos ahora al problema del uso del modelo de regresión para la *contrastación de hipótesis*. El tipo de hipótesis que más corrientemente se considera entre la variable explicativa X y la variable dependiente Y . Puede darse una interpretación más precisa a esta hipótesis especificando, en primer lugar, los ingredientes de la hipótesis sostenida a la que se halla asociada; es decir,

... que estamos dispuestos a efectuar acerca de los supuestos son todos los subyacentes en el modelo clásico de regresión lineal, presentados en las expresiones (7.1) a (7.6). Bajo esta hipótesis sostenida, la relación entre X e Y viene dada por la dependencia lineal del valor medio de Y_i respecto de X_i , es decir, por $E(Y_i) = \alpha + \beta X_i$. Por tanto, la afirmación de que «no existe relación entre X e Y » debe interpretarse en el sentido de que el valor medio de Y_i no es linealmente dependiente de X_i —es decir, que la recta de regresión de la población es horizontal—. Pero esto es simplemente otra forma de decir que β es igual a cero. Por tanto, la hipótesis nula de no relación entre X e Y es

$$H_0: \beta = 0.$$

Si no tenemos ningún conocimiento previo respecto a los valores de los parámetros de regresión, la hipótesis alternativa será

$$H_A: \beta \neq 0.$$

Si sabemos a priori que β no puede ser positiva, o negativa, la hipótesis alternativa deberá modificarse de acuerdo con ello. Para contrastar H_0 tenemos que obtener un estadístico de prueba y calcular las regiones crítica y de aceptación. El estadístico de prueba se puede obtener a partir del estimador minimocuadrático de β , que bajo los supuestos efectuados posee todas las propiedades óptimas. Podemos utilizar simplemente el hecho de que

$$\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \sim t_{n-2},$$

como se expresa en (7.38). Bajo la hipótesis nula, $\beta = 0$ y, por tanto, el estadístico de prueba correspondiente es

$$\frac{\hat{\beta}}{s_{\hat{\beta}}}$$

que tiene una distribución t con $(n - 2)$ grados de libertad. El límite entre la región de aceptación y la región crítica puede obtenerse a partir de la tabla de la distribución t , para cualquier nivel de significación y cualquier número de grados de libertad. Para una contrastación de dos extremos, al nivel de significación λ y con $(n - 2)$ grados de libertad, la región de aceptación se define por

$$(7.46) \quad -t_{n-2, \lambda/2} \leq \frac{\hat{\beta}}{s_{\hat{\beta}}} \leq +t_{n-2, \lambda/2}.$$

EJEMPLO. A modo de ejemplo, consideremos la demanda de naranjas estimada mediante un modelo de regresión lineal a partir de los datos presentados en la sección 7.3. La hipótesis nula de no relación entre X (precio) e Y (cantidad demandada) equivale a la afirmación de que la demanda de naranjas no está influida por el precio. Tenemos, pues,

$$H_0: \beta = 0.$$

Como hipótesis alternativa tomamos

$$H_A: \beta \neq 0.$$

Queremos contrastar H_0 al nivel de significación del 1%. Como la distribución t se extiende entre $-\infty$ y $+\infty$, todos los valores de $\hat{\beta}/s_{\hat{\beta}}$ son consistentes con la hipótesis nula, pero si H_0 es cierta, los valores de $\hat{\beta}/s_{\hat{\beta}}$ que están muy «alejados» de cero no son muy probables. La decisión de utilizar el nivel de significación del 1% quiere decir que si la desviación de $\hat{\beta}/s_{\hat{\beta}}$ respecto de cero es lo suficientemente grande como para que sólo ocurra por casualidad en un 1% de las veces, rechazaremos H_0 . Como disponemos de 12 observaciones, el valor apropiado del estadístico t para una contrastación de dos extremos es el correspondiente a $t_{10, 0.005}$. A partir de la tabla de la distribución t vemos que dicho valor es 3,169. La región de aceptación para la contrastación es, pues,

$$-3.169 \leq \frac{\hat{\beta}}{s_{\hat{\beta}}} \leq +3.169.$$

A partir de cálculos anteriores tenemos

$$\hat{\beta} = -1.578 \quad \text{y} \quad s_{\hat{\beta}} = 0.176,$$

o sea, que

$$\frac{\hat{\beta}}{s_{\hat{\beta}}} = -8.965,$$

que claramente se encuentra fuera de la región de aceptación. Por tanto, rechazaremos la hipótesis de que no existe relación entre X e Y . En realidad, es probable que en un caso como éste queramos utilizar una hipótesis alternativa con una sola posibilidad, ya que los valores positivos de β (es decir, una curva de demanda creciente) se pueden eliminar en base a principios teóricos. Por tanto, podemos efectuar una contrastación «más estricta» planteando la hipótesis alternativa en la forma

$$H_A: \beta < 0,$$

región de aceptación

$$-t_{n-2, \alpha} \leq \frac{\hat{\beta}}{s_{\hat{\beta}}}.$$

$$-2.764 \leq \frac{\hat{\beta}}{s_{\hat{\beta}}}.$$

Como -8.965 se halla fuera de esta región, la conclusión de rechazar H_0 queda inalterada.

Debe observarse que la hipótesis de no relación entre X e Y puede contrastarse también por otro procedimiento que no utiliza la distribución t . Si la hipótesis nula es cierta, la variación de Y entre una observación y otra no se verá afectada por los cambios de X , sino que debe explicarse por completo por la perturbación aleatoria. Esto significa que la «suma de cuadrados debida a la regresión» (SCR) sólo es distinta de cero porque estamos observando una muestra en lugar de la población total.¹¹ En realidad, a partir de la fórmula (7.44) tenemos

$$SCR = \hat{\beta}^2 \sum x_i^2.$$

Si $\beta = 0$, el valor de SCR en la población sería también cero. Además, según la expresión (7.43) tenemos

$$STC = SCR + SCE,$$

y, por tanto, la implicación de un valor de $SCR = 0$ es que STC y SCE son iguales. Por tanto, si no hay ninguna relación entre X e Y , la razón SCR/SCE sólo sería distinta de cero a consecuencia del muestreo. Ahora bien, puede demostrarse¹¹ que, bajo la hipótesis nula, el estadístico

$$\frac{SCR/1}{SCE/(n-2)}$$

es el cociente de dos variables independientes ji-cuadrado, cada una de las cuales está dividida por el número de grados de libertad correspondiente. Esto

11. Ver, por ejemplo, R. L. Anderson y T. A. Bancroft, *Statistical Theory in Research*. Nueva York, McGraw-Hill, 1952, pp. 158-60.

significa que dicho cociente tiene una distribución F (ver pág. 171), o sea que

$$(7.47) \quad \frac{SCR/1}{SCE/(n-2)} \sim F_{1, n-2}.$$

La región de aceptación para la hipótesis nula de no relación entre X e Y será en este caso

$$(7.48) \quad \frac{SCR/1}{SCE/(n-2)} \leq F_{1, n-2}^{(\lambda)}$$

en la que $F_{1, n-2}^{(\lambda)}$ es el valor del estadístico F con 1 y $(n-2)$ grados de libertad, correspondiente al nivel de significación λ . La contrastación (7.48) equivale a la contrastación de dos extremos utilizando la distribución t (7.46) en el sentido de que ambas proporcionan la misma respuesta siempre que el nivel de significación y los datos muestrales sean los mismo. Como veremos, la diferencia consiste en que la contrastación que utiliza la variable F se puede generalizar fácilmente con el fin de aplicarla a un modelo de regresión con más de una variable explicativa, mientras que la distribución t sólo se puede aplicar en el caso en que existe un solo coeficiente de regresión.

EJEMPLO. En el ejemplo numérico presentado en el apartado anterior tenemos

$$\begin{aligned} SCR &= 5\,602, \\ SCE &= 698, \end{aligned}$$

o sea, que

$$\frac{SCR/1}{SCE/(n-2)} = \frac{5\,602}{698/10} = 80,26.$$

El valor del estadístico F con 1 y 10 grados de libertad, al nivel de significación del 1 %, es 10,0 y, por tanto, es evidente que debemos rechazar la hipótesis nula.

Además de contrastar la hipótesis acerca de la existencia de relación entre X e Y , podemos efectuar contrastaciones para todos los valores específicos de los coeficientes de regresión. Por ejemplo, la hipótesis de que α es igual a cero es, en realidad, la hipótesis de que la recta de regresión de la población pasa por el origen. El estadístico de prueba adecuado sería

$$\frac{\hat{\alpha}}{s_{\hat{\alpha}}}$$

que tiene una distribución t con $(n-2)$ grados de libertad. O también podemos estar interesados en contrastar la hipótesis de que la pendiente β de la

recta de regresión es igual a cierto valor β_0 . En este caso, el estadístico de

$$\frac{\hat{\beta} - \beta_0}{s_{\hat{\beta}}} \sim t_{n-2}.$$

Predicción

Además de utilizarse para la estimación y la contrastación de hipótesis, el modelo de regresión puede utilizarse también para la predicción. Concretamente, a menudo nos interesa «pronosticar» el valor de Y correspondiente a un valor determinado de X . Por ejemplo, el gerente de un supermercado puede estar interesado en conocer la cantidad de naranjas que confía vender si establece el precio a un cierto nivel. Para concretar, supongamos que el valor dado de la variable explicativa es X_0 , o sea que nuestra tarea consiste en predecir el valor de Y_0 . Como Y_0 es una variable aleatoria cuyos valores están dispersos en torno al punto correspondiente a X_0 sobre la recta de regresión de la población, no podremos conocer nunca su valor con anterioridad al experimento, ni aun en el caso que conociéramos todos los parámetros de la población. Si se conocieran los parámetros de la población, el predictor de Y_0 sería su media,

$$E(Y_0) = \alpha + \beta X_0,$$

que define un punto sobre la recta de regresión de la población. Este es el predictor óptimo de Y_0 en el sentido de que la variancia de Y_0 respecto a $E(Y_0)$ es menor que su variancia respecto a cualquier otro punto. Los valores de Y_0 se distribuirán normalmente con variancia igual a σ^2 . Esto es debido a la presencia de la perturbación aleatoria ε en la ecuación de regresión. En la realidad, no conocemos $E(Y_0)$ y tenemos que estimarla. Su estimador es el punto correspondiente sobre la recta de regresión de la muestra,

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0,$$

ya que $\hat{\alpha}$ y $\hat{\beta}$ son los mejores estimadores de α y β que podemos construir bajo determinados supuestos. Ahora bien, el valor real de Y_0 diferirá del valor de la predicción $\hat{Y}_0$ por las dos siguientes razones:

1. El valor de Y_0 no será igual a $E(Y_0)$ —es decir, no se hallará sobre la recta de regresión de la población— debido a la perturbación aleatoria ε_0 .
2. La regresión muestral no será igual a la recta de regresión de la población debido a la existencia del error muestral.

Formalmente, podemos escribir

$$Y_0 - \hat{Y}_0 = [Y_0 - E(Y_0)] + [E(Y_0) - \hat{Y}_0].$$

El primer tipo de error es inherente al mecanismo por medio del cual se generan los valores de la variable dependiente, y no podemos hacer nada para reducirlo. Sin embargo, el segundo tipo de error se reduciría si aumentásemos la precisión de la estimación de la recta de regresión de la población mediante un aumento en el tamaño de la muestra.

La diferencia entre el valor real de Y_0 y el valor de la predicción $\hat{Y}_0$ se conoce con el nombre de *error del pronóstico*. Obsérvese que

$$Y_0 - \hat{Y}_0 = (\alpha + \beta X_0 + \varepsilon_0) - (\hat{\alpha} + \hat{\beta} X_0)$$

es una combinación lineal de las variables aleatorias independientes $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ que se distribuyen normalmente y, por tanto, el error del pronóstico es también una variable aleatoria con una distribución normal. En consecuencia, su distribución queda totalmente determinada a través de su media y su variancia. La media puede obtenerse de la forma siguiente:

$$\begin{aligned} E(Y_0 - \hat{Y}_0) &= E(\alpha + \beta X_0 + \varepsilon_0 - \hat{\alpha} - \hat{\beta} X_0) \\ &= \alpha + \beta X_0 + E(\varepsilon_0) - E(\hat{\alpha}) - E(\hat{\beta}) X_0 = 0. \end{aligned}$$

La variancia es

$$\begin{aligned} E[(Y_0 - \hat{Y}_0) - E(Y_0 - \hat{Y}_0)]^2 &= E(Y_0 - \hat{Y}_0)^2 \\ &= E\{[Y_0 - E(Y_0)] + [E(Y_0) - \hat{Y}_0]\}^2 \\ &= E[Y_0 - E(Y_0)]^2 + E[E(Y_0) - \hat{Y}_0]^2 \\ &\quad + 2E[Y_0 - E(Y_0)][E(Y_0) - \hat{Y}_0]. \end{aligned}$$

Tomando el último término de lado derecho de la igualdad,

$$\begin{aligned} 2E[Y_0 - E(Y_0)][E(Y_0) - \hat{Y}_0] &= 2E(\varepsilon_0)[E(Y_0) - \hat{Y}_0] \\ &= 2E(\varepsilon_0)(\alpha + \beta X_0 - \hat{\alpha} - \hat{\beta} X_0) \\ &= 2E(\varepsilon_0)[-(\hat{\alpha} - \alpha) - (\hat{\beta} - \beta) X_0] \\ &= 0, \end{aligned}$$

dependen únicamente de las perturbaciones muestrales, las cuales son independientes de ε_0 . Por tanto,

$$E(Y_0 - \hat{Y}_0)^2 = E[Y_0 - E(Y_0)]^2 + E[E(Y_0) - \hat{Y}_0]^2$$

Variancia total del error del pronóstico (σ_f^2)	Variancia debida a la perturbación aleatoria (σ^2)	Variancia debida al error mues- tral ($\sigma_{\hat{Y}_0}^2$)
---	--	---

o

$$\sigma_f^2 = \sigma^2 + \sigma_{\hat{Y}_0}^2.$$

Es decir, la variancia del error del pronóstico consta de dos partes, una de las cuales es igual a la variancia de la perturbación y la otra a la variancia del predictor $\hat{Y}_0$ respecto de su media $E(Y_0)$. La variancia de la perturbación queda fuera de nuestro control, pero la variancia del predictor se puede disminuir aumentando el tamaño de la muestra utilizada para la estimación de la recta de regresión de la población.

Utilizando la siguiente fórmula para la variancia del error del pronóstico:

$$(7.50) \quad \sigma_f^2 = \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i'^2} \right] = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i'^2} \right].$$

Esto significa que la variancia del error del pronóstico será menor:

1. Cuanto mayor sea el tamaño muestral n .
2. Cuanto mayor sea la dispersión de la variable explicativa en la muestra (es decir, cuanto mayor sea $\sum x_i'^2$).
3. Cuanto menor sea la distancia entre X_0 y la media de la muestra $\bar{X}$.

Las dos primeras conclusiones son inmediatas y reflejan el hecho de que, cuanto mejor sea la estimación de la recta de regresión de la población, menor será la variancia del error del pronóstico. La tercera conclusión resulta más interesante y significa que el pronóstico será mejor para los valores de X , que se hallen próximos a $\bar{X}$ que para los que se encuentran más alejados de dicho punto. Esto es consistente con el argumento, que resulta intuitivamente plausible, de que somos más capaces de efectuar pronósticos que se hallen dentro del ámbito de nuestra experiencia que de realizarlos fuera de ella. En este caso, el ámbito de nuestra «experiencia» está representado por los valores muestrales de la variable explicativa X , y el punto central de dicho ámbito es $\bar{X}$. Cuanto más lejos nos aventuremos al pronosticar, menos fiable resultará el pronóstico.

En general, la expresión de σ_F^2 presentada en la ecuación (7.50) será desconocida, y tendrá que ser estimada. Esto puede hacerse simplemente sustituyendo σ^2 por su estimador s^2 , lo cual producirá un estimador insesgado, consistente, y asintóticamente eficiente, de σ_F^2 , al que podemos llamar s_F^2 que se define en la forma siguiente

$$(7.51) \quad s_F^2 = s^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i'^2} \right].$$

Resumiendo, hemos visto que el error del pronóstico ($Y_0 - \hat{Y}_0$) se distribuye normalmente, tiene una media cero y su variancia es σ_F^2 . Es decir,

$$(Y_0 - \hat{Y}_0) \sim N(0, \sigma_F^2);$$

$$\text{y, por tanto,} \quad \frac{Y_0 - \hat{Y}_0}{\sigma_F} \sim N(0, 1).$$

Sustituyendo σ_F por s_F

$$\frac{Y_0 - \hat{Y}_0}{s_F} \sim t_{n-2}.$$

El último resultado nos permite efectuar afirmaciones específicas de tipo probabilístico acerca de los pronósticos. Concretamente, podemos establecer un intervalo con una probabilidad determinada de que contenga el valor real de Y_0 . Sea este nivel de probabilidad $(1 - \lambda)$, siendo λ cualquier número determinado que elijamos entre 0 y 1. En este caso podemos escribir

$$P \left[-t_{n-2, \lambda/2} \leq \frac{Y_0 - \hat{Y}_0}{s_F} \leq +t_{n-2, \lambda/2} \right] = 1 - \lambda.$$

El correspondiente intervalo de confianza para Y_0 es

$$\hat{Y}_0 - t_{n-2, \lambda/2} s_F \leq Y_0 \leq \hat{Y}_0 + t_{n-2, \lambda/2} s_F.$$

Este intervalo es simétrico en torno al predictor Y_0 , y puede esperarse que contenga el valor real de Y_0 con una probabilidad $(1 - \lambda)$.

EJEMPLO. Supongamos que queremos efectuar una predicción de la demanda de naranjas al precio de 110 centavos por libra, utilizando la curva de demanda estimada anteriormente. La predicción de la cantidad demandada será

$$\hat{Y}_0 = 210.460 - 1.578 \times 110 = 36.88,$$

$$s_F = \sqrt{10.9 \left[1 + \frac{1}{12} + \frac{(110 - 70)^2}{2250} \right]} = 11.20.$$

El intervalo de confianza del 95 % para Y_0 es $36.88 \pm 2.228 \times 11.20$, o sea entre 12 y 62 libras.

Presentación de los resultados de la regresión

Un último punto a mencionar antes de finalizar este capítulo hace referencia a la presentación de los resultados del análisis de regresión. Sea cual sea el objetivo que se pretenda, al operar con datos muestrales se efectuarán estimaciones de los coeficientes de regresión y de los errores standard de dichos estimadores. Estos resultados están completados generalmente por R^2 , el coeficiente de determinación, que indica en qué medida la recta de regresión de la muestra se ajusta a las observaciones. Se ha convertido en práctica corriente la presentación de estos resultados escribiendo la ecuación de regresión estimada con los errores standard estimados, entre paréntesis, debajo de los coeficientes respectivos, y a continuación el valor de R^2 . Es decir, escribimos

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i, \quad R^2 = \dots$$

(s_α) (s_β)

En el caso del ejemplo correspondiente a la demanda de naranjas, tenemos

$$\hat{Y}_i = 210.460 - 1.578 X_i, \quad R^2 = 0.889.$$

(12.563) (0.176)

El «acento circunflejo» sobre $\hat{Y}_i$ indica que la ecuación se cumple sólo para los valores ajustados de la variable dependiente y no para los valores realmente observados. De forma alternativa, podemos escribir

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i, \quad R^2 = \dots$$

(s_α) (s_β)

en la que e_i representa los residuos minimocuadráticos. Pero obsérvese que es incorrecto escribir $Y_i = \hat{\alpha} + \hat{\beta} X_i$. Algunos investigadores presentan los coeficientes de los coeficientes estimados respecto a sus correspondientes errores standard estimados, en lugar de escribir directamente estos últimos, lo que constituye una alternativa aceptable.

EJERCICIOS:

Todos los problemas de este grupo de ejercicios se refieren al modelo de regresión lineal simple

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

para el cual se cumplen todos los supuestos (7.2) a (7.6).

- 7-1. Obtener el estimador lineal insesgado óptimo de α y su variancia.
- 7-2. Considérense dos perturbaciones ε_t y ε_s ($t \neq s$) cualesquiera de la regresión. Según los supuestos, estas perturbaciones tienen la misma variancia y son independientes entre sí. ¿Puede afirmarse lo mismo acerca de los residuos minimocuadráticos e_t y e_s respectivos?
- 7-3. Utilizando el método de las semi-medias, y los datos sobre la demanda de naranjas presentados en el texto, hemos estimado la recta de regresión

$$\hat{Y}_i = 205 - 1.5X_i.$$

Calcular el valor de R^2 .

- 7-4. ¿Qué supuestos de entre los presentados en (7.2) a (7.6) son necesarios para que el estimador minimocuadrático de β sea a) insesgado; b) lineal insesgado y óptimo; c) consistente?
- 7-5. Supóngase que $Y_i = \log$ del valor de la producción por trabajador, $X_i = \log$ de la tasa de salarios, y que el subíndice i hace referencia a la empresa i -ésima. El parámetro β puede considerarse como una medida de la elasticidad de sustitución entre trabajo y capital. Los resultados minimocuadráticos correspondientes a la industria A son los siguientes

$$Y_i = -0.4 + 1.0X_i + \varepsilon_i \quad (n = 50). \quad (0.1)$$

Los resultados correspondientes a la industria B son

$$Y_i = -0.3 + 0.8X_i + \varepsilon_i \quad (n = 50). \quad (0.1)$$

... las ε_i son estadísticamente independientes

- a. Verificar que R^2_A igual a $2/3$, y $R^2_B = 4/7$
- b. Verificar la hipótesis de que ambas industrias tienen la misma elasticidad de sustitución, es decir que $\beta_A = \beta_B$.

- 7-6. Supóngase que se han calculado los valores de $\hat{\alpha}$, $\hat{\beta}$, $s_{\hat{\alpha}}$, $s_{\hat{\beta}}$ y R^2 , en base a una muestra de n observaciones. Supóngase ahora que añadimos una nueva observación a la muestra y calculamos nuevamente estos cinco valores. Comparar los valores obtenidos ahora con los originales —señalando si experimentan un aumento, una disminución, ningún cambio, o un resultado incierto— dadas las situaciones siguientes:

$$a. X_{n+1} = \frac{1}{n} \sum_{i=1}^n X_i.$$

$$b. X_{n+1} = \max\{X_1, X_2, \dots, X_n\}.$$

$$c. X_{n+1} = 0.$$

- 7-7. Supóngase que la variable explicativa X puede tomar únicamente los valores 0 y 1. La muestra está formada por n_1 observaciones en las que $X = 0$, y n_2 observaciones en las que $X = 1$. Sea Y_1 el valor medio de Y para las n_1 observaciones en las que $X = 0$, e Y_2 el valor medio de Y para las n_2 observaciones en las que $X = 1$. Obtener $\hat{\alpha}$, $\hat{\beta}$, $\text{Var}(\hat{\alpha})$ y $\text{Var}(\hat{\beta})$.
- 7-8. Se han propuesto tres modelos alternativos, que pretenden explicar los cambios en el gasto agregado en alimentos, en los EE. UU.:

$$\text{Modelo A: } Y_t = \alpha + \beta X_t + \varepsilon_{1t}$$

$$\text{Modelo B: } Y_t = \gamma + \delta Z_t + \varepsilon_{2t}$$

$$\text{Modelo C: } \log Y_t = \alpha^* + \beta^* \log X_t + \varepsilon_{3t},$$

siendo Y = gasto privado en alimentos (excluyendo las bebidas alcohólicas), X = renta personal disponible, Z = gastos totales personales en consumo, y t = tiempo expresado en años. A partir de los datos del cuadro 7-4,

- a. Estimar las tres ecuaciones de regresión por el método de los mínimos cuadrados.
- b. Obtener, para cada modelo, una estimación (expresada por un valor numérico) de la elasticidad-renta de la demanda de alimentos, definida en la forma

$$\frac{dQ}{dX} \frac{X}{Q},$$

en la que Q es la cantidad de alimentos consumidos, es decir, su importe dividido por el precio.

- c. Examinar los resultados obtenidos en a, intentando descubrir si pueden ser de utilidad para elegir entre los tres modelos.

- 7-9. La ecuación de regresión minimocuadrática estimada a partir de 22 observaciones es

$$Y_i = 10 + 5X_i + e_i, \quad R^2 = 0.8.$$

Efectuar una contrastación de la existencia o no existencia de relación entre X e Y , utilizando a) la distribución t ; b) la distribución F .

- 7-10. Dos investigadores que trabajan de forma independiente han estimado los coeficientes de

$$Y_i = \alpha + \beta X_i + e_i$$

por el método de los mínimos cuadrados. Al enterarse, cada uno de ellos del trabajo del otro, decidieron agregar sus resultados para obtener una estimación conjunta de β , considerando dos procedimientos posibles para realizarla:

1. Tomar la media aritmética simple de sus estimaciones.
2. Combinar las dos muestras y obtener una nueva estimación de β .

¿Existe alguna diferencia entre ambos procedimientos? De ser así, ¿cuál de ellos produciría una estimación mejor de β ?

- 7-11. Una de las contrastaciones relativas a la existencia de relación entre X e Y se efectúa construyendo el estadístico t , mediante $\hat{\beta}/s_{\hat{\beta}}$. Mostrar que esto es exactamente igual a

$$[R^2(n-2)/(1-R^2)]^{1/2}.$$

- 7-12. En una población, la verdadera relación entre X e Y viene dada por la ecuación

$$Y_i = 2 + 3X_i + e_i.$$

Cuadro 7-4

Año	Gastos personales de consumo *		Renta personal disponible *
	Total	Alimentos	
1929	79.0	19.5	83.1
1930	71.0	18.0	74.4
1931	61.3	14.7	63.8
1932	49.3	11.4	48.7
1933	46.4	10.9	45.7
1934	51.9	12.2	52.0
1935	56.3	13.6	58.3
1936	62.6	15.2	66.2
1937	67.3	16.4	71.0
1938	64.6	15.6	65.7
1939	67.6	15.7	70.4
1940	71.9	16.7	76.1
1946	147.1	40.7	160.6
1947	165.4	45.8	170.1
1948	178.3	48.2	189.3
1949	181.2	46.4	189.7
1950	195.0	47.4	207.7
1951	209.8	53.4	227.5
1952	219.8	55.8	238.7
1953	232.6	56.6	252.5
1954	238.0	57.7	256.9
1955	256.9	59.2	274.4
1956	269.9	62.2	292.9
1957	285.2	65.2	308.8
1958	293.2	67.4	317.9
1959	313.5	68.1	337.1
1960	328.2	69.7	349.9
1961	337.3	71.0	364.7
1962	356.8	73.9	384.6
1963	375.0	76.0	402.5
1964	399.2	80.0	431.8

Fuente: *Economic Report of the President*, enero 1965, cuadros B-7 y B-9.

* En miles de millones de dólares

Supóngase que los valores de X en una muestra de 10 observaciones son 1, 2, ..., 10. Los valores de los términos de perturbación se obtienen de forma aleatoria a partir de una población normal con media cero y variancia unitaria:

$\epsilon_1 = 0.464$	$\epsilon_6 = 0.906$
$\epsilon_2 = 0.060$	$\epsilon_7 = -1.501$
$\epsilon_3 = 1.486$	$\epsilon_8 = -0.690$
$\epsilon_4 = 1.022$	$\epsilon_9 = 0.179$
$\epsilon_5 = 1.394$	$\epsilon_{10} = -1.372$

- Presentar los 10 valores observados de X e Y .
- Estimar los coeficientes de regresión y sus errores standard utilizando las fórmulas minimocuadráticas, comparando los resultados con los verdaderos valores.
- Contrastar la existencia de relación entre X e Y .
- Obtener la predicción del valor de Y —y los límites para el intervalo de confianza del 95 %— para $X = 12$.

Violaciones de los supuestos básicos

En el capítulo 7 hemos desarrollado el llamado modelo clásico de regresión normal lineal, indicando cómo puede utilizarse dicho modelo para la estimación, la contrastación de hipótesis y la predicción. Para obtener los resultados se han utilizado determinados supuestos acerca de la perturbación estocástica ϵ_i y la variable explicativa X_i a los que hemos llamado *supuestos básicos*. Los cuatro primeros supuestos respecto a la perturbación estocástica son los de normalidad, media cero, homoscedasticidad y no existencia de autorregresión. El quinto supuesto se refiere a la variable explicativa, que se supone que es no estocástica y que su variancia muestral es un número finito, sea cual sea el tamaño de la muestra. Basándonos en estos supuestos, hemos podido mostrar que los estimadores minimocuadráticos de los parámetros de regresión poseen todas las propiedades deseables. El principal objetivo de este capítulo consiste en descubrir cómo se ven afectadas dichas propiedades de los estimadores minimocuadráticos cuando no se cumple alguno de los supuestos básicos. Además, en el caso de que nos encontremos con que las consecuencias de la violación de tales supuestos son serias, intentaremos construir métodos de estimación alternativos que proporcionen resultados más satisfactorios.

Quizá sea conveniente empezar con una advertencia. Todos los resultados correspondientes al modelo clásico de regresión normal lineal estudiado en el capítulo anterior están bien establecidos, pero los problemas que se plantean en las situaciones en que no se cumplen los supuestos básicos no se han resuelto de forma satisfactoria en todos los casos. En realidad, lo que distingue a un econométra de un estadístico es la preocupación por este tipo de problemas. La principal razón para ello estriba en que, debido a la naturaleza particular de las relaciones económicas, es muy probable que algunos de los supuestos básicos dejen de cumplirse y, por tanto, los problemas resultantes para la estimación y la contrastación de hipótesis son especialmente relevantes en economía. Muchos de los resultados que presentamos aquí son relativamente recientes, algunos de los problemas que siguen sin resolver son objeto de es-

tudio en la actualidad y es de esperar que se resuelvan en un futuro próximo, y otros, por el contrario, pueden seguir pendientes durante largo tiempo.

De los cinco supuestos básicos, los dos primeros, normalidad y media cero de la perturbación, no necesitan ser estudiados con gran amplitud. Si se elimina el supuesto de que la perturbación se distribuye normalmente, los estimadores minimocuadráticos de los coeficientes de regresión siguen siendo ELIO, ya que dicha propiedad es independiente de la forma de la población originaria. Esto significa que, aun prescindiendo del supuesto de normalidad, los estimadores minimocuadráticos son insesgados y poseen una variancia mínima entre todos los estimadores lineales insesgados de los parámetros respectivos. Sin embargo, ya no puede afirmarse que son eficientes, porque al no especificarse la forma de la distribución, no se conocen los límites inferiores de Cramer-Rao de sus variancias. Además, los estimadores minimocuadráticos ya no son estimadores maximoverosímiles, puesto que la función de verosimilitud, que se basa en el supuesto de normalidad, ya no puede aplicarse. Con respecto a las propiedades asintóticas, los estimadores minimocuadráticos son consistente y asintóticamente eficientes, independientemente de la distribución del término de perturbación. Esto se deriva del hecho de que la distribución de los estimadores minimocuadráticos de los coeficientes de regresión se aproxima a una normal cuando $n \rightarrow \infty$. Esta afirmación puede demostrarse con la ayuda del teorema central del límite,¹ y ello implica que los estimadores minimocuadráticos poseen *asintóticamente* la misma distribución que los estimadores maximoverosímiles basados en el supuesto de normalidad; por tanto, poseen las mismas propiedades asintóticas deseables. Resumiendo, cuando se elimina el supuesto de normalidad del término de perturbación, los estimadores minimocuadráticos de los coeficientes de regresión siguen poseyendo la mayor parte de sus propiedades deseables, y las fórmulas correspondientes a las variancias de dichos estimadores quedan inalteradas. Sin embargo, los intervalos de confianza y las contrastaciones de la significación de α y β dependen crucialmente del supuesto de normalidad. Sin dicho supuesto, los estimadores minimocuadráticos no se distribuyen normalmente en las muestras pequeñas y, por tanto, en sentido estricto, los límites de confianza y las contrastaciones presentadas en la sección 7.4 no son aplicables en este caso. Afortunadamente, si la distribución del término de perturbación no difiere muy radicalmente de una normal, estos límites de confianza y las contrastaciones respecto a la significación no se ven afectados muy seriamente y pueden utilizarse como aproximaciones razonables.²

El segundo supuesto —media cero de la perturbación de la regresión— se efectúa en base a la especificación de que la recta de regresión de la población es

$$E(Y_i) = \alpha + \beta X_i.$$

Si la media de la perturbación no es cero, sino, por ejemplo, μ_i , tenemos

$$E(Y_i) = \alpha + \beta X_i + \mu_i.$$

Las implicaciones de ello dependen de la naturaleza de μ_i . En especial, tenemos que distinguir entre el caso en que μ_i tiene el mismo valor para todas las observaciones, y el caso en que este valor puede variar. En el primero, podemos escribir $\mu_i = \mu$ y, por tanto, la verdadera recta de regresión de la población es

$$E(Y_i) = \alpha + \mu + \beta X_i$$

o

$$E(Y_i) = \alpha^* + \beta X_i.$$

Es evidente, pues, que, aunque el estimador minimocuadrático β no se ve afectado, la fórmula minimocuadrática para la estimación de la ordenada en el origen nos proporciona una estimación de α^* y no de α . No existe *ningún* procedimiento para estimar α y μ por separado y obtener estimaciones insesgadas o por lo menos consistentes. En el segundo caso, en el que μ_i no es una constante, la ordenada en el origen se convierte en $(\alpha + \mu_i)$, de modo que puede variar de una observación a otra. Ello significa que el valor medio de la variable dependiente, $E(Y_i)$, no sólo varía debido a los cambios en X_i , sino también por otras razones; o sea, no se ha especificado correctamente la relación entre X_i e Y_i . Este y otros tipos de errores de especificación se estudiarán detenidamente en el capítulo 10.

Tras este primer análisis, nos quedan los tres supuestos restantes. Respecto al primero de ellos debe señalarse que, cuando al término de perturbación no es homoscedástico, se denomina «heteroscedástico», y en la sección 8.1 estudiaremos este caso. En las secciones 8.2 y 8.3 nos ocuparemos de las perturbaciones autorregresivas y de las variables explicativas estocásticas.

8.1 Heteroscedasticidad

Según el supuesto (7.4) del modelo clásico de regresión normal lineal, tenemos

$$E(\epsilon_i^2) = \sigma^2 \text{ para todos los valores de } i.$$

1. E. Malinvaud, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1960, pp. 195-97.

2. *Ibid.*, pp. 251-53.

Como se supone que la media de ϵ_i es cero, podemos escribir

$$\text{Var}(\epsilon_i) = \sigma^2.$$

Esta característica del término de perturbación de la regresión recibe el nombre de homoscedasticidad e implica que la variancia de la perturbación es constante en todas las observaciones. En los modelos que incluyen observaciones de agregados a lo largo del tiempo, este supuesto puede no presentar graves problemas, ya que los valores de la variable explicativa son, típicamente, de un orden de magnitud parecido en todos los puntos de observación, y lo mismo ocurre con los valores de la variable dependiente. Por ejemplo, en una función agregada de consumo el nivel de consumo en años recientes es de un orden de magnitudes parecido al nivel de consumo de hace veinte años, y lo mismo ocurre con la renta. A menos que se presenten algunas circunstancias especiales, el supuesto de homoscedasticidad resulta plausible en los modelos agregados. Sin embargo, cuando se trata de datos microeconómicos, las observaciones pueden acarrear diferencias sustanciales de magnitud, como, por ejemplo, en el caso de los datos correspondientes a las rentas y gastos de familias individuales. En este caso, el supuesto de homoscedasticidad no resulta muy plausible a priori, ya que podríamos esperar una menor variación en el consumo de las familias de rentas pequeñas que en el de familias de renta elevada. En los niveles pequeños de renta, el nivel medio de consumo es reducido y la variación en torno a este nivel se ve restringida: el consumo no puede reducirse muy por debajo del nivel medio, porque ello podría significar inanición, y no puede crecer muy por encima, ya que los activos y las posibilidades de crédito no lo permiten. Es probable que estas restricciones sean menos efectivas en los niveles de renta elevada, y la evidencia empírica sugiere que tales consideraciones a priori coinciden con la conducta real.³ En este caso, y en otros similares, el modelo apropiado debe incluir perturbaciones *heteroscedásticas*.

Propiedades de los estimadores minimocuadráticos

Si el término de perturbación de la regresión es heteroscedástico, tenemos

$$E(\epsilon_i^2) = \sigma_i^2.$$

Ello implica que la variación de la perturbación puede variar de una observación a otra, y queremos saber el efecto de esta conducta de la variancia sobre las propiedades de los estimadores minimocuadráticos de los coeficientes de

regresión. Consideremos en primer lugar la propiedad de insesgadez. El estimador minimocuadrático de β es

$$\hat{\beta} = \frac{\sum x_i' y_i'}{\sum x_i' x_i'} = \beta + \frac{\sum x_i' \epsilon_i}{\sum x_i' x_i'},$$

como se indica en la expresión (7.32). Por tanto,

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum x_i' \epsilon_i}{\sum x_i' x_i'}\right] = \beta.$$

Del mismo modo, $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = (\alpha + \beta\bar{X} + \bar{\epsilon}) - \hat{\beta}\bar{X}$,

y

$$E(\hat{\alpha}) = \alpha + \beta\bar{X} + E(\bar{\epsilon}) - E(\hat{\beta})\bar{X} = \alpha.$$

Es decir, los estimadores minimocuadráticos son *insesgados* incluso en condiciones de heteroscedasticidad.

A continuación, veamos si los estimadores minimocuadráticos siguen siendo los estimadores lineales insesgados óptimos (ELIO). Lo podemos efectuar obteniendo las fórmulas de los ELIO cuando existe heteroscedasticidad, y comparándolas con las fórmulas minimocuadráticas. Si existe alguna diferencia, los estimadores minimocuadráticos no serán ELIO. Si empezamos considerando el ELIO de β , al que llamaremos $\tilde{\beta}$, tenemos

$$\tilde{\beta} = \sum_i a_i Y_i$$

según la condición de linealidad. En esta expresión, a_i ($i = 1, 2, \dots, n$) son constantes que se deben calcular. Además,

$$E(\tilde{\beta}) = E\left(\sum_i a_i Y_i\right) = \sum_i a_i(\alpha + \beta X_i) = \alpha\left(\sum_i a_i\right) + \beta\left(\sum_i a_i X_i\right).$$

Para que $\tilde{\beta}$ sea insesgado es preciso que

$$\sum_i a_i = 0 \quad \text{y} \quad \sum_i a_i X_i = 1.$$

como en el caso de homoscedasticidad. La variancia de $\tilde{\beta}$ viene dada por la expresión

$$\text{Var}(\tilde{\beta}) = E\left[\sum_i a_i Y_i - E\left(\sum_i a_i Y_i\right)\right]^2 = E\left[\sum_i a_i (Y_i - E Y_i)\right]^2$$

$$E\left[\sum_i a_i \epsilon_i\right]^2 = E\sum_i a_i^2 \epsilon_i^2 + 2E\sum_{i \neq j} a_i a_j \epsilon_i \epsilon_j,$$

3. Ver S. J. Prais y H. S. Houthakker *The Analysis of Family Budgets*. Cambridge, Inglaterra, The University Press, 1955.

considerando el resultado obtenido en (7.17). Como $E(\varepsilon_i^2) = \sigma_i^2$ y $E(\varepsilon_i \varepsilon_j) = 0$ ($i < j$), obtenemos

$$\text{Var}(\beta) = \sum_i a_i^2 \sigma_i^2.$$

Ahora tenemos que obtener los valores $a_1, a_2, \dots, a_n$ que minimicen la expresión anterior y, al mismo tiempo, cumplan las condiciones $\sum_i a_i = 0$ y $\sum_i a_i X_i = 1$. Utilizando el método de los multiplicadores de Lagrange, formamos una función

$$H = \sum_i a_i^2 \sigma_i^2 - \lambda_1 \left(\sum_i a_i \right) - \lambda_2 \left(\sum_i a_i X_i - 1 \right),$$

que debe minimizarse respecto de $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 . Derivando H e igualando a cero las primeras derivadas resultantes, obtenemos

$$2a_1\sigma_1^2 - \lambda_1 - \lambda_2 X_1 = 0,$$

$$2a_2\sigma_2^2 - \lambda_1 - \lambda_2 X_2 = 0,$$

$$\vdots$$

$$2a_n\sigma_n^2 - \lambda_1 - \lambda_2 X_n = 0,$$

$$-\sum_i a_i = 0,$$

$$-(\sum_i a_i X_i) + 1 = 0.$$

Ello proporciona ($n+2$) ecuaciones con las incógnitas $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 . Ahora podemos introducir una notación más conveniente escribiendo

$$\frac{1}{\sigma_i^2} = w_i.$$

Por tanto, las primeras n ecuaciones obtenidas se pueden escribir en la forma

$$a_1 = \frac{1}{2}(\lambda_1 w_1 + \lambda_2 w_1 X_1),$$

$$a_2 = \frac{1}{2}(\lambda_1 w_2 + \lambda_2 w_2 X_2),$$

$$\vdots$$

$$a_n = \frac{1}{2}(\lambda_1 w_n + \lambda_2 w_n X_n).$$

Sumándolas, obtenemos

$$\sum_i a_i = \frac{1}{2}(\lambda_1 \sum_i w_i + \lambda_2 \sum_i w_i X_i).$$

Además, si multiplicamos la primera ecuación por X_1 , la segunda por X_2 , la tercera por X_3 , y así sucesivamente, y a continuación las sumamos, obtendremos

$$\sum_i a_i X_i = \frac{1}{2}(\lambda_1 \sum_i w_i X_i + \lambda_2 \sum_i w_i X_i^2).$$

Sustituyendo $\sum_i a_i$ y $\sum_i a_i X_i$ por sus respectivos valores, 0 y 1, obtendremos

$$0 = \frac{1}{2}(\lambda_1 \sum_i w_i + \lambda_2 \sum_i w_i X_i),$$

$$1 = \frac{1}{2}(\lambda_1 \sum_i w_i X_i + \lambda_2 \sum_i w_i X_i^2).$$

La solución para λ_1 y λ_2 es

$$\lambda_1 = \frac{-2 \sum_i w_i X_i}{(\sum_i w_i)(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2},$$

$$\lambda_2 = \frac{2 \sum_i w_i}{(\sum_i w_i)(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2}.$$

Estas expresiones se pueden sustituir en las soluciones correspondientes a $a_1, a_2, \dots, a_n$,

$$a_i = \frac{1}{2}(\lambda_1 w_i + \lambda_2 w_i X_i) \quad (i = 1, 2, \dots, n),$$

para obtener

$$a_i = \frac{-w_i(\sum_i w_i X_i) + w_i X_i(\sum_i w_i)}{(\sum_i w_i)(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2}.$$

Estas son las constantes que minimizan la variancia de $\tilde{\beta}$ y, al mismo tiempo, lo convierten en un estimador insesgado. Sustituyendo a_i por su valor en la expresión $\tilde{\beta} = \sum_i a_i Y_i$ da como resultado

$$(8.1) \quad \tilde{\beta} = \frac{(\sum_i w_i)(\sum_i w_i X_i Y_i) - (\sum_i w_i X_i)(\sum_i w_i Y_i)}{(\sum_i w_i)(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2},$$

que es, obviamente, una fórmula distinta de la correspondiente al estimador minimocuadrático de β . La variancia de $\tilde{\beta}$ es

$$\text{Var}(\tilde{\beta}) = \sum_i a_i^2 Y_i^2 = \frac{(\sum_i w_i Y_i^2) - \frac{(\sum_i w_i Y_i)^2}{(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2}}{(\sum_i w_i)(\sum_i w_i X_i^2) - (\sum_i w_i X_i)^2}.$$

Por un procedimiento similar, podemos obtener el estimador lineal insesgado óptimo de α y su variancia cuando el término de perturbación es heteroscedástico. Este estimador es

$$(8.3) \quad \bar{\alpha} = \frac{\sum w_i Y_i}{\sum w_i} - \beta \left[\frac{\sum w_i X_i}{\sum w_i} \right],$$

que nuevamente es distinto del estimador minimocuadrático de α . La variancia de $\bar{\alpha}$ es

$$(8.4) \quad \text{Var}(\bar{\alpha}) = \frac{\sum w_i X_i^2}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}.$$

Dados estos resultados, debemos llegar a la conclusión de que, cuando no se cumple el supuesto de homoscedasticidad, los estimadores minimocuadráticos de los coeficientes de regresión no son los ELIO. Ello significa que los estimadores minimocuadráticos no poseen la menor variancia dentro de la clase de los estimadores insesgados y que, por tanto, no son eficientes.

Pasando a la consideración de las propiedades asintóticas, podemos analizar si los estimadores minimocuadráticos son consistentes en condiciones de heteroscedasticidad, estableciendo sus límites probabilísticos en la forma presentada en la sección 6.1. Tenemos

$$\begin{aligned} \lim \text{prob } \beta &= \lim \text{prob } \left[\frac{\sum x'_i y'_i}{\sum x'^2_i} \right] = \beta + \lim \text{prob } \left[\frac{\sum x'_i \epsilon_i}{\sum x'^2_i} \right] \\ &= \beta + \frac{\lim \text{prob } (\sum x'_i \epsilon_i / n)}{\lim \text{prob } (\sum x'^2_i / n)} \end{aligned}$$

según el teorema de Slutsky (teorema 17 de la sección 6.1). En el último término del lado derecho, el numerador es igual a cero porque $(\sum x'_i \epsilon_i / n)$ es un estimador consistente de la «covariancia» entre X_i y ϵ_i , que es cero, y el denominador es un número finito distinto de cero, según el supuesto (7.6).

Por tanto,

$$\lim \text{prob } \bar{\beta} = \beta,$$

lo que indica que $\bar{\beta}$ es un estimador consistente de β . Del mismo modo,

$$\begin{aligned} \lim \text{prob } \bar{\alpha} &= \lim \text{prob } (\bar{Y} - \bar{\beta} \bar{X}) = \lim \text{prob } (\alpha + \beta \bar{X} + \epsilon - \beta \bar{X}) \\ &= \alpha + \beta \bar{X} + \lim \text{prob } (\epsilon) - \lim \text{prob } (\beta \bar{X}) = \alpha \end{aligned}$$

Es decir, los estimadores minimocuadráticos de los coeficientes de regresión son consistentes aunque el término de perturbación no sea homoscedástico.

Para descubrir si los estimadores minimocuadráticos son asintóticamente eficientes en condiciones de heteroscedasticidad, tenemos que obtener los estimadores maximoverosímiles apropiados, que sabemos son asintóticamente eficientes. A continuación, comprobaremos si las variancias de dichos estimadores maximoverosímiles equivalen asintóticamente a las de los estimadores minimocuadráticos. De no ser así, estos últimos no serán asintóticamente eficientes. Expresando la función de verosimilitud como en (7.26), pero permitiendo la existencia de heteroscedasticidad, obtenemos

$$(8.5) \quad L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log \sigma_i^2 - \frac{1}{2} \sum_{i=1}^n \left[\frac{Y_i - \alpha - \beta X_i}{\sigma_i} \right]^2.$$

Las primeras derivadas de L respecto de α y β son

$$\frac{\partial L}{\partial \alpha} = \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{\sigma_i^2} \right],$$

$$\frac{\partial L}{\partial \beta} = \sum_i \left[\frac{X_i (Y_i - \alpha - \beta X_i)}{\sigma_i^2} \right].$$

Igualando a cero estas derivadas y resolviendo las ecuaciones resultantes respecto de los estimadores de α y β da como resultado las fórmulas (8.1) y (8.3) de los estimadores lineales insesgados óptimos. Por tanto, las variancias de los estimadores maximoverosímiles también son iguales a las de los estimadores lineales insesgados óptimos. Estas son las indicadas en las fórmulas (8.2) y (8.4), y tenemos que comparar estas variancias con las de los estimadores minimocuadráticos. Cuando existe heteroscedasticidad, la variancia del estimador minimocuadrático de β es

$$\begin{aligned} (8.6) \quad \text{Var}(\bar{\beta}) &= E(\bar{\beta} - \beta)^2 = E \left[\frac{\sum x'_i \epsilon_i}{\sum x'^2_i} \right]^2 \\ &= E \left[\frac{\sum x'^2_i \epsilon_i^2}{(\sum x'^2_i)^2} \right] + 2E \left[\frac{x'_i \epsilon_i x'_j \epsilon_j}{(\sum x'^2_i)^2} \right] \\ &= \frac{\sum x'^2_i \sigma_i^2}{(\sum x'^2_i)^2} = \frac{\sum (x'^2_i / w_i)}{(\sum x'^2_i)^2}. \end{aligned}$$

Esta variancia es claramente distinta de la variancia de $\bar{\beta}$ presentada en la fórmula (8.2), sea cual sea el tamaño de la muestra. Se puede obtener una conclusión similar respecto a la variancia del estimador minimocuadrático de α . Por tanto, como las variancias de los estimadores minimocuadráticos no son asintóticamente equivalentes a las de los estimadores maximoverosímiles, cuando la perturbación no es homoscedástica los primeros no son asintóticamente eficientes.

Propiedades de las variancias estimadas de los estimadores minimocuadráticos

Hemos visto que, en condiciones de heteroscedasticidad, los estimadores minimocuadráticos de los coeficientes de regresión son insesgados y consistentes pero no eficientes ni asintóticamente eficientes. Por tanto, si el término de perturbación es heteroscedástico y no lo sabemos (o lo sabemos, pero prescindimos de ello) y usamos las fórmulas minimocuadráticas, los estimadores resultantes tendrán aún algunas de las propiedades deseables. Pero cuando se trata de utilizar estos estimadores para la contrastación de hipótesis o para contruir intervalos de confianza, no sólo es preciso que los estimadores sean insesgados, sino que también lo sean sus variancias estimadas. De lo contrario, las contrastaciones no son válidas y los intervalos de confianza contruidos resultan incorrectos. Por lo tanto, el problema siguiente se refiere a la sesgadez o insesgadez de las variancias estimadas obtenidas a partir de las fórmulas convencionales de los estimadores minimocuadráticos. La fórmula convencional para calcular la variancia del estimador minimocuadrático de la pendiente de la regresión, $\hat{\beta}$, aparece en la expresión (7.37) en la forma

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum x_i'^2},$$

siendo $s^2 = [\sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2] / (n - 2)$. En condiciones de homoscedasticidad, esto es un estimador insesgado de la variancia de $\hat{\beta}$, y queremos saber si la propiedad de insesgadez de $s_{\hat{\beta}}^2$ se conserva cuando no se cumple el supuesto de homoscedasticidad. Para responder a esta pregunta tenemos que obtener la esperanza matemática de s^2 . Tenemos

$$\begin{aligned} E(s^2) &= E \left[\frac{1}{n-2} \right] \sum_i [\alpha + \beta X_i + \varepsilon_i - \hat{\alpha} - \hat{\beta} X_i]^2 \\ &= \frac{1}{n-2} \sum_i E[-(\hat{\alpha} - \alpha) - (\hat{\beta} - \beta)X_i + \varepsilon_i]^2. \end{aligned}$$

Sustituyendo $(\hat{\alpha} - \alpha)$ por su valor en la expresión (7.30) obtenemos

$$\begin{aligned} E(s^2) &= \frac{1}{n-2} \sum_i E[-(\hat{\beta} - \beta)x_i' + \varepsilon_i']^2 \\ &= \frac{1}{n-2} [E(\hat{\beta} - \beta)^2 \sum_i x_i'^2 + E \sum_i \varepsilon_i'^2 - 2E(\hat{\beta} - \beta) \sum_i x_i' \varepsilon_i']. \end{aligned}$$

Ahora bien, cuando existe heteroscedasticidad

$$E(\hat{\beta} - \beta)^2 = \frac{\sum x_i'^2 \sigma_i^2}{(\sum x_i'^2)^2}$$

en (8.6). Además,

$$\begin{aligned} E(\varepsilon_i'^2) &= E(\varepsilon_i^2) + E(\bar{\varepsilon})^2 - 2E(\varepsilon_i \bar{\varepsilon}) = E(\varepsilon_i^2) + \frac{1}{n^2} E \left[\sum_i \varepsilon_i^2 + 2 \sum_{i < j} \varepsilon_i \varepsilon_j \right] \\ &\quad - \frac{2}{n} E \varepsilon_i (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_i + \dots + \varepsilon_n) \\ &= \sigma_i^2 + \frac{\sum \sigma_i^2}{n^2} - \frac{2\sigma_i^2}{n} = \left[\frac{n-2}{n} \right] \sigma_i^2 + \frac{\sum \sigma_i^2}{n^2}. \end{aligned}$$

Por último,

$$\begin{aligned} E(\hat{\beta} - \beta) \sum_i x_i' \varepsilon_i &= E \left[\frac{\sum x_i' \varepsilon_i}{\sum x_i'^2} \right] (\sum x_i' \varepsilon_i) = E \left[\frac{\sum x_i' \varepsilon_i}{\sum x_i'^2} \right]^2 (\sum x_i'^2) \\ &= [\text{Var}(\hat{\beta})] (\sum x_i'^2) = \left[\frac{\sum x_i'^2 \sigma_i^2}{(\sum x_i'^2)^2} \right] (\sum x_i'^2). \end{aligned}$$

Sustituyendo estos resultados en la expresión correspondiente a $E(s^2)$, obtenemos

$$\begin{aligned} (8.7) \quad E(s^2) &= \frac{1}{n-2} \left\{ \left[\frac{\sum x_i'^2 \sigma_i^2}{(\sum x_i'^2)^2} \right] (\sum x_i'^2) + \sum \left[\sigma_i^2 + \frac{\sum \sigma_i^2}{n^2} - \frac{2\sigma_i^2}{n} \right] \right. \\ &\quad \left. - 2 \left[\frac{\sum x_i'^2 \sigma_i^2}{(\sum x_i'^2)^2} \right] (\sum x_i'^2) \right\} \\ &= \frac{1}{n-2} \left[-\frac{\sum x_i'^2 \sigma_i^2}{\sum x_i'^2} + \sum \sigma_i^2 + \frac{\sum \sigma_i^2}{n} - \frac{2 \sum \sigma_i^2}{n} \right] \\ &= \frac{-n(\sum x_i'^2 \sigma_i^2) + (n-1)(\sum x_i'^2)(\sum \sigma_i^2)}{n(n-2)(\sum x_i'^2)} \end{aligned}$$

y

$$(8.8) \quad E(s_{\hat{\beta}}^2) = \frac{E(s^2)}{\sum x_i'^2} = \frac{-n(\sum x_i'^2 \sigma_i^2) + (n-1)(\sum x_i'^2)(\sum \sigma_i^2)}{n(n-2)(\sum x_i'^2)^2}.$$

Como esta expresión no es igual a la expresión correspondiente a la variancia de $\hat{\beta}$ presentada en (8.6), tenemos que llegar a la conclusión de que la variancia de $\hat{\beta}$ calculada de forma convencional está *sesgada* cuando la perturbación es heteroscedástica, y puede obtenerse una conclusión similar respecto de $\hat{\alpha}$.

La consecuencia del resultado anterior es que, si utilizamos los estimadores minimocuadráticos de los coeficientes de regresión cuando no se satisface el supuesto de homoscedasticidad, los límites de confianza y las contrastaciones de hipótesis obtenidas en el capítulo 7 no son de aplicación en este caso.

Ello significa que si efectuamos el análisis de regresión basándonos en la idea equivocada de que la perturbación es homoscedástica, las inferencias acerca de los coeficientes de la población serán incorrectos —es decir, los intervalos de confianza y las regiones de aceptación calculadas serán erróneas—. En este caso, sería interesante saber la dirección del error, ya que, de ser así, podríamos decir si los intervalos de confianza y las regiones de aceptación incorrectas son probablemente más amplias y más estrechas que las correctas. Podemos obtener la respuesta calculando la dirección del sesgo de la variancia calculada. Si el sesgo es positivo, las regiones de aceptación y los intervalos incorrectos serán más amplios que los correctos; si el sesgo es negativo, serán más estrechos. El sesgo viene dado por

$$(8.9) \quad E(s^2) - \text{Var}(\hat{\beta}) = \frac{-n(\sum x_i^2 \sigma_i^2) + (n-1)(\sum x_i^2)(\sum \sigma_i^2)}{n(n-2)(\sum x_i^2)^2} - \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \\ = \frac{-n(n-1)(\sum x_i^2 \sigma_i^2) + (n-1)(\sum x_i^2)(\sum \sigma_i^2)}{n(n-2)(\sum x_i^2)^2}.$$

La dirección del sesgo depende del signo de esta expresión. Para $n > 2$, el denominador será siempre positivo y, por tanto, el signo decisivo será el del numerador. Este signo depende de la dirección de la asociación entre x_i^2 y σ_i^2 . Puede mostrarse —estableciendo $\sigma_i^2 = a + bx_i^2$ ($b > 0$) y efectuando la sustitución correspondiente en la fórmula (8.9)— que cuando x_i^2 y σ_i^2 se hallan asociadas positivamente, el sesgo es negativo. En estos casos, el uso de los errores standard calculado de forma convencional tenderá a producir regiones de aceptación e intervalos de confianza más estrechos que los correctos, lo cual significa que los estimadores se considerarán poseedores de una mayor precisión de la que está justificada por el nivel de confianza elegido, y que la probabilidad de rechazar la hipótesis nula será mayor que la indicada por el nivel de significación que se mencione.

Supuestos acerca de σ_i^2

Una vez examinadas las propiedades de los estimadores minimocuadráticos de los coeficientes de regresión en condiciones de heteroscedasticidad, pasamos a considerar los métodos de estimación alternativos. El lector recordará que ya hemos presentado fórmulas de estimación para el caso de heteroscedasticidad en esta misma sección. Estas fórmulas, presentadas en (8.1) y (8.3), satisfacen a la vez los principios de la máxima verosimilitud y de los ELIO. Ahora podemos considerar un procedimiento alternativo y más sencillo para obtener dichas fórmulas. La ecuación de regresión es

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

cuando existe heteroscedasticidad, $\varepsilon_i \sim N(0, \sigma_i^2)$. Escribamos ahora

$$\varepsilon_i = \frac{u_i}{\sigma \sqrt{w_i}},$$

en la que $u_i \sim N(0, \sigma^2)$ y w_i son cantidades no estocásticas. En este caso,

$$\sigma_i^2 = \frac{E(u_i^2)}{\sigma^2 w_i} = \frac{\sigma^2}{\sigma^2 w_i} = \frac{1}{w_i}.$$

La ecuación de regresión puede escribirse, pues, en la forma siguiente

$$Y_i = \alpha + \beta X_i + \frac{u_i}{\sigma \sqrt{w_i}}.$$

Si multiplicamos los dos lados de esta ecuación por $\sigma \sqrt{w_i}$, obtenemos

$$(8.10) \quad (Y_i \sigma \sqrt{w_i}) = \alpha(\sigma \sqrt{w_i}) + \beta(X_i \sigma \sqrt{w_i}) + u_i.$$

Podemos considerar a la ecuación (8.10) como una ecuación de regresión en la que la variable dependiente es $(Y_i \sigma \sqrt{w_i})$, y que incluye dos variables explicativas, $(\sigma \sqrt{w_i})$ y $(X_i \sigma \sqrt{w_i})$. Obsérvese que dichas variables sólo se pueden medir si se conoce el valor de $\sigma \sqrt{w_i}$, y que la ordenada en el origen de la ecuación de regresión es cero. La transformación que da como resultado la ecuación (8.10) tiene por objeto lograr que la ecuación de regresión satisfaga la condición de homoscedasticidad, ya que la nueva perturbación u_i posee una variancia constante. Si aplicamos ahora el principio de los mínimos cuadrados, obtendremos las «ecuaciones normales minimocuadráticas» siguientes (ver nota 4, pág. 243)

$$(8.11) \quad \sum_i (w_i Y_i) = \bar{\alpha} \sum_i (w_i) + \bar{\beta} \sum_i (w_i X_i) \\ \sum_i (w_i X_i Y_i) = \bar{\alpha} \sum_i (w_i X_i) + \bar{\beta} \sum_i (w_i X_i^2)$$

ya que el término σ^2 desaparece. Despejando $\bar{\alpha}$ y $\bar{\beta}$ en estas ecuaciones, obtenemos las fórmulas (8.1) y (8.3). Puede ser interesante mencionar que estas fórmulas se llaman a veces fórmulas de los «mínimos cuadrados ponderados», ya que, en realidad, se dan ponderaciones distintas a las observaciones muestrales.

Las fórmulas de estimación (8.1) y (8.3) presentan la dificultad de incluir en ellos los términos σ_i^2 (o sus recíprocos w_i), que generalmente no se conocen. Por tanto, estas expresiones no pueden ser consideradas estimadores, ya que no son susceptibles de evaluación. En muchos casos es posible superar

esta dificultad, ya sea efectuando determinados supuestos acerca de σ_i^2 , o estimando estas variancias a partir de la muestra. En primer lugar, estudiaremos las situaciones en las que se puede disponer de información adicional acerca de σ_i^2 . Esta información toma a menudo la forma de un supuesto que señala que σ_i^2 está asociada a alguna variable a la que podemos llamar Z_i . Por ejemplo, en el caso de la función microeconómica de consumo, la variancia de la perturbación se considera a menudo asociada positivamente con el nivel de renta; en tal caso, la variable explicativa de la función de regresión, X_i (renta), tomaría el papel de Z_i . Un supuesto alternativo, aunque similar, sería el de que la variancia de la perturbación está asociada positivamente con el nivel medio de consumo, en cuyo caso sería $E(Y_i)$ la que desempeñaría el papel de Z_i . Puede ocurrir también que se considere que los cambios en la variancia de la perturbación están asociados a los cambios en alguna variable «externa», como, por ejemplo, el tamaño de la familia. La utilización de Z_i nos proporciona simplemente un procedimiento para formular el supuesto acerca de σ_i^2 de una forma muy general, pero para que este supuesto resulte operativo debemos especificar la forma de la asociación. Un procedimiento muy conveniente para ello y que podemos considerar muy plausible es

$$(8.12) \quad \sigma_i^2 = \sigma^2 Z_i^\delta,$$

que incluye dos parámetros, σ^2 y δ . El parámetro δ , que mide el grado de heteroscedasticidad, tiene particular importancia: cuanto menor sea su magnitud, más pequeñas serán las diferencias entre las variancias individuales y, cuando $\delta = 0$, el modelo es homoscedástico.

Puede ocurrir que se desconozcan los dos parámetros de la ecuación (8.12), en cuyo caso deben estimarse conjuntamente con los coeficientes de regresión α y β , o que puede especificarse a priori el valor de uno de ellos por lo menos. Por ejemplo, a veces se supone que el valor de δ es 2, ya que ello hace que la desviación standard de la perturbación sea proporcional a Z_i . Estudiaremos en primer lugar el problema de la estimación en general y después consideraremos algunos casos especiales. El modelo de regresión completo para este tipo de condición heteroscedástica es, pues,

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma_i^2),$$

$$\sigma_i^2 = \sigma^2 Z_i^\delta \quad (\sigma > 0; Z_i > 0).$$

Naturalmente, se supone que la perturbación ε_i no es autorregresiva, y se considera que tanto X_i como Z_i son no estocásticas. En este caso, podemos obtener

los estadísticos máximoverosimiles de α , β , σ^2 y δ substituyendo simplemente la expresión de σ_i^2 en la función de verosimilitud (8.5), que se convierte en

$$(8.13) \quad L = -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^n (\log \sigma^2 + \delta \log Z_i) - \frac{1}{2} \sum_{i=1}^n \left[\frac{Y_i - \alpha - \beta X_i}{\sigma Z_i^{\delta/2}} \right]^2.$$

Las primeras derivadas de L son

$$(8.14) \quad \begin{aligned} \frac{\partial L}{\partial \alpha} &= \frac{1}{\sigma^2} \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{Z_i^\delta} \right] \\ \frac{\partial L}{\partial \beta} &= \frac{1}{\sigma^2} \sum_i \left[\frac{(Y_i - \alpha - \beta X_i) X_i}{Z_i^\delta} \right] \\ \frac{\partial L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{Z_i^{\delta/2}} \right]^2 \\ \frac{\partial L}{\partial \delta} &= -\frac{1}{2} \sum_i \log Z_i + \frac{1}{2\sigma^2} \sum_i \left[\frac{(Y_i - \alpha - \beta X_i)^2 \log Z_i}{Z_i^\delta} \right]. \end{aligned}$$

Igualando a cero estas derivadas obtenemos cuatro ecuaciones a partir de las cuales podemos obtener los valores de las cuatro incógnitas α , β , σ^2 y δ . Sin embargo, estas ecuaciones son muy complejas y pueden ser muy difíciles de resolver.

Consideremos ahora algunos casos especiales de la expresión (8.12) que se han utilizado en aplicaciones prácticas.

Caso A: $\delta = 2$ y $Z_i = X_i$. En este caso, $\sigma^2 = \sigma^2 X_i^2$; es decir, se supone que la variancia del término de perturbación es proporcional al cuadrado de la variable explicativa. La función de verosimilitud (8.13) sólo contiene ahora tres parámetros desconocidos α , β y σ^2 . Podemos substituir δ y Z_i por sus valores en las tres primeras ecuaciones de (8.14), igualarlas a cero y resolverlas, obteniendo los valores de las tres incógnitas. Puede verse fácilmente que la solución es exactamente igual a la obtenida en las ecuaciones (8.1) y (8.3), excepto por el hecho de que, en lugar de w_i , tenemos ahora $1/(\sigma^2 X_i^2)$. Es decir, tenemos

$$(8.15) \quad \hat{\beta}_A = \frac{[\sum (1/X_i^2)][\sum (Y_i/X_i)] - [\sum (1/X_i)][\sum (Y_i/X_i^2)]}{n \sum (1/X_i^2) - [\sum (1/X_i)]^2};$$

$$(8.16) \quad \hat{\alpha}_A = \frac{\sum (Y_i/X_i^2)}{\sum (1/X_i^2)} - \hat{\beta}_A \left[\frac{\sum (1/X_i)}{\sum (1/X_i^2)} \right],$$

en las que el subíndice A indica que estos estimadores se han obtenido para la especificación del caso A. Si la especificación del modelo es correcta, estos estimadores son ELIO, consistentes, y asintóticamente eficientes. El estimador de σ^2 es

$$(8.17) \quad \hat{\sigma}_A^2 = \frac{1}{n} \sum_i \left[\left(\frac{Y_i}{X_i} \right) - \hat{\alpha}_A \left(\frac{1}{X_i} \right) - \hat{\beta}_A \right]^2$$

$$= \frac{1}{n} \left[\sum_i \left(\frac{Y_i}{X_i} \right)^2 - \hat{\alpha}_A \sum_i \left(\frac{Y_i}{X_i^2} \right) - \hat{\beta}_A \sum_i \left(\frac{Y_i}{X_i} \right) \right].$$

Este estimador es consistente y asintóticamente eficiente, pero no es insesgado. Un estimador insesgado de σ^2 , al que podemos llamar s_A^2 , es

$$(8.18) \quad s_A^2 = \left[\frac{n}{n-2} \right] \hat{\sigma}_A^2.$$

Las variancias de $\hat{\alpha}_A$ y $\hat{\beta}_A$ vienen dadas por las ecuaciones (8.2) y (8.4), sustituyendo w_i por $1/(\sigma^2 X_i^2)$, y se pueden obtener estimadores insesgados de estas variancias sustituyendo σ^2 por s_A^2 . El resultado es

$$(8.19) \quad s_{\hat{\beta}_A}^2 = \frac{s_A^2 \sum (1/X_i^2)}{n \sum (1/X_i^2) - [\sum (1/X_i)]^2}$$

y

$$(8.20) \quad s_{\hat{\alpha}_A}^2 = \frac{n s_A^2}{n \sum (1/X_i^2) - [\sum (1/X_i)]^2}.$$

EJEMPLO. Consideremos los datos muestrales siguientes, correspondientes a la renta y los gastos anuales en vestido, obtenidos a partir de una muestra de 20 familias

Renta en dólares	Número de familias	Gasto en vestido en dólares
2000	8	160, 160, 180, 200, 210, 220, 230, 250
4000	7	200, 220, 230, 300, 310, 340, 350
6000	5	300, 300, 400, 450, 540

En la figura 8-1 se presentan gráficamente estas observaciones. Se postula que la relación entre el gasto en vestido y la renta es

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

siendo Y = gasto en vestido, X = renta, ϵ = perturbación aleatoria, y el subíndice i indica la familia i -ésima. Se considera que la variable explicativa X_i no es estocástica, y se supone que la perturbación ϵ_i es una variable aleatoria no autorre-

correlada, que se distribuye normalmente con media cero y variancia σ_i^2 . Los estimadores minimocuadráticos ordinarios de los coeficientes de regresión son

$$\hat{\beta} = \frac{\sum Y_i X_i}{\sum X_i^2} = \frac{2,425,000}{50,200,000} = 0.0483,$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = 277.5 - 0.0483 \times 3700 = 98.79,$$

y el coeficiente de determinación es

$$R^2 = \frac{\beta \sum X_i Y_i}{\sum Y_i^2} = \frac{0.0483 \times 2,425,000}{190,975} = 0.6133.$$

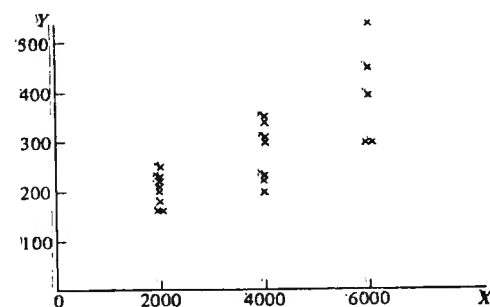


Figura 8-1

Como el modelo es heteroscedástico, los estimadores minimocuadráticos son insesgados, pero no son eficientes ni asintóticamente eficientes. Sus correspondientes errores standard, calculados de la forma convencional, son sesgados y, por tanto, no los presentamos. Obsérvese que el valor de R^2 que hemos presentado anteriormente es el valor máximo para esta muestra concreta, ya que la recta de regresión minimocuadrática proporciona, por definición, el mejor ajuste posible a una recta.

Ahora bien, si suponemos que

$$\sigma_i^2 = \sigma^2 X_i^2,$$

podemos obtener estimadores eficientes de los coeficientes de regresión utilizando las fórmulas (8.15) y (8.16). Estas fórmulas utilizan las cantidades siguientes, que se han calculado a partir de los datos muestrales:

$$\sum_i \left[\frac{1}{X_i} \right] = \frac{8}{2000} + \frac{7}{4000} + \frac{5}{6000} = 0.006583;$$

$$\sum_i \left[\frac{1}{X_i^2} \right] = \frac{8}{2000^2} + \frac{7}{4000^2} + \frac{5}{6000^2} = 0.00002567;$$

$$\sum_i \left[\frac{Y_i}{X_i} \right] = \frac{160}{2000} + \frac{160}{2000} + \dots + \frac{540}{6000} = 1.624167;$$

$$\sum_i \left[\frac{Y_i}{X_i^2} \right] = \frac{160}{2000^2} + \frac{160}{2000^2} + \dots + \frac{540}{6000^2} = 0.000579630.$$

También podemos observar que

$$\sum_i \left[\frac{Y_i}{X_i} \right]^2 = 0.141388.$$

Efectuando las sustituciones necesarias obtenemos

$$\hat{\beta}_A = 0.0451 \quad \text{y} \quad \hat{\alpha}_A = 109.82.$$

Además, utilizando las fórmulas (8.19) y (8.20), obtenemos estimaciones de los errores standard de $\hat{\alpha}_A$ y $\hat{\beta}_A$:

$$s_{\hat{\beta}_A}^2 = 0.00017932, \quad s_{\hat{\beta}_A} = 0.0134,$$

$$s_{\hat{\alpha}_A}^2 = 1,391.7020, \quad s_{\hat{\alpha}_A} = 37.30$$

Por último, $R^2 = 1 - \frac{\sum (Y_i - \hat{\alpha}_A - \hat{\beta}_A X_i)^2}{\sum Y_i^2} = 1 - \frac{74,364}{190,975} = 0.6106.$

Podemos resumir estos resultados en la forma siguiente:

$$Y_i = 109.82 + 0.0451 X_i + e_i, \quad R^2 = 0.6106. \\ (37.30) \quad (0.0134)$$

Como era de esperar, el valor de R^2 es menor que el obtenido en la regresión minimocuadrática, pero la diferencia es muy pequeña. Obsérvese que la estimación implícita de la elasticidad-renta de la demanda de vestido correspondiente a la renta media es

$$\hat{\eta}_{INC} = \hat{\beta}_A \left[\frac{X}{\bar{Y}} \right] = 0.6013,$$

lo cual sugiere que la demanda de vestido es inelástica respecto a la renta.

Caso B: $\delta = 2$ y $Z_i = E(Y_i)$. En este caso, se supone que la variancia de la perturbación es proporcional al cuadrado de la media de Y_i . Es decir, suponemos que

$$\sigma_i^2 = \sigma^2 [E(Y_i)]^2 = \sigma^2 (\alpha + \beta X_i)^2$$

La función de verosimilitud (8.13) se convierte aquí en

$$(8.21) \quad L = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \sum_i \log(\alpha + \beta X_i) \\ - \frac{1}{2\sigma^2} \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{(\alpha + \beta X_i)} \right]^2.$$

que en el caso A, esta función incluye solamente tres parámetros desconocidos: α , β y σ^2 , y podríamos tomar las primeras derivadas de L respecto a cada uno de los tres parámetros, igualarlas a cero y resolverlas obteniendo los valores de las tres incógnitas.⁴ Sin embargo, las ecuaciones resultantes son muy complejas y su solución resulta muy difícil. Podemos simplificar el problema sustituyendo $E(Y_i)$ por $\hat{Y}_i$, que representa los valores de Y calculados a partir de la regresión minimocuadrática. Es decir, consideramos que la desviación standard de la perturbación es proporcional a la distancia entre la recta de regresión minimocuadrática y el eje horizontal. Como $\hat{Y}_i$ es un estimador consistente de $E(Y_i)$, podemos escribir

$$\sigma_i = \sigma E(Y_i)$$

en la forma

$$\sigma_i = \sigma \lim \text{prob } \hat{Y}_i$$

y, por tanto, al escribir

$$\sigma_i = \sigma \hat{Y}_i$$

no se altera la consistencia de los correspondientes estimadores de los coeficientes de regresión. En sentido estricto, estos estimadores no son *exactamente* estimadores maximoverosímiles, pero sirven como aproximación a ellos.

Si utilizamos la simplificación anterior, podemos hacer uso de las fórmulas (8.1) y (8.3) para estimar α y β , a condición de que reemplazemos w_i por $1/(\sigma^2 \hat{Y}_i^2)$. De esta forma, obtenemos

$$(8.22) \quad \hat{\beta}_B = \frac{\sum (1/\hat{Y}_i^2) \sum (X_i Y_i / \hat{Y}_i^2) - \sum (X_i / \hat{Y}_i^2) \sum (Y_i / \hat{Y}_i^2)}{\sum (1/\hat{Y}_i^2) \sum (X_i^2 / \hat{Y}_i^2) - [\sum (X_i / \hat{Y}_i^2)]^2},$$

y

$$(8.23) \quad \hat{\alpha}_B = \frac{\sum (Y_i / \hat{Y}_i^2)}{\sum (1/\hat{Y}_i^2)} - \hat{\beta}_B \left[\frac{\sum (X_i / \hat{Y}_i^2)}{\sum (1/\hat{Y}_i^2)} \right],$$

en las que el subíndice B indica que los estimadores corresponden a la especificación del caso B. Si la especificación del modelo es correcta, estos estimadores se aproximan a los estimadores maximoverosímiles, de los que se sabe que son consistentes y asintóticamente eficientes. El estimador de σ^2 es

$$(8.24) \quad s_B^2 = \frac{1}{n-2} \left[\sum_i \left(\frac{Y_i^2}{\hat{Y}_i^2} \right) - \hat{\alpha}_B \sum_i \left(\frac{Y_i}{\hat{Y}_i^2} \right) - \hat{\beta}_B \sum_i \left(\frac{X_i Y_i}{\hat{Y}_i^2} \right) \right].$$

⁴ Obsérvese que sería incorrecto sustituir simplemente Z_i (y δ) por sus valores, en las expresiones de las primeras derivadas presentadas en (8.14), ya que éstas se obtuvieron partiendo de la base de que Z_i no incluía ninguno de los parámetros desconocidos. Esto sólo cumple cuando, como en el caso actual, $Z_i = \alpha + \beta X_i$.

Por último, los estimadores de los errores standard de $\hat{\alpha}_B$ y $\hat{\beta}_B$ vienen dados por

$$(8.25) \quad s_{\hat{\beta}_B}^2 = \frac{s_B^2 \sum (1/\hat{Y}_i^2)}{\sum (1/\hat{Y}_i^2) \sum (X_i^2/\hat{Y}_i^2) - [\sum (X_i/\hat{Y}_i^2)]^2}$$

y

$$(8.26) \quad s_{\hat{\alpha}_B}^2 = \frac{s_B^2 \sum (X_i^2/\hat{Y}_i^2)}{\sum (1/\hat{Y}_i^2) \sum (X_i^2/\hat{Y}_i^2) - [\sum (X_i/\hat{Y}_i^2)]^2}.$$

EJEMPLO. En el ejemplo de la relación entre gasto familiar en vestido y renta, presentado al estudiar el caso A, hemos supuesto que $\sigma_i = \sigma X_i$. Cambiemos ahora este supuesto, y consideremos, de acuerdo con el caso B, $\sigma_i = \sigma E(Y_i)$. Al efectuar los cálculos, aproximaremos $E(Y_i)$ por medio de $\hat{Y}_i$, que, como se ha calculado previamente, es

$$\hat{Y}_i = 98.79 + 0.0483 X_i.$$

A partir de los datos muestrales se han calculado las siguientes cantidades:

$$\sum_i \left[\frac{1}{\hat{Y}_i^2} \right] = 0.000322729,$$

$$\sum_i \left[\frac{X_i}{\hat{Y}_i^2} \right] = 0.946107,$$

$$\sum_i \left[\frac{X_i^2}{\hat{Y}_i^2} \right] = 3,271.421,$$

$$\sum_i \left[\frac{Y_i}{\hat{Y}_i^2} \right] = 0.077417,$$

$$\sum_i \left[\frac{Y_i^2}{\hat{Y}_i^2} \right] = 20.503404,$$

$$\sum_i \left[\frac{X_i Y_i}{\hat{Y}_i^2} \right] = 250.091731.$$

Efectuando las sustituciones correspondientes en las fórmulas (8.22) a (8.26), obtenemos

$$\hat{\beta}_B = 0.0465,$$

$$\hat{\alpha}_B = 103.56,$$

$$s_{\hat{\beta}_B}^2 = 0.00009558, \quad s_{\hat{\alpha}_B}^2 = 0.0098,$$

$$s_{\hat{\beta}_B}^2 = 968.9202, \quad s_{\hat{\alpha}_B}^2 = 31.13.$$

$$R^2 = 1 - \frac{\sum (Y_i - \hat{\alpha}_B - \hat{\beta}_B X_i)^2}{\sum Y_i^2} = 1 - \frac{74,063}{190,975} = 0.6122.$$

Resumiendo, los resultados correspondientes al caso B son

$$Y_i = 103.56 + 0.0465 X_i + e_i, \quad R^2 = 0.6122. \\ (31.13) \quad (0.0098)$$

Caso C: $Z_i = X_i$. Este caso es más general que los dos anteriores, ya que no efectuamos especificación alguna a priori respecto al valor de δ . En este caso, el supuesto relativo a la naturaleza de la heteroscedasticidad es

$$\sigma_i^2 = \sigma^2 X_i^\delta,$$

o también $\log \sigma_i = \log \sigma + \frac{\delta}{2} \log X_i$.

La función de verosimilitud (8.13) incluye aquí cuatro parámetros desconocidos, α , β , σ^2 y δ . Como ya se ha mencionado con anterioridad, δ mide el grado de heteroscedasticidad, en el sentido de que, cuanto menor sea el valor de δ , menores serán las diferencias entre los valores individuales de la variancia de la perturbación. En el caso C, contrariamente a lo que ocurre en los casos A y B, dejamos que sean los datos los que nos indiquen cuál parece ser este grado, en lugar de especificarlo a priori. La única especificación que efectuamos previamente es la de que la desviación standard de la perturbación está relacionada con la variable explicativa —y que esta relación es lineal en términos de los logaritmos—. El precio que tendremos que pagar por la mayor generalidad de este caso es el mayor grado de complejidad de la estimación.

El problema de obtener estimaciones maximoverosímiles de los coeficientes de regresión, aunque más complejo que en los dos casos anteriores, puede resolverse utilizando una calculadora electrónica. Una forma relativamente sencilla de hacerlo es la siguiente. Si igualamos a cero cada una de las tres primeras derivadas de la expresión (8.14), sustituyendo Z_i por X_i , y reordenando los términos, obtendremos las tres ecuaciones siguientes:

$$(8.27) \quad \sum_i \left[\frac{Y_i}{X_i^\delta} \right] = \hat{\alpha} \sum_i \left[\frac{1}{X_i^\delta} \right] + \hat{\beta} \sum_i \left[\frac{X_i}{X_i^\delta} \right],$$

$$(8.28) \quad \sum_i \left[\frac{X_i Y_i}{X_i^\delta} \right] = \hat{\alpha} \sum_i \left[\frac{X_i}{X_i^\delta} \right] + \hat{\beta} \sum_i \left[\frac{X_i^2}{X_i^\delta} \right],$$

$$(8.29) \quad \hat{\sigma}^2 = \frac{1}{n} \left[\sum_i \left(\frac{Y_i^2}{X_i^\delta} \right) - \hat{\alpha} \sum_i \left(\frac{Y_i}{X_i^\delta} \right) - \hat{\beta} \sum_i \left(\frac{X_i Y_i}{X_i^\delta} \right) \right].$$

Si se conociera el valor de δ , se podrían obtener los valores de $\hat{\alpha}$, $\hat{\beta}$ y $\hat{\sigma}^2$, a partir de las ecuaciones (8.27) a (8.29). Por tanto, tomando distintos valores de δ obtenemos estimaciones distintas de los tres parámetros restantes. Por este procedimiento podemos obtener estimaciones de α , β y σ^2 correspondientes a $\delta = 0$, $\delta = 0,1$, $\delta = 0,2$, y así sucesivamente hasta que lleguemos a valores de δ que son claramente inadecuados (en la mayor parte de las situaciones prácticas, esto sería quizá 5,0 o 6,0). Para cada valor de δ —y los valores de $\hat{\alpha}$, $\hat{\beta}$ y $\hat{\sigma}^2$ correspondientes— calculamos también el valor de L tal como aparece en la ecuación (8.13). A continuación, seleccionamos de entre todas las soluciones la que proporciona el valor más alto de L . Esta solución maximizará la función de verosimilitud tal como deseábamos. Los errores standard de $\hat{\alpha}$ y $\hat{\beta}$ se pueden estimar utilizando las fórmulas (8.2) y (8.4), sustituyendo w_i por $(1/\hat{\sigma}^2 X_i \hat{\sigma})$. La variancia de $\hat{\sigma}^2$ puede estimarse mediante $s_{\hat{\sigma}^2}^2 = 2/[\sum_i (\log_e X_i)^2]$, en la que $\log_e X_i = 2,302585 \log_{10} X_i$. Esta fórmula representa la variancia asintótica de $\hat{\sigma}^2$ tal como se definió en la pág. 214. Naturalmente, si la precisión obtenida por los intervalos de una décima en el valor de $\hat{\sigma}^2$ no resulta suficiente, podemos elegir intervalos más pequeños para los valores sucesivos de δ próximos a la solución.

Estimación de σ_i^2

Cuando no se efectúa ningún supuesto acerca de la naturaleza de la heteroscedasticidad, es preciso basarse por completo en la información proporcionada por la muestra, estimando las variancias de las perturbaciones a partir de los datos. Como el valor de la variancia de la perturbación puede ser distinto para cada valor concreto de X , necesitamos varias observaciones de la variable dependiente correspondiente a cada valor X_i . Si sólo tuviésemos una observación para cada valor de X , no podríamos, evidentemente, obtener información acerca de la dispersión de todas las observaciones posibles. Supongamos, pues, que la muestra tiene m valores distintos de X , y que para cada X_i tenemos n_i observaciones de la variable dependiente. En este caso, la ecuación de regresión se puede escribir en la forma

$$Y_{ij} = \alpha + \beta X_i + \varepsilon_{ij} \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n_i),$$

$$n = \sum_{i=1}^m n_i,$$

$$X_i \neq X_j \quad (i \neq j).$$

Para ilustrar el nuevo uso de subíndices, podemos usar los datos correspondien-

tes para los gastos familiares en vestido, de la pág. 308. Tenemos pues

$$m = 3,$$

$$n_1 = 8,$$

$$n_2 = 7,$$

$$n_3 = 5,$$

$$n = 20.$$

y, por ejemplo, $Y_{25} = 310$ es el valor de la variable dependiente correspondiente a la familia número 5 del grupo de renta de 4000 dólares. En general, hay que estimar $(m+2)$ parámetros: α , β , σ_1^2 , σ_2^2 , ..., σ_m^2 . La función de verosimilitud correspondiente viene dada por

$$(8.30) \quad L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^m n_i \log \sigma_i^2 - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} \left[\frac{Y_{ij} - \alpha - \beta X_i}{\sigma_i} \right]^2.$$

Derivando L respecto a los parámetros desconocidos, e igualando a cero dichas derivadas obtendríamos $(m+2)$ ecuaciones a partir de las cuales podrían obtenerse los $(m+2)$ valores de los parámetros; sin embargo, esta solución presenta grandes dificultades de cálculo.

Un procedimiento que permite superar dichas dificultades de cálculo consiste en utilizar un procedimiento iterativo, de la forma siguiente.

1. Obtener las estimaciones minimocuadráticas ordinarias de α y β , a las que llamaremos $\hat{\alpha}$ y $\hat{\beta}$. A partir de ellas se obtiene la «primera vuelta» de estimaciones de σ_i^2 , a las que llamaremos $\hat{\sigma}_i^2$. Estas vienen dadas por

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha} - \hat{\beta} X_i)^2.$$

2. Sustituir w_i por $1/\hat{\sigma}_i$ en las fórmulas (8.1) y (8.3), y obtener nuevas estimaciones de α y β , $\hat{\alpha}$ y $\hat{\beta}$. A partir de ellas se obtiene la «segunda vuelta» de estimaciones de σ_i^2 , $\hat{\sigma}_i^2$, que vienen dadas por

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha} - \hat{\beta} X_i)^2.$$

3. Sustituir w_i por $1/\hat{\sigma}_i$ en las fórmulas (8.1) y (8.3), y obtener un nuevo conjunto de estimaciones de α y β . A partir de ellas se obtiene la «tercera vuelta» de estimaciones de σ_i^2 .

Este proceso debe repetirse hasta que los valores de las estimaciones converjan, es decir, hasta que las diferencias entre grupos sucesivos de estimaciones sean

negligibles.⁵ Los errores standard de los coeficientes de regresión estimados se pueden estimar tomando $w_i = 1/\bar{\sigma}_i^2$ (siendo $\bar{\sigma}_i^2$ las estimaciones de σ_i^2 en la «última vuelta») en las fórmulas (8.2) y (8.4).

Este procedimiento iterativo resulta evidentemente muy laborioso, y una alternativa simple consiste en estimar σ_i^2 por medio de

$$(8.31) \quad s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2,$$

$$\text{en la que} \quad \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}.$$

Se puede mostrar fácilmente que s_i^2 es un estimador consistente de σ_i^2 . Sustituyendo w_i por $1/s_i^2$ en (8.1) y (8.3), obtenemos estimaciones de los coeficientes de regresión, y efectuando la misma sustitución en (8.2) y (8.4) obtenemos sus errores standard estimados. El precio a pagar por esta simplificación es que las estimaciones resultantes no son exactamente iguales a las estimaciones maximoverosimiles, sino sólo asintóticamente.

EJEMPLO. Volviendo de nuevo a los datos sobre renta y gasto familiares en vestido, podemos estimar las variancias de la perturbación a partir de los datos, en lugar de efectuar supuestos acerca de ellas. Utilizando la fórmula (8.31) obtenemos

$$s_1^2 = 935.94, \quad s_2^2 = 3183.67, \quad s_3^2 = 8416.00.$$

Las fórmulas (8.1) a (8.4) requieren la obtención de las cantidades siguientes:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^{n_i} w_i &= \sum_{i=1}^m n_i w_i = \frac{n_1}{s_1^2} + \frac{n_2}{s_2^2} + \frac{n_3}{s_3^2} = 0.011340725; \\ \sum_{i=1}^m \sum_{j=1}^{n_i} w_i X_i Y_{ij} &= \sum_{i=1}^m w_i X_i \sum_{j=1}^{n_i} Y_{ij} = \frac{X_1}{s_1^2} \sum_{j=1}^8 Y_{1j} + \frac{X_2}{s_2^2} \sum_{j=1}^7 Y_{2j} + \frac{X_3}{s_3^2} \sum_{j=1}^5 Y_{3j} \\ &= 7,309.240; \\ \sum_{i=1}^m \sum_{j=1}^{n_i} w_i X_i^2 &= \sum_{i=1}^m n_i w_i X_i^2 = \frac{n_1 X_1^2}{s_1^2} + \frac{n_2 X_2^2}{s_2^2} + \frac{n_3 X_3^2}{s_3^2} = 29.455270; \\ \sum_{i=1}^m \sum_{j=1}^{n_i} w_i Y_{ij} &= \sum_{i=1}^m w_i \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{s_1^2} \sum_{j=1}^8 Y_{1j} + \frac{1}{s_2^2} \sum_{j=1}^7 Y_{2j} + \frac{1}{s_3^2} \sum_{j=1}^5 Y_{3j} \\ &= 2.569218; \\ \sum_{i=1}^m \sum_{j=1}^{n_i} w_i X_i^2 &= \sum_{i=1}^m n_i w_i X_i^2 = \frac{n_1 X_1^2}{s_1^2} + \frac{n_2 X_2^2}{s_2^2} + \frac{n_3 X_3^2}{s_3^2} = 90,758.660. \end{aligned}$$

5. La corrección de este procedimiento se basa en dos afirmaciones: a) que el procedimiento da lugar siempre a un proceso convergente, y b) que los valores finales de las estimaciones coinciden con los valores de los estimadores maximoverosimiles. Este autor no conoce la existencia de ninguna prueba formal de estas afirmaciones, pero se han cumplido en todos los casos en que se han probado.

Las estimaciones de los coeficientes de regresión son

$$\hat{\beta} = 0.0446 \quad \text{y} \quad \hat{\alpha} = 110.63$$

$$s_{\hat{\beta}}^2 = 0.00007015, \quad s_{\hat{\alpha}}^2 = 0.0084;$$

$$s_{\hat{\beta}}^2 = 561.4306, \quad s_{\hat{\alpha}}^2 = 23.69.$$

$$\text{Por último,} \quad R^2 = 1 - \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \hat{\alpha} - \hat{\beta} X_i)^2}{\sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2} = 1 - \frac{74,591}{190,975} = 0.6094.$$

En resumen, los resultados son

$$Y_{ij} = 110.63 + 0.0446 X_i + e_{ij}, \quad R^2 = 0.6094. \\ (23.69) \quad (0.0084)$$

Estos resultados son similares a los obtenidos en el caso A y en el caso B al efectuar distintos supuestos acerca de σ_i^2 .

Contrastación de la homoscedasticidad

Hasta este momento, nuestro análisis se ha ocupado de las implicaciones de la heteroscedasticidad en los modelos de regresión lineal. Hemos examinado los efectos de la heteroscedasticidad sobre las propiedades de los estimadores minimocuadráticos ordinarios y sus errores standard calculados en la forma convencional, y hemos estudiado varios estimadores alternativos diseñados para los modelos heteroscedásticos. Sin embargo, si no sabemos si el modelo que estamos investigando es homoscedástico o no, y no queremos efectuar un supuesto a este respecto, es posible que no podamos decidir qué procedimiento de estimación debemos utilizar. En tales casos tenemos que acudir a la información proporcionada por la muestra, y efectuar una contrastación. Concretamente, podemos estar interesados en contrastar la hipótesis nula

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 \quad (m \leq n),$$

en la que m es el número de valores distintos de la variable explicativa. La naturaleza de la contrastación dependerá de la forma en que se especifique la hipótesis sostenida. Si esta última es

$$\sigma_i^2 = \sigma^2 Z_i^0 \quad (i = 1, 2, \dots, m),$$

en la que Z_i es alguna variable no estocástica conocida (por ejemplo, X_i), la hipótesis nula puede presentarse como

$$H_0: \delta = 0.$$

Esto es debido a que si $\delta = 0$, $\sigma^2 = \sigma^2$, lo cual significa que existe homoscedasticidad. Ya hemos estudiado previamente el problema de obtener un estimador maximoverosímil de δ y su error standard estimado. En las muestras grandes, este estimador se distribuirá normalmente (por lo menos, de forma aproximada) de modo que la contrastación de H_0 resulta muy sencilla. Al nivel de significación del 5 %, la región de aceptación de H_0 será

$$-1.96 \leq \frac{\hat{\delta}}{s_{\hat{\delta}}} \leq +1.96,$$

en la que $\hat{\delta}$ es el estimador maximoverosímil de δ , y $s_{\hat{\delta}}$ es su error standard estimado.

Cuando no se efectúa ninguna especificación acerca de la naturaleza de la heteroscedasticidad, se puede efectuar una contrastación de la hipótesis de homoscedasticidad si se dispone de varias observaciones de la variable dependiente para cada valor específico de la variable explicativa. En tal caso, podemos formar el siguiente estadístico de prueba:

$$(8.32) \quad \lambda = \frac{-4.60517 \log M}{1 + N},$$

en la que

$$\log M = \left\{ \sum_i \left[\frac{n_i - 1}{2} \right] \log \left[\frac{n_i s_i^2}{n_i - 1} \right] \right\} - \left\{ \sum_i \left[\frac{n_i - 1}{2} \right] \right\} \left\{ \log \left[\frac{\sum n_i s_i^2}{\sum (n_i - 1)} \right] \right\},$$

$$N = \frac{1}{3(m-1)} \left\{ \sum_i \left[\frac{1}{n_i} \right] - \frac{1}{n} \right\} \quad (i = 1, 2, \dots, m),$$

y s_i^2 se define como en la fórmula (8.31). Esta estadística de prueba se distribuye aproximadamente con una distribución ji-cuadrado con $(m-1)$ grados de libertad.⁶ La región de aceptación de H_0 al nivel de significación del 5 % es en este caso

$$\lambda \leq \chi_{m-1, 0.05}^2.$$

6. Ver Paul G. Hoel, *Introduction to Mathematical Statistics*, 2ª ed., Nueva York, Wiley, 1955, p. 195. Obsérvese que en la definición (8.32) utilizamos logaritmos decimales. J. B. Ramsey en «Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis», *Journal of the Royal Statistical Society*, serie B, vol. 31, 1969, pp. 350-71, ha propuesto una modificación de (8.32) para el caso en que sólo observemos un valor de Y para cada valor concreto de X . También pueden verse algunas contrastaciones especiales en S. M. Goldfeld y R. E. Quandt, «Some Tests for Homoscedasticity», *Journal of the American Statistical Association*, vol. 60, septiembre 1965, pp. 539-47.

En nuestro ejemplo de la renta y gastos familiares tenemos tres valores de X , de modo que el estadístico de prueba λ tendrá una distribución ji-cuadrado con dos grados de libertad. La región de aceptación del 5 % para la hipótesis de homoscedasticidad es pues

$$\lambda \leq 5.991,$$

en la que el valor 5.991 se ha obtenido a partir de la tabla de la ji-cuadrado. Sustituyendo en los datos de la muestra obtenemos

$$\begin{aligned} \log M &= \frac{7}{2} \log \left[\frac{8 \times 935.94}{7} \right] + \frac{6}{2} \log \left[\frac{7 \times 3183.67}{6} \right] + \frac{4}{2} \log \left[\frac{5 \times 8416}{4} \right] \\ &\quad - \left[\frac{7}{2} + \frac{6}{2} + \frac{4}{2} \right] \left\{ \log \left[\frac{8 \times 935.94 + 7 \times 3183.67 + 5 \times 8416}{7 + 6 + 4} \right] \right\} \\ &= -1.46505 \\ N &= \frac{1}{3 \times 2} \left[\frac{1}{8} + \frac{1}{7} + \frac{1}{5} - \frac{1}{20} \right] = 0.06964. \end{aligned}$$

Por tanto,

$$\lambda = \frac{(-4.60517) \times (-1.46505)}{1.06964} = 6.305.$$

Como el valor de λ se halla fuera de la región de aceptación, rechazamos la hipótesis de homoscedasticidad, al nivel de significación del 5 %. La evidencia empírica sugiere que la variancia de la perturbación no es constante.

8.2 Perturbaciones autorregresivas

Según el supuesto (7.5) del modelo clásico de regresión normal lineal, tenemos

$$E(\epsilon_i \epsilon_j) = 0 \text{ para todo } i \neq j.$$

Como se supone que la media de ϵ_i y la de ϵ_j son cero, podemos escribir

$$\text{Cov}(\epsilon_i, \epsilon_j) = E[\epsilon_i - E(\epsilon_i)][\epsilon_j - E(\epsilon_j)] = E(\epsilon_i \epsilon_j) = 0.$$

Esta característica de los términos de perturbación de la regresión se conoce con el nombre de no autorregresión, y algunos autores lo denominan ausencia de correlación serial. Ello implica que la perturbación que tiene lugar en un punto de la observación no está correlacionado con ninguna otra perturbación. Esto significa que cuando se efectúan observaciones a lo largo del tiempo, el

efecto de la perturbación que ocurre en un período no se transmite a ningún otro período. Por ejemplo, en el estudio de la relación entre inputs y outputs de una empresa o industria a partir de observaciones mensuales, la no autorregresión de la perturbación implica que el efecto de una avería en la maquinaria es estrictamente temporal en el sentido de que sólo se ve afectada la producción del mes en curso. En el caso de las observaciones de corte transversal, como las de la renta y gastos de distintas familias, el supuesto de no autorregresión significa que, si la pauta de gasto de una familia se ve «perturbada» —por ejemplo, por la visita de un familiar—, ello no afecta a la pauta de gasto de ninguna otra familia.

Nuestra tarea actual consiste en considerar la plausibilidad del supuesto de no autorregresión, examinar las consecuencias de su violación sobre las propiedades de los estimadores minimocuadráticos y obtener métodos de estimación alternativos en caso de que sean necesarios. En la sección 8.1 se ha afirmado que el supuesto de homocedasticidad resulta razonable frecuentemente en el caso de los modelos que describen la conducta de los agregados a lo largo del tiempo, pero que su plausibilidad es dudosa cuando se estiman relaciones microeconómicas a partir de datos de corte transversal. Aquí, cuando se trata del supuesto de no autorregresión, el argumento es precisamente lo contrario. La afirmación usual es la de que el supuesto de no autorregresión se infringe más a menudo en el caso de las relaciones estimadas a partir de datos de series temporales que en el caso de las relaciones estimadas a partir de datos de corte transversal. Esta afirmación se basa principalmente en la interpretación del término de perturbación como el resumen de un gran número de factores aleatorios e independientes que entran en la relación estudiada, pero no son susceptibles de medición. En tal caso, se puede suponer que el efecto de los factores que operen en un período se transmitirá, en parte, a los períodos siguientes. Esto parece más probable que el hecho de que los efectos se transmitan de una familia, empresa, u otra unidad económica similar, a otra —suponiendo, naturalmente, que las unidades se han seleccionado aleatoriamente a partir de una población grande—.

La autorregresión de los términos de perturbación puede compararse al efecto sonoro al pulsar una cuerda musical: el sonido es máximo en el momento del impacto, pero no se interrumpe inmediatamente, sino que persiste durante un tiempo hasta que, por último, desaparece. Esta puede ser también la característica de las perturbaciones, ya que su efecto puede prolongarse por algún tiempo después de que hayan ocurrido. Pero mientras persiste el efecto de una perturbación, ocurren otras perturbaciones, como si la cuerda musical se pulsara una y otra vez, algunas veces más intensamente que otras. Cuanto más corto es el período de tiempo entre las pulsaciones, mayor es la probabilidad de que aún se oiga el sonido anterior. Del mismo modo, cuanto más cortos son los períodos de las observaciones individuales, mayor es la probabilidad de que nos encontremos con perturbaciones autorregresivas. Por tanto,

las predicciones acerca de la existencia de autorregresión serán mayores cuando se basen en observaciones mensuales o trimestrales que cuando los datos correspondan a intervalos anuales.

La presunción de que las relaciones estimadas a partir de observaciones a lo largo del tiempo incluyen perturbaciones autorregresivas es tan corriente que, en todos los estudios de la autorregresión, las variables tienen el subíndice t (representando «tiempo») en lugar del subíndice i que se utiliza en el caso general. En nuestro estudio seguiremos esta costumbre y, por tanto, si las perturbaciones son autorregresivas, tenemos

$$E(\epsilon_t \epsilon_{t-s}) \neq 0 \quad (t > s).$$

Esta expresión implica que la perturbación correspondiente al tiempo t está relacionada con la perturbación correspondiente al tiempo $(t - s)$. Las consecuencias de la autorregresión en la estimación son difíciles de descubrir, a menos que se especifique con mayor precisión su naturaleza. La mayor parte de estudios en este sentido se basan en el supuesto de que

$$(8.33) \quad E(\epsilon_t \epsilon_{t-s}) = \rho^s \sigma^2 \quad (s < t)$$

$$\text{Cov}(\epsilon_t, \epsilon_{t-s}) = \rho^s \sigma^2,$$

o

en la que ρ es un parámetro cuyo valor es menor que $+1$ y mayor que -1 , y σ^2 es la variancia de ϵ_t como en el caso anterior.⁷ Las covariancias sucesivas de las perturbaciones son

$$\text{Cov}(\epsilon_t, \epsilon_{t-1}) = \rho \sigma^2,$$

$$\text{Cov}(\epsilon_t, \epsilon_{t-2}) = \rho^2 \sigma^2,$$

$$\text{Cov}(\epsilon_t, \epsilon_{t-3}) = \rho^3 \sigma^2,$$

⋮

$$\text{Cov}(\epsilon_t, \epsilon_1) = \rho^{t-1} \sigma^2.$$

Si el valor de ρ es igual a algún número entre 0 y 1, ρ^2 será menor que ρ , ρ^3 menor que ρ^2 , y así sucesivamente. Ello significa que, cuanto mayor sea el número de períodos entre dos perturbaciones, menor será su covariancia. Si ρ es igual a cero, cada covariancia será igual a cero, de modo que se mantiene el supuesto de no autorregresión. Si ρ está comprendida entre -1 y 0, los valores de ρ , ρ^2 , ρ^3 , ... disminuirán en valor absoluto, pero alternarán de signo y lo mismo ocurrirá con las covariancias respectivas.

⁷ El supuesto $|\rho| = 1$ significaría que las covariancias no disminuyen al aumentar s , implicaría también que la variancia de la perturbación se haría infinitamente grande (p. 771). Esto resulta implausible en la mayor parte de los contextos económicos.

Generación de las perturbaciones

La pregunta que debe responderse ahora se refiere a la forma en que se generan las perturbaciones de modo que estén relacionadas unas con otras según se especifica en la ecuación (8.33). En el caso de que se cumplan todos los supuestos básicos, cada perturbación representa un valor aleatorio e independiente procedente de una población normal con media cero y variancia σ_u^2 . Cuando las perturbaciones son autorregresivas, las extracciones ya no son independientes. Concretamente, postularemos que las perturbaciones se generan según el esquema siguiente:

$$(8.34) \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (t = 1, 2, \dots),$$

en la que u_t es una variable aleatoria independiente que se distribuye normalmente con media cero y variancia σ_u^2 que se supone independiente de ε_{t-1} . Es decir,

$$(8.35) \quad u_t \sim N(0, \sigma_u^2) \quad \text{para todo } t,$$

$$E(u_t u_s) = 0 \quad \text{para todo } t \neq s.$$

$$E(u_t \varepsilon_{t-1}) = 0 \quad \text{para todo } t.$$

$$E(\varepsilon_t \varepsilon_{t-s}) = 0 \quad \forall s \neq 0, s \leq t$$

Las relaciones del tipo de la ecuación (8.34) se conocen con el nombre de *esquemas autorregresivos de primer orden*. Ello implica que cada perturbación actual es igual a una «porción» de la perturbación anterior más un efecto aleatorio representado por u_t . Sustituyendo sucesivamente ε_{t-1} , ε_{t-2} , ... por sus valores, obtenemos

$$\begin{aligned} \varepsilon_t &= \rho \varepsilon_{t-1} + u_t \\ &= \rho(\rho \varepsilon_{t-2} + u_{t-1}) + u_t \\ &= \rho^2 \varepsilon_{t-2} + \rho u_{t-1} + u_t \\ &= \rho^2(\rho \varepsilon_{t-3} + u_{t-2}) + \rho u_{t-1} + u_t \\ &= \rho^3 \varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t \\ &\vdots \\ &= \rho^t \varepsilon_0 + \rho^{t-1} u_1 + \rho^{t-2} u_2 + \dots + \rho^2 u_{t-2} + \rho u_{t-1} + u_t. \end{aligned}$$

Esto significa que cada perturbación ε_t se genera como una función lineal de los efectos aleatorios $u_1, u_2, \dots, u_t$ y la «perturbación inicial» ε_0 . Ya hemos especificado cómo se generan las u_t y para completar la descripción de ε_t tenemos que efectuar una especificación adicional relativa al valor inicial de ε_0 . Por

como veremos más adelante, suponemos que ε_0 es una variable aleatoria que se distribuye normalmente con media cero y variancia $\sigma_u^2/(1 - \rho^2)$. Es decir,

$$(8.36) \quad \varepsilon_0 \sim N\left[0, \frac{\sigma_u^2}{1 - \rho^2}\right].$$

Demostremos ahora que el esquema autorregresivo de primer orden especificado en las ecuaciones (8.34) a (8.36) conduce en realidad a unas covariancias entre las perturbaciones sucesivas de la forma presentada en la ecuación (8.33) y que, al mismo tiempo, no se halla en conflicto con los supuestos de normalidad, media cero y homoscedasticidad de las ε_t . En primer lugar, como ε_t es una combinación lineal de $\varepsilon_0, u_1, u_2, \dots, u_t$, todas las cuales son normales e independientes entre sí, la propia ε_t debe ser normal. Además, tenemos

$$E(\varepsilon_t) = \rho^t E(\varepsilon_0) + \rho^{t-1} E(u_1) + \dots + \rho E(u_{t-1}) + E(u_t) = 0,$$

o sea, que también se mantiene el supuesto de media cero para ε_t . A continuación, obsérvese que

$$(8.37) \quad \text{Var}(\varepsilon_t) = (\rho^t)^2 \text{Var}(\varepsilon_0) + (\rho^{t-1})^2 \text{Var}(u_1) + (\rho^{t-2})^2 \text{Var}(u_2) + \dots$$

$$\begin{aligned} &= \rho^{2t} \left[\frac{\sigma_u^2}{1 - \rho^2} \right] + \rho^{2(t-1)} \sigma_u^2 + \dots + \rho^2 \sigma_u^2 + \sigma_u^2 \\ &= \rho^{2t} \left[\frac{\sigma_u^2}{1 - \rho^2} \right] + \sigma_u^2 [\rho^{2(t-1)} + \rho^{2(t-2)} + \dots + \rho^2 + 1] \\ &= \sigma_u^2 \left[\frac{\rho^{2t}}{1 - \rho^2} + \frac{1 - \rho^{2t}}{1 - \rho^2} \right] = \frac{\sigma_u^2}{1 - \rho^2}. \end{aligned}$$

Como σ_u^2 es constante a lo largo del tiempo, $\text{Var}(\varepsilon_t)$ será también constante y, por tanto, también se cumple el supuesto de homoscedasticidad. Obsérvese que este resultado se ha obtenido haciendo uso de la ecuación (8.36). Por último, a partir de la ecuación (8.34) tenemos

$$\begin{aligned} \varepsilon_t &= \rho \varepsilon_{t-1} + u_t \\ &= \rho^2 \varepsilon_{t-2} + \rho u_{t-1} + u_t \\ &\vdots \\ &= \rho^s \varepsilon_{t-s} + \rho^{s-1} u_{t-s+1} + \dots + \rho u_{t-1} + u_t. \end{aligned}$$

Multiplicando ambos lados de la expresión anterior por ε_{t-s} y tomando esperanzas matemáticas, obtenemos

$$E(\varepsilon_t \varepsilon_{t-s}) = \rho^s E(\varepsilon_t^2) + \rho^{s-1} E(\varepsilon_{t-s} \varepsilon_{t-s+1}) + \dots + \rho E(\varepsilon_{t-s} \varepsilon_{t-s-1}) + E(\varepsilon_{t-s} \varepsilon_t) \\ = \rho^s \text{Var}(\varepsilon_{t-s}) = \rho^s \sigma^2,$$

que es igual que la especificación (8.33). Por tanto, el mecanismo generador de las ecuaciones (8.34) a (8.36) produce realmente la correlación entre las perturbaciones que se ha especificado antes, y no presenta ningún conflicto con los supuestos básicos restantes.

Los comentarios anteriores indican claramente que las relaciones entre las perturbaciones dependen crucialmente del valor del parámetro ρ . Esta dependencia se ve subrayada de forma especial si interpretamos dicho parámetro de la forma siguiente. En la ecuación (8.33), la covarianza entre dos perturbaciones sucesivas cualesquiera, por ejemplo ε_t y ε_{t-1} , viene dada por

$$\rho = \frac{\text{Cov}(\varepsilon_t, \varepsilon_{t-1})}{\sigma^2},$$

Por tanto,

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho \sigma^2.$$

lo cual puede escribirse también como

$$\rho = \frac{\text{Cov}(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{\text{Var}(\varepsilon_t)} \sqrt{\text{Var}(\varepsilon_{t-1})}}.$$

ya que $\sigma^2 = \text{Var}(\varepsilon_t) = \text{Var}(\varepsilon_{t-1})$. Ahora bien, la expresión en la que se divide la covarianza por el producto de las desviaciones standard de dichas variables se conoce con el nombre de *coeficiente de correlación* entre las dos variables. Este coeficiente mide el grado de la relación entre dos variables aleatorias, y sus valores oscilan entre -1 y $+1$. Los valores positivos del coeficiente reflejan la existencia de una relación positiva, y los valores negativos la presencia de una relación negativa. Cuando el coeficiente de correlación tiene un valor próximo a $+1$ o a -1 , ello indica un elevado grado de relación entre las variables, y cuando el coeficiente tiene un valor próximo a cero, ello indica un escaso grado de relación. Esto significa que ρ es, en realidad, el coeficiente de correlación entre ε_t y ε_{t-1} , ρ^2 es el coeficiente de correlación entre ε_t y ε_{t-2} , ρ^3 es el coeficiente de correlación entre ε_t y ε_{t-3} , y así sucesivamente. Obsérvese que la posibilidad de que $\rho = +1$ o $\rho = -1$ queda eliminada la

hipótesis sostenida que se ha especificado en relación con la ecuación (8.33). Cuando ρ es igual a cero, tenemos

$$\varepsilon_t = u_t,$$

$$\text{Var}(\varepsilon_t) = \sigma_u^2,$$

$$\text{Var}(\varepsilon_0) = \sigma_u^2.$$

y como u_t es una variable independiente que se distribuye normalmente con media cero y variancia constante, se cumplen todos los supuestos básicos correspondientes a ε .

Propiedades de los estimadores minimocuadráticos

Examinemos ahora las propiedades de los estimadores minimocuadráticos de α y β en

$$Y_t = \alpha + \beta X_t + \varepsilon_t,$$

cundo la perturbación ε_t es autorregresiva. El estimador minimocuadrático de β es

$$\hat{\beta} = \frac{\sum x'_i y'_i}{\sum x'^2_i} = \beta + \frac{\sum x'_i \varepsilon_i}{\sum x'^2_i},$$

como indica la ecuación (7.32). Por tanto,

$$E(\hat{\beta}) = \beta + \frac{\sum x'_i E(\varepsilon_i)}{\sum x'^2_i} = \beta.$$

El estimador minimocuadrático de α es

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = (\alpha + \beta \bar{X} + \bar{\varepsilon}) - \hat{\beta} \bar{X};$$

$$y \quad E(\hat{\alpha}) = \alpha + \beta \bar{X} + E(\bar{\varepsilon}) - E(\hat{\beta}) \bar{X} = \alpha.$$

Esto significa que los estimadores minimocuadráticos son *insesgados* aun en el caso de que las perturbaciones sean autorregresivas.

A continuación, examinaremos si los estimadores minimocuadráticos son todavía estimadores lineales insesgados óptimos (ELIO), obteniendo las fórmulas de los ELIO para el caso autorregresivo y comparándolas con las fórmulas minimocuadráticas. Si los dos conjuntos de fórmulas no coinciden, los estima-

dores minimocuadráticos no son ELIO. Empecemos por considerar el ELIO de β . Por la condición de linealidad, tenemos

$$\beta = \sum_{t=1}^n a_t Y_t,$$

en la que a_t ($t = 1, 2, \dots, n$) son constantes que deben calcularse. Además,

$$E(\beta) = E\left(\sum_t a_t Y_t\right) = \sum_t a_t(\alpha + \beta X_t) = \alpha \left(\sum_t a_t\right) + \beta \left(\sum_t a_t X_t\right).$$

Según la condición de insesgadez, es preciso que

$$\sum a_t = 0 \quad \text{y} \quad \sum a_t X_t = 1.$$

La variancia de β viene dada por la expresión

$$\begin{aligned} \text{Var}(\beta) &= E\left[\sum_t a_t Y_t - E\left(\sum_t a_t Y_t\right)\right]^2 = E\left[\sum_t a_t \varepsilon_t\right]^2 \\ &= E \sum_t a_t^2 \varepsilon_t^2 + 2E \sum_{s < t} a_t \varepsilon_t a_s \varepsilon_s \\ &= \sigma^2 \sum_t a_t^2 + 2\sigma^2 \sum_{s < t} a_t a_s \rho^s. \end{aligned}$$

Para minimizar $\text{Var}(\beta)$ condicionada a $\sum a_t = 0$ y $\sum a_t X_t = 1$, utilizamos la técnica de los multiplicadores de Lagrange, y formamos la función H de la forma siguiente:

$$H = \sigma^2 \sum_t a_t^2 + 2\sigma^2 \sum_{s < t} a_t a_s \rho^s - \lambda_1 \left(\sum_t a_t\right) - \lambda_2 \left(\sum_t a_t X_t - 1\right).$$

Esta función debe minimizarse respecto de $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 . Derivando e igualando a cero cada una de las derivadas, obtenemos

$$\begin{aligned} 2a_1\sigma^2 + 2\sigma^2(a_2\rho + a_3\rho^2 + \dots + a_n\rho^{n-1}) - \lambda_1 - \lambda_2 X_1 &= 0 \\ 2a_2\sigma^2 + 2\sigma^2(a_1\rho + a_3\rho + \dots + a_n\rho^{n-2}) - \lambda_1 - \lambda_2 X_2 &= 0 \\ \vdots \\ 2a_n\sigma^2 + 2\sigma^2(a_1\rho^{n-1} + a_2\rho^{n-2} + \dots + a_{n-1}\rho) - \lambda_1 - \lambda_2 X_n &= 0 \\ -\sum_t a_t &= 0 \\ -\sum_t a_t X_t + 1 &= 0 \end{aligned}$$

Estas $(n+2)$ ecuaciones pueden resolverse obteniendo los $(n+2)$ valores desconocidos de $a_1, a_2, \dots, a_n, \lambda_1$ y λ_2 . Algebraicamente, la solución resulta muy laboriosa y no la vamos a desarrollar aquí,⁸ pero simplemente observando las ecuaciones resulta obvio que las soluciones para $a_1, a_2, \dots, a_n$ incluirán el parámetro ρ . Ello significa que la fórmula del ELIO de β también incluirá ρ y, por tanto, será distinta de la fórmula del estimador minimocuadrático de β . En realidad, puede mostrarse⁹ que el ELIO de β viene dado por

$$\begin{aligned} (8.38) \quad \beta &= \frac{(1-\rho^2)x'_1 y'_1 + \sum (x'_i - \rho x'_{i-1})(y'_i - \rho y'_{i-1})[1 + (1+\rho)/(n-1)(1-\rho)]}{(1-\rho^2)x'^2_1 + \sum (x'_i - \rho x'_{i-1})^2 + (1+\rho)/(n-1)(1-\rho)} \\ &= \frac{\sum (x'_i - \rho x'_{i-1})(y'_i - \rho y'_{i-1})}{\sum (x'_i - \rho x'_{i-1})^2} + \delta_\beta, \end{aligned}$$

en la que el subíndice i se extiende entre $i = 2$ y $i = n$ y δ es un factor de corrección del que, para la mayor parte de fines prácticos, se puede prescindir. Otros resultados son

$$(8.39) \quad \text{Var}(\beta) = \frac{\sigma_u^2}{\sum (x'_i - \rho x'_{i-1})^2} + \delta_{\text{Var}(\beta)},$$

$$\begin{aligned} (8.40) \quad \bar{\alpha} &= \frac{1}{1-\rho} \left\{ \frac{1}{n-1} \sum (Y_t - \rho Y_{t-1}) \right. \\ &\quad \left. - (\beta - \delta_\beta) \left[\frac{1}{n-1} \sum (X_t - \rho X_{t-1}) \right] \right\} + \delta_\alpha, \end{aligned}$$

y

$$(8.41) \quad \text{Var}(\bar{\alpha}) = \frac{\sigma_u^2}{(1-\rho)^2} \left\{ \frac{1}{n-1} + \frac{[\sum (X_t - \rho X_{t-1})/(n-1)]^2}{\sum (x'_i - \rho x'_{i-1})^2} \right\} + \delta_{\text{Var}(\bar{\alpha})},$$

en los que, al igual que en (8.38), el subíndice i toma los valores entre $i = 2, i = n$ y δ representa un factor de corrección del que puede prescindirse en la práctica. Estos resultados son claramente distintos de los presentados por las fórmulas minimocuadráticas y, por tanto, llegamos a la conclusión de que cuando las perturbaciones son autorregresivas los estimadores minimocuadráticos no son ELIO. Ello implica que, en este caso, los estimadores minimocuadráticos no son eficientes. (Intuitivamente, la pérdida de eficiencia puede explicarse como resultado del hecho de que la dependencia entre las pertur-

8. El lector interesado puede consultar, por ejemplo, J. Johnston, *Econometric Methods* Nueva York, McGraw-Hill, 1965, pp. 179-87 (ed. en español: *Métodos de Econometría* Barcelona, Ed. Vicens-Vives, 1972).

9. *Ibid.*

baciones disminuye el número *efectivo* de piezas de información independientes en la muestra.) Puede señalarse que cuando $\rho = 0$, las fórmulas (8.38) a (8.41) —incluyendo los factores de corrección— se reducen a las fórmulas minimocuadráticas; si se prescinde de los factores de corrección, estas fórmulas son iguales a las obtenidas por el método de los mínimos cuadrados eliminando la primera observación.

Pasemos ahora a las propiedades asintóticas de los estimadores minimocuadráticos de los coeficientes de regresión cuando las perturbaciones son autorregresivas. Respecto a la consistencia, podemos comprobar si las variancias de estos estimadores tienden a cero cuando el tamaño de la muestra tiende a infinito. Como los estimadores minimocuadráticos son insesgados, ésta es una condición suficiente para la consistencia. Empezando por la variancia de $\hat{\beta}$, el estimador minimocuadrático de β , tenemos

$$\begin{aligned}(8.42) \quad \text{Var}(\hat{\beta}) &= E(\hat{\beta} - \beta)^2 = E \left[\frac{\sum x'_i e_i}{\sum x_i'^2} \right]^2 \\&= \frac{1}{(\sum x_i'^2)^2} E \left[\sum x_i'^2 e_i^2 + 2 \sum_{s < t} x'_i e_i x'_{t-s} e_{t-s} \right] \\&= \frac{\sigma^2}{(\sum x_i'^2)^2} \left[\sum x_i'^2 + 2 \sum_{s < t} x'_i x'_{t-s} \rho^s \right] \\&= \frac{\sigma^2}{\sum x_i'^2} + \frac{2\sigma^2}{(\sum x_i'^2)^2} \left[\rho \sum_{t=2}^n x'_i x'_{t-1} + \rho^2 \sum_{t=3}^n x'_i x'_{t-2} + \dots \right].\end{aligned}$$

Para simplificar la notación, introducimos el coeficiente de correlación entre X_t y X_{t-s} , al que llamaremos r_s , definiéndolo en la forma siguiente

$$r_s = \frac{(1/n) \sum x'_i x'_{t-s}}{\sqrt{(1/n) \sum x_i'^2} \sqrt{(1/n) \sum x_{t-s}'^2}} = \frac{\sum x'_i x'_{t-s}}{\sqrt{\sum x_i'^2} \sqrt{\sum x_{t-s}'^2}},$$

en la que $s = 1, 2, \dots, n-1$; $t = s+1, s+2, \dots, n$; y $s < t$. Puede mostrarse fácilmente que el valor máximo de r_s^2 (como el del cuadrado de todos los coeficientes de correlación) es la unidad. Por tanto, podemos escribir $\text{Var}(\hat{\beta})$ en la forma

$$\begin{aligned}(8.42a) \quad \text{Var}(\hat{\beta}) &= \frac{\sigma^2}{\sum x_i'^2} \\&+ \frac{2\sigma^2}{(\sum x_i'^2)^2} \left[\rho r_1 \sqrt{\sum x_i'^2} \sqrt{\sum x_{t-1}'^2} + \rho^2 r_2 \sqrt{\sum x_i'^2} \sqrt{\sum x_{t-2}'^2} + \dots \right] \\&= \frac{(\sigma^2/n)}{(1/n) \sum x_i'^2} + \frac{2(\sigma^2/n)}{[(1/n) \sum x_i'^2]^2} \left[\rho r_1 \sqrt{\frac{1}{n} \sum x_i'^2} \sqrt{\frac{1}{n} \sum x_{t-1}'^2} \right. \\&\quad \left. + \rho^2 r_2 \sqrt{\frac{1}{n} \sum x_i'^2} \sqrt{\frac{1}{n} \sum x_{t-2}'^2} + \dots \right].\end{aligned}$$

cuando n tiende a infinito, los términos

$$\frac{1}{n} \sum x_i'^2, \quad \frac{1}{n} \sum x_{t-1}'^2, \quad \frac{1}{n} \sum x_{t-2}'^2, \quad \dots,$$

tenderán todos al mismo número finito positivo, por ejemplo m_{xx} , y los términos $r_1, r_2, r_3, \dots$, tenderán a ciertos números menores o iguales que uno en valor absoluto, por ejemplo $r_1^*, r_2^*, r_3^*, \dots$. Por tanto, tenemos

$$\begin{aligned}\lim_{n \rightarrow \infty} \text{Var}(\hat{\beta}) &= \frac{\lim (\sigma^2/n)}{m_{xx}} + \frac{2 \lim (\sigma^2/n)}{m_{xx}} [\rho r_1^* + \rho^2 r_2^* + \dots] \\&= \frac{\lim (\sigma^2/n)}{m_{xx}} [1 + 2\rho r_1^* + 2\rho^2 r_2^* + \dots].\end{aligned}$$

Ahora bien, como ρ se halla entre -1 y $+1$, y $r_1^*, r_2^*, \dots$, son todos menores o iguales que uno en valor absoluto, la suma de la serie infinita

$$[1 + 2\rho r_1^* + 2\rho^2 r_2^* + \dots]$$

será un número finito.

Por tanto, como $\lim_{n \rightarrow \infty} (\sigma^2/n) = 0$,

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\beta}) = 0.$$

Utilizando un argumento similar, podemos mostrar también que

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\alpha}) = 0.$$

Ello significa que los estimadores minimocuadráticos de los coeficientes de regresión son consistentes aunque los términos de perturbación de la regresión sean autorregresivos.

La última propiedad que nos interesa es la eficiencia asintótica. Esta puede examinarse comparando las variancias asintóticas de los estimadores minimocuadráticos con las de los estimadores lineales insesgados óptimos (o, mejor dicho, sus aproximaciones) indicadas por las ecuaciones (8.38) y (8.40). Utilizando la fórmula de la variancia asintótica de un estimador presentada en

la sección 6.1, podemos obtener la variancia asintótica del estimador mínimo-cuadrático de β en la forma siguiente:

$$\begin{aligned}\text{Var. asintótica}(\hat{\beta}) &= \frac{1}{n} \lim_{n \rightarrow \infty} E n \left[\frac{\sum x'_t e_t}{\sum x'_t{}^2} \right]^2 \\ &= \frac{1}{n} \lim_{n \rightarrow \infty} \frac{n}{(\sum x'_t{}^2)^2} \left[\sigma^2 \sum x'_t{}^2 + 2\sigma^2 \sum_{s < t} x'_t x'_{t-s} \rho^s \right] \\ &= \frac{\sigma^2}{n} \left[\frac{m_{xx} + 2\rho r_1^* m_{xx} + 2\rho^2 r_2^* m_{xx} + \dots}{m_{xx}^2} \right] \\ &= \frac{\sigma^2}{nm_{xx}} [1 + 2\rho r_1^* + 2\rho^2 r_2^* + \dots].\end{aligned}$$

Como estimador alternativo de β , consideremos $\tilde{\beta}$ definida como

$$(8.43) \quad \tilde{\beta} = \hat{\beta} - \delta_{\hat{\beta}},$$

en la que $\hat{\beta}$ es el estimador lineal insesgado óptimo indicado en (8.38), y $\delta_{\hat{\beta}}$ es el factor de corrección, que se considera sin importancia. La variancia de $\tilde{\beta}$ es

$$(8.44) \quad \text{Var}(\tilde{\beta}) = \text{Var}(\hat{\beta}) - \delta_{\text{Var}(\hat{\beta})},$$

como se definió en (8.39). La variancia asintótica de $\tilde{\beta}$ es, en este caso,

$$\begin{aligned}\text{Var. asintótica}(\tilde{\beta}) &= \frac{1}{n} \lim_{n \rightarrow \infty} n \left[\frac{\sigma^2(1 - \rho^2)}{\sum (x'_t - \rho x'_{t-1})^2} \right] = \frac{1}{n} \left[\frac{\sigma^2(1 - \rho^2)}{m_{xx} - 2\rho r_1^* m_{xx} + \rho^2 m_{xx}} \right] \\ &= \frac{\sigma^2}{nm_{xx}} \left[\frac{1 - \rho^2}{1 - 2\rho r_1^* + \rho^2} \right]\end{aligned}$$

Las variancias asintóticas de $\hat{\beta}$ y $\tilde{\beta}$ pueden compararse formando el cociente

$$\begin{aligned}\frac{\text{Var. asintótica}(\hat{\beta})}{\text{Var. asintótica}(\tilde{\beta})} &= \frac{(\sigma^2/nm_{xx})[1 + 2\rho r_1^* + 2\rho^2 r_2^* + \dots]}{(\sigma^2/nm_{xx})[(1 - \rho^2)/(1 - 2\rho r_1^* + \rho^2)]} \\ &= \frac{1 + 2\rho r_1^* + 2\rho^2 r_2^* + \dots}{[(1 - \rho^2)/(1 - 2\rho r_1^* + \rho^2)]}.\end{aligned}$$

Si este cociente es mayor que la unidad, $\tilde{\beta}$ no puede considerarse asintóticamente eficiente. (En sentido estricto, esta afirmación sólo es cierta si se conoce el valor de ρ o si se puede estimar de forma consistente; de lo contrario, $\tilde{\beta}$ no puede considerarse como estimador. El problema de obtener un estimador consistente de ρ se estudiará en la última parte de esta sección.) Supon-

gamos que evaluamos el cociente anterior para $1 > \rho > 0$ y $r_2^* = r_1^{*2}$, $r_3^* = r_1^{*3}$, ... Es decir, consideramos una situación en la que los términos de perturbación están autocorrelacionados positivamente, y los coeficientes de correlación entre X_t y X_{t-1} , X_t y X_{t-2} , etc., siguen una progresión geométrica. Este es el tipo de situaciones que se considera muy común cuando se trata de series económicas temporales.¹⁰ Con esta especificación obtenemos

$$\begin{aligned}\frac{\text{Var. asintótica}(\hat{\beta})}{\text{Var. asintótica}(\tilde{\beta})} &= \frac{1 + 2\rho r_1^* + 2\rho^2 r_1^{*2} + \dots}{[(1 - \rho^2)/(1 - 2\rho r_1^* + \rho^2)]} \\ &= \frac{1 - \rho r_1^* - 2\rho^2 r_1^{*2} + \rho^2 + \rho^3 r_1^*}{1 - \rho r_1^* - \rho^2 + \rho^3 r_1^*}.\end{aligned}$$

Esta expresión será mayor o igual que la unidad si

$$1 - \rho r_1^* - 2\rho^2 r_1^{*2} + \rho^2 + \rho^3 r_1^* \geq 1 - \rho r_1^* - \rho^2 + \rho^3 r_1^*$$

o

$$-2\rho^2 r_1^{*2} + \rho^2 \geq -\rho^2;$$

es decir, si

$$2\rho^2(1 - r_1^{*2}) \geq 0.$$

Esta condición se satisface siempre. Por ejemplo, cuando $\rho = 0,6$ y $r_1^* = 0,8$, $r_2^* = 0,64$, $r_3^* = 0,512$, etc., el cociente entre las dos variancias asintóticas es igual a 1,78, es decir, la variancia asintótica de $\hat{\beta}$ es un 78 por ciento mayor que la de $\tilde{\beta}$. Con respecto a α puede obtenerse un resultado similar. Por tanto, llegamos a la conclusión de que los estimadores minimocuadráticos de los coeficientes de regresión *no son asintóticamente eficientes* cuando las perturbaciones son autorregresivas.

Propiedades de las variancias estimadas de los estimadores minimocuadráticos

En resumen, hemos establecido que cuando los términos de perturbación son autorregresivos, los estimadores minimocuadráticos de los coeficientes de regresión son insesgados y consistentes pero no son eficientes ni asintóticamente eficientes. Por tanto, si utilizamos las fórmulas minimocuadráticas cuando las perturbaciones son autorregresivas, los estimadores resultantes poseerán

¹⁰ Ver E. Ames y S. Reiter, «Distributions of Correlation Coefficients in Economic Time Series», *Journal of the American Statistical Association*, vol. 56, septiembre 1961, p. 675. Los autores consideran 100 series anuales de 25 observaciones seleccionadas al azar de un resumen de estadísticas de Estados Unidos descubriendo que, en promedio, los coeficientes de correlación son 0,84, 0,71, 0,60, 0,53 y 0,45.

aún algunas de las propiedades deseables. Sin embargo, si queremos utilizar estos estimadores con el fin de contrastar hipótesis o para construir intervalos de confianza, no sólo precisamos la insesgadez de los propios estimadores, sino también la de sus variancias estimadas. El problema, pues, es saber si las fórmulas convencionales para la estimación de las variancias de los estimadores minimocuadráticos garantizan en realidad la insesgadez aun en el caso de que exista autorregresión en los términos de perturbación. Obsérvese que la fórmula minimocuadrática convencional para la estimación de la variancia de β es

$$s_{\beta}^2 = \frac{s^2}{\sum x_i^2},$$

en la que s^2 es un estimador de σ^2 definido como la suma de los cuadrados de los residuos minimocuadráticos dividida por $(n-2)$. Como $\sum x_i^2$ es no estocástico, sólo tenemos que ocuparnos de s^2 . Para este término tenemos

$$\begin{aligned} s^2 &= \frac{1}{n-2} \sum_i (y_i - \beta x_i)^2 = \frac{1}{n-2} \sum_i (\beta x_i + \epsilon_i - \beta x_i)^2 \\ &= \frac{1}{n-2} \sum_i [-(\hat{\beta} - \beta)x_i + \epsilon_i]^2 \\ &= \frac{1}{n-2} [(\hat{\beta} - \beta)^2 \sum_i x_i^2 + \sum_i \epsilon_i^2 - 2(\hat{\beta} - \beta) \sum_i x_i \epsilon_i] \\ &= \frac{1}{n-2} [\sum_i \epsilon_i^2 - (\hat{\beta} - \beta)^2 \sum_i x_i^2] \end{aligned}$$

$$y \quad E(s^2) = \frac{1}{n-2} [\sum_i E(\epsilon_i^2) - (\sum_i x_i^2) \text{Var}(\hat{\beta})].$$

A partir de la expresión (8.42) conocemos $\text{Var}(\hat{\beta})$, de modo que nuestro único problema consiste en hallar $E(\epsilon_i^2)$. Tenemos

$$\begin{aligned} E(\epsilon_i^2) &= E(\epsilon_i - \bar{\epsilon})^2 = E(\epsilon_i^2) + E\left[\frac{1}{n} \sum_i \epsilon_i\right]^2 - 2E\left[\epsilon_i \frac{1}{n} \sum_i \epsilon_i\right] \\ &= \sigma^2 + \frac{\sigma^2}{n} + \frac{2}{n^2} E\left[\sum_{s < i} \epsilon_s \epsilon_i\right] \\ &\quad - \frac{2}{n} E[\epsilon_i(\epsilon_1 + \epsilon_2 + \dots + \epsilon_i + \dots + \epsilon_n)] \end{aligned}$$

$$\begin{aligned} &= \sigma^2 + \frac{\sigma^2}{n} + \frac{2}{n^2} [E(\epsilon_1 \epsilon_2) + E(\epsilon_1 \epsilon_3) + \dots + E(\epsilon_1 \epsilon_n) + E(\epsilon_2 \epsilon_3) \\ &\quad + E(\epsilon_2 \epsilon_4) + \dots + E(\epsilon_2 \epsilon_n) + E(\epsilon_3 \epsilon_4) + \dots + E(\epsilon_{n-1} \epsilon_n)] \\ &\quad - \frac{2}{n} [E(\epsilon_i \epsilon_1) + E(\epsilon_i \epsilon_2) + \dots + E(\epsilon_i^2) + \dots + E(\epsilon_i \epsilon_n)] \\ &= \frac{(n+1)\sigma^2}{n} + \frac{2\sigma^2}{n^2} [(n-1)\rho + (n-2)\rho^2 + \dots + \rho^{n-1}] \\ &\quad - \frac{2\sigma^2}{n} [\rho^{i-1} + \rho^{i-2} + \dots + \rho + 1 + \rho + \dots + \rho^{n-i}]. \end{aligned}$$

Por tanto,

$$\begin{aligned} \sum_i E(\epsilon_i^2) &= (n+1)\sigma^2 + \frac{2\sigma^2}{n} [(n-1)\rho + (n-2)\rho^2 + \dots + \rho^{n-1}] \\ &\quad - \frac{2\sigma^2}{n} \sum_i [\rho^{i-1} + \rho^{i-2} + \dots + \rho + 1 + \rho + \dots + \rho^{n-i}]. \end{aligned}$$

Consideremos el último sumatorio:

$$\begin{aligned} \sum_i [\rho^{i-1} + \rho^{i-2} + \dots + \rho + 1 + \rho + \dots + \rho^{n-i}] &= (1 + \rho + \rho^2 + \dots + \rho^{n-1}) \\ &\quad + (\rho + 1 + \rho + \dots + \rho^{n-2}) \\ &\quad + (\rho^2 + \rho + 1 + \dots + \rho^{n-3}) \\ &\quad \vdots \\ &\quad + (\rho^{n-1} + \rho^{n-2} + \rho^{n-3} + \dots + 1) \\ &= n + 2(n-1)\rho + 2(n-2)\rho^2 + \dots + 4\rho^{n-2} + 2\rho^{n-1}. \end{aligned}$$

Por tanto,

$$\begin{aligned} \sum_i E(\epsilon_i^2) &= (n+1)\sigma^2 + \frac{2\sigma^2}{n} \{[(n-1)\rho + (n-2)\rho^2 + \dots + \rho^{n-1}] \\ &\quad - [n + 2(n-1)\rho + 2(n-2)\rho^2 + \dots + 4\rho^{n-2} + 2\rho^{n-1}]\} \\ &= (n+1)\sigma^2 - \frac{2\sigma^2}{n} A, \end{aligned}$$

en la que $A = n + (n-1)\rho + (n-2)\rho^2 + \dots + 2\rho^{n-2} + \rho^{n-1}$.

Ahora bien, como

$$\begin{aligned} A - \rho A &= n + (n-1)\rho + (n-2)\rho^2 + \dots + 2\rho^{n-2} + \rho^{n-1} \\ &\quad - \rho n - \rho(n-1)\rho - \dots - \rho 3\rho^{n-2} - 2\rho^{n-1} - \rho^n \\ &= n - \rho - \rho^2 - \dots - \rho^{n-1} - \rho^{n-2} - \rho^n = n - \frac{\rho(1-\rho^n)}{(1-\rho)} \\ &= \frac{n(1-\rho) - \rho(1-\rho^n)}{(1-\rho)} \end{aligned}$$

tenemos

$$A = \frac{n(1-\rho) - \rho(1-\rho^n)}{(1-\rho)^2}$$

y

$$\begin{aligned} \sum_i E(\epsilon_i^2) &= (n+1)\sigma^2 - \frac{2\sigma^2}{n} \left[\frac{n(1-\rho) - \rho(1-\rho^n)}{(1-\rho)^2} \right] \\ &= (n-1)\sigma^2 - \frac{2\sigma^2}{n} \left[\frac{n\rho(1-\rho) - \rho(1-\rho^n)}{(1-\rho)^2} \right]. \end{aligned}$$

Sustituyendo este resultado y la expresión correspondiente a $\text{Var}(\hat{\beta})$ en la fórmula de $E(s^2)$, obtenemos

$$\begin{aligned} (8.45) \quad E(s^2) &= \frac{1}{n-2} \left\{ (n-1)\sigma^2 - \frac{2\sigma^2}{n} \left[\frac{n\rho(1-\rho) - \rho(1-\rho^n)}{(1-\rho)^2} \right] \right. \\ &\quad \left. - \frac{1}{n-2} \sum_i x_i'^2 \left\{ \frac{\sigma^2}{\sum x_i'^2} + \frac{2\sigma^2}{(\sum x_i'^2)^2} \left[\rho \sum_{i=2}^n x_i' x_{i-1}' \right. \right. \right. \\ &\quad \left. \left. \left. + \rho^2 \sum_{i=3}^n x_i' x_{i-2}' + \dots \right] \right\} \right\} \\ &= \sigma^2 - \frac{2\sigma^2}{n-2} \left[\frac{n\rho(1-\rho) - \rho(1-\rho^n)}{n(1-\rho)^2} \right. \\ &\quad \left. + \frac{\rho \sum x_i' x_{i-1}' + \rho^2 \sum x_i' x_{i-2}' + \dots}{\sum x_i'^2} \right], \end{aligned}$$

es decir, que

$$(8.46) \quad E(s_\beta^2) = \frac{\sigma^2}{\sum x_i'^2} - \frac{2\sigma^2}{(n-2) \sum x_i'^2} \left[\frac{n\rho(1-\rho) - \rho(1-\rho^n)}{n(1-\rho)^2} \right. \\ \left. + \frac{\rho \sum x_i' x_{i-1}' + \rho^2 \sum x_i' x_{i-2}' + \dots}{\sum x_i'^2} \right].$$

Como la expresión correspondiente a $E(s_\beta^2)$ es distinta de la de $\text{Var}(\hat{\beta})$ indicada en la ecuación (8.42), llegamos a la conclusión de que, cuando las perturbaciones son autorregresivas, la variancia de $\hat{\beta}$ calculada de forma convencional es sesgada. Puede obtenerse un resultado similar en relación a $\tilde{\alpha}$.

El resultado anterior implica que, cuando las perturbaciones son autorregresivas, las fórmulas convencionales para efectuar contrastaciones de significación o para construir intervalos de confianza respecto a los coeficientes de regresión conducen a afirmaciones incorrectas. Es decir, las regiones de aceptación o los intervalos de confianza calculados serán más reducidos o más amplios que los correctos según que el sesgo en la estimación de la variancia sea negativo o positivo. Por lo tanto, veamos si podemos determinar la dirección de dicho sesgo. Restando (8.42) de (8.46) obtenemos una expresión del sesgo en la estimación de la variancia de $\hat{\beta}$ que viene dado por

$$\begin{aligned} (8.47) \quad E(s_\beta^2) - \text{Var}(\hat{\beta}) &= \frac{-2\sigma^2}{\sum x_i'^2} \left\{ \left[\frac{n\rho(1-\rho) - \rho(1-\rho^n)}{n(n-2)(1-\rho)^2} \right] \right. \\ &\quad \left. + \frac{(n-1)}{(n-2) \sum x_i'^2} \left[\rho \sum_{i=2}^n x_i' x_{i-1}' \right. \right. \\ &\quad \left. \left. + \rho^2 \sum_{i=3}^n x_i' x_{i-2}' + \dots \right] \right\}. \end{aligned}$$

A partir de ello puede verse que cuando $\rho > 0$ y X_t está positivamente correlacionada con X_{t-1} , X_{t-2} , ..., el sesgo es negativo. Como se ha señalado anteriormente, este tipo de situación es muy corriente en las series económicas temporales y, por tanto, si las perturbaciones son autorregresivas y persistimos en utilizar las fórmulas minimocuadráticas convencionales, las regiones de aceptación o los intervalos de confianza calculados serán a menudo más estrechos de lo que deberían ser para el nivel de significación o de confianza que se ha especificado. Con el fin de obtener una idea acerca de la magnitud de este sesgo, considérese, por ejemplo, el caso en que $\rho = 0.6$ y $r_1 = 0.8$, $r_2 = (0.8)^2$, $r_3 = (0.8)^3$, etc. Suponiendo que $\sum x_i'^2$, $\sum x_i' x_{i-1}'$, $\sum x_i' x_{i-2}'$, ... son todos aproximadamente iguales, los sesgos en la estimación de $\text{Var}(\hat{\beta})$ para distintos tama-

ñes muestrales serían los presentados en el cuadro 8-1. Estos resultados indican que la magnitud del sesgo puede ser muy sustancial.

Cuadro 8-1

Tamaño de la muestra	$E(s_{\beta}^2)$	$\text{Var}(\hat{\beta})$	Sesgo	$\frac{E(s_{\beta}^2)}{\text{Var}(\hat{\beta})}$
20	$0.752 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$2.846 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$-2.094 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	0.264
50	$0.902 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$2.846 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$-1.944 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	0.317
100	$0.951 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$2.846 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	$-1.895 \left[\frac{\sigma^2}{\sum x_i^2} \right]$	0.334

Estimación maximoverosmil y LIO

La discusión anterior nos ha permitido poner de manifiesto los defectos de los estimadores minimocuadráticos de los coeficientes de regresión en el caso en que el término de perturbación no satisface el supuesto de no autorregresión. Ahora nos ocuparemos del problema de la obtención de estimadores alternativos que no tengan tales inconvenientes. En este proceso nos limitaremos al caso en que la perturbación ϵ_t sigue un esquema autorregresivo de primer orden del tipo descrito por las expresiones (8.34) y (8.35). El lector recordará que ya hemos obtenido las fórmulas de los estimadores lineales insesgados óptimos para el modelo autorregresivo. Estos estimadores, a los que denominaremos $\tilde{\alpha}$ y $\tilde{\beta}$, y sus variancias vienen dados por las expresiones (8.38) a (8.41). Si prescindimos de los factores de corrección incluidos en cada fórmula, las expresiones resultantes son más sencillas, aunque, naturalmente, ya no son exactamente ELIO. Si se utilizan $\tilde{\alpha}$ y $\tilde{\beta}$ para indicar $\hat{\alpha}$ y $\hat{\beta}$ menos los factores de corrección respectivos, tenemos

$$(8.48) \quad \tilde{\beta} = \frac{\sum (y_i' - \rho y_{i-1}') (x_i' - \rho x_{i-1}')}{\sum (x_i' - \rho x_{i-1}')^2},$$

$$(8.49) \quad \text{Var}(\tilde{\beta}) = \frac{\sigma_u^2}{\sum (x_i' - \rho x_{i-1}')^2},$$

$$(8.50) \quad \tilde{\alpha} = \frac{1}{1 - \rho} [(\bar{Y} - \rho \bar{Y}_{-1}) - \tilde{\beta}(\bar{X} - \rho \bar{X}_{-1})],$$

y

$$(8.51) \quad \text{Var}(\tilde{\alpha}) = \frac{\sigma_u^2}{(1 - \rho)^2} \left[\frac{1}{n - 1} + \frac{(\bar{X} - \rho \bar{X}_{-1})^2}{\sum (x_i' - \rho x_{i-1}')^2} \right],$$

en la que el subíndice t toma los valores entre $t = 2$ y $t = n$, y

$$\bar{Y} = \frac{1}{n - 1} \sum_i Y_i,$$

$$\bar{Y}_{-1} = \frac{1}{n - 1} \sum_i Y_{i-1},$$

$$\bar{X} = \frac{1}{n - 1} \sum_i X_i,$$

$$\bar{X}_{-1} = \frac{1}{n - 1} \sum_i X_{i-1}.$$

Antes de considerar la posibilidad de calcular la utilidad de estos estimadores, podemos mencionar un procedimiento alternativo, y más sencillo, para obtenerlos. El valor de la variable dependiente en el período t está determinado por

$$Y_t = \alpha + \beta X_t + \epsilon_t,$$

y su valor en el período $t - 1$, por

$$Y_{t-1} = \alpha + \beta X_{t-1} + \epsilon_{t-1}.$$

Multiplicando por ρ esta última ecuación y restándola de la primera, obtenemos

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (\epsilon_t - \rho \epsilon_{t-1}).$$

Pero como, según la ecuación (8.34),

$$\epsilon_t - \rho \epsilon_{t-1} = u_t,$$

podemos escribir

$$(8.52) \quad Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + u_t \quad (t = 2, 3, \dots, n).$$

Podemos considerar la ecuación (8.52) como una ecuación de regresión en la que la variable dependiente es $(Y_t - \rho Y_{t-1})$ y la variable explicativa es $(X_t - \rho X_{t-1})$. Estas dos variables son susceptibles de medición si ρ tiene un valor conocido. Como u_t es una variable aleatoria independiente que se distribuye normalmente con media cero y variancia constante, y $(X_t - \rho X_{t-1})$ es no estocástica y está limitada, se cumplen todos los supuestos básicos del modelo clásico de regresión normal lineal. Por tanto, los estimadores mínimo-

cuadráticos de la ecuación (8.52) poseerán todas las propiedades deseables salvo por el hecho de que, al utilizar las variables $(Y_t - \rho Y_{t-1})$ en lugar de Y_t , y $(X_t - \rho X_{t-1})$ en lugar de X_t , perdemos una observación. Esta pérdida implica que tales estimadores no son exactamente ELIO, lo cual ya se ha señalado anteriormente.

La dificultad existente en relación con las fórmulas de estimación (8.48) y (8.50) se debe a que incluyen el parámetro ρ cuyo valor no se conoce casi nunca. Si se desconoce dicho valor, no se pueden calcular las fórmulas, y en consecuencia $\hat{\alpha}$ y $\hat{\beta}$ no pueden considerarse estimadores. Esta dificultad se puede superar estimando ρ (junto con α y β) a partir de las observaciones muestrales. Es posible utilizar distintos métodos de estimación y vamos a estudiar cada uno de ellos. En primer lugar, consideremos la estimación maximoverosímil de (8.52). Como

$$u_t = (Y_t - \rho Y_{t-1}) - \alpha(1 - \rho) - \beta(X_t - \rho X_{t-1}) \quad (t = 2, 3, \dots, n),$$

la función de verosimilitud logarítmica para $Y_2, Y_3, \dots, Y_n$ es

$$(8.53) \quad L = -\frac{n-1}{2} \log(2\pi) - \frac{n-1}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_{t=2}^n [(Y_t - \rho Y_{t-1}) - \alpha(1 - \rho) - \beta(X_t - \rho X_{t-1})]^2.$$

Obsérvese que

$$f(y_2, y_3, \dots, y_n) = \left| \frac{\partial u}{\partial Y} \right| f(u_2, u_3, \dots, u_n),$$

en la que $|\partial u / \partial Y|$ es el valor absoluto del determinante

$$\begin{vmatrix} \frac{\partial u_2}{\partial Y_2} & \frac{\partial u_2}{\partial Y_3} & \dots & \frac{\partial u_2}{\partial Y_n} \\ \frac{\partial u_3}{\partial Y_2} & \frac{\partial u_3}{\partial Y_3} & \dots & \frac{\partial u_3}{\partial Y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial u_n}{\partial Y_2} & \frac{\partial u_n}{\partial Y_3} & \dots & \frac{\partial u_n}{\partial Y_n} \end{vmatrix}.$$

Esto es una ampliación del teorema 18 (el teorema del «cambio de la variable») de una variable a varias. El determinante recibe el nombre de *Jacobiano* de la transformación de $u_2, u_3, \dots, u_n$ en $Y_2, Y_3, \dots, Y_n$. Se ha presentado *condicionado* a un determinado valor de Y_1 , es decir, como si Y_1 fuese una cons-

ta, lo cual nos permite prescindir de $\partial u_2 / \partial Y_1$. Puede mostrarse fácilmente que $\partial u / \partial Y = 1$ de modo que $f(u_2, u_3, \dots, u_n) = f(y_2, y_3, \dots, y_n)$.

Si derivamos L respecto de α , β , ρ , y σu^2 , e igualamos a cero cada una de estas derivadas, obtenemos cuatro ecuaciones a partir de las cuales pueden obtenerse los valores de los cuatro parámetros desconocidos. Sin embargo, estas ecuaciones son altamente no lineales, y su solución resulta muy difícil. Un procedimiento relativamente simple para resolver este problema consiste en utilizar el hecho de que si conociésemos el valor de ρ , los estimadores maximoverosímiles de α y β serían iguales a los valores de $\hat{\alpha}$ y $\hat{\beta}$ correspondientes a las ecuaciones (8.48) y (8.50), y el estimador maximoverosímil de σu^2 , al que podemos llamar $\hat{\sigma}_u^2$, sería

$$(8.54) \quad \hat{\sigma}_u^2 = \frac{1}{n-1} \sum_{t=2}^n [(Y_t - \rho Y_{t-1}) - \hat{\alpha}(1 - \rho) - \hat{\beta}(X_t - \rho X_{t-1})]^2.$$

Ello significa que distintos valores de ρ producen distintos valores de $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_u^2$. Por tanto, podemos obtener soluciones distintas para, por ejemplo, $\rho = -0.95, -0.90, -0.85, \dots, +0.85, +0.90, +0.95$, seleccionando después el valor de ρ —y los correspondientes valores de $\hat{\alpha}$ y $\hat{\beta}$ — que produzcan el valor más pequeño de $\hat{\sigma}_u^2$. Esta solución hará máxima la función de verosimilitud L tal como pretendemos. Los estimadores obtenidos de este modo se denominarán $\hat{\alpha}$, $\hat{\beta}$, $\hat{\rho}$ y $\hat{\sigma}_u^2$. Los errores standard estimados de $\hat{\alpha}$ y $\hat{\beta}$ pueden obtenerse por medio de las fórmulas (8.49) y (8.51), sustituyendo ρ por $\hat{\rho}$ y σu^2 por $\hat{\sigma}_u^2$. La variancia de $\hat{\rho}$ se puede estimar mediante

$$(8.55) \quad s_{\hat{\rho}}^2 = \frac{1 - \hat{\rho}^2}{n}.$$

La fórmula (8.55) representa la variancia asintótica estimada tal como se ha definido en la sección 6.1.

Sabemos que, bajo condiciones muy generales, los estimadores maximoverosímiles poseen las propiedades óptimas de las muestras grandes. Estas propiedades óptimas están garantizadas por los teoremas generales de la estimación maximoverosímil. Sin embargo, en el caso actual existe algún problema acerca de la aplicabilidad de los teoremas generales, ya que $Y_2, Y_3, \dots, Y_n$ no son independientes, es decir,

$$f(y_2, y_3, \dots, y_n) \neq f(y_2)f(y_3) \dots f(y_n).$$

Esto significa que la función de verosimilitud no es el producto de unas distribuciones idénticas independientes;¹¹ sin embargo, puede mostrarse que los

11. C. Hildreth y J. Y. Lu pusieron particularmente de relieve este punto en «Demand Relations with Autocorrelated Disturbances», *Technical Bulletin* 276, Michigan State University Agricultural Experiment Station, noviembre 1960, p. 14.

estimadores maximoverosímiles de (8.52) son consistentes y equivalen asintóticamente a los estimadores lineales insesgados óptimos.¹² En la página 349 se indicarán las propiedades de los estimadores maximoverosímiles en las muestras pequeñas.

EJEMPLO. Friedman y Meiselman¹³ estimaron una ecuación representativa de la forma simple de la teoría cuantitativa del dinero:

$$C_t = \alpha + \beta M_t + \varepsilon_t$$

en la que C = gastos de consumo y M = cantidad de dinero, ambas medidas en miles de millones de dólares corrientes. Estimaremos de nuevo esta relación utilizando los datos trimestrales del cuadro (8-2) y suponiendo que la perturbación sigue un esquema autorregresivo de primer orden.

Cuadro 8-2

Año y trimestre	Gastos de consumo *	Cantidad de dinero *	Año y trimestre	Gastos de consumo *	Cantidad de dinero *
1952 I	214.6	159.3	1954 III	238.7	173.9
II	217.7	161.2	IV	243.2	176.1
III	219.6	162.8			
IV	227.2	164.6	1955 I	249.4	178.0
			II	254.3	179.1
1953 I	230.9	165.9	III	260.9	180.2
II	233.3	167.9	IV	263.3	181.2
III	234.1	168.3			
IV	232.3	169.7	1956 I	265.6	181.6
			II	268.2	182.5
1954 I	233.7	170.5	II	270.4	183.3
II	236.5	171.6	IV	275.6	184.3

Fuente: Milton Friedman y David Meiselman, «The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958», en Commission on Money and Credit, *Stabilization Policies*. Englewood Cliffs, N. J., Prentice-Hall, 1963, p. 266.

* En miles de millones de dólares.

Si siguiendo el procedimiento para la obtención de las estimaciones maximoverosímiles de α y β , calculamos los valores de $\hat{\sigma}_u^2$ —o, mejor dicho, de $(n-1) \hat{\sigma}_u^2$ —para distintos valores de ρ . Los resultados de los cálculos se presentan en el cuadro 8-3. Podemos ver que el valor de ρ que produce el valor más pequeño de $\hat{\sigma}_u^2$,

12. La consistencia se ha demostrado; ver *ibid.*, Apéndice B. Malinvaud, *op. cit.*, pp. 440-41 ha demostrado la equivalencia asintótica respecto a los ELIO.

13. Milton Friedman y David Meiselman, «The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958» en Commission on Money and Credit, *Stabilization Policies*. Englewood Cliffs, N. J., Prentice-Hall, 1963.

es 0.85. Las correspondientes estimaciones minimocuadráticas de los coeficientes y sus errores standard estimados son

$$(C_t - 0.85C_{t-1}) = -38.256 + 2.859(M_t - 0.85M_{t-1}) + \hat{u}_t, \quad R^2 = 0.647, \\ (13.926) \quad (0.512)$$

Cuadro 8-3

ρ	$(n-1)\hat{\sigma}_u^2$	ρ	$(n-1)\hat{\sigma}_u^2$	ρ	$(n-1)\hat{\sigma}_u^2$
-0.95	919.30	-0.05	302.75	0.75	85.26
-0.90	874.77	0.00	279.99	0.80	82.71
-0.85	831.45	0.05	258.44	0.85	82.14
:	:	0.10	238.10	0.90	84.48
-0.10	326.73	:	:	0.95	88.67

A partir de estos resultados obtenemos:

$$\hat{\beta} = 2.859, \quad y \quad s_{\hat{\beta}} = 0.512, \\ \hat{\alpha} = \frac{-38.256}{1-0.85} = -255.040, \quad y \quad s_{\hat{\alpha}} = \frac{13.926}{1-0.85} = 92.840. \\ \hat{\rho} = 0.85, \quad y \quad s_{\hat{\rho}} = 0.178.$$

Para tener una idea del ajuste de la recta de regresión estimada a las observaciones muestrales, calculamos también el valor de R^2 . Según la fórmula (7.45b)

$$R^2 = 1 - \frac{SCE}{STC}$$

$$\text{en la que } SCE = \sum_{i=2}^{20} (C_i - \hat{\alpha} - \hat{\beta}M_i)^2 = \sum_{i=2}^{20} (c'_i - \hat{\beta}m'_i)^2 \\ = \sum_{i=2}^{20} c_i'^2 - 2\hat{\beta} \sum_{i=2}^{20} c'_i m'_i + \hat{\beta}^2 \sum_{i=2}^{20} m_i'^2,$$

y

$$STC = \sum_{i=2}^{20} c_i'^2.$$

En los cálculos siguientes, los sumatorios se extienden desde $t = 2$ hasta $t = n$, ya

que disponemos de los valores de dichos sumatorios como subproductos de la estimación minimocuadrática con $\rho = 0$. En este caso, tenemos

$$\sum_{i=2}^{20} e_i'^2 = 5,803.63,$$

$$\sum_{i=2}^{20} m_i'^2 = 1,013.07,$$

y

$$\sum_{i=2}^{20} e_i' m_i' = 2.335 \times 1,013.07 = 2,365.52,$$

o sea, que

$$R^2 = 0.904.$$

El resultado completo es, pues,

$$C_t = -255.040 + 2.859 M_t + e_t, \quad R^2 = 0.904, \\ (92.840) \quad (0.512)$$

Estimación iterativa y en dos etapas

Un procedimiento de estimación alternativo para la estimación de ecuaciones de regresión cuyos términos de perturbación son autorregresivos es el *método iterativo*,¹⁴ que consiste en efectuar los siguientes pasos:

1. Obtener las estimaciones minimocuadráticas ordinarias de

$$Y_t = \alpha + \beta X_t + e_t,$$

y calcular los residuos $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n$. Con estos residuos se obtiene la «primera vuelta» de la estimación de $\rho, \hat{\rho}$, que viene dada por

$$\hat{\rho} = \frac{\sum \hat{e}_t \hat{e}_{t-1}}{\sum \hat{e}_{t-1}^2} \quad (t = 2, 3, \dots, n).$$

2. Construir las nuevas variables $(Y_t - \hat{\rho} Y_{t-1})$ y $(X_t - \hat{\rho} X_{t-1})$, y obtener las estimaciones minimocuadráticas ordinarias de

$$(Y_t - \hat{\rho} Y_{t-1}) = \alpha^* + \beta(X_t - \hat{\rho} X_{t-1}) + u_t,$$

en la que α^* es igual a $\alpha(1 - \hat{\rho})$. Estas estimaciones de la «segunda vuelta», a las que podemos llamar $\hat{\alpha}$ y $\hat{\beta}$, producen los residuos de la «segunda vuelta» $\hat{\hat{e}}_1, \hat{\hat{e}}_2, \dots, \hat{\hat{e}}_n$ (calculados en la forma $\hat{\hat{e}}_t = Y_t - \hat{\alpha} - \hat{\beta} X_t$). Estos últimos se utilizan ahora para obtener una nueva estimación de ρ :

$$\hat{\hat{\rho}} = \frac{\sum \hat{\hat{e}}_t \hat{\hat{e}}_{t-1}}{\sum \hat{\hat{e}}_{t-1}^2} \quad (t = 2, 3, \dots, n).$$

3. Construir las nuevas variables $(Y_t - \hat{\hat{\rho}} Y_{t-1})$ y $(X_t - \hat{\hat{\rho}} X_{t-1})$, y proceder a continuación como en el apartado 2.

Estos pasos deben seguirse hasta que los valores de los estimadores convergen. Puede mostrarse que el procedimiento es convergente y que, en realidad, los estimadores de α y β de la «última vuelta» coinciden con los valores de los estimadores maximoverosímiles anteriormente descritos.¹⁵ Por tanto, la única diferencia entre los estimadores maximoverosímiles que hemos obtenido antes y los estimadores iterativos propuestos por Orcutt y otros autores se encuentra en el diseño del método de cálculo.

El procedimiento iterativo puede reducirse a un *procedimiento en dos etapas* si se interrumpe el proceso tras haber obtenido las estimaciones de $\hat{\alpha}$ y $\hat{\beta}$ de la «segunda vuelta», basándonos en el valor de ρ de la «primera vuelta». Los estimadores de dos etapas poseerán las mismas propiedades asintóticas que los estimadores maximoverosímiles, y en la página 349 se presentan algunos resultados acerca de sus propiedades en las muestras pequeñas. Utilizando las fórmulas (8.49) y (8.51), y sustituyendo ρ por $\hat{\rho}$, pueden obtenerse las estimaciones de los errores standard de $\hat{\alpha}$ y $\hat{\beta}$.

EJEMPLO. Podemos utilizar la relación de la «teoría cuantitativa» y los datos del ejemplo anterior para ilustrar el procedimiento de la estimación en dos etapas. La estimación de ρ de la «primera vuelta» es

$$\hat{\rho} = 0.827.$$

14. Ver D. Cochrane y G. H. Orcutt, «Application of Least Squares Regressions to Relationships Containing Autorrelated Error Terms», *Journal of the American Statistical Association*, vol. 44, marzo 1949, pp. 32-61.

15. Ver J. D. Sargan, «Wages and Prices in the United Kingdom: A Study in Econometric Methodology», en P. L. Hart, G. Mills y J. K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Londres, Butterworths, 1964.

Obsérvese que este valor se halla muy próximo a la estimación maximoverosímil de ρ del ejemplo anterior. Las estimaciones minimocuadráticas de los coeficientes de regresión basadas en los datos transformados son

$$(C_t - 0.827C_{t-1}) = -42.290 + 2.805(M_t - 0.827M_{t-1}) + e_t, \quad R^2 = 0.703.$$

(13.760) (0.442)

Esto nos conduce a las siguientes estimaciones de las observaciones no transformadas:

$$C_t = -244.450 + 2.805M_t + e_t, \quad R^2 = 0.912.$$

(79.537) (0.442)

Estos resultados son muy parecidos a los obtenidos anteriormente a través del método de la máxima verosimilitud.

Método de Durbin

Durbin ha propuesto un método de estimación distinto.¹⁶ El *procedimiento de Durbin* consta, al igual que el método anterior, de dos etapas. En primer lugar, escribimos de nuevo la ecuación (8.52) en la forma

$$Y_t = \alpha(1 - \rho) + \rho Y_{t-1} + \beta X_t - \beta \rho X_{t-1} + u_t$$

o
$$Y_t = \alpha^* + \rho Y_{t-1} + \beta X_t + \gamma X_{t-1} + u_t.$$

Podemos considerar esta expresión como una ecuación de regresión con tres variables explicativas, X_t , X_{t-1} e Y_{t-1} , y la podemos estimar por el método de los mínimos cuadrados ordinarios (tal como se explica en el capítulo 10). El estimador resultante de ρ , es decir $\hat{\rho}$, debe utilizarse para construir las nuevas variables $(Y_t - \hat{\rho}Y_{t-1})$ y $(X_t - \hat{\rho}X_{t-1})$. En el segundo paso, estimamos

$$(Y_t - \hat{\rho}Y_{t-1}) = \alpha^* + \beta(X_t - \hat{\rho}X_{t-1}) + u_t^*,$$

en la que $\alpha^* = \alpha(1 - \rho)$. Los estimadores que obtenemos de α y β poseerán las mismas propiedades *asintóticas* que los estimadores maximoverosímiles anteriormente descritos.

Utilización de las primeras diferencias

En los primeros estudios aplicados, los investigadores intentaron a menudo resolver el problema de la autorregresión en el término de perturbación utilizando el *método de las primeras diferencias*. Este método consiste en transformar los datos originales respecto de X e Y en sus primeras diferencias $(X_t - X_{t-1})$ e $(Y_t - Y_{t-1})$ y plantear la ecuación de regresión en la forma

$$(8.56) \quad (Y_t - Y_{t-1}) = \alpha^{**} + \beta(X_t - X_{t-1}) + v_t.$$

estimando luego α^{**} y β por el método de los mínimos cuadrados. Obsérvese que, como

$$Y_t = \alpha + \beta X_t + e_t$$

y

$$Y_{t-1} = \alpha + \beta X_{t-1} + e_{t-1},$$

se deduce que $\alpha^{**} = 0$ y $v_t = e_t - e_{t-1}$. La lógica del método de las primeras diferencias consiste en la creencia de que el verdadero valor de ρ es muy próximo a la unidad. Como $\alpha^{**} = 0$, no es de esperar que su estimación de la ecuación (8.56) sea significativamente distinta de cero, y la implicación es que con este método no podemos estimar α . Señalemos de paso que a menudo se han racionalizado los resultados en que la estimación de α^{**} es significativamente distinta a cero afirmando que se había especificado de forma incorrecta el modelo original y que, además de X_t , debería haberse incluido una tendencia como variable explicativa. En este caso, y si se midiese la tendencia mediante la «variable tiempo» t , tendríamos

$$Y_t = \alpha + \beta X_t + \delta t + e_t,$$

$$Y_{t-1} = \alpha + \beta X_{t-1} + \delta(t-1) + e_{t-1},$$

y

$$\bar{Y}_t - Y_{t-1} = \beta(X_t - X_{t-1}) + \delta + (e_t - e_{t-1}),$$

o sea, que la ordenada en el origen en la ecuación (8.56) mediría el coeficiente de la variable de la tendencia.

Consideremos las propiedades del estimador de β cuando se utiliza el método de las primeras diferencias. Igual que antes, suponemos que e_t sigue un esquema autorregresivo de primer orden del tipo presentado en las ecuaciones (8.34) a (8.36). El término de perturbación de la ecuación (8.56) es

$$v_t = e_t - e_{t-1},$$

16. J. Durbin, «Estimation of Parameters in Time Series Regression Models», *Journal of the Royal Statistical Society*, serie B, vol. 22, enero 1960, pp. 159-55.

o sea, que

$$E(v_t) = 0,$$

$$E(v_t^2) = E(e_t^2 + e_{t-1}^2 - 2e_t e_{t-1}) = \sigma^2 + \sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1 - \rho)$$

$$= \frac{2\sigma_u^2(1 - \rho)}{(1 - \rho^2)} = \frac{2\sigma_u^2}{(1 + \rho)},$$

$$E(v_t v_{t-1}) = E(e_t - e_{t-1})(e_{t-1} - e_{t-2})$$

$$= E(e_t e_{t-1} - e_{t-1}^2 - e_t e_{t-2} + e_{t-1} e_{t-2}) = \rho\sigma^2 - \sigma^2 - \rho^2\sigma^2 + \rho\sigma^2$$

$$= -\sigma^2(1 - \rho)^2 = \frac{-\sigma_u^2(1 - \rho)}{(1 + \rho)}.$$

Se deduce, pues, que el término de perturbación de la ecuación (8.56) tiene media cero y variancia constante, pero aún es autorregresivo, aunque, naturalmente, esta autorregresión será pequeña si ρ está próxima a la unidad.

El estimador minimocuadrático de β basado en las primeras diferencias, al que llamaremos $\tilde{\beta}$, es

$$(8.57) \quad \tilde{\beta} = \frac{\sum (y'_t - y'_{t-1})(x'_t - x'_{t-1})}{\sum (x'_t - x'_{t-1})^2}$$

$$= \beta + \frac{\sum z'_t(e_t - e_{t-1})}{\sum z_t'^2} \quad (t = 2, 3, \dots, n)$$

en la que

$$z'_t = x'_t - x'_{t-1}.$$

Y

$$E(\tilde{\beta}) = \beta,$$

o sea, que $\tilde{\beta}$ es insesgado. Además,

$$\text{Var}(\tilde{\beta}) = E \left[\frac{\sum z'_t(e_t - e_{t-1})}{\sum z_t'^2} \right]^2 = \frac{1}{(\sum z_t'^2)^2} [E \sum z_t'^2 (e_t^2 - 2e_t e_{t-1} + e_{t-1}^2)$$

$$+ 2E \sum_{s < t} z'_t z'_s (e_t e_{t-s} - e_{t-1} e_{t-s} - e_t e_{t-s-1} + e_{t-1} e_{t-s-1})]$$

$$= \frac{2\sigma^2}{(\sum z_t'^2)^2} [(1 - \rho) \sum z_t'^2 + \sum_{s < t} z'_t z'_{t-s} \rho^{s-1} (2\rho - 1 - \rho^2)]$$

$$= \frac{2\sigma^2(1 - \rho)}{\sum z_t'^2} \left\{ 1 - \frac{(1 - \rho)}{\sum z_t'^2} \left[\sum_{t=3}^n z'_t z'_{t-1} + \rho \sum_{t=3}^n z'_t z'_{t-2} + \dots \right] \right\}.$$

El estimador $\tilde{\beta}$ se basa en el supuesto de que $\rho \rightarrow 1$, mientras que el estimador minimocuadrático ordinario β se puede considerar basado en el supuesto de que $\rho = 0$. Por tanto, puede ser interesante comparar las variancias de estos dos estimadores para distintos valores de ρ . Limitaremos esta comparación a las situaciones que presentan las siguientes características.

1. Las muestras son grandes, de modo que, para todos los fines prácticos,

$$\sum_t x_t'^2 = \sum_t x_{t-1}'^2 = \sum_t x_{t-2}'^2 = \dots$$

2. $0 \leq \rho < 1$.

3. $r_s = r^s$ ($s = 1, 2, \dots$), siendo

$$r_s = \frac{\sum x'_t x'_{t-s}}{\sqrt{\sum x_t'^2} \sqrt{\sum x_{t-s}'^2}} = \frac{\sum x'_t x'_{t-s}}{\sum x_t'^2} \quad (\text{utilizando 1})$$

Con estas simplificaciones tenemos

$$\sum z_t'^2 = \sum x_t'^2 - 2 \sum x'_t x'_{t-1} + \sum x_{t-1}'^2 = 2(1 - r) \sum x_t'^2,$$

$$\sum z'_t z'_{t-1} = \sum x'_t x'_{t-1} - \sum x_{t-1}'^2 - \sum x'_t x'_{t-2} + \sum x_{t-1}' x'_{t-2} = -(1 - r)^2 \sum x_t'^2,$$

$$\sum z'_t z'_{t-2} = -r(1 - r)^2 \sum x_t'^2,$$

$$\sum z'_t z'_{t-3} = -r^2(1 - r)^2 \sum x_t'^2,$$

...

o sea, que

$$\text{Var}(\tilde{\beta}) = \frac{2\sigma^2(1 - \rho)}{2(1 - r) \sum x_t'^2} \left\{ 1 + \frac{(1 - \rho)(1 - r)^2 \sum x_t'^2}{2(1 - r) \sum x_t'^2} [1 + \rho r + \rho^2 r^2 + \dots] \right\}$$

$$= \frac{\sigma^2(1 - \rho)}{\sum x_t'^2} \left[\frac{1}{1 - r} + \frac{(1 - \rho)}{2(1 - \rho r)} \right].$$

La variancia del estimador minimocuadrático ordinario de β viene dado por la ecuación (8.42) y con la simplificación anterior se convierte en

$$\text{Var}(\beta) = \frac{\sigma^2}{\sum x_t'^2} [1 + 2\rho r + 2\rho^2 r^2 + \dots] = \frac{\sigma^2}{\sum x_t'^2} \left[\frac{1 + \rho r}{1 - \rho r} \right].$$

Con el fin de facilitar la comparación, se han calculado estas variancias para algunos valores seleccionados de ρ y r . En el cuadro 8-4 se presentan los resul-

tados de estos cálculos, y λ representa $\sigma^2/\Sigma x_i^2$. Tal como se ha anticipado, $\text{Var}(\hat{\beta})$ es relativamente pequeña

Cuadro 8-4

Valor de ρ	$i = 0,4$		$i = 0,8$	
	$\text{Var}(\hat{\beta})$	$\text{Var}(\hat{\beta})$	$\text{Var}(\hat{\beta})$	$\text{Var}(\hat{\beta})$
0.0	2.17 λ	1.00 λ	5.50 λ	1.00 λ
0.3	1.44 λ	1.27 λ	3.82 λ	1.63 λ
0.6	0.77 λ	1.63 λ	2.15 λ	2.85 λ
0.9	0.17 λ	2.12 λ	0.52 λ	6.14 λ

cuando ρ es próxima a la unidad y relativamente grande cuando ρ es pequeña. Sin embargo, debe ponerse de relieve que el *estimador* de la variancia de $\hat{\beta}$, al igual que el de la variancia de $\hat{\beta}$, está sesgado. Por tanto, resulta inapropiado utilizar $\hat{\beta}$ para la contrastación de hipótesis o la construcción de intervalos de confianza,¹⁷ y por este motivo no es recomendable el uso del método de las primeras diferencias, a menos que se crea realmente que el valor de ρ se halla muy próximo a la unidad.

Propiedades de los estimadores alternativos en las muestras pequeñas

Generalmente, se desconocen las propiedades correspondientes a las muestras de los estimadores alternativos de los coeficientes de regresión en los modelos con perturbaciones autorregresivas, ya que es muy difícil obtener las distribuciones muestrales de tales estimadores. Sin embargo, podemos hallar alguna idea acerca del comportamiento de tales estimadores en las muestras pequeñas si obtenemos *experimentalmente* sus distribuciones muestrales. Naturalmente, esto sólo puede hacerse respecto a modelos específicos y a poblaciones concretas de los términos de perturbación. La obtención experimental de las distribuciones muestrales de variables discretas se estudió en la sección 2.2. En el caso de perturbaciones que se distribuyen normalmente, las variables son continuas, pero el principio de muestreo experimental permanece inalterado. Los experimentos muestrales de este tipo se conocen con el nombre de *experimentos de Monte Carlo* debido a su parecido con los juegos de azar. A conti-

17. El lector puede calcular como ejercicio el sesgo en la estimación de la variancia de $\hat{\beta}$ por medio de la fórmula convencional. También es interesante señalar que si ρ es positivo, los valores de $\hat{\beta}$ y $\hat{\beta}$ contienen entre ellos el valor de $\hat{\beta}$ que se hubiese obtenido en caso de conocerse ρ . Ver Malinvaud, *op. cit.*, pp. 444-45.

nuación describiremos uno de estos experimentos, que se diseñó con el fin de comparar distintos estimadores de los coeficientes de regresión en un modelo con perturbaciones autorregresivas.

Consideremos la ecuación de regresión de la población

$$Y_t = 10 + 2X_t + \varepsilon_t \quad (t = 1, 2, \dots)$$

con

$$\varepsilon_t = 0.8\varepsilon_{t-1} + u_t,$$

en la que

$$u_t \sim N(0, 0.36),$$

$$E(u_t \varepsilon_{t-1}) = 0,$$

y

$$\varepsilon_0 \sim N(0, 1).$$

Por tanto,

$$E(\varepsilon_t) = 0,$$

$$\text{Var}(\varepsilon_t) = 1,$$

$$\rho = 0.8.$$

Los valores de X , que se seleccionaron de forma más o menos arbitraria, son

$X_1 = 2,$	$X_6 = 1,$	$X_{11} = X_1,$
$X_2 = 0,$	$X_7 = 4,$	$X_{12} = X_2,$
$X_3 = 1,$	$X_8 = 9,$	$\vdots$
$X_4 = 3,$	$X_9 = 4,$	$X_{99} = X_9,$
$X_5 = 9,$	$X_{10} = 3,$	$X_{100} = X_{10}.$

El experimento consiste en extraer 100 muestras de tamaño 10, 20 y 100. Los valores aleatorios de las perturbaciones se obtuvieron a partir de las tablas

Cuadro 8-5

Tamaño muestral	Estimador					
	MCO		MV		MCZE	
	Media	D.S.	Media	D.S.	Media	D.S.
10	2.0070	0.1029	2.0079	0.0634	2.0058	0.0754
20	2.0008	0.0634	1.9995	0.0402	1.9997	0.0412
100	2.0001	0.0201	1.9990	0.0136	1.9990	0.0136

1. Kmenta y R. F. Gilbert, «Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances», Michigan State University, Econometrics Workshop, Paper No. 18, diciembre 1968, p. 18.

One Million Random Digits and One Hundred Thousand Deviates (publicadas por Rand Corporation, Santa Mónica, California, 1950). Para cada muestra se han calculado los valores de los estimadores de los coeficientes de regresión por el método de los mínimos cuadrados ordinarios (MCO), de la máxima verosimilitud (MV) y de los mínimos cuadrados en dos etapas (MC2E). Una vez extraídas 100 muestras de un determinado tamaño, se calculó la media y la desviación standard de cada uno de los estimadores. Los resultados de la estimación de la pendiente de la regresión —cuyo verdadero valor es 2— aparecen en el cuadro 8-5. A partir de estos resultados puede apreciarse que los tres estimadores son insesgados, y que el estimador MCO es ineficiente en relación con los estimadores MV y MC2E. Estos dos últimos poseen idénticas desviaciones standard en las muestras grandes, y casi idénticas en las muestras de tamaño 20, pero cuando el tamaño de la muestra es 10, el estimador MV es más eficiente. Las generalizaciones a partir de experimentos concretos son siempre un tanto arriesgadas pero, no obstante, los resultados nos proporcionan alguna evidencia acerca de la conducta de los estimadores en las muestras pequeñas.

Contrastación de la no existencia de autorregresión

En toda la parte anterior de esta sección nos hemos ocupado de las implicaciones debidas a la presencia de perturbaciones autorregresivas en el modelo de regresión lineal. Hemos examinado las propiedades de los estimadores mínimos cuadrados ordinarios de los coeficientes de regresión, y al descubrir sus limitaciones hemos estudiado varios métodos alternativos de estimación. Sin embargo, si no sabemos si el modelo es o no autorregresivo —o no queremos efectuar el supuesto—, probablemente no podremos decidir qué método de estimación debemos utilizar. En tal caso, debemos acudir a la información proporcionada por la muestra y, en concreto, nos puede interesar contrastar la hipótesis de que no exista autorregresión,

$$H_0: \rho = 0,$$

respecto a una hipótesis alternativa en una o ambas direcciones. En las relaciones económicas, la hipótesis alternativa corriente es la de autorregresión positiva, es decir,

$$H_A: \rho > 0.$$

A partir del estudio de la estimación maximo-verosímil puede obtenerse, al menos de forma implícita, un procedimiento para la contrastación de la existencia de autorregresión. Cuando utilizamos el método de la máxima verosimi-

litud estimamos el valor de ρ juntamente con los valores de los coeficientes de regresión y, por tanto, podemos contrastar simplemente si el valor estimado de ρ es distinto de cero de forma significativa. En las muestras grandes, el estimador maximo-verosímil de ρ se distribuye de forma aproximadamente normal, con una variancia indicada en la ecuación (8.55), de la forma

$$s_{\hat{\rho}}^2 = \frac{1 - \hat{\rho}^2}{n}.$$

En las muestras más pequeñas puede utilizarse esta fórmula como aproximación.

EJEMPLO. Como ejemplo de la estimación maximo-verosímil cuando existe autorregresión, hemos utilizado la ecuación de la «teoría cuantitativa» de Friedman y Meiselman. Vimos que la estimación maximo-verosímil de ρ era 0.85 y su error standard estimado en las muestras grandes es

$$\sqrt{\frac{1 - 0.85^2}{20}} = 0.178.$$

Si utilizamos una contrastación al 5 % respecto a una alternativa unilateral de autorregresión positiva, la región de aceptación de la hipótesis nula es

$$\frac{\hat{\rho}}{s_{\hat{\rho}}} < 1.645.$$

En nuestro caso tenemos

$$\frac{\hat{\rho}}{s_{\hat{\rho}}} = \frac{0.85}{0.178} = 4.77.$$

Como este valor se halla claramente fuera de la región de aceptación, debemos rechazar la hipótesis de que no existe autorregresión.

Una contrastación alternativa que se ha utilizado muy ampliamente en las aplicaciones econométricas es la que se conoce con el nombre de *contrastación de Durbin-Watson*. Para efectuar esta contrastación calculamos el valor del estadístico d que viene dado por

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2},$$

en la que las e indican los residuos minimocuadráticos ordinarios. Si la hipótesis alternativa es la de autorregresión positiva, las reglas de decisión son las siguientes:

1. Rechazar la hipótesis si $d < d_L$.
2. No rechazar la hipótesis si $d > d_U$.
3. Si $d_L \leq d \leq d_U$, el resultado es incierto.

Los valores de d_L («límite inferior») y d_U («límite superior») vienen dados en la tabla calculada por Durbin y Watson y reproducida en el Apéndice D.¹⁸ Estos valores varían con el número de observaciones y el número de variables explicativas en la ecuación de regresión. Si la hipótesis alternativa abarca las dos direcciones, las reglas de decisión para la contrastación de Durbin-Watson son:

1. Rechazar la hipótesis si $d < d_L$, o si $d > 4 - d_L$.
2. No rechazar la hipótesis si $d_U < d < 4 - d_U$.
3. Si $d_L \leq d \leq d_U$, o si $4 - d_U \leq d \leq 4 - d_L$, el resultado es incierto.

En la figura 8-2 se representa gráficamente esta contrastación. Debe mencionarse que la contrastación de Durbin-Watson no es aplicable a las ecuaciones de regresión en las que la variable explicativa es un valor retardado de la variable dependiente.

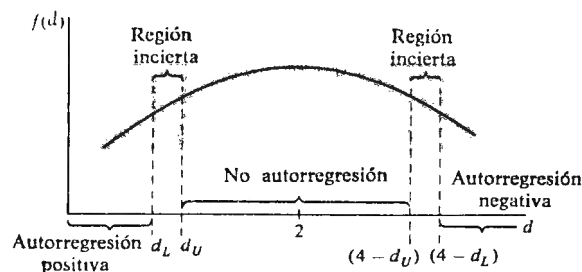


Figura 8-2

18. En H. Theil y A. L. Nagar, «Testing the Independence of Regression Disturbances», *Journal of the American Statistical Association*, vol. 56, diciembre 1961, pp. 795-806, se presenta una tabla en la que no existe ninguna región incierta, pero está basada en ciertos supuestos acerca de la variable o variables explicativas.

EjemPLO. Consideremos de nuevo la estimación de la ecuación de la «teoría cuantitativa». Supongamos que queremos contrastar la hipótesis

$$H_0: \rho = 0$$

respecto a

$$H_A: \rho > 0$$

al nivel de significación del 5 %, por medio de la contrastación de Durbin-Watson. A partir de la tabla, los valores correspondientes a d_L y d_U para 20 observaciones y una variable explicativa son

$$d_L = 1,20 \quad \text{y} \quad d_U = 1,41.$$

Ello significa que rechazaremos H_0 si $d < 1,20$ y no la rechazaremos si $d > 1,41$. El valor calculado del estadístico d es

$$d = 0,321$$

o sea, que según esta contrastación también debe rechazarse la hipótesis de no autorregresión ($\rho = 0$).

Como se ha mencionado anteriormente, la contrastación de Durbin-Watson se ha utilizado ampliamente en aplicaciones econométricas. En realidad, en la mayor parte de los estudios relacionados con la estimación de ecuaciones de regresión a partir de los datos de series temporales, se presenta el valor del estadístico d junto a las demás estimaciones. El problema, pues, consiste en saber qué acción debe tomarse (si es que debe tomarse alguna) en respuesta a un determinado resultado de esta contrastación. Si no existe indicación de autorregresión, podemos mantener las estimaciones minimocuadráticas sin temor a una pérdida de eficiencia ni a la existencia de sesgo en los errores standard estimados. Sin embargo, si la contrastación señala la existencia de autorregresión, tenemos motivos de preocupación. Una respuesta consiste en reestimar la ecuación utilizando uno de los métodos de estimación diseñados para este tipo de situaciones (por ejemplo, el procedimiento de la máxima verosimilitud o el de los mínimos cuadrados en dos etapas). De forma alternativa, podemos fijarnos de nuevo en la especificación del modelo de regresión, ya que la autorregresión del término de perturbación puede reflejar simplemente la presencia de alguna influencia sistemática no explicada sobre la variable dependiente. Un procedimiento *ad hoc* muy comúnmente adoptado para intentar descubrir esta influencia consiste en representar gráficamente los residuos y tratar de hallar una variable cuyos puntos máximos y mínimos se parezcan a los de los residuos representados. Por último, si el resultado de la contrastación es incierto, podemos tomar alguna acción o no tomar ninguna. Existe

cierta evidencia¹⁹ de que, cuando la variable explicativa varía gradualmente sin fluctuaciones regulares, agudas y cortas, el límite apropiado de la región se halla más próximo a d_u .

8.3 Variable explicativa estocástica

El supuesto 7.6 del modelo clásico de regresión normal lineal consta de los siguientes puntos:

1. X es no estocástica.
2. Los valores de X están fijos en las muestras repetidas.
3. $(1/n) \sum_i (X_i - \bar{X})^2$ es igual a un número finito y distinto de cero para todos los tamaños muestrales.

En esta sección nos ocuparemos principalmente del apartado 1 y los apartados 2 y 3 recibirán únicamente una breve mención. Para empezar, el requisito de que los valores de X estén fijos en repetidas muestras tiene sólo un interés teórico, puesto que, en realidad, raramente extraemos u observamos más de una muestra para un determinado conjunto de valores de X . El propósito de tal requisito es establecer el marco para las distribuciones muestrales de los distintos estimadores que consideramos, y en realidad equivale a decir que si obtuviésemos un número infinito de muestras de cierto tamaño (es decir, un número infinito de conjuntos de valores de $Y_1, Y_2, \dots, Y_n$) para un conjunto fijo de valores de X (es decir, para un conjunto de valores de $X_1, X_2, \dots, X_n$), las distribuciones muestrales y las propiedades de los estimadores serían las señaladas. De este modo se evitan las complicaciones que surgirían en caso de que los valores de X variasen de una muestra a otra. Obsérvese que se puede considerar que los valores de X permanecen constantes en diferentes muestras aun en el caso de que X sea una variable estocástica (siempre y cuando el hecho de mantener constantes los valores de X no haga que los de Y también sean constantes).

Según el supuesto de que $\sum_i (X_i - \bar{X})^2/n$ es un número finito y distinto de cero para todo tamaño muestral, es preciso que los valores de X de la muestra no sean todos iguales y que no aumenten ni disminuyan de forma ilimitada. El primer requisito —que no todos los valores de X sean iguales— resulta crucial, ya que, de lo contrario, sería imposible la obtención de los coeficientes de la regresión minimocuadrática, como se ha señalado en la sección 7.4. Si los valores de X no son iguales, pero las diferencias entre ellos son muy

pequeñas, el término $\sum_i (X_i - \bar{X})^2$ será también pequeño, y es probable que las variancias de los estimadores sean muy grandes. Esto, a su vez, implica que la probabilidad del error del tipo II (aceptar H_0 siendo falsa) en las pruebas de significación es grande, y que los intervalos de confianza para los parámetros de la población son amplios. A esto es a lo que se refieren los investigadores cuando, comentando los resultados de sus regresiones, dicen: «Lamentablemente, existía muy poca variación en la variable explicativa». Por lo que respecta al segundo requisito, la restricción de que los valores de X deben estar limitados resulta menos crucial, y se utiliza básicamente para demostrar que el estimador es consistente, pero *no* es una condición necesaria para la consistencia en todas las ocasiones.

Como ejemplo de un caso en el que los valores de la variable explicativa no están limitados, podemos considerar la ecuación de una «tendencia lineal» representada por

$$Y_t = \alpha + \beta t + \epsilon_t \quad (t = 1, 2, \dots, n),$$

en la que t representa tiempo en periodos especificados. Supongamos que se cumplen todos los supuestos básicos acerca del término de perturbación ϵ_t . En este caso, naturalmente, a medida que n tiende hacia infinito, lo mismo ocurre con los valores de t . En realidad,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (t - \bar{t})^2 &= \frac{1}{n} \sum_{t=1}^n t^2 - \left[\frac{1}{n} \sum_{t=1}^n t \right]^2 \\ &= \frac{1}{n} [1^2 + 2^2 + \dots + n^2] - \frac{1}{n^2} [1 + 2 + \dots + n]^2 \\ &= \frac{1}{n} \left[\frac{n}{6} (n+1)(2n+1) \right] - \frac{1}{n^2} \left[\frac{n}{2} (1+n) \right]^2 = \frac{n^2 - 1}{12}. \end{aligned}$$

Las propiedades de los estimadores minimocuadráticos en las muestras pequeñas no se ven afectadas por ello. Por lo que respecta a la consistencia, tenemos

$$\text{Var}(\beta) = \frac{\sigma^2}{\sum (t - \bar{t})^2} = \frac{\sigma^2/n}{\sum (t - \bar{t})^2/n} = \frac{12\sigma^2}{n(n^2 - 1)},$$

o sea, que

$$\lim_{n \rightarrow \infty} \text{Var}(\beta) = 0.$$

Ello implica que $\hat{\beta}$ es un estimador consistente de β .

Tras la discusión anterior, sólo queda por considerar el requisito de que la variable explicativa sea no estocástica. Ahora nos ocuparemos de los pro-

19. Ver *ibid.*; también Malinvaud, *op. cit.*, p. 425.

blemas de la estimación de los coeficientes de la ecuación de regresión cuando no se cumple dicho requisito, es decir, cuando X es una variable estocástica. En dicho caso, X no toma unos valores fijos, sino que se presentan distintos valores —o intervalos de valores— de X con distintas probabilidades. En la econometría, las regresiones con variables explicativas estocásticas resultan muy frecuentes, o incluso predominantes. En muchas relaciones económicas, los valores de la variable explicativa se determinan como resultado de ciertos mecanismos probabilísticos, junto a los valores de la variable dependiente, y no están bajo el control del experimentador ni de otras personas o instituciones. Si la variable X es estocástica, el punto más importante a considerar es su dependencia o independencia respecto del término de perturbación ε , y en caso de que sea dependiente, cuál es la naturaleza de dicha dependencia. En este sentido, distinguiremos tres posibilidades:

1. X y ε son independientes.
2. X y ε en un mismo período tienen correlación nula.
3. X y ε son independientes y en un mismo período *no* tienen correlación nula.

En cada caso nos ocuparemos de las propiedades de los estimadores minimocuadráticos de los coeficientes de regresión, considerando que se cumplen todo los supuestos básicos respecto al término de perturbación. Efectuaremos también el supuesto de que la variancia de X es un número finito distinto de cero.

Empecemos por el caso en que X y ε son *independientes*. A modo de ejemplo, tomemos la relación

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

en la que X_i e Y_i se definen de la forma siguiente:

$$Y_i = \log \left[\frac{V}{PL} \right]_i,$$

$$X_i = \log \left[\frac{W}{P} \right]_i,$$

siendo V = valor añadido en la producción, L = cantidad de trabajo, P = precio del producto, W = tasa de salario monetario, y el subíndice i hace referencia a la región i . Se efectúan las observaciones para una industria concreta y un período determinado de tiempo. El coeficiente β representa la elasticidad

de sustitución entre trabajo y capital.²⁰ En este caso, X es una variable estocástica y puede suponerse independiente de ε . El estimador minimocuadrático de β es

$$\hat{\beta} = \beta + \frac{\sum x_i' \varepsilon_i}{\sum x_i'^2}$$

y

$$E(\hat{\beta}) = \beta + E \left[\frac{\sum x_i' \varepsilon_i}{\sum x_i'^2} \right] E(\varepsilon_i) = \beta.$$

Este resultado procede de los supuestos de que X y ε son independientes y de que $E(\varepsilon_i) = 0$, y del uso del teorema 10 de la sección 3.6. Del mismo modo,

$$\begin{aligned} E(\hat{\alpha}) &= E(\bar{Y} - \hat{\beta} \bar{X}) = E(\alpha + \beta \bar{X} + \bar{\varepsilon} - \hat{\beta} \bar{X}) \\ &= \alpha + \beta E(\bar{X}) + E(\bar{\varepsilon}) - E \left[\beta + \frac{\sum x_i' \varepsilon_i}{\sum x_i'^2} \right] \bar{X} = \alpha, \end{aligned}$$

de modo que $\hat{\alpha}$ y $\hat{\beta}$ mantienen la propiedad de insesgadez.

Como X es una variable estocástica, $\hat{\alpha}$ y $\hat{\beta}$ ya no son funciones lineales de $Y_1, Y_2, \dots, Y_n$; es decir, ya no se pueden definir como iguales a $\sum a_i Y_i$ en las que $a_1, a_2, \dots, a_n$ son *constantes*. Por lo tanto, en el sentido más estricto, no puede considerarse que $\hat{\alpha}$ y $\hat{\beta}$ son los estimadores *lineales* insesgados óptimos. Sin embargo, puede verse fácilmente que $\hat{\alpha}$ y $\hat{\beta}$ son eficientes si consideramos que sus variancias están condicionadas a un conjunto determinado de valores de $X_1, X_2, \dots, X_n$. Podemos imaginar la población de todos los pares de valores de X e Y como una enorme cantidad de fichas en la que cada una de ellas representa un par de valores concretos. Entonces podemos especificar un conjunto de valores de $X_1, X_2, \dots, X_n$ y hacer que la máquina clasificadora separe todas las fichas con estos valores de X . Este subconjunto de fichas representa la población de la cual se van a extraer las muestras. Cada una de estas muestras puede utilizarse para estimar los coeficientes de regresión. Si extrásemos un número infinito de muestras podríamos construir las distribuciones muestrales de estos estimadores y obtener sus propiedades. Obviamente, las propiedades de los estimadores obtenidos de esta forma estarían condicionados a los valores elegidos $X_1, X_2, \dots, X_n$. Ello significa que podemos tratar estos valores *como si* fuesen números fijos, lo que son en realidad *trata* haber separado todas las fichas apropiadas del grupo original. Sabemos que los estimadores minimocuadráticos de los coeficientes de regresión son eficientes

²⁰ J. H. Duesenberry, "Capital-Labor Substitution and Economic Efficiency", *Journal of Political Economy*, Vol. 45, pp. 1-14, p. 2, 1937, p. 14.

para un conjunto fijo de valores $X_1, X_2, \dots, X_n$, y sabemos también que el es cierto para *cualquier* conjunto fijo de valores de X . Por tanto, como los estimadores minimocuadráticos de los coeficientes de regresión son insesgados incondicionalmente, y son eficientes para *todos* los conjuntos de valores de $X_1, X_2, \dots, X_n$, de ello se deriva que también son incondicionalmente eficientes.

Resulta fácil mostrar que los estimadores minimocuadráticos de los coeficientes de regresión siguen poseyendo también sus propiedades asintóticas deseables cuando X es estocástica pero independiente de ϵ . Una forma de hacerlo consiste en apoyarse en el principio de la máxima verosimilitud. La distribución de probabilidad para una observación, por ejemplo la i -ésima, viene dada por $f(x_i, \epsilon_i)$. Para n observaciones independientes tenemos

$$f(x_1, \epsilon_1)f(x_2, \epsilon_2) \dots f(x_n, \epsilon_n),$$

lo cual, como X y ϵ son independientes, puede escribirse en la forma

$$f(x_1)f(x_2) \dots f(x_n)f(\epsilon_1)f(\epsilon_2) \dots f(\epsilon_n).$$

La transformación de probabilidad a partir de las ϵ , que no son observables, a las Y es

$$f(y_i) = \left| \frac{\partial \epsilon_i}{\partial Y_i} \right| f(\epsilon_i) = f(\epsilon_i),$$

o sea, que la función logarítmica de verosimilitud se convierte en

$$\begin{aligned} L &= \log [f(x_1)f(x_2) \dots f(x_n)] + \log [f(\epsilon_1)f(\epsilon_2) \dots f(\epsilon_n)] \\ &= \sum_i \log f(x_i) - \frac{n}{2} \log (2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i (Y_i - \alpha - \beta X_i)^2. \end{aligned}$$

En la medida en que la distribución de X no incluya ninguno de los parámetros α , β o σ^2 , la maximización de la función de verosimilitud respecto a dichos parámetros no se verá afectada por la presencia del término $\sum \log f(x_i)$. Los estimadores resultantes de α y β serán, en este caso, los mismos que los estimadores minimocuadráticos. Por tanto, los estimadores minimocuadráticos de los coeficientes de regresión son consistentes, asintóticamente eficientes, y asintóticamente normales tanto si X no es estocástica como si lo es, siempre y cuando X y ϵ sean independientes.

Como último punto, consideremos las variancias de los estimadores minimocuadráticos de α y β . Para $\hat{\beta}$ tenemos

$$\text{Var}(\hat{\beta}) = E \left[\frac{\sum x_i' \epsilon_i}{\sum x_i'^2} \right]^2 - E \left[\frac{\sum x_i'^2 \epsilon_i^2}{(\sum x_i'^2)^2} \right] + 2E \left[\frac{\sum x_i' x_i' \epsilon_i \epsilon_j}{(\sum x_i'^2)^2} \right] = \sigma^2 E \left[\frac{1}{\sum x_i'^2} \right]$$

De un modo más preciso, puede mostrarse que

$$\text{Var}(\hat{\alpha}) = \sigma^2 E \left[\frac{1}{n} + \frac{\lambda^2}{\sum x_i'^2} \right],$$

es decir, las variancias de $\hat{\alpha}$ y $\hat{\beta}$ son iguales que en el caso en que X es no estocástica, salvo por el hecho de que los términos que incluyen X se ven sustituidos por sus esperanzas matemáticas. Las fórmulas de los estimadores insesgados de $\text{Var}(\hat{\alpha})$ y $\text{Var}(\hat{\beta})$ vienen dadas por (7.37). Además, los procedimientos clásicos para la estimación de intervalos y la contrastación de hipótesis que se han desarrollado en el capítulo 7 siguen siendo válidas cuando X es estocástica pero independiente de ϵ . Por tanto, la eliminación del supuesto de que X es no estocástica sustituyéndolo por el supuesto de que es estocástica pero independiente de ϵ no altera las propiedades deseables ni la viabilidad de la estimación minimocuadrática.

Ahora se plantea la segunda posibilidad respecto a la variable explicativa, es decir, el caso en el que X y ϵ en un mismo período tienen correlación nula. Es decir, suponemos que

$$\text{Cov}(X_1, \epsilon_1) = \text{Cov}(X_2, \epsilon_2) = \dots = \text{Cov}(X_n, \epsilon_n) = 0.$$

Obsérvese que no se supone que X y ϵ en un mismo período son independientes, lo cual constituiría un requisito más estricto. A modo de ejemplo, consideremos el siguiente modelo muy simplificado de la determinación de la renta:

$$C_t = \gamma_0 + \gamma_1 Y_t + \epsilon_{1t} \quad (\text{función de consumo}),$$

$$I_t = \delta_0 + \delta_1 Y_{t-1} + \epsilon_{2t} \quad (\text{función de inversión}),$$

$$Y_t = C_t + I_t \quad (\text{identidad de la renta})$$

siendo C = consumo, Y = renta, I = inversión, y ϵ_{1t} y ϵ_{2t} perturbaciones aleatorias que cumplen los supuestos básicos. El modelo de tres ecuaciones puede reducirse a una sola ecuación despejando Y_t para obtener

$$Y_t = (\gamma_0 + \gamma_1 Y_t + \epsilon_{1t}) + (\delta_0 + \delta_1 Y_{t-1} + \epsilon_{2t}).$$

Esto puede escribirse en la forma

(8.58)

$$Y_t = \alpha + \beta Y_{t-1} + \epsilon_t,$$

en la que

$$\alpha = \frac{\gamma_0 + \delta_0}{1 - \gamma_1},$$

$$\beta = \frac{\delta_1}{1 - \gamma_1},$$

y

$$\varepsilon_t = \frac{\varepsilon_{1t} + \varepsilon_{2t}}{1 - \gamma_1}.$$

En la ecuación (8.58), la variable explicativa está representada por el valor retardado de la variable dependiente. Los modelos de este tipo se conocen generalmente con el nombre de «modelos autorregresivos en las variables». En nuestro caso, nos limitaremos a las situaciones en que $-1 < \beta < 1$. Efectuando sustituciones sucesivas, podemos expresar Y_t en la forma

$$Y_t = \alpha(1 + \beta + \beta^2 + \dots + \beta^{t-1}) + \beta^t Y_0 + \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2\varepsilon_{t-2} + \dots + \beta^{t-1}\varepsilon_1.$$

Cuando $t \rightarrow \infty$, Y_t se convierte en

$$Y_t = \frac{\alpha}{1 - \beta} + \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2\varepsilon_{t-2} + \dots;$$

es decir, a largo plazo Y_t se «establece», fluctuando aleatoriamente en torno a un nivel fijo. Obsérvese que en la ecuación (8.58) la variable Y_{t-1} no está correlacionada con la perturbación del período actual ε_t , ya que Y_{t-1} depende de $Y_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{t-1}$, pero no de ε_t .

Consideremos ahora el estimador minimocuadrático de β en la ecuación (8.58). Tenemos

$$\hat{\beta} = \beta + \frac{\sum y'_{t-1}\varepsilon_t}{\sum y'^2_{t-1}}.$$

Ahora bien, Y'_{t-1} y ε_t no son independientes, ya que

$$y'_{t-1} = Y_{t-1} - \frac{1}{n}(Y_0 + Y_1 + \dots + Y_t + \dots + Y_{n-1});$$

es decir, Y'_{t-1} incorpora Y_t , que no es independiente de ε_t . Esto significa que cuando tomamos la esperanza matemática de $\hat{\beta}$ no podemos separar ε_t como hacíamos en el caso de independencia y, por tanto, no puede afirmarse que $\hat{\beta}$ sea insesgado. Sin embargo, el límite probabilístico de $\hat{\beta}$ es

$$\lim \text{prob } \hat{\beta} = \beta + \frac{\lim \text{prob } (\sum y'_{t-1}\varepsilon_t)/n}{\lim \text{prob } (\sum y'^2_{t-1})/n} = \beta,$$

ya que $\sum y'_{t-1}\varepsilon_t$ es un estimador consistente de la covarianza entre y'_{t-1} y la perturbación, que es cero, y $(\sum y'^2_{t-1})/n$ es un estimador consistente de la variancia de y'_{t-1} , que es un número finito distinto de cero. Por tanto, $\hat{\beta}$ es un estimador consistente de β y, del mismo modo, puede mostrarse que $\hat{\alpha}$ es un estimador consistente de α . La eficiencia y la normalidad asintóticas de los estimadores minimocuadráticos de α y β resultan más difíciles de demostrar, pero también se han podido establecer.²¹ La conclusión es, pues, que cuando la variable explicativa y el término de perturbación en el mismo período tienen correlación nula (como en el modelo autorregresivo en las variables), los resultados clásicos de la estimación por mínimos cuadrados establecidos en el capítulo 7 se cumplen sólo asintóticamente.

La última posibilidad respecto a la variable explicativa es el caso en que X y ε no son independientes ni su correlación en el mismo período es nula. En este caso,

$$\lim \text{prob } \hat{\beta} = \beta + \frac{\lim \text{prob } (\sum x'_t\varepsilon_t)/n}{\lim \text{prob } (\sum x'^2_t)/n} \neq \beta$$

y

$$\lim \text{prob } \hat{\alpha} \neq \alpha,$$

de modo que los estimadores minimocuadráticos de los coeficientes de regresión no son ni siquiera consistentes. Una explicación intuitiva de ello es que el método de estimación minimocuadrática está construido de forma que la variación total de Y (STC) puede dividirse siempre en dos partes, una de las cuales representa la variación debida a la variable explicativa (SCR) y la otra representa la variación debida a otros factores. Pero cuando la variable explicativa y el término de perturbación están correlacionados, esta división no resulta válida, ya que no tiene en cuenta el efecto conjunto de X y ε sobre Y .

EJEMPLO. Como ejemplo de una situación en la que la variable explicativa y el término de perturbación en el mismo período están correlacionados, consideremos una ecuación de demanda de mercado representada por

$$Q_t = \alpha + \beta P_t + \varepsilon_t \quad (\alpha > 0, \beta < 0),$$

siendo Q = cantidad y P = precio de una mercancía determinada. Si el mercado está en equilibrio, la cantidad de mercancía vendida en el mercado y el precio de equilibrio están determinados por la intersección de las funciones de oferta y demanda. Supongamos que la función de oferta es

$$Q_t = \gamma + \delta P_t + \eta_t \quad (\gamma > 0, \delta > 0),$$

²¹ Para los detalles y referencias importantes, ver Malinvaud, *op. cit.*, pp. 452-53.

en la que $\eta_i \sim N(0, \sigma_\eta^2)$ es un término de perturbación estocástico, no autorregresivo e independiente de ε_i . Las ecuaciones de oferta y demanda pueden resolverse despejando el precio y la cantidad de equilibrio, obteniéndose

$$Q_i = \frac{\alpha\delta - \beta\gamma}{\delta - \beta} + \frac{\delta\varepsilon_i - \beta\eta_i}{\delta - \beta},$$

$$P_i = \frac{\alpha - \gamma}{\delta - \beta} + \frac{\varepsilon_i - \eta_i}{\delta - \beta}.$$

A partir del último resultado podemos ver que P_i y ε_i están correlacionadas; concretamente,

$$\text{Cov}(P_i, \varepsilon_i) = E[P_i - E(P_i)][\varepsilon_i - E(\varepsilon_i)] = E\left[\frac{\varepsilon_i - \eta_i}{\delta - \beta}\varepsilon_i\right] = \frac{\sigma^2}{\delta - \beta}.$$

El estimador minimocuadrático de la pendiente de la ecuación de demanda es

$$\hat{\beta} = \beta + \frac{\sum p'_i \varepsilon'_i}{\sum p'^2_i},$$

en la que

$$p'_i = P_i - \bar{P} = \frac{\varepsilon'_i - \eta'_i}{\delta - \beta}.$$

Sustituyendo el valor de p'_i en la expresión de $\hat{\beta}$ obtenemos

$$\hat{\beta} = \beta + \frac{\left[\frac{1}{\delta - \beta}\right] \sum \varepsilon'_i(\varepsilon'_i - \eta'_i)}{\left[\frac{1}{\delta - \beta}\right]^2 \sum (\varepsilon'_i - \eta'_i)^2} = \beta + \frac{(\delta - \beta) \left[\frac{1}{n} \sum \varepsilon'^2_i - \frac{1}{n} \sum \varepsilon'_i \eta'_i \right]}{\frac{1}{n} \sum \varepsilon'^2_i + \frac{1}{n} \sum \eta'^2_i - \frac{2}{n} \sum \varepsilon'_i \eta'_i}.$$

Por tanto,

$$\lim \text{prob} \quad \hat{\beta} = \beta + \frac{(\delta - \beta)\sigma^2}{\sigma^2 + \sigma_\eta^2},$$

lo cual demuestra que el estimador minimocuadrático de β es inconsistente. Este ejemplo ilustra el llamado «problema de las ecuaciones simultáneas» que se presenta en la estimación de ciertas relaciones económicas.

Las ecuaciones de regresión en las que la variable explicativa y el término de perturbación están correlacionadas resultan muy frecuentes en la econometría, y especialmente en el área de los modelos de ecuaciones simultáneas, como se ha puesto de manifiesto en el ejemplo anterior. Sin embargo, este problema se plantea también en relación con algunos modelos uniecuacionales, como el modelo de «retardos distribuidos» o el de «errores en las variables», que se estudiarán en capítulos posteriores. En estos casos, el fracaso del método de estimación de los mínimos cuadrados ha dado lugar al desarrollo de métodos de estimación alternativos que proporcione estimaciones consistentes.

tes. Estos métodos se estudiarán en relación con los modelos para los cuales se construyeron o a los que han resultado ser aplicables.

EJERCICIOS:

8-1 Obtener las fórmulas (8.3) y (8.4).

8-2. Demostrar que si x_i^2 y σ_i^2 están positivamente asociadas, el sesgo en la ecuación (8.9) es negativo.

8-3. Considerar el modelo siguiente:

$$Y_i = \alpha + \varepsilon_i,$$

en el que

$$\varepsilon_i \sim N(0, \sigma^2 X_i),$$

$$E(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j),$$

y X_i es no estocástica. Obtener el estimador lineal insesgado óptimo de α y su variancia.

8-4. Utilizando los datos sobre la renta (X) y el gasto (Y) de una muestra de economías domésticas, que aparecen en el cuadro 8-6, estimar la ecuación de regresión

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

para los siguientes modelos heteroscedásticos:

a. $\text{Var}(\varepsilon_i) = \sigma^2 X_i^2$. b. $\text{Var}(\varepsilon_i) = \sigma^2 [E(Y_i)]^2$. c. $\text{Var}(\varepsilon_i) = \sigma_i^2$.

Cuadro 8-6

X	Y										Total
	14	19	21	23	25	27	29	31	33	35	
18	74	13	7	1							95
23	6	4	2	7	4						23
25	2	3	2	2	4						13
27	1	1	2	3	3	2					12
29	2		1	3	2		6				14
31	2		2	1			1	2	1		9
33				2			1	1	3		7
35		1		1					2		4
37					1	1		1		1	4

8-5 Demostrar la validez de la fórmula (8.38).

8-6. Demostrar que, cuando $\rho = 0$, las fórmulas (8.38) a (8.41) se convierten en las fórmulas minimocuadráticas.

8-7. Demostrar que la fórmula minimocuadrática convencional para la estimación de la variancia del estimador da lugar a resultados sesgados en el caso (8.57).

8-8. Una posible representación de la función agregada de consumo para la economía de los Estados Unidos es:

$$C_t = \alpha + \beta Y_t + \varepsilon_t,$$

siendo C = gastos personales en consumo a precios constantes, Y = renta personal disponible a precios constantes, y t = tiempo expresado en años.

a. Suponiendo que se satisfacen todos los supuestos básicos del modelo de regresión simple, estimar la relación a partir de los datos del cuadro 8-7, siguiendo el método de cálculo indicado en el Apéndice C.

Cuadro 8-7

Año	Gastos personales de consumo *		Renta personal disponible ^a a precios corrientes †
	a precios corrientes (1)	a precios constantes de 1954 (2)	
1948	178.3	199.3	189.3
1949	181.2	204.3	189.7
1950	195.0	216.8	207.7
1951	209.8	218.5	227.5
1952	219.8	224.2	238.7
1953	232.6	235.1	252.5
1954	238.0	238.0	256.9
1955	256.9	256.0	274.4
1956	269.9	264.3	292.9
1957	285.2	271.2	308.8
1958	293.2	273.2	317.9
1959	313.5	288.9	337.1
1960	328.2	298.1	349.9
1961	337.3	303.8	364.7
1962	356.8	318.5	384.6
1963	375.0	330.6	402.5
1964	399.3	347.5	431.8

Fuentes: 1948-1960: *Business Statistics*, 1963 ed., U.S. Department of Commerce; 1961-1964: *Statistical Abstract of the United States*, 1965, U.S. Department of Commerce.

* En miles de millones de dólares.

† Para obtener la renta personal disponible «real», puede dividirse por el deflector de precios implícito en las columnas (1) y (2).

b. En base a los resultados obtenidos en a, efectuar la contrastación de Durbin-Watson para la existencia de autorregresión en los términos de perturbación.

c. Suponiendo que los términos de perturbación siguen un esquema autorregresivo de primer orden, obtener estimaciones asintóticamente eficientes de los coeficientes de regresión y de sus errores standard estimados.

8-9. Obtener la fórmula indicada en (8.55).

8-10. Se ha especificado una función agregada de consumo de alimentos en la forma

$$Y_t = \alpha + \beta Z_t + \varepsilon_t,$$

siendo Y = gastos personales en alimentos, y Z = gastos totales personales de consumo. Los datos correspondientes a los años 1929-1964, para los Estados Unidos, se han presentado en el ejercicio 7-8 (cuadro 7-4). Efectuar una contrastación de la ausencia de autorregresión positiva

a. Mediante la contrastación de Durbin-Watson.

b. Mediante la contrastación de la máxima verosimilitud para muestras grandes.

Estimación con datos deficientes

En el capítulo 8 hemos tratado el problema de la estimación de los coeficientes de las ecuaciones de regresión lineales, cuando no se cumple alguno de los supuestos básicos del modelo clásico de regresión normal lineal. En este capítulo nos ocuparemos de la estimación en situaciones en que los datos muestrales resultan deficientes en algún sentido. Las deficiencias concretas que vamos a estudiar

1. Existencia de errores de medición.
2. Datos agrupados.
3. Carencia de algunas observaciones.

Estas deficiencias, con las que nos podemos encontrar fácilmente en la práctica, dan lugar a problemas de estimación que merecen especial atención.

9.1 Errores de medición

Hasta este momento, hemos dado siempre por supuesto que en el modelo de regresión los valores de las variables se miden sin error alguno. Esta presunción ha estado implícita a lo largo de todo el análisis, y todas las fórmulas obtenidas se basan en ella. En este apartado vamos a ocuparnos de lo que les ocurre a los estimadores de los coeficientes de regresión cuando esto no es cierto, y de qué puede hacerse para solucionar este tipo de situaciones. Para ello nos limitaremos únicamente a los casos en los que puede suponerse que los errores de medición son aleatorios y poseen unas características probabilísticas determinadas. Este tratamiento convencional de los errores de medición en la literatura estadística y econométrica corresponde a una gran variedad de situaciones que se presentan en la vida real.

Empecemos por considerar el problema de los errores de medición en el

contexto del modelo clásico de regresión normal lineal estudiado en el capítulo 7. Como en aquel caso, la ecuación de regresión es

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Y se mantienen también todos los supuestos básicos (7.2) a (7.6). Supongamos ahora que las observaciones de X e Y contienen errores, de modo que, en lugar de observar X_i e Y_i , observamos X_i^* e Y_i^* , que pueden definirse como

$$Y_i^* = Y_i + v_i,$$

$$X_i^* = X_i + w_i,$$

en las que v_i y w_i representan los errores en la medición del valor i -ésimo de Y y de X . Suponemos que las características de estos errores son las siguientes:

$$(9.1) \quad v_i \sim N(0, \sigma_v^2),$$

$$w_i \sim N(0, \sigma_w^2);$$

$$(9.2) \quad E(v_i v_j) = 0 \quad (i \neq j),$$

$$E(w_i w_j) = 0 \quad (i \neq j);$$

$$(9.3) \quad E(v_i w_i) = 0,$$

$$E(v_i \varepsilon_i) = 0,$$

$$E(w_i \varepsilon_i) = 0.$$

El supuesto (9.1) indica que cada error es una variable aleatoria normal con media cero y variancia constante. El supuesto (9.2) elimina las situaciones en las que los errores son de tipo autorregresivo, y considerado conjuntamente con el supuesto (9.1) implica que los errores efectuados en un punto de observación son independientes de los efectuados en todos los demás puntos. El supuesto (9.3) indica cómo están relacionados —o mejor dicho, no relacionados— los errores entre sí. Los supuestos (9.1), (9.2), (9.3) y los supuestos básicos acerca de la perturbación tomados conjuntamente implican que los errores de medición son independientes entre sí, independientes respecto al término de perturbación de la ecuación de regresión, e independientes de los verdaderos valores de X e Y .

Consideremos ahora la estimación de los coeficientes de regresión a partir de datos de X^* e Y^* . Como

$$Y_i = Y_i^* - v_i$$

y

$$X_i = X_i^* - w_i,$$

la ecuación de regresión puede escribirse en la forma

$$(Y_i^* - v_i) = \alpha + \beta(X_i^* - w_i) + \varepsilon_i$$

(9.4)

$$Y_i^* = \alpha + \beta X_i^* + \varepsilon_i^*,$$

siendo

$$\varepsilon_i^* = \varepsilon_i + v_i - \beta w_i.$$

Ahora bien, cuando el modelo de regresión se escribe en la forma (9.4), en la que la variable explicativa y la dependiente son observables, la variable explicativa X_i^* está correlacionada con el término de perturbación ε_i^* del mismo período. Concretamente,

$$\text{Cov}(X_i^*, \varepsilon_i^*) = E[X_i^* - E(X_i^*)]\varepsilon_i^* = E[w_i(\varepsilon_i + v_i - \beta w_i)] = -\beta\sigma_w^2.$$

Esto implica que el estimador minimocuadrático de β es *inconsistente*, y lo mismo ocurre con el estimador minimocuadrático de α .

Estimación por medio de variables instrumentales

La ecuación (9.4) se denomina a veces modelo con *errores en las variables y errores en la ecuación*. Este modelo no puede estimarse por el método de la máxima verosimilitud, ya que los estimadores resultantes no son consistentes.¹ Un método que sabemos que proporciona estimaciones consistentes en este caso es el *método de las variables instrumentales*. Este método puede utilizarse siempre que podamos hallar una nueva variable Z_i tal que

$$1. \quad \lim \text{prob} (\sum Z_i' \varepsilon_i^*)/n = 0.$$

$$2. \quad \lim \text{prob} (\sum Z_i' X_i^*)/n \text{ es un número finito distinto de cero.}$$

De acuerdo con la notación que ya hemos adoptado anteriormente, definimos

$$z_i' = Z_i - \bar{Z},$$

$$x_i^{*'} = X_i^* - \bar{X}^*.$$

La primera condición se cumplirá si Z_i no está correlacionada con ε_i , v_i , ni w_i . La segunda condición se cumplirá si Z_i y X_i^* están correlacionadas entre sí. Una condición adicional, no necesariamente para la consistencia pero que ayuda a reducir la variancia asintótica del estimador por variables instrumen-

1. Ver J. Johnston, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, pp. 151-52. El fallo del método de la máxima verosimilitud se debe al hecho de que la función de verosimilitud incluye los valores «verdaderos» de la variable explicativa, que no se conocen y, por tanto, deben ser considerados como parámetros desconocidos. Esto significa que, a medida que aumenta el tamaño de la muestra, también aumenta el número de parámetros desconocidos y, en este caso, la función de verosimilitud no tiene un

tales, es que $\lim \text{prob} (\sum z_i' x_i^{**})/n$ sea lo más grande posible —es decir, que el grado de correlación entre Z_i y X_i^* sea grande—. La variable Z_i , que recibe el nombre de «variable instrumental», puede ser estocástica o no estocástica y, por último, no importa que Z_i se mida con error o sin él siempre y cuando se satisfagan las condiciones que se han especificado respecto a los valores observables de Z . Los estimadores de α y β , a los que llamaremos $\alpha^\dagger$ y $\beta^\dagger$ se definen de la forma siguiente:

$$(9.5a) \quad \beta^\dagger = \frac{\sum y_i^* z_i'}{\sum x_i^* z_i'}$$

$$(9.5b) \quad \alpha^\dagger = \bar{Y}^* - \beta^\dagger \bar{X}^*,$$

$$\text{siendo} \quad y_i^* = Y_i^* - \bar{Y}^*,$$

Por razones de conveniencia de cálculo, los valores de la variable instrumental se miden en unidades tales que $\bar{Z}$ sea igual a $\bar{X}^*$. Esto se puede conseguir fácilmente multiplicando cada valor de la variable instrumental por el factor $(\bar{X}^*/\bar{Z})$, siendo $\bar{Z}^*$ la media muestral de los valores de la variable instrumental medidos en términos de las unidades originarias. Esta transformación no altera para nada las correlaciones entre Z y ϵ^* y entre Z y X^* .

La idea subyacente en el método de las variables instrumentales puede explicarse haciendo referencia a las «ecuaciones normales» minimocuadráticas presentadas en (7.10) y (7.11). En el caso de la ecuación de regresión (9.4), las ecuaciones normales minimocuadráticas son:

$$\sum Y_i^* = \hat{\alpha}n + \hat{\beta} \sum X_i^*, \quad \sum Y_i^* X_i^* = \hat{\alpha} \sum X_i^* + \hat{\beta} \sum X_i^{*2}.$$

El procedimiento para obtener estas ecuaciones es el siguiente. Para obtener la primera ecuación normal, se multiplican los dos lados de la ecuación (9.4) por el multiplicador de α (que es 1), se suman todas las observaciones, y se prescinde del último término correspondiente a la perturbación. Para obtener la segunda ecuación normal, se multiplican los dos lados de la ecuación (9.4) por el multiplicador de β (que es X_i^*), se suman todas las observaciones y se prescinde del último término correspondiente a la perturbación. Para obtener las ecuaciones normales correspondientes a los estimadores del método de las variables instrumentales, procedemos de un modo similar. La primera ecuación normal se obtiene exactamente de la misma forma que la primera ecuación normal del método de los mínimos cuadrados, y la segunda ecuación normal se obtiene multiplicando ambos lados de (9.4) por Z_i en lugar de X_i^* como hacíamos en el caso de los mínimos cuadrados. El resultado es:

$$\sum Y_i^* = \alpha^\dagger n + \beta^\dagger \sum X_i^*, \quad \sum Y_i^* Z_i = \alpha^\dagger \sum Z_i + \beta^\dagger \sum X_i^* Z_i.$$

Las ecuaciones dan lugar a las fórmulas de los estimadores correspondientes al método de las variables instrumentales, presentadas en (9.5a) y (9.5b).

Obsérvese que en el caso de las ecuaciones de regresión en que la variable explicativa X_i es independiente de la perturbación ϵ_i , la propia X_i constituiría una variable instrumental adecuada, ya que no está correlacionada con ϵ_i y, al mismo tiempo, está correlacionada (perfectamente) consigo misma, satisfaciéndose así las condiciones 1 y 2. Así pues, los estimadores minimocuadráticos ordinarios pueden considerarse estimadores por variables instrumentales en los que $Z_i = X_i$.

Vamos a demostrar ahora que los estimadores por variables instrumentales son consistentes. Respecto a $\beta^\dagger$, tenemos

$$\beta^\dagger = \frac{\sum y_i^* z_i'}{\sum x_i^* z_i'} = \beta + \frac{\sum \epsilon_i^* z_i'}{\sum x_i^* z_i'}$$

$$y \quad \lim \text{prob} \beta^\dagger = \beta + \frac{\lim \text{prob} (\sum \epsilon_i^* z_i')/n}{\lim \text{prob} (\sum x_i^* z_i')/n} = \beta$$

según las condiciones que se supone que Z_i debe satisfacer. Del mismo modo,

$$\lim \text{prob} \alpha^\dagger = \lim \text{prob} (\bar{Y}^* - \beta^\dagger \bar{X}^*) = \lim \text{prob} (\alpha + \beta \bar{X}^* + \bar{\epsilon}^* - \beta^\dagger \bar{X}^*) = \alpha$$

Esto muestra que tanto $\alpha^\dagger$ como $\beta^\dagger$ son consistentes. Obsérvese que no podemos demostrar que $\beta^\dagger$ es insesgada, ya que en el último término de

$$\beta^\dagger = \beta + \frac{\sum \epsilon_i^* z_i'}{\sum x_i^* z_i'}$$

no podemos separar z_i' del resto de la expresión al tomar la esperanza matemática (ni aun en el caso en que Z_i es no estocástica).

Las variancias asintóticas de los estimadores por variables instrumentales pueden obtenerse a partir de las fórmulas correspondientes.² Los resultados son:

$$(9.6a) \quad \text{Var asintótica} (\beta^\dagger) = \frac{\sigma_{\epsilon^*}^2 \sum z_i'^2}{(\sum x_i^* z_i')^2},$$

$$(9.6b) \quad \text{Var asintótica} (\alpha^\dagger) = \sigma_{\epsilon^*}^2 \left[\frac{1}{n} + \frac{\bar{X}^{*2} \sum z_i'^2}{(\sum x_i^* z_i')^2} \right],$$

$$\text{siendo} \quad \sigma_{\epsilon^*}^2 = \text{Var}(\epsilon_i^*) = \sigma^2 + \sigma_v^2 + \sigma_w^2.$$

² Para los detalles, véase A. S. Goldberger, *Econometric Theory*. Nueva York, Wiley, 1968, pp. 285-86.

Para obtener estas fórmulas hemos utilizado el hecho de que Z se mide de una forma tal que $\bar{Z} = \bar{X}^*$. El valor de $\sigma_{\epsilon^*}^2$ es desconocido y, por tanto, debe ser estimado. Un estimador consistente de $\sigma_{\epsilon^*}^2$ al que llamaremos s^{*2} viene dado por la expresión

$$(9.7) \quad s^{*2} = \frac{1}{n-2} \sum_i (Y_i^* - \alpha^\dagger - \beta^\dagger X_i^*)^2,$$

es decir, dividiendo la suma de los cuadrados de los residuos por $(n-2)$. Esto nos permite obtener estimadores de las variancias asintóticas de $\alpha^\dagger$ y $\beta^\dagger$ mediante las fórmulas

$$(9.8a) \quad s_{\beta^\dagger}^2 = \frac{s^{*2} \sum z_i'^2}{(\sum x_i^* z_i')^2},$$

$$(9.8b) \quad s_{\alpha^\dagger}^2 = s^{*2} \left[\frac{1}{n} + \frac{\bar{X}^{*2} \sum z_i'^2}{(\sum x_i^* z_i')^2} \right].$$

Estos estimadores hacen referencia a las variancias asintóticas, pero se han utilizado como aproximaciones en muestras finitas. Se puede demostrar que los resultados (9.5), (9.7) y (9.8) se cumplen también si la «verdadera» variable explicativa X es estocástica pero independiente de ϵ^* .

La única limitación en la elección de la variable instrumental viene dada por el requisito de que esta última debe estar correlacionada con X y no correlacionada con ϵ^* . Pueden existir, y en general existirán, muchas variables que pueden utilizarse como instrumentos, y trataremos de elegir la que posea una mayor correlación con X^* , ya que ésta será la que proporciona unas variancias asintóticas de los estimadores más pequeñas. Por ejemplo, cuando los datos estén en forma de series temporales, una elección adecuada puede ser la de $Z_t = X_{t-1}^*$, es decir el valor retardado de la variable explicativa. Si ϵ_t^* no es autorregresiva, X_{t-1}^* no estará correlacionada con ϵ_t^* y, al mismo tiempo, es probable que esté muy correlacionada con X_t^* . Sin embargo, no podemos estar nunca seguros de que es imposible obtener mejores resultados eligiendo una variable instrumental distinta de la que hemos seleccionado. Por tanto, no podemos afirmar que un estimador particular por variables instrumentales posee la mínima variancia asintótica, es decir, que es asintóticamente eficiente. La única propiedad deseable de los estimadores por variables instrumentales es, pues, la consistencia. Esta limitación, junto a la arbitrariedad en la elección de las variables instrumentales y a la dificultad de comprobar que la variable elegida no está realmente correlacionada con ϵ^* , hacen que el método no sea muy atractivo. Desgraciadamente, cuando se trata de modelos con errores en las variables y errores en la ecuación, no disponemos de ningún método mejor —por lo menos en el contexto de la teoría clásica del muestreo—. La tendencia existente en los trabajos econométricos previos ha sido la de esquivar el pro-

blema suponiendo que los errores de medición son tan pequeños que puede prescindirse de ellos sin graves inconvenientes.

EjemPlo. En un famoso artículo sobre la sustitución entre trabajo y capital,³ sus autores estiman la relación siguiente:

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

en la que X_i e Y_i se definen en la forma siguiente

$$X_i = \log(W/P)_i,$$

$$Y_i = \log(V/PL)_i,$$

siendo W = tasa de salario monetario, P = precio del producto, V = valor añadido en la producción, L = cantidad de trabajo, y el subíndice i se refiere al país i -ésimo. Cada observación corresponde a una industria concreta en un año determinado. En la estimación, los autores utilizaron el método de los mínimos cuadrados. Como es muy probable que las observaciones de X e Y contengan errores de medición, podemos estimar de nuevo la relación utilizando el método de las variables instrumentales. Nos limitaremos únicamente a la industria de la fabricación de muebles (grupo 260), que es una de las dos industrias para las cuales el artículo proporciona los datos correspondientes a precios. Como variable instrumental utilizaremos $\log(W/P)_i$ correspondiente a la producción de géneros de punto (grupo 232), efectuando el ajuste correspondiente para que $\bar{Z} = \bar{X}$. Es muy probable que esta variable instrumental satisfaga los dos requisitos para la consistencia: es muy poco probable que la tasa de salarios en las fábricas de géneros de punto esté relacionada con las perturbaciones (y los errores de medición) de la industria de la fabricación de muebles, pero es probable que estén relacionadas con la tasa de salarios en dicha industria. En el cuadro 9-1 se presentan los datos correspondientes. La variable precio está expresada en forma de índice con el valor correspondiente a los Estados Unidos igual a 1. Si los errores de medición son insignificantes, es decir, si $X_i^* = X_i$ y $Y_i^* = Y_i$, los estimadores minimocuadráticos serán adecuados. Estos estimadores son

$$Y_i = -2.2877 + 0.8401 X_i + \epsilon_i, \quad R^2 = 0.986.$$

(0.0996) (0.0331)

Por otra parte, si existen serios errores de medición, serían preferibles las estimaciones obtenidas por el método de las variables instrumentales, ya que éstas son consistentes. La ecuación de regresión estimada por este método es

$$Y_i^* = -2.2978 + 0.8435 X_i^* + \epsilon_i^*, \quad R^2 = 0.985.$$

(0.1025) (0.0342)

Los resultados obtenidos por el método de las variables instrumentales son muy similares a los obtenidos por el método de los mínimos cuadrados.

3. K. J. Arrow, et al., «Capital-Labor Substitution and Economic Efficiency», *Review of Economics and Statistics*, vol. 43, agosto 1961, pp. 225-50.

Cuadro 9-1

País	Industria de la fabricación de muebles				Fabricación de géneros de punto			
	L/V	P	W (\$)	$\log(V/PL) = Y^*$	$\log(W/P) = X^*$	P	W (\$)	$\log(W/P) = Z^*$
Estados Unidos	0.1706	1.0000	3515	0.7680	3.5459	1.0000	2698	3.4310
Canadá	0.2385	1.5470	2668	0.4330	3.2367	1.4891	2260	3.1812
Nueva Zelanda	0.3678	0.9482	1834	0.4575	3.2865	1.0346	1548	3.1750
Australia	0.3857	0.8195	1713	0.5002	3.3202	0.7358	1487	3.3055
Dinamarca	0.5040	0.8941	1288	0.3462	3.1585	0.7713	1169	3.1806
Noruega	0.5228	0.9437	1342	0.3068	3.1529	0.8973	1021	3.0553
Reino Unido	0.6291	0.6646	1078	0.3787	3.2101	0.6030	802	3.1238
Colombia	0.7200	1.8260	738	-0.1188	2.6066	2.2570	845	2.5733
Brasil	0.9415	1.4590	448	-0.1379	2.4872	0.9720	364	2.5734
México	0.9017	1.7580	471	-0.2001	2.4280	1.2458	546	2.6417
Argentina	1.0863	2.2300	464	-0.3845	2.3182	1.3901	523	2.5755
								3.4241
								3.1748
								3.1686
								3.2989
								3.1742
								3.0492
								3.1175
								2.5681
								2.5682
								2.6364
								2.5703

Modelo generalizado con errores en la ecuación

Al describir la ecuación (9.4), nos referimos a ella como un modelo con «errores en las variables y errores en la ecuación». La expresión *errores en las variables* se utiliza para describir los errores de medición v_i y w_i , mientras que la expresión *errores en la ecuación* se refiere al término estocástico de perturbación ε_i . Por tanto, un modelo de regresión en el que las variables se miden sin errores puede denominarse modelo con «errores en la ecuación». Consideremos ahora un modelo en el que los errores de medición sólo ocurren en relación con la variable dependiente. Es decir, consideremos

$$w_i = 0 \quad \text{para todo } i.$$

En este caso, sustituyendo

$$Y_i = Y_i^* - v_i$$

en la ecuación de regresión

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

da como resultado

$$(9.9) \quad Y_i^* = \alpha + \beta X_i + \eta_i,$$

siendo

$$\eta_i = \varepsilon_i + v_i.$$

Como suponemos que ε_i y v_i son normales e independientes entre sí, tenemos

$$\eta_i \sim N(0, \sigma_\eta^2),$$

en la que

$$\sigma_\eta^2 = \sigma^2 + \sigma_v^2.$$

La ecuación (9.9) se puede denominar *modelo generalizado con errores en la ecuación*. El término η_i puede considerarse como una variable aleatoria que se comporta exactamente de la misma forma que ε_i y, por tanto, la ecuación (9.9) equivale formalmente a la ecuación de regresión (7.1) del modelo clásico de regresión normal lineal. Esto significa que los estimadores mínimo-cuadráticos de los coeficientes de regresión poseen las propiedades deseadas. Sin embargo, η_i y ε_i difieren en dos aspectos que resulta interesante señalar:

- La variancia de ε_i refleja las características de los factores que influyen en la variable dependiente, junto a la influencia sistemática de la variable explicativa. Como regla general, el observador no puede

alterar esta variancia. Por otra parte, la variancia del error de medición v_i —y en consecuencia, la variancia η_i — se puede reducir mejorando los métodos de medición.

- B. Como en la práctica los métodos de medición de las variables económicas han mejorado durante las últimas décadas, es posible que la variancia de v_i —y en consecuencia, la variancia de η_i — haya disminuido a lo largo del tiempo. Por tanto, si v_i no es insignificante, cuando las observaciones toman la forma de series temporales η_i puede ser heteroscedástica.

Sin embargo, estas diferencias no dan lugar a ningún problema de estimación suficientemente importante como para dedicarle mayor discusión.

Supongamos ahora que, en lugar de la variable dependiente, es la variable explicativa la que presenta errores de medición. Es decir, consideremos la situación en que

$$v_i = 0 \quad \text{para todo } i.$$

En este caso, tenemos

$$X_i = X_i^* - w_i,$$

que puede sustituirse en la ecuación de regresión,

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

obteniendo

$$(9.10) \quad Y_i = \alpha + \beta X_i^* + \zeta_i,$$

en la que

$$\zeta_i = \varepsilon_i - \beta w_i.$$

En este caso, X_i^* está correlacionada con ζ_i en el mismo período. En realidad,

$$\text{Cov}(X_i^*, \zeta_i) = E[X_i^* - E(X_i^*)][\zeta_i - E(\zeta_i)] = E(w_i \zeta_i) = -\beta \sigma_w^2 \neq 0,$$

o sea, que los estimadores minimocuadráticos de α y β en (9.10) son *inconsistentes*. Se pueden obtener estimadores consistentes utilizando el método de las variables instrumentales que se ha descrito anteriormente.

Modelos con errores en las variables

Por último, llegamos a un modelo que se utiliza muy a menudo en las ciencias naturales, pero raramente en la econometría moderna. Lo incluimos aquí

en parte para tener en cuenta todas las posibilidades, y, en parte, debido a que posee una importante aplicación en relación con la llamada «hipótesis de la renta permanente». Este modelo se conoce generalmente con el nombre de modelo con *errores en las variables*, y la idea subyacente es la de que la relación entre la variable explicativa y la variable dependiente es de tipo determinístico, y que la única razón por la que los puntos de la muestra no se hallan situados todos sobre una línea recta es porque no disponemos de mediciones exactas de X e Y . En la terminología que venimos utilizando, el modelo de los errores en las variables se caracteriza por el supuesto de que

$$\varepsilon_i = 0 \quad \text{para todo } i.$$

Generalmente, este modelo no se utiliza en econometría debido a la opinión general de que la perturbación estocástica es una parte esencial de las relaciones económicas. Una excepción a esta regla la constituye la formulación de la hipótesis de la «renta permanente» por Friedman,⁴ que postuló una relación determinística entre «consumo permanente» y «renta permanente». En este caso, las mediciones del consumo y la renta son erróneas, ya que contienen «componentes transitorios» que pueden considerarse como errores de medición. Existen también otros casos en los que cabe esperar razonablemente que los errores de medición sean sustanciales, mientras que los términos de perturbación de la regresión pueden considerarse relativamente sin importancia. Por ejemplo, las estadísticas internacionales son a menudo muy poco precisas, y en tales casos puede ser preferible el uso del modelo de errores en las variables, basándonos en el supuesto de que las inexactitudes debidas a la no consideración del término de perturbación de la regresión no son importantes.

Supongamos, pues, que la relación entre X e Y es

$$(9.11) \quad Y_i = \alpha + \beta X_i,$$

en la que X_i puede ser estocástica o no estocástica. Si X_i es estocástica, suponemos que es independiente de v_i y w_i . Como no observamos X_i e Y_i , sino X_i^* e Y_i^* , tenemos que sustituir

$$Y_i = Y_i^* - v_i \quad \text{y} \quad X_i = X_i^* - w_i,$$

para obtener

$$(9.12) \quad Y_i^* = \alpha + \beta X_i^* + \xi_i,$$

en la que

$$\xi_i = v_i - \beta w_i.$$

⁴ Milton Friedman, *A Theory of the Consumption Function*. Princeton, Princeton University Press, 1957.

Como X_i^* y ζ_i están claramente correlacionadas, los estimadores minimocuadráticos de α y β son *inconsistentes*. Para obtener estimadores consistentes podemos utilizar el método de las variables instrumentales, suponiendo que podamos hallar un instrumento que satisfaga las condiciones de no estar correlacionado con ζ_i y estar correlacionado con X_i^* .

Método de las medias de grupo

Un procedimiento alternativo para estimar (9.12) es el *método de las medias de grupos*, el cual, en su forma más sencilla, requiere que los pares de valores observados (X_i^* , Y_i^*) se ordenen según la magnitud de X_i^* , de modo que

$$X_1^* \leq X_2^* \leq \dots \leq X_n^*.$$

A continuación, estos pares de valores se dividen en tres grupos, aproximadamente del mismo tamaño. Los estimadores de los coeficientes de regresión por el método de las medias de grupos son

$$(9.13a) \quad \beta^{++} = \frac{\bar{Y}_3^* - \bar{Y}_1^*}{\bar{X}_3^* - \bar{X}_1^*},$$

$$(9.13b) \quad \alpha^{++} = \bar{Y}^* - \beta^{++}\bar{X}^*,$$

en las que $\bar{Y}^*$ e $\bar{X}_1^*$ son las medias de Y correspondientes al primer y tercer grupo respectivamente, y $\bar{X}_1^*$ y $\bar{X}_3^*$ son las correspondientes medias de X . Los estimadores por medias de grupo son consistentes siempre que la agrupación se haga de forma que, de haberse agrupado los datos en base a las X no observables en lugar de según las X^* observadas, ninguno de los pares de valores del primer grupo quedase reasignado al tercero, y viceversa. Es interesante señalar que el método de las medias de grupos se puede considerar como un caso especial del método de las variables instrumentales. Si los tres grupos de observaciones son del mismo tamaño, y los valores de la variable instrumental Z_i son tales que

$$\begin{aligned} Z_i &= -1 \text{ si } i \text{ pertenece al primer grupo,} \\ &= 0 \text{ si } i \text{ pertenece al segundo grupo,} \\ &= 1 \text{ si } i \text{ pertenece al tercer grupo,} \end{aligned}$$

venimos que las expresiones (9.5a) y (9.5b) equivalen a las expresiones (9.13a) y (9.13b), respectivamente. En este caso, podemos estimar las variaciones de α^{++} y β^{++} utilizando las fórmulas (9.8a) y (9.8b).

EJEMPLO. Calculemos los estimadores por medias de grupos, correspondientes y (9.13b), respectivamente. En este caso, podemos estimar las variancias de

a la relación entre $\log(V/PL)$ y $\log(W/P)$ que hemos estudiado en el ejemplo anterior. A partir de los datos del cuadro 9-1 vemos que se dispone de once observaciones; utilizaremos cuatro observaciones para el primer grupo y cuatro para el tercero, dejando tres sin utilizar. Los cálculos son los siguientes

$$\bar{X}_1^* = \frac{1}{4}(2.3182 + 2.4280 + 2.4872 + 2.6066) = 2.4600,$$

$$\bar{X}_3^* = \frac{1}{4}(3.2367 + 3.2865 + 3.3202 + 3.5459) = 3.3473,$$

$$\bar{Y}_1^* = \frac{1}{4}(-0.3845 - 0.2001 - 0.1379 - 0.1188) = -0.2103$$

$$\bar{Y}_3^* = \frac{1}{4}(0.4330 + 0.4575 + 0.5002 + 0.7680) = 0.5397.$$

Y, por tanto,

$$\beta^{++} = \frac{0.5397 - (-0.2103)}{3.3473 - 2.4600} = 0.8453$$

$$\alpha^{++} = \frac{1}{11}(2.3491 - 0.8453 \times 32.7508) = -2.3031.$$

En este caso, la relación estimada sería

$$Y_i^* = -2.3031 + 0.8453X_i^* + \zeta_i,$$

que, numéricamente, es muy parecida al resultado anteriormente obtenido.

Regresión ponderada

Si podemos efectuar un supuesto a priori respecto a la razón entre las variancias de los errores, existe otro método que nos permite obtener estimadores de los coeficientes del modelo con errores en las variables. Concretamente, supongamos

$$\frac{\sigma_w^2}{\sigma_v^2} = \lambda,$$

en la que λ tiene un valor conocido. Por ejemplo, en ocasiones cabe esperar razonablemente que los errores en la medición de X tengan aproximadamente la misma dispersión que los errores efectuados al medir Y . En este caso, puede ser conveniente suponer que $\lambda = 1$. Obtengamos ahora estimadores consistentes de α y β para un valor determinado de λ . Frecuentemente, estos esti-

madores reciben el nombre de *estimadores de regresión ponderada*. Podemos empezar por considerar el estimador minimocuadrático de β , que es

$$\hat{\beta} = \frac{\sum y_i^{*'} x_i^{*'}}{\sum x_i^{*'^2}}.$$

Como

$$y_i^{*'} = y_i' + v_i' = \beta x_i' + v_i',$$

y

$$x_i^{*'} = x_i' + w_i',$$

podemos escribir $\hat{\beta}$ en la forma

$$\hat{\beta} = \frac{\sum (\beta x_i' + v_i')(x_i' + w_i')}{\sum (x_i' + w_i')^2} = \frac{\beta \sum x_i'^2 + \sum x_i' v_i' + \beta \sum x_i' w_i' + \sum v_i' w_i'}{\sum x_i'^2 + 2 \sum x_i' w_i' + \sum w_i'^2}.$$

El límite probabilístico de $\hat{\beta}$ es

$$(9.14) \quad \lim \text{prob } \hat{\beta} = \frac{\beta \sigma_x^2}{\sigma_x^2 + \sigma_w^2},$$

en la que el término σ_x^2 representa la variancia de X si ésta es estocástica, o $\lim (\sum x_i'^2)/n$ si es no estocástica. Para obtener (9.14) hemos utilizado los supuestos de independencia entre v_i , w_i y X_i . La fórmula (9.14) indica que la causa de la inconsistencia de $\hat{\beta}$ es la presencia del término σ_w^2 en el denominador. Si este término no apareciera ahí, el término σ_x^2 del numerador se eliminaría con el del denominador y el estimador sería consistente. Esto sugiere que podríamos obtener un estimador consistente de β a través de

$$(9.15) \quad \hat{\beta} = \frac{(\sum y_i^{*'} x_i^{*'})/n}{(\sum x_i^{*'^2})/n - \hat{\sigma}_w^2},$$

en la que $\hat{\sigma}_w^2$ es un estimador consistente de σ_w^2 . El problema se reduce, pues, a encontrar una expresión correspondiente a un estimador consistente de σ_w^2 en términos de X^* , Y^* y β . Sustituyendo esta expresión en (9.15) podríamos despejar $\hat{\beta}$ que, en tal caso, sería consistente.

Si conocemos λ , tenemos

$$\sigma_w^2 = \lambda \sigma_v^2,$$

que puede escribirse en la forma

$$\begin{aligned} \lim \text{prob } \frac{1}{n} \sum w_i'^2 &= \lambda \lim \text{prob } \frac{1}{n} \sum v_i'^2 = \lambda \lim \text{prob } \frac{1}{n} \sum (y_i^{*'} - y_i')^2 \\ &= \lambda \lim \text{prob } \frac{1}{n} \sum (y_i^{*'} - \beta x_i')^2 \\ &= \lambda \left[\lim \text{prob } \frac{1}{n} \sum y_i^{*'^2} - 2\beta \lim \text{prob } \frac{1}{n} \sum y_i^{*'} x_i' + \beta^2 \lim \text{prob } \frac{1}{n} \sum x_i'^2 \right] \end{aligned}$$

Ahora es preciso eliminar x_i' , ya que no es observable. Sustituyendo $x_i' = x_i^{*'} - w_i'$ obtenemos

$$\lim \text{prob } \frac{1}{n} \sum y_i^{*'} x_i' = \lim \text{prob } \frac{1}{n} \sum y_i^{*'} (x_i^{*'} - w_i') = \lim \text{prob } \frac{1}{n} \sum y_i^{*'} x_i^{*'} - \lim \text{prob } \frac{1}{n} \sum y_i^{*'} w_i'$$

$$\text{y} \quad \lim \text{prob } \frac{1}{n} \sum x_i'^2 = \lim \text{prob } \frac{1}{n} \sum (x_i^{*'} - w_i')^2 =$$

$$\lim \text{prob } \frac{1}{n} \sum x_i^{*'^2} - \lim \text{prob } \frac{1}{n} \sum w_i'^2.$$

Vamos a introducir ahora la notación abreviada siguiente:

$$\frac{1}{n} \sum y_i^{*'^2} = m_{yy}^*,$$

$$\frac{1}{n} \sum x_i^{*'^2} = m_{xx}^*,$$

$$\frac{1}{n} \sum x_i^{*'} y_i^{*'} = m_{xy}^*.$$

Utilizando esta notación, y sustituyendo los límites probabilísticos por sus contrapartidas muestrales, obtenemos

$$\hat{\sigma}_w^2 = \lambda [m_{yy}^* - 2\beta m_{xy}^* + \beta^2 (m_{xx}^* - \hat{\sigma}_w^2)].$$

Despejando $\hat{\sigma}_w^2$ en esta expresión tenemos

$$(9.16) \quad \hat{\sigma}_w^2 = \frac{\lambda}{1 + \lambda \beta^2} [m_{yy}^* - 2\beta m_{xy}^* + \beta^2 m_{xx}^*].$$

Esta es la expresión de $\hat{\sigma}_w^2$ que deseábamos obtener, en términos de las cantidades observables m_{yy}^* , m_{xy}^* , m_{xx}^* , y en términos de β . Sustituyendo $\hat{\sigma}_w^2$ en (9.15) por su valor según la expresión (9.16), obtenemos

$$\hat{\beta} = \frac{m_{xy}^*}{m_{xx}^* - [\lambda/(1 + \lambda\beta^2)](m_{yy}^* - 2\beta m_{xy}^* + \beta^2 m_{xx}^*)}$$

que puede expresarse en forma de la ecuación cuadrática

$$(9.17) \quad \beta^2 \lambda m_{xy}^* + \beta(m_{xx}^* - \lambda m_{yy}^*) - m_{xy}^* = 0.$$

Esta ecuación proporciona dos valores de $\hat{\beta}$, uno positivo y el otro negativo. La elección entre estos dos viene determinada por el valor de

$$\sum (y_i^* - \beta x_i^*)^2 = m_{yy}^* - 2\beta m_{xy}^* + \beta^2 m_{xx}^*,$$

es decir, por el valor de la suma de los cuadrados de las desviaciones respecto de la recta de regresión ajustada. Como queremos que este valor sea pequeño, elegiremos el valor de $\hat{\beta}$ que tenga el mismo signo que m_{xy}^* . El correspondiente estimador consistente de α es

$$(9.18) \quad \hat{\alpha} = \bar{Y}^* - \hat{\beta} \bar{X}^*.$$

Puede mostrarse que $\hat{\alpha}$ y $\hat{\beta}$ constituyen los estimadores maximoverosímiles de los parámetros respectivos y, por tanto, son asintóticamente eficientes.⁵

Podemos obtener una cierta comprensión del principio de la estimación por medio de regresión ponderada —y explicar la razón de este nombre— si la comparamos con la estimación por medio de mínimos cuadrados. Supongamos que la variable «explicativa» X se mide sin ningún error. En este caso, el modelo de los errores en las variables se reduce a

$$Y_i^* = \alpha + \beta X_i + v_i,$$

y los estimadores minimocuadráticos ordinarios son consistentes, además de tener todas las propiedades deseables. Obsérvese que, en este caso, la estimación minimocuadrática se basa en la idea de minimizar la suma de los cuadrados de las desviaciones de las Y^* observadas, respecto de la recta de regresión. Este procedimiento se denomina «minimización en la dirección de Y »,

5. Ver Johnston, *op. cit.*, pp. 152-55. A Madansky, «The Fitting of Straight Lines When Both Variables Are Subject to Error», *Journal of the American Statistical Association*, vol. 54, marzo 1959, pp. 179-80, presenta una fórmula aproximada (y muy complicada) de las variancias asintóticas de estos estimadores.

y consiste en minimizar la suma de los cuadrados de las distancias *verticales* entre los puntos observados y la recta de regresión. Los casos en los que X se observa sin ningún error son aquellos en los que σ_w^2 es igual a cero, ya que, si no existen errores de medición, su variancia es obviamente cero. Esto implica que $\lambda = 0$, y sustituyendo este valor en la ecuación (9.17) obtenemos

$$\hat{\beta} = \frac{m_{xy}^*}{m_{xx}^*},$$

que es la fórmula del estimador minimocuadrático. En el otro extremo, supongamos que no existe ningún error en la medición de Y . En este caso, el modelo de los errores en las variables se reduce a

$$Y_i = \alpha + \beta(X_i^* - w_i)$$

o

$$X_i^* = -\frac{\alpha}{\beta} + \frac{1}{\beta} Y_i + w_i.$$

Ahora podemos considerar a X_i^* como la variable «dependiente», y a Y_i como la variable «explicativa». El estimador minimocuadrático de $(1/\beta)$, que es consistente, es

$$\left(\frac{1}{\hat{\beta}}\right) = \frac{m_{xy}^*}{m_{yy}^*}.$$

A partir de este estimador, podemos obtener un estimador consistente de β , que vendrá dado por

$$\hat{\beta} = \frac{m_{yy}^*}{m_{xy}^*}.$$

En realidad, al obtener este estimador estamos minimizando la suma de los cuadrados de las desviaciones de las X^* observadas, respecto de la recta de regresión, es decir, estamos «minimizando en la dirección de X ». Esto significa minimizar la suma de los cuadrados de las distancias *horizontales* entre los puntos observados y la recta de regresión. Obsérvese que, al no cometer errores en la observación de Y , esto significa que $\sigma_v^2 = 0$ y, por tanto, $\lambda \rightarrow \infty$. Dividiendo la ecuación (9.17) por λ obtenemos

$$\beta^2 m_{xy}^* + \beta \left(\frac{m_{xx}^*}{\lambda} - m_{yy}^* \right) - \frac{m_{xy}^*}{\lambda} = 0.$$

Cuando $\lambda \rightarrow \infty$, esto se reduce a

$$\beta^2 m_{xy}^* - \beta m_{yy}^* = 0$$

o

$$\beta = \frac{m_{yy}^*}{m_{xy}^*},$$

que es exactamente el mismo resultado que hemos obtenido para $\hat{\beta}$. Esto indica que el tamaño de λ determina la dirección de la minimización. Cuando el valor de λ está comprendido entre 0 e ∞ , servirá como factor de «ponderación» señalando la dirección de la minimización entre X e Y .

EJEMPLO. Siguiendo con el ejemplo de la relación entre $\log(V/PL)$ y $\log(W/P)$, calculemos las estimaciones de α y β , por el método de la regresión ponderada, en el supuesto de que $\lambda = 1$. Los cálculos correspondientes son

$$m_{xx}^* = 20.1834/11^2,$$

$$m_{yy}^* = 14.4456/11^2,$$

$$m_{xy}^* = 16.9577/11^2,$$

y, por tanto,

$$m_{xx}^* - m_{yy}^* = 5.7378/11^2.$$

Sustituyendo este valor en la ecuación (9.17) obtenemos

$$\beta = \frac{-5.7378 + \sqrt{5.7378^2 + 4 \times 16.9577^2}}{2 \times 16.9577} = 0.8452.$$

Del mismo modo, si utilizamos la ecuación (9.18), obtenemos

$$\hat{\alpha} = -2.3023.$$

Conclusión

Esto finaliza nuestro análisis de los errores de medición. Como hemos señalado al principio, es probable que estos errores estén presentes en la mayor parte de las observaciones económicas, o en todas ellas. Hemos visto que es muy difícil resolver el problema de la estimación cuando existen perturbaciones estocásticas y errores de medición —la situación descrita por el modelo con «errores en las variables y errores en la ecuación». La práctica común en la econometría moderna ha consistido en prescindir de los errores de medición, concentrándose en el «modelo con errores en la ecuación». La razón para ello no ha sido la opinión de que los datos son perfectos, sino la de que los errores de medición son de poca importancia en comparación con el papel que desempeñan las perturbaciones estocásticas en las relaciones económicas.

Por nuestra parte, seguiremos adelante con este mismo espíritu, constatando que el área de los errores de medición representa un desafío para la investigación econométrica del futuro.

9.2 Estimación a partir de datos agrupados

Los datos obtenidos en estudios o encuestas se presentan a menudo en forma de tablas que resumen los valores correspondientes a las observaciones individuales. Para estimar los coeficientes de las ecuaciones de regresión, se utilizan a menudo estas informaciones tabuladas. Una razón para ello es la de que, probablemente, los resultados presentados en esta forma son fácilmente accesibles, mientras que la recuperación de las observaciones individuales puede resultar costosa y requerir mucho tiempo. Otro motivo para utilizar resúmenes condensados de las observaciones muestrales es el deseo de evitar los cálculos en gran escala que pueden ser necesarios si se utilizan los datos en su forma original. En esta sección estudiaremos en qué forma se ven afectadas las propiedades de los estimadores de los coeficientes de regresión, por el uso de información muestral condensada, en el modelo de regresión simple. Consideramos dos tipos de información tabulada, uno de los cuales representa el caso en el que las observaciones se agrupan según el valor de una de las variables, mientras que en el otro las observaciones se agrupan según los valores de las dos variables. Los cuadros respectivos se conocen generalmente con el nombre de cuadros de clasificación *simple* y cuadros de *doble entrada*. Naturalmente, los cuadros de clasificación simple representan un mayor grado de agrupación de la información muestral que los de doble entrada, y otra forma de expresar lo mismo, es decir, que los cuadros de clasificación simple implican una mayor pérdida de información que los de doble entrada.

Agrupación simple

Consideremos un modelo de regresión lineal simple como el descrito por las condiciones (7.1) a (7.6). Como se recordará, esto significa que la variable explicativa X es no estocástica, y que el término de perturbación de la regresión, ϵ , tiene media cero y es normal, homoscedástico y no autorregresivo. (Sin embargo, las conclusiones de esta sección se aplican también al caso en que X es estocástica pero independiente de ϵ .) Supongamos que las n observaciones muestrales se dividen en G grupos. Sea n_1 el número de observaciones del primer grupo, n_2 el número de observaciones del segundo grupo, y así sucesivamente. Como en total existen n observaciones, debemos tener

$$\sum_{g=1}^G n_g = n.$$

Indiquemos la observación i -ésima del grupo g mediante el doble subíndice ig , de modo que la ecuación de regresión puede escribirse en la forma

$$(9.19) \quad Y_{ig} = \alpha + \beta X_{ig} + \varepsilon_{ig} \quad (i = 1, 2, \dots, n_g; g = 1, 2, \dots, G).$$

Supongamos ahora que en lugar de disponer de una enumeración completa de todas las observaciones en cada grupo, se nos da únicamente su número y los valores medios (o totales) de X e Y , presentados en la forma siguiente:

Grupo	Número de observaciones	Media de X	Media de Y
1	n_1	$\bar{X}_1$	$\bar{Y}_1$
2	n_2	$\bar{X}_2$	$\bar{Y}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
G	n_G	$\bar{X}_G$	$\bar{Y}_G$

siendo

$$\bar{X}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} X_{ig} \quad \text{y} \quad \bar{Y}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} Y_{ig} \quad (g = 1, 2, \dots, G).$$

El problema consiste ahora en obtener las fórmulas de estimación de los coeficientes de regresión, utilizando las medias de los grupos, y en descubrir cómo se comparan las propiedades de los estimadores resultantes con las de los estimadores minimocuadráticos ordinarios basados en las observaciones individuales. Empleemos la ecuación de regresión (9.19) y «condensémosla» tomando la media respecto a todas las observaciones dentro de cada grupo. De esta forma obtenemos

$$(9.20) \quad \bar{Y}_g = \alpha + \beta \bar{X}_g + \bar{\varepsilon}_g \quad (g = 1, 2, \dots, G).$$

Es decir, hemos sustituido las n observaciones originales por las G medias de los grupos. Si X_{ig} no es estocástica, $\bar{X}_g$ será también no estocástica, de modo que sólo nos tenemos que ocupar de $\bar{\varepsilon}_g$. Observemos en primer lugar que

$$E(\bar{\varepsilon}_g) = E \frac{1}{n_g} (\varepsilon_{1g} + \varepsilon_{2g} + \dots + \varepsilon_{n_g g}) = 0,$$

lo que significa que los estimadores minimocuadráticos ordinarios de α y β basados en las medias de los grupos son insesgados. En segundo lugar, para $g \neq h$,

$$E(\bar{\varepsilon}_g \bar{\varepsilon}_h) = E \left[\frac{1}{n_g} (\varepsilon_{1g} + \varepsilon_{2g} + \dots + \varepsilon_{n_g g}) \right] \left[\frac{1}{n_h} (\varepsilon_{1h} + \varepsilon_{2h} + \dots + \varepsilon_{n_h h}) \right] = 0,$$

lo que significa que $\bar{\varepsilon}_g$ no es autorregresivo. Por último,

$$\text{Var}(\bar{\varepsilon}_g) = \frac{1}{n_g^2} (\sigma^2 + \sigma^2 + \dots + \sigma^2) = \frac{n_g \sigma^2}{n_g^2} = \frac{\sigma^2}{n_g},$$

lo que significa que el término de perturbación en (9.20) es *heteroscedástico*, a menos que el número de observaciones sea el mismo en todos los grupos.

La naturaleza heteroscedástica de $\bar{\varepsilon}_g$ implica que los estimadores minimocuadráticos de α y β obtenidos utilizando como «observaciones» las medias de los grupos no son eficientes. Para utilizar de forma eficiente las medias de los grupos tendremos que emplear las fórmulas de estimación obtenidas para las regresiones heteroscedásticas. Estas fórmulas se han obtenido anteriormente y se han presentado en las expresiones (8.1) a (8.4). Por lo que respecta a $\bar{\varepsilon}_g$ estamos en la afortunada posición de conocer exactamente cómo se altera su variancia de una «observación» a otra, es decir, de un grupo a otro, ya que conocemos n_g . Con el fin de adaptar las fórmulas (8.1) a (8.4) al uso de las medias de los grupos, sustituiremos el subíndice i por g , X e Y por $\bar{X}$ e $\bar{Y}$, y ν por (n_g/σ^2) . Los estimadores resultantes de α y β , a los que llamaremos $\tilde{\alpha}$ y $\tilde{\beta}$, son

$$(9.21) \quad \tilde{\beta} = \frac{[(\sum n_g)(\sum n_g \bar{X}_g \bar{Y}_g) - (\sum n_g \bar{X}_g)(\sum n_g \bar{Y}_g)]/\sigma^4}{[(\sum n_g)(\sum n_g \bar{X}_g^2) - (\sum n_g \bar{X}_g)^2]/\sigma^4} = \frac{\sum n_g \bar{X}_g \bar{Y}_g - n \bar{X} \bar{Y}}{\sum n_g \bar{X}_g^2 - n \bar{X}^2} \\ = \frac{\sum n_g (\bar{X}_g - \bar{X})(\bar{Y}_g - \bar{Y})}{\sum n_g (\bar{X}_g - \bar{X})^2}$$

y

$$(9.22) \quad \tilde{\alpha} = \frac{(\sum n_g \bar{Y}_g)/\sigma^2}{(\sum n_g)/\sigma^2} - \tilde{\beta} \frac{(\sum n_g \bar{X}_g)/\sigma^2}{(\sum n_g)/\sigma^2} = \bar{Y} - \tilde{\beta} \bar{X},$$

en la que $\bar{X}$ es la media muestral global de X e $\bar{Y}$ la de Y . Para simplificar las expresiones correspondientes a $\tilde{\alpha}$ y $\tilde{\beta}$ se han utilizado las igualdades siguientes:

$$\sum_g n_g = n_1 + n_2 + \dots + n_G = n,$$

$$\sum_g n_g \bar{X}_g = n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_G \bar{X}_G = \sum_i X_{i1} + \sum_i X_{i2} + \dots + \sum_i X_{iG}$$

$$= \sum_i \sum_g X_{ig} = n \bar{X},$$

$$\sum_g n_g \bar{Y}_g = n_1 \bar{Y}_1 + n_2 \bar{Y}_2 + \dots + n_G \bar{Y}_G = n \bar{Y}.$$

Las variancias de $\bar{\alpha}$ y $\bar{\beta}$ son

$$(9.23) \quad \text{Var}(\bar{\beta}) = \frac{(\sum n_g)/\sigma^2}{[(\sum n_g)(\sum n_g \bar{X}_g^2) - (\sum n_g \bar{X}_g)^2]/\sigma^4}$$

$$= \frac{\sigma^2}{\sum n_g(\bar{X}_g - \bar{X})^2}$$

y

$$(9.24) \quad \text{Var}(\bar{\alpha}) = \frac{(\sum n_g \bar{X}_g^2)/\sigma^2}{[(\sum n_g)(\sum n_g \bar{X}_g^2) - (\sum n_g \bar{X}_g)^2]/\sigma^4}$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum n_g(\bar{X}_g - \bar{X})^2} \right].$$

Estamos particularmente interesados en comparar las variancias de $\bar{\alpha}$ y $\bar{\beta}$ con las de los estimadores minimocuadráticos ordinarios basados en las observaciones sin agrupar. Sabemos que, al agrupar las observaciones y estimar los coeficientes de regresión a partir de las medias de los grupos en lugar de hacerlo en base a las observaciones individuales, estamos perdiendo parte de la información contenida en la muestra, y en concreto la relativa a la variación de las observaciones *dentro* de cada grupo. Es de esperar, pues, que al pasar de la estimación basada en todas las observaciones individuales a la basada en las medias de los grupos se pierda cierta eficiencia. Para ver si ello es cierto, y en qué medida, calcularemos el cociente entre $\text{V.}(\bar{\beta})$ y $\text{Var}(\beta)$, siendo β el estimador minimocuadrático de β basado en las observaciones individuales. De acuerdo con la fórmula (7.23), la variancia de β es

$$\text{Var}(\beta) = \frac{\sigma^2}{\sum_i \sum_g (X_{ig} - \bar{X})^2}.$$

Obsérvese que el denominador del lado derecho de esta igualdad se puede escribir en la forma

$$\begin{aligned} \sum_i \sum_g (X_{ig} - \bar{X})^2 &= \sum_i \sum_g [(X_{ig} - \bar{X}_g) + (\bar{X}_g - \bar{X})]^2 \\ &= \sum_i \sum_g (X_{ig} - \bar{X}_g)^2 + \sum_i \sum_g (\bar{X}_g - \bar{X})^2 \\ &\quad + 2 \sum_i \sum_g (X_{ig} - \bar{X}_g)(\bar{X}_g - \bar{X}) \\ &= \sum_i \sum_g (X_{ig} - \bar{X}_g)^2 + \sum_g n_g (\bar{X}_g - \bar{X})^2. \end{aligned}$$

El cociente entre las dos variancias es, pues,

$$(9.25) \quad \frac{\text{Var}(\bar{\beta})}{\text{Var}(\beta)} = \frac{\sigma^2 / \sum_g n_g (\bar{X}_g - \bar{X})^2}{\sigma^2 / [\sum_i \sum_g (X_{ig} - \bar{X}_g)^2 + \sum_g n_g (\bar{X}_g - \bar{X})^2]}$$

$$= 1 + \frac{\sum_i \sum_g (X_{ig} - \bar{X}_g)^2}{\sum_g n_g (\bar{X}_g - \bar{X})^2}.$$

Este cociente es siempre mayor que la unidad o, en el mejor de los casos, igual a cero. El último término del lado derecho indica la pérdida de eficiencia debida al uso de datos agrupados en lugar de observaciones individuales. Obsérvese que el tamaño del numerador refleja la variación de los valores de X dentro de cada grupo, respecto a la media del grupo, mientras que el tamaño del denominador indica la variación de las medias de cada grupo de X respecto a la media muestral global. Por tanto, si no existe variación en los valores de X dentro de cada grupo no perderemos ninguna eficiencia a consecuencia de la agrupación, y si esta variación es pequeña en comparación con la de las medias de los grupos respecto a la media global, la pérdida de eficiencia será también pequeña. En otras palabras, al pasar de las observaciones individuales a los grupos siempre se producirá alguna pérdida de eficiencia a menos que todas las X en el interior de cada grupo sean iguales. *Esta conclusión se cumple tanto si todos los grupos contienen el mismo número de observaciones como si no.* El hecho de que los grupos sean todos de igual tamaño hará que $\bar{\epsilon}_g$ sea homocedástica, pero no impedirá la pérdida de eficiencia debido a la agrupación.

Al estimar los coeficientes de regresión a partir de datos agrupados podemos utilizar las fórmulas (9.21) y (9.22), ya que su cálculo resulta fácil, pero las expresiones correspondientes a las variancias de dichos estimadores contienen el parámetro desconocido σ^2 . Para obtener un estimador insesgado de σ^2 , calcularemos en primer lugar la esperanza matemática de la suma ponderada de los cuadrados de los residuos, utilizando como ponderaciones el número de observaciones en cada grupo. Es decir,

$$\begin{aligned} E \sum_{g=1}^G n_g (\bar{Y}_g - \bar{\alpha} - \bar{\beta} \bar{X}_g)^2 &= E \sum_g n_g [(\bar{Y}_g - \bar{Y}) - \bar{\beta}(\bar{X}_g - \bar{X})]^2 \\ &= E \sum_g n_g [-(\bar{\beta} - \beta)(\bar{X}_g - \bar{X}) + (\bar{\epsilon}_g - \bar{\epsilon})]^2 \\ &= E(\bar{\beta} - \beta)^2 \sum_g n_g (\bar{X}_g - \bar{X})^2 + E \sum_g n_g (\bar{\epsilon}_g - \bar{\epsilon})^2 \\ &\quad - 2E(\bar{\beta} - \beta) \sum_g n_g (\bar{X}_g - \bar{X})(\bar{\epsilon}_g - \bar{\epsilon}). \end{aligned}$$

Ahora bien,

$$E(\hat{\beta} - \beta)^2 \sum_g n_g (\bar{X}_g - \bar{X})^2 = \left[\frac{\sigma^2}{\sum_g n_g (\bar{X}_g - \bar{X})^2} \right] \sum_g n_g (\bar{X}_g - \bar{X})^2 = \sigma^2,$$

$$\begin{aligned} E \sum_g n_g (\bar{\varepsilon}_g - \bar{\varepsilon})^2 &= E \sum_g n_g \bar{\varepsilon}_g^2 + E \bar{\varepsilon}^2 \sum_g n_g - 2E \bar{\varepsilon} \sum_g n_g \bar{\varepsilon}_g \\ &= \sum_g n_g \left(\frac{\sigma^2}{n_g} \right) + \left(\frac{\sigma^2}{n} \right) \sum_g n_g - 2 \left(\frac{\sigma^2}{n} \right) \sum_g n_g \\ &= G\sigma^2 + \sigma^2 - 2\sigma^2 = (G-1)\sigma^2, \end{aligned}$$

y

$$-2E(\hat{\beta} - \beta) \sum_g n_g (\bar{X}_g - \bar{X})(\bar{\varepsilon}_g - \bar{\varepsilon}) = -2E(\hat{\beta} - \beta)^2 \sum_g n_g (\bar{X}_g - \bar{X})^2 = -2\sigma^2.$$

Por tanto,

$$E \sum_g n_g (\bar{Y}_g - \bar{\alpha} - \hat{\beta} \bar{X}_g)^2 = \sigma^2 + (G-1)\sigma^2 - 2\sigma^2 = (G-2)\sigma^2.$$

El estimador insesgado de σ^2 , al que llamaremos $\hat{s}^2$, es, pues,

$$(9.26) \quad \hat{s}^2 = \frac{1}{G-2} \sum_g n_g (\bar{Y}_g - \bar{\alpha} - \hat{\beta} \bar{X}_g)^2.$$

A efectos de los cálculos, esta expresión puede simplificarse de la forma siguiente:

$$\begin{aligned} (9.26a) \quad \hat{s}^2 &= \frac{1}{G-2} \sum_g n_g [\bar{Y}_g - (\bar{Y} - \hat{\beta} \bar{X}) - \hat{\beta} \bar{X}_g]^2 \\ &= \frac{1}{G-2} \left[\sum_g n_g (\bar{Y}_g - \bar{Y})^2 - \hat{\beta}^2 \sum_g n_g (\bar{X}_g - \bar{X})^2 \right]. \end{aligned}$$

Utilizando $\hat{s}^2$ como estimador de σ^2 en las fórmulas (9.23) y (9.24), obtenemos los estimadores de las variancias de α y β . Dichos estimadores son

$$(9.27) \quad \hat{s}_{\hat{\beta}}^2 = \frac{\hat{s}^2}{\sum_g n_g (\bar{X}_g - \bar{X})^2},$$

$$(9.28) \quad \hat{s}_{\hat{\alpha}}^2 = \hat{s}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_g n_g (\bar{X}_g - \bar{X})^2} \right].$$

Estos estimadores son insesgados, consistentes, y asintóticamente eficientes dentro de la clase de todos los estimadores basados en la misma información.

El último punto de interés en relación con los datos muestrales agrupados es la forma de los cuadros de clasificación simple hace referencia al comportamiento del coeficiente de determinación (R^2) al pasar de las observaciones individuales a las medias de los grupos. Cuando se efectúa la estimación en base a las observaciones individuales, el valor de R^2 se calcula según la fórmula

$$(9.29) \quad R^2 = \frac{\hat{\beta}^2 \sum_i \sum_g (X_{ig} - \bar{X})^2}{\sum_i \sum_g (Y_{ig} - \bar{Y})^2},$$

en la que $\hat{\beta}$ es el estimador minimocuadrático ordinario de β basado en las observaciones individuales. Por otra parte, cuando los coeficientes de regresión se estiman a través de las medias de los grupos, el valor de R^2 se calcula en la forma

$$(9.30) \quad R^2 = 1 - \frac{\sum_g (\bar{Y}_g - \bar{\alpha} - \hat{\beta} \bar{X}_g)^2}{\sum_g (\bar{Y}_g - \bar{Y})^2},$$

en la que

$$\begin{aligned} \sum_g (\bar{Y}_g - \bar{\alpha} - \hat{\beta} \bar{X}_g)^2 &= \sum_g [\bar{Y}_g - (\bar{Y} - \hat{\beta} \bar{X}) - \hat{\beta} \bar{X}_g]^2 \\ &= \sum_g (\bar{Y}_g - \bar{Y})^2 - 2\hat{\beta} \sum_g (\bar{Y}_g - \bar{Y})(\bar{X}_g - \bar{X}) \\ &\quad + \hat{\beta}^2 \sum_g (\bar{X}_g - \bar{X})^2 \end{aligned}$$

y

$$\sum_g (\bar{Y}_g - \bar{Y})^2 = \sum_g \bar{Y}_g^2 - \frac{1}{G} (\sum_g \bar{Y}_g)^2.$$

Se ha podido mostrar que el valor de R^2 calculado según la fórmula (9.30) tiende a ser mayor que el calculado a partir de (9.29).⁶ Es decir, el valor de R^2 tiende a aumentar cuando pasamos de las observaciones individuales a las medias de los grupos. Como la muestra subyacente es la misma tanto si utilizamos las observaciones individuales como las medias de los grupos, el incremento en el valor de R^2 se debe únicamente a la agrupación, y debe interpretarse así. Este incremento refleja simplemente el hecho de que las medias de los grupos tienden a estar menos dispersas que las observaciones individuales en torno a la recta de regresión ajustada.

EMPLO. Una encuesta efectuada en 1959 en una muestra de inmigrantes por el Departamento de demografía de la Universidad Nacional de Australia incluía in-

⁶ J. S. Cramer, "Efficient Grouping, Regression and Correlation in Engel Curve Data," *Journal of the American Statistical Association*, vol. 59, marzo 1964, pp. 223-30.

formación acerca de la renta y los gastos de consumo semanales de las familias inmigrantes. En el cuadro 9-2 se presentan los resultados de la encuesta correspondientes a 181 inmigrantes británicos.

Cuadro 9-2

Grupo de renta	Número de observaciones (n_g)	Renta media* (X_g)	Gastos medios de consumo* (Y_g)
Menos de 18	51	15.5	13.900
18 a menos de 20	22	19.0	15.291
20 a menos de 22	22	21.0	18.195
22 a menos de 24	23	23.0	20.104
24 a menos de 26	13	25.0	20.985
26 a menos de 28	12	27.0	22.742
28 a menos de 30	14	29.0	24.414
30 a menos de 32	9	31.0	24.089
32 a menos de 34	7	33.0	29.286
34 a menos de 36	4	35.0	27.000
36 o más	4	37.0	29.500

Datos facilitados por el Dr. J. Zubrzycki de la Australian National University.

* En libras australianas por semana.

Supongamos que queremos estimar los coeficientes de la función lineal de consumo,

$$Y = \alpha + \beta X + \varepsilon,$$

a partir de dicha información. Los resultados de los cálculos básicos son

$$\bar{X} = \frac{1}{181} \sum_g n_g X_g = 22.390, \quad \bar{Y} = \frac{1}{181} \sum_g n_g Y_g = 19.024, \quad \bar{Y} = \frac{1}{11} \sum_g Y_g = 22.319,$$

$$\sum_g n_g (X_g - \bar{X})(Y_g - \bar{Y}) = \sum_g n_g X_g Y_g - 181(\bar{X}\bar{Y}) = 82,153.2 - 77,096.5 = 5056.7,$$

$$\sum_g n_g (X_g - \bar{X})^2 = \sum_g n_g X_g^2 - 181(\bar{X}^2) = 97,358.8 - 90,735.5 = 6623.3,$$

$$\sum_g n_g (Y_g - \bar{Y})^2 = \sum_g n_g Y_g^2 - 181(\bar{Y}^2) = 69,475.8 - 65,506.2 = 3969.6,$$

$$\sum_g (Y_g - \bar{Y})^2 = 273.7, \quad \sum_g (Y_g - \bar{\alpha} - \beta X_g)^2 = 12.069.$$

por tanto,

$$\beta = \frac{5056.7}{6623.3} = 0.763,$$

$$\bar{\alpha} = 19.024 - 0.763 \times 22.390 = 1.940,$$

$$s^2 = \frac{113.7201}{9} = 12.6356,$$

$$s_a^2 = 12.6356 \left[\frac{1}{181} + \frac{22.390^2}{6623.3} \right] = 1.026188,$$

$$s_b^2 = \frac{12.6356}{6623.3} = 0.001908,$$

$$R^2 = 1 - \frac{12.069}{273.7} = 0.956.$$

La ecuación de regresión estimada es, pues,

$$Y_g = 1.940 + 0.763 X_g + e_g, \quad R^2 = 0.956.$$

(1.012) (0.044)

Agrupación de doble entrada

Consideremos ahora el caso en que la información muestral se condensa en la forma de los cuadros de «doble entrada». En este caso, las observaciones se agrupan según los valores de la variable explicativa y de la variable dependiente. Supongamos que los valores de X se dividen en G grupos, y los valores de Y en H grupos. Cada uno de los G grupos está representado por la respectiva media de X en el grupo, y corresponde a una fila, mientras que cada uno de los H grupos está representado por la media respectiva de Y en el grupo y corresponde a una columna. El cuerpo central del cuadro está compuesto de $G \times H$ casillas, cada una de las cuales contiene el número de observaciones correspondientes. Las casillas sin ninguna información se denominan «vacías». En cada fila y en cada columna hay por lo menos una casilla no vacía. En lo que sigue utilizaremos la siguiente notación:

- $\bar{X}_g$ media de X en el grupo de la fila g
- $\bar{Y}_{gh}$ media de Y en el grupo de la columna h
- J número total de casillas no vacías ($J \leq G \times H$)
- n_{gh} número de observaciones en la casilla correspondiente a la fila g y columna h
- $n_{g\cdot}$ número de observaciones en la fila g
- $n_{\cdot h}$ número de observaciones en la columna h
- n número total de observaciones en la muestra

Esto implica que

$$\sum_{g=1}^G \sum_{h=1}^H n_{gh} = n,$$

$$\sum_{g=1}^G n_{g\cdot} = n,$$

y

$$\sum_{h=1}^H n_{\cdot h} = n.$$

La información tabulada se presenta en la forma siguiente:

	$Y_{\cdot 1}$	$Y_{\cdot 2}$	$\dots$	$Y_{\cdot H}$	
$\bar{X}_{1\cdot}$	n_{11}	n_{12}	$\dots$	n_{1H}	$n_{1\cdot}$
$\bar{X}_{2\cdot}$	n_{21}	n_{22}	$\dots$	n_{2H}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\bar{X}_{G\cdot}$	n_{G1}	n_{G2}	$\dots$	n_{GH}	$n_{G\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	$\dots$	$n_{\cdot H}$	n

Utilizando el triple subíndice igh para indicar la observación i -ésima de la fila g y la columna h , podemos escribir la ecuación de regresión que describe la relación entre X e Y en la forma

$$(9.31) \quad Y_{igh} = \alpha + \beta X_{igh} + \epsilon_{igh} \quad (i = 1, 2, \dots, n_{gh}; \\ g = 1, 2, \dots, G; h = 1, 2, \dots, H).$$

Como no disponemos de las observaciones individuales, tenemos que estimar los coeficientes de regresión α y β a partir de las medias de cada fila y cada columna. Consideremos esto con mayor detalle. Para cada fila (es decir, para cada media de un grupo de X) podemos tener H valores distintos de la variable dependiente, es decir

$$\bar{Y}_{g1} = \alpha + \beta \bar{X}_{g\cdot} + \bar{\epsilon}_{g1}$$

$$\bar{Y}_{g2} = \alpha + \beta \bar{X}_{g\cdot} + \bar{\epsilon}_{g2}$$

$\vdots$

$$\bar{Y}_{gH} = \alpha + \beta \bar{X}_{g\cdot} + \bar{\epsilon}_{gH},$$

$$Y_{gh} = \alpha + \beta X_{gh} + \epsilon_{gh} \quad (g = 1, 2, \dots, G; h = 1, 2, \dots, H),$$

$$\bar{\epsilon}_{gh} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} \epsilon_{igh}.$$

Se puede ver fácilmente que $\bar{\epsilon}_{gh}$ satisface todos los supuestos básicos *excepto* el de homoscedasticidad. En realidad, tenemos

$$\text{Var}(\bar{\epsilon}_{gh}) = \frac{\sigma^2}{n_{gh}},$$

lo que significa que, en general, la variancia de $\bar{\epsilon}_{gh}$ varía de una casilla a otra.

Debido a la heteroscedasticidad de $\bar{\epsilon}_{gh}$, las fórmulas adecuadas para la estimación de los coeficientes de regresión y de sus variancias son las (8.1) a (8.4). Para adaptar estas fórmulas al problema actual, debemos sustituir el subíndice i por g , en el caso de X , y por h en el caso de Y , las X e Y por $\bar{X}$ e $\bar{Y}$, y el término w por (n_{gh}/σ^2) , tomando los sumatorios respecto de g y de h . Los estimadores de α y β resultantes, a los que llamaremos $\tilde{\alpha}$ y $\tilde{\beta}$, son

$$(9.33) \quad \tilde{\beta} = \frac{[(\sum_g \sum_h n_{gh})(\sum_g \sum_h n_{gh} \bar{X}_g \bar{Y}_{gh}) - (\sum_g \sum_h n_{gh} \bar{X}_g)(\sum_g \sum_h n_{gh} \bar{Y}_{gh})]/\sigma^4}{[(\sum_g \sum_h n_{gh})(\sum_g \sum_h n_{gh} \bar{X}_g^2) - (\sum_g \sum_h n_{gh} \bar{X}_g)^2]/\sigma^4}$$

y

$$(9.34) \quad \tilde{\alpha} = \frac{(\sum_g \sum_h n_{gh} \bar{Y}_{gh})/\sigma^2}{(\sum_g \sum_h n_{gh})/\sigma^2} - \tilde{\beta} \frac{(\sum_g \sum_h n_{gh} \bar{X}_g)/\sigma^2}{(\sum_g \sum_h n_{gh})/\sigma^2}.$$

En estas expresiones, tenemos

$$\begin{aligned} \sum_g \sum_h n_{gh} \bar{X}_g \bar{Y}_{gh} &= \sum_g (n_{g1} \bar{X}_g \bar{Y}_{g1} + n_{g2} \bar{X}_g \bar{Y}_{g2} + \dots + n_{gH} \bar{X}_g \bar{Y}_{gH}) \\ &= \sum_g n_g \bar{X}_g \left[\frac{n_{g1} \bar{Y}_{g1} + n_{g2} \bar{Y}_{g2} + \dots + n_{gH} \bar{Y}_{gH}}{n_g} \right] \\ &= \sum_g n_g \bar{X}_g \bar{Y}_{g\cdot}, \end{aligned}$$

Esto implica que

$$\sum_{g=1}^G \sum_{h=1}^H n_{gh} = n,$$

$$\sum_{g=1}^G n_g = n,$$

y

$$\sum_{h=1}^H n_h = n.$$

La información tabulada se presenta en la forma siguiente:

	$Y_{\cdot 1}$	$Y_{\cdot 2}$	...	$Y_{\cdot H}$	
$X_{1\cdot}$	n_{11}	n_{12}	...	n_{1H}	$n_{1\cdot}$
$X_{2\cdot}$	n_{21}	n_{22}	...	n_{2H}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$X_{G\cdot}$	n_{G1}	n_{G2}	...	n_{GH}	$n_{G\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot H}$	n

Utilizando el triple subíndice igh para indicar la observación i -ésima de la fila g y la columna h , podemos escribir la ecuación de regresión que describe la relación entre X e Y en la forma

$$(9.31) \quad Y_{igh} = \alpha + \beta X_{igh} + e_{igh} \quad (i = 1, 2, \dots, n_{gh}; \\ g = 1, 2, \dots, G; h = 1, 2, \dots, H).$$

Como no disponemos de las observaciones individuales, tenemos que estimar los coeficientes de regresión α y β a partir de las medias de cada fila y cada columna. Consideremos esto con mayor detalle. Para cada fila (es decir, para cada media de un grupo de X) podemos tener H valores distintos de la variable dependiente, es decir

$$\bar{Y}_{g1} = \alpha + \beta \bar{X}_g + \bar{e}_{g1}$$

$$\bar{Y}_{g2} = \alpha + \beta \bar{X}_g + \bar{e}_{g2}$$

$\vdots$

$$\bar{Y}_{gH} = \alpha + \beta \bar{X}_g + \bar{e}_{gH},$$

$$Y_{gh} = \alpha + \beta X_{gh} + e_{gh} \quad (g = 1, 2, \dots, G; h = 1, 2, \dots, H),$$

$$\bar{e}_{gh} = \frac{1}{n_{gh}} \sum_{i=1}^{n_{gh}} e_{igh}.$$

Se puede ver fácilmente que $\bar{e}_{gh}$ satisface todos los supuestos básicos *excepto* el de homoscedasticidad. En realidad, tenemos

$$\text{Var}(\bar{e}_{gh}) = \frac{\sigma^2}{n_{gh}},$$

lo que significa que, en general, la variancia de $\bar{e}_{gh}$ varía de una casilla a otra.

Debido a la heteroscedasticidad de $\bar{e}_{gh}$, las fórmulas adecuadas para la estimación de los coeficientes de regresión y de sus variancias son las (8.1) a (8.4). Para adaptar estas fórmulas al problema actual, debemos sustituir el subíndice i por g , en el caso de X , y por h en el caso de Y , las X e Y por $\bar{X}$ e $\bar{Y}$, y el término w por (n_{gh}/σ^2) , tomando los sumatorios respecto de g y de h . Los estimadores de α y β resultantes, a los que llamaremos $\tilde{\alpha}$ y $\tilde{\beta}$, son

$$(9.33) \quad \tilde{\beta} = \frac{[(\sum_g \sum_h n_{gh})(\sum_g \sum_h n_{gh} \bar{X}_g \bar{Y}_{gh}) - (\sum_g \sum_h n_{gh} \bar{X}_g)(\sum_g \sum_h n_{gh} \bar{Y}_{gh})]/\sigma^4}{[(\sum_g \sum_h n_{gh})(\sum_g \sum_h n_{gh} \bar{X}_g^2) - (\sum_g \sum_h n_{gh} \bar{X}_g)^2]/\sigma^4}$$

y

$$(9.34) \quad \tilde{\alpha} = \frac{(\sum_g \sum_h n_{gh} \bar{Y}_{gh})/\sigma^2}{(\sum_g \sum_h n_{gh})/\sigma^2} - \tilde{\beta} \frac{(\sum_g \sum_h n_{gh} \bar{X}_g)/\sigma^2}{(\sum_g \sum_h n_{gh})/\sigma^2}.$$

En estas expresiones, tenemos

$$\begin{aligned} \sum_g \sum_h n_{gh} \bar{X}_g \bar{Y}_{gh} &= \sum_g (n_{g1} \bar{X}_g \bar{Y}_{g1} + n_{g2} \bar{X}_g \bar{Y}_{g2} + \dots + n_{gH} \bar{X}_g \bar{Y}_{gH}) \\ &= \sum_g n_g \bar{X}_g \left[\frac{n_{g1} \bar{Y}_{g1} + n_{g2} \bar{Y}_{g2} + \dots + n_{gH} \bar{Y}_{gH}}{n_g} \right] \\ &= \sum_g n_g \bar{X}_g \bar{Y}_g, \end{aligned}$$

y es distinto de $\bar{s}^2$, el estimador de σ^2 y se obtuvo a partir del cuadro de clasificación simple. En relación a los cálculos, obsérvese que

$$\begin{aligned}\sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2 - \beta^2 \sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})^2 \\ &= \sum_g n_{g\cdot} (\bar{Y}_{g\cdot} - \bar{Y})^2 - \beta^2 \sum_g n_{g\cdot} (\bar{X}_g - \bar{X})^2.\end{aligned}$$

Aunque no vamos a presentar una prueba formal de ello, resulta evidente que $\bar{s}^2$ es un estimador más eficiente de σ^2 que $\bar{s}^2$. Esto es debido a que $\bar{s}^2$ utiliza mayor información acerca de la variación de la variable dependiente. Al utilizar $\bar{s}^2$ como estimador de σ^2 en las ecuaciones (9.23) y (9.24) se obtienen unos nuevos estimadores de $\text{Var}(\bar{\alpha})$ y $\text{Var}(\bar{\beta})$ que vendrán dados por las expresiones

$$(9.38) \quad \bar{s}_{\beta}^2 = \frac{\bar{s}^2}{\sum_g n_{g\cdot} (\bar{X}_g - \bar{X})^2}$$

y

$$(9.39) \quad \bar{s}_{\alpha}^2 = \bar{s}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_g n_{g\cdot} (\bar{X}_g - \bar{X})^2} \right].$$

Nuevamente, estos estimadores son insesgados, consistentes, y asintóticamente eficientes respecto a todos los estimadores que utilizan la misma información.

Por último, podemos estar interesados en la obtención del valor del coeficiente de determinación. La fórmula de R^2 calculada a partir de un cuadro de doble entrada es, por analogía con la ecuación (9.30),

$$(9.40) \quad R^2 = 1 - \frac{\sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2}{\sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2},$$

en la que

$$\begin{aligned}\sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 + \beta^2 \sum_g \sum_h (\bar{X}_g - \bar{X})^2 \\ &\quad - 2\beta \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})(\bar{X}_g - \bar{X})\end{aligned}$$

y

$$\sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 = \sum_g \sum_h \bar{Y}_{gh}^2 - \frac{1}{J} \sum_g \sum_h \bar{Y}_{gh}.$$

Todos estos sumatorios se efectúan únicamente respecto a las casillas no vacías. Como la clasificación de doble entrada permite una mayor dispersión de las informaciones que los cuadros de clasificación simple, cabe esperar que el valor

calculado a partir de la fórmula (9.40) sea menor que el calculado por

cuando X es una variable estocástica (independiente del término de perturbación de la regresión), podemos estar interesados en el cálculo del valor del coeficiente de correlación entre X e Y . En el caso de los cuadros de clasificación de doble entrada, este valor se calcula normalmente teniendo en cuenta las frecuencias de las casillas individuales contrariamente a lo que ocurre cuando se calcula R^2 en base a la fórmula (9.40). Representando por r_{XY} el coeficiente de correlación X e Y , su valor puede obtenerse a partir de

$$(9.41) \quad r_{XY}^2 = \frac{[\sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})(\bar{Y}_{gh} - \bar{Y})]^2}{[\sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})^2][\sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2]}$$

Naturalmente, las fórmulas de los coeficientes de regresión estimados y de sus errores standard permanecen inalteradas.

Ejem p l o. Los resultados acerca de la renta y el gasto, correspondientes a los 181 emigrantes británicos en Australia de la encuesta utilizada en el ejemplo anterior, se hallan también disponibles en la forma de un cuadro de doble entrada, y se presentan en el cuadro 9-3. En este caso, X = renta, e Y = gastos de consumo. Las estimaciones de los coeficientes de la función lineal de consumo serán las mismas que se han calculado a partir del cuadro 9-2. Sin embargo, tenemos que efectuar nuevos cálculos de las estimaciones de los errores standard y de R^2 . Los resultados básicos son los siguientes

$$\begin{aligned}J &= 45, \\ \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2 &= 6150.8, \\ \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= 2292.5, \\ \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 &= 1538.0, \\ \sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= 983.0.\end{aligned}$$

Y, por tanto,

$$\bar{s}^2 = \frac{1}{45 - 2} \times 2292.5 = 53.31,$$

$$\bar{s}_{\alpha}^2 = 4.3298,$$

$$\bar{s}_{\beta}^2 = 0.008050,$$

$$R^2 = 1 - \frac{983.0}{1538.0} = 0.361.$$

y es distinto de $\hat{s}^2$, el estimador de σ^2 y se obtuvo a partir del cuadro de clasificación simple. En relación a los cálculos, obsérvese que

$$\begin{aligned}\sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2 - \beta^2 \sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})^2 \\ &= \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2 - \beta^2 \sum_g n_{gh} (\bar{X}_g - \bar{X})^2.\end{aligned}$$

Aunque no vamos a presentar una prueba formal de ello, resulta evidente que $\hat{s}^2$ es un estimador más eficiente de σ^2 que $\hat{s}^2$. Esto es debido a que $\hat{s}^2$ utiliza mayor información acerca de la variación de la variable dependiente. Al utilizar $\hat{s}^2$ como estimador de σ^2 en las ecuaciones (9.23) y (9.24) se obtienen unos nuevos estimadores de $\text{Var}(\hat{\alpha})$ y $\text{Var}(\hat{\beta})$ que vendrán dados por las expresiones

$$(9.38) \quad \hat{s}_{\hat{\beta}}^2 = \frac{\hat{s}^2}{\sum_g n_{gh} (\bar{X}_g - \bar{X})^2}$$

y

$$(9.39) \quad \hat{s}_{\hat{\alpha}}^2 = \hat{s}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_g n_{gh} (\bar{X}_g - \bar{X})^2} \right].$$

Nuevamente, estos estimadores son insesgados, consistentes, y asintóticamente eficientes respecto a todos los estimadores que utilizan la misma información.

Por último, podemos estar interesados en la obtención del valor del coeficiente de determinación. La fórmula de R^2 calculada a partir de un cuadro de doble entrada es, por analogía con la ecuación (9.10),

$$(9.40) \quad R^2 = 1 - \frac{\sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2}{\sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2},$$

en la que

$$\begin{aligned}\sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 + \beta^2 \sum_g \sum_h (\bar{X}_g - \bar{X})^2 \\ &\quad - 2\beta \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})(\bar{X}_g - \bar{X})\end{aligned}$$

y

$$\sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 = \sum_g \sum_h \bar{Y}_{gh}^2 - \frac{1}{J} \sum_g \sum_h \bar{Y}_{gh}.$$

Todos estos sumatorios se efectúan únicamente respecto a las casillas no vacías. Como la clasificación de doble entrada permite una mayor dispersión de las informaciones que los cuadros de clasificación simple, cabe esperar que el valor

calculado a partir de la fórmula (9.40) sea menor que el calculado por el cuadro de clasificación simple.

Si $\bar{Y}_{gh}$ es una variable estocástica (independiente del término de perturbación de la regresión), podemos estar interesados en el cálculo del valor del coeficiente de correlación entre X e Y . En el caso de los cuadros de clasificación de doble entrada, este valor se calcula normalmente teniendo en cuenta las frecuencias de las casillas individuales contrariamente a lo que ocurre cuando se calcula R^2 en base a la fórmula (9.40). Representando por r_{XY} el coeficiente de correlación X e Y , su valor puede obtenerse a partir de

$$(9.41) \quad r_{XY}^2 = \frac{[\sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})(\bar{Y}_{gh} - \bar{Y})]^2}{[\sum_g \sum_h n_{gh} (\bar{X}_g - \bar{X})^2][\sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2]}.$$

Naturalmente, las fórmulas de los coeficientes de regresión estimados y de sus errores standard permanecen inalteradas.

EJEMPLO. Los resultados acerca de la renta y el gasto, correspondientes a los 181 emigrantes británicos en Australia de la encuesta utilizada en el ejemplo anterior, se hallan también disponibles en la forma de un cuadro de doble entrada, y se presentan en el cuadro 9-3. En este caso, X = renta, e Y = gastos de consumo. Las estimaciones de los coeficientes de la función lineal de consumo serán las mismas que se han calculado a partir del cuadro 9-2. Sin embargo, tenemos que efectuar nuevos cálculos de las estimaciones de los errores standard y de R^2 . Los resultados básicos son los siguientes

$$\begin{aligned}J &= 45, \\ \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{Y})^2 &= 6150.8, \\ \sum_g \sum_h n_{gh} (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= 2292.5, \\ \sum_g \sum_h (\bar{Y}_{gh} - \bar{Y})^2 &= 1538.0, \\ \sum_g \sum_h (\bar{Y}_{gh} - \bar{\alpha} - \beta \bar{X}_g)^2 &= 983.0.\end{aligned}$$

Y, por tanto,

$$\hat{s}^2 = \frac{1}{45 - 2} \times 2292.5 = 53.31,$$

$$\hat{s}_{\hat{\alpha}}^2 = 4.3298,$$

$$\hat{s}_{\hat{\beta}}^2 = 0.008050,$$

$$R^2 = 1 - \frac{983.0}{1538.0} = 0.361.$$

en las que $\bar{X}_c$ e $\bar{Y}_c$ son las medias muestrales de X e Y calculadas a partir de los pares completos, y Σ_c representa la suma respecto a todos estos pares. Los estimadores $\hat{\alpha}_c$ y $\hat{\beta}_c$ son insesgados y eficientes respecto a todos los estimadores de α y β que utilizan la misma información.

Variable explicativa no estocástica

En el caso en que X es no estocástica, sus valores están bajo el control del investigador o del «experimentador» originario. Naturalmente, esto debe interpretarse en sentido amplio —por ejemplo, considerando que el Gobierno efectúa un «experimento» siempre que realiza algún gasto—. La implicación de esto es que los pares de valores en los que no aparece el valor de Y no proporcionan información alguna acerca del resultado del «experimento», y no deben contarse en absoluto como formando parte de la muestra. Por tanto, en este contexto, el único caso interesante es aquel en que faltan algunos de los valores de X pero se conocen todos los valores de Y . Los pares incompletos nos proporcionan información acerca de Y , o sea acerca del resultado del «experimento», pero no acerca de la variable condicionante X . En primer lugar calcularemos la pérdida de eficiencia que resulta del uso de los pares completos solamente en vez de utilizar todos los pares si todos estuviesen completos. Después trataremos de utilizar los pares incompletos con el fin de reducir dicha pérdida de eficiencia. Durante este proceso, y a lo largo de toda esta sección, utilizaremos la notación siguiente, además de los símbolos ya utilizados en las fórmulas (9.42) y (9.43):

Σ_x	sumatorio respecto a todos los pares en los que se observa X
Σ_y	sumatorio respecto a todos los pares en los que se observa Y
Σ_{0x}	sumatorio respecto a todos los pares en los que no se observa X
Σ_{0y}	sumatorio respecto a todos los pares en los que no se observa Y
n_c	número de pares completos
n_x	número de pares en los que se observa X
n_y	número de pares en los que se observa Y
m_x	número de pares en los que no se observa X
m_y	número de pares en los que no se observa Y
$\bar{X}_x = \frac{1}{n_x} \Sigma_x X_i$	etc.

Obsérvese que

$$n_c + m_x + m_y = n,$$

$$n_c + m_y = n_x,$$

$$n_c + m_x = n_y.$$

En este contexto, en el que disponemos de todos los valores de Y , tenemos

$$n_x = n_c,$$

$$n_y = n,$$

$$m_y = 0,$$

$$\bar{X}_x = \bar{X}_c.$$

Si utilizamos únicamente los pares de informaciones completos, la variancia del estimador minimocuadrático de β es

$$(9.44) \quad \text{Var}(\hat{\beta}_c) = \frac{\sigma^2}{\sum_x (X_i - \bar{X}_x)^2},$$

Si todos los pares estuviesen completos, la variancia del estimador minimocuadrático de β , al que en este caso llamaremos $\hat{\beta}_y$, sería

$$(9.45) \quad \text{Var}(\hat{\beta}_y) = \frac{\sigma^2}{\sum_y (X_i - \bar{X}_y)^2}.$$

La pérdida de eficiencia debida al hecho de que algunos de los pares de observaciones no indican el valor de X (y sólo utilizamos los pares completos) puede medirse por la razón entre $\text{Var}(\hat{\beta}_x)$ y $\text{Var}(\hat{\beta}_y)$, o sea,

$$(9.46) \quad \frac{\text{Var}(\hat{\beta}_c)}{\text{Var}(\hat{\beta}_y)} = \frac{\sum_y (X_i - \bar{X}_y)^2}{\sum_x (X_i - \bar{X}_x)^2}.$$

En esta expresión, tenemos

$$\begin{aligned} \sum_y (X_i - \bar{X}_y)^2 &= \sum_x (X_i - \bar{X}_y)^2 + \sum_{0x} (X_i - \bar{X}_y)^2 \\ &= \sum_x [(X_i - \bar{X}_x) + (\bar{X}_x - \bar{X}_y)]^2 \\ &\quad + \sum_{0x} [(X_i - \bar{X}_{0x}) + (\bar{X}_{0x} - \bar{X}_y)]^2 \\ &= \sum_x (X_i - \bar{X}_x)^2 + n_x(\bar{X}_x - \bar{X}_y)^2 + \sum_{0x} (X_i - \bar{X}_{0x})^2 \\ &\quad + m_x(\bar{X}_{0x} - \bar{X}_y)^2. \end{aligned}$$

Sabiendo que

$$\bar{X}_y = \frac{1}{n} (n_x \bar{X}_x + m_x \bar{X}_{0x}),$$

en las que $\bar{X}_c$ e $\bar{Y}_c$ son las medias muestrales de X e Y calculadas a partir de los pares completos, y Σ_c representa la suma respecto a todos estos pares. Los estimadores $\hat{\alpha}_c$ y $\hat{\beta}_c$ son insesgados y eficientes respecto a todos los estimadores de α y β que utilizan la misma información.

Variable explicativa no estocástica

En el caso en que X es no estocástica, sus valores están bajo el control del investigador o del «experimentador» originario. Naturalmente, esto debe interpretarse en sentido amplio —por ejemplo, considerando que el Gobierno efectúa un «experimento» siempre que realiza algún gasto—. La implicación de esto es que los pares de valores en los que no aparece el valor de Y no proporcionan información alguna acerca del resultado del «experimento», y no deben considerarse como formando parte de la muestra. Por tanto, en este contexto, el único caso interesante es aquel en que faltan algunos de los valores de X pero se conocen todos los valores de Y . Los pares incompletos nos proporcionan información acerca de Y , o sea acerca del resultado del «experimento», pero no acerca de la variable condicionante X . En primer lugar calcularemos la pérdida de eficiencia que resulta del uso de los pares completos solamente en vez de utilizar todos los pares si todos estuviesen completos. Después trataremos de utilizar los pares incompletos con el fin de reducir dicha pérdida de eficiencia. Durante este proceso, y a lo largo de toda esta sección, utilizaremos la notación siguiente, además de los símbolos ya utilizados en las fórmulas (9.42) y (9.43):

$\sum_x$	sumatorio respecto a todos los pares en los que se observa X
$\sum_y$	sumatorio respecto a todos los pares en los que se observa Y
$\sum_{0x}$	sumatorio respecto a todos los pares en los que no se observa X
$\sum_{0y}$	sumatorio respecto a todos los pares en los que no se observa Y
n_c	número de pares completos
n_x	número de pares en los que se observa X
n_y	número de pares en los que se observa Y
m_x	número de pares en los que no se observa X
m_y	número de pares en los que no se observa Y
$\bar{X}_x = \frac{1}{n_x} \sum_x X_i$, etc.	

Obsérvese que

$$n_c + m_x + m_y = n,$$

$$n_c + m_y = n_x,$$

$$n_c + m_x = n_y.$$

Si, en el que disponemos de todos los valores de Y , tenemos

$$n_x = n_c,$$

$$n_y = n,$$

$$m_y = 0,$$

$$\bar{X}_x = \bar{X}_c.$$

Si utilizamos únicamente los pares de informaciones completos, la variancia del estimador minimocuadrático de β es

$$(9.44) \quad \text{Var}(\hat{\beta}_c) = \frac{\sigma^2}{\sum_x (X_i - \bar{X}_x)^2}.$$

Si todos los pares estuviesen completos, la variancia del estimador minimocuadrático de β , al que en este caso llamaremos $\hat{\beta}_y$, sería

$$(9.45) \quad \text{Var}(\hat{\beta}_y) = \frac{\sigma^2}{\sum_y (X_i - \bar{X}_y)^2}.$$

La pérdida de eficiencia debida al hecho de que algunos de los pares de observaciones no indican el valor de X (y sólo utilizamos los pares completos) puede medirse por la razón entre $\text{Var}(\hat{\beta}_x)$ y $\text{Var}(\hat{\beta}_y)$, o sea,

$$(9.46) \quad \frac{\text{Var}(\hat{\beta}_x)}{\text{Var}(\hat{\beta}_y)} = \frac{\sum_y (X_i - \bar{X}_y)^2}{\sum_x (X_i - \bar{X}_x)^2}.$$

En esta expresión, tenemos

$$\begin{aligned} \sum_y (X_i - \bar{X}_y)^2 &= \sum_x (X_i - \bar{X}_y)^2 + \sum_{0x} (X_i - \bar{X}_y)^2 \\ &= \sum_x [(X_i - \bar{X}_x) + (\bar{X}_x - \bar{X}_y)]^2 \\ &\quad + \sum_{0x} [(X_i - \bar{X}_{0x}) + (\bar{X}_{0x} - \bar{X}_y)]^2 \\ &= \sum_x (X_i - \bar{X}_x)^2 + n_x (\bar{X}_x - \bar{X}_y)^2 + \sum_{0x} (X_i - \bar{X}_{0x})^2 \\ &\quad + m_x (\bar{X}_{0x} - \bar{X}_y)^2. \end{aligned}$$

Sabiendo que

$$\bar{X}_y = \frac{1}{n} (n_x \bar{X}_x + m_x \bar{X}_{0x}),$$

podemos escribir

$$n_x(\bar{X}_x - \bar{X}_y)^2 = n_x \left[\bar{X}_x - \frac{1}{n}(n_x \bar{X}_x + m_x \bar{X}_{0x}) \right]^2 = \frac{n_x m_x^2}{n^2} (\bar{X}_x - \bar{X}_{0x})^2,$$

y

$$m_x(\bar{X}_{0x} - \bar{X}_y)^2 = m_x \left[\bar{X}_{0x} - \frac{1}{n}(n_x \bar{X}_x + m_x \bar{X}_{0x}) \right]^2 = \frac{n_x^2 m_x}{n^2} (\bar{X}_x - \bar{X}_{0x})^2.$$

Por tanto,

$$(9.46a) \quad \frac{\text{Var}(\beta_c)}{\text{Var}(\beta_y)} = 1 + \frac{\sum_{0x} (X_i - \bar{X}_{0x})^2 + (n_x m_x / n) (\bar{X}_x - \bar{X}_{0x})^2}{\sum_x (X_i - \bar{X}_x)^2}.$$

Este resultado indica que la pérdida de eficiencia será pequeña si los valores de X que faltan tienen una dispersión pequeña y, al mismo tiempo, la media de dichos valores no difiere mucho de la media de los valores de X que conocemos. En las muestras finitas no existirá pérdida de eficiencia alguna si, y solamente si, todos los valores de X que no conocemos son iguales a la media de los valores conocidos de X . Naturalmente, como no conocemos los valores de X que faltan, no podemos calcular la razón (9.46a), pero la podemos estimar a partir de la información muestral disponible, como se muestra más adelante. Tratemos ahora de utilizar la información contenida en los pares de observaciones en los que falta el valor de X . Podemos considerar estos valores que nos faltan como unos parámetros desconocidos que se pueden estimar al mismo tiempo que los coeficientes de regresión y σ^2 . Representaremos por ξ_i estos valores de X que faltan, y según la notación que hemos indicado, su número será m . La función de verosimilitud para $(Y_1, Y_2, \dots, Y_n)$ será

$$L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_x (Y_i - \alpha - \beta X_i)^2 - \frac{1}{2\sigma^2} \sum_{0x} (Y_i - \alpha - \beta \xi_i)^2.$$

Derivando L respecto de α , β , σ^2 , y cada una de las ξ_i , e igualando a cero estas derivadas y despejando los valores de los parámetros desconocidos, obtendremos los respectivos estimadores maximoverosímiles. Es simplemente cuestión de álgebra el mostrar que los estimadores maximoverosímiles de α y β son exactamente iguales a los estimadores minimocuadráticos (9.42) y (9.43), que el estimador de σ^2 es igual al que se basa únicamente en los pares completos, y que los estimadores de ξ_i son

$$(9.47) \quad \hat{\xi}_i = \frac{Y_i - \hat{\alpha}_c}{\hat{\beta}_c}.$$

Esto significa que el método de la estimación maximoverosímil aplicado a todas las observaciones en las que se conoce Y proporciona estimaciones de

los valores que faltan de X , pero deja a las estimaciones de α , β y σ^2 exactamente como si se estimaban en base a los pares completos solamente. Esto es algo descorazonador, pero, sin embargo, hemos adelantado un poco, por lo menos podemos utilizar las estimaciones de los valores de X que faltan para obtener alguna idea acerca de la pérdida de eficiencia debida a la presencia de pares de observaciones incompletos. Esto se puede hacer sustituyendo los valores desconocidos de X en la ecuación (9.46a) por sus estimaciones $\hat{\xi}_i$. El resultado es

$$(9.48) \quad \text{Est.} \left[\frac{\text{Var}(\beta_c)}{\text{Var}(\beta_y)} \right] = 1 + \frac{\sum_{0x} (\hat{\xi}_i - \bar{\hat{\xi}})^2 + (n_x m_x / n) (\bar{X}_x - \bar{\hat{\xi}})^2}{\sum_x (X_i - \bar{X}_x)^2}.$$

El estimador $\hat{\xi}_i$ tiene las propiedades asintóticas deseables que poseen los demás estimadores maximoverosímiles, a condición de que el número de valores de X que faltan no aumente al aumentar el tamaño de la muestra.

EJEMPLO. En el ejemplo de la sección 7.3, nos ocupamos de la estimación de los coeficientes de una relación lineal entre el precio (X) y la cantidad de naranjas vendidas (Y) en un supermercado durante doce días consecutivos. Las observaciones fueron:

X :	100	90	80	70	70	70	70	65	60	60	55	50
Y :	55	70	90	100	90	105	80	110	125	115	130	130

Los resultados de los cálculos apropiados fueron los siguientes:

$$\begin{aligned} \bar{X} &= 70, \\ \sum (X_i - \bar{X})^2 &= 2250, \\ \hat{\alpha} &= 210.460, \\ \hat{\beta} &= -1.578. \end{aligned}$$

Supongamos ahora que, además de los 12 pares de observaciones, también tenemos la información de que la cantidad vendida en el día número trece fue de 37 libras, pero no tenemos noticias del precio. Es decir, $Y_{13} = 37$, y se ha prescindido de esta observación. Queremos saber qué eficiencia hubiésemos ganado en la estimación de β si hubiéramos conocido X_{13} . En primer lugar, utilizamos la ecuación (9.47) para estimar X_{13} , de modo que

$$\hat{\xi}_{13} = \frac{Y_{13} - \hat{\alpha}}{\hat{\beta}} = \frac{37 - 210.460}{-1.578} = 110.$$

En tal caso, la razón estimada entre $\text{Var}(\hat{\beta}_1)$ y $\text{Var}(\hat{\beta}_1)$ es

$$1 + \frac{0 + [(12 \times 1)/13](70 - 110)^2}{2250} = 1.6564,$$

lo que significa que la pérdida de eficiencia se estima en un 65,64 %.

Una forma alternativa de utilizar la información contenida en los pares de observaciones incompletos consiste en rellenar los espacios vacíos utilizando alguna aproximación de los valores de X que faltan. Este procedimiento es, probablemente, muy corriente en la práctica. Las aproximaciones pueden obtenerse, por ejemplo, por interpolación a partir de los valores observados de X , o con referencia a alguna variable Z que esté correlacionada con X . Sin embargo, si sustituimos los valores de X que faltan por alguna aproximación, introducimos errores de medición en los valores de la variable explicativa y , en consecuencia, obtenemos estimaciones inconsistentes de los coeficientes de regresión, tal como se explicó detalladamente en la sección 9.1. Naturalmente, el tamaño de esta inconsistencia dependerá de la medida de los errores efectuados en la aproximación. En realidad, lo que estamos haciendo en este caso es perder consistencia, esperando reducir con ello la variancia del estimador. Si estamos razonablemente seguros de que los errores de aproximación son pequeños mientras que el aumento de eficiencia es potencialmente grande, esto puede ser un procedimiento racional. De lo contrario, este truco puede producir pérdidas.

Variable explicativa estocástica

Pasemos ahora al caso en que X es una variable estocástica que se distribuye independientemente del término de perturbación. Las fórmulas de los estimadores minimocuadráticos de los coeficientes de regresión basadas en los pares de observaciones completos permanecen inalteradas y lo mismo ocurre con las fórmulas de sus variancias —salvo por el hecho de que estas últimas deben considerarse condicionadas al conjunto específico de valores de X de que se dispone—. Cada par de valores observados de X e Y procede ahora de una distribución de probabilidad bivalente, y el problema consiste en estimar los coeficientes de regresión cuando algunos pares de observaciones están incompletos. Aparte de prescindir de los pares incompletos, también podemos tratar de rellenar los espacios vacíos aplicando a continuación la estimación por mínimos cuadrados. Una forma de rellenar los espacios vacíos consiste en preguntarnos qué valores de X , o de Y , esperaríamos observar antes de efectuar la observación. Generalmente, estos valores serían las esperanzas matemáticas de X o Y , es decir, sus medias. Como las medias son desconocidas podemos utilizar como estimadores las medias muestrales disponibles, o sea que podemos completar las observaciones que faltan en los pares incompletos utilizando las medias muestrales de las variables respectivas. Los estimadores minimocuadráticos

de α y β obtenidos a partir de la muestra completada de esta forma reciben el nombre de *estimadores de regresión de orden cero*,⁷ y se definen de la forma siguiente:

$$(9.49) \quad \beta_0 = \frac{\sum_c (X_i - \bar{X}_x)(Y_i - \bar{Y}_y) + \sum_{0x} (\bar{X}_x - \bar{X}_x)(Y_i - \bar{Y}_y)}{\sum_c (X_i - \bar{X}_x)^2 + \sum_{0x} (\bar{X}_x - \bar{X}_x)^2 + \sum_{0y} (X_i - \bar{X}_x)^2} \\ + \frac{\sum_{0y} (X_i - \bar{X}_x)(\bar{Y}_y - \bar{Y}_y)}{\sum_c (X_i - \bar{X}_x)^2 + \sum_{0x} (\bar{X}_x - \bar{X}_x)^2 + \sum_{0y} (X_i - \bar{X}_x)^2} \\ = \frac{\sum_c (X_i - \bar{X}_x)(Y_i - \bar{Y}_y)}{\sum_x (X_i - \bar{X}_x)^2}$$

y

$$(9.50) \quad \hat{\alpha}_0 = \bar{Y}_y - \hat{\beta}_0 \bar{X}_x.$$

Para descubrir si estos estimadores son insesgados, sustituiremos

$$Y_i - \bar{Y}_y = \beta(X_i - \bar{X}_x) + (\epsilon_i - \bar{\epsilon}_y)$$

en la ecuación (9.49) para obtener

$$\hat{\beta}_0 = \frac{\sum_c (X_i - \bar{X}_x)[\beta(X_i - \bar{X}_x) + (\epsilon_i - \bar{\epsilon}_y)]}{\sum_x (X_i - \bar{X}_x)^2}.$$

La esperanza matemática de $\hat{\beta}_0$ condicionada a los valores observados de X es

$$E(\hat{\beta}_0) = \frac{\beta \sum_c (X_i - \bar{X}_c + \bar{X}_c - \bar{X}_x)[X_i - \bar{X}_c + \bar{X}_c - E(\bar{X}_y)]}{\sum_x (X_i - \bar{X}_x)^2} \\ = \frac{\beta \sum_c (X_i - \bar{X}_c)^2 + n_c(\bar{X}_c - \bar{X}_x)[\bar{X}_x - E(\bar{X}_y)]}{\sum_x (X_i - \bar{X}_x)^2}.$$

⁷ Ver A. A. Aftab y R. M. Hershoff, "Missing Observations in Multivariate Statistics," *Journal of the American Statistical Association*, vol. 62, marzo 1967, pp. 10-29.

Pero

$$E(\bar{X}_y) = E\left[\frac{1}{n_y} \left(\sum_c X_i + \sum_{0x} X_i\right)\right] = \frac{1}{n_y} (n_c \bar{X}_c + m_x \mu_x),$$

en la que se utiliza μ_x , que es la media de la población de X , para sustituir a $E(\bar{X}_{0x})$, ya que $\bar{X}_{0x}$ no se ha observado. Por tanto,

$$E(\beta_0) = \frac{\beta \left[\sum_c (\bar{X}_i - \bar{X}_c)^2 + (n_c m_x / n_y) (\bar{X}_c - \bar{X}_x) (\bar{X}_c - \mu_x) \right]}{\sum_x (X_i - \bar{X}_x)^2} \neq \beta.$$

La conclusión es, pues, que, en general, el estimador de regresión de orden cero de β es *sesgado*, y lo mismo ocurre con respecto al estimador de regresión de orden cero de α .

Antes de finalizar con el método de regresión de orden cero, consideremos algunos casos especiales. En primer lugar, supongamos que se dispone de todos los valores de X y que solamente faltan algunos valores de Y . En este caso,

$$\begin{aligned} n_x &= n, \\ n_y &= n_c, \\ m_x &= 0, \\ \bar{X}_{0x} &= 0. \end{aligned}$$

Así pues,

$$E(\beta_0) = \beta \frac{\sum_c (X_i - \bar{X}_c)^2}{\sum_x (X_i - \bar{X}_x)^2},$$

de modo que β_0 sigue siendo sesgado a menos que $\bar{X}_c = \bar{X}_x$. Alternativamente, supongamos que disponemos de todos los valores de Y , pero que faltan algunos de los valores de X . En este caso,

$$\begin{aligned} n_x &= n_c, \\ n_y &= n, \\ m_y &= 0, \\ \bar{X}_x &= \bar{X}_c, \end{aligned}$$

y

$$E(\beta_0) = \frac{\beta \sum_c (X_i - \bar{X}_c)^2}{\sum_c (X_i - \bar{X}_c)^2} = \beta,$$

de modo que, en este caso, el estimador de β de la regresión de orden cero es insesgado. Sin embargo, la variancia de β_0 condicionada a los valores observados de X es

$$\text{Var}(\beta_0) = E(\beta_0 - \beta)^2 = E \left[\frac{\sum_c (X_i - \bar{X}_c)(e_i - \bar{e}_y)}{\sum_x (X_i - \bar{X}_x)^2} \right]^2 = \frac{\sigma^2}{\sum_x (X_i - \bar{X}_x)^2},$$

que es igual a la expresión de $\text{Var}(\hat{\beta}_c)$ indicada en la fórmula (9.44).⁵ Esto significa que en términos de eficiencia no ganamos nada usando β_0 en lugar de $\hat{\beta}_c$.

El método de estimación de la regresión de orden cero se basa en la idea de sustituir cada uno de los valores que faltan de X por $\bar{X}_x$, y los valores que faltan de Y por $\bar{Y}_y$. Como idea alternativa, se pueden sustituir los valores desconocidos de X por el parámetro ξ , y los de Y por el parámetro η . Como todos los valores de X se sustituyen por un solo parámetro, y todos los de Y por otro, este procedimiento incorpora solamente dos parámetros desconocidos adicionales independientemente del tamaño de la muestra y del número de valores que faltan. A continuación, se pueden estimar simultáneamente los coeficientes de regresión α y β , y los parámetros ξ y η . Esto puede hacerse minimizando

$$\sum_c (Y_i - \alpha - \beta X_i)^2 + \sum_{0x} (Y_i - \alpha - \beta \xi)^2 + \sum_{0y} (\eta - \alpha - \beta X_i)^2$$

respecto a α , β , ξ y η . Los estimadores resultantes, llamados *estimadores modificados de regresión de orden cero*,⁶ son

$$(9.51) \quad \hat{\beta}_m = \frac{\sum_c (X_i - \bar{X}_c)(Y_i - \bar{Y}_c)}{\sum_c (X_i - \bar{X}_c)^2 + \sum_{0y} (X_i - \bar{X}_{0y})^2}$$

y

$$(9.52) \quad \hat{\alpha}_m = \bar{Y}_c - \hat{\beta}_m \bar{X}_c.$$

Los estimadores de ξ y η , que tienen solamente un interés muy reducido, son

$$\hat{\xi} = \frac{\bar{Y}_{0x} - \hat{\alpha}_m}{\hat{\beta}_m}$$

y

$$\hat{\eta} = \hat{\alpha}_m + \hat{\beta}_m \bar{X}_{0y}.$$

Examinemos ahora la sesgadez o insesgadez de $\hat{\alpha}_m$ y $\hat{\beta}_m$. Respecto a $\hat{\beta}_m$ tenemos

$$(9.51a) \quad \hat{\beta}_m = \frac{\beta \sum_c (X_i - \bar{X}_c)^2 + \sum_c (X_i - \bar{X}_c)(e_i - \bar{e}_c)}{\sum_c (X_i - \bar{X}_c)^2 + \sum_{oy} (X_i - \bar{X}_{oy})^2}$$

y la esperanza matemática de $\hat{\beta}_m$, condicionada a los valores observados de X , es

$$E(\hat{\beta}_m) = \frac{\beta \sum_c (X_i - \bar{X}_c)^2}{\sum_c (X_i - \bar{X}_c)^2 + \sum_{oy} (X_i - \bar{X}_{oy})^2} \neq \beta.$$

Esto significa que el estimador modificado de la regresión de orden cero para β es, en general, *sesgado*, y lo mismo ocurre con respecto al estimador correspondiente de α .

Examinemos de nuevo algunos casos especiales. En primer lugar, supongamos que disponemos de todos los valores de X y que sólo faltan algunos valores de Y . En este caso, se puede mostrar fácilmente que las fórmulas (9.51) y (9.52) quedan inalteradas, lo que significa que no conseguimos mejora alguna. Por otra parte, supongamos que disponemos de todos los valores de Y y que solamente faltan algunos valores de X . En tal caso, las fórmulas (9.51) y (9.52) se convierten en las fórmulas (9.42) y (9.43), lo cual significa que los estimadores $\hat{\alpha}_m$ y $\hat{\beta}_m$ son exactamente iguales a los estimadores minimocuadráticos ordinarios basados únicamente en los pares de observaciones completas.

Resumen

Podemos resumir todo este análisis señalando que, cuando trabajamos con muestras en las que algunos pares de observaciones están incompletos, la información contenida en estos pares tiene una utilidad relativamente pequeña para la estimación de los coeficientes de regresión. Cuando X es no estocástica, la información contenida en los pares en los que únicamente se conoce el valor de Y nos permite obtener una estimación de la pérdida de eficiencia debida a la falta de algunos de los valores de X . Si esta pérdida es considerable, puede resultar conveniente tomarse la molestia de intentar recuperar los valores que faltan de X , o encontrar buenas aproximaciones de ellos. Cuando X es estocástica y utilizamos el método de la regresión de orden cero o su versión modificada, obtenemos estimadores generalmente sesgados. Si solamente faltan valores de X , los dos métodos producirán estimaciones insesgadas de β , pero

las estimaciones no serán más eficientes que las estimaciones minimocuadráticas ordinarias basadas únicamente en los pares completos. Un punto a favor de los estimadores de los coeficientes de regresión obtenidos por el método de la regresión de orden cero o su versión modificada es el hecho de que, cuando la correlación entre X e Y es pequeña, la media del error al cuadrado de estos estimadores es menor que la de los estimadores minimocuadráticos ordinarios basados en los pares completos.⁹ Por tanto, en algunas circunstancias, cualquiera de los métodos anteriores puede ser más conveniente que la estimación a partir de los pares completos solamente.

EJERCICIOS:

9-1. Considerando el modelo de los «errores en las variables», estimar la relación entre $\log(V/PL)$ y $\log(W/P)$ a partir de los datos de la industria de la fabricación de muebles que aparecen en el cuadro 9.1. Utilizar el método de la regresión ponderada con $\lambda = 2$.

9-2. Supóngase que los grupos de renta que aparecen en el cuadro 9-2 del texto se combinan de la forma siguiente:

Grupos de renta	Número de observaciones (n_r)
Menos de 18	51
18 a menos de 22	44
22 a menos de 26	36
26 a menos de 30	26
30 a menos de 34	16
34 o más	8

Calcular los valores apropiados de $\bar{X}_g$ e $\bar{Y}_g$ y utilizarlos para estimar los coeficientes de

$$Y = \alpha + \beta X + \varepsilon$$

y sus errores standard. Comparar los resultados con los basados en la información originariamente en el cuadro 9-2.

9-3. Efectuar la obtención de la fórmula (9.47).

⁹ Ver *ibid.* para la demostración y ampliación de esta afirmación.

9-4. Considerar las siguientes observaciones de X (precio de las naranjas) e Y (cantidad vendida de naranjas):

X	Y
100.	55
90	70
80	90
70	100
70	90
70	105
70	80
65	110
60	125
60	115
55	130
50	130
—	130
—	140

Estimar la pérdida de eficiencia en la estimación de β debida a la no consideración de las dos últimas observaciones incompletas.

9-5. Dados los momentos muestrales de los valores observados de X e Y , el estimador de regresión ponderada de β —definido en la ecuación (9.7)— se convierte en función de λ . En el ejemplo presentado en el texto hemos obtenido

$$m_{xx} = 20.1834/121,$$

$$m_{yy} = 14.4456/121,$$

$$m_{xy} = 16.9577/121.$$

Calcular los valores de $\hat{\beta}$ correspondientes a distintos valores de λ y representar gráficamente los resultados.

Regresión múltiple

El modelo de regresión presentado en el capítulo 7 se aplica a las regresiones que tienen una sola variable explicativa. Cuando el modelo se amplía con el fin de introducir más de una variable explicativa, hablamos de modelos de *regresión múltiple*. Las relaciones que se pueden representar a través de los modelos de regresión múltiple son muy corrientes en economía. Por ejemplo, en las funciones de producción, la cantidad producida es generalmente función de varios factores; en las funciones de consumo, la variable dependiente puede estar influida por la renta y por otros factores, y en las funciones de demanda, las variables explicativas tradicionales son el precio del producto, los precios de los sustitutivos, y la renta.

El modelo de regresión múltiple utilizado para representar este tipo de relaciones es una ampliación natural del modelo de regresión simple. En realidad, la mayor parte de los resultados obtenidos en el modelo de regresión simple se pueden generalizar fácilmente con el fin de hacerlos aplicables al caso de la regresión múltiple. En la sección 10.1 se presentan los resultados básicos relacionados con la estimación y en la sección 10.2 se analiza la contrastación de hipótesis y la predicción. La sección 10.3 está dedicada a la multicolinealidad —característica que sólo puede presentarse en los modelos de regresión con dos o más variables explicativas—. Por último, en la sección 10.4 se examina la validez de los resultados de las secciones anteriores en los casos en los que no se especifica correctamente la ecuación de regresión.

10.1 Estimación de los parámetros de la regresión

En la economía, un tipo corriente de proposición teórica afirma que los cambios de una variable se pueden explicar a partir de los cambios en varias va-

Examinemos ahora la sesgadez o insesgadez de $\hat{\alpha}_m$ y $\hat{\beta}_m$. Respecto a $\hat{\beta}_m$ tenemos

$$(9.51a) \quad \hat{\beta}_m = \frac{\beta \sum_c (X_i - \bar{X}_c)^2 + \sum_c (X_i - \bar{X}_c)(e_i - \bar{e}_c)}{\sum_c (X_i - \bar{X}_c)^2 + \sum_{oy} (X_i - \bar{X}_{oy})^2}$$

y la esperanza matemática de $\hat{\beta}_m$, condicionada a los valores observados de X , es

$$E(\hat{\beta}_m) = \frac{\beta \sum_c (X_i - \bar{X}_c)^2}{\sum_c (X_i - \bar{X}_c)^2 + \sum_{oy} (X_i - \bar{X}_{oy})^2} \neq \beta.$$

Esto significa que el estimador modificado de la regresión de orden cero para β es, en general, *sesgado*, y lo mismo ocurre con respecto al estimador correspondiente de α .

Examinemos de nuevo algunos casos especiales. En primer lugar, supongamos que disponemos de todos los valores de X y que sólo faltan algunos valores de Y . En este caso, se puede mostrar fácilmente que las fórmulas (9.51) y (9.52) quedan inalteradas, lo que significa que no conseguimos mejora alguna. Por otra parte, supongamos que disponemos de todos los valores de Y y que solamente faltan algunos valores de X . En tal caso, las fórmulas (9.51) y (9.52) se convierten en las fórmulas (9.42) y (9.43), lo cual significa que los estimadores $\hat{\alpha}_m$ y $\hat{\beta}_m$ son exactamente iguales a los estimadores minimocuadráticos ordinarios basados únicamente en los pares de observaciones completas.

Resumen

Podemos resumir todo este análisis señalando que, cuando trabajamos con muestras en las que algunos pares de observaciones están incompletos, la información contenida en estos pares tiene una utilidad relativamente pequeña para la estimación de los coeficientes de regresión. Cuando X es no estocástica, la información contenida en los pares en los que únicamente se conoce el valor de Y nos permite obtener una estimación de la pérdida de eficiencia debida a la falta de algunos de los valores de X . Si esta pérdida es considerable, puede resultar conveniente tomarse la molestia de intentar recuperar los valores que faltan de X , o encontrar buenas aproximaciones de ellos. Cuando X es estocástica y utilizamos el método de la regresión de orden cero o su versión modificada, obtenemos estimadores generalmente sesgados. Si solamente faltan valores de X , los dos métodos producirán estimaciones insesgadas de β , pero

estas estimaciones no serán más eficientes que las estimaciones minimocuadráticas ordinarias basadas únicamente en los pares completos. Un punto a favor de los estimadores de los coeficientes de regresión obtenidos por el método de la regresión de orden cero o su versión modificada es el hecho de que, cuando la correlación entre X e Y es pequeña, la media del error al cuadrado de estos estimadores es menor que la de los estimadores minimocuadráticos ordinarios basados en los pares completos.⁹ Por tanto, en algunas circunstancias, cualquiera de los métodos anteriores puede ser más conveniente que la estimación a partir de los pares completos solamente.

EJERCICIOS:

9-1. Considerando el modelo de los «errores en las variables», estimar la relación entre $\log(V/PL)$ y $\log(W/P)$ a partir de los datos de la industria de la fabricación de muebles que aparecen en el cuadro 9.1. Utilizar el método de la regresión ponderada con $\lambda = 2$.

9-2. Supóngase que los grupos de renta que aparecen en el cuadro 9-2 del texto se combinan de la forma siguiente:

Grupos de renta	Número de observaciones (n_c)
Menos de 18	51
18 a menos de 22	44
22 a menos de 26	36
26 a menos de 30	26
30 a menos de 34	16
34 o más	8

Calcular los valores apropiados de $\bar{X}_c$ e $\bar{Y}_c$ y utilizarlos para estimar los coeficientes de

$$Y = \alpha + \beta X + e$$

y sus errores standard. Comparar los resultados con los basados en la información originariamente en el cuadro 9-2.

9-3. Efectuar la obtención de la fórmula (9.47).

⁹ Véase el apéndice para la demostración y ampliación de esta afirmación.

riables distintas. Este tipo de relaciones se puede expresar sencillamente, por medio de una ecuación de regresión lineal múltiple de la forma

$$(10.1) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK} + \varepsilon_i,$$

en la que Y representa la variable dependiente, las X indican las variables explicativas, y ε es una perturbación estocástica. El subíndice i hace referencia a la observación i -ésima, y el segundo subíndice utilizado en la descripción de las variables explicativas sirve para identificar a la variable en cuestión. El número de variables explicativas es $K - 1$, de modo que en el caso en que $K = 2$, la ecuación (10.1) se reduce a la ecuación de regresión simple. Una forma alternativa de expresar la ecuación (10.1) es

$$(10.1a) \quad Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + \varepsilon_i,$$

en la que $X_{i1} = 1$ para todo $i = 1, 2, \dots, n$. Al escribir el factor por el que se multiplica β_1 en la forma X_{i1} en lugar de simplemente 1, hacemos que la ecuación aparezca simétrica, sin introducir en realidad ningún cambio. Para completar la especificación del modelo de regresión, añadiremos los siguientes *supuestos básicos*:

(10.2) ε_i se distribuye normalmente

$$(10.3) \quad E(\varepsilon_i) = 0,$$

$$(10.4) \quad E(\varepsilon_i^2) = \sigma^2,$$

$$(10.5) \quad E(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j),$$

(10.6) Todas las variables explicativas son no estocásticas, con valores fijos en repetidas muestras y tales que, para todo tamaño muestral $\Sigma_{i=1}^n = (X_{ik} - \bar{X}_k)^2/n$ es un número finito distinto de cero para todos los valores de $k = 2, 3, \dots, K$.

(10.7) El número de observaciones es mayor que el número de coeficientes a estimar.

(10.8) No existe ninguna relación lineal exacta entre las variables explicativas.

Estos supuestos se consideran válidos para todas las observaciones. La especificación completa del modelo indicada en (10.1) a (10.8) describe el llamado «modelo clásico de regresión normal lineal» en el contexto de la regresión múltiple. Los supuestos (10.2) a (10.5) se refieren al término de perturbación y

los supuestos (7.2) a (7.5) correspondientes al modelo de regresión simple, mientras que los tres últimos se refieren a las variables explicativas. El supuesto (10.6) es igual al supuesto (7.6), salvo por el hecho de que se aplica a un mayor número de variables explicativas. Los supuestos (10.7) y (10.8) son nuevos. El primero de estos dos asegura la existencia de un número suficiente de «grados de libertad» en la estimación, mientras que el supuesto (10.8) señala que ninguna de las variables explicativas está perfectamente correlacionada con ninguna de las demás variables explicativas o con cualquier combinación lineal de estas últimas. Como se verá dentro de poco, este último supuesto es también necesario para la estimación.

Cuando el modelo de regresión múltiple se especifica de la forma anterior, la distribución de Y_i es normal, igual que en el caso del modelo de regresión simple.

La media de Y_i es

$$(10.9) \quad E(Y_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK},$$

y su variancia es

$$(10.10) \quad \text{Var}(Y_i) = E[Y_i - E(Y_i)]^2 = \sigma^2.$$

Obsérvese que, si utilizamos la ecuación (10.9), podemos interpretar los coeficientes de regresión en la forma siguiente:

β_1 = media de Y_i cuando todas las variables explicativas son iguales a cero;

β_k = cambio en $E(Y_i)$ correspondiente a un cambio unitario en la variable explicativa k , permaneciendo constantes todas las demás variables explicativas

$$= \frac{\delta E(Y_i)}{\delta X_{ik}} \quad (k = 2, 3, \dots, K).$$

En algunas ocasiones β_1 recibe el nombre de *ordenada en el origen* (o *constante de regresión*) y $\beta_2, \beta_3, \dots, \beta_K$ se denominan *pendientes de la regresión* (o *coeficientes de regresión parcial*).

Esta interpretación de los coeficientes de regresión posee una importante implicación respecto a su estimación. Considérese, por ejemplo, el problema consistente en estimar β_K . Como sabemos que β_K mide el efecto de X_{iK} sobre $E(Y_i)$ cuando se mantienen constantes las demás variables $X_{i2}, X_{i3}, \dots, X_{iK-1}$, una forma obvia para estimar β_K consistiría en utilizar observaciones efectuadas cuando todas las variables explicativas excepto X_{iK} se mantienen realmente constantes. Es decir, se obtendrían las observaciones a partir de un experimento

controlado en el cual todas las variables explicativas menos X_{ik} se mantendrían constantes a unos niveles determinados. Veamos lo que ocurriría con la estimación en este caso. Concretamente, supongamos que el nivel de X_{i2} se mantiene constante a ξ_2 , el de X_{i3} a ξ_3 , y así sucesivamente, y hagamos que X_{ik} varíe. En este caso, la ecuación de regresión (10.1) puede escribirse en la forma

$$Y_i = \beta_1 + \beta_2 \xi_2 + \beta_3 \xi_3 + \dots + \beta_{K-1} \xi_{K-1} + \beta_K X_{ik} + \epsilon_i$$

$$\text{o} \quad Y_i = \alpha + \beta_K X_{ik} + \epsilon_i,$$

lo cual muestra claramente que nos encontramos de nuevo en las condiciones de la regresión simple. Esto es precisamente lo que los laboratorios que realizan experimentos en las ciencias naturales tratan frecuentemente de conseguir. De esto se deriva que, si queremos mantener el supuesto de variables explicativas no estocásticas, y a la vez queremos mantener una justificación de la existencia del modelo de regresión múltiple, tenemos que excluir la posibilidad de que los valores de las variables explicativas puedan ser controlados por el investigador. Por tanto, solamente vamos a considerar las situaciones en las que el «experimento» lo ha llevado a cabo alguna otra persona distinta del econométra, y con finalidades distintas de la estimación de los coeficientes de regresión o de la contrastación de hipótesis acerca de ellos. Naturalmente, éste es el procedimiento normal por el cual se adquieren los datos económicos. El «laboratorio» es la sociedad, y el econométra es, en general, un simple espectador. La descripción del modelo clásico de regresión normal lineal se presenta habitualmente en forma de *notación matricial*. En primer lugar, la ecuación (10.1a) puede escribirse en la forma

$$(10.1b) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

en la que

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1K} \\ X_{21} & X_{22} & \dots & X_{2K} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Esto significa que las dimensiones de las matrices y vectores correspondientes son las siguientes:

$$\mathbf{Y} \rightarrow (n \times 1), \quad \boldsymbol{\beta} \rightarrow (K \times 1),$$

$$\mathbf{X} \rightarrow (n \times K), \quad \boldsymbol{\epsilon} \rightarrow (n \times 1).$$

Obsérvese particularmente que cada fila de la matriz $\mathbf{X}$ representa un conjunto de valores de las variables explicativas correspondientes a una observación, mientras que cada columna representa el conjunto de valores de una variable explicativa correspondientes a las n observaciones muestrales. La primera columna de $\mathbf{X}$ está compuesta únicamente por unos. Los supuestos (10.2) a (10.5) pueden presentarse en notación matricial de la forma siguiente

$$(10.2a-3a) \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \text{ siendo } \mathbf{0} \text{ un vector columna de ceros, y } \boldsymbol{\Sigma} \rightarrow (n \times n),$$

$$(10.4a-5a) \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n \text{ siendo } \mathbf{I}_n \text{ una matriz identidad de orden } (n \times n), \text{ con unos a lo largo de la diagonal principal y ceros en todos los lugares restantes.}$$

La afirmación presentada en (10.4a-5a) combina los supuestos de homoscedasticidad y no autorregresión; las perturbaciones que cumplen ambos supuestos se denominan «esféricas». Por último, los supuestos (10.6) a (10.8) acerca de las variables explicativas pueden expresarse nuevamente en la forma

$$(10.6a-8a) \quad \text{Los elementos de la matriz } \mathbf{X} \text{ son no estocásticos, con valores fijos en repetidas muestras, y la matriz } (1/n)(\mathbf{X}'\mathbf{X}) \text{ es no singular y sus elementos son finitos para todo tamaño muestral.}$$

Estimación por mínimos cuadrados

Consideremos ahora la obtención de los *estimadores minimocuadráticos* de los coeficientes de regresión. La suma de cuadrados que queremos minimizar es

$$S = \sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK})^2.$$

Derivando S respecto de $\beta_1, \beta_2, \dots, \beta_K$, tenemos

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK}),$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum_i X_{i2} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK}),$$

$$\vdots$$

$$\frac{\partial S}{\partial \beta_K} = -2 \sum_i X_{iK} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK}).$$

Igualando a cero estas derivadas, y reordenando los términos, se obtienen las siguientes ecuaciones normales minimocuadráticas:

$$\begin{aligned} \sum_i Y_i &= \beta_1 n + \beta_2 \sum_i X_{i2} + \beta_3 \sum_i X_{i3} + \cdots + \beta_K \sum_i X_{iK}, \\ \sum_i X_{i2} Y_i &= \beta_1 \sum_i X_{i2} + \beta_2 \sum_i X_{i2}^2 + \beta_3 \sum_i X_{i2} X_{i3} + \cdots + \beta_K \sum_i X_{i2} X_{iK}, \\ &\vdots \\ \sum_i X_{iK} Y_i &= \beta_1 \sum_i X_{iK} + \beta_2 \sum_i X_{i2} X_{iK} + \beta_3 \sum_i X_{i3} X_{iK} + \cdots + \beta_K \sum_i X_{iK}^2. \end{aligned}$$

Estas ecuaciones representan una simple generalización de las ecuaciones (7.10) y (7.11). (En la nota 4 de la página 243 se indica una regla sencilla para recordar la formación de las ecuaciones normales minimocuadráticas.) Obsérvese que en este sistema de ecuaciones normales no se pueden despejar las β desconocidas si 1) el número de variables explicativas más una es mayor que el número de observaciones, o si 2) una de las variables explicativas es una combinación lineal exacta de algunas de las restantes. En el primer caso, el número de ecuaciones sería menor que el número de incógnitas, y en el segundo las ecuaciones no serían independientes.

Con el fin de resolver las ecuaciones normales minimocuadráticas, obsérvese que la primera ecuación puede escribirse en la forma

$$(10.11) \quad \beta_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \bar{X}_3 - \cdots - \beta_K \bar{X}_K,$$

$$\text{en la que} \quad \bar{Y} = \frac{1}{n} \sum_i Y_i \quad \text{y} \quad \bar{X}_k = \frac{1}{n} \sum_i X_{ik} \quad (k = 2, 3, \dots, K).$$

Sustituyendo la ecuación (10.11) en las restantes ecuaciones normales y efectuando algunas simplificaciones, obtenemos

$$\begin{aligned} m_{Y2} &= m_{22}\beta_2 + m_{23}\beta_3 + \cdots + m_{2K}\beta_K, \\ m_{Y3} &= m_{32}\beta_2 + m_{33}\beta_3 + \cdots + m_{3K}\beta_K, \\ &\vdots \\ m_{YK} &= m_{K2}\beta_2 + m_{K3}\beta_3 + \cdots + m_{KK}\beta_K, \end{aligned}$$

$$\text{en la que} \quad m_{Yk} = \sum_i (Y_i - \bar{Y})(X_{ik} - \bar{X}_k)$$

$$\text{y} \quad m_{jk} = \sum_i (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \quad (j, k = 2, 3, \dots, K).$$

A partir de estas ecuaciones podemos despejar $\hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_K$. La solución es inmediata, pero puede resultar algo laboriosa. En el caso en que existen dos variables explicativas (es decir, cuando $K = 3$) tenemos

$$(10.12) \quad \hat{\beta}_2 = \frac{\begin{vmatrix} m_{Y2} & m_{23} \\ m_{Y3} & m_{33} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{23} & m_{33} \end{vmatrix}} = \frac{m_{Y2}m_{33} - m_{Y3}m_{23}}{m_{22}m_{33} - m_{23}^2},$$

$$(10.13) \quad \hat{\beta}_3 = \frac{\begin{vmatrix} m_{22} & m_{Y2} \\ m_{23} & m_{Y3} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{23} & m_{33} \end{vmatrix}} = \frac{m_{Y3}m_{22} - m_{Y2}m_{23}}{m_{22}m_{33} - m_{23}^2}.$$

Las ecuaciones normales minimocuadráticas se pueden expresar en notación matricial en la forma siguiente

$$(X'Y) = (X'X)\beta,$$

en la que

$$(X'Y) = \begin{bmatrix} \sum Y_i \\ \sum X_{i2} Y_i \\ \vdots \\ \sum X_{iK} Y_i \end{bmatrix}, \quad (X'X) = \begin{bmatrix} n & \sum X_{i2} & \cdots & \sum X_{iK} \\ \sum X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2} X_{iK} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{iK} & \sum X_{i2} X_{iK} & \cdots & \sum X_{iK}^2 \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}.$$

La solución para $\hat{\beta}$ se convierte simplemente en

$$(10.14) \quad \hat{\beta} = (X'X)^{-1}(X'Y).$$

De forma alternativa, podemos eliminar $\hat{\beta}_1$ por sustitución a partir de la ecuación (10.11), resolviendo después el sistema de ecuaciones reducido para obtener

$$\hat{\beta} = (X'X)^{-1}(X'Y).$$

en la que

$$\underline{\beta} = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{bmatrix}, \quad (\underline{X}'\underline{X}) = \begin{bmatrix} m_{22} & m_{23} & \dots & m_{2K} \\ m_{23} & m_{33} & \dots & m_{3K} \\ \vdots & \vdots & \ddots & \vdots \\ m_{2K} & m_{3K} & \dots & m_{KK} \end{bmatrix}, \quad (\underline{X}'\underline{Y}) = \begin{bmatrix} m_{Y2} \\ m_{Y3} \\ \vdots \\ m_{YK} \end{bmatrix}.$$

EJEMPLO. En la sección 7.3 se ilustró el método de los mínimos cuadrados en el contexto de la regresión simple estimando la relación entre el precio y la cantidad de naranjas vendidas en un supermercado durante doce días consecutivos. Modifiquemos ahora este supuesto, postulando que la cantidad vendida depende no solamente del precio, sino también del importe gastado en publicidad del producto. Es decir, supongamos

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

siendo Y_i = cantidad (libras) de naranjas vendidas, X_{i2} = precio en centavos por libra, y X_{i3} = gastos en publicidad, en dólares. Los datos aparecen en el cuadro 10-1.

Cuadro 10-1

Cantidad (en libras)	Precio (centavos/libra)	Gastos en publicidad (dólares)
55	100	5.50
70	90	6.30
90	80	7.20
100	70	7.00
90	70	6.30
105	70	7.35
80	70	5.60
110	65	7.15
125	60	7.50
115	60	6.90
130	55	7.15
130	50	6.50

Los resultados de los cálculos básicos son los siguientes

$$\begin{aligned} \bar{Y} &= 100, & m_{22} &= 2250, & m_{Y2} &= -3550, \\ \bar{X}_2 &= 70, & m_{33} &= 4.86, & m_{Y3} &= 125.25, \\ \bar{X}_3 &= 6.7, & m_{23} &= -54, & m_{YY} &= 6300. \end{aligned}$$

La cantidad $m_{YY} = \sum (Y_i - \bar{Y})^2$ la utilizaremos en posteriores ejemplos.) Las ecuaciones de los coeficientes de regresión son

$$\begin{aligned} \hat{\beta}_2 &= \frac{-3550 \times 4.86 - (-54) \times 125.25}{2250 \times 4.86 - (-54)^2} = \frac{-10,631.5}{8019} = -1.326, \\ \hat{\beta}_3 &= \frac{2250 \times 125.25 - (-54) \times (-3550)}{8019} = \frac{90,112.5}{8019} = 11.237, \end{aligned}$$

$$y \quad \hat{\beta}_1 = 100 - (-1.326) \times 70 - 11.237 \times 6.7 = 117.532.$$

Por tanto, la ecuación de regresión estimada es

$$Y_i = 117.532 - 1.326 X_{i2} + 11.237 X_{i3} + \varepsilon_i.$$

Esto indica que hemos estimado que una disminución de 10 centavos en el precio de las naranjas aumentará las ventas en unas 13 libras aproximadamente si el gasto en publicidad permanece constante, mientras que un aumento de 1 dólar en publicidad aumentaría las ventas en unas 11 libras si el precio permaneciese inalterado.

Estimación lineal insesgada óptima

Vamos a obtener ahora los *estimadores lineales insesgados óptimos* (ELIO) del modelo de regresión múltiple. Por razones de simplicidad utilizaremos un modelo con dos variables explicativas, ya que la ampliación al caso general requiere únicamente mayor cantidad de expresiones algebraicas. Supongamos, pues, que queremos obtener el ELIO de β_2 en el modelo

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i.$$

Según la condición de linealidad, tenemos

$$\hat{\beta}_2 = \sum_i a_i Y_i,$$

en la que a_i ($i = 1, 2, \dots, n$) son constantes que debemos obtener. La esperanza matemática de $\hat{\beta}_2$ es

$$E(\hat{\beta}_2) = E \sum_i a_i Y_i = E \sum_i a_i (\beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i)$$

$$\beta_1 \sum_i a_i + \beta_2 \sum_i a_i X_{i2} + \beta_3 \sum_i a_i X_{i3} + \sum_i a_i \varepsilon_i.$$

de modo que, para que $\tilde{\beta}_2$ sea insesgada, es preciso que

$$\begin{aligned}\sum_i a_i &= 0, \\ \sum_i a_i X_{i2} &= 1, \\ \sum_i a_i X_{i3} &= 0.\end{aligned}$$

La variancia de $\tilde{\beta}_2$ es

$$\text{Var}(\tilde{\beta}_2) = E\left[\sum_i a_i Y_i - E\left(\sum_i a_i Y_i\right)\right]^2 = E\left(\sum_i a_i \varepsilon_i\right)^2 = \sigma^2 \sum_i a_i^2.$$

Es decir, tenemos que obtener los valores de a_i que minimizan

$$\sigma^2 \sum_i a_i^2,$$

condicionada a

$$\begin{aligned}\sum_i a_i &= 0, \\ \sum_i a_i X_{i2} &= 1, \\ \sum_i a_i X_{i3} &= 0.\end{aligned}$$

Utilizando el método de los multiplicadores de Lagrange obtenemos

$$H = \sigma^2 \sum_i a_i^2 - \lambda_1 \left(\sum_i a_i\right) - \lambda_2 \left(\sum_i a_i X_{i2} - 1\right) - \lambda_3 \left(\sum_i a_i X_{i3}\right).$$

Derivando H respecto de $a_1, a_2, \dots, a_n, \lambda_1, \lambda_2$ y λ_3 e igualando a cero todas estas derivadas, obtenemos

$$(10.16) \quad 2a_i \sigma^2 - \lambda_1 - \lambda_2 X_{i2} - \lambda_3 X_{i3} = 0 \quad (i = 1, 2, \dots, n),$$

$$(10.17) \quad -\left(\sum_i a_i\right) = 0,$$

$$(10.18) \quad -\left(\sum_i a_i X_{i2}\right) + 1 = 0,$$

$$(10.19) \quad -\left(\sum_i a_i X_{i3}\right) = 0.$$

Las primeras $(n+3)$ ecuaciones a partir de las cuales tenemos que resolver las incógnitas $a_1, a_2, \dots, a_n, \lambda_1, \lambda_2$ y λ_3 . Impezaremos sumando las n ecuaciones de (10.16). Entonces obtenemos

$$(10.20) \quad 2\sigma^2 \sum_i a_i - n\lambda_1 - \lambda_2 \sum_i X_{i2} - \lambda_3 \sum_i X_{i3} = 0.$$

A continuación, multiplicando ambos lados de (10.16) por X_{i2} y sumando las ecuaciones resultantes obtenemos

$$(10.21) \quad 2\sigma^2 \sum_i a_i X_{i2} - \lambda_1 \sum_i X_{i2} - \lambda_2 \sum_i X_{i2}^2 - \lambda_3 \sum_i X_{i2} X_{i3} = 0.$$

Por último, multiplicamos los dos lados de (10.16) por X_{i3} y sumamos de nuevo estas ecuaciones:

$$(10.22) \quad 2\sigma^2 \sum_i a_i X_{i3} - \lambda_1 \sum_i X_{i3} - \lambda_2 \sum_i X_{i2} X_{i3} - \lambda_3 \sum_i X_{i3}^2 = 0,$$

Sustituyendo $\sum_i a_i, \sum_i a_i X_{i2}$, y $\sum_i a_i X_{i3}$ por sus valores según las ecuaciones (10.17), (10.18) y (10.19), en las ecuaciones (10.20), (10.21) y (10.22), obtenemos

$$(10.20a) \quad -n\lambda_1 - \lambda_2 \sum_i X_{i2} - \lambda_3 \sum_i X_{i3} = 0,$$

$$(10.21a) \quad -\lambda_1 \sum_i X_{i2} - \lambda_2 \sum_i X_{i2}^2 - \lambda_3 \sum_i X_{i2} X_{i3} = -2\sigma^2,$$

$$(10.22a) \quad -\lambda_1 \sum_i X_{i3} - \lambda_2 \sum_i X_{i2} X_{i3} - \lambda_3 \sum_i X_{i3}^2 = 0.$$

A partir de (10.20a) obtenemos

$$\lambda_1 = -\lambda_2 \bar{X}_2 - \lambda_3 \bar{X}_3.$$

podemos utilizar esta expresión de λ_1 en las ecuaciones (10.21a) y (10.22a), con lo cual obtenemos

$$\lambda_2 m_{22} + \lambda_3 m_{23} = 2\sigma^2,$$

$$\lambda_2 m_{23} + \lambda_3 m_{33} = 0.$$

Por tanto,

$$\lambda_2 = \frac{2\sigma^2 m_{33}}{m_{22}m_{33} - m_{23}^2}$$

$$\lambda_3 = \frac{-2\sigma^2 m_{23}}{m_{22}m_{33} - m_{23}^2}.$$

lo que implica que

$$\lambda_1 = \frac{2\sigma^2(-\bar{X}_2 m_{33} + \bar{X}_3 m_{23})}{m_{22}m_{33} - m_{23}^2}.$$

Por este camino hemos obtenido la solución de λ_1 , λ_2 y λ_3 . A partir de (10.16) tenemos

$$a_i = \frac{1}{2\sigma^2}(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 X_{i3}),$$

y sustituyendo las λ por sus valores esta expresión se convierte en

$$(10.23) \quad a_i = \frac{-\bar{X}_2 m_{33} + \bar{X}_3 m_{23} + m_{33} X_{i2} - m_{23} X_{i3}}{m_{22}m_{33} - m_{23}^2}.$$

Por tanto,

$$\begin{aligned} \beta_2 &= \sum_i a_i Y_i = \frac{-n\bar{Y}\bar{X}_2 m_{33} + n\bar{Y}\bar{X}_3 m_{23} + m_{33} \sum_i X_{i2} Y_i - m_{23} \sum_i X_{i3} Y_i}{m_{22}m_{33} - m_{23}^2} \\ &= \frac{m_{33}m_{Y2} - m_{23}m_{Y3}}{m_{22}m_{33} - m_{23}^2}, \end{aligned}$$

que es precisamente la fórmula que obtuvimos en (10.12) para el estimador minimocuadrático de β_2 . Del mismo modo, podríamos mostrar que $\tilde{\beta}_3 = \hat{\beta}_3$ y $\tilde{\beta}_1 = \hat{\beta}_1$. Así pues, *los estimadores minimocuadráticos de los coeficientes de regresión son ELIO*. Esta conclusión es válida para los modelos de regresión con cualquier número de variables explicativas.

El método ELIO nos proporciona no solamente la fórmula de $\tilde{\beta}_2$, sino también la de su variancia. Concretamente,

$$\text{Var}(\tilde{\beta}_2) = \sigma^2 \sum_i a_i^2.$$

Multiplicando los dos lados de (10.23) por a_i , sumando respecto a todas las observaciones y sustituyendo $\sum_i a_i$, $\sum_i a_i X_{i2}$, y $\sum_i a_i X_{i3}$ por sus valores obtenemos

$$\sum_i a_i^2 = \frac{m_{33}}{m_{22}m_{33} - m_{23}^2}.$$

Esto significa que

$$(10.24) \quad \text{Var}(\tilde{\beta}_2) = \frac{\sigma^2 m_{33}}{m_{22}m_{33} - m_{23}^2}.$$

el resultado correspondiente a la variancia de $\tilde{\beta}_3$, que no vamos a obtener aquí, es

$$(10.25) \quad \text{Var}(\tilde{\beta}_3) = \frac{\sigma^2 m_{22}}{m_{22}m_{33} - m_{23}^2}.$$

Estimación maximoverosímil

A continuación, pasamos a ocuparnos de la obtención de los *estimadores maximoverosímiles* de los parámetros del modelo de regresión normal lineal múltiple. Esto representa una ampliación directa del procedimiento utilizado en el modelo de regresión simple. Al establecer la función de verosimilitud como en el caso de (7.26), pero teniendo en cuenta más de una variable explicativa, obtenemos

$$\begin{aligned} L &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 \\ &\quad - \frac{1}{2\sigma^2} \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK})^2. \end{aligned}$$

Los estimadores maximoverosímiles se obtienen derivando L respecto de β_1 , $\beta_2, \dots, \beta_K$ y σ^2 , e igualando a cero cada una de estas derivadas. Puede verse fácilmente que las primeras K ecuaciones, a partir de las cuales pueden despejarse los valores de las β , son exactamente iguales que las ecuaciones normales minimocuadráticas. Esto significa que *los estimadores maximoverosímiles de los coeficientes de regresión equivalen a los estimadores minimocuadráticos*. El estimador maximoverosímil de σ^2 es

$$(10.26) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK})^2$$

o

$$(10.26a) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_i e_i^2,$$

en las que los términos e_i representan los residuos minimocuadráticos.

10.2 Otros resultados de la inferencia estadística

A partir de la discusión anterior, acerca de la estimación de los coeficientes del modelo de regresión múltiple, llegamos a la conclusión de que los tres métodos de estimación que hemos considerado dan como resultado exactamente las mismas fórmulas, la igual que sucedía en el caso del modelo de regresión simple. Una ampliación de este resultado es que los estimadores minimocuadráticos

dráticos de los coeficientes de regresión poseen todas las propiedades deseables. A continuación, vamos a presentar algunas características adicionales de estos estimadores y a estudiar su utilización en la contrastación de hipótesis y la predicción.

Variancias y covariancias de los estimadores minimocuadráticos

Las fórmulas de las variancias de los coeficientes de regresión estimados en el modelo con dos variables explicativas se han presentado en (10.24) y (10.25). Estas fórmulas se pueden generalizar de forma inmediata a los modelos en los que existe un número mayor de variables explicativas. En el caso en el que tenemos $(K - 1)$ variables el resultado es

$$(10.27) \quad E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2(\mathbf{X}'\mathbf{X})^{-1},$$

o, expresado explícitamente,

(10.27a)

$$\begin{bmatrix} \text{Var}(\hat{\beta}_2) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_K) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) & \text{Var}(\hat{\beta}_3) & \dots & \text{Cov}(\hat{\beta}_3, \hat{\beta}_K) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_K) & \text{Cov}(\hat{\beta}_3, \hat{\beta}_K) & \dots & \text{Var}(\hat{\beta}_K) \end{bmatrix} = \sigma^2 \begin{bmatrix} m_{22} & m_{23} & \dots & m_{2K} \\ m_{23} & m_{33} & \dots & m_{3K} \\ \vdots & \vdots & \ddots & \vdots \\ m_{2K} & m_{3K} & \dots & m_{KK} \end{bmatrix}^{-1}$$

La matriz $[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$, cuya dimensión es $(K - 1) \times (K - 1)$, es la llamada *matriz de variancias y covariancias* de los estimadores minimocuadráticos de las pendientes de regresión. En esta matriz, las variancias de los estimadores aparecen a lo largo de la diagonal principal, mientras que las covariancias son los términos restantes.

Hasta este momento no hemos obtenido la fórmula de la variancia de la ordenada en el origen, y vamos a hacerlo ahora, empezando en primer lugar con el modelo de dos variables explicativas, y pasando a continuación al caso general. En el caso en el que sólo existen dos variables explicativas, tenemos

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + \bar{\varepsilon} - \bar{\beta}_2 \bar{X}_2 - \bar{\beta}_3 \bar{X}_3,$$

o sea, que

$$\hat{\beta}_1 - \beta_1 = -(\hat{\beta}_2 - \beta_2)\bar{X}_2 - (\hat{\beta}_3 - \beta_3)\bar{X}_3 + \bar{\varepsilon}.$$

lo que significa que

$$(10.28) \quad \text{Var}(\hat{\beta}_1) = \bar{X}_2^2 \text{Var}(\hat{\beta}_2) + \bar{X}_3^2 \text{Var}(\hat{\beta}_3) + 2\bar{X}_2 \bar{X}_3 \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) - 2\bar{X}_2 E\bar{\varepsilon}(\hat{\beta}_2 - \beta_2) - 2\bar{X}_3 E\bar{\varepsilon}(\hat{\beta}_3 - \beta_3) + \text{Var}(\bar{\varepsilon}).$$

Los tres primeros términos del lado derecho de la ecuación (10.28) pueden obtenerse a partir de (10.27). Para calcular los dos términos que aparecen a continuación, obsérvese en primer lugar que

$$\begin{aligned} m_{Y2} &= \sum_i (Y_i - \bar{Y})(X_{i2} - \bar{X}_2) \\ &= \sum_i [\beta_2(X_{i2} - \bar{X}_2) + \beta_3(X_{i3} - \bar{X}_3) + (\varepsilon_i - \bar{\varepsilon})](X_{i2} - \bar{X}_2) \\ &= \beta_2 m_{22} + \beta_3 m_{23} + \sum_i (X_{i2} - \bar{X}_2)\varepsilon_i \end{aligned}$$

$$\text{y} \quad m_{Y3} = \beta_2 m_{23} + \beta_3 m_{33} + \sum_i (X_{i3} - \bar{X}_3)\varepsilon_i.$$

Podemos sustituir estas expresiones en la fórmula (10.12) para obtener

$$\begin{aligned} \hat{\beta}_2 &= \frac{m_{33}[\beta_2 m_{22} + \beta_3 m_{23} + \sum_i (X_{i2} - \bar{X}_2)\varepsilon_i]}{m_{22}m_{33} - m_{23}^2} \\ &\quad - \frac{m_{23}[\beta_2 m_{22} + \beta_3 m_{23} + \sum_i (X_{i3} - \bar{X}_3)\varepsilon_i]}{m_{22}m_{33} - m_{23}^2} \end{aligned}$$

o

$$(10.29) \quad \hat{\beta}_2 - \beta_2 = \frac{m_{33} \sum_i (X_{i2} - \bar{X}_2)\varepsilon_i - m_{23} \sum_i (X_{i3} - \bar{X}_3)\varepsilon_i}{m_{22}m_{33} - m_{23}^2}.$$

Del mismo modo, sustituyendo m_{Y2} y m_{Y3} por sus valores en la ecuación (10.13) obtenemos

$$(10.30) \quad \hat{\beta}_3 - \beta_3 = \frac{m_{22} \sum_i (X_{i3} - \bar{X}_3)\varepsilon_i - m_{23} \sum_i (X_{i2} - \bar{X}_2)\varepsilon_i}{m_{22}m_{33} - m_{23}^2}.$$

Por tanto,

$$E(\hat{\beta}_2 - \beta_2)\bar{\varepsilon} = \frac{m_{33} \sum_i (X_{i2} - \bar{X}_2)E(\varepsilon_i \bar{\varepsilon}) - m_{23} \sum_i (X_{i3} - \bar{X}_3)E(\varepsilon_i \bar{\varepsilon})}{m_{22}m_{33} - m_{23}^2} = 0,$$

y también

$$E(\hat{\beta}_3 - \beta_3)\bar{\varepsilon} = 0.$$

Por último, sabemos que

$$\text{Var}(\bar{\varepsilon}) = \frac{\sigma^2}{n}.$$

Además,

$$E \sum_i (e_i - \bar{e})^2 = E(\sum_i e_i^2) - 2E(\sum_i e_i \bar{e}) + nE(\bar{e}^2) = n\sigma^2 - 2\sigma^2 + \sigma^2 \\ = (n-1)\sigma^2.$$

Y por tanto,

$$E \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3})^2 = \frac{\sigma^2(m_{22}m_{33} + m_{23}m_{32} - 2m_{23}^2)}{m_{22}m_{33} - m_{23}^2} \\ - 2\sigma^2 - 2\sigma^2 + (n-1)\sigma^2 \\ = (n-3)\sigma^2.$$

El estimador insesgado de σ^2 , al que llamaremos s^2 , en el modelo de regresión con dos variables explicativas es, pues,

$$(10.34) \quad s^2 = \frac{1}{n-3} \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3})^2.$$

Esta fórmula puede generalizarse al caso en que existen $(K-1)$ variables explicativas en la forma siguiente:

$$(10.35) \quad s^2 = \frac{1}{n-K} \sum_i (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \dots - \beta_K X_{iK})^2.$$

Como s^2 equivale asintóticamente al estimador maximoverosímil de σ^2 , $\hat{\sigma}^2$ se deduce que s^2 posee las mismas propiedades asintóticas óptimas que dicho estimador.

A efectos del cálculo del valor de s^2 las fórmulas (10.34) y (10.35) se pueden simplificar con el fin de evitar el cálculo de los residuos individuales. Veámoslo, en primer lugar, para el caso de dos variables explicativas,

$$s^2 = \frac{1}{n-3} \sum_i [(Y_i - \bar{Y}) - \beta_2(X_{i2} - \bar{X}_2) - \beta_3(X_{i3} - \bar{X}_3)]^2 \\ = \frac{1}{n-3} (m_{YY} + \beta_2^2 m_{22} + \beta_3^2 m_{33} - 2\beta_2 m_{Y2} - 2\beta_3 m_{Y3} + 2\beta_2 \beta_3 m_{23}).$$

Pero, a partir de las ecuaciones normales minimocuadráticas, sabemos que

$$m_{Y2} = \beta_2 m_{22} + \beta_3 m_{23},$$

$$m_{Y3} = \beta_2 m_{23} + \beta_3 m_{33}.$$

Por tanto,

$$\beta_2^2 m_{22} + \beta_2 \beta_3 m_{23} - \beta_2 m_{Y2} = 0,$$

$$\beta_2 \beta_3 m_{23} + \beta_3^2 m_{33} - \beta_3 m_{Y3} = 0.$$

Y, pues, s^2 se convierte en

$$(10.34a) \quad s^2 = \frac{1}{n-3} (m_{YY} - \beta_2 m_{Y2} - \beta_3 m_{Y3}).$$

Esto se puede generalizar al caso en que existen $(K-1)$ variables explicativas, de modo que la ecuación (10.35) se convierte en

$$(10.35a) \quad s^2 = \frac{1}{n-K} (m_{YY} - \beta_2 m_{Y2} - \beta_3 m_{Y3} - \dots - \beta_K m_{YK}).$$

A efectos de cálculo, estas fórmulas resultan más convenientes que las ecuaciones (10.34) y (10.35). Al utilizar s^2 en lugar de σ^2 en las fórmulas (10.27) y (10.32) —o en la (10.33)—, obtenemos estimadores insesgados de las variancias y covariancias de los estimadores minimocuadráticos de los coeficientes de regresión. Podemos construir *intervalos de confianza* para estos coeficientes de regresión observando que

$$\frac{\hat{\beta}_k - \beta_k}{s_{\hat{\beta}_k}} \sim t_{n-K} \quad (k = 1, 2, \dots, K),$$

en la que $s_{\hat{\beta}_k}$ representa el error standard estimado de $\hat{\beta}_k$.

EJEMPLO. Con el fin de ilustrar el uso de las fórmulas anteriores, construyamos los intervalos de confianza del 95 % correspondientes a los coeficientes del modelo de regresión que hemos presentado en el ejemplo anterior. Sustituyendo los resultados numéricos en la fórmula (10.34a), obtenemos

$$s^2 = \frac{1}{12-3} [5300 - (-1.326) \times (-3550) - 11.237 \times 125.25] \\ = \frac{185.27}{9} = 20.586.$$

Este resultado, junto a las fórmulas (10.24), (10.25) y (10.31), nos proporciona

$$s_{\hat{\beta}_2}^2 = \frac{20.586 \times 4.86}{2250 \times 4.86 - (-54)^2} = 0.012476,$$

$$s_{\hat{\beta}_3}^2 = \frac{20.586 \times 2250}{2250 \times 4.86 - (-54)^2} = 5.7761,$$

$$s_{\hat{\beta}_1}^2 = 20.586 \left[\frac{1}{12} + \frac{70^2 \times 4.86 + 6.7^2 \times 2250 - 2 \times 70 \times 6.7 \times (-54)}{2250 \times 4.86 - (-54)^2} \right] \\ = 452.1700.$$

Por tanto, los errores standard estimados de $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$ son

$$s_{\hat{\beta}_1} = 21.264,$$

$$s_{\hat{\beta}_2} = 0.112,$$

$$s_{\hat{\beta}_3} = 2.403.$$

En las tablas, el valor de t correspondiente a $12 - 3 = 9$ grados de libertad y una probabilidad de 0,025 en los dos extremos es 2,262. Por tanto, los intervalos de confianza del 95 % para los coeficientes de regresión son

$$117.532 - 2.262 \times 21.264 \leq \beta_1 \leq 117.532 + 2.262 \times 21.264$$

$$69.43 \leq \beta_1 \leq 165.63,$$

$$-1.326 - 2.262 \times 0.112 \leq \beta_2 \leq -1.326 + 2.262 \times 0.112$$

$$-1.58 \leq \beta_2 \leq -1.07,$$

$$11.237 - 2.262 \times 2.403 \leq \beta_3 \leq 11.237 + 2.262 \times 2.403$$

$$5.80 \leq \beta_3 \leq 16.67.$$

Intervalos de confianza para $E(Y_i)$

Los intervalos de confianza que acabamos de presentar indican la precisión de cada uno de los estimadores de los coeficientes de regresión por separado. La precisión de la ecuación de regresión estimada, en su conjunto, se indica por medio de la *zona de confianza*. Esta puede obtenerse como un intervalo de confianza para $E(Y_i)$, utilizando como estimador el valor ajustado $\hat{Y}_i$. Señalemos que $\hat{Y}_i$ se distribuye normalmente con media igual a $E(Y_i)$. En el caso de dos variables explicativas, su variancia es

$$\begin{aligned}\sigma_{\hat{Y}_i}^2 &= E[\hat{Y}_i - E(\hat{Y}_i)]^2 = E[\hat{Y}_i - E(Y_i)]^2 \\ &= E[(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)X_{i2} + (\hat{\beta}_3 - \beta_3)X_{i3}]^2.\end{aligned}$$

$$\text{Pero } \hat{\beta}_1 - \beta_1 = -(\hat{\beta}_2 - \beta_2)\bar{X}_2 - (\hat{\beta}_3 - \beta_3)\bar{X}_3 + \bar{e},$$

lo que proporciona

$$\begin{aligned}(10.36) \quad \sigma_{\hat{Y}_i}^2 &= E[(\hat{\beta}_2 - \beta_2)(X_{i2} - \bar{X}_2) + (\hat{\beta}_3 - \beta_3)(X_{i3} - \bar{X}_3) + \bar{e}]^2 \\ &= (X_{i2} - \bar{X}_2)^2 \text{Var}(\hat{\beta}_2) + (X_{i3} - \bar{X}_3)^2 \text{Var}(\hat{\beta}_3) \\ &\quad + 2(X_{i2} - \bar{X}_2)(X_{i3} - \bar{X}_3) \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) + \frac{\sigma^2}{n}.\end{aligned}$$

En el modelo con $(K-1)$ variables explicativas tenemos

$$\begin{aligned}(10.37) \quad \sigma_{\hat{Y}_i}^2 &= \sum_k (X_{ik} - \bar{X}_k)^2 \text{Var}(\hat{\beta}_k) \\ &\quad + 2 \sum_{j < k} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \text{Cov}(\hat{\beta}_j, \hat{\beta}_k) + \frac{\sigma^2}{n} \\ &\quad (j, k = 2, 3, \dots, K; j < k),\end{aligned}$$

En notación matricial,

$$\begin{aligned}(10.37a) \quad \sigma_{\hat{Y}_i}^2 &= (\underline{X}_i - \bar{\underline{X}})' E(\underline{\hat{\beta}} - \underline{\beta})(\underline{\hat{\beta}} - \underline{\beta})' (\underline{X}_i - \bar{\underline{X}}) + \frac{\sigma^2}{n} \\ &= \sigma^2 \left[(\underline{X}_i - \bar{\underline{X}})' (\underline{X}' \underline{X})^{-1} (\underline{X}_i - \bar{\underline{X}}) + \frac{1}{n} \right],\end{aligned}$$

en la que

$$(\underline{X}_i - \bar{\underline{X}}) = \begin{bmatrix} X_{i2} - \bar{X}_2 \\ X_{i3} - \bar{X}_3 \\ \vdots \\ X_{iK} - \bar{X}_K \end{bmatrix}$$

La expresión correspondiente a $\sigma_{\hat{Y}_i}^2$ incluye el parámetro σ^2 , cuyo valor se desconoce. Sustituyendo σ^2 por su estimador insesgado s^2 , obtendremos un estimador insesgado de $\sigma_{\hat{Y}_i}^2$ al que podemos llamar $s_{\hat{Y}_i}^2$. La zona de confianza correspondiente a $E(Y_i)$ se puede construir, pues, observando que

$$\frac{\hat{Y}_i - E(Y_i)}{s_{\hat{Y}_i}} \sim t_{n-K}.$$

A partir de ello podremos calcular el intervalo de confianza para $E(Y_i)$ correspondiente a *cualquier* conjunto de valores de las variables explicativas. Hay que señalar que este intervalo tendrá su punto más estrecho allí donde el valor de la variable explicativa sea igual a la media muestral (en cuyo caso $s_{\hat{Y}_i}$ será igual a $s/\sqrt{n}$), y que el intervalo se hará cada vez más amplio a medida que nos alejemos de la media muestral de las X en cualquier dirección.

Descomposición de la variación muestral de Y

Si definimos la variación muestral de Y como $\sum (Y_i - \bar{Y})^2$, al igual que en el caso de la regresión simple, tendremos

$$\begin{aligned}\sum_i (Y_i - \bar{Y})^2 &= \sum_i [(\hat{Y}_i + e_i) - \bar{Y}]^2 \\ &= \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i e_i^2 + 2 \sum_i (\hat{Y}_i - \bar{Y})e_i,\end{aligned}$$

Los términos e_i representan los residuos minimocuadráticos igual que en la regresión simple. A partir de las ecuaciones normales minimocuadráticas vemos que

$$\sum_i (\hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \dots + \hat{\beta}_K X_{iK})e_i - \bar{Y} \sum_i e_i = 0.$$

lo que indica que el 98 % de la variación muestral en la cantidad de naranjas vendidas puede atribuirse a los efectos estimados de la variación en el precio y la variación en el gasto en publicidad.

Contrastación de hipótesis

Por lo que respecta a la *contrastación de hipótesis*, el modelo de regresión múltiple ofrece mayor número de oportunidades que el modelo de regresión simple. En primer lugar, aparece la contrastación de la hipótesis de que el valor de β_k es igual a algún número específico. Es decir,

$$H_0: \beta_k = \gamma_k,$$

que debe contrastarse con respecto a una alternativa con uno o dos extremos. El estadístico que vamos a utilizar para contrastar H_0 es

$$(10.42) \quad \frac{\hat{\beta}_k - \gamma_k}{s_k} \sim t_{n-K} \quad (k = 2, 3, \dots, K).$$

Con mayor frecuencia nos interesa contrastar la hipótesis de que β_k es igual a cero. Cuando $k = 1$, esta hipótesis implica que la ordenada en el origen es igual a cero, o sea que el plano de regresión pasa por el origen. Cuando $k = 2, 3, \dots, K$, la hipótesis de que $\beta_k = 0$ significa que la variable X_{ik} no ejerce ninguna influencia sobre la media de Y_i ; si no se rechaza la hipótesis, llegamos a la conclusión de que X_{ik} no es una variable relevante en la ecuación de regresión.

Otra hipótesis de mayor amplitud es la de que *ninguna* de las variables explicativas ejerce influencia sobre la media de Y_i . En este caso, contrastamos

$$H_0: \beta_2 = \beta_3 = \dots = \beta_K = 0$$

contra la alternativa de que H_0 no es cierta; es decir, de que por lo menos una de las pendientes de regresión es distinta de cero. Si H_0 es cierta, la variación de Y de una observación a otra no está afectada por los cambios en ninguna de las variables explicativas, sino que es puramente aleatoria. Pero, de ser así, el valor observado de SCR («suma de los cuadrados debidos a la regresión») sólo es distinto de cero a causa del muestreo. Según (10.39),

$$SCR = \sum_k \hat{\beta}_k^2 m_{kk} + 2 \sum_{j < k} \hat{\beta}_j \hat{\beta}_k m_{jk} \quad (j, k = 2, 3, \dots, K; j < k),$$

es evidente que si substituyésemos las $\hat{\beta}$ por β y la hipótesis nula fuese cierta, SCR sería igual a cero y, por tanto, SCE sería igual a STC. Ahora puede mostrarse¹ que si la hipótesis nula es cierta,

$$(10.43) \quad \frac{SCR/(K-1)}{SCE/(n-K)} \sim F_{K-1, n-K}$$

en la que $F_{K-1, n-K}$ representa la distribución F con $(K-1)$ y $(n-K)$ grados de libertad. Si el valor de la expresión (10.43) no es significativamente distinto de cero, la muestra no ofrece ninguna evidencia de que las variables explicativas tengan algún efecto en la media de Y_i . Cabe señalar que el valor del estadístico F en la expresión (10.43) se puede calcular utilizando R^2 , ya que

$$\frac{SCR/(K-1)}{SCE/(n-K)} = \frac{[n-K]}{[K-1]} \left[\frac{SCR/STC}{1 - (SCR/STC)} \right] = \frac{[n-K]}{[K-1]} \left[\frac{R^2}{1 - R^2} \right].$$

Obsérvese que, si algunos de los coeficientes de regresión estimados $\hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_K$ es significativamente distinto de cero según la contrastación basada en la distribución t de (10.42), el valor del estadístico F en la expresión (10.43) será también significativamente distinto de cero, siempre que las contrastaciones se efectúen al mismo nivel de significación y utilizando la misma alternativa.² Sin embargo, es perfectamente posible encontrar que *ninguna* de las $\hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_K$ es significativamente distinta de cero según la contrastación de la t , y al mismo tiempo, rechazar la hipótesis de que $\beta_2 = \beta_3 = \dots = \beta_K = 0$ según la contrastación basada en la F . Esto puede suceder en los casos en que las variables explicativas están muy correlacionadas entre sí. En estas condiciones, las influencias separadas de cada una de las variables explicativas sobre la variable dependiente pueden ser débiles, mientras que su influencia conjunta puede ser muy fuerte.

La relación entre los valores de los estadísticos t en la contrastación de

$$H_0: \beta_k = 0 \quad (k = 2, 3, \dots, K)$$

1. Ver, por ejemplo, J. Johnston, *Econometric Methods*. Nueva York, McGraw-Hill, 1965, pp. 119-22.

2. En sentido estricto, esto no es del todo cierto. R. C. Geary y C. E. V. Leser en «Significance Tests in Multiple Regression», *The American Statistician*, vol. 22, febrero 1968, pp. 20-21, proporcionan un ejemplo en el que todos los valores de las t son significativos, pero el valor de la F no lo es. Sin embargo, los autores señalan que esta situación es muy poco frecuente y que «la posibilidad de su existencia constituye poco más que una curiosidad».

y el valor del estadístico F en la contrastación de

$$H_0: \beta_2 = \beta_3 = \dots = \beta_K$$

se puede poner en forma explícita expresando el estadístico F en términos de las t . Estableceremos esta relación para el caso en que existen dos variables explicativas. Según (10.38) tenemos

$$SCR = \beta_2^2 m_{22} + \beta_3^2 m_{33} + 2\beta_2 \beta_3 m_{23},$$

y según (10.34) sabemos que

$$\frac{SCE}{(n-3)} = s^2.$$

Por tanto,

$$F_{2, n-3} = \frac{SCR/2}{SCE/(n-3)} = \frac{\beta_2^2 m_{22} + \beta_3^2 m_{33} + 2\beta_2 \beta_3 m_{23}}{2s^2}.$$

Representemos ahora los valores del estadístico t correspondientes a β_2 y β_3 , por t_2 y t_3 , respectivamente. En este caso,

$$t_2 = \frac{\hat{\beta}_2}{s_{\hat{\beta}_2}} \quad \text{y} \quad t_3 = \frac{\hat{\beta}_3}{s_{\hat{\beta}_3}}.$$

Basándonos en (10.24) y (10.25), podemos escribir

$$t_2^2 = \frac{\hat{\beta}_2^2 (m_{22} m_{33} - m_{23}^2)}{s^2 m_{33}},$$

$$t_3^2 = \frac{\hat{\beta}_3^2 (m_{22} m_{33} - m_{23}^2)}{s^2 m_{22}},$$

o

$$\hat{\beta}_2^2 = \frac{t_2^2 s^2 m_{33}}{m_{22} m_{33} - m_{23}^2},$$

$$\hat{\beta}_3^2 = \frac{t_3^2 s^2 m_{22}}{m_{22} m_{33} - m_{23}^2},$$

y también

$$\hat{\beta}_2 \hat{\beta}_3 = \frac{t_2 t_3 s^2 \sqrt{m_{22}} \sqrt{m_{33}}}{m_{22} m_{33} - m_{23}^2}.$$

Sustituyendo en la ecuación (10.44) tenemos

$$(10.44a) \quad F_{2, n-3} = \frac{t_2^2 s^2 m_{22} m_{33} + t_3^2 s^2 m_{22} m_{33} + 2 t_2 t_3 s^2 (\sqrt{m_{22}} \sqrt{m_{33}}) m_{23}}{2 s^2 (m_{22} m_{33} - m_{23}^2)}.$$

Si llamamos al coeficiente de correlación entre X_2 y X_3 en la muestra r_{23} , es decir, si definimos

$$r_{23} = \frac{m_{23}}{\sqrt{m_{22}} \sqrt{m_{33}}},$$

la ecuación (10.44a) se reduce a

$$(10.44b) \quad F_{2, n-3} = \frac{t_2^2 + t_3^2 + 2 t_2 t_3 r_{23}}{2(1 - r_{23}^2)}.$$

Esto muestra claramente que si r_{23} no es muy distinto de la unidad, el valor de $F_{2, n-3}$ puede ser muy grande, aunque tanto t_2 como t_3 sean pequeñas. Por tanto, $\hat{\beta}_2$ y $\hat{\beta}_3$ pueden no ser significativamente distintas de cero y, sin embargo, el valor de $F_{2, n-3}$ puede ser muy significativo.³

EJEMPLO. Consideremos el modelo de regresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + e_i.$$

Supongamos que $n = 20$, y los momentos muestrales son los siguientes

$$m_{YY} = 100, \quad m_{22} = 100,$$

$$m_{Y2} = 90, \quad m_{33} = 100,$$

$$m_{Y3} = 90, \quad m_{23} = 95.$$

En este caso,

$$\hat{\beta}_2 = \frac{90 \times 100 - 90 \times 95}{100 \times 100 - 95^2} = 0.461,$$

y

$$\hat{\beta}_3 = \frac{90 \times 100 - 90 \times 95}{100 \times 100 - 95^2} = 0.461.$$

Además,

$$SCR = 0.461^2 \times 100 + 0.461^2 \times 100 + 2 \times 0.461 \times 0.461 \times 95 = 83,$$

$$STC = 100,$$

$$SCE = 17.$$

3. Para una exploración sistemática de este punto, véase K. Fox, *Intermediate Economic Statistics*. Nueva York, Wiley, 1968, pp. 259-65. El ejemplo presentado en el texto constituye una pequeña modificación de una de las múltiples ilustraciones indicadas por Fox.

Por tanto,

$$s^2 = \frac{17}{17} = 1,$$

y

$$s_{\beta_2}^2 = \frac{1 \times 100}{100 \times 100 - 95^2} = 0.102564,$$

$$s_{\beta_3}^2 = \frac{1 \times 100}{100 \times 100 - 95^2} = 0.102564.$$

En consecuencia,

$$t_2 = \frac{0.461}{\sqrt{0.102564}} = 1.444,$$

$$t_3 = \frac{0.461}{\sqrt{0.102564}} = 1.444.$$

Como en las tablas el valor de la t correspondiente a 17 grados de libertad al nivel de significación del 5% (o sea, $t_{17, 0.025}$) es 2.110, y el valor calculado (1.444) es inferior, ni β_2 , ni β_3 son significativamente distintas de cero al nivel de significación del 5%. Por otra parte, el valor

$$F = \frac{83/2}{1} = 41.5$$

es mucho mayor que el valor tabulado de la $F_{2, 17}$ al nivel de significación del 5%, que es 3.59. Por tanto, en base a la contrastación por medio de la t no podemos rechazar la hipótesis de que $\beta_2 = 0$ o $\beta_3 = 0$, pero por la contrastación basada en la F rechazamos la hipótesis de que $\beta_2 = \beta_3 = 0$. La razón de ello es el hecho de que las contribuciones de X_{i2} y X_{i3} por separado a la explicación de la variación de Y_i son débiles, pero su contribución conjunta, que no se puede descomponer, es muy fuerte. En realidad,

$$\beta_2 m_{22} = 21.3$$

$$\beta_3 m_{33} = 21.3$$

$$2\beta_2 \beta_3 m_{23} = 40.4$$

$$SCR = 83.0$$

lo que indica que la contribución de X_{i2} y X_{i3} a la SCR es casi el doble de la contribución de X_{i2} o X_{i3} por separado. Obsérvese que, en este ejemplo, $r_{23} = 0.95$, lo que señala un elevado grado de correlación muestral entre las dos variables explicativas.

Existe también una contrastación algo distinta que se refiere a la influencia de otras variables explicativas adicionales sobre la media de Y_i . Concretamente, consideremos dos teorías, una de las cuales afirma que la ecuación de regresión es

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{iK} + \varepsilon_i,$$

La otra afirma que Y_i no depende solamente de $X_{i2}, X_{i3}, \dots$, también de las variables explicativas adicionales $X_{i, K+1}, X_{i, K+2}, \dots, X_{i, (K+Q)} = X_{iQ}$, es decir

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{iK} + \beta_{K+1} X_{i, K+1} + \dots + \beta_Q X_{iQ} + \varepsilon_i.$$

En tal caso, la segunda teoría se puede contrastar por medio de la hipótesis

$$H_0: \beta_{K+1} = \beta_{K+2} = \dots = \beta_Q = 0$$

respecto a la alternativa de que H_0 no es cierta. Con el fin de formular la contrastación adecuada, introduciremos una nueva notación utilizando el subíndice K para indicar los valores correspondientes al conjunto original de variables explicativas, y el subíndice Q para representar los valores correspondientes al nuevo conjunto, ampliado, de variables explicativas. Los valores sin ningún subíndice son válidos para cualquiera de los dos conjuntos. Podemos escribir

$$STC = SCR_K + SCE_K$$

y también

$$STC = SCR_Q + SCE_Q.$$

Si las variables explicativas adicionales no son relevantes para la explicación de la variable de Y_i , SCR_K y SCR_Q tendrán el mismo valor en la población, y la diferencia que observemos entre ellas se deberá por completo a los errores muestrales. Si la hipótesis nula es cierta, se puede mostrar⁴ que

$$(10.45) \quad \frac{(SCR_Q - SCR_K)/(Q - K)}{SCE_Q/(n - Q)} \sim F_{Q-K, n-Q}.$$

Podemos utilizar esta expresión para contrastar H_0 a un nivel de significación concreto, y se puede mostrar fácilmente que

$$SCR_Q \geq SCR_K$$

de modo que la expresión (10.45) no puede ser nunca negativa. Una implicación de este resultado es que

$$\frac{SCR_Q}{STC} \geq \frac{SCR_K}{STC};$$

⁴ Ver, por ejemplo, A. S. Goldberger, *Econometric Theory*. Nueva York, Wiley, 1964, p. 475.

o sea, que

$$R^2_Q \geq R^2_K.$$

Esto significa que cuando añadimos nuevas variables explicativas a una ecuación de regresión, el resultado no puede ser nunca una disminución en el valor del coeficiente de determinación. Sin embargo, esto no es necesariamente cierto para $\bar{R}^2$.

EJEMPLO. En un artículo sobre las funciones de consumo a corto plazo en los Estados Unidos,⁵ se presentaron, entre otras, las dos funciones siguientes

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$$

$$y \quad C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + \beta_4 L_{t-1} + \varepsilon_t,$$

siendo C = consumo, Y = renta, y L = activos líquidos. Contrastemos la hipótesis de que

$$H_0: \beta_3 = \beta_4 = 0$$

respecto a la alternativa de que H_0 no es cierta al nivel de significación del 1 % por ejemplo. Para utilizar la contrastación por medio de la F , de (10.45), tenemos que calcular

$$F = \left[\frac{SCR_Q - SCR_K}{SCE_Q} \right] \left[\frac{n-Q}{Q-K} \right].$$

Sabemos que $n = 31$, $Q = 4$, y $K = 2$. En el artículo no aparecen los valores de SCR y SCE , pero los podemos obtener a partir de los valores de los $\bar{R}^2$, que son

$$\bar{R}^2_Q = 0.984 \quad y \quad \bar{R}^2_K = 0.944.$$

En primer lugar, vamos a obtener los valores de los R^2 «no corregidos». Como

$$\bar{R}^2_K = R^2_K - \frac{K-1}{n-K} (1 - R^2_K),$$

podemos despejar R^2_K , obteniendo

$$R^2_K = \frac{(n-K)\bar{R}^2_K + (K-1)}{n-1} = 0.946$$

y, del mismo modo,

$$R^2_Q = 0.986.$$

5. Arnold Zellner, «The Short-Run Consumption Function», *Econometrica*, vol. 25, octubre 1957, pp. 552-67.

del correspondiente estadístico F es, pues,

$$F = \frac{SCR_Q - SCR_K}{1 - (SCR_Q/STC)} \left[\frac{n-Q}{Q-K} \right] = \frac{R^2_Q - R^2_K}{1 - R^2_Q} \left[\frac{n-Q}{Q-K} \right] = 38.57.$$

En las tablas, el valor de $F_{2,27}$ al nivel de significación del 1 %, es 5.49 y, por tanto, tenemos que rechazar la hipótesis nula. Es decir, la evidencia sugiere con firmeza que la incorporación del consumo y los activos líquidos con un retardo de un período contribuye a la explicación de las variaciones en el consumo actual.

Otra hipótesis en la que estamos interesados en algunas ocasiones es

$$H_0: \beta_j = \beta_k \quad (j \neq k),$$

y la podemos contrastar respecto a una alternativa con uno o dos extremos. Por ejemplo, una de las formas que se han propuesto para representar la función agregada de consumo es

$$C_t = \beta_1 + \beta_2 W_t + \beta_3 P_t + \beta_4 C_{t-1} + \varepsilon_t,$$

siendo C = consumo, W = renta salarial, y P = renta no salarial. Una hipótesis interesante en relación con esta formulación es la de que la propensión marginal al consumo de los perceptores de salarios es igual a la de los que perciben otros ingresos, es decir, que $\beta_2 = \beta_3$. Para contrastar la hipótesis de que los dos coeficientes de regresión son iguales, estudiaremos la distribución de la diferencia entre los dos estimadores minimocuadráticos correspondientes. En general, tenemos

$$E(\hat{\beta}_j - \hat{\beta}_k) = \beta_j - \beta_k$$

$$y \quad \text{Var}(\hat{\beta}_j - \hat{\beta}_k) = \text{Var}(\hat{\beta}_j) + \text{Var}(\hat{\beta}_k) - 2 \text{Cov}(\hat{\beta}_j, \hat{\beta}_k).$$

Por tanto, si la hipótesis nula es cierta,

$$(\hat{\beta}_j - \hat{\beta}_k) \sim N(0, \sigma^2_{\hat{\beta}_j - \hat{\beta}_k}).$$

Podemos obtener un estimador insesgado de $\sigma^2_{\hat{\beta}_j - \hat{\beta}_k}$, al que llamaremos $s^2_{\hat{\beta}_j - \hat{\beta}_k}$ utilizando s^2 como estimador de σ^2 . De esto se deriva que

$$(10.46) \quad \frac{\hat{\beta}_j - \hat{\beta}_k}{s_{\hat{\beta}_j - \hat{\beta}_k}} \sim t_{n-K}.$$

Si el valor de este estadístico es significativamente distinto de cero, rechazamos la hipótesis de que los dos coeficientes de regresión son iguales. Goldberger⁶ presenta una contrastación de la hipótesis de que más de dos coeficientes de regresión son iguales.

Goldberger, *op. cit.*, p. 175, teorema (7.19).

Otra hipótesis similar a la del caso anterior es la proposición de que la suma de dos coeficientes de regresión tiene un valor determinado. Por ejemplo, en la función de producción de Cobb-Douglas

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

en la que $Y_i = \log$ producto, $X_{i2} = \log$ factor trabajo, y $X_{i3} = \log$ factor capital, la hipótesis de que existen rendimientos constantes de escala equivale a la hipótesis

$$H_0: \beta_2 + \beta_3 = 1.$$

En general, podemos contrastar la hipótesis

$$H_0: \beta_j + \beta_k = a$$

teniendo en cuenta que

$$(10.47) \quad \left[\frac{\hat{\beta}_j + \hat{\beta}_k - a}{s_{\hat{\beta}_j + \hat{\beta}_k}} \right] \sim t_{n-k},$$

en la que

$$s_{\hat{\beta}_j + \hat{\beta}_k} = \sqrt{s_{\hat{\beta}_j}^2 + s_{\hat{\beta}_k}^2 + 2 \text{Est. Cov}(\hat{\beta}_j, \hat{\beta}_k)}.$$

Esta contrastación se puede ampliar fácilmente a la suma de más de dos coeficientes de regresión.

La última contrastación estudia la igualdad de dos ecuaciones de regresión. Concretamente, consideremos la ecuación de regresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{iK} + \varepsilon_i,$$

estimada a partir de una muestra de n observaciones. Supongamos ahora que obtenemos $m(>K)$ observaciones adicionales y queremos contrastar la hipótesis de que dichas observaciones proceden de la misma población que las primeras n observaciones. Por ejemplo, los datos correspondientes a la función agregada de consumo abarcan frecuentemente el período anterior a la Segunda Guerra Mundial así como el posterior. Si admitimos la posibilidad de que los parámetros de la función de consumo del segundo período sean distintos de los de la función en el primer período, contrastaríamos la hipótesis nula de que los parámetros de dicha función de consumo *no* han variado. Concretamente, podemos escribir

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{iK} + \varepsilon_i \quad (i = 1, 2, \dots, n)$$

y

$$Y_i = \gamma_1 + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \dots + \gamma_K X_{iK} + \varepsilon_i \quad (i = n+1, n+2, \dots, n+m).$$

La hipótesis nula es

$$H_0: \beta_1 = \gamma_1, \beta_2 = \gamma_2, \dots, \beta_K = \gamma_K.$$

La tenemos que contrastar con la hipótesis de que H_0 no es cierta. La obtención de la contrastación adecuada en este caso ha sido expuesta por otros autores.⁸ El correspondiente estadístico de prueba se obtiene aplicando el método de estimación por mínimos cuadrados al primer conjunto de datos ($i = 1, 2, \dots, n$), al segundo conjunto de datos ($i = n+1, n+2, \dots, n+m$), y a los dos grupos de datos combinados ($i = 1, 2, \dots, n+m$). Para estimar los coeficientes de regresión a partir de la combinación de los dos grupos de datos, escribiremos la ecuación de regresión en la forma

$$Y_i = \delta_1 + \delta_2 X_{i2} + \delta_3 X_{i3} + \dots + \delta_K X_{iK} + \varepsilon_i \quad (i = 1, 2, \dots, n+m).$$

Además, representaremos las sumas de los cuadrados de los residuos mínimo-cuadráticos en la forma siguiente:

$$SCE_1 = \sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{i2} - \hat{\beta}_3 X_{i3} - \dots - \hat{\beta}_K X_{iK})^2,$$

$$SCE_2 = \sum_{i=n+1}^{n+m} (Y_i - \hat{\gamma}_1 - \hat{\gamma}_2 X_{i2} - \hat{\gamma}_3 X_{i3} - \dots - \hat{\gamma}_K X_{iK})^2,$$

$$y \quad SCE_C = \sum_{i=1}^{n+m} (Y_i - \hat{\delta}_1 - \hat{\delta}_2 X_{i2} - \hat{\delta}_3 X_{i3} - \dots - \hat{\delta}_K X_{iK})^2.$$

Se ha mostrado que

$$(10.48) \quad \frac{(SCE_C - SCE_1 - SCE_2)/K}{(SCE_1 + SCE_2)/(n+m-2K)} \sim F_{K, n+m-2K}.$$

Podemos utilizar esta expresión para contrastar H_0 , pero deben señalarse dos puntos:

1. El estadístico de prueba de la expresión (10.48) sólo es aplicable en el caso en que el número de observaciones adicionales es mayor que el número de variables explicativas *más una*; es decir, si $m > K$. En el caso en que $m < K$, debemos utilizar una contrastación algo distinta.⁸

⁸ Ver, por ejemplo, Johnston, *op. cit.*, pp. 136-38.
Ídem, p. 158, ecuación (4.65).

2. La formulación de la hipótesis que nos conduce a la expresión (10.48) admite la posibilidad de que, al pasar de un grupo de datos a otro, varíen *todos* los parámetros de la regresión. En la sección 11.1, que se ocupa de las variables binarias, presentaremos algunas contrastaciones de hipótesis relativas solamente a *algunos* coeficientes de regresión.

Predicción

Tras haber estudiado algunas de las contrastaciones más útiles que se pueden efectuar a través del modelo de regresión múltiple, pasamos a considerar el problema de *pronosticar* el valor de la variable dependiente correspondiente a un determinado conjunto de valores de las variables explicativas. Más formalmente, sean los valores de las variables explicativas $X_{02}, X_{03}, \dots, X_{0K}$, y el correspondiente valor de la variable dependiente Y_0 . Queremos pronosticar Y_0 , y como se ha señalado al ocuparnos del modelo de regresión simple, el mejor predictor de Y_0 es $E(Y_0)$, ya que la variancia de Y_0 es mínima en este punto. Como no conocemos $E(Y_0)$, utilizaremos en su lugar el valor de Y_0 ajustado por los mínimos cuadrados, o sea $\hat{Y}_0$. Sabemos que

$$\hat{Y}_0 = \beta_1 + \beta_2 X_{02} + \beta_3 X_{03} + \dots + \beta_K X_{0K},$$

y, por tanto, se deriva que $\hat{Y}_0$ se distribuye normalmente con media

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_{02} + \beta_3 X_{03} + \dots + \beta_K X_{0K}.$$

Según la expresión (10.37), la variancia de $\hat{Y}_0$ es

$$\begin{aligned} \text{Var}(\hat{Y}_0) &= \sigma_{\hat{Y}_0}^2 \\ &= \sum_k (X_{0k} - \bar{X}_k)^2 \text{Var}(\beta_k) + 2 \sum_{j < k} (X_{0j} - \bar{X}_j)(X_{0k} - \bar{X}_k) \text{Cov}(\beta_j, \beta_k) \\ &\quad + \frac{\sigma^2}{n} \quad (j, k = 2, 3, \dots, K; j < k), \end{aligned}$$

o, en notación matricial,

$$\text{Var}(\hat{Y}_0) = \sigma^2 \left[(\underline{X}_0 - \underline{\bar{X}})' (\underline{X}' \underline{X})^{-1} (\underline{X}_0 - \underline{\bar{X}}) + \frac{1}{n} \right],$$

en la que

$$(\underline{X}_0 - \underline{\bar{X}}) = \begin{bmatrix} X_{02} - \bar{X}_2 \\ X_{03} - \bar{X}_3 \\ \vdots \\ X_{0K} - \bar{X}_K \end{bmatrix}$$

que nos interesa fundamentalmente es el *error del pronóstico*, $Y_0 - \hat{Y}_0$. Esta variable aleatoria se distribuye normalmente con

$$E(Y_0 - \hat{Y}_0) = 0$$

variancia

$$\sigma_F^2 = \text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(Y_0) + \text{Var}(\hat{Y}_0) - 2 \text{Cov}(Y_0, \hat{Y}_0).$$

Pero sabemos que

$$\text{Var}(Y_0) = \sigma^2,$$

$$\text{Var}(\hat{Y}_0) = \sigma_{\hat{Y}_0}^2,$$

y

$$\begin{aligned} -2 \text{Cov}(Y_0, \hat{Y}_0) &= -2 E[(Y_0 - E(Y_0))(\hat{Y}_0 - E(\hat{Y}_0))] \\ &= -2 E \epsilon_0 [\hat{Y}_0 - E(Y_0)] = 0. \end{aligned}$$

Por tanto,

$$\begin{aligned} (10.49) \quad \sigma_F^2 &= \sigma^2 + \frac{\sigma^2}{n} + \sum_k (X_{0k} - \bar{X}_k)^2 \text{Var}(\beta_k) \\ &\quad + 2 \sum_{j < k} (X_{0j} - \bar{X}_j)(X_{0k} - \bar{X}_k) \text{Cov}(\beta_j, \beta_k), \end{aligned}$$

o, en notación matricial,

$$(10.49a) \quad \sigma_F^2 = \sigma^2 \left[1 + \frac{1}{n} + (\underline{X}_0 - \underline{\bar{X}})' (\underline{X}' \underline{X})^{-1} (\underline{X}_0 - \underline{\bar{X}}) \right].$$

Al igual que en el caso del modelo de regresión simple, cuanto menor sea la distancia entre los valores dados de las variables explicativas y sus medias muestrales respectivas, menor será la variancia del error del pronóstico. Podemos obtener un estimador insesgado de σ_F^2 sustituyendo σ^2 por s^2 . Si llamamos s_F^2 al estimador resultante, tenemos

$$(10.50) \quad \frac{Y_0 - \hat{Y}_0}{s_F} \sim t_{n-K}.$$

A partir de este resultado, podemos construir un intervalo para el pronóstico, con cualquier probabilidad que deseemos de que contenga el verdadero valor

Si llamamos λ ($0 < \lambda < 1$) al valor uno menos el nivel de probabilidad elegido, tendremos

$$\hat{Y}_0 - t_{n-K, \lambda/2} s_F \leq Y_0 \leq \hat{Y}_0 + t_{n-K, \lambda/2} s_F.$$

La expresión (10.50) puede utilizarse también para contrastar la hipótesis de que una nueva observación, por ejemplo la $(n+1)$, procede de la misma población que las n observaciones utilizadas para estimar los parámetros de la regresión.

EJEMPLO. En el primer ejemplo de esta sección, se ha estimado una ecuación de regresión que expresa la demanda de naranjas. La muestra consistía en doce observaciones relativas a la cantidad de naranjas vendidas, su precio, y la cantidad gastada en publicidad (ver cuadro 10-1). La ecuación de regresión estimada es

$$\hat{Y}_i = 117.532 - 1.326X_{i2} + 11.237X_{i3},$$

en la que Y_i = cantidad de naranjas vendidas, X_{i2} = precio, y X_{i3} = gasto en publicidad. Supongamos ahora que no se venden naranjas hasta que llega una nueva partida unas semanas más tarde. Entonces, los resultados para el primer día de venta son

$$Y_0 = 100, \quad X_{02} = 80, \quad \text{y} \quad X_{03} = 7.$$

El problema consiste en descubrir si la función de demanda ha variado desde que se terminó la partida anterior de naranjas. Aquí tenemos

$$\hat{Y}_0 = 117.532 - 1.326 \times 80 + 11.237 \times 7 = 90.111.$$

Este es el «pronóstico» de Y . La variancia estimada del error del pronóstico es

$$s_p^2 = \left[s^2 + \frac{s^2}{n} + (X_{02} - \bar{X}_2)^2 s_{\beta_2}^2 + (X_{03} - \bar{X}_3)^2 s_{\beta_3}^2 + 2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3) \text{Est. Cov}(\beta_2, \beta_3) \right].$$

Ahora bien, según los cálculos anteriores, sabemos que

$$\begin{aligned} s^2 &= 20.586, & s_{\beta_2}^2 &= 0.012476, \\ \bar{X}_2 &= 70, & s_{\beta_3}^2 &= 5.7761, \\ \bar{X}_3 &= 6.7, & \text{Est. Cov}(\beta_2, \beta_3) &= 0.138628. \end{aligned}$$

Y, por tanto,

$$\begin{aligned} s_p^2 &= 20.586 \left[1 + \frac{1}{12} + (80 - 70)^2 \times 0.012476 + (7 - 6.7)^2 \times 5.7761 \right. \\ &\quad \left. + 2 \times (80 - 70) \times (7 - 6.7) \times 0.138628 \right] \\ &= 75.808974, \\ s_p &= 8.706. \end{aligned}$$

El intervalo de confianza del 95% para Y_0 sabiendo que el valor observado de Y_0 es 100 y, por tanto, se halla en el interior del intervalo del 95%.

$$90.111 - 2.262 \times 8.706 \leq Y_0 \leq 90.111 + 2.262 \times 8.706$$

$$70.418 \leq Y_0 \leq 109.804.$$

El valor observado de Y_0 es 100 y, por tanto, se halla en el interior del intervalo del 95%.

Cambio en las unidades de medida

Analicemos ahora los efectos de un *cambio en las unidades de medida* sobre los resultados que hemos obtenido anteriormente. Sean

$$Y_i, X_{i2}, X_{i3}, \dots, X_{ik}$$

los valores de las respectivas variables medidas en términos de las unidades «originarias», y

$$Y_i^*, X_{i2}^*, X_{i3}^*, \dots, X_{ik}^*$$

los valores de estas variables medidos en términos de unas «nuevas» unidades. En primer lugar, estudiaremos el caso

$$Y_i^* = d_Y Y_i, \quad X_{i2}^* = d_2 X_{i2}, \quad X_{i3}^* = d_3 X_{i3}, \quad \vdots \quad X_{ik}^* = d_k X_{ik},$$

en el que las d son cantidades conocidas positivas. (Obsérvese que cuando $d = 1$ para algunas variable concreta, no se produce ningún cambio en la unidad de medida.) Este caso de cambios multiplicativos en las unidades de medida abarca cambios tales como el paso de centavos a dólares ($d = 0.01$) o la expresión de una variable en términos de un índice simple (por ejemplo, $d_k = 1/X_{ik}$). Las medias y los momentos muestrales de segundo orden de las variables con «asterisco» son

$$\bar{Y}^* = d_Y \bar{Y},$$

$$\bar{X}_k^* = d_k \bar{X}_k,$$

$$m_{jk}^* = \sum_i (d_j X_{ij} - d_j \bar{X}_j)(d_k X_{ik} - d_k \bar{X}_k) = d_j d_k m_{jk},$$

$$m_{Yk}^* = d_Y d_k m_{Yk},$$

$$m_{Yk}^* = d_Y^2 m_{Yk} \quad (j, k = 2, 3, \dots, K).$$

Veamos ahora cómo se ven afectados los valores de los coeficientes estimados y de sus errores standard a consecuencia de estos cambios. En el modelo de regresión con dos variables explicativas, según las ecuaciones (10.11), (10.12), y (10.13), tenemos

$$\beta_2^* = \frac{(d_Y d_2 m_{Y2})(d_3^2 m_{33}) - (d_Y d_3 m_{Y3})(d_2 d_3 m_{23})}{(d_2^2 m_{22})(d_3^2 m_{33}) - (d_2 d_3 m_{23})^2} = \left(\frac{d_Y}{d_2}\right) \beta_2,$$

$$\beta_3^* = \frac{(d_Y d_3 m_{Y3})(d_2^2 m_{22}) - (d_Y d_2 m_{Y2})(d_2 d_3 m_{23})}{(d_2^2 m_{22})(d_3^2 m_{33}) - (d_2 d_3 m_{23})^2} = \left(\frac{d_Y}{d_3}\right) \beta_3,$$

$$y \quad \beta_1^* = d_Y \bar{Y} - \beta_2^*(d_2 \bar{X}_2) - \beta_3^*(d_3 \bar{X}_3) = d_Y \beta_1.$$

Para el estimador de σ^2 tenemos, según la ecuación (10.34a),

$$s^{*2} = \frac{1}{n-K} (m_{YY}^* - \beta_2^* m_{Y2}^* - \beta_3^* m_{Y3}^*) = d_Y^2 s^2.$$

Por tanto, en base a las ecuaciones (10.24) y (10.25),

$$s_{\beta_2}^{*2} = \frac{(d_Y^2 s^2)(d_3^2 m_{33})}{(d_2^2 m_{22})(d_3^2 m_{33}) - (d_2 d_3 m_{23})^2} = \left(\frac{d_Y}{d_2}\right)^2 s_{\beta_2}^2,$$

$$s_{\beta_3}^{*2} = \frac{(d_Y^2 s^2)(d_2^2 m_{22})}{(d_2^2 m_{22})(d_3^2 m_{33}) - (d_2 d_3 m_{23})^2} = \left(\frac{d_Y}{d_3}\right)^2 s_{\beta_3}^2,$$

y, según (10.31),

$$s_{\beta_1}^{*2} = (d_2 \bar{X}_2)^2 \left(\frac{d_Y}{d_2}\right)^2 s_{\beta_2}^2 + (d_3 \bar{X}_3)^2 \left(\frac{d_Y}{d_3}\right)^2 s_{\beta_3}^2 + 2(d_2 \bar{X}_2)(d_3 \bar{X}_3) \left[\frac{(d_Y^2 s^2)(d_2 d_3 m_{23})}{(d_2^2 m_{22})(d_3^2 m_{33}) - (d_2 d_3 m_{23})^2} \right] + \frac{d_Y^2 s^2}{n} = d_Y^2 s_{\beta_1}^2.$$

Esto significa que los cambios en las unidades de medida afectan a los valores de los coeficientes de regresión estimados y de sus errores standard. Sin embargo, estos cambios no alteran el valor del estadístico t , que se utiliza en la contrastación de la hipótesis de que $\beta_k = 0$ ($k = 1, 2, \dots, K$):

$$\frac{\beta_k^*}{s_{\beta_k}^*} = \frac{(d_Y/d_k) \beta_k}{(d_Y/d_k) s_{\beta_k}} = \frac{\beta_k}{s_{\beta_k}}.$$

La ecuación se convierte en relación con el estadístico F :

$$\begin{aligned} F^* &= \frac{(K-1)}{(n-K)} = \frac{[n-K]}{[K-1]} \left[\frac{\text{STC}^* - \text{SCE}^*}{\text{SCE}^*} \right] = \frac{[n-K]}{[K-1]} \left[\frac{m_{YY}^* - (n-K)s^{*2}}{(n-K)s^{*2}} \right] \\ &= \frac{[n-K]}{[K-1]} \left[\frac{(d_Y^2 m_{YY}) - (n-K)(d_Y^2 s^2)}{(n-K)(d_Y^2 s^2)} \right] = \frac{\text{SCR}/(K-1)}{\text{SCE}/(n-K)}. \end{aligned}$$

Por último, el valor de R^2 también permanece inalterado:

$$R^{*2} = 1 - \frac{\text{SCE}^*}{\text{STC}^*} = 1 - \frac{(n-K)d_Y^2 s^2}{d_Y^2 m_{YY}} = R^2.$$

Puede demostrarse que estos resultados se aplican también a las ecuaciones de regresión con *cualquier* número de variables explicativas.

EJEMPLO. En un ejemplo anterior hemos estimado la ecuación de la demanda de naranjas en la forma

$$Y_i = 117.532 - 1.326 X_{i2} + 11.237 X_{i3} + e_i, \quad R^2 = 0.981, \\ (21.264) \quad (0.112) \quad (2.403)$$

en la que Y_i = cantidad en libras, X_{i2} = precio en centavos, y X_{i3} = gastos en publicidad (en dólares). Supongamos que queremos medir el precio (X_{i2}) en dólares en lugar de centavos. En este caso tenemos

$$d_Y = 1, \quad d_2 = 0.01, \quad y \quad d_3 = 1.$$

La ecuación de regresión estimada se convierte en

$$Y_i = 117.532 - 132.6 X_{i2}^* + 11.237 X_{i3} + e_i, \quad R^2 = 0.981. \\ (21.264) \quad (11.2) \quad (2.403)$$

En ocasiones, el cambio en las unidades de medida es de tipo aditivo en lugar de multiplicativo. Esto sucede con frecuencia cuando las variables aparecen en la ecuación de regresión en forma logarítmica. Por ejemplo, en el caso de la función de producción de Cobb-Douglas, el modelo tradicional es

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + e_i.$$

en la que $Y_i = \log$ de la cantidad producida, $X_{i2} = \log$ de la cantidad de trabajo, y $X_{i3} = \log$ de la cantidad de capital. En tal caso, si cambiamos la unidad de medida de la cantidad de trabajo, pasando de miles de trabajadores a millones, tendremos

$$X_{i2}^* = \log L_i^* = \log d_L L_i = \log d_L + \log L_i = (\log d_L) + X_{i2}.$$

En términos generales, sean

$$Y_i^* = g_Y + Y_i,$$

$$X_{i2}^* = g_2 + X_{i2},$$

$$X_{i3}^* = g_3 + X_{i3}.$$

⋮

$$X_{iK}^* = g_K + X_{iK}.$$

En este caso,

$$\bar{Y}^* = g_Y + \bar{Y},$$

$$\bar{X}_k^* = g_k + \bar{X}_k,$$

$$m_{jk}^* = m_{jk},$$

$$m_{Yk}^* = m_{Yk},$$

$$m_{Yj}^* = m_{Yj} \quad (j, k = 2, 3, \dots, K).$$

Y, por tanto, $\hat{\beta}_k^* = \hat{\beta}_k$ para $k = 2, 3, \dots, K$;

pero

$$\begin{aligned} \beta_1^* &= \bar{Y}^* - \beta_2^* \bar{X}_2^* - \beta_3^* \bar{X}_3^* - \dots - \beta_K^* \bar{X}_K^* \\ &= \beta_1 + (g_Y - \beta_2 g_2 - \beta_3 g_3 - \dots - \beta_K g_K). \end{aligned}$$

Es decir, las estimaciones de las pendientes de regresión quedan inalteradas, pero varía la estimación de la ordenada en el origen. Del mismo modo, las estimaciones de las variancias de $\hat{\beta}_2^*$, $\hat{\beta}_3^*$, ..., $\hat{\beta}_K^*$ no varían, pero sí la de la variancia estimada de $\hat{\beta}_1^*$. Los valores de los estadísticos t y F , y el de R^2 siguen siendo los mismos.

En resumen, el principio general acerca de las unidades de medida nos dice que, en las ecuaciones de regresión lineal, las transformaciones lineales

de los valores de R^2 ni de los estadísticos de prueba fundamentales. Los valores de los coeficientes de regresión estimados pueden variar, pero el R^2 puede obtenerse fácilmente en base a los parámetros de la transformación lineal.

Breve comentario sobre los supuestos básicos y los datos

En el capítulo 8 hemos estudiado las consecuencias del incumplimiento de los supuestos básicos del modelo clásico de regresión normal lineal en el caso de la regresión simple. Las conclusiones que obtuvimos allí se cumplen totalmente, o efectuando solamente pequeñas modificaciones, en el contexto de la regresión múltiple. Concretamente, los comentarios relativos al supuesto de la distribución normal y la media cero de ϵ , son también de aplicación en el modelo de regresión múltiple. Respecto a la heteroscedasticidad, se pueden ampliar fácilmente el análisis y los resultados de la sección 8.1 y aplicarlo al caso de la regresión múltiple con sólo considerar la posibilidad de que exista más de una variable explicativa, y lo mismo ocurre con respecto a los resultados de los modelos con perturbaciones autorregresivas (sección 8.2) o con regresores estocásticos (sección 8.3). Los problemas debidos a las deficiencias en los datos, que hemos estudiado en el capítulo 9, aparecen también en los modelos de regresión múltiple. La mayor parte de los resultados y afirmaciones efectuadas en el contexto de los modelos de regresión simple se pueden ampliar sin gran dificultad para aplicarlos a los modelos de regresión múltiple.

10.3 Multicolinealidad

Según el supuesto (10.8) del modelo clásico de regresión normal lineal, es preciso que ninguna de las variables explicativas esté perfectamente correlacionada con otra de ellas, o con cualquier combinación lineal de varias de ellas. Cuando este supuesto no se cumple, decimos que existe *multicolinealidad perfecta*, y cuando, por el contrario, las variables explicativas no están correlacionadas entre sí, hablamos de *ausencia de multicolinealidad*. Resulta muy interesante estudiar los casos en los que existe un *elevado grado de multicolinealidad*, es decir, cuando una variable explicativa está muy correlacionada con otra variable explicativa o con una combinación lineal de varias de ellas.

Antes de estudiar con detalle la multicolinealidad, es preciso aclarar dos puntos.

1. La multicolinealidad es una cuestión de grado, no de existencia. La distinción importante no es entre presencia y ausencia, sino entre los distintos grados de multicolinealidad.

2. La multicolinealidad hace referencia a una característica de las variables explicativas, que se suponen no estocásticas, y, por tanto, es una característica de la muestra y no de la población.⁹

Por tanto, no vamos a contrastar la «existencia de multicolinealidad» sino que, en caso de descartarlo, mediremos su importancia en una muestra concreta. En lo que sigue, nos ocuparemos de las implicaciones de los distintos grados de multicolinealidad para la estimación de los coeficientes de regresión.

Ausencia de multicolinealidad

Empecemos con el caso en que *no existe multicolinealidad*, es decir, cuando las variables explicativas no están correlacionadas entre sí. En tal caso, la matriz $(\mathbf{X}'\mathbf{X})$ es una matriz diagonal. En el modelo de regresión con dos variables explicativas

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

en el que $m_{23} = 0$, las ecuaciones normales minimocuadráticas de $\hat{\beta}_2$ y $\hat{\beta}_3$ se convierten en

$$m_{Y2} = \beta_2 m_{22}, \quad m_{Y3} = \beta_3 m_{33}.$$

Por tanto,
$$\hat{\beta}_2 = \frac{m_{Y2}}{m_{22}}, \quad \hat{\beta}_3 = \frac{m_{Y3}}{m_{33}}.$$

Estas fórmulas son exactamente iguales a las correspondientes a la regresión simple de Y_i sobre X_{i2} y de Y_i sobre X_{i3} .

El resultado anterior parece indicar que, cuando X_{i2} y X_{i3} no están correlacionadas, podríamos abandonar el modelo de regresión múltiple con dos variables explicativas y sustituirlo por los dos modelos de regresión simple

$$Y_i = \alpha_2 + \beta_2 X_{i2} + \varepsilon_i,$$

y
$$Y_i = \alpha_3 + \beta_3 X_{i3} + \varepsilon_i.$$

dos regresiones simples nos permitiría obtener un estimador de la constante de regresión β_1 , aunque en realidad esto no es muy importante, ya que el estimador minimocuadrático de β_1 es simplemente

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3.$$

9. Si las variables explicativas son estocásticas y en la población existe una relación subyacente entre ellas, dicha relación tendría que especificarse como parte del modelo. Si en la población no existe tal relación, aún es posible que encontremos alguna relación entre las variables de la muestra (y generalmente ocurrirá así). En este caso, la multicolinealidad es nuevamente una característica de la muestra, no de la población.

que es más importante es que la utilización de las regresiones simples para estimar las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$ da como resultado unas estimaciones sesgadas. Esto puede verse de la forma siguiente. Según la ecuación (10.24) sabemos que la variancia de $\hat{\beta}_2$ es

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2 m_{33}}{m_{22} m_{33} - m_{23}^2},$$

que cuando $m_{23} = 0$ se convierte en

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{m_{22}}.$$

El estimador de $\text{Var}(\hat{\beta}_2)$ en la regresión simple es

$$s_{\hat{\beta}_2}^2 = \frac{s_2^2}{m_{22}},$$

en la que
$$s_2^2 = \frac{1}{n-2} \sum_i [(Y_i - \bar{Y}) - \hat{\beta}_2 (X_{i2} - \bar{X}_2)]^2.$$

Tomando la esperanza matemática de s_2^2 , obtenemos

$$\begin{aligned} E(s_2^2) &= \frac{1}{n-2} E \sum_i [\beta_2 (X_{i2} - \bar{X}_2) + \beta_3 (X_{i3} - \bar{X}_3) \\ &\quad + (\varepsilon_i - \bar{\varepsilon}) - \beta_2 (X_{i2} - \bar{X}_2)]^2 \\ &= \frac{1}{n-2} [m_{22} \text{Var}(\hat{\beta}_2) + \beta_3^2 m_{33} - 2m_{22} \text{Var}(\hat{\beta}_2) + (n-1)\sigma^2] \\ &= \frac{1}{n-2} [(n-2)\sigma^2 + \beta_3^2 m_{33}], \end{aligned}$$

o sea, que
$$E(s_{\hat{\beta}_2}^2) = \text{Var}(\hat{\beta}_2) + \frac{\beta_3^2 m_{33}}{m_{22}(n-2)}.$$

Del mismo modo,
$$E(s_{\hat{\beta}_3}^2) = \text{Var}(\hat{\beta}_3) + \frac{\beta_2^2 m_{22}}{m_{33}(n-2)}.$$

Esto significa que los estimadores de las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$ obtenidos en la regresión simple tienen un sesgo positivo. Este resultado se puede generalizar para aplicarlo a los modelos de regresión con cualquier número de variables explicativas no correlacionadas entre sí.

Multicolinealidad perfecta

A continuación pasamos al caso en el que existe *multicolinealidad perfecta*. En el modelo de regresión múltiple con dos variables explicativas, la multicolinealidad perfecta significa que podemos escribir

$$(10.51) \quad X_{i2} = a + bX_{i3},$$

en la que a y b son ciertos números fijos, y $b \neq 0$. En este caso, el coeficiente de correlación entre X_{i2} y X_{i3} en la muestra es

$$r_{23} = \frac{\sum (a + bX_{i3} - a - b\bar{X}_3)(X_{i3} - \bar{X}_3)}{\sqrt{\sum (a + bX_{i3} - a - b\bar{X}_3)^2} \sqrt{\sum (X_{i3} - \bar{X}_3)^2}} = \frac{bm_{33}}{\sqrt{b^2 m_{33}} \sqrt{m_{33}}},$$

o sea, que $r_{23}^2 = 1$.

Analicemos ahora los estimadores minimocuadráticos de los coeficientes de regresión. Las ecuaciones normales minimocuadráticas del modelo con dos variables explicativas son

$$m_{Y2} = \beta_2 m_{22} + \beta_3 m_{23},$$

$$m_{Y3} = \beta_2 m_{23} + \beta_3 m_{33}.$$

Pero, según la ecuación (10.51), sabemos que

$$m_{Y2} = bm_{Y3},$$

$$m_{22} = b^2 m_{33},$$

$$m_{23} = bm_{33}.$$

Y, por tanto, las ecuaciones normales minimocuadráticas se convierten en

$$bm_{Y3} = b(\beta_2 bm_{33} + \beta_3 m_{33}),$$

$$m_{Y3} = \beta_2 bm_{33} + \beta_3 m_{33}.$$

Esto pone de manifiesto que la primera ecuación normal es igual a la segunda multiplicada por b y, por tanto, las dos ecuaciones no son independientes, lo que implica que la solución para β_2 y β_3 es indeterminada.

Consideremos ahora el caso de tres variables explicativas, con el fin de ilustrar una característica especial de la multicolinealidad, que no se pone de

manifiesto en un modelo con sólo dos variables. En este caso, la existencia de multicolinealidad perfecta significa que podemos escribir

$$(10.52) \quad X_{i2} = a + b_3 X_{i3} + b_4 X_{i4},$$

en la que a , b_3 y b_4 son ciertas cantidades fijas. Supongamos que tanto b_3 como b_4 son distintas de cero. El coeficiente de correlación muestral entre X_{i2} y X_{i3} es

$$\begin{aligned} r_{23} &= \frac{m_{23}}{\sqrt{m_{22}} \sqrt{m_{33}}}, \\ &= \frac{b_3 m_{33} + b_4 m_{34}}{\sqrt{b_3^2 m_{33} + b_4^2 m_{44} + 2b_3 b_4 m_{34}} \sqrt{m_{33}}}, \\ &= \frac{b_3 \sqrt{m_{33}} + b_4 r_{34} \sqrt{m_{44}}}{\sqrt{(b_3 \sqrt{m_{33}} + b_4 r_{34} \sqrt{m_{44}})^2 + b_4^2 m_{44} (1 - r_{34}^2)}}, \end{aligned}$$

en el que r_{34} es el coeficiente de correlación muestral entre X_{i3} y X_{i4} . Del mismo modo, el coeficiente de correlación entre X_{i2} y X_{i4} en la muestra es

$$r_{24} = \frac{b_3 r_{34} \sqrt{m_{33}} + b_4 \sqrt{m_{44}}}{\sqrt{(b_3 r_{34} \sqrt{m_{33}} + b_4 \sqrt{m_{44}})^2 + b_3^2 m_{33} (1 - r_{34}^2)}}.$$

Este resultado muestra claramente que la existencia de multicolinealidad perfecta *no* significa necesariamente que la correlación entre cualquier par de variables explicativas tenga que ser perfecta, ni siquiera especialmente alta, si el número total de variables explicativas es mayor que dos. Por ejemplo, cuando

$$X_{i2} = X_{i3} + X_{i4},$$

$$m_{33} = m_{44},$$

$$r_{34} = -0.5,$$

y

tenemos

$$r_{23} = \frac{1 - 0.5}{\sqrt{(1 - 0.5)^2 + (1 - 0.5^2)}} = 0.5$$

$$r_{24} = \frac{-0.5 + 1}{\sqrt{(-0.5 + 1)^2 + (1 - 0.5^2)}} = 0.5.$$

En este caso, existe multicolinealidad perfecta, pero, sin embargo, ninguno de los coeficientes de correlación es mayor que una mitad en valor absoluto. Esto es importante porque significa que, cuando existen más de dos variables expli-

cativas, no basta con mirar simplemente los coeficientes de correlación, y llegar a la conclusión de que *no* existe multicolinealidad perfecta (o muy elevada) en la muestra. Por otra parte, si la correlación entre cualquier par de variables explicativas es perfecta, la multicolinealidad de la muestra *es* perfecta, ya que la correlación perfecta entre (por ejemplo) X_{12} y X_{13} implica que la primera de estas dos variables es una función lineal exacta de la segunda, y podemos escribir

$$X_{12} = a + b_3 X_{13} \quad (b_3 \neq 0),$$

que equivale a la expresión (10.52) con $b_4 = 0$. Por tanto, cuando el número de variables explicativas es mayor que dos, la correlación perfecta entre dos variables explicativas es una condición suficiente, pero no necesaria, para la existencia de multicolinealidad perfecta en la muestra.

Veamos ahora lo que ocurre con los estimadores minimocuadráticos de los coeficientes de regresión, en condiciones de multicolinealidad perfecta, cuando el modelo contiene tres variables explicativas. Las ecuaciones normales minimocuadráticas son

$$m_{Y2} = \beta_2 m_{22} + \beta_3 m_{23} + \beta_4 m_{24},$$

$$m_{Y3} = \beta_2 m_{23} + \beta_3 m_{33} + \beta_4 m_{34},$$

$$m_{Y4} = \beta_2 m_{24} + \beta_3 m_{34} + \beta_4 m_{44}.$$

Pero, según la ecuación (10.52), tenemos

$$m_{Y2} = b_3 m_{Y3} + b_4 m_{Y4},$$

$$m_{22} = b_3^2 m_{33} + b_4^2 m_{44} + 2b_3 b_4 m_{34},$$

$$m_{23} = b_3 m_{33} + b_4 m_{34},$$

$$m_{24} = b_3 m_{34} + b_4 m_{44}.$$

Sustituyendo estas expresiones en las ecuaciones normales minimocuadráticas, obtenemos

$$b_3 m_{Y3} + b_4 m_{Y4} = \beta_2 (b_3^2 m_{33} + b_4^2 m_{44} + 2b_3 b_4 m_{34}) + \beta_3 (b_3 m_{33} + b_4 m_{34}) + \beta_4 (b_3 m_{34} + b_4 m_{44}),$$

$$m_{Y3} = \beta_2 (b_3 m_{33} + b_4 m_{34}) + \beta_3 m_{33} + \beta_4 m_{34},$$

$$m_{Y4} = \beta_2 (b_3 m_{34} + b_4 m_{44}) + \beta_3 m_{34} + \beta_4 m_{44}.$$

Por tanto, la primera ecuación normal es igual a la segunda multiplicada por b_1 más la tercera multiplicada por b_4 , lo que indica que, cuando existe multicolinealidad perfecta, las tres ecuaciones normales no son independientes y no

podemos despejar los valores de $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_4$ a partir de ellos. Este resultado puede ampliarse a los modelos de regresión con cualquier número de variables explicativas. Según la expresión (10.14), el vector de los estimadores minimocuadráticos de los coeficientes de regresión es

$$\hat{\beta} = (X'X)^{-1}(X'Y),$$

y, por tanto, la existencia de una relación lineal exacta entre las variables explicativas significa que una de las columnas de $(X'X)$ es una función lineal exacta de otra columna o de varias de ellas. En consecuencia, la matriz $(X'X)$ es una matriz singular y carece de inversa.

Así pues, cuando en la muestra existe multicolinealidad perfecta, los estimadores minimocuadráticos de los coeficientes de regresión son indeterminados. Desgraciadamente, en estas circunstancias no es posible obtener otros resultados. Al existir una multicolinealidad perfecta, la muestra no nos proporciona información alguna acerca de la respuesta de la variable dependiente debida a los cambios en una de las variables explicativas «manteniéndose constantes las demás». La sugerencia tradicional consiste en utilizar información procedente de otras fuentes acerca de los coeficientes de regresión. Por ejemplo, si en el modelo

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

las dos variables explicativas están perfectamente correlacionadas, pero sabemos o podemos suponer que

$$\frac{\beta_3}{\beta_2} = k,$$

siendo k una constante conocida, la estimación de los coeficientes de regresión se hace posible. Sustituyendo

$$\beta_3 = k\beta_2$$

en el modelo de regresión, obtenemos

$$Y_i = \beta_1 + \beta_2 X_{i2} + k\beta_2 X_{i3} + \varepsilon_i$$

o

$$Y_i = \beta_1 + \beta_2 Z_i + \varepsilon_i,$$

en la que Z_i corresponde a $(X_{i2} + kX_{i3})$. En tal caso, podemos obtener los estimadores minimocuadráticos de β_1 y β_2 , y obtener por inferencia el estimador de β_3 . Estos estimadores pueden considerarse condicionados al valor conocido de k . Si disponemos de varios valores alternativos de k , podemos obtener

un conjunto de estimadores distintos, cada uno de ellos condicionado a un valor distinto de k .¹⁰

Los casos en los que conocemos la razón entre los dos coeficientes de regresión, o tenemos algún otro tipo de información exacta acerca de ellos, son relativamente escasos. Con mayor frecuencia, podemos disponer de una estimación de uno o más coeficientes de regresión procedente de una muestra distinta. Por ejemplo, podemos disponer de una muestra de observaciones procedentes de una serie temporal para estimar un modelo de regresión, junto a una estimación insesgada de uno de los coeficientes de regresión, procedente de los datos de corte transversal de una encuesta. Este podría ser el caso en la estimación de una función de demanda de un producto cualquiera, como función del precio y de la renta. Si se efectúan las observaciones a lo largo del tiempo, generalmente se podrá disponer de una estimación del coeficiente de la renta a partir de una muestra de corte transversal. Supongamos que el modelo de regresión que vamos a estimar es

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

y disponemos de un estimador insesgado de β_3 , al que llamaremos $\hat{\beta}_3$, procedente de una muestra independiente de corte transversal. En este caso, el modelo de regresión puede escribirse de nuevo en la forma

$$(Y_i - \hat{\beta}_3 X_{i3}) = \beta_1 + \beta_2 X_{i2} - (\hat{\beta}_3 - \beta_3) X_{i3} + \varepsilon_i$$

o

$$Y_i^* = \beta_1 + \beta_2 X_{i2} + u_i.$$

El estimador minimocuadrático de β_2 se convierte en

$$\begin{aligned} (10.53) \quad \hat{\beta}_2 &= \frac{\sum (Y_i^* - \bar{Y}^*)(X_{i2} - \bar{X}_2)}{\sum (X_{i2} - \bar{X}_2)^2} \\ &= \frac{\sum (Y_i - \hat{\beta}_3 X_{i3} - \bar{Y} + \hat{\beta}_3 \bar{X}_3)(X_{i2} - \bar{X}_2)}{\sum (X_{i2} - \bar{X}_2)^2} \\ &= \frac{m_{Y2} - \hat{\beta}_3 m_{23}}{m_{22}}. \end{aligned}$$

La media y la variancia de $\hat{\beta}_2$ son

$$(10.54) \quad E(\hat{\beta}_2) = \frac{1}{m_{22}} E[\beta_2 m_{22} + \beta_3 m_{33} + \sum_i (X_{i2} - \bar{X}_2) \varepsilon_i - \hat{\beta}_3 m_{23}] = \beta_2,$$

10. En estos casos, el enfoque bayesiano de la estimación puede resultar particularmente atractivo. Ver, por ejemplo, D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2 — Inference*, Cambridge, Inglaterra, The University Press, 1964, cap. 8.

$$\text{Var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2$$

$$= \frac{1}{m_{22}^2} E\{-(\hat{\beta}_3 - \beta_3)m_{23} + \sum_i (X_{i2} - \bar{X}_2)\varepsilon_i\}^2$$

$$= \frac{1}{m_{22}^2} [m_{23}^2 \text{Var}(\hat{\beta}_3) + \sigma^2 m_{22}]$$

$$= \frac{\sigma^2}{m_{22}} + \frac{m_{23}^2 \text{Var}(\hat{\beta}_3)}{m_{22}^2}.$$

En condiciones de multicolinealidad perfecta,

$$m_{23}^2 = m_{22}m_{33},$$

y, por tanto, la variancia de $\hat{\beta}_2$ se convierte en

$$(10.56) \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{m_{22}} + \frac{m_{33} \text{Var}(\hat{\beta}_3)}{m_{22}}.$$

En las aplicaciones empíricas, a menudo se simplifica la fórmula correspondiente a la variancia de $\hat{\beta}_2$ considerándola condicionada al valor de $\hat{\beta}_3$, lo que equivale a tratar a este último como una constante.

La objeción que puede plantearse a este tratamiento del problema de la multicolinealidad, consistente en la obtención de información procedente de otras fuentes además de la muestra de que disponemos, es que se trata de algo que debería hacerse en cualquier caso. Normalmente, suponemos que al especificar el modelo de regresión y el procedimiento de estimación utilizamos toda la información disponible acerca de la población. No hay ninguna razón para esperar hasta que se presente la multicolinealidad perfecta para ir en busca de toda la información relevante, salvo en el caso en que su obtención resulte muy costosa. Si realmente hubiésemos agotado todas las fuentes de información antes de pasar a la estimación, el remedio que se ha propuesto para el problema de la multicolinealidad no se podría aplicar. Además, por lo que respecta al uso de una estimación procedente de otra muestra, podemos también poner en duda el buen criterio implícito en el uso de esta información solamente, prescindiendo del resto de la muestra.

Alto grado de multicolinealidad

Aquí hemos considerado solamente los dos casos extremos de multicolinealidad: el caso en que no existía en absoluto y el caso en el que la multicolinealidad era perfecta. En las aplicaciones prácticas no es frecuente encontrar uno de estos dos extremos, pero la mayor parte de los datos presentan multicolinealidad (no perfecta). En el modelo de regresión con

dos variables explicativas, la relación que existe entre estas dos últimas en la muestra se puede describir de forma general como

$$(10.57) \quad X_{i2} = a + bX_{i3} + v_i,$$

en la que a y b son ciertas constantes, y v_i es un «residuo» tal que $\sum v_i = \sum v_i X_{i3} = 0$. Si definimos m_{vv} como

$$m_{vv} = \sum v_i^2,$$

el caso en que $m_{vv} = 0$ representa la multicolinealidad perfecta, y el caso en que $m_{vv} = m_{22}$ representa la ausencia de multicolinealidad. (Obsérvese que m_{22} es el valor máximo que puede tomar m_{vv} .) Estos dos casos constituyen casos especiales de la ecuación (10.57). El coeficiente de correlación entre X_{i2} y X_{i3} , en la muestra, es

$$\begin{aligned} r_{23} &= \frac{\sum (a + bX_{i3} + v_i - a - b\bar{X}_3)(X_{i3} - \bar{X}_3)}{\sqrt{\sum (a + bX_{i3} + v_i - a - b\bar{X}_3)^2} \sqrt{\sum (X_{i3} - \bar{X}_3)^2}} \\ &= \frac{bm_{33}}{\sqrt{b^2 m_{33} + m_{vv}} \sqrt{m_{33}}} \\ \text{o} \quad r_{23}^2 &= \frac{b^2 m_{33}}{b^2 m_{33} + m_{vv}}. \end{aligned}$$

Cuando $m_{vv} = 0$, tendremos $r_{23}^2 = 1$, y cuando $m_{vv} = m_{22}$ (lo que implica que $b = 0$), $r_{23}^2 = 0$. En la mayor parte de los casos, m_{vv} estará entre estos dos extremos, y r_{23}^2 tomará un valor entre 0 y 1. Consideremos ahora las ecuaciones normales minimocuadráticas del modelo con dos variables explicativas. En primer lugar, a partir de la ecuación (10.57), tenemos

$$\begin{aligned} m_{Y2} &= bm_{Y3} + m_{Yv} \quad (\text{en la que } m_{Yv} = \sum (Y_i - \bar{Y})v_i), \\ m_{22} &= b^2 m_{33} + m_{vv}, \\ m_{23} &= bm_{33}. \end{aligned}$$

Por tanto, las ecuaciones normales minimocuadráticas se pueden escribir en la forma

$$\begin{aligned} bm_{Y3} + m_{Yv} &= \beta_2(b^2 m_{33} + m_{vv}) + \beta_3 bm_{33}, \\ m_{Y3} &= \beta_2 bm_{33} + \beta_3 m_{33}. \end{aligned}$$

Cuando $m_{vv} = 0$ (y, por tanto $m_{Yv} = 0$), la solución de $\hat{\beta}_2$ y $\hat{\beta}_3$ se hace indeterminada, y para todos los demás valores de m_{vv} existirá una solución. Puede verse fácilmente que esto también se cumple para los modelos con más de dos

variables explicativas. Esto significa que, mientras la multicolinealidad no sea perfecta (y $n \geq K$), podremos obtener siempre una solución determinada para los estimadores minimocuadráticos de los coeficientes de regresión, a menos que nos encontremos con el problema de los errores de redondeo.

Examinemos ahora la relación existente entre el grado de multicolinealidad y las propiedades de los estimadores minimocuadráticos de los coeficientes de regresión. Bajo los supuestos básicos del modelo clásico de regresión normal lineal, los estimadores minimocuadráticos de los coeficientes de regresión poseen todas las propiedades deseables. Pero, como se ha señalado en la sección 7.4, el hecho de saber que los estimadores poseen tales propiedades sirve de escaso consuelo si sus variancias tienen un tamaño tal que las estimaciones resultantes son poco fiables. Es decir, el conocimiento de que nuestros estimadores tienen la menor variancia posible (entre todos los estimadores insesgados) no es de gran ayuda si, al mismo tiempo, esta variancia resulta ser muy grande, y es aquí donde interviene la multicolinealidad. Consideremos un modelo de regresión con dos variables explicativas. Según las ecuaciones (10.24) y (10.25), las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$ son

$$\begin{aligned} \text{Var}(\hat{\beta}_2) &= \frac{\sigma^2 m_{33}}{m_{22} m_{33} - m_{23}^2} = \frac{\sigma^2}{m_{22}(1 - r_{23}^2)}, \\ \text{Var}(\hat{\beta}_3) &= \frac{\sigma^2 m_{22}}{m_{22} m_{33} - m_{23}^2} = \frac{\sigma^2}{m_{33}(1 - r_{23}^2)}, \end{aligned}$$

y, según la ecuación (10.27), su covariancia es

$$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-\sigma^2 m_{23}}{m_{22} m_{33} - m_{23}^2} = \frac{-\sigma^2 r_{23}}{\sqrt{m_{22}} \sqrt{m_{33}} (1 - r_{23}^2)}.$$

Esto muestra claramente que cuando r_{23}^2 está muy cerca de la unidad, las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$, y su covariancia, son muy grandes (cuando $r_{23}^2 = 1$, son infinitas). En el caso en que sólo hay dos variables explicativas, el valor de r_{23}^2 indica el grado de multicolinealidad y, por tanto, el resultado anterior implica que, cuanto mayor sea dicho grado de multicolinealidad, mayores serán las variancias y la covariancia de $\hat{\beta}_2$ y $\hat{\beta}_3$. Cuando hay más de dos variables explicativas, la matriz de variancias y covariancias de los estimadores minimocuadráticos de los coeficientes de regresión es

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 (X'X)^{-1},$$

lo que indica en la ecuación (10.33). La existencia de un elevado grado de multicolinealidad significa que una columna de la matriz $(X'X)$ es casi una combinación lineal de una o varias de las columnas restantes. Esto implica que la matriz $(X'X)$ es numéricamente pequeña y que, por tanto, los

elementos de la matriz $(X'X)^{-1}$ —y, con ellos, las variancias y covariancias de los coeficientes de regresión estimados— son grandes.

Así pues, llegamos a la conclusión de que un elevado grado de multicolinealidad resulta perjudicial en el sentido de que hace que las estimaciones de los coeficientes de regresión sean muy poco precisas. La imprecisión surge a consecuencia de las grandes variancias de los estimadores minimocuadráticos. Aunque estas variancias no se conocen, pueden obtenerse estimaciones de ellas utilizando la fórmula (10.35a) para estimar σ^2 . No obstante, debe señalarse que las variancias de los coeficientes de regresión estimados pueden ser grandes aunque no exista multicolinealidad en absoluto, ya sea debido a que las variables explicativas tienen una dispersión pequeña, o a que la propia σ^2 es grande. Si queremos culpar de este problema a la multicolinealidad, tenemos que ser capaces de medir su grado. Para ello, en el caso de los modelos con dos variables explicativas, podemos utilizar el valor de r^2_{23} , pero, cuando existen más de dos variables explicativas, la medición del grado de multicolinealidad se hace mucho más complicada. Esto se debe a que, como hemos demostrado anteriormente, la existencia de un elevado grado de multicolinealidad, o incluso de multicolinealidad perfecta, generalmente no implica que la correlación entre cada par de variables explicativas tenga que ser particularmente elevada.

Medidas de multicolinealidad

El problema de la medición de la multicolinealidad en los modelos con más de dos variables explicativas se ha planteado de muchas formas distintas. Algunos investigadores han utilizado el valor del determinante de $(X'X)$, ya que dicho valor es pequeño cuando el grado de multicolinealidad es elevado, y cero cuando la multicolinealidad es perfecta. Esta medida tiene el inconveniente de no poseer límites y el de verse afectada por la dispersión de las variables explicativas, además de por su interrelación. Por ejemplo, si utilizásemos esta medida en el modelo con dos variables explicativas, obtendríamos

$$\text{Det} \begin{bmatrix} n & \sum X_{12} & \sum X_{13} \\ \sum X_{12} & \sum X_{12}^2 & \sum X_{12}X_{13} \\ \sum X_{13} & \sum X_{12}X_{13} & \sum X_{13}^2 \end{bmatrix} = n(m_{22}m_{33} - m_{23}^2) = nm_{22}m_{33}(1 - r_{23}^2).$$

Por tanto, dos conjuntos de datos muestrales con el mismo número de observaciones y el mismo valor de r^2_{23} producirían dos valores distintos del determinante si el producto de m_{22} y m_{33} no fuese igual en las dos muestras.

Otra medida de multicolinealidad, incorporada en algunos programas de regresión minimocuadrática para calculadoras electrónicas, se define en términos

llamado « R^2 con omisión». En estos programas no sólo se calcula el R^2 correspondiente a la ecuación de regresión en cuestión, sino también los R^2 que se obtienen omitiendo una de las variables explicativas cada vez. Los R^2 obtenidos de esta forma se denominan « R^2 con omisión». Por ejemplo, en una ecuación de regresión con tres variables explicativas, X_{2i} , X_{3i} y X_{4i} , tendríamos tres « R^2 con omisión», uno correspondiente a la regresión de Y_i respecto de X_{2i} y X_{4i} , otro a la regresión de Y_i respecto de X_{2i} y X_{3i} , y otro a la regresión de Y_i sobre X_{3i} y X_{4i} . Si en la muestra existe un fuerte grado de multicolinealidad, al menos una de las variables explicativas se moverá básicamente de acuerdo con la variación de otra o varias de las restantes variables explicativas. En consecuencia, la introducción de esta variable en la ecuación de regresión producirá solamente un pequeño aumento en el valor de R^2 . Un elevado grado de multicolinealidad se reflejará, pues, en el hecho de que la diferencia entre R^2 y el mayor de los « R^2 con omisión» será pequeña.

El hecho de que un elevado grado de multicolinealidad significa simplemente que al menos una de las variables explicativas se puede representar como una función lineal de otra o varias de las restantes variables explicativas más un pequeño residuo, nos sugiere una medida relativamente simple del grado de multicolinealidad. Si efectuamos una regresión de cada una de las variables explicativas respecto a las restantes, podremos obtener una medida de la «bondad del ajuste» calculando el valor de R^2 en cada caso. Si alguno de estos valores está próximo a la unidad, el grado de multicolinealidad es elevado o, en general, podemos considerar a los valores más altos de estos R^2 como una medida del grado de multicolinealidad existente en la muestra. Sin embargo, esta medida y todas las demás que hemos presentado adolecen de la falta de una definición clara de lo que constituye un grado «elevado» de multicolinealidad. Teniendo en cuenta que casi siempre existe cierta multicolinealidad, la pregunta relevante es: ¿en qué punto deja de ser «normal» esta multicolinealidad, para pasar a ser «perjudicial»? Esta pregunta no ha podido ser resuelta de modo satisfactorio. Según el criterio utilizado a veces en la práctica, se considera que la multicolinealidad es perjudicial si a un nivel de significación de, por ejemplo, el 5 %, el valor del estadístico F difiere de cero de forma significativa, pero ninguno de los estadísticos t correspondientes a los coeficientes de regresión (aparte del de la constante de regresión) difiere de dicho valor. En tal caso, rechazaríamos la hipótesis de que no existe relación entre Y_i por una parte y X_{2i} , X_{3i} , ..., X_{ki} por otra, pero no rechazaríamos la hipótesis de que cada una de las variables explicativas ejerce una influencia irrelevante sobre Y_i . Esta situación indica que la influencia de cada una de las variables explicativas por separado es muy débil en relación a su influencia conjunta sobre Y_i . Esto es un síntoma de la existencia de un grado elevado de multicolinealidad, que nos impide desentrañar las influencias individuales de las variables explicativas. El inconveniente de este criterio con respecto a la multicolinealidad es demasiado estricto, en el sentido de que sólo consideramos

que la multicolinealidad es perjudicial cuando no se pueden distinguir las distintas influencias de las variables explicativas sobre Y .

Supongamos ahora que tenemos una muestra con un grado de multicolinealidad considerado definitivamente perjudicial. ¿Qué podemos hacer en tal caso? Si ya hemos utilizado toda la información previa al especificar el modelo, y no disponemos de información procedente de otras muestras, no podemos hacer gran cosa. Algunos investigadores, al encontrarse con este problema en observaciones procedentes de series temporales, han transformado los datos convirtiéndolos en primeras diferencias. Es decir, en lugar de estimar

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \dots + \beta_K X_{tK} + \epsilon_t,$$

se aplica el método de estimación minimocuadrática al modelo

$$Y_t - Y_{t-1} = \alpha + \beta_2(X_{t2} - X_{t-1,2}) + \beta_3(X_{t3} - X_{t-1,3})$$

$$+ \dots + \beta_K(X_{tK} - X_{t-1,K}) + (\epsilon_t - \epsilon_{t-1}).$$

En algunas circunstancias, esta transformación puede reducir el grado de multicolinealidad, pero introduce autorregresión en los términos de perturbación, que, de lo contrario, serían independientes. Como se ha señalado, la autorregresión posee unas consecuencias perjudiciales en relación con las propiedades de los estimadores minimocuadráticos, lo cual hace que la práctica consistente en utilizar primeras diferencias en lugar de los datos originales resulte poco convincente.

Como es muy poco lo que puede hacerse respecto a la multicolinealidad una vez que se ha utilizado toda la información de la que se dispone, lo único que podemos efectuar es aumentar el volumen de información. Una forma de hacerlo consiste en aumentar el tamaño de la muestra. Por ejemplo, consideremos las variancias de β_2 y β_3 en un modelo con dos variables explicativas. Estas variancias son

$$\text{Var}(\beta_2) = \frac{\sigma^2}{m_{22}(1 - r_{23}^2)},$$

$$\text{Var}(\beta_3) = \frac{\sigma^2}{m_{33}(1 - r_{23}^2)}.$$

Un aumento en el tamaño de la muestra puede hacer aumentar m_{22} y m_{33} , o reducir r_{23}^2 , o ambas cosas a la vez. El aumento en m_{22} y m_{33} ocurrirá en los casos en que los valores adicionales de X_{t2} y X_{t3} sean distintos de $\bar{X}_2$ y $\bar{X}_3$, como probablemente ocurra. Por otra parte, es difícil prever lo que ocurrirá con r_{23}^2 al aumentar n , ya que las X son no estocásticas y no se hallan bajo nuestro control. En realidad, r_{23} podría aumentar hasta el punto que las variancias de β_2 y β_3 realmente aumentasen. Una forma de evitar que ocurra esto es incrementar el tamaño de la muestra en forma tal que m_{22} y m_{33} aumenten sustancialmente, por ejemplo, combinando datos de series temporales y de corte transversal.

En la literatura aplicada, existe a menudo bastante confusión acerca del grado de encontrar un alto grado de multicolinealidad en la muestra. Hay una tendencia por parte de algunos investigadores a señalar que el elevado grado de multicolinealidad en la muestra es la causa de que los coeficientes de regresión estimados no sean significativamente distintos de cero, y se hace pensar al lector que, si la multicolinealidad fuese menor, los coeficientes de regresión estimados resultarían significativos. Esto puede ser así, pero, desde luego, no se deriva de los resultados que se han presentado. Por este motivo, es importante darse cuenta de que un elevado grado de multicolinealidad es simplemente una característica de la muestra que contribuye a reducir la habilidad de los coeficientes estimados, pero no ejerce influencia alguna respecto a las conclusiones a las que se llegue como resultado de esta escasa fiabilidad. Si los coeficientes de regresión estimados son muy poco fiables —es decir, si sus variancias son muy grandes—, la región de aceptación correspondiente a la hipótesis de que un determinado coeficiente es cero será muy amplia. A su vez, esto significa que la potencia de la contrastación es muy pequeña y, por tanto, no es de gran ayuda para discriminar entre hipótesis ciertas y falsas. Esto es lo único que se puede decir, independientemente del motivo que hace que dichas variancias sean grandes.

10.4 Errores de especificación

La especificación de un modelo de regresión consta de la formulación de la ecuación de regresión, y de las afirmaciones o supuestos acerca de los regresores y el término de perturbación. En el sentido amplio de la expresión, existirá un «error de especificación» siempre que sea incorrecta la formulación de la ecuación de regresión o uno de los supuestos subyacentes. Sin embargo, en sentido estricto, el *error de especificación* se refiere únicamente a los errores en la formulación de la ecuación de regresión apropiada, y ésta es la interpretación que adoptaremos. Consideraremos varias clases distintas de estos errores, y concretamente los que ocurren a consecuencia de:

- 1. Omisión de una variable explicativa relevante.
2. No consideración de un cambio cualitativo en una de las variables explicativas.
3. Inclusión de una variable explicativa irrelevante.
4. Formulación matemática incorrecta de la ecuación de regresión.
5. Especificación incorrecta de la forma en que el término de perturbación aparece en la ecuación de regresión.

En este capítulo estudiaremos los casos en que las variables explicativas son no estocásticas, las conclusiones permanecerían básicamente inalteradas si di-

chas variables fuesen estocásticas, siempre que fuesen independientes del término de perturbación de la regresión. Nuestro objetivo principal consistirá en descubrir las consecuencias de cada uno de los tipos de errores de especificación sobre los estimadores minimocuadráticos de los coeficientes de regresión y sus errores standard.

Omisión de una variable explicativa relevante

Consideremos, en primer lugar, los errores de especificación debidos a la omisión de una variable explicativa relevante en la ecuación de regresión. Concretamente, supongamos que la especificación correcta de la ecuación de regresión es

$$(10.58) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

pero, en cambio, estimamos

$$(10.59) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \epsilon_i^*.$$

Este tipo de error puede cometerse si no se dispone de observaciones de X_{i3} , o si el investigador no es consciente del hecho de que para especificar correctamente la hipótesis sostenida debería incluir dicha variable en la ecuación de regresión. Si la ecuación (10.59) fuese correcta, los estimadores minimocuadráticos de β_1 y β_2 serían insesgados y eficientes para todos los tamaños muestrales. Veamos ahora lo que ocurre con dichos estimadores cuando la formulación correcta es (10.58) en lugar de (10.59). Para $\hat{\beta}_2$ tenemos

$$E(\hat{\beta}_2) = E\left[\frac{\sum (X_{i2} - \bar{X}_2)(Y_i - \bar{Y})}{\sum (X_{i2} - \bar{X}_2)^2}\right].$$

Pero, según (10.58), sabemos que

$$(Y_i - \bar{Y}) = \beta_2(X_{i2} - \bar{X}_2) + \beta_3(X_{i3} - \bar{X}_3) + (\epsilon_i - \bar{\epsilon})$$

de modo que

$$E(\hat{\beta}_2) = \beta_2 + \beta_3 d_{32},$$

en la que

$$d_{32} = \frac{\sum (X_{i2} - \bar{X}_2)(X_{i3} - \bar{X}_3)}{\sum (X_{i2} - \bar{X}_2)^2}.$$

Del mismo modo, para $\hat{\beta}_1$ tenemos

$$E(\hat{\beta}_1) = E(\bar{Y} - \hat{\beta}_2 \bar{X}_2) = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 - (\beta_2 + \beta_3 d_{32}) \bar{X}_2$$

$$= \beta_1 + \beta_3 d_{31},$$

en la que

$$d_{31} = \bar{X}_3 - d_{32} \bar{X}_2.$$

que las expresiones correspondientes a d_{31} y d_{32} son, en realidad, las de los coeficientes minimocuadráticos de la ecuación

$$(10.60) \quad X_{i3} = d_{31} + d_{32} X_{i2} + \text{residuo}.$$

Como X_{i2} y X_{i3} no son estocásticas, la ecuación (10.60) se puede considerar solamente como una ecuación de regresión puramente descriptiva. En esta ecuación, la variable «dependiente» está representada por la variable omitida X_{i3} , y la «explicativa» por la variable incluida X_{i2} . Sabiendo que β_3 es distinta de cero, el estimador minimocuadrático de β_2 basado en la ecuación (10.59) será *sesgado* a menos que d_{32} sea igual a cero, es decir, a menos que X_{i2} y X_{i3} no estén correlacionadas. Si β_3 y d_{32} tienen el mismo signo, el sesgo de β_2 será positivo y, en caso contrario, negativo. Esto significa que la dirección del sesgo de $\hat{\beta}_2$ depende de

1. El signo de β_3 .
2. La dirección de la correlación entre la variable explicativa omitida y la incluida.

Si la correlación entre X_{i2} y X_{i3} no desaparece al aumentar el tamaño de la muestra, es decir, si

$$\lim_{n \rightarrow \infty} d_{32} \neq 0,$$

$\hat{\beta}_2$ será también inconsistente. Además, el estimador minimocuadrático de β_1 basado en la ecuación (10.59) será sesgado si

$$\bar{X}_3 - d_{32} \bar{X}_2 \neq 0,$$

e inconsistente si

$$\lim_{n \rightarrow \infty} (\bar{X}_3 - d_{32} \bar{X}_2) \neq 0.$$

Estos resultados parecen sugerir que, si la variable explicativa omitida *no está correlacionada* con la variable explicativa incluida, su omisión puede no dar lugar a graves consecuencias para la estimación minimocuadrática. Examinemos este punto con mayor detalle. Supongamos que X_{i2} y X_{i3} no están correlacionadas, o sea que $d_{32} = 0$. En este caso, $\hat{\beta}_2$ basada en la ecuación (10.59) es insesgada, y su variancia es

$$\text{Var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2}{\sum (X_{i2} - \bar{X}_2)^2}.$$

El estimador de $\text{Var}(\hat{\beta}_2)$ basado en (10.59) es

$$\begin{aligned} s_{\hat{\beta}_2}^2 &= \frac{s^2}{\sum (X_{i2} - \bar{X}_2)^2} = \frac{\sum [(Y_i - \bar{Y}) - \beta_2(X_{i2} - \bar{X}_2)]^2 / (n-2)}{\sum (X_{i2} - \bar{X}_2)^2} \\ &= \frac{\sum [-(\hat{\beta}_2 - \beta_2)(X_{i2} - \bar{X}_2) + \beta_3(X_{i3} - \bar{X}_3) + (\varepsilon_i - \bar{\varepsilon})]^2}{(n-2) \sum (X_{i2} - \bar{X}_2)^2}. \end{aligned}$$

La esperanza matemática de $s_{\hat{\beta}_2}^2$ viene dada por

$$\begin{aligned} E(s_{\hat{\beta}_2}^2) &= \frac{m_{22}\text{Var}(\hat{\beta}_2) + \beta_3^2 m_{33} - 2m_{23}\text{Var}(\hat{\beta}_2) + (n-1)\sigma^2}{(n-2)m_{22}} \\ &= \text{Var}(\hat{\beta}_2) + \frac{\beta_3^2 m_{33}}{(n-2)m_{22}}. \end{aligned}$$

Esto implica que, en tal caso, el estimador de $\text{Var}(\hat{\beta}_2)$ tiene un sesgo positivo. Por tanto, las contrastaciones de significación habituales para β_2 no son válidas, ya que tenderán a rechazar la hipótesis nula con mayor frecuencia de la que está justificada por el nivel de significación particular. Además, la esperanza matemática del estimador minimocuadrático de β_1 basado en (10.59) es

$$E(\hat{\beta}_1) = E(\bar{Y} - \hat{\beta}_2 \bar{X}_2) = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 - \beta_2 \bar{X}_2 = \beta_1 + \beta_3 \bar{X}_3,$$

lo que significa que $\hat{\beta}_1$ es sesgado a menos que $\bar{X}_3 = 0$.

Los resultados correspondientes a la estimación minimocuadrática de β_1 y β_2 basada en la especificación (10.59), cuando la especificación correcta es (10.58), pueden resumirse de la forma siguiente: si la variable explicativa omitida está correlacionada con la incluida, los estimadores de β_1 y β_2 serán sesgados e inconsistentes. Si la variable omitida *no* está correlacionada con la incluida, el estimador de β_1 seguirá siendo generalmente sesgado e inconsistente, pero el estimador de β_2 será insesgado. No obstante, el estimador de la *variancia* de β_2 tendrá un sesgo positivo, de modo que las pruebas de significación y los intervalos de confianza para β_2 tenderán a producir conclusiones demasiado conservadoras.

El análisis anterior se puede ampliar fácilmente al caso en que existe mayor cantidad de variables explicativas. Supongamos que la especificación correcta de la ecuación de regresión es

$$(10.61) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i,$$

pero, en cambio, estimamos

$$(10.62) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i^*.$$

El estimador minimocuadrático de β_2 basado en (10.62) es

$$\hat{\beta}_2 = \frac{m_{Y2}m_{33} - m_{Y3}m_{23}}{m_{22}m_{33} - m_{23}^2}.$$

Ahora bien, $E(m_{Y2}) = \beta_2 m_{22} + \beta_3 m_{23} + \beta_4 m_{24},$

$$E(m_{Y3}) = \beta_2 m_{23} + \beta_3 m_{33} + \beta_4 m_{34},$$

o sea, que

$$E(m_{Y2}m_{33} - m_{Y3}m_{23}) = \beta_2(m_{22}m_{33} - m_{23}^2) + \beta_4(m_{24}m_{33} - m_{34}m_{23}).$$

Por tanto,

$$E(\hat{\beta}_2) = \beta_2 + \beta_4 d_{42},$$

en la que

$$d_{42} = \frac{m_{42}m_{33} - m_{43}m_{23}}{m_{22}m_{33} - m_{23}^2}.$$

Obsérvese que la expresión correspondiente a d_{42} es exactamente igual a la fórmula del coeficiente minimocuadrático de X_{i4} en la ecuación

$$(10.63) \quad X_{i4} = d_{41} + d_{42} X_{i2} + \text{residuo},$$

en la ecuación (10.63) se ha efectuado una «regresión» de la variable explicativa omitida X_{i4} , respecto a las dos variables incluidas X_{i2} y X_{i3} . Del mismo modo,

$$E(\hat{\beta}_3) = \beta_3 + \beta_4 d_{43} \quad \text{y} \quad E(\hat{\beta}_1) = \beta_1 + \beta_4 d_{41}.$$

Por tanto, llegamos a la misma conclusión que en el caso en que existía un menor número de variables explicativas.

EJEMPLO. En un estudio de las funciones de producción en la industria de la India, Murti y Sastry¹¹ utilizaron datos correspondientes a la producción y los factores de una muestra de 320 empresas, obteniendo las siguientes estimaciones de la función de producción de Cobb-Douglas:

$$\log x_i = \log 0.68 + 0.53 \log n_i + 0.50 \log k_i + \varepsilon_i,$$

donde x_i = valor neto de la producción, n = sueldos y salarios, k = valor de los factores netos, y el subíndice i hace referencia a la empresa i -ésima. Si el factor «ges-

¹¹ Murti, M. y Sastry, S. «Production Functions for Indian Industry», *Econometrica*, vol. 29, pp. 205-21.

ción» difiere sistemáticamente de una empresa a otra y, por tanto, no puede considerarse como parte de la perturbación aleatoria, la ecuación anterior no está bien especificada. Si se introdujese dicho factor —que, naturalmente, es de difícil medición— en la función de producción en la misma forma que los factores n_i y k_i , el signo de su coeficiente sería claramente positivo. La regresión minimocuadrática de carácter descriptivo que relaciona el factor «gestión» m_i con las dos variables explicativas incluidas, n_i y k_i , es

$$\log m_i = d_{41} + d_{42} \log n_i + d_{43} \log k_i + \text{residuo.}$$

En este caso, podemos efectuar especulaciones acerca de los signos d_{42} y d_{43} . Si las empresas con un elevado nivel de capital poseen una gestión superior en comparación con las que utilizan una mayor proporción de trabajo, d_{42} sería negativa y d_{43} positiva. En estas circunstancias, y de no existir otras violaciones de los supuestos del modelo clásico de regresión normal lineal, la estimación del coeficiente de $\log n_i$ tendría un sesgo negativo, y la del coeficiente de $\log k_i$ un sesgo positivo.

Cambios cualitativos en las variables explicativas

Consideremos ahora los errores de especificación que surgen a consecuencia de un *cambio cualitativo* en una o más variables explicativas, al pasar de una observación a otra. Concretamente, supongamos que la ecuación de regresión especificada correctamente es

$$(10.64) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

pero no podemos observar X_{i2} y utilizamos X_{i2}^* que se define en la forma siguiente

$$(10.65) \quad X_{i2}^* = X_{i2} + Q_{i2},$$

en la que Q_{i2} es un factor que representa el cambio cualitativo en X_{i2}^* . Supongamos también que Q_{i2} es no estocástico y no se puede medir. Esta situación puede presentarse en el caso de una función de producción en la que la calidad del factor trabajo ha ido mejorando a lo largo del tiempo. Sin embargo, el concepto teórico de función de producción se refiere a una relación entre un producto homogéneo y unos factores homogéneos. El efecto de un cambio de este tipo sobre la estimación por mínimos cuadrados puede obtenerse de la forma siguiente. Sustituyendo X_{i2} por su valor según la ecuación (10.65) en la ecuación (10.64), obtenemos

$$(10.66) \quad \begin{aligned} Y_i &= \beta_1 + \beta_2 (X_{i2}^* - Q_{i2}) + \beta_3 X_{i3} + \epsilon_i \\ &= \beta_1 + \beta_2 X_{i2}^* + \beta_3 X_{i3} - \beta_2 Q_{i2} + \epsilon_i. \end{aligned}$$

los estimamos

$$(10.67) \quad Y_i = \beta_1 + \beta_2 X_{i2}^* + \beta_3 X_{i3} + \epsilon_i^*,$$

en lugar de (10.66), encontraremos el mismo error de especificación que si hubiésemos dejado de incluir una variable explicativa relevante. En este caso, las esperanzas matemáticas de los estimadores minimocuadráticos de β_1 , β_2 y β_3 basados en (10.67) son

$$E(\hat{\beta}_1) = \beta_1 - \beta_2 d_{41},$$

$$E(\hat{\beta}_2) = \beta_2 - \beta_2 d_{42},$$

$$E(\hat{\beta}_3) = \beta_3 - \beta_2 d_{43},$$

en las que las d son los coeficientes minimocuadráticos de

$$(10.68) \quad Q_{i2} = d_{41} + d_{42} X_{i2}^* + d_{43} X_{i3} + \text{residuo.}$$

En el caso de una función de producción, todas las β serían positivas. Además, es de esperar que en las observaciones en forma de serie temporal, d_{42} y d_{43} sean positivas, ya que las tres variables de (10.68) han ido creciendo a lo largo del tiempo, al menos después de la Segunda Guerra Mundial. Ello implica que tanto $\hat{\beta}_2$ como $\hat{\beta}_3$ tendrían un sesgo negativo. El análisis se puede ampliar fácilmente al caso en que el número de variables explicativas es mayor, o al caso en que se produce un cambio cualitativo en más de una variable explicativa.

Inclusión de una variable explicativa irrelevante

Otro de los tipos de errores de especificación se presenta cuando el grupo de variables explicativas significativas se ve aumentado por la inclusión de una o más *variables irrelevantes*. (Las ecuaciones de regresión formuladas por el procedimiento de incluir todos los posibles candidatos en el grupo de variables explicativas sin prestar gran atención a la teoría subyacente se llaman a veces, por razones obvias, «modelos de fregadero».) Por ejemplo, supongamos que la especificación correcta de la ecuación de regresión es

$$(10.69) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \epsilon_i,$$

los estimamos

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i^*.$$

Al utilizar la ecuación (10.70) en lugar de la (10.69) cometemos un error de especificación, ya que en la formulación de la hipótesis sostenida ignoramos la restricción de que $\beta_1 = 0$. Veamos lo que ocurre en tal caso con los estimadores minimocuadráticos de los coeficientes de regresión. En primer lugar, la esperanza matemática de $\hat{\beta}_3$ es

$$E(\hat{\beta}_3) = E \left[\frac{m_{Y3}m_{22} - m_{Y2}m_{23}}{m_{22}m_{33} - m_{23}^2} \right].$$

Pero, según (10.69), sabemos que

$$E(m_{Y2}) = \beta_2 m_{22} \quad \text{y} \quad E(m_{Y3}) = \beta_2 m_{23}$$

o sea, que

$$E(\hat{\beta}_3) = \frac{\beta_2 m_{22}m_{23} - \beta_2 m_{22}m_{23}}{m_{22}m_{33} - m_{23}^2} = 0.$$

Es decir, la media de $\hat{\beta}_3$ es igual al valor real de β_3 , que es cero. En una muestra determinada, la probabilidad de observar un valor de $\hat{\beta}_3$ significativamente distinto de cero es igual al nivel de significación elegido. Por otra parte, la esperanza matemática de $\hat{\beta}_2$ es

$$E(\hat{\beta}_2) = E \left[\frac{m_{Y2}m_{33} - m_{Y3}m_{23}}{m_{22}m_{33} - m_{23}^2} \right] = \beta_2.$$

Asimismo, $E(\hat{\beta}_1) = E(\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) = (\beta_1 + \beta_2 \bar{X}_2) - \beta_2 \bar{X}_2 = \beta_1$.

Estos resultados muestran que todos los estimadores de los coeficientes de la ecuación (10.70) son insesgados. En cuanto a sus variancias, tenemos

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2 m_{33}}{m_{22}m_{33} - m_{23}^2} = \frac{\sigma^2}{m_{22}(1 - r_{23}^2)},$$

en la que r_{23} es el coeficiente de correlación entre X_{i2} y X_{i3} . Ahora bien, si estimásemos β_2 en base a la ecuación de regresión especificada en la forma correcta, (10.69), su variancia, a la que llamaremos $\text{Var}(\hat{\beta}_2^*)$, sería

$$\text{Var}(\hat{\beta}_2^*) = \frac{\sigma^2}{m_{22}}.$$

La razón entre las dos variancias es

$$\frac{\text{Var}(\hat{\beta}_2)}{\text{Var}(\hat{\beta}_2^*)} = \frac{1}{1 - r_{23}^2}.$$

Como $0 \leq r_{23}^2 \leq 1$, podemos ver que

$$\frac{\text{Var}(\hat{\beta}_2)}{\text{Var}(\hat{\beta}_2^*)} \geq 1,$$

es la que la igualdad sólo se cumple cuando $r_{23} = 0$, es decir cuando X_{i2} y X_{i3} no están correlacionadas. La implicación de este resultado es que, generalmente, $\hat{\beta}_2$ no es eficiente. Del mismo modo, si operamos a partir de las fórmulas de $\text{Var}(\hat{\beta}_1)$ y $\text{Var}(\hat{\beta}_1^*)$ —en las que $\hat{\beta}_1$ y $\hat{\beta}_1^*$ representan los estimadores minimocuadráticos de β_1 en (10.70) y (10.69) respectivamente—, se puede mostrar que $\hat{\beta}_1$ tampoco es eficiente a menos que

$$(\bar{X}_3 \sqrt{m_{22}} - \bar{X}_2 r_{23} \sqrt{m_{33}}) = 0.$$

Consideremos ahora las variancias estimadas de los coeficientes minimocuadráticos de (10.70). Las fórmulas de $\text{Var}(\hat{\beta}_1)$, $\text{Var}(\hat{\beta}_2)$ y $\text{Var}(\hat{\beta}_3)$ sólo incluyen un parámetro desconocido, σ^2 , que se puede estimar por medio de

$$\begin{aligned} s^2 &= \frac{1}{n-3} \sum_i [(Y_i - \bar{Y}) - \hat{\beta}_2(X_{i2} - \bar{X}_2) - \hat{\beta}_3(X_{i3} - \bar{X}_3)]^2 \\ &= \frac{1}{n-3} \sum_i [\beta_2(X_{i2} - \bar{X}_2) + (\epsilon_i - \bar{\epsilon}) - \hat{\beta}_2(X_{i2} - \bar{X}_2) - \hat{\beta}_3(X_{i3} - \bar{X}_3)]^2 \\ &= \frac{1}{n-3} \sum_i [-(\hat{\beta}_2 - \beta_2)(X_{i2} - \bar{X}_2) - \hat{\beta}_3(X_{i3} - \bar{X}_3) + (\epsilon_i - \bar{\epsilon})]^2. \end{aligned}$$

La esperanza matemática de s^2 es

$$\begin{aligned} E(s^2) &= \frac{1}{n-3} [m_{22}\text{Var}(\hat{\beta}_2) + m_{33}\text{Var}(\hat{\beta}_3) + (n-1)\sigma^2 + 2m_{23}\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) \\ &\quad - 2E(\hat{\beta}_2 - \beta_2) \sum_i (X_{i2} - \bar{X}_2)\epsilon_i - 2E\hat{\beta}_3 \sum_i (X_{i3} - \bar{X}_3)\epsilon_i]. \end{aligned}$$

Pero

$$\begin{aligned} -2E(\hat{\beta}_2 - \beta_2) \sum_i (X_{i2} - \bar{X}_2)\epsilon_i \\ -2E \left[\frac{m_{Y3} \sum_i (X_{i2} - \bar{X}_2)\epsilon_i - m_{23} \sum_i (X_{i3} - \bar{X}_3)\epsilon_i}{m_{22}m_{33} - m_{23}^2} \right] \sum_i (X_{i2} - \bar{X}_2)\epsilon_i &= -2\sigma^2, \end{aligned}$$

$$-2 \sum_i (\hat{\beta}_3 - \beta_3) (X_{i3} - \bar{X}_3)\epsilon_i$$

$$-2E \left[\frac{m_{Y2} \sum_i (X_{i3} - \bar{X}_3)\epsilon_i - m_{23} \sum_i (X_{i2} - \bar{X}_2)\epsilon_i}{m_{22}m_{33} - m_{23}^2} \right] \sum_i (X_{i3} - \bar{X}_3)\epsilon_i = -2\sigma^2,$$

de modo que

$$E(s^2) = \frac{1}{n-3} \left[\frac{\sigma^2(m_{22}m_{33} + m_{33}m_{22} - 2m_{23}^2)}{m_{22}m_{33} - m_{23}^2} + (n-1)\sigma^2 - 4\sigma^2 \right] = \sigma^2.$$

Esto significa que el estimador de σ^2 basado en la ecuación (10.70) es insesgado y, en consecuencia, los estimadores de $\text{Var}(\beta_1)$, $\text{Var}(\beta_2)$ y $\text{Var}(\beta_3)$ también son insesgados. Este resultado se cumple tanto si X_{12} y X_{13} están correlacionadas, como si no lo están. La conclusión, pues, es que si el error de especificación consiste en la inclusión de algunas variables explicativas irrelevantes en la ecuación de regresión, los estimadores minimocuadráticos de los coeficientes de regresión son insesgados, pero no eficientes. Los estimadores de sus variancias son también insesgados, de modo que, si no existen otras complicaciones, las pruebas de significación y los intervalos de confianza para los parámetros de regresión son válidos. Estas conclusiones se han obtenido para un modelo de regresión simple, pero se puede mostrar fácilmente que también se pueden aplicar a los modelos con mayor número de variables explicativas.

No linealidad

Otro de los errores de especificación se presenta en el caso en que la especificación correcta en la ecuación de regresión *no es lineal*, pero estimamos lo que se puede considerar como su aproximación lineal. Como en la investigación econométrica aplicada se utilizan ampliamente las relaciones lineales, es fácil que este tipo de error se cometa con gran frecuencia. Supongamos que la especificación correcta de la ecuación de regresión viene dada por

$$(10.71) \quad Y_i = f(X_i) + \varepsilon_i,$$

en la que $f(X_i)$ es alguna función de X_i . Supongamos que esta función es continua y posee p derivadas continuas, siendo p algún número entero positivo. En tal caso, podemos efectuar una expansión de $f(X_i)$ respecto de $\bar{X}$, según el teorema de Taylor,¹² escribiendo la ecuación (10.71) en la forma

$$(10.71a) \quad Y_i = f(\bar{X}) + (X_i - \bar{X})f'(\bar{X}) + \frac{(X_i - \bar{X})^2}{2!}f''(\bar{X}) + \dots + \frac{(X_i - \bar{X})^p}{p!}f^{(p)}(\bar{X}) + R_{p+1} + \varepsilon_i,$$

en la que

$$f(\bar{X}) = f(X_i) \Big|_{X_i = \bar{X}},$$

$$f'(\bar{X}) = \frac{df(X_i)}{dX_i} \Big|_{X_i = \bar{X}},$$

$$f''(\bar{X}) = \frac{d^2f(X_i)}{dX_i^2} \Big|_{X_i = \bar{X}},$$

$\vdots$

$$R_{p+1} = \text{resto}.$$

Reordenando los términos del lado derecho de la ecuación (10.71a), obtenemos

$$(10.71b) \quad Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \dots + \varepsilon_i,$$

$$\text{en la que} \quad \beta_1 = f(\bar{X}) - \bar{X}f'(\bar{X}) + \frac{\bar{X}^2}{2!}f''(\bar{X}) - \dots,$$

$$\beta_2 = f'(\bar{X}) - \bar{X}f''(\bar{X}) + \frac{\bar{X}^2}{2!}f'''(\bar{X}) - \dots,$$

y así sucesivamente. Es decir, las β son expresiones en términos de $\bar{X}$ y de los parámetros de $f(X_i)$, y se pueden considerar como coeficientes de regresión paramétricos. Si ahora estimamos una aproximación lineal de (10.71), es decir,

$$(10.72) \quad Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i^*,$$

en realidad estamos omitiendo «variables explicativas» relevantes en la ecuación de regresión. Ya hemos estudiado las consecuencias de este tipo de error en la primera parte de esta sección.

Los resultados anteriores se pueden ilustrar de la forma siguiente. Supongamos que la ecuación de regresión correctamente especificada se puede representar por medio de una función parabólica,

$$(10.73) \quad Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i.$$

Si utilizamos (10.72) como aproximación de (10.73) y empleamos el método de estimación por mínimos cuadrados, podemos obtener las propiedades de los estimadores resultantes por el mismo procedimiento que hemos utilizado en (10.58) y (10.59). El error de especificación cometido al estimar (10.72) en lugar de (10.73) es simplemente la omisión de X_i^2 en la ecuación de regresión. Por lo tanto, tenemos

$$E(\hat{\beta}_1) = \beta_1 + \beta_3 d_{11}, \quad \text{y} \quad E(\hat{\beta}_2) = \beta_2 + \beta_3 d_{21},$$

12. Ver, por ejemplo, A. C. Chiang, *Fundamental Methods of Mathematical Economics*. Nueva York, McGraw-Hill, 1967, pp. 256-60.

en las que d_{31} y d_{32} representan los coeficientes minimocuadráticos de

$$(10.74) \quad X_i^2 = d_{31} + d_{32} X_i + \text{residuo}.$$

En general, d_{32} será distinta de cero, de modo que el estimador de β_2 (y el de β_1) será sesgado e inconsistente. La magnitud del sesgo dependerá del tamaño de β_3 , que determina la curvatura de la ecuación de regresión especificada correctamente, y de los valores de X en la muestra.

Especificación incorrecta del término de perturbación

Para ver otro tipo de error de especificación, consideremos el caso en que la perturbación estocástica se introduce de forma incorrecta en la ecuación de regresión. En concreto, supongamos que la especificación correcta de la ecuación de regresión es

$$(10.75) \quad Y_i = f(X_i)\varepsilon_i,$$

en la que el logaritmo de ε_i se distribuye normalmente con media cero y variancia σ^2 , es decir

$$\log_e \varepsilon_i \sim N(0, \sigma^2).$$

Obsérvese que, en tal caso, tenemos¹³

$$E(\varepsilon_i) = e^{\sigma^2/2},$$

$$\text{Var}(\varepsilon_i) = e^{\sigma^2}(e^{\sigma^2} - 1).$$

Supongamos ahora que en lugar de utilizar (10.75) tomamos

$$(10.76) \quad Y_i = f(X_i) + \varepsilon_i^*.$$

En este caso tenemos

$$\varepsilon_i^* = Y_i - f(X_i) = f(X_i)\varepsilon_i - f(X_i) = f(X_i)(\varepsilon_i - 1).$$

Y_i , en consecuencia, $E(\varepsilon_i^*) = f(X_i)(e^{\sigma^2/2} - 1)$

y $\text{Var}(\varepsilon_i^*) = [f(X_i)]^2[e^{\sigma^2}(e^{\sigma^2} - 1)]$.

Esto significa que ε_i^* no cumple las condiciones de media cero y homoscedasticidad. Además, si el logaritmo de ε_i es una normal, ε_i^* no se puede distribuir

Las implicaciones de estos resultados sobre las propiedades de los estimadores de los parámetros de (10.75) se pueden calcular con detalle aplicando la forma matemática de $f(X_i)$. Por ejemplo, sea la ecuación de regresión correctamente especificada

$$(10.77) \quad Y_i = \gamma X_i \varepsilon_i,$$

en la que $\log_e \varepsilon_i$ cumple todos los supuestos del modelo clásico de regresión normal. Supongamos que, en lugar de (10.77), postulamos

$$(10.78) \quad Y_i = \gamma X_i + \varepsilon_i^*.$$

En este caso, el estimador minimocuadrático de γ basado en la ecuación (10.78) es

$$\hat{\gamma} = \frac{\sum Y_i X_i}{\sum X_i^2} = \frac{\gamma \sum X_i^2 \varepsilon_i}{\sum X_i^2}.$$

La esperanza matemática de $\hat{\gamma}$ es

$$E(\hat{\gamma}) = \gamma e^{\sigma^2/2},$$

lo que significa que $\hat{\gamma}$ es sesgada. Además,

$$\lim \text{prob } \hat{\gamma} = \frac{\lim \text{prob}(\sum X_i^2 \varepsilon_i)/n}{\lim \text{prob}(\sum X_i^2)/n} = \gamma e^{\sigma^2/2},$$

lo que significa que $\hat{\gamma}$ es inconsistente.

En el caso anterior, hemos supuesto que, en la ecuación de regresión especificada correctamente, el término de perturbación aparece en forma multiplicativa, mientras que en la ecuación de regresión mal especificada aparece en forma aditiva. Examinemos ahora el caso contrario. Supongamos que la especificación correcta de la ecuación de regresión es

$$(10.79) \quad Y_i = f(X_i) + \varepsilon_i,$$

en la que ε_i cumple todos los supuestos del modelo clásico de regresión normal, y supongamos que, por error, efectuamos el supuesto

$$(10.80) \quad Y_i = f(X_i)\varepsilon_i^*.$$

En este caso, tenemos

$$\varepsilon_i^* = \frac{Y_i}{f(X_i)} = 1 + \frac{\varepsilon_i}{f(X_i)}.$$

13. Véase Goldberger, *op. cit.*, p. 215.

Y, por tanto, $E(\epsilon_i^*) = 1$

$$y \quad \text{Var}(\epsilon_i^*) = \frac{\sigma^2}{[f(X)]^2}.$$

Para ilustrar este caso, consideremos la ecuación de regresión especificada correctamente en la forma

$$(10.81) \quad Y_i = \alpha X_i^\beta + \epsilon_i,$$

pero estimada en la forma

$$(10.82) \quad \log Y_i = \log \alpha + \beta \log X_i + \epsilon_i^*.$$

El estimador minimocuadrático de β basado en (10.82) es

$$\hat{\beta} = \frac{\sum [\log Y_i - (\sum \log Y_i)/n][\log X_i - (\sum \log X_i)/n]}{\sum [\log X_i - (\sum \log X_i)/n]^2}.$$

Pero, a partir de (10.81), tenemos

$$\log Y_i = \log(\alpha X_i^\beta + \epsilon_i) = \log \alpha + \beta \log X_i + \log [1 + (\epsilon_i/\alpha X_i^\beta)].$$

Por tanto,

$$\hat{\beta} = \beta + \frac{\{\sum [\log X_i - (\sum \log X_i)/n] \log [1 + (\epsilon_i/\alpha X_i^\beta)]\}/n}{\{\sum [\log X_i - (\sum \log X_i)/n]^2\}/n}.$$

Según la expansión en serie de Taylor, podemos escribir¹⁴

$$\begin{aligned} E \log [1 + (\epsilon_i/\alpha X_i^\beta)] &= E[(\epsilon_i/\alpha X_i^\beta) - \frac{1}{2}(\epsilon_i/\alpha X_i^\beta)^2 + \frac{1}{3}(\epsilon_i/\alpha X_i^\beta)^3 - \dots] \\ &= -[\frac{1}{2}E(\epsilon_i^2)/(\alpha X_i^\beta)^2 + \frac{1}{3}E(\epsilon_i^3)/(\alpha X_i^\beta)^3 + \dots]. \end{aligned}$$

Por tanto, vemos que

$$E(\hat{\beta}) \neq \beta$$

y además $\lim \text{prob } \hat{\beta} \neq \beta.$

Así pues, llegamos a la conclusión de que $\hat{\beta}$ es sesgado e inconsistente.

Contrastación de hipótesis sobre los errores de especificación

El análisis que acabamos de efectuar de las consecuencias de las distintas clases de errores de especificación ha mostrado que, salvo en el caso en que

14. Ver, por ejemplo, T. Yamane, *Mathematics for Economists*. Englewood Cliffs, N. J., Prentice-Hall, 1962, p. 170.

se incluyan variables explicativas irrelevantes en la ecuación de regresión, todos los errores de especificación que hemos considerado convertirían a los estimadores minimocuadráticos en sesgados e inconsistentes. En el caso en que se incluyan variables explicativas irrelevantes, los estimadores minimocuadráticos eran insesgados y consistentes, pero no eficientes. Por tanto, es importante que tratemos de evitar los errores de especificación en la medida de lo posible, y que, si son inevitables, seamos por lo menos conscientes de su existencia. Ya hemos mencionado dos razones posibles por las que se cometen errores de especificación. Una de ellas es que las ecuaciones de regresión no pueden estimarse según su especificación correcta debido a las limitaciones de los datos. En tal caso, el análisis anterior adaptado al problema concreto nos puede dar alguna idea de la seriedad del sesgo y de su dirección. La segunda razón por la que se cometen errores de especificación es la falta de conocimiento acerca de la especificación correcta de la ecuación de regresión. En este caso, nos interesaría poder *contrastar* si hemos especificado correctamente o no la ecuación de regresión. Esta es precisamente la forma en que se planteó el problema en relación con los supuestos acerca del término de perturbación en el capítulo 8, al estudiar algunas de las contrastaciones relativas a la homocedasticidad y a la no existencia de autorregresión. En este apartado, vamos a considerar algunas contrastaciones acerca de los errores de especificación correspondientes a la inclusión de variables irrelevantes y a la omisión de variables relevantes; las contrastaciones de la linealidad de la ecuación de regresión se estudiarán en la sección 11.3.

Empecemos con el problema de *contrastar la hipótesis de que la forma en que se ha especificado la ecuación de regresión incluye algunas variables explicativas irrelevantes*. Supongamos que la ecuación de regresión se especifica en la forma

$$Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_K X_{iK} + \beta_{K+1} X_{i,K+1} + \dots + \beta_Q X_{iQ} + \epsilon_i.$$

Supongamos, además, que existe una teoría alternativa según la cual las variables $X_{i,K+1}, X_{i,K+2}, \dots, X_{i,Q}$ son irrelevantes. En este caso, la hipótesis nula se puede expresar en la forma

$$H_0: \beta_{K+1} = \beta_{K+2} = \dots = \beta_Q = 0,$$

y la hipótesis alternativa como

$$H_A: H_0 \text{ no es cierta.}$$

La contrastación adecuada es la que utiliza el estadístico F y se ha presentado en (10.43). Esta contrastación nos permitirá discriminar entre las dos hipótesis. La limitación existente en la aplicación de esta contrastación es que, en general, las teorías no suelen estar formuladas de forma tan defi-

nida como se supone aquí. Frecuentemente, no disponemos de un procedimiento único para clasificar las variables explicativas en dos grupos, uno de ellos conteniendo las variables que deberían incluirse y el otro con las variables que suponemos se pueden excluir. Por ejemplo, si la ecuación de regresión se especifica en la forma

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i,$$

podemos considerar, como alternativas posibles, las ecuaciones de regresión que contienen una o dos de las variables X_{i2} , X_{i3} y X_{i4} como variables explicativas. Si estas tres variables están muy correlacionadas entre sí, no servirá de gran cosa probar todas las regresiones posibles. Lo que puede ocurrir es que, en las regresiones con más de una variable explicativa, *no rechazemos* las hipótesis

$$\beta_2 = 0,$$

$$\beta_3 = 0,$$

$$\beta_4 = 0,$$

pero *rechacemos* las hipótesis

$$\beta_2 = \beta_3 = \beta_4 = 0,$$

$$\beta_2 = \beta_3 = 0,$$

$$\beta_2 = \beta_4 = 0,$$

$$\beta_3 = \beta_4 = 0.$$

Es decir, puede ocurrir que la variación conjunta de las X explique una gran proporción de la variación de Y , pero que no se pueda desentrañar esta variación conjunta para asignar las partes correspondientes a las variables individuales.

El problema anterior es de difícil solución a partir de los datos muestrales existentes solamente. Sin embargo, puede ser posible distinguir entre los modelos de regresión alternativos en base a su actuación respectiva al efectuar pronósticos. Para ilustrar esto, consideremos el siguiente conjunto de modelos alternativos:

Modelo 1: $Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i;$

Modelo 2: $Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i;$

Modelo 3: $Y_i = \beta_1 + \beta_2 X_{i2} + \beta_4 X_{i4} + \varepsilon_i;$

Modelo 4: $Y_i = \beta_1 + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i.$

Supongamos ahora que esperamos observar un valor adicional de Y , por ejemplo Y_{n+1} , para el cual conocemos los valores apropiados de las X . En este caso, podemos construir el intervalo correspondiente al pronóstico de cada uno de los modelos

$$\hat{Y}_{n+1} \pm t s_F,$$

tal como se ha indicado en la ecuación (10.50). El intervalo correspondiente al pronóstico del modelo correcto contendrá el valor de Y_{n+1} con una probabilidad determinada, por ejemplo el 95 %. Tras observar Y_{n+1} , obtendremos uno de los resultados siguientes:

- 1) Ninguno de los intervalos contiene Y_{n+1} .
- 2) Y_{n+1} se halla en el interior del intervalo correspondiente al modelo 1 y no está comprendido en ningún otro intervalo.
- 3) No ocurre ninguno de los dos resultados anteriores.

Si el resultado es el del apartado 1), tenemos que rechazar todos los modelos y buscar otra teoría. Si sucede el resultado del apartado 2), rechazamos los modelos 2, 3 y 4 y hemos acabado la contrastación. Si el resultado es 3), la contrastación no es concluyente. En este caso tenemos que esperar a obtener nuevas observaciones y continuar la contrastación. Si el modelo correcto es el 1, los demás modelos se podrán eliminar a lo largo de este proceso. Si el modelo correcto es, por ejemplo, el modelo 2, los modelos 3 y 4 se podrán eliminar y la discusión quedará limitada a los modelos 1 y 2. Como el modelo correcto es el 2, tanto éste como el modelo 1 nos proporcionarán pronósticos insesgados, pero la variancia del error del pronóstico será mayor para el modelo 1. Sin embargo, hay que señalar que, si los pronósticos se efectúan más de una vez, la contrastación adecuada se efectúa por medio del estadístico F , y tal contrastación ha sido expuesta por otros autores.¹⁵

Consideremos ahora la *contrastación del error de especificación consistente en omitir una variable explicativa relevante en la ecuación de regresión*. Por ejemplo, supongamos que tenemos dos modelos alternativos,

Modelo A: $Y_i = \beta_1 + \beta_2 X_{i2} + \varepsilon_i,$

Modelo B: $Y_i = \beta_1 + \beta_3 X_{i3} + \varepsilon_i.$

Si el modelo A es correcto, β_3 debe ser igual a cero. Para efectuar una contrastación que nos permita discriminar entre estos dos modelos, podemos formular el modelo «general»,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

y contrastar la hipótesis de que $\beta_3 = 0$. Sin embargo, si X_{12} y X_{13} están muy correlacionadas, es muy posible que nos encontremos con que ni $\hat{\beta}_2$ ni $\hat{\beta}_3$ son significativamente distintos de cero, pero el efecto conjunto de X_{12} y X_{13} es decididamente significativo. Cuando esto ocurre, es posible que tengamos que esperar a disponer de observaciones adicionales. Si sólo existe un modelo, el problema de contrastar la existencia de un error de especificación consistente en la omisión de una variable explicativa relevante se hace más difícil. Una posibilidad estriba en esperar a disponer de observaciones adicionales considerando cada una de ellas como evidencia empírica potencialmente contraria al modelo. Si el modelo es correcto, la probabilidad de que un valor futuro de Y se halle fuera del intervalo del pronóstico es pequeña. La dificultad consiste en que es posible que existan otros modelos —que no conocemos— cuyos intervalos de pronóstico se hallen superpuestos al intervalo correspondiente al modelo que estamos contrastando. En este caso, pueden ser necesarias muchas observaciones adicionales antes de que se pueda rechazar este modelo aunque sea incorrecto.

La necesidad de esperar a que existan observaciones adicionales antes de tomar una decisión acerca de la existencia o no de errores de especificación presenta unos inconvenientes obvios. Existe un enfoque alternativo que se apoya en el hecho de que, si el modelo está correctamente especificado, los términos de perturbación de la regresión (y, por tanto, los residuos que representan sus contrapartidas en la muestra) tienen que presentar determinadas propiedades. Concretamente, si omitimos alguna variable explicativa relevante en la ecuación de regresión, la media de la perturbación ya no será cero y su variancia no será constante. Así pues, podemos contrastar la existencia de esta clase de errores de especificación examinando la conducta de los residuos de la regresión. J. B. Ramsey ha desarrollado recientemente varias de estas contrastaciones que pueden utilizarse en este contexto.¹⁶ Dichas contrastaciones precisan de una descripción algo compleja, que no vamos a presentar aquí, pero resultan de gran utilidad y representan un importante avance en el tratamiento del problema de los errores de especificación.

EJERCICIOS:

10-1. A partir de la ecuación de regresión múltiple:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK} + \beta_{K+1} X_{i,K+1} + \cdots + \beta_Q X_{iQ} + e_i.$$

16. Ver James B. Ramsey, «Tests for Specification Errors, in Classical Linear Least Squares Regression Analysis», *Journal of the Royal Statistical Society, serie B*, vol. 31, 1969, pp. 350-71.

Demstrar que

- $SCR_Q \geq SCR_K$.
- $R^2_Q \geq R^2_K$.

10-2. A partir de las definiciones de $\hat{\beta}$, X e Y presentadas en relación a la ecuación (10.15), demostrar que

$$a. \quad \hat{\beta} = (X'X)^{-1}(X'Y).$$

$$b. \quad \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

$$c. \quad Y'Y = Y'Y - \left(\sum_i Y_i\right)^2/n.$$

10-3. Obtener el valor de R^2 a partir de las estimaciones minimocuadráticas

$$Y_i = 5 + 3X_{i1} + 10X_{i2} + e_i \quad (n = 100),$$

(1) (2)

sabiendo que el coeficiente de correlación muestral entre las dos variables explicativas es 0,5.

10-4. A partir de la ecuación $Y = X\beta + \epsilon$, con K (número de variables explicativas incluyendo la constante) igual a 2, mostrar, escribiendo las matrices, que las fórmulas minimocuadráticas correspondientes al modelo de regresión múltiple proporcionan la misma respuesta que las fórmulas correspondientes al modelo de regresión simple obtenidas en la sección 7.3 sin utilizar el álgebra matricial.

10-5. Se ha sugerido que el volumen de inversión de las sociedades se puede explicar mediante la relación

$$I_t = \beta_1 + \beta_2 F_{t-1} + \beta_3 K_{t-1} + e_t,$$

siendo I_t = Inversión bruta actual, F_{t-1} = Valor de las acciones en circulación al final del período anterior, y K_{t-1} = Cantidad de capital al final del período anterior. A partir de los datos del cuadro 10-2, correspondientes a la General Motors Corporation:

- Obtener las estimaciones minimocuadráticas de los coeficientes de regresión y de sus errores standard estimados.
- Calcular el valor de R^2 y efectuar una contrastación relativa a la existencia de relación entre estas variables.
- Los valores de F_{t-1} y K_{t-1} para 1954 son 5 593,6 y 2 226,3, respectivamente. Calcular el valor del pronóstico para I_{1954} y su intervalo al nivel de confianza del 95% y el valor real de I_{1954} fue 1 450,7.

- 10-6. Obtener la fórmula del estimador lineal insesgado óptimo de β_1 , y de su variancia, en el modelo de regresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

- 10-7. Obtener la expresión correspondiente al estadístico F ,

$$F_{3, n-4} = \frac{\text{SCR}/3}{\text{SCE}/(n-4)}$$

en términos de los estadísticos t , en un modelo de regresión con tres variables explicativas (es decir, $K = 4$.)

Cuadro 10-2*

Año	I_t	F_{t-1}	K_{t-1}
1935	317.6	3078.5	2.8
1936	391.8	4661.7	52.6
1937	410.6	5387.1	156.9
1938	257.7	2792.2	209.2
1939	330.8	4313.2	203.4
1940	461.2	4643.9	207.2
1941	512.0	4551.2	255.2
1942	448.0	3244.1	303.7
1943	499.6	4053.7	264.1
1944	547.5	4379.3	201.6
1945	561.2	4840.9	265.0
1946	688.1	4900.9	402.2
1947	568.9	3526.5	761.5
1948	529.2	3254.7	922.4
1949	555.1	3700.2	1020.1
1950	642.9	3755.6	1099.0
1951	755.9	4833.0	1207.7
1952	891.2	4924.9	1430.5
1953	1304.4	6241.7	1777.3

Fuente: J. C. G. Boot y G. M. deWitt, «Investment Demand: An Empirical Contribution to the Aggregation Problem», *International Economic Review*, vol. 1, enero 1960, pp. 3-30.

* Todos los valores (en millones de dólares) se han dividido por los correspondientes índices de precios.

- 10-8. Considerar un modelo de regresión correctamente especificado en la forma

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

en el que se cumplen todos los supuestos del modelo clásico de regresión normal lineal. Supóngase también que no se dispone de observaciones

en X_i , que en su lugar, se utiliza alguna variable no estocástica Z como «sustituto», es decir, se estima

$$Y_i = \alpha + \beta Z_i + \varepsilon_i^*.$$

- ¿Cuáles son las propiedades de las estimaciones minimocuadráticas resultantes de α y β , y de sus variancias estimadas?
- ¿Se ve afectada la validez de la contrastación habitual de la hipótesis de que $\beta = 0$ a consecuencia de la sustitución de X_i por Z_i ? ¿Existiría alguna diferencia si Z_i no fuese un «buen sustituto» de X_i , es decir si X_i y Z_i no estuviesen muy correlacionadas entre sí?

- 10-9. Sea E_i = gasto total semanal planeado de la i -ésima economía doméstica, X_{ig} = gasto de la i -ésima economía doméstica en el bien g , y N_i = número de personas que componen la i -ésima economía doméstica. (La variable N_i es no estocástica.) Considerar el modelo siguiente:

$$X_{ig} = \alpha_g + \beta_g E_i + \gamma_g N_i + u_{ig} \quad (i = 1, 2, \dots, n; g = 1, 2, \dots, G; n > G).$$

Además,

$$E_i = \sum_{g=1}^G X_{ig} + v_i,$$

$$v_i = - \sum_{g=1}^G u_{ig}.$$

Suponemos que

$$E(u_{ig}) = 0,$$

$$E(u_{ig} u'_{ih}) = \sigma_{gg} \mathbf{I}_n,$$

$$E(u_{ig} u'_{ih}) = 0 \quad (g \neq h),$$

en las que

$$u'_g = [u_{1g} \quad u_{2g} \quad \dots \quad u_{ng}].$$

Supóngase que se han obtenido observaciones correspondientes a E_i y X_{ig} , y se ha ajustado la ecuación siguiente por el método de los mínimos cuadrados ordinarios:

$$X_{ig} = a_g + b_g E_i + e_{ig}, \quad g = 1, 2, \dots, G.$$

- ¿Cuáles son las condiciones, si es que las hay, en que b_g será un estimador insesgado de β_g ?
- ¿Cuáles son las condiciones, si es que las hay, en que a_g será un estimador insesgado de α_g ?

Formulación y estimación de modelos especiales

Debido a su flexibilidad, la regresión múltiple constituye un instrumento analítico adecuado para los problemas de la inferencia estadística en las condiciones en que la imposición de determinados conocimientos previos acerca de la ecuación de regresión hace abandonar el modelo ordinario de regresión lineal. En este capítulo estudiaremos cuatro casos especiales. La sección 11.1 se ocupa de la formulación y estimación de las relaciones que incluyen variables cualitativas o binarias. En dicha sección se analizan las conocidas técnicas estadísticas del «análisis de la variancia» y «análisis de la covariancia», las cuales pueden considerarse como modelos de regresión con variables explicativas de tipo cualitativo. En la sección 11.2 se considera el problema de la incorporación de distintas restricciones previas sobre los coeficientes de las relaciones lineales al efectuar su estimación. La sección 11.3 contiene una descripción de algunos modelos no lineales y se ocupa de los problemas de la estimación y de la contrastación de la linealidad. Por último, en la sección 11.4 examinamos algunos modelos en los que la respuesta de la variable dependiente a los cambios en las variables explicativas puede ocurrir con retardos. Estos modelos que se conocen con el nombre de «modelos con retardos distribuidos» han adquirido recientemente un importante papel en muchos sectores de la investigación económica aplicada.

11.1 Modelos con variables binarias

Algunos de los fenómenos que observamos no los podemos medir, y lo único que podemos hacer es contarlos. Esto ocurre con todas las características cualitativas de los objetos, de las personas, de los períodos de tiempo, etc. En estos casos, la observación consiste en señalar si una característica determinada se halla presente o no. Por ejemplo, cuando la unidad de observación es un individuo adulto, podemos observar si es o no propietario de

una casa, si posee o no un título universitario, si fuma o no fuma, o cualquier otra característica que sea relevante para el problema que nos ocupa. Como podemos asignar el valor 1 a la presencia del atributo en cuestión, y el valor 0 a su ausencia, podemos considerarlo como una variable que está limitada a estos dos valores. (Naturalmente, no es preciso que los dos valores sean 0 y 1. Si lo deseamos podemos elegir cualquier otro par de valores para representar la presencia y ausencia de un determinado atributo, pero, por razones obvias, los valores elegidos con mayor frecuencia son 0 y 1.) Este tipo de variables se denominan «binarias» o «ficticias». A continuación, vamos a estudiar el problema de la formulación y estimación de modelos en los que aparecen variables cualitativas en algún lado de la ecuación de regresión, partiendo de los modelos más simples y pasando después a algunos más complejos.

Una variable cualitativa

El modelo de regresión simple en que la variable explicativa es de tipo binario se puede ilustrar con el ejemplo de los salarios que se ofrecen a los estudiantes graduados de economía al entrar en el mercado de trabajo académico. Supongamos que estos salarios se distribuyen normalmente con variancia σ^2 media igual a μ_1 para los candidatos que ya han obtenido su Ph. D.,* y μ_0 para los que no lo han obtenido. Esta situación se puede representar por medio de un modelo de regresión en el que la variable dependiente es el salario y la variable explicativa el grado que se posee. Formalmente,

$$(11.1) \quad Y_i = \alpha + \beta X_i + \varepsilon_i,$$

en la que Y_i es el salario del candidato i y X_i es una variable binaria tal que

$$\begin{aligned} X_i &= 1 && \text{si el candidato posee un Ph. D.,} \\ &= 0 && \text{en caso contrario.} \end{aligned}$$

El término de perturbación ε_i es una variable aleatoria que cumple todos los supuestos básicos del modelo clásico de regresión normal lineal.

Los valores medios de Y_i correspondientes a los dos valores de X_i son

$$\begin{aligned} E(Y_i | X_i = 0) &= \alpha, \\ E(Y_i | X_i = 1) &= \alpha + \beta. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \alpha &= \mu_0, \\ \alpha + \beta &= \mu_1, \\ \text{o} \quad \beta &= \mu_1 - \mu_0. \end{aligned}$$

Esto significa que la ordenada en el origen de la recta de regresión de la población (11.1) mide el salario medio de los Ph. D., mientras que la pendiente mide la diferencia entre el salario medio de los que poseen un Ph. D. y el de los que no lo poseen. La contrastación de la hipótesis de que β es 0 equivale a la contrastación de que no existe diferencia entre el salario medio de los que poseen un Ph. D. y el de los que no lo poseen.

Los coeficientes de la ecuación de regresión (11.1) pueden estimarse por el método de los mínimos cuadrados. Bajo las condiciones de los supuestos del modelo clásico de regresión normal lineal, las estimaciones resultantes poseerán todas las propiedades deseables. Como se recordará, las fórmulas de los estimadores minimocuadráticos son

$$\begin{aligned} \hat{\beta} &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \\ \hat{\alpha} &= \bar{Y} - \hat{\beta} \bar{X}. \end{aligned}$$

Sea n_1 = número de candidatos con Ph. D., en la muestra,
 n_0 = número de candidatos sin Ph. D., en la muestra,
 Y_1 = salario medio muestral de los candidatos con Ph. D.,
 Y_0 = salario medio muestral de los candidatos sin Ph. D.

En este caso,

$$\begin{aligned} \sum_{i=1}^n X_i &= n_1, \\ \sum_{i=1}^n X_i^2 &= n_1, \\ \sum_{i=1}^n Y_i &= n_1 \bar{Y}_1 + n_0 \bar{Y}_0, \\ \sum_{i=1}^n X_i Y_i &= n_1 \bar{Y}_1. \end{aligned}$$

Por tanto,

$$\begin{aligned} \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_i X_i Y_i - \frac{1}{n} (\sum_i X_i) (\sum_i Y_i) \\ &= n_1 \bar{Y}_1 - \frac{n_1}{n} (n_1 \bar{Y}_1 + n_0 \bar{Y}_0) = \frac{n_0 n_1}{n} (\bar{Y}_1 - \bar{Y}_0) \end{aligned}$$

* N. del T.: Título universitario superior en Estados Unidos (Doctor of Philosophy). En muchas materias los otros dos títulos son, por orden decreciente, M. A. (Master of Arts) y B. A. (Bachelor of Arts).

$$y \quad \sum_i (X_i - \bar{X})^2 = \sum_i X_i^2 - \frac{1}{n} (\sum_i X_i)^2 = n_1 - \frac{n_1^2}{n} = \frac{n_0 n_1}{n}.$$

Lo que significa que

$$\beta = \frac{(n_0 n_1 / n)(\bar{Y}_1 - \bar{Y}_0)}{n_0 n_1 / n} = \bar{Y}_1 - \bar{Y}_0$$

$$y \quad \hat{\alpha} = \frac{1}{n} (n_1 \bar{Y}_1 + n_0 \bar{Y}_0) - (\bar{Y}_1 - \bar{Y}_0) \frac{n_1}{n} = \bar{Y}_0.$$

Por tanto, el estimador minimocuadrático de la pendiente de regresión es igual a la diferencia entre el salario medio muestral de quienes poseen un Ph.D. y el de los que no lo poseen, y el estimador minimocuadrático de la ordenada en el origen es igual al salario medio muestral de los que no han obtenido su Ph. D. La contrastación de la hipótesis de que β es igual a cero, basada en la variable t , y presentada en la expresión (7.46) es en este caso exactamente igual a la contrastación de la hipótesis de que las medias de las dos poblaciones son iguales, presentada esta última en la expresión (5.2) (el estadístico de prueba $\hat{\beta}/s_{\hat{\beta}}$ es igual al estadístico de prueba de la ecuación (5.2) sustituyendo σ_1^2 y σ_2^2 por s^2).

En el ejemplo anterior, la característica representada por la variable explicativa presentaba una clasificación dual; es decir, sólo existían dos posibilidades relevantes. Pero podemos utilizar también modelos en los que la característica explicativa presente una clasificación múltiple. La única consecuencia de esta complicación es que necesitaremos más de una variable binaria para describir la característica. Por ejemplo, supongamos que los salarios iniciales de los profesores de inglés de escuelas superiores se distribuyen normalmente con variancia σ^2 y que la media depende de si el título superior obtenido por el candidato es un B.A., un M.A., o un Ph.D. Sea μ_A el salario medio inicial de los B.A.; μ_B el de los M.A. y μ_C el de los Ph.D. La ecuación de regresión correspondiente se puede representar por

$$(11.2) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i,$$

en la que Y_i es el salario del candidato i -ésimo, y

$$\begin{aligned} X_{i2} &= 1 \text{ si el título del candidato es un Ph.D.,} \\ &= 0 \text{ si no lo es;} \\ X_{i3} &= 1 \text{ si el título del candidato es un M.A.,} \\ &= 0 \text{ si no lo es.} \end{aligned}$$

(Observe que si $X_{i2} = 1$, X_{i3} tiene que ser igual a cero, y viceversa. Los valores medios de Y_i correspondientes a los distintos valores de los regresores son

$$E(Y_i | X_{i2} = 1, X_{i3} = 0) = \beta_1 + \beta_2,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 1) = \beta_1 + \beta_3,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 0) = \beta_1.$$

De lo que se deriva que

$$\beta_1 = \mu_A,$$

$$\beta_2 = \mu_C - \mu_A,$$

$$\beta_3 = \mu_B - \mu_A.$$

Este resultado es análogo al obtenido en la clasificación dual de la expresión (11.1).

Obsérvese que los tres apartados del modelo anterior se representan por medio de *dos* variables binarias, cada una de las cuales puede tomar los valores 0 ó 1. En este caso, no sería correcto utilizar *una* variable con tres valores, por ejemplo, 0 para B.A., 1 para M.A., y 2 para Ph.D. Si lo hiciésemos así y formásemos el modelo de regresión en la forma

$$Y_i = \alpha + \beta W_i + \varepsilon_i,$$

en la que W_i es la variable explicativa que puede tomar los valores 0, 1, y 2 tendríamos

$$E(Y_i | W_i = 0) = \alpha,$$

$$E(Y_i | W_i = 1) = \alpha + \beta,$$

$$E(Y_i | W_i = 2) = \alpha + 2\beta.$$

Sin embargo, esto implica que la diferencia entre el salario medio de los M.A. y el de los B.A. es

$$(\alpha + \beta) - \alpha = \beta,$$

y la diferencia entre el salario medio de los Ph.D. y el de los M.A. es

$$(\alpha + 2\beta) - (\alpha + \beta) = \beta.$$

Es decir, si utilizamos una variable que toma los valores 0, 1, y 2 (o cualesquiera otros tres valores equidistantes entre sí), estamos suponiendo en realidad que la diferencia entre el salario de los Ph.D. y el de los M.A. es igual a la que existe entre el salario de estos últimos y el de los B.A., y, a menos que sepamos previamente que esto es así, no resulta justificado efectuar dicho supuesto.

Obsérvese también que los tres apartados no se pueden representar mediante tres variables binarias en lugar de dos (a menos que eliminemos la constante de la ecuación de regresión), ya que, si lo hiciésemos así y formásemos el modelo de regresión en la forma

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i,$$

en la que X_{i2} y X_{i3} se definen como en el caso anterior, y

$$X_{i4} = 1 \text{ si el título del candidato es un B.A.,}$$

$$= 0 \text{ si no lo es.}$$

la solución para $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, y $\hat{\beta}_4$ sería indeterminada. Esto se debe a que

$$X_{i4} = 1 - X_{i2} - X_{i3}$$

y, por tanto, las ecuaciones normales minimocuadráticas no son independientes o (lo que equivale a decir lo mismo) la matriz $X'X$ es singular. Este resultado tiene carácter general: cuando la característica explicativa origina una clasificación en G apartados, utilizaremos $(G-1)$ variables binarias para representarla.

Los modelos con una sola variable cualitativa se han formulado tradicionalmente en los libros de texto de estadística como modelos del «análisis de la variancia para una sola variable» y no como modelos de regresión con regresores binarios. Los dos enfoques son equivalentes en el sentido de que describen el mismo fenómeno y producen los mismos resultados cuando se efectúan las contrastaciones. Consideremos una variable aleatoria Y que se distribuye normalmente, y cuya media depende de cierta característica que da lugar a una clasificación con G apartados. Supongamos que la variancia de Y es constante, y que las observaciones son independientes. Según el enfoque del «análisis de la variancia», dividiremos todos los valores observados de Y en los G grupos, según sus características, y formularemos el modelo en la forma siguiente

$$(11.3) \quad Y_{tg} = \mu + \alpha_g + \varepsilon_{tg} \quad (t = 1, 2, \dots, n_g; g = 1, 2, \dots, G).$$

En esta expresión, Y_{tg} es la observación t de Y en el grupo g , μ es la «media general», α_g es la desviación de la media del grupo g respecto de μ , y ε_{tg} es una perturbación estocástica. Obsérvese que

$$\sum_{g=1}^G \alpha_g = 0.$$

Comparemos ahora el modelo del «análisis de la variancia» con el correspondiente modelo de regresión. Este último viene dado por

$$(11.4) \quad Y_{ig} = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_G X_{iG} + \varepsilon_{ig},$$

en la que $X_{ig} = 1$ si la observación pertenece al grupo g ,
 $= 0$ en caso contrario ($g = 2, 3, \dots, G$).

A fin de especificar los coeficientes de (11.4) en términos de los parámetros de (11.3) y viceversa, compararemos por separado las dos formulaciones para cada grupo (cuadro 11-1).

Cuadro 11-1

Grupo	Modelo del análisis de la variancia	Modelo de regresión
1	$E(Y_{11}) = \mu + \alpha_1$	$E(Y_{11}) = \beta_1$
2	$E(Y_{12}) = \mu + \alpha_2$	$E(Y_{12}) = \beta_1 + \beta_2$
$\vdots$	$\vdots$	$\vdots$
G	$E(Y_{1G}) = \mu + \alpha_G$	$E(Y_{1G}) = \beta_1 + \beta_G$

Es decir,

$$\mu + \alpha_1 = \beta_1,$$

$$\mu + \alpha_2 = \beta_1 + \beta_2,$$

$$\vdots$$

$$\mu + \alpha_G = \beta_1 + \beta_G,$$

$$\mu = \beta_1 + \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G),$$

$$\alpha_1 = -\frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G),$$

$$\alpha_2 = \beta_2 - \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G),$$

$$\vdots$$

$$\alpha_G = \beta_G - \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G).$$

La hipótesis, contrastada con la ayuda del modelo del análisis de la variancia, es que no existe diferencia entre las medias de los grupos; es decir,

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_G = 0,$$

$$H_A: H_0 \text{ no es cierta.}$$

Se puede ver fácilmente que esta hipótesis nula equivale exactamente a la hipótesis de que todas las pendientes de la ecuación de regresión (11.4) son cero: es decir,

$$H_0: \beta_2 = \beta_3 = \dots = \beta_G = 0,$$

$$H_A: H_0 \text{ no es cierta.}$$

Como se ha mostrado anteriormente, para contrastar H_0 en el marco del modelo de regresión, utilizamos la contrastación presentada en la expresión (10.43), que se basa en la distribución F . Dicha contrastación equivale a la contrastación de la F de los modelos del análisis de la variancia que se presenta generalmente en los libros de texto de estadística.¹ El modelo de regresión se puede utilizar también para contrastar la hipótesis de que alguno de los coeficientes β es igual a cero, pero el modelo del análisis de la variancia no se puede adaptar fácilmente para efectuar esta contrastación.

Varias variables cualitativas

Vamos a ampliar ahora la formulación de los modelos de regresión con variables explicativas de tipo cualitativo con el fin de considerar más de una característica. Modifiquemos el ejemplo anterior correspondiente a los salarios de los profesores de inglés en las escuelas superiores, y supongamos que el salario medio depende también, además del título que poseen los candidatos, de si la escuela que efectúa la oferta es pública o privada. Supondremos que las diferencias entre los salarios medios de las escuelas públicas y las privadas son las mismas cualquiera que sea el título del candidato. Igual que en el caso anterior, suponemos que los salarios se distribuyen normalmente con variancia σ^2 y que las observaciones son independientes. Sea

- μ_{A0} = salario medio de los B.A. en las escuelas privadas;
- μ_{A1} = salario medio de los B.A. en las escuelas públicas;
- μ_{B0} = salario medio de los M.A. en las escuelas privadas;
- μ_{B1} = salario medio de los M.A. en las escuelas públicas;
- μ_{C0} = salario medio de los Ph.D. en las escuelas privadas;
- μ_{C1} = salario medio de los Ph.D. en las escuelas públicas.

1. Ver, por ejemplo, J. E. Freund, *Mathematical Statistics*. Englewood Cliffs, Prentice-Hall, 1962, p. 335, fórmula (14.2.5.).

en este caso, se puede formular un modelo de regresión en la forma

$$(11.5) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \gamma Z_i + \epsilon_i,$$

en la que Y_i , X_{i2} , y X_{i3} se definen igual que en (11.2), y

$Z_i = 1$ si la escuela a la que el candidato presenta su solicitud es una escuela pública,
 $= 0$ en caso contrario.

Obsérvese nuevamente que cuando $X_{i2} = 1$, X_{i3} debe ser igual a cero. Los valores medios de Y_i correspondientes a los distintos valores de los regresores son

$$E(Y_i | X_{i2} = 1, X_{i3} = 0, Z_i = 1) = \beta_1 + \beta_2 + \gamma,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 1, Z_i = 1) = \beta_1 + \beta_3 + \gamma,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 0, Z_i = 1) = \beta_1 + \gamma,$$

$$E(Y_i | X_{i2} = 1, X_{i3} = 0, Z_i = 0) = \beta_1 + \beta_2,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 1, Z_i = 0) = \beta_1 + \beta_3,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 0, Z_i = 0) = \beta_1.$$

De lo que se sigue que

$$\beta_1 = \mu_{A0},$$

$$\beta_2 = \mu_{C0} - \mu_{A0} = \mu_{C1} - \mu_{A1},$$

$$\beta_3 = \mu_{B0} - \mu_{A0} = \mu_{B1} - \mu_{A1},$$

$$\gamma = \mu_{A1} - \mu_{A0} = \mu_{B1} - \mu_{B0} = \mu_{C1} - \mu_{C0}.$$

Es decir, β_1 indica el salario medio de los B.A. en las escuelas privadas, β_2 es la diferencia entre el salario medio de los Ph.D. y el de los B.A. (que se supone igual en los dos tipos de escuelas), β_3 mide la diferencia entre los salarios medios de los M.A. y los B.A. (también igual en los dos tipos de escuelas), y γ representa la diferencia entre los salarios medios en las escuelas públicas y las escuelas privadas.

Este modelo de regresión se presenta en los libros de texto de estadística como el modelo del «análisis de la variancia para dos variables». De nuevo, los dos enfoques son equivalentes. Consideremos una variable aleatoria Y cuyos valores se pueden clasificar según dos criterios, uno de los cuales da lugar a una clasificación con G apartados, y el otro a una clasificación con H apartados. Las observaciones de Y se pueden presentar en forma de un cuadro de doble entrada con G columnas y H filas. Cada casilla de este cuadro contiene las ob-

servaciones correspondientes a la fila y columna apropiadas (el modelo del «análisis de la variancia» supone generalmente que el número de observaciones es el mismo en todas las casillas, pero este supuesto no es necesario en la formulación basada en la regresión). El modelo del análisis de la variancia para dos variables se presenta en la forma

$$(11.6) \quad Y_{tgh} = \mu + \alpha_g + \lambda_h + \varepsilon_{tgh} \\ (t = 1, 2, \dots, n_{gh}; g = 1, 2, \dots, G; h = 1, 2, \dots, H).$$

En esta expresión, Y_{tgh} es la observación t de la casilla correspondiente a la columna g y la fila h , μ es la «media general», α_g es la desviación de la media de la columna g respecto de μ , λ_h es la desviación de la media de la fila h respecto de μ , y ε_{tgh} es una perturbación estocástica. Obsérvese que

$$\sum_{g=1}^G \alpha_g = 0, \quad \sum_{h=1}^H \lambda_h = 0.$$

Podemos comparar esta formulación con el correspondiente modelo de regresión:

$$(11.7) \quad Y_{tgh} = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \dots + \beta_G X_{tG} \\ + \gamma_2 Z_{t2} + \gamma_3 Z_{t3} + \dots + \gamma_H Z_{tH} + \varepsilon_{tgh},$$

en el que $X_{tg} = 1$ si la observación pertenece a la columna g ,
 $= 0$ en caso contrario ($g = 2, 3, \dots, G$);
 $Z_{th} = 1$ si la observación pertenece a la fila h ,
 $= 0$ en caso contrario ($h = 2, 3, \dots, H$).

Para obtener la relación entre los coeficientes de (11.6) y los parámetros de (11.7), podemos comparar las medias de Y para cada casilla tal como se indica en el cuadro 11-2.

Cuadro 11-2

Columna	Fila	Modelo del análisis de la variancia	Modelo de regresión
1	1	$E(Y_{111}) = \mu + \alpha_1 + \lambda_1$	$E(Y_{111}) = \beta_1$
1	2	$E(Y_{112}) = \mu + \alpha_1 + \lambda_2$	$E(Y_{112}) = \beta_1 + \gamma_2$
⋮	⋮	⋮	⋮
1	H	$E(Y_{11H}) = \mu + \alpha_1 + \lambda_H$	$E(Y_{11H}) = \beta_1 + \gamma_H$
2	1	$E(Y_{221}) = \mu + \alpha_2 + \lambda_1$	$E(Y_{221}) = \beta_1 + \beta_2$
2	2	$E(Y_{222}) = \mu + \alpha_2 + \lambda_2$	$E(Y_{222}) = \beta_1 + \beta_2 + \gamma_2$
⋮	⋮	⋮	⋮
G	H	$E(Y_{GHH}) = \mu + \alpha_G + \lambda_H$	$E(Y_{GHH}) = \beta_1 + \beta_G + \gamma_H$

A partir de ello, vemos que

$$\mu = \beta_1 + \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G) + \frac{1}{H}(\gamma_2 + \gamma_3 + \dots + \gamma_H), \\ \alpha_1 = -\frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G), \\ \alpha_2 = \beta_2 - \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G), \\ \vdots \\ \alpha_G = \beta_G - \frac{1}{G}(\beta_2 + \beta_3 + \dots + \beta_G), \\ \lambda_1 = -\frac{1}{H}(\gamma_2 + \gamma_3 + \dots + \gamma_H), \\ \lambda_2 = \gamma_2 - \frac{1}{H}(\gamma_2 + \gamma_3 + \dots + \gamma_H), \\ \vdots \\ \lambda_H = \gamma_H - \frac{1}{H}(\gamma_2 + \gamma_3 + \dots + \gamma_H).$$

El modelo del análisis de la variancia para dos variables se utiliza para contrastar las hipótesis

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_G = 0$$

y

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_H = 0.$$

Estas hipótesis tienen sus contrapartidas exactas en términos de los coeficientes del modelo de regresión (11.7). La primera hipótesis equivale a la hipótesis de que $\beta_2, \beta_3, \dots, \beta_G$ son todos cero; es decir,

$$H_0: \beta_2 = \beta_3 = \dots = \beta_G = 0.$$

La segunda hipótesis equivale a la de que $\gamma_2, \gamma_3, \dots, \gamma_H$ son todos cero; es decir

$$H_0: \gamma_2 = \gamma_3 = \dots = \gamma_H = 0.$$

La contrastación adecuada para estas hipótesis es la contrastación de la F presentada en (10.45), y equivale a la contrastación de la F de los modelos del análisis de la variancia que aparece en los libros de texto de estadística. Los modelos (11.6) y (11.7), y su comparación, se pueden generalizar a un número mayor de características explicativas por simple analogía.

Términos de interacción

Los modelos de regresión con dos o más características explicativas se pueden generalizar de forma adicional por medio de la introducción de los «términos de interacción». Consideremos el ejemplo anterior de los salarios iniciales de los profesores de inglés en las escuelas superiores. En dicho ejemplo se efectuó el supuesto de que el salario medio depende del título que poseen los candidatos y del tipo de escuela, y que la diferencia entre los salarios medios en las escuelas públicas y las privadas es la misma para todos los títulos. Supongamos ahora que no queremos efectuar este último supuesto. En tal caso, el modelo de regresión (11.5) se puede modificar de la forma siguiente:

$$(11.8) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \gamma Z_i + \delta_2 X_{i2} Z_i + \delta_3 X_{i3} Z_i + \epsilon_i,$$

en el que todas las variables se definen como en el modelo (11.5). Los valores medios de Y_i correspondientes a los distintos valores de los regresores son

$$E(Y_i | X_{i2} = 1, X_{i3} = 0, Z_i = 1) = \beta_1 + \beta_2 + \gamma + \delta_2,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 1, Z_i = 1) = \beta_1 + \beta_3 + \gamma + \delta_3,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 0, Z_i = 1) = \beta_1 + \gamma,$$

$$E(Y_i | X_{i2} = 1, X_{i3} = 0, Z_i = 0) = \beta_1 + \beta_2,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 1, Z_i = 0) = \beta_1 + \beta_3,$$

$$E(Y_i | X_{i2} = 0, X_{i3} = 0, Z_i = 0) = \beta_1.$$

Esto significa que podemos definir los coeficientes de regresión en términos de los salarios medios en la forma siguiente:

$$\beta_1 = \mu_{A0},$$

$$\beta_2 = \mu_{C0} - \mu_{A0},$$

$$\beta_3 = \mu_{H0} - \mu_{A0},$$

$$\gamma = \mu_{A1} - \mu_{A0},$$

$$\delta_2 = (\mu_{C1} - \mu_{C0}) - (\mu_{A1} - \mu_{A0}),$$

$$\delta_3 = (\mu_{H1} - \mu_{H0}) - (\mu_{A1} - \mu_{A0}).$$

Las diferencias entre los salarios medios de las escuelas públicas y los de las privadas son

$$\text{B.A.: } \mu_{A1} - \mu_{A0} = \gamma,$$

$$\text{M.A.: } \mu_{H1} - \mu_{H0} = \gamma + \delta_3,$$

$$\text{Ph.D.: } \mu_{C1} - \mu_{C0} = \gamma + \delta_2.$$

El modelo de regresión (11.8) equivale al modelo presentado en la literatura estadística como «análisis de la variancia para dos variables con interacción». Siguiendo el mismo camino que en los dos casos analizados previamente, se podría demostrar la equivalencia entre estos dos modelos, pero no vamos a efectuarlo aquí.

Variables explicativas cuantitativas y cualitativas

En los modelos anteriores todos los regresores que hemos incluido estaban representados por variables binarias, pero estos modelos no son realmente muy frecuentes en economía. Más a menudo nos encontramos con modelos en los que *algunos* regresores son binarios pero otros no. Un ejemplo tradicional es la función de consumo estimada a partir de datos de series temporales que comprenden un importante período de guerra. En este modelo se supone que el consumo medio depende de la renta y de si se trata de un período de paz o de guerra. Una forma sencilla de representar este modelo es

$$(11.9) \quad C_t = \beta_1 + \beta_2 Y_t + \gamma Z_t + \epsilon_t,$$

en la que C representa el consumo, Y representa la renta y Z es una variable binaria tal que

$$Z_t = 1 \quad \text{si } t \text{ es un período de guerra,}$$

$$= 0 \quad \text{en caso contrario.}$$

En tal caso, tenemos

$$C_t = (\beta_1 + \gamma) + \beta_2 Y_t + \epsilon_t \quad (\text{período de guerra}),$$

$$C_t = \beta_1 + \beta_2 Y_t + \epsilon_t \quad (\text{período de paz}).$$

pues, en realidad estamos postulando que en los períodos de guerra, la renta en el origen de la función de consumo pasa de β_1 a $\beta_1 + \gamma$. En la

figura 11-1 se representa gráficamente dicho modelo. Si se considera que la

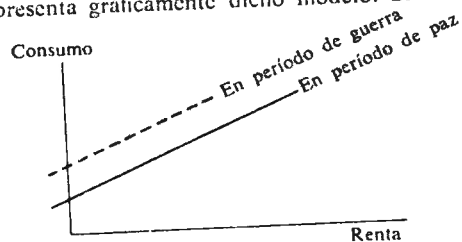


Figura 11-1

ordenada en el origen representa el «nivel de subsistencia» de consumo, este modelo implica que dicho nivel de subsistencia varía durante la guerra. Este cambio se puede presentar en forma de hipótesis a contrastar, es decir,

$$H_0: \gamma = 0,$$

$$H_A: \gamma \neq 0.$$

Los libros de estadística presentan una formulación equivalente de los modelos del tipo del (11.9) con el nombre de «análisis de la covariancia».

El efecto de la guerra sobre la función de consumo se puede expresar de forma distinta si postulamos que las condiciones de guerra afectan a la pendiente y no a la ordenada en el origen de la función. Con esta formulación teórica, el modelo de regresión es

$$(11.10) \quad C_t = \beta_1 + \beta_2 Y_t + \delta Y_t Z_t + \varepsilon_t,$$

en el que las variables se definen igual que antes. En este caso, tenemos

$$C_t = \beta_1 + (\beta_2 + \delta) Y_t + \varepsilon_t \quad (\text{en periodo de guerra}),$$

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t \quad (\text{en periodo de paz}).$$

La ecuación (11.10) implica que el efecto de la guerra consiste en cambiar la propensión marginal al consumo, como se indica en la figura 11-2. Esta implicación puede probarse contrastando la hipótesis de que δ es cero.

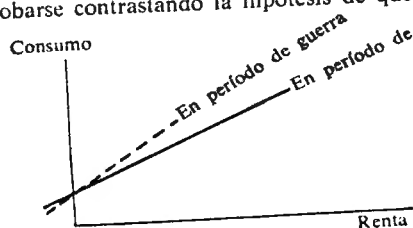


Figura 11-2

La tercera y última posibilidad para distinguir entre las observaciones correspondientes a tiempo de paz y a tiempo de guerra consiste en suponer que en el período de guerra varían la ordenada en el origen y la pendiente de la función de consumo. En este caso la ecuación de regresión sería

$$(11.11) \quad C_t = \beta_1 + \beta_2 Y_t + \gamma Z_t + \delta Y_t Z_t + \varepsilon_t.$$

Y tendríamos

$$C_t = (\beta_1 + \gamma) + (\beta_2 + \delta) Y_t + \varepsilon_t \quad (\text{en periodo de paz}).$$

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t \quad (\text{en periodo de guerra}).$$

Estas relaciones se representan en la figura 11-3. El punto a señalar en relación a la ecuación (11.11) es que los estimadores minimocuadráticos de los coeficientes de regresión son exactamente iguales a los que se obtendrían mediante dos regresiones separadas de C_t respecto a Y_t , una a partir de las observaciones en período de paz y la otra con las observaciones del período de guerra. La demostración de esta afirmación puede obtenerse por medio de una aplicación directa de las fórmulas minimocuadráticas, y no la vamos a presentar aquí. La única diferencia entre los dos enfoques corresponde a σ^2 . Si, como suponemos normalmente, la variancia de ε_t permanece inalterada a lo largo de todo el período, su estimación a partir de la ecuación (11.11), basada en *todas* las observaciones, será eficiente, pero las dos estimaciones obtenidas a partir de las dos submuestras no lo serán. Esto se debe a que la estimación de σ^2 basada en cada una de las submuestras no utiliza la información acerca de esta variancia, contenida en la otra submuestra.

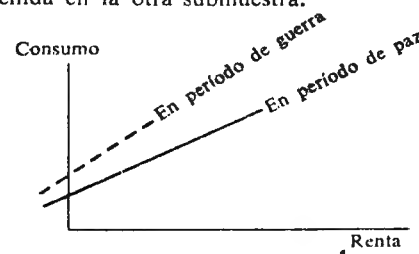


Figura 11-3

Otra aplicación de los modelos con variables explicativas cuantitativas y cualitativas aparece al contrastar la hipótesis de respuesta *asimétrica*. Según esta hipótesis, la variación de $E(Y_t)$ a consecuencia de un *aumento* de una unidad en una variable explicativa determinada es distinta en valor absoluto a la variación debida a una *disminución* de una unidad en la misma variable explicativa. Este tipo de conducta está implícito en las afirmaciones del tipo

de «los salarios son rígidos a la baja», o cuando se señala que los consumidores responden con mayor prontitud a un aumento de la renta que a una disminución. La formulación habitual de la ecuación de regresión lineal resultaría inadecuada en tales casos, pero podemos incorporar esta respuesta asimétrica de $E(Y_t)$ a las variaciones en la variable explicativa introduciendo una variable binaria. Consideremos el modelo de regresión

$$Y_t = \alpha + \beta X_t + \varepsilon_t,$$

que cumple todos los supuestos básicos. En este modelo, un aumento unitario de X_t produce un cambio β en $E(Y_t)$, y una disminución unitaria en X_t produce un cambio igual a $-\beta$. Si queremos que la respuesta de $E(Y_t)$ sea asimétrica, podemos escribir

$$(11.12) \quad Y_t = \alpha + \beta X_t + \gamma X_t Z_t + \varepsilon_t$$

en la que

$$Z_t = 1 \text{ si } X_t \leq X_{t-1},$$

$$= 0 \text{ en caso contrario.}$$

La contrastación de la hipótesis de respuesta asimétrica equivale a contrastar la hipótesis de que $\gamma = 0$. Obsérvese que el modelo de respuesta asimétrica puede utilizarse no sólo en las series temporales, sino también con los datos de corte transversal, a condición de que se posea información acerca del valor de la variable explicativa relevante en el período anterior al de la encuesta.

Los modelos con variables explicativas cualitativas y cuantitativas se utilizan frecuentemente en relación a los modelos de regresión que incorporan efectos de carácter estacional. Estos modelos se estiman generalmente a partir de observaciones mensuales o trimestrales. Por ejemplo, consideremos las siguientes regresiones simples:

$$(11.13) \quad Y_t = \alpha_1 + \beta X_t + \varepsilon_t \quad (\text{primer trimestre}),$$

$$Y_t = \alpha_2 + \beta X_t + \varepsilon_t \quad (\text{segundo trimestre}),$$

$$Y_t = \alpha_3 + \beta X_t + \varepsilon_t \quad (\text{tercer trimestre}),$$

$$Y_t = \alpha_4 + \beta X_t + \varepsilon_t \quad (\text{cuarto trimestre}).$$

En este caso, suponemos que los efectos estacionales desplazan la ordenada en el origen de la función de regresión. El modelo puede expresarse también mediante una sola ecuación de regresión ampliada con la introducción de regresores binarios para representar los factores estacionales:

$$(11.14) \quad Y_t = \alpha + \beta X_t + \gamma_2 Q_{12} + \gamma_3 Q_{13} + \gamma_4 Q_{14} + \varepsilon_t,$$

la que

$$Q_{12} = 1 \text{ si } t \text{ es un segundo trimestre,} \\ = 0 \text{ si no lo es;}$$

$$Q_{13} = 1 \text{ si } t \text{ es un tercer trimestre,} \\ = 0 \text{ si no lo es;}$$

$$Q_{14} = 1 \text{ si } t \text{ es un cuarto trimestre,} \\ = 0 \text{ si no lo es.}$$

Obsérvese que los efectos estacionales de los trimestres están representados por tres regresores binarios y no por cuatro; de no ser así, los estimadores mínimo-cuadráticos de los coeficientes de regresión serían indeterminados. Comparando (11.14) y (11.13) podemos ver que

$$\alpha = \alpha_1,$$

$$\alpha + \gamma_2 = \alpha_2,$$

$$\alpha + \gamma_3 = \alpha_3,$$

$$\alpha + \gamma_4 = \alpha_4.$$

Para examinar la importancia de los efectos estacionales en los modelos de regresión del tipo del (11.14), contrastamos la hipótesis de que γ_2 , γ_3 , y γ_4 son todos cero; es decir,

$$H_0: \gamma_2 = \gamma_3 = \gamma_4 = 0,$$

$$H_A: H_0 \text{ no es cierta.}$$

En este caso, la contrastación adecuada es la contrastación de la F presentada en (10.45).

Un procedimiento alternativo para introducir los factores estacionales en el modelo de regresión es

$$(11.15) \quad Y_t = \alpha + \beta_1 X_t + \varepsilon_t \quad (\text{primer trimestre}),$$

$$Y_t = \alpha + \beta_2 X_t + \varepsilon_t \quad (\text{segundo trimestre}),$$

$$Y_t = \alpha + \beta_3 X_t + \varepsilon_t \quad (\text{tercer trimestre}),$$

$$Y_t = \alpha + \beta_4 X_t + \varepsilon_t \quad (\text{cuarto trimestre}).$$

En este caso, suponemos que los factores estacionales afectan a la pendiente de la ecuación de regresión, pero no a su ordenada en el origen. El modelo (11.15) se puede representar en forma compacta como

$$(11.16) \quad Y_t = \alpha + \beta X_t + \delta_2 Q_{12} X_t + \delta_3 Q_{13} X_t + \delta_4 Q_{14} X_t + \varepsilon_t,$$

en la que las Q se definen igual que en (11.14). Comparando (11.16) y (11.15) podemos ver que

$$\begin{aligned}\beta &= \beta_1, \\ \beta + \delta_2 &= \beta_2, \\ \beta + \delta_3 &= \beta_3, \\ \beta + \delta_4 &= \beta_4.\end{aligned}$$

La importancia de los efectos estacionales se puede estudiar contrastando la hipótesis de que δ_2 , δ_3 , y δ_4 son todas cero; es decir,

$$H_0: \delta_2 = \delta_3 = \delta_4 = 0,$$

$$H_A: H_0 \text{ no es cierta.}$$

Como última posibilidad, podríamos suponer que los factores estacionales afectan a la vez a la ordenada en el origen y a la pendiente de la ecuación de regresión. Desde el punto de vista de la estimación de los coeficientes de regresión, esta formulación equivale a dividir la muestra en cuatro submuestras, como se ha estudiado anteriormente en el caso de la ecuación (11.11).

Nota acerca de la utilización de datos desestacionalizados

Siguiendo con el tema de los factores estacionales en las ecuaciones de regresión, podemos analizar las implicaciones del uso de datos «desestacionalizados», procedentes de series temporales, en el análisis de regresión. Un procedimiento común consiste en formar la ecuación de regresión en la forma

$$\tilde{Y}_t = \alpha + \beta \tilde{X}_t + \epsilon_t,$$

en la que Y y X son los valores «desestacionalizados» de X e Y . La «desestacionalización» consiste en eliminar los movimientos oscilatorios regulares de período anual en las series temporales originales. Esto lo puede realizar el organismo estadístico que facilita los datos, o el propio econométra. En el análisis supondremos que la desestacionalización ha tenido éxito, en el sentido de que los elementos estacionales se han eliminado por completo de las series. Estos elementos estacionales se consideran generalmente aditivos o multiplicativos, y en el primer caso podemos escribir

$$Y_t = \tilde{Y}_t + D_{tY},$$

$$X_t = \tilde{X}_t + D_{tX},$$

en las que D_{tY} y D_{tX} representan las desviaciones estacionales existentes en las series respectivas. Obsérvese que, por ejemplo, en los datos trimestrales,

$$D_{tY} = D_{t+4,Y},$$

$$D_{tX} = D_{t+4,X},$$

$$y \quad D_{tY} + D_{t+1,Y} + D_{t+2,Y} + D_{t+3,Y} = 0,$$

$$D_{tX} + D_{t+1,X} + D_{t+2,X} + D_{t+3,X} = 0,$$

para todos los valores de t . Si alguna serie no contiene elementos estacionales, los valores correspondientes de las D serán cero. La ecuación de regresión

$$\tilde{Y}_t = \alpha + \beta \tilde{X}_t + \epsilon_t$$

implica que

$$(11.17) \quad (Y_t - D_{tY}) = \alpha + \beta(X_t - D_{tX}) + \epsilon_t$$

o

$$(11.17a) \quad Y_t = (\alpha + D_{tY} - \beta D_{tX}) + \beta X_t + \epsilon_t.$$

Es decir, en este caso suponemos que los factores estacionales operan a través de un desplazamiento de la ordenada en el origen de la función de regresión, lo cual resulta equivalente a la formulación presentada en (11.13) o (11.14).

En el caso en que los elementos estacionales son de carácter multiplicativo, podemos escribir

$$Y_t = \tilde{Y}_t S_{tY} \quad y \quad X_t = \tilde{X}_t S_{tX},$$

en las que S_{tY} y S_{tX} representan los índices estacionales correspondientes a las series respectivas. Obsérvese que, por ejemplo, en los datos trimestrales,

$$S_{tY} = S_{t+4,Y} \quad y \quad S_{tX} = S_{t+4,X},$$

y

$$S_{tY} \times S_{t+1,Y} \times S_{t+2,Y} \times S_{t+3,Y} = 1.$$

$$S_{tX} \times S_{t+1,X} \times S_{t+2,X} \times S_{t+3,X} = 1$$

para todos los valores de t . Si alguna serie no contiene elementos estacionales, los valores correspondientes de las S serán la unidad. La ecuación de regresión

$$\tilde{Y}_t = \alpha + \beta \tilde{X}_t + \epsilon_t$$

implica en este caso

$$(11.18) \quad \frac{Y_i}{S_{iY}} = \alpha + \beta \frac{X_i}{S_{iX}} + \epsilon_i$$

o

$$(11.18a) \quad Y_i = \alpha S_{iY} + \beta \frac{S_{iY}}{S_{iX}} X_i + \epsilon_i^*,$$

en la que

$$\epsilon_i^* = S_{iY} \epsilon_i.$$

En esta ocasión, suponemos que los elementos estacionales afectan a la vez a la ordenada en el origen y a la pendiente de la ecuación de regresión, y que el término de perturbación es heteroscedástico, con variancia igual a $\sigma^2 S_{iY}^2$. Cuando los investigadores aplican el método de los mínimos cuadrados a los datos desestacionalizados y efectúan a la vez los supuestos básicos habituales acerca del término de perturbación como en el caso de (11.18), esto tiene ciertas implicaciones para la relación existente entre X e Y en el mundo real, «no desestacionalizado». Estas implicaciones aparecen detalladas en (11.18a), y si esta ecuación no proporciona una descripción correcta de la relación existente entre X e Y , los estimadores minimocuadráticos de α y β basados en los datos desestacionalizados no poseerán, en general, las propiedades deseables que tendrían en caso contrario.

Variable dependiente binaria

Vamos a dedicar ahora nuestra atención a los modelos de regresión en los que la variable *dependiente* es binaria, mientras que las variables explicativas son cuantitativas. A modo de ejemplo, consideremos un modelo de regresión que trata de explicar la posesión o no posesión de determinado tipo de electrodoméstico —por ejemplo, un lavavajillas—. Supongamos que la ecuación de regresión que se postula es

$$(11.19) \quad Y_i = \alpha + \beta X_i + \epsilon_i,$$

en la que X_i representa la renta de la familia i -ésima, e Y_i es una variable binaria tal que

$$Y_i = 1 \text{ si la familia } i\text{-ésima posee un lavavajillas,} \\ = 0 \text{ en caso contrario.}$$

Suponemos también que la variable explicativa X_i es no estocástica, o que si es estocástica es independiente de ϵ_i . El término de perturbación ϵ_i es una variable aleatoria con media cero, y es, además, independiente de ϵ_j ($i \neq j$)

Como Y_i sólo puede tomar los valores 0 y 1, según la definición de esperanza matemática presentada en (3.23), tenemos

$$E(Y_i) = 1 \times f_i(1) + 0 \times f_i(0) = f_i(1),$$

en la que $f_i(1)$ es la probabilidad de que una familia cuya renta es X_i posea un lavavajillas. Obsérvese que, debido a que, según la ecuación (11.19),

$$E(Y_i) = \alpha + \beta X_i,$$

suponemos que la probabilidad $f_i(1)$ es distinta para los distintos niveles de renta. Por tanto, podemos interpretar $E(Y_i)$ como una medida de la proporción de familias con renta X_i que poseen lavavajillas. Esto implica que

$$0 \leq \alpha + \beta X_i \leq 1.$$

Consideremos ahora el término de perturbación ϵ_i . Según la ecuación (11.19),

$$\epsilon_i = Y_i - \alpha - \beta X_i,$$

y, por tanto, como Y_i sólo puede tomar los valores 0 y 1, a cada nivel de renta X_i la perturbación puede tomar únicamente los dos valores ($-\alpha - \beta X_i$) o $(1 - \alpha - \beta X_i)$. Esto significa que ϵ_i no se distribuye normalmente, sino que posee una distribución discreta que se define de la forma siguiente

$$\frac{\epsilon_i}{-\alpha - \beta X_i} \quad \frac{f(\epsilon_i)}{f}$$

$$1 - \alpha - \beta X_i \quad \frac{1 - f}{1}$$

Las probabilidades f y $(1-f)$ se pueden calcular utilizando el supuesto de que $E(\epsilon_i) = 0$. Esto significa que

$$(-\alpha - \beta X_i)f + (1 - \alpha - \beta X_i)(1 - f) = 0,$$

lo que indica que

$$f = 1 - \alpha - \beta X_i.$$

Por tanto, la variancia de ϵ_i es

$$(11.20) \quad E(\epsilon_i^2) = (-\alpha - \beta X_i)^2(1 - \alpha - \beta X_i) + (1 - \alpha - \beta X_i)^2(\alpha + \beta X_i) \\ = (\alpha + \beta X_i)(1 - \alpha - \beta X_i) = E(Y_i)[1 - E(Y_i)].$$

Esto significa que ϵ_i es heteroscedástica, ya que su variancia depende de $E(Y_i)$. Debido a la especial naturaleza de la variable dependiente del modelo

(11.19). existen ciertos problemas en la estimación y predicción. El primero de ellos es que, a causa de la naturaleza heteroscedástica del término de perturbación, los estimadores minimocuadráticos de α y β serán insesgados pero no eficientes. Este problema se puede resolver, al menos para muestras grandes, utilizando las fórmulas de estimación (8.1) y (8.3), que están especialmente concebidas para los modelos heteroscedásticos, e incluyen las variancias de ϵ_i , que no conocemos. En nuestro caso, podemos obtener estimaciones consistentes de estas variancias por medio de

$$(11.21) \quad \hat{\sigma}_i^2 = \hat{Y}_i(1 - \hat{Y}_i),$$

en la que $\hat{Y}_i$ es el valor de Y_i ajustado por mínimos cuadrados. El segundo problema se refiere a la distribución de los estimadores de α y β . Como ϵ_i no se distribuye normalmente, los estimadores de α y β tampoco tienen esta distribución y, en consecuencia, no podemos aplicar las contrastaciones de significación usuales. En el caso de que queramos contrastar una hipótesis acerca de α o β , será preciso obtener la región de aceptación a partir de la distribución conocida de ϵ_i . Sin embargo, no hay ninguna dificultad en el cálculo de las medias y variancias asintóticas de los estimadores de α y β : las medias asintóticas son iguales a los valores verdaderos y las variancias asintóticas se pueden obtener a partir de las fórmulas (8.2) y (8.4). El tercero y último de los problemas se refiere a la predicción. Como $E(Y_i)$ se interpreta como una probabilidad, su valor queda limitado al intervalo entre 0 y 1. Sin embargo, el valor correspondiente a la predicción de Y es un punto sobre una línea recta y, por tanto, su valor puede hallarse entre $-\infty$ y $+\infty$ (a menos que la recta de regresión de la muestra sea perfectamente horizontal, claro está). Existen varios procedimientos para superar dicha dificultad. El más sencillo de ellos es quizás el siguiente: sean $\hat{\alpha}$ y $\hat{\beta}$ las estimaciones de α y β , y además,

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i.$$

Utilicemos ahora como predictor del valor de Y_0 correspondiente a un valor determinado de X (por ejemplo, X_0) el valor $\hat{Y}_0$ definido de la forma siguiente:

$$\hat{Y}_0 = \hat{Y}_0 \quad \text{si} \quad 0 < \hat{Y}_0 < 1,$$

$$= 1 \quad \text{si} \quad \hat{Y}_0 \geq 1,$$

$$= 0 \quad \text{si} \quad \hat{Y}_0 \leq 0,$$

Este método de predicción se representa gráficamente en la figura 11-4.

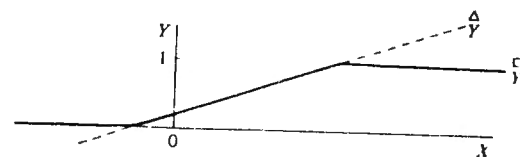


Figura 11-4

Naturalmente, existen ciertas dificultades para efectuar afirmaciones probabilísticas exactas acerca de Y_0 , ya que la distribución de Y_0 no es determinada.²

Variables explicativas y dependiente, cualitativas

Consideremos ahora los modelos en que tanto la variable dependiente como las variables explicativas son de tipo cualitativo. En el caso anterior en que la variable dependiente era cualitativa, partimos del supuesto de que la característica en cuestión correspondía a una clasificación dual. Eliminemos ahora esta restricción, permitiendo la posibilidad de que la característica dependiente presente una clasificación múltiple. Sin embargo, vamos a limitar el análisis al caso en que sólo existe una característica explicativa. Un ejemplo de este caso sería un modelo en el que se considera que la renta —clasificada en grandes grupos tales como baja, media y elevada— depende del nivel de educación. Los resultados muestrales de un modelo de este tipo se presentan generalmente en forma de las llamadas *tablas de contingencia* (ver cuadro 11-3). Estas con-

Cuadro 11-3

		Característica explicativa (X)						
		1	2	...	h	...	H	
Característica dependiente (Y)	1	n_{11}	n_{12}	...	n_{1h}	...	n_{1H}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2h}	...	n_{2H}	$n_{2.}$
	...	...	...	...	...	...	...	...
	g	n_{g1}	n_{g2}	...	n_{gh}	...	n_{gH}	$n_{g.}$
	...	...	...	...	...	...	...	...
	G	n_{G1}	n_{G2}	...	n_{Gh}	...	n_{GH}	$n_{G.}$
		$n_{.1}$	$n_{.2}$	...	$n_{.h}$	...	$n_{.H}$	n

2. Un enfoque alternativo del problema correspondiente al caso en que la variable dependiente es binaria lo proporciona el *modelo del análisis «probit»*. Ver, por ejemplo, J. J. Bin, «Estimation of relationships for limited dependent variables», *Econometrica*, enero 1958, pp. 24-36.

sisten en cuadros de doble entrada en los que las filas representan las clasificaciones basadas, por ejemplo, en la característica dependiente, mientras que las columnas corresponden a las clasificaciones según la característica explicativa. El número que aparece en cada casilla es el número de observaciones correspondiente a la fila y columna de que se trate. Las anotaciones en el centro del cuadro representan el número de observaciones en cada casilla y las anotaciones en el margen son los totales de cada fila o columna.

La razón por la que se construye una tabla de contingencia es la opinión de que las dos características están relacionadas, y para contrastar esta proposición es preciso examinar los datos muestrales para ver si refutan o no dicha proposición. El modelo de regresión no se adapta con facilidad a esta contrastación, pero existe una *contrastación de la ji-cuadrado de las tablas de contingencia* diseñada precisamente con tal propósito. En esta contrastación vamos a utilizar un estadístico de prueba que se construye comparando el número real de observaciones, o frecuencias, de cada casilla con el número que obtendríamos si no existiese relación entre X e Y , llamado normalmente frecuencia «teórica». Las frecuencias «teóricas» se obtienen en base a la «independencia en sentido probabilístico» definida en el teorema 5 de la sección 3.3. Según este teorema, si dos sucesos son independientes, su probabilidad conjunta es igual al producto de las probabilidades marginales respectivas. Esto significa que si X e Y son independientes, la probabilidad de que una observación obtenida al azar pertenezca a la fila g y a la columna h es

$$P(g \cap h) = P(g)P(h).$$

En esta expresión, $P(g)$ es la probabilidad de que una observación obtenida al azar pertenezca a la fila g , y $P(h)$ es la probabilidad de que una observación obtenida al azar pertenezca a la columna h . Por tanto, si conocemos $P(g)$ y $P(h)$, podremos obtener el valor de $P(g \cap h)$ cuando X e Y son independientes. Como no conocemos los valores $P(g)$ y $P(h)$, los estimaremos por medio de las proporciones muestrales respectivas, es decir

$$\text{Est. } P(g) = \frac{n_{g.}}{n},$$

$$\text{Est. } P(h) = \frac{n_{.h}}{n}.$$

Representemos ahora la frecuencia «teórica» deseada de la casilla correspondiente a la fila g y a la columna h por e_{gh} . En este caso, podemos escribir

$$(11.22) \quad \frac{e_{gh}}{n} = \text{Est. } P(g \cap h) = \text{Est. } P(g) \times \text{Est. } P(h) = \frac{n_{g.}}{n} \times \frac{n_{.h}}{n},$$

o

$$(11.22a) \quad e_{gh} = \frac{n_{g.} \times n_{.h}}{n}.$$

Se puede presentar una justificación intuitiva de la expresión (11.22) en la forma siguiente. Supongamos que Y es la renta y X es la educación, y supongamos también que la fila g representa el grupo de «renta media» y la columna h el grupo con «título de escuela superior». Si no existiese ninguna relación entre la educación y la renta, sería de esperar que la proporción de perceptores de renta media entre los poseedores de un título de escuela superior fuese la misma que la proporción de perceptores de renta media en la población total. Utilizando las proporciones muestrales en lugar de las de la población, tenemos

$$\frac{e_{gh}}{n_{.h}} = \frac{n_{g.}}{n},$$

en la que $n_{.h}$ representa el número total de titulados de escuela superior y $n_{g.}$ el número total de perceptores de renta media. Esta ecuación equivale a las (11.22) o (11.22a).

La contrastación de la ji-cuadrado en la tabla de contingencia es en realidad una contrastación de la hipótesis de que *no* existe relación entre X e Y . Si la hipótesis es cierta, la razón principal de la diferencia entre las frecuencias reales y las «teóricas» serán los errores debidos al muestreo. El estadístico que se utiliza en la contrastación es

$$\sum_{g=1}^G \sum_{h=1}^H \frac{(n_{gh} - e_{gh})^2}{e_{gh}}.$$

Puede mostrarse que dicho estadístico presenta una distribución que, en las muestras grandes, se puede aproximar a través de una distribución ji-cuadrado.³ Concretamente, si la muestra es grande, tenemos

$$(11.23) \quad \sum_{g=1}^G \sum_{h=1}^H \frac{(n_{gh} - e_{gh})^2}{e_{gh}} \sim \chi_{(G-1) \times (H-1)}^2,$$

en la que el subíndice $(G-1) \times (H-1)$ se refiere al número de grados de libertad correspondiente. Este número está determinado por el hecho de que, dados los totales correspondientes a las filas y columnas, sólo podemos rellenar a voluntad $(G-1) \times (H-1)$ casillas; una vez hecho esto, los números de las restantes casillas quedan automáticamente determinados. Por lo que respecta al requisito de un tamaño muestral «grande», el criterio utilizado en la práctica es que ninguna de las frecuencias «teóricas» sea menor que 5. En consecuencia, a veces resulta necesario agrupar varias casillas para hacer que el número de frecuencias satisfaga el mínimo requerido. Cuando el valor del estadístico de prueba, calculado a partir de las observaciones muestrales, es mayor que el valor de la ji-cuadrado correspondiente al nivel de significación elegido

3. Ver, por ejemplo, A. M. Mood y F. A. Graybill, *Introduction to the Theory of Statistics*, Nueva York, McGraw-Hill, 1963, p. 318.

que obtenemos en las tablas, se rechaza la hipótesis nula. Obsérvese que la evidencia en contra de la hipótesis nula sólo puede tomar la forma de un valor grande del estadístico de prueba calculado y, por tanto, la contrastación adecuada es la de *un extremo*. La contrastación de la ji-cuadrado en las tablas de contingencia es muy parecida a la contrastación de la «bondad del ajuste», por medio de la ji-cuadrado, estudiada al final de la sección 5.2.

11.2 Modelos con restricciones en los coeficientes

El problema que vamos a tratar aquí es el de cómo incorporar al proceso de estimación la información previa de que dispongamos acerca del coeficiente o coeficientes de regresión. Podemos considerar que esta información introduce ciertas restricciones en uno o más de los coeficientes de la regresión. Por ejemplo, si sabemos con anterioridad que la recta de regresión de la población pasa por el origen, si utilizamos este conocimiento previo nos vemos obligados a efectuar la restricción de que el valor de la ordenada en el origen de la recta de regresión de la muestra sea cero. La incorporación de este conocimiento previo al procedimiento de estimación presenta dos aspectos distintos. En primer lugar, no estimamos la ordenada en el origen que conocemos su valor, y, en segundo lugar, estimamos la pendiente de la recta de regresión de forma que el estimador posea las propiedades deseables habituales *bajo la condición* de que la ordenada en el origen es cero. El segundo aspecto resulta muy importante, ya que un estimador que posea todas las propiedades deseables en condiciones generales no sigue manteniendo necesariamente dichas propiedades si las condiciones se hacen especiales en algún sentido. En nuestro ejemplo, las «condiciones generales» significan que la recta de regresión puede estar en cualquier lugar del plano, pero cuando especificamos que la recta de regresión *debe* pasar por el origen, imponemos una restricción que reduce el grado de generalidad.

En esta sección vamos a considerar distintos tipos de restricciones, algunas de ellas impuestas por la información previa acerca del *valor* de los coeficientes de regresión individuales, y otras impuestas por las *relaciones* entre dichos coeficientes de regresión.

Restricciones de valor fijo

Con el fin de ilustrar el caso en que poseemos información exacta acerca del valor de uno de los coeficientes de regresión, vamos a utilizar una ecuación de regresión con dos variables explicativas,

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

pero el análisis se aplica también, y se puede extender fácilmente, a las ecuaciones con mayor número de variables explicativas. Supongamos ahora que

conocemos a priori el valor de la ordenada en el origen; en particular, supongamos

$$\beta_1 = 0,$$

como es generalmente el caso. La ecuación de regresión se convierte pues en

$$(11.24) \quad Y_i = \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i.$$

Esta especificación de la ecuación de regresión incorpora ahora nuestro conocimiento previo acerca del valor de la ordenada en el origen, tal como pretendíamos. Podemos obtener los estimadores minimocuadráticos de β_2 y β_3 minimizando la suma de cuadrados S representada por

$$S = \sum_i (Y_i - \beta_2 X_{i2} - \beta_3 X_{i3})^2.$$

Derivando S respecto de β_2 y β_3 , obtenemos

$$\frac{\partial S}{\partial \beta_2} = -2 \sum_i X_{i2} (Y_i - \beta_2 X_{i2} - \beta_3 X_{i3}),$$

$$\frac{\partial S}{\partial \beta_3} = -2 \sum_i X_{i3} (Y_i - \beta_2 X_{i2} - \beta_3 X_{i3}).$$

Igualando a cero estas derivadas y reordenando los términos, obtenemos las ecuaciones normales minimocuadráticas siguientes:

$$\sum_i Y_i X_{i2} = \beta_2 \sum_i X_{i2}^2 + \beta_3 \sum_i X_{i2} X_{i3},$$

$$\sum_i Y_i X_{i3} = \beta_2 \sum_i X_{i2} X_{i3} + \beta_3 \sum_i X_{i3}^2.$$

(Comparando estas ecuaciones con las (10.12) y (10.13), vemos que son iguales a las que se obtendrían para los estimadores minimocuadráticos sin restricciones, en el caso de que $\bar{Y}$, $\bar{X}_2$ y $\bar{X}_3$ fuesen cero.) Los estimadores resultantes de β_2 y β_3 son

$$(11.25) \quad \begin{aligned} \hat{\beta}_2 &= \frac{(\sum Y_i X_{i2})(\sum X_{i3}^2) - (\sum Y_i X_{i3})(\sum X_{i2} X_{i3})}{(\sum X_{i2}^2)(\sum X_{i3}^2) - (\sum X_{i2} X_{i3})^2}, \\ \hat{\beta}_3 &= \frac{(\sum Y_i X_{i3})(\sum X_{i2}^2) - (\sum Y_i X_{i2})(\sum X_{i2} X_{i3})}{(\sum X_{i2}^2)(\sum X_{i3}^2) - (\sum X_{i2} X_{i3})^2}. \end{aligned}$$

Para obtener las fórmulas de las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$, señalemos que

$$\hat{\beta}_2 = \beta_2 + \frac{(\sum X_{i3}^2)(\sum X_{i2} \epsilon_i) - (\sum X_{i2} X_{i3})(\sum X_{i3} \epsilon_i)}{(\sum X_{i2}^2)(\sum X_{i3}^2) - (\sum X_{i2} X_{i3})^2},$$

$$\hat{\beta}_3 = \beta_3 + \frac{(\sum X_{i2}^2)(\sum X_{i3} \epsilon_i) - (\sum X_{i2} X_{i3})(\sum X_{i2} \epsilon_i)}{(\sum X_{i2}^2)(\sum X_{i3}^2) - (\sum X_{i2} X_{i3})^2}.$$

Por tanto,

$$(11.26) \quad \begin{aligned} \text{Var}(\hat{\beta}_2) &= E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2 \sum X_{13}^2}{(\sum X_{12}^2)(\sum X_{13}^2) - (\sum X_{12}X_{13})^2}, \\ \text{Var}(\hat{\beta}_3) &= E(\hat{\beta}_3 - \beta_3)^2 = \frac{\sigma^2 \sum X_{12}^2}{(\sum X_{12}^2)(\sum X_{13}^2) - (\sum X_{12}X_{13})^2}, \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) &= E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_3 - \beta_3) = \frac{-\sigma^2 \sum X_{12}X_{13}}{(\sum X_{12}^2)(\sum X_{13}^2) - (\sum X_{12}X_{13})^2}. \end{aligned}$$

Siguendo los pasos enumerados en la primera parte de la sección 10.1 puede mostrarse que $\hat{\beta}_2$ y $\hat{\beta}_3$ poseen todas las propiedades deseables. Las variancias de $\hat{\beta}_2$ y $\hat{\beta}_3$ se pueden estimar a partir de (11.26), y sustituyendo σ^2 por su estimador insesgado s^2 que se define como

$$(11.27) \quad s^2 = \frac{1}{n-2} \sum_i (Y_i - \hat{\beta}_2 X_{i2} - \hat{\beta}_3 X_{i3})^2.$$

Obsérvese que en la ecuación (11.27), la suma de cuadrados de los residuos se divide por $(n-2)$ —y no por $(n-3)$ como en el caso de (10.34)—, ya que sólo se ignoran dos coeficientes de regresión. Todos estos resultados se pueden generalizar fácilmente con el fin de aplicarlos a los modelos de regresión con más de dos variables explicativas.

El caso anterior demuestra la mejora conseguida al incorporar el conocimiento previo acerca de un coeficiente de regresión al procedimiento de estimación. Esta mejoría tiene básicamente dos aspectos. En primer lugar, no tenemos necesidad de estimar β_1 , ya que su valor es conocido, y, en segundo lugar, los estimadores minimocuadráticos de β_2 y β_3 con esta restricción, es decir, los estimadores que incorporan la información de que $\beta_1 = 0$, poseen una variancia menor que los estimadores minimocuadráticos ordinarios sin restricciones. Esta segunda mejoría representa una característica menos obvia, pero muy importante, de la estimación con restricciones. Por ejemplo, sea $\hat{\beta}_2$ el estimador con restricciones de β_2 definido en la ecuación (11.25), y $\hat{\beta}_2^{ORD}$ el estimador minimocuadrático ordinario de β_2 sin restricciones, tal como se ha definido en la ecuación (10.12). El cociente entre las variancias de estos dos estimadores es

$$(11.28) \quad \begin{aligned} \frac{\text{Var}(\hat{\beta}_2^{ORD})}{\text{Var}(\hat{\beta}_2)} &= \frac{\sigma^2 m_{33}/(m_{22}m_{33} - m_{23}^2)}{\sigma^2 \sum X_{13}^2 / [(\sum X_{12}^2)(\sum X_{13}^2) - (\sum X_{12}X_{13})^2]} \\ &= \frac{m_{33}[(m_{22} + n\bar{X}_2^2)(m_{33} + n\bar{X}_3^2) - (m_{23} + n\bar{X}_2\bar{X}_3)^2]}{(m_{33} + n\bar{X}_3^2)(m_{22}m_{33} - m_{23}^2)} \\ &= 1 + \frac{n(\bar{X}_2 m_{33} - \bar{X}_3 m_{23})^2}{(m_{33} + n\bar{X}_3^2)(m_{22}m_{33} - m_{23}^2)}, \end{aligned}$$

lógicamente, es mayor o igual que la unidad. Por tanto,

$$\text{Var}(\hat{\beta}_2^{ORD}) \geq \text{Var}(\hat{\beta}_2).$$

Se puede presentar una demostración similar para β_3 . Sin embargo, obsérvese que, aunque los estimadores con restricciones de β_2 y β_3 poseen menor variancia que sus contrapartidas sin restricciones, el valor de R^2 correspondiente a la recta de regresión de la muestra sin restricciones es mayor que el de la regresión con la restricción. Esto se debe a que la estimación minimocuadrática sin restricciones produce el valor máximo de R^2 y, por tanto, *cualquier* otro procedimiento originará una disminución.

Supongamos ahora que la información previa acerca del valor de un coeficiente de regresión no se refiere a la ordenada en el origen, sino, por ejemplo, a β_3 . En este caso, la ecuación de regresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + e_i$$

se puede escribir en la forma

$$(11.29) \quad Y_i^* = \beta_1 + \beta_2 X_{i2} + e_i,$$

en la que Y_i^* corresponde a $(Y_i - \beta_3 X_{i3})$. Los estimadores minimocuadráticos de β_1 y β_2 en la ecuación (11.29), a los que llamaremos $\hat{\beta}_1$ y $\hat{\beta}_2$, poseen todas las propiedades deseables. Obsérvese que, en este caso, la variancia de $\hat{\beta}_2$ es

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{m_{22}},$$

mientras que la variancia del estimador minimocuadrático sin restricciones de β_2 es

$$\text{Var}(\hat{\beta}_2^{ORD}) = \frac{\sigma^2 m_{33}}{m_{22}m_{33} - m_{23}^2}.$$

El cociente entre estas variancias es, pues,

$$(11.30) \quad \begin{aligned} \frac{\text{Var}(\hat{\beta}_2^{ORD})}{\text{Var}(\hat{\beta}_2)} &= \frac{\sigma^2 m_{33}/(m_{22}m_{33} - m_{23}^2)}{\sigma^2/m_{22}} = \frac{m_{22}m_{33}}{m_{22}m_{33} - m_{23}^2} \\ &= \frac{1}{1 - r_{23}^2}, \end{aligned}$$

en el que r_{23}^2 es el coeficiente de correlación entre X_{12} y X_{13} en la muestra. Por tanto, si al estimar β_2 utilizamos la información acerca de β_3 , obtendremos siempre un aumento de eficiencia a menos que r_{23}^2 sea cero. Este aumento de eficiencia es especialmente importante si r_{23}^2 está próximo a la unidad, es decir, si las dos variables explicativas están muy correlacionadas en la muestra.

Restricciones en forma de desigualdad

En muchos casos, la información previa acerca de un coeficiente de regresión no viene dada en forma de una igualdad definida, sino que toma la forma de una desigualdad. Por ejemplo, en el caso de la función lineal de consumo, sabemos a priori que el coeficiente correspondiente a la renta (es decir, la propensión marginal al consumo) se halla entre 0 y 1. En general, supongamos que la información previa acerca del coeficiente de regresión β_k es

$$a \leq \beta_k \leq b,$$

en la que a y b son ciertos valores conocidos, y $b > a$. En este caso, nos encontramos con el problema de estimar los coeficientes de una ecuación de regresión sujeta a una restricción en forma de desigualdad sobre uno o más de sus coeficientes. Este tipo de restricción resulta bastante más difícil de incorporar al procedimiento de estimación que cuando se trataba de una igualdad definida. El procedimiento más sencillo consiste en obtener el estimador minimocuadrático de β_k ignorando por completo la restricción, y definir el estimador de β_k con la restricción, al que llamaremos $\hat{\beta}_k$, de la forma siguiente:

(11.31)

$$\begin{aligned} \hat{\beta}_k &= \hat{\beta}_k^{ORD} & \text{si} & \quad a \leq \hat{\beta}_k^{ORD} \leq b, \\ &= a & \text{si} & \quad \hat{\beta}_k^{ORD} < a, \\ &= b & \text{si} & \quad \hat{\beta}_k^{ORD} > b, \end{aligned}$$

siendo $\hat{\beta}_k^{ORD}$ el estimador ordinario, sin restricciones, de β_k . Naturalmente, como β_k está limitada a un intervalo concreto, su distribución deja de ser normal. Sin embargo, a medida que aumenta el tamaño de la muestra, la probabilidad de que las estimaciones ordinarias, sin restricciones, se hallen fuera del intervalo especificado por la desigualdad, se hace más pequeña, a menos que β_k sea igual a a o a b . Por tanto, en las muestras grandes podemos considerar que la distribución normal es una aproximación razonable de la distribución de $\hat{\beta}_k$ correspondiente a (11.31). Naturalmente, cuanto más cerca esté el verdadero valor de β_k del centro del intervalo, mejor será la aproximación. Este procedimiento de estimación de los coeficientes de regresión sujetos a una restricción en forma de desigualdad presenta el inconveniente de que, al estimar los coeficientes de regresión restantes, prescindimos de la información acerca de β_k . Un enfoque alternativo, que no posee dicho inconveniente, consiste en obtener los estimadores del coeficiente de regresión maximizando la correspondiente función de verosimilitud sujeta a la restricción (o restricciones) en forma

de desigualdad. Esto representa un problema de programación cuadrática y su solución es muy complicada.⁴

Otro procedimiento distinto para la estimación de los coeficientes de regresión sujetos a restricciones en forma de desigualdad es el propuesto por Theil y Goldberger, conocido con el nombre de «estimación mixta».⁵ Los autores señalan que la desigualdad

$$a \leq \beta_k \leq b$$

es en realidad una afirmación en el sentido de que consideramos que β_k se halla con certeza entre a y b . Esto se puede considerar como si efectuásemos una afirmación de tipo probabilístico acerca de β_k , es decir, como si estuviésemos diciendo que la probabilidad de que β_k se halle fuera del intervalo $a \leq \beta_k \leq b$ es cero, y que la probabilidad de que β_k esté en el interior de dicho intervalo es la unidad. Sin embargo, si efectuamos estas afirmaciones al distinguir entre los valores de β_k dentro y fuera del intervalo $a \leq \beta_k \leq b$, podemos también estar dispuestos a efectuar otra afirmación de tipo probabilístico acerca de β_k en el interior del citado intervalo. Una afirmación de este tipo (probablemente muy común) es la de que no tenemos ninguna idea acerca de la situación de β_k en el interior del intervalo. Esto se puede interpretar en el sentido de que, por lo que sabemos, todos los valores de β_k en el interior del intervalo son igualmente probables, de modo que la distribución de β_k dentro del intervalo es uniforme. En este caso, podemos escribir

(11.32)

$$\beta_k = \frac{a+b}{2} + u,$$

en la que u tiene una distribución uniforme continua con media

$$E(u) = 0$$

y variancia⁶

$$\text{Var}(u) = \frac{(b-a)^2}{12}.$$

4. Este enfoque ha sido desarrollado por G. C. Judge y T. Takayama, en «Inequality Restrictions in Regression Analysis», *Journal of the American Statistical Association*, vol. 61, marzo 1966, pp. 166-81.

5. H. Theil y A. S. Goldberger, «On Pure and Mixed Statistical Estimation in Econometrics», *International Economic Review*, vol. 2, enero 1961, pp. 65-78. Este método se denomina «mixto» porque incluye elementos del enfoque clásico y del enfoque bayesiano de estimación.

6. Ver, por ejemplo, Freund, *op. cit.*, p. 146.

Esta especificación se puede incorporar a nuestro procedimiento de estimación utilizando el método «mixto» de Theil y Goldberger.⁷

La idea fundamental del método de estimación «mixta» consiste en combinar la información previa acerca de uno o más de los coeficientes de regresión con la información proporcionada por la muestra. Esto se puede conseguir utilizando la restricción en forma de desigualdad como si fuese una observación adicional, aplicando entonces el método de los mínimos cuadrados a la muestra «ampliada». En primer lugar, podemos escribir nuevamente la fórmula (11.32) en la forma

$$(11.32a) \quad \frac{a+b}{2} = \beta_1 \times 0 + \beta_2 \times 0 + \dots + \beta_k \times 1 + \dots + \beta_K \times 0 + (-u),$$

que puede considerarse como la observación $(n+1)$ correspondiente a

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \dots + \beta_K X_{iK} + \epsilon_i,$$

utilizando $(-u)$ como término de perturbación. Sin embargo, como todas las ϵ_i tienen una variancia igual a σ^2 mientras que la variancia de $(-u)$ es $(b-a)^2/12$, esta última observación presentada en la forma de la ecuación (11.32a) introduciría heteroscedasticidad. Esto se puede remediar fácilmente multiplicando ambos lados de la ecuación (11.32a) por $\sigma\sqrt{12}/(b-a)$ con el fin de obtener

(11.32b)

$$\frac{(b+a)\sigma\sqrt{3}}{b-a} = \beta_1 \times 0 + \beta_2 \times 0 + \dots + \beta_k \times \left[\frac{\sigma\sqrt{12}}{b-a} \right] + \dots + \beta_K \times 0 + u^*$$

en la que

$$u^* = \frac{-u\sigma\sqrt{12}}{b-a},$$

o sea, que

$$\text{Var}(u^*) = \left[\frac{12\sigma^2}{(b-a)^2} \right] \text{Var}(u) = \sigma^2,$$

7. En caso de considerar que los valores de β_k en el centro del intervalo $a \leq \beta_k \leq b$ son más probables que los valores cercanos a los límites del intervalo, obtendríamos una especificación distinta. En este caso, podríamos especificar la variancia de u de forma que

$$\frac{a+b}{2} + 2\sqrt{\text{Var}(u)} = b \quad \text{y} \quad \frac{a+b}{2} - 2\sqrt{\text{Var}(u)} = a,$$

basándonos en el hecho de que, en muchas distribuciones, los valores de la variable no comprendidos en el intervalo delimitado por la media más dos veces la desviación standard y la media menos dos veces la desviación standard, se presentan muy raramente. De esta forma, obtendríamos

$$\text{Var}(u) = \frac{(b-a)^2}{16}.$$

Esta es la especificación utilizada por Theil y Goldberger, *op. cit.*

como queríamos obtener. Así pues, las observaciones de la muestra «ampliada» son

i	Y_i	X_{i1}	X_{i2}	...	X_{ik}	...	X_{iK}
1	Y_1	1	X_{12}	...	X_{1k}	...	X_{1K}
2	Y_2	1	X_{22}	...	X_{2k}	...	X_{2K}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
n	Y_n	1	X_{n2}	...	X_{nk}	...	X_{nK}
$n+1$	$\frac{(b+a)\sigma\sqrt{3}}{b-a}$	0	0	...	$\frac{\sigma\sqrt{12}}{b-a}$	...	0

El valor de la observación $(n+1)$ de Y no se conoce porque no sabemos cuál es el valor de σ . Sin embargo, esta última puede estimarse a través de s [definida en (10.34)] en las primeras n observaciones. Como se ha mostrado, s^2 es un estimador insesgado, consistente y asintóticamente eficiente de σ^2 . Las ecuaciones normales minimocuadráticas de los estimadores «mixtos» de los coeficientes de regresión son

(11.33)

$$\sum_i Y_i = \beta_1 n + \beta_2 \sum_i X_{i2} + \dots + \beta_k \sum_i X_{ik}$$

$$+ \dots + \beta_K \sum_i X_{iK},$$

$$\sum_i Y_i X_{i2} = \beta_1 \sum_i X_{i2} + \beta_2 \sum_i X_{i2}^2 + \dots + \beta_k \sum_i X_{i2} X_{ik}$$

$$+ \dots + \beta_K \sum_i X_{i2} X_{iK},$$

$\vdots$

$$\sum_i Y_i X_{ik} + \frac{6s^2(b+a)}{(b-a)^2} = \beta_1 \sum_i X_{ik} + \beta_2 \sum_i X_{i2} X_{ik}$$

$$+ \dots + \beta_k \left[\sum_i X_{ik}^2 + \frac{12s^2}{(b-a)^2} \right]$$

$$+ \dots + \beta_K \sum_i X_{ik} X_{iK},$$

$\vdots$

$$\sum_i Y_i X_{iK} = \beta_1 \sum_i X_{iK} + \beta_2 \sum_i X_{i2} X_{iK} + \dots + \beta_k \sum_i X_{ik} X_{iK}$$

$$+ \dots + \beta_K \sum_i X_{iK}^2,$$

en las que todos los sumatorios se extienden entre $i=1$ e $i=n$. En el caso de dos variables explicativas con la restricción

$$a \leq \beta_2 \leq b,$$

los estimadores «mixtos» son

$$(11.34) \quad \begin{aligned} \hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3, \\ \hat{\beta}_2 &= \frac{m_{Y2}^* m_{33} - m_{Y3} m_{23}}{m_{22}^* m_{33} - m_{23}^2}, \\ \hat{\beta}_3 &= \frac{m_{Y3} m_{22}^* - m_{Y2} m_{23}}{m_{22}^* m_{33} - m_{23}^2}, \end{aligned}$$

en las que

$$\begin{aligned} m_{Y2}^* &= m_{Y2} + \frac{6s^2(b+a)}{(b-a)^2}, \\ m_{22}^* &= m_{22} + \frac{12s^2}{(b-a)^2}. \end{aligned}$$

Aparte de las modificaciones obvias, estas fórmulas son las mismas que las de los estimadores minimocuadráticos ordinarios, sin restricciones, presentados en (10.11) a (10.13). Las variancias de los estimadores «mixtos» se pueden estimar de la forma siguiente

$$(11.35) \quad \begin{aligned} \text{Est. Var}(\hat{\beta}_2) &= \frac{s^2 m_{33}}{m_{22}^* m_{33} - m_{23}^2}, \\ \text{Est. Var}(\hat{\beta}_3) &= \frac{s^2 m_{22}^*}{m_{22}^* m_{33} - m_{23}^2}. \end{aligned}$$

Estos resultados se han obtenido por analogía con los de (10.24) y (10.25). Con esta especificación, los estimadores «mixtos» definidos en (11.34) son consistentes y asintóticamente eficientes, pero la insesgadez resulta difícil de demostrar ya que las fórmulas incluyen s^2 en forma no lineal. Nótese, sin embargo, que no podemos eliminar la posibilidad de que el estimador «mixto» de β_2 se halle fuera del intervalo especificado. Si sucede así, esto reflejará el hecho de que, dada la hipótesis sostenida del modelo, la muestra está en desacuerdo con la información previa acerca de β_2 .

EJEMPLO. Podemos ilustrar el método de la estimación «mixta» estimando los coeficientes de la función de consumo trimestral en los Estados Unidos,

$$C_t = \alpha + \beta Y_t + \varepsilon_t,$$

en la que C = consumo e Y = renta. La información previa acerca de β es

$$0 \leq \beta \leq 1.$$

Zellner ha estimado esta función⁸ sin la restricción, obteniendo los resultados siguientes:

$$C_t = 38.09 + 0.747 Y_t + \varepsilon_t, \quad R^2 = 0.944. \\ (0.033)$$

Si queremos imponer la restricción en forma de desigualdad sobre β y utilizar el método de estimación «mixta», escribiremos

$$\beta = \frac{1}{2} + u,$$

en la que $E(u) = 0$ y $\text{Var}(u) = 1/12$.

Las ecuaciones normales minimocuadráticas se convierten en

$$\sum_i C_i = \hat{\alpha} n + \hat{\beta} \sum_i Y_i,$$

$$\sum_i C_i Y_i + 6s^2 = \hat{\alpha} \sum_i Y_i + \hat{\beta} (\sum_i Y_i^2 + 12s^2).$$

Eliminando $\hat{\alpha}$, obtenemos

$$\sum (C_i - \bar{C})(Y_i - \bar{Y}) + 6s^2 = \hat{\beta} [\sum (Y_i - \bar{Y})^2 + 12s^2].$$

Ahora bien, a partir de los resultados de Zellner, vemos que

$$\sum (C_i - \bar{C})(Y_i - \bar{Y}) = 0.747 \times \sum (Y_i - \bar{Y})^2$$

$$\frac{s^2}{\sum (Y_i - \bar{Y})^2} = 0.033^2,$$

$$s^2 = 0.033^2 \times \sum (Y_i - \bar{Y})^2.$$

Por tanto,

$$\begin{aligned} 0.747 \times \sum (Y_i - \bar{Y})^2 + 6 \times 0.033^2 \times \sum (Y_i - \bar{Y})^2 \\ = \hat{\beta} [\sum (Y_i - \bar{Y})^2 + 12 \times 0.033^2 \times \sum (Y_i - \bar{Y})^2]. \end{aligned}$$

En esta expresión podemos despejar $\hat{\beta}$ porque los términos $\sum (Y_i - \bar{Y})^2$ se cancelan. El resultado es

$$\hat{\beta} = 0.744$$

$$\begin{aligned} \hat{\alpha} &= \bar{C} - 0.744 \bar{Y} = (38.09 + 0.747 \bar{Y}) - 0.744 \bar{Y} \\ &= 38.09 + 0.003 \times 200.55 = 38.69. \end{aligned}$$

⁸ A. Zellner, «The Short Run Consumption Function», *Econometrica*, vol. 25, octu-
pp. 552-67.

Por último,

$$\text{Est. Var}(\hat{\beta}) = \frac{s^2}{\sum (Y_i - \bar{Y})^2 + 12s^2} = \frac{0.033^2 \times \sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2 + 12 \times 0.033^2 \times \sum (Y_i - \bar{Y})^2} = 0.001075.$$

El resultado de la estimación «mixta» de la función de consumo es, pues,

$$C_t = 38.69 + 0.744 Y_t + e_t. \\ (0.033)$$

Como puede verse, la incorporación de la restricción en forma de desigualdad al procedimiento de estimación ha tenido como resultado un cambio muy pequeño en la ecuación de regresión estimada en este caso.

Restricciones lineales

Frecuentemente, la información previa de que se dispone no se refiere a los valores de los coeficientes de regresión individuales, sino a las relaciones entre ellos. Por ejemplo, podemos saber a priori que un coeficiente es igual a la suma de otros dos, o que determinados coeficientes forman una progresión geométrica. Estas restricciones pueden dividirse en dos grupos, según que la relación entre los coeficientes sea lineal o no. Vamos a estudiar en primer lugar las *restricciones lineales*, ya que son las más sencillas. En ambos casos, resulta conveniente comparar la ecuación de regresión en la que se tienen en cuenta explícitamente las restricciones con la ecuación de regresión en que se ignoran dichas restricciones. Los parámetros de la primera ecuación se llamarán «con restricciones» y los de la segunda, «sin restricciones». Esta yuxtaposición nos permitirá introducir el concepto de «identificación» que juega un importante papel en muchos problemas econométricos. Efectuaremos el análisis mediante ejemplos que ilustrarán el tipo de problema y su solución, y que pueden modificarse fácilmente para su aplicación a otros casos.

Consideremos el problema de estimar un modelo de regresión con una variable explicativa y con efectos estacionales aditivos. Es decir,

$$\begin{aligned} Y_t &= \alpha_1 + \beta X_t + \varepsilon_t && \text{(primer trimestre),} \\ Y_t &= \alpha_2 + \beta X_t + \varepsilon_t && \text{(segundo trimestre),} \\ Y_t &= \alpha_3 + \beta X_t + \varepsilon_t && \text{(tercer trimestre),} \\ Y_t &= \alpha_4 + \beta X_t + \varepsilon_t && \text{(cuarto trimestre).} \end{aligned}$$

Si introducimos variables binarias para los tres últimos trimestres, podremos escribir de nuevo las ecuaciones anteriores en forma compacta:

$$(11.36) \quad Y_t = \alpha_1 + \beta X_t + (\alpha_2 - \alpha_1)Q_{t2} + (\alpha_3 - \alpha_1)Q_{t3} + (\alpha_4 - \alpha_1)Q_{t4} + \varepsilon_t,$$

en la que

$Q_{t2} = 1$ si t es un segundo trimestre,

$= 0$ si no lo es;

$Q_{t3} = 1$ si t es un tercer trimestre,

$= 0$ si no lo es;

$Q_{t4} = 1$ si t es un cuarto trimestre,

$= 0$ si no lo es;

En este caso tenemos cinco parámetros con restricciones: α_1 , α_2 , α_3 , α_4 y β . Las restricciones consisten en que α_2 tiene que ser igual al coeficiente de Q_{t2} más la ordenada en el origen, α_3 igual al de Q_{t3} más la ordenada en el origen, y α_4 al coeficiente de Q_{t4} más la ordenada en el origen. La forma de (11.36) sin restricciones es

$$(11.36a) \quad Y_t = \beta_1 + \beta_2 Q_{t2} + \beta_3 Q_{t3} + \beta_4 Q_{t4} + \beta_5 X_t + \varepsilon_t.$$

Comparando los coeficientes sin restricciones de (11.36a) con los parámetros de (11.36) con restricciones, podemos ver que

$$\begin{aligned} \beta_1 &= \alpha_1, \\ \beta_2 &= \alpha_2 - \alpha_1, \\ \beta_3 &= \alpha_3 - \alpha_1, \\ \beta_4 &= \alpha_4 - \alpha_1, \\ \beta_5 &= \beta. \end{aligned}$$

En este caso, existe una correspondencia de uno a uno entre los parámetros, en el sentido de que existe una única solución para los parámetros con restricciones en términos de los parámetros sin restricciones. Concretamente,

$$\begin{aligned} \alpha_1 &= \beta_1, \\ \alpha_2 &= \beta_2 + \beta_1, \\ \alpha_3 &= \beta_3 + \beta_1, \\ \alpha_4 &= \beta_4 + \beta_1, \\ \beta &= \beta_5. \end{aligned}$$

Este caso recibe el nombre de *identificación exacta* para indicar el hecho de que los parámetros con restricciones se pueden «identificar» unívocamente en términos de los coeficientes sin restricciones.

La importancia práctica de la identificación exacta radica en el hecho de que podemos obtener estimaciones minimocuadráticas de los coeficientes sin restricciones utilizándolas después para obtener las estimaciones de los parámetros con restricción. Como estas últimas son funciones lineales de las primeras, todas las propiedades deseables de éstas se transmitirán a las de los parámetros con restricción. Asimismo, las variancias de las estimaciones de los parámetros con restricciones pueden obtenerse a partir de las variancias y covariancias de los coeficientes sin restricciones estimados. Concretamente,

$$\begin{aligned}\text{Var}(\hat{\alpha}_1) &= \text{Var}(\hat{\beta}_1), \\ \text{Var}(\hat{\alpha}_2) &= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2 \text{Cov}(\hat{\beta}_1, \hat{\beta}_2), \\ \text{Var}(\hat{\alpha}_3) &= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_3) + 2 \text{Cov}(\hat{\beta}_1, \hat{\beta}_3), \\ \text{Var}(\hat{\alpha}_4) &= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_4) + 2 \text{Cov}(\hat{\beta}_1, \hat{\beta}_4), \\ \text{Var}(\hat{\beta}) &= \text{Var}(\hat{\beta}_5).\end{aligned}$$

Y entre las respectivas estimaciones de las variancias se cumplen las mismas relaciones. Obsérvese que en las estimaciones de los parámetros con restricciones obtendríamos los mismos resultados si escribiésemos la fórmula (11.36) en la forma

$$(11.36b) \quad Y_t = \alpha_1 Q_{t1} + \alpha_2 Q_{t2} + \alpha_3 Q_{t3} + \alpha_4 Q_{t4} + \beta X_t + \epsilon_t.$$

en la que $Q_{ti} = 1$ si t es un primer trimestre,
 $= 0$ si no lo es.

La ecuación (11.36b) tiene la restricción de pasar por el origen, y puede estimarse por el método de los mínimos cuadrados en la forma explicada al principio de esta sección. Las estimaciones resultantes y sus variancias son exactamente iguales a las obtenidas a partir de las estimaciones sin restricciones de (11.36a)

Consideremos ahora un caso distinto de restricciones lineales, en las que la suma de dos o más coeficientes de regresión tiene que ser igual a un número determinado. Un típico ejemplo de esta restricción es la función de producción de Cobb-Douglas con rendimientos constantes de escala. En este caso, queremos que la suma de las pendientes de regresión sea igual a la unidad. De modo específico,

$$(11.37) \quad Y_t = \alpha_1 + \alpha_2 X_{t2} + (1 - \alpha_2) X_{t3} + \epsilon_t,$$

en la que $Y_t = \log$ de la cantidad producida, $X_{t2} = \log$ de la cantidad de trabajo, y $X_{t3} = \log$ de la cantidad de capital. Tenemos dos parámetros con restricciones: α_1 y α_2 . La forma sin restricciones de (11.37) es

$$(11.37a) \quad Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \epsilon_t.$$

La relación entre los parámetros sin restricciones y con restricciones es

$$\begin{aligned}\beta_1 &= \alpha_1, \\ \beta_2 &= \alpha_2, \\ \beta_3 &= 1 - \alpha_2.\end{aligned}$$

En este caso, el número de coeficientes sin restricciones es mayor que el número de parámetros con restricciones, y no existe una solución única para α_2 . En realidad,

$$\begin{aligned}\alpha_1 &= \beta_1, \\ \alpha_2 &= \beta_2, \\ \alpha_2 &= 1 - \beta_3.\end{aligned}$$

y

Este caso se denomina *superidentificación*, aludiendo al hecho de que existe más de una solución para «identificar» el parámetro con restricciones α_2 .

En condiciones de superidentificación no podemos resolver el problema como hicimos en el caso de la identificación exacta, estimando la ecuación sin restricciones y utilizando los resultados para obtener estimaciones de los parámetros con restricciones. Por el contrario, debemos utilizar directamente la ecuación con restricciones (11.37). Los estimadores minimocuadráticos de los parámetros con restricciones pueden obtenerse minimizando

$$\sum_t [Y_t - \alpha_1 - \alpha_2 X_{t2} - (1 - \alpha_2) X_{t3}]^2$$

respecto de α_1 y α_2 . Puede mostrarse fácilmente que las estimaciones resultantes son iguales a las obtenidas aplicando el método de los mínimos cuadrados a

$$(11.37b) \quad Y_t^* = \alpha_1 + \alpha_2 X_{t2}^* + \epsilon_t,$$

en la que Y_t^* equivale a $(Y_t - X_{t3})$, y X_{t2}^* a $(X_{t2} - X_{t3})$. La ecuación (11.37b) representa una forma alternativa de escribir (11.37) reordenando sus términos. Si hay superidentificación y las restricciones son lineales, siempre existe una solución.

Como tercero y último caso de restricciones lineales, consideremos aquel en el que el número de parámetros con restricciones es mayor que el número de coeficientes sin restricciones. Por ejemplo, supongamos que el gasto de una familia en fruta puede expresarse a través de la siguiente ecuación de regresión:

$$F_i = \alpha_F + \beta_F Y_i + \varepsilon_{iF},$$

en la que F_i = gasto familiar en fruta, e Y_i = renta familiar. Supongamos además que el gasto familiar en vegetales puede expresarse a través de

$$V_i = \alpha_V + \beta_V Y_i + \varepsilon_{iV},$$

en la que V_i = gasto familiar en vegetales. Ahora bien si, como ocurre normalmente, la muestra no proporciona información separada del gasto en fruta y en vegetales, sino solamente el total, tenemos que combinar las regresiones anteriores para obtener

$$(11.38) \quad G_i = (\alpha_F + \alpha_V) + (\beta_F + \beta_V)Y_i + \varepsilon_i,$$

en la que $G_i = F_i + V_i$, y $\varepsilon_i = \varepsilon_{iF} + \varepsilon_{iV}$. La versión sin restricciones de (11.38) es

$$(11.38a) \quad G_i = \alpha + \beta Y_i + \varepsilon_i.$$

Si comparamos los coeficientes de las dos ecuaciones, vemos que

$$\alpha = \alpha_F + \alpha_V \quad \beta = \beta_F + \beta_V.$$

En este modelo tenemos cuatro parámetros con restricciones, pero solamente dos coeficientes sin restricciones y, obviamente, no podemos expresar los primeros en términos de los segundos. Este caso se conoce con el nombre de *subidentificación* y en él los parámetros con restricciones no se pueden estimar a partir de la información muestral disponible.

Restricciones no lineales

Como primer caso de las restricciones no lineales, vamos a considerar la *identificación exacta*. Esta situación puede ilustrarse a través de una versión simple del llamado «modelo de ajuste de inventarios». Supongamos que el volumen de inventarios de una mercancía que una empresa «desea» mantener es una función lineal de las ventas, es decir

$$Y_t^* = \alpha + \beta X_t,$$

en la que Y_t^* = nivel deseado de inventarios al finalizar el período t , y X_t = ventas durante el período t . En general, Y_t^* no es observable. Supongamos además que los ajustes realizados por cada empresa al nivel deseado en cada período no son completos, de modo que

$$Y_t - Y_{t-1} = \gamma(Y_t^* - Y_{t-1}) + \varepsilon_t,$$

en la que Y_t = nivel real de inventarios al finalizar el período t . El parámetro γ se denomina «coeficiente de ajuste» y su valor se halla entre 0 y 1. Un valor de γ próximo a cero indica que durante cada período sólo se elimina una parte muy pequeña de la diferencia entre el nivel real y el nivel deseado de inventarios, mientras que un valor de γ próximo a la unidad indica que se elimina gran parte de esta diferencia. El término de perturbación ε_t se introduce para tener en cuenta las influencias aleatorias en el proceso de ajuste. Sustituyendo Y_t^* por su expresión, y reordenando los términos, obtenemos

$$(11.39) \quad Y_t = \alpha\gamma + \beta\gamma X_t + (1 - \gamma)Y_{t-1} + \varepsilon_t.$$

Esta ecuación explica la inversión en inventarios y, todas sus variables excepto ε_t son observables. En (11.39) tenemos tres parámetros «con restricciones»: α , β y γ . La contrapartida sin restricciones de (11.39) es

$$(11.39a) \quad Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + \varepsilon_t.$$

Los coeficientes de (11.39a) se relacionan con los parámetros de (11.39) en la forma siguiente:

$$\beta_1 = \alpha\gamma,$$

$$\beta_2 = \beta\gamma,$$

$$\beta_3 = 1 - \gamma.$$

En este caso, podemos obtener una solución única para los parámetros con restricciones en términos de los coeficientes sin restricción, o sea que tenemos *identificación exacta*. La solución es

$$\alpha = \frac{\beta_1}{1 - \beta_3},$$

$$\beta = \frac{\beta_2}{1 - \beta_3},$$

$$\gamma = 1 - \beta_3.$$

Obsérvese que α y β son funciones *no lineales* de las β sin restricciones.

Para estimar los parámetros de (11.39), obtendremos las estimaciones minimocuadráticas de los coeficientes sin restricciones de (11.39a) y utilizaremos las ecuaciones de α , β y γ para obtener las correspondientes estimaciones de estos parámetros. [Las restricciones en forma de desigualdad con respecto a γ , y en consecuencia a β_3 , pueden manejarse siguiendo la regla especificada en (11.51).] Por lo que respecta a las propiedades de los estimadores resultantes, hay que señalar que estos estimadores, que son funciones no lineales de los coeficientes sin restricciones, reciben las propiedades asintóticas de estos últimos, pero *no* las propiedades de las muestras pequeñas. La razón de ello es que la insesgadez no «persiste» a través de las funciones no lineales (véase el teorema 17 y los comentarios que le siguen en la sección 6.1). En el caso que nos ocupa, los propios estimadores sin restricciones no son insesgados debido a la presencia de Y_{t-1} entre las variables explicativas, de modo que *ninguno* de los estimadores con restricciones puede pretender ser insesgado. La variancia del estimador con restricciones de γ se puede obtener a partir de la variancia de β_3 sin restricciones, observando que

$$\text{Var}(\hat{\gamma}) = \text{Var}(\hat{\beta}_3).$$

La obtención de las variancias de $\hat{\alpha}$ y $\hat{\beta}$ resulta algo más compleja, ya que $\hat{\alpha}$ y $\hat{\beta}$ son funciones no lineales de los estimadores sin restricciones. Sin embargo, existe una fórmula aproximada que se puede utilizar en este caso. La fórmula se refiere al caso general en el que un estimador, por ejemplo $\hat{\alpha}$, es función de otros k estimadores tales como $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, es decir,

$$\hat{\alpha} = f(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k).$$

Entonces, la variancia de $\hat{\alpha}$ en las muestras grandes se puede aproximar⁹ a través de

$$(11.40) \quad \text{Var}(\hat{\alpha}) \approx \sum_k \left[\frac{\partial f}{\partial \hat{\beta}_k} \right]^2 \text{Var}(\hat{\beta}_k) + 2 \sum_{j < k} \left[\frac{\partial f}{\partial \hat{\beta}_j} \right] \left[\frac{\partial f}{\partial \hat{\beta}_k} \right] \text{Cov}(\hat{\beta}_j, \hat{\beta}_k)$$

($j, k = 1, 2, \dots, K; j < k$).

(Esta aproximación se obtiene utilizando la expansión en serie de Taylor de $f(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$ en torno a $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, eliminando los términos de segundo orden o de orden superior, y obteniendo después la variancia por medio de la fórmula habitual.) Por ejemplo, para

$$\hat{\alpha} = \frac{\hat{\beta}_1}{1 - \hat{\beta}_3},$$

tenemos

$$\begin{aligned} \text{Var}(\hat{\alpha}) \approx & \left[\frac{1}{1 - \hat{\beta}_3} \right]^2 \text{Var}(\hat{\beta}_1) + \left[\frac{\hat{\beta}_1}{(1 - \hat{\beta}_3)^2} \right]^2 \text{Var}(\hat{\beta}_3) \\ & + 2 \left[\frac{1}{1 - \hat{\beta}_3} \right] \left[\frac{\hat{\beta}_1}{(1 - \hat{\beta}_3)^2} \right] \text{Cov}(\hat{\beta}_1, \hat{\beta}_3). \end{aligned}$$

Esta fórmula puede utilizarse para obtener una aproximación de la variancia de $\hat{\alpha}$ en las muestras grandes. Como las variancias y covariancias de los estimadores sin restricciones en las muestras grandes pueden obtenerse fácilmente a través de las fórmulas corrientes, no hay ningún problema para estimar las variancias de los estimadores con restricciones en las muestras grandes.

Un enfoque alternativo para estimar los parámetros de (11.39) consiste en minimizar la suma de cuadrados S que viene dada por

$$S = \sum_{t=2}^n [Y_t - \alpha\gamma - \beta\gamma X_t - (1 - \gamma)Y_{t-1}]^2$$

respecto de α , β y γ . Los estimadores resultantes de estos parámetros reciben el nombre de *estimadores minimocuadráticos no lineales*. Se puede mostrar fácilmente que estos estimadores son iguales a los obtenidos a partir de los coeficientes sin restricciones estimados en base a (11.39a). Además, como la función logarística de verosimilitud de $Y_2, Y_3, \dots, Y_n$ (relacionada a Y_1) es

$$L = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^n [Y_t - \alpha\gamma - \beta\gamma X_t - (1 - \gamma)Y_{t-1}]^2,$$

minimizar S respecto de α , β y γ equivale a maximizar L respecto a los mismos parámetros. Por tanto, si el término de perturbación de la regresión se distribuye normalmente, los estimadores minimocuadráticos no lineales son iguales a los estimadores maximoverosímiles y, en consecuencia, sus variancias asintóticas se pueden estimar a través de la matriz de información apropiada.

EJEMPLO. El caso de restricciones no lineales en condiciones de identificación exacta se presenta en relación con uno de los modelos de la función de consumo analizados por Zellner¹⁰. No se trata de un «modelo de ajuste de inventarios» en sentido estricto, pero posee las mismas características básicas. Utilizando una versión sencilla de la «hipótesis de la renta permanente», podemos postular la función de consumo en la forma

$$C_t = \alpha + k(1 - \lambda)Y_t + \lambda C_{t-1} + u_t$$

¹⁰ Op. cit. En el artículo de Zellner no aparece el análisis teórico, pero puede verse Zellner, D. S. Huang y L. C. Chau, «Further Analysis of the Short-Run Consumption with Emphasis on the Role of Liquid Assets», *Econometrica*, vol. 33, julio 1965, pp. 571-81.

⁹ Ver L. R. Klein, *A Textbook of Econometrics*. Evanston, Illinois: R. A. Irwin, 1953, p. 258.

en la que C = consumo real, e Y = renta real. La contrapartida de esta ecuación, sin restricciones, es

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + u_t.$$

Suponemos que el término de perturbación u_t no es autorregresivo. Zellner estimó la función sin restricciones a partir de 31 observaciones trimestrales correspondientes a los Estados Unidos, obteniendo el siguiente resultado:

$$C_t = 0.10 + 0.128 Y_t + 0.870 C_{t-1} + e_t, \quad \bar{R}^2 = 0.978, \\ (0.093) \quad (0.127)$$

Por tanto, las estimaciones de los parámetros con restricciones son

$$\hat{\alpha} = \hat{\beta}_1 = 0.10,$$

$$\hat{k} = \frac{\hat{\beta}_2}{1 - \hat{\beta}_3} = 0.985,$$

$$\hat{\lambda} = \hat{\beta}_3 = 0.870.$$

La estimación del error standard de $\hat{\lambda}$ en las muestras grandes puede obtenerse directamente a partir del resultado sin restricciones. El error standard estimado de k , correspondiente a las muestras grandes, puede obtenerse utilizando la fórmula aproximada (11.40), es decir, observando que

$$\text{Var}(k) = \left[\frac{1}{1 - \hat{\beta}_3} \right]^2 \text{Var}(\hat{\beta}_2) + \left[\frac{\hat{\beta}_2}{(1 - \hat{\beta}_3)^2} \right]^2 \text{Var}(\hat{\beta}_3) \\ + 2 \left[\frac{1}{1 - \hat{\beta}_3} \right] \left[\frac{\hat{\beta}_2}{(1 - \hat{\beta}_3)^2} \right] \text{Cov}(\hat{\beta}_2, \hat{\beta}_3).$$

Las estimaciones de $\text{Var}(\hat{\beta}_2)$ y $\text{Var}(\hat{\beta}_3)$ pueden obtenerse directamente a partir de los resultados de Zellner, y la estimación de $\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)$ se puede obtener sabiendo que

$$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-\sigma^2 m_{23}}{m_{22}m_{33} - m_{23}^2} = \frac{-\sigma^2 r_{23} \sqrt{m_{22}} \sqrt{m_{33}}}{m_{22}m_{33} - m_{23}^2} = -r_{23} \sqrt{\text{Var}(\hat{\beta}_2)} \sqrt{\text{Var}(\hat{\beta}_3)},$$

de modo que la única información adicional que precisamos es el valor de r_{23} , el coeficiente de correlación entre Y_t y C_{t-1} en la muestra. Este valor puede obtenerse a partir de la relación existente entre el estadístico F y los dos estadísticos t . En primer lugar, como $R^2 = 0.978$, según la expresión (10.41) tenemos que

$$R^2 = 0.9795.$$

Por tanto, según (10.43),

$$F = \frac{R^2/(3-1)}{(1-R^2)/(31-3)} = 669.$$

Esto nos permite utilizar la expresión (10.44b):

$$669 = \frac{(0.128/0.093)^2 + (0.870/0.127)^2 + 2(0.128/0.093)(0.870/0.127)r_{23}}{2(1 - r_{23}^2)}.$$

Esto es una ecuación cuadrática en la que r_{23} tiene dos soluciones, una positiva y otra negativa. Elegimos la solución positiva porque la correlación muestral entre Y_t y C_{t-1} es claramente positiva. Esto da como resultado

$$r_{23} = 0.975$$

y

$$\text{Est. Cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.975 \times 0.093 \times 0.127 = -0.011515.$$

Por tanto,

$$\text{Est. Var}(k) = \left[\frac{1}{1 - 0.870} \right]^2 0.093^2 + \left[\frac{0.128}{(1 - 0.870)^2} \right]^2 0.127^2 \\ + \left[\frac{1}{1 - 0.870} \right] \left[\frac{0.128}{(1 - 0.870)^2} \right] (-0.011515) = 0.095175.$$

Así pues, los errores standard estimados, correspondientes a las muestras grandes, de los estimadores con restricciones son

$$s_{\hat{\lambda}} = 0.127 \quad y \quad s_{\hat{k}} = 0.308.$$

lo que significa que tanto $\hat{\lambda}$ como $\hat{k}$ son muy significativos.

Para ilustrar el caso con *restricciones no lineales y superidentificación*, utilizaremos un modelo de regresión simple en el que el término de perturbación sigue un esquema autorregresivo de primer orden. En particular, supongamos que tenemos

$$Y_t = \alpha + \beta X_t + e_t,$$

$$e_t = \rho e_{t-1} + u_t,$$

en el que $u_t \sim N(0, \sigma^2_u)$, y $E(u_t e_{t-1}) = 0$. Tal como se mencionó en la sección 8.2, la ecuación de regresión se puede transformar de modo que se elimine la perturbación autorregresiva e_t . Si tomamos la ecuación de regresión retardada un período, la multiplicamos por ρ y restamos el resultado de la forma original de la ecuación de regresión, obtenemos

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + u_t.$$

Esta ecuación es la (8.52) de la sección 8.2. De modo alternativo,

$$(11.41) \quad Y_t = \alpha(1 - \rho) + \beta X_t - \beta\rho X_{t-1} + \rho Y_{t-1} + u_t,$$

que es una ecuación de regresión múltiple con los parámetros α , β y ρ . La contrapartida de (11.41) sin restricciones es

$$(11.41a) \quad Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + u_t.$$

Comparando las dos versiones, vemos que

$$\beta_1 = \alpha(1 - \rho)$$

$$\beta_2 = \beta,$$

$$\beta_3 = -\beta\rho,$$

$$\beta_4 = \rho.$$

Es decir, tenemos cuatro coeficientes sin restricciones y solamente tres parámetros con restricciones. Como no existe una solución única para ninguno de los parámetros con restricción en términos de los coeficientes sin restricciones, nos hallamos claramente ante un caso de superidentificación. Sin embargo, no existe gran dificultad para estimar los parámetros con restricciones por el método de los mínimos cuadrados no lineales; es decir, podemos minimizar

$$S = \sum_{t=2}^n [Y_t - \alpha(1 - \rho) - \beta X_t + \beta\rho X_{t-1} - \rho Y_{t-1}]^2$$

respecto de α , β y ρ . Las estimaciones resultantes son equivalentes a las estimaciones maximoverosímiles (condicionadas a Y_1), y sus variancias asintóticas pueden estimarse a partir de la matriz de información. En la sección 8.2 se ha presentado una forma sencilla de efectuar los cálculos por medio de una calculadora y se han estudiado también varios métodos alternativos de estimación diseñados para este modelo. El método de la estimación maximoverosímil puede utilizarse con todas las clases de superidentificación, pero en algunos casos los cálculos resultan muy complicados y no siempre existe la garantía de que el máximo de la función de verosimilitud no sea local en lugar de global.

Por último, consideremos la *subidentificación*. Ilustraremos este caso en el contexto de un modelo en el que algunos parámetros están superidentificados, mientras que otros están subidentificados. Consideremos otro modelo de la función de consumo elaborado y estimado por Zellner, Huang y Chau:¹¹

11. *Ibid.*

$$(11.42) \quad C_t = (k - \alpha\eta)(1 - \lambda)Y_t + \alpha L_{t-1} - \alpha\lambda L_{t-2} + \lambda C_{t-1} + u_t,$$

siendo

$$u_t = e_t - \lambda e_{t-1},$$

y L = tenencias reales de activos líquidos al final del período. Por tanto, tenemos que estimar cuatro parámetros, α , λ , k y η . La versión de (11.42) sin restricciones es

$$(11.42a) \quad C_t = \beta_1 Y_t + \beta_2 L_{t-1} + \beta_3 L_{t-2} + \beta_4 C_{t-1} + u_t.$$

En (11.42a) hay cuatro coeficientes sin restricciones, y la relación entre estos últimos y los parámetros de (11.42) es

$$\beta_1 = (k - \alpha\eta)(1 - \lambda),$$

$$\beta_2 = \alpha,$$

$$\beta_3 = -\alpha\lambda,$$

$$\beta_4 = \lambda.$$

A partir de ello, podemos ver que

$$y \quad \alpha = \beta_2 \quad \text{or} \quad \alpha = -\frac{\beta_3}{\beta_4},$$

$$\lambda = \beta_4 \quad \text{or} \quad \lambda = -\frac{\beta_3}{\beta_2}.$$

Por tanto, α y λ están superidentificados, pero no existe solución para k y η . Esto significa que estos dos últimos parámetros están subidentificados. La implicación es que α y λ pueden estimarse por el método de los mínimos cuadrados no lineales, pero a partir de la muestra no se pueden obtener estimaciones de k y η , sino que, como máximo, podemos obtener una estimación de $(k - \alpha\eta)$.

EJEMPLO. Zellner, Huang y Chau han obtenido estimaciones de los parámetros de (11.42) procedentes de datos trimestrales de los Estados Unidos, suponiendo que u_t no es autorregresiva. El resultado de la estimación minimocuadrática no lineal es

$$C_t = 0.475 Y_t + 0.226 L_{t-1} - 0.085 L_{t-2} + 0.378 C_{t-1} + e_t.$$

(0.085) (0.045) (0.106)

Por tanto,

$$\begin{aligned}\hat{\alpha} &= 0.226, \\ \hat{\lambda} &= 0.378, \\ k - 0.226\hat{\eta} &= 0.763.\end{aligned}$$

Se estimó también esta ecuación por el método de los mínimos cuadrados no lineales añadiendo un término constante:

$$C_t = -12.470 + 0.538 Y_t + 0.376 L_{t-1} - 0.091 L_{t-2} + 0.242 C_{t-1} + e_t.$$

(0.080) (0.050) (0.105)

El resultado de la estimación minimocuadrática sin restricción de la ecuación anterior es

$$C_t = -1.082 + 0.517 Y_t + 0.560 L_{t-1} - 0.296 L_{t-2} + 0.273 C_{t-1} + e_t.$$

(3.606) (0.085) (0.172) (0.181) (0.110)

Obsérvese que las estimaciones sin restricciones son numéricamente distintas de las estimaciones con restricciones, pero en algunos coeficientes esta diferencia no es extraordinaria.

Efectos de las restricciones sobre R^2 y la variancia del error del pronóstico

En el caso en que los parámetros con restricciones están exactamente identificados en términos de los coeficientes sin restricciones, las estimaciones de los primeros conducen a la misma ecuación de regresión muestral que las estimaciones de los segundos. Esto significa que el valor de R^2 y el pronóstico del valor de la variable dependiente son los mismos en los dos conjuntos de estimaciones. Sin embargo, si la estimación con restricciones se efectúa en condiciones de superidentificación, la recta de regresión de la muestra basada en las estimaciones con restricciones será distinta de la que está basada en las estimaciones sin restricciones. Las estimaciones minimocuadráticas sin restricciones minimizan la suma de los cuadrados de los residuos y, por tanto, producen el valor máximo posible de R^2 . Esto implica necesariamente que el valor de R^2 correspondiente a la ecuación de regresión muestral con restricciones será menor o igual que el valor de R^2 correspondiente a la ecuación sin restricciones. Este hecho se ha utilizado a veces como argumento a favor de la idea de que los pronósticos del valor de la variable dependiente basados en las estimaciones sin restricciones son mejores que los basados en las estimaciones con restricciones. Si interpretamos el término «mejores» en el sentido de que «la variancia del error del pronóstico es menor (o igual)», el argumento es engañoso. Vamos a demostrarlo en el caso en que existen restricciones lineales y superidentificación, pero la demostración se puede ampliar con el fin de abarcar también a las restricciones no lineales. (En tal caso, la demo-

tración se ve complicada por el hecho de que los estimadores con restricciones son funciones no lineales de las perturbaciones, pero, sin embargo, es perfectamente posible efectuar la demostración en términos de las variancias y covariancias asintóticas.)

Consideremos, por ejemplo, la ecuación de regresión (11.37), que se ha utilizado para ilustrar el caso de la superidentificación con restricciones lineales. La ecuación se presentó en la forma

$$Y_t = \alpha_1 + \alpha_2 X_{t2} + (1 - \alpha_2) X_{t3} + e_t,$$

y la versión sin restricciones se presentó en (11.37a) en la forma

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + e_t.$$

Supongamos ahora que queremos pronosticar el valor de Y correspondiente a determinados valores de las dos variables explicativas, por ejemplo X_{02} y X_{03} . Si utilizamos la ecuación de regresión sin restricciones, efectuaríamos un pronóstico del valor de Y , por ejemplo $\hat{Y}_0$, definido en la forma

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{02} + \hat{\beta}_3 X_{03},$$

en la que las $\hat{\beta}$ representan las estimaciones minimocuadráticas sin restricciones de las β . Mediante la fórmula (10.49), podemos obtener la variancia del error del pronóstico de $\hat{Y}_0$ según la expresión

$$\begin{aligned}(11.43) \quad E(\hat{Y}_0 - Y_0)^2 &= \sigma^2 + \frac{\sigma^2}{n} + (X_{02} - \bar{X}_2)^2 \text{Var}(\hat{\beta}_2) + (X_{03} - \bar{X}_3)^2 \text{Var}(\hat{\beta}_3) \\ &\quad + 2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3) \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_{02} - \bar{X}_2)^2 m_{33} + (X_{03} - \bar{X}_3)^2 m_{22}}{m_{22} m_{33} - m_{23}^2} \right. \\ &\quad \left. - \frac{2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3) m_{23}}{m_{22} m_{33} - m_{23}^2} \right].\end{aligned}$$

Por otra parte, si efectuamos el pronóstico mediante la ecuación de regresión con restricciones, utilizaremos $\hat{Y}_0$ definido como

$$\hat{Y}_0 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{02} + (1 - \hat{\alpha}_2) X_{03},$$

en la que las $\hat{\alpha}$ representan las estimaciones minimocuadráticas con restricciones de las α . En este caso, la variancia del error del pronóstico es

$$\begin{aligned}
 (11.44) \quad E(\tilde{Y}_0 - Y_0)^2 &= E[(\tilde{Y}_0 - X_{03}) - (Y_0 - X_{03})]^2 \\
 &= \sigma^2 + \frac{\sigma^2}{n} + [(X_{02} - X_{03}) - (\bar{X}_2 - \bar{X}_3)]^2 \text{Var}(\hat{\alpha}_2) \\
 &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_{02} - \bar{X}_2)^2 + (X_{03} - \bar{X}_3)^2}{m_{22} + m_{33} - 2m_{23}} \right. \\
 &\quad \left. - \frac{2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3)}{m_{22} + m_{33} - 2m_{23}} \right].
 \end{aligned}$$

Comparemos ahora (11.43) con (11.44). La diferencia entre estas dos variancias es

$$\begin{aligned}
 (11.45) \quad E(\hat{Y}_0 - Y_0)^2 - E(\tilde{Y}_0 - Y_0)^2 &= \sigma^2 \left[\frac{(X_{02} - \bar{X}_2)^2 m_{33} + (X_{03} - \bar{X}_3)^2 m_{22} - 2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3)m_{23}}{m_{22}m_{33} - m_{23}^2} \right. \\
 &\quad \left. - \frac{(X_{02} - \bar{X}_2)^2 + (X_{03} - \bar{X}_3)^2 - 2(X_{02} - \bar{X}_2)(X_{03} - \bar{X}_3)}{m_{22} + m_{33} - 2m_{23}} \right] \\
 &= \sigma^2 \left\{ \frac{[(X_{02} - \bar{X}_2)(m_{33} - m_{23}) + (X_{03} - \bar{X}_3)(m_{22} - m_{23})]^2}{(m_{22}m_{33} - m_{23}^2)(m_{22} + m_{33} - 2m_{23})} \right\}.
 \end{aligned}$$

Esta expresión es claramente positiva o cero, lo que indica que la variancia del error del pronóstico correspondiente al predictor sin restricciones $\tilde{Y}_0$ es mayor, o en el mejor de los casos igual, que la del error del pronóstico correspondiente al predictor con restricciones $\hat{Y}_0$. Esto significa que la ecuación de regresión con restricciones permite efectuar mejores pronósticos que la ecuación sin restricciones.

Función de las restricciones previas en la estimación

Podemos finalizar esta sección efectuando algunos comentarios sobre la importancia de las restricciones previas en la estimación. Las restricciones sobre los coeficientes de la regresión representan información acerca de la ecuación de regresión de la población y, si esta información es correcta, podemos

la eficiencia de los estimadores incorporándola al procedimiento de estimación. El tamaño de este aumento de eficiencia depende del tipo de información (es decir, de lo *específica* que sea) y de las variancias de los estimadores sin restricciones, suponiendo que dichas variancias existen y que los estimadores sin restricciones poseen la propiedad de la consistencia del error al cuadrado. Por ejemplo, supongamos que la restricción toma la forma de la afirmación de que un coeficiente de regresión determinado, por ejemplo β , es positivo. En este caso, si la variancia del estimador sin restricciones de β es pequeña, sólo obtendremos muestras con una estimación de β negativa en muy raras ocasiones. Por tanto, a todos los efectos prácticos, la restricción puede no ser en absoluto «restrictiva». En tal caso, el valor de la información previa puede resultar insignificante. Esto plantea otro de los aspectos acerca de la información previa, consistente en el hecho de que su valor disminuye al aumentar el tamaño de la muestra. Esto se debe a que, en general, las variancias de los estimadores sin restricciones disminuyen cuando el tamaño de la muestra aumenta, y resulta cierto independientemente de lo específica que sea la restricción. Por ejemplo, incluso una restricción tan específica como la que limita β a un número concreto no tiene mucho valor si, en realidad, las estimaciones de β están muy concentradas en la proximidad inmediata del verdadero valor del parámetro. En el límite (es decir, cuando $n \rightarrow \infty$), la información previa no tiene ningún valor. Sin embargo, si las estimaciones tienden a estar muy concentradas en torno a un valor distinto del especificado por la restricción, o bien la información previa es falsa, o bien el modelo está mal especificado y la afirmación de que el estimador sin restricciones es consistente no está justificada. Desgraciadamente, las muestras de que se dispone en economía son típicamente pequeñas y la generación de observaciones adicionales se ve severamente limitada, de modo que la información previa resulta muy valiosa. Por este motivo, debería recomendarse encarecidamente su utilización (olvidada frecuentemente en los trabajos econométricos aplicados).

11.3 Modelos no lineales

Los modelos que hemos considerado hasta aquí se han caracterizado, con escasas excepciones, por tener ecuaciones de regresión lineal. Sin embargo, esto no es tan restrictivo como puede parecer a primera vista, por varias razones. En primer lugar, todos los resultados correspondientes a los modelos de regresión lineales que hemos obtenido en los capítulos anteriores se pueden aplicar sin modificación alguna a los modelos de regresión no lineales respecto a las variables pero lineales respecto a los parámetros que vamos a estimar. En segundo lugar, los modelos de regresión no lineales respecto a las variables y a los parámetros a estimar pueden analizarse utilizando muchos de los principios básicos y de los resultados obtenidos en relación con los modelos puramente lineales. Así pues, los modelos de regresión no lineales se pueden cla-

sificar en dos grupos según que sean o no lineales respecto a los parámetros a estimar. Denominaremos «modelos intrínsecamente lineales» al primero de estos dos tipos, y «modelos intrínsecamente no lineales» a los del segundo tipo, estudiando cada grupo por separado. Además, estudiaremos también el problema de contrastar la existencia de linealidad en los casos en que la forma funcional de la ecuación de regresión resulta dudosa.

Modelos intrínsecamente lineales

Los modelos no lineales pero intrínsecamente lineales son, como hemos indicado, los que resultan *no lineales* respecto a las variables, pero *lineales* respecto a los parámetros a estimar. La característica básica común a estos modelos consiste en que pueden convertirse en modelos lineales ordinarios mediante una transformación adecuada de las variables. Frecuentemente, esta transformación consiste simplemente en redefinir una o más de las variables. Ya hemos utilizado alguno de estos modelos sin ni siquiera detenernos a considerar el hecho de que algunas de las variables aparecían en la ecuación de regresión en forma no lineal. En este apartado vamos a considerar los modelos intrínsecamente lineales de forma sistemática y prestando cierta atención a los detalles. Como primer caso, tomaremos un modelo en el que los regresores vienen representados por una *serie potencial* de X_i :

$$(11.46) \quad Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \cdots + \beta_K X_i^{K-1} + \varepsilon_i,$$

en la que X_i es no estocástica y está limitada, y ε_i cumple todos los supuestos del modelo clásico de regresión normal lineal. Por ejemplo, la especificación puede requerir en primer lugar un aumento y a continuación una disminución en $E(Y_i)$ a consecuencia de los aumentos en X_i , como es el caso en la relación entre la renta y la edad de un individuo. En este caso, la serie potencial vendría expresada por una parábola

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i,$$

representada en la figura 11-5. Pero, cualquiera que sea el grado del polinomio en (11.46), la ecuación puede escribirse nuevamente en la forma

$$(11.46a) \quad Y_i = \beta_1 + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \cdots + \beta_K Z_{iK} + \varepsilon_i,$$

en la que $Z_{i2} = X_i$, $Z_{i3} = X_i^2, \dots, Z_{iK} = X_i^{K-1}$. En este caso, si el número de observaciones es mayor que K , los estimadores minimocuadráticos ordinarios de los coeficientes de regresión de (11.46a) poseerán las propiedades deseables. Sin embargo, las Z estarán a menudo muy correlacionadas, de modo que

las variancias de los coeficientes de regresión estimados pueden ser muy grandes.¹² Una característica muy importante de la función potencial (11.46) o de

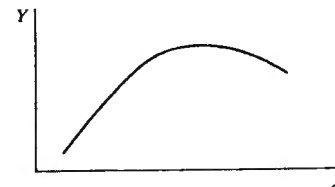


Figura 11-5

su equivalente (11.46a) es que podemos contrastar fácilmente la hipótesis de que el grado del polinomio es menor que $(K-1)$. Por ejemplo, podemos estar interesados en contrastar la hipótesis de que la ecuación de regresión es lineal, respecto a la hipótesis alternativa de que la ecuación de regresión está representada por un polinomio de grado $(K-1)$. La formulación sería

$$H_0: \beta_3 = \beta_4 = \cdots = \beta_K = 0,$$

$$H_1: H_0 \text{ no es cierta.}$$

El correspondiente procedimiento de contrastación se ha presentado en la sección 10.2, y el estadístico de prueba es el de la fórmula (10.45).

La discusión anterior es importante no sólo para los modelos de regresión en que la variable (o variables) explicativa entra en forma de una función potencial, sino para todos los modelos de regresión no lineales en general. La razón de ello es que toda función continua $f(x)$ que posea p derivadas continuas se puede escribir en la forma

$$(11.47) \quad f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \frac{(x-a)^3}{3!}f'''(a) + \cdots + \frac{(x-a)^p}{p!}f^{(p)}(a) + R_{p+1},$$

12. Si los valores muestrales consecutivos de X_i son equidistantes, podemos simplificar los cálculos de las estimaciones minimocuadráticas utilizando la llamada «transformación polinómica ortogonal»; ver, por ejemplo, R. L. Anderson y T. A. Bancroft, *Statistical Theory in Research*, Nueva York, McGraw-Hill, 1952, pp. 207-14. Naturalmente, esto no reduce a las variancias de las estimaciones minimocuadráticas de las β .

en la que a es cualquier valor fijo en el dominio de X y, además,

$$\begin{aligned} f(a) &= f(x) \Big|_{x=a}, \\ f'(a) &= \frac{df(x)}{dx} \Big|_{x=a}, \\ f''(a) &= \frac{d^2f(x)}{dx^2} \Big|_{x=a}, \\ &\vdots \\ R_{p+1} &= \text{residuo.} \end{aligned}$$

La serie presentada en (11.47) se denomina *expansión en serie de Taylor*, de $f(x)$ respecto al valor $x = a$. Si p es suficientemente grande, el residuo R_{p+1} será pequeño y, por tanto, prescindiendo de dicho residuo obtendremos una aproximación de $f(x)$ que puede ser tan exacta como queramos tomando un valor adecuado de p .

EJEMPLO 1. Sea $f(x) = b_0 + b_1x + b_2x^2 + b_3x^3$, y $p = 3$. Supongamos que efectuamos la expansión de $f(x)$ alrededor de $x = 0$. En este caso, tenemos

$$\begin{aligned} f(0) &= b_0, \\ f'(0) &= (b_1 + 2b_2x + 3b_3x^2)_{x=0} = b_1, \\ f''(0) &= (2b_2 + 6b_3x)_{x=0} = 2b_2, \\ &\vdots \\ f'''(0) &= 6b_3, \end{aligned}$$

o sea, que

$$\begin{aligned} f(x) &= b_0 + (x-0)b_1 + \frac{(x-0)^2}{2} 2b_2 + \frac{(x-0)^3}{6} 6b_3 \\ &= b_0 + b_1x + b_2x^2 + b_3x^3, \end{aligned}$$

que es exactamente correcta.

EJEMPLO 2. Sea $f(x) = e^{bx}$, $p = 3$, y $a = 0$. En este caso, tenemos

$$\begin{aligned} f(0) &= 1, \\ f'(0) &= (be^{bx})_{x=0} = b, \\ f''(0) &= (b^2e^{bx})_{x=0} = b^2, \\ f'''(0) &= (b^3e^{bx})_{x=0} = b^3, \end{aligned}$$

o sea, que

$$f(x) \approx 1 + bx + \left[\frac{b^2}{2}\right]x^2 + \left[\frac{b^3}{6}\right]x^3.$$

Si, por ejemplo, $b = 1$, los valores aproximados de $f(x)$ correspondientes a varios valores de x son

x	$f(x)$ exacta	$f(x)$ aproximada
0	1.0000	1.0000
0.5	1.6487	1.6458
1	2.7183	2.6667
2	7.3891	6.3333
3	20.0860	13.0000

Como puede verse, la aproximación se hace cada vez más inexacta a medida que nos alejamos del valor en torno al cual se ha efectuado la expansión.

La expansión en serie de Taylor puede ampliarse con el fin de aplicarla a una función con más de una variable. Concretamente, la expansión correspondiente a dos variables viene dada por

$$\begin{aligned} (11.48) \quad f(x, z) &= f(a, b) + f_x(a, b)(x - a) + f_z(a, b)(z - b) \\ &+ \frac{1}{2!} [f_{xx}(a, b)(x - a)^2 + 2f_{xz}(a, b)(x - a)(z - b) \\ &+ f_{zz}(a, b)(z - b)^2] + \dots, \end{aligned}$$

en la que

$$\begin{aligned} f_x(a, b) &= \frac{\partial f(x, z)}{\partial x} \Big|_{\substack{x=a \\ z=b}}, & f_{xx}(a, b) &= \frac{\partial^2 f(x, z)}{\partial x^2} \Big|_{\substack{x=a \\ z=b}}, \\ f_z(a, b) &= \frac{\partial f(x, z)}{\partial z} \Big|_{\substack{x=a \\ z=b}}, & f_{xz}(a, b) &= \frac{\partial^2 f(x, z)}{\partial x \partial z} \Big|_{\substack{x=a \\ z=b}}, \end{aligned}$$

$$f_{zz}(a,b) = \left. \frac{\partial^2 f(x,z)}{\partial z^2} \right|_{\substack{x=a \\ z=b}},$$

etcétera. La fórmula (11.48) se ha utilizado para la obtención de aproximaciones lineales de los productos y cocientes de dos variables, efectuando expansiones en serie respecto a las medias muestrales de las dos variables y tomando sólo los tres primeros términos de la expansión.¹³ De forma específica,

$$(11.49) \quad X_i Z_i \approx \bar{X}\bar{Z} + \bar{Z}(X_i - \bar{X}) + \bar{X}(Z_i - \bar{Z}) \approx \bar{Z}X_i + \bar{X}Z_i - \bar{X}\bar{Z},$$

$$(11.50) \quad \frac{X_i}{Z_i} \approx \frac{\bar{X}}{\bar{Z}} + \frac{1}{\bar{Z}}(X_i - \bar{X}) + \left[-\frac{\bar{X}}{\bar{Z}^2}\right](Z_i - \bar{Z}) \\ \approx \frac{\bar{X}}{\bar{Z}} + \left[\frac{1}{\bar{Z}}\right]X_i - \left[\frac{\bar{X}}{\bar{Z}^2}\right]Z_i.$$

Consideremos ahora una ecuación de regresión en la que el cambio en $E(Y_i)$ correspondiente a un cambio unitario en X_{ij} depende del valor de X_{ik} . Formalmente, supongamos que el valor medio de la variable dependiente es una función de las dos variables explicativas no estocásticas X_{i2} y X_{i3}

$$E(Y_i) = f(X_{i2}, X_{i3})$$

tal que
$$\frac{\partial E(Y_i)}{\partial X_{i2}} = g(X_{i3})$$

y
$$\frac{\partial E(Y_i)}{\partial X_{i3}} = h(X_{i2}),$$

en las que $g(X_{i3})$ y $h(X_{i2})$ son funciones cuya forma matemática se especifica previamente. Existen muchas situaciones en las que este tipo de modelo resulta adecuado. Por ejemplo, en las estimaciones de las curvas de Engel a partir de los datos de corte transversal, los gastos familiares en un producto determinado se consideran muy a menudo dependientes de la renta familiar y del tamaño de la familia. Puede ser razonable suponer que las familias de distinto tamaño responden de forma distinta a un determinado cambio de renta y que las familias en distintos grupos de renta responden de forma distinta a un cambio en el tamaño familiar. La formulación más simple de este tipo de modelos se obtiene introduciendo en la ecuación lineal de regresión un tér-

mino de interacción, definido como un múltiplo del producto de las dos variables explicativas:

$$(11.51) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i2} X_{i3} + \varepsilon_i.$$

En este modelo
$$\frac{\partial E(Y_i)}{\partial X_{i2}} = \beta_2 + \beta_4 X_{i3}$$

y
$$\frac{\partial E(Y_i)}{\partial X_{i3}} = \beta_3 + \beta_4 X_{i2}.$$

Las dos funciones $g(X_{i3})$ y $h(X_{i2})$ son funciones lineales de las variables respectivas y tienen la misma pendiente β_4 . La ecuación (11.51) es intrínsecamente lineal, ya que podemos escribir $X_{i2} X_{i3} = X_{i4}$.

La existencia de términos de interacción en las ecuaciones de regresión tiene una implicación importante para la contrastación de la hipótesis de que una variable explicativa determinada no resulta relevante en el modelo de regresión de que se trata, es decir, de que no ejerce influencia alguna sobre $E(Y_i)$. Cuando en una ecuación aparecen términos de interacción, cada variable explicativa puede estar representada no por un regresor, sino por varios. La hipótesis de que esta variable no ejerce ninguna influencia sobre $E(Y_i)$ significa que los coeficientes de *todas* los regresores en los que interviene esta variable son cero. Por tanto, la contrastación apropiada para esta hipótesis es la contrastación de la F presentada en (10.45). Normalmente, nos interesa también contrastar la hipótesis de que la variación en $E(Y_i)$ correspondiente a un aumento de una unidad en X_{ik} es constante. Esto equivale a contrastar la hipótesis de que los coeficientes de las «variables» de interacción son cero.

EJEMPLO. La forma de utilizar el modelo de regresión múltiple de la sección 10.1 se ilustró estimando la demanda de naranjas como función lineal del precio y del importe gastado en publicidad. Modifiquemos ahora este ejemplo especificando que la ecuación de regresión debe incluir también un término de interacción. Es decir, supongamos que

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i2} X_{i3} + \varepsilon_i.$$

... 13. Ver Klein, *op. cit.*, pp. 120-21.

en la que Y_i = cantidad vendida de naranjas, X_{i2} = precio, y X_{i3} = gasto en publicidad. Los detalles correspondientes a las doce observaciones muestrales de que se dispone aparecen en el cuadro 10-1. Los resultados de los cálculos básicos son

$$\begin{aligned} m_{22} &= 2250, \\ m_{23} &= -54, & m_{33} &= 4.8573, \\ m_{24} &= 10,237.5, & m_{34} &= 15.6854, & m_{44} &= 66,503.85, \\ m_{Y2} &= -3550, & m_{Y3} &= 125.2500, & m_{Y4} &= -13,542.50, \\ m_{YY} &= 6300. \end{aligned}$$

(El subíndice «4» representa la «variable» de interacción $X_{i2}X_{i3}$.) En este caso, la ecuación de regresión estimada es

$$Y_i = -93.100 + 1.508X_{i2} + 43.990X_{i3} - 0.446X_{i2}X_{i3} + e_i, \quad R^2 = 0.984.$$

(0.832) (9.803) (0.119)

Estos resultados pueden utilizarse para la contrastación de varias hipótesis interesantes. En primer lugar, contrastaremos la hipótesis de que X_{i2} no ejerce ninguna influencia sobre $E(Y_i)$. Formalmente,

$$H_0: \beta_2 = \beta_4 = 0,$$

$$H_A: H_0 \text{ no es cierta.}$$

Según la contrastación presentada en (10.45), la región de aceptación correspondiente a un nivel de significación del 5 %, por ejemplo, es

$$\frac{(SCR_Q - SCR_K)/(Q - K)}{SCE_Q/(n - Q)} \leq F_{Q-K, n-Q, 0.05}.$$

En este problema tenemos $n = 12$, $Q = 4$, y $K = 2$, de modo que $F_{2,8,0.05} = 4.46$, como puede verse en la tabla de la distribución F . Además, a partir de la muestra tenemos

$$\begin{aligned} SCR_Q &= 1,508 \times (-3550) + 43,990 \times 125,25 + (-0,446) \times (-13542,5) = 6198, \\ SCE_Q &= 6300 - 6198 = 102. \end{aligned}$$

El valor de SCR_K se obtiene a través de una regresión de Y_i sobre X_{i3} únicamente, es decir, aplicando el método de estimación por mínimos cuadrados a

$$Y_i = \beta_1^* + \beta_3^* X_{i3} + e_i^*.$$

En este caso,

$$SCR_K = \hat{\beta}_3^* m_{Y3} = \left[\frac{m_{Y3}}{m_{33}} \right] m_{Y3} = \frac{125,25^2}{4,86} = 3228.$$

Y, por tanto,

$$\frac{(SCR_Q - SCR_K)/(Q - K)}{SCE_Q/(n - Q)} = \frac{(6198 - 3228)/(4 - 2)}{102/(12 - 4)} = 116,5,$$

que es considerablemente mayor que el valor crítico 4.46. Por tanto, rechazamos la hipótesis de que X_{i2} no ejerce influencia sobre $E(Y_i)$, aunque la razón t en el coeficiente de X_{i2} es menor que 2,0. A continuación, contrastaremos la misma hipótesis para X_{i3} . La región de aceptación es exactamente igual a la de la primera contrastación, y lo mismo ocurre con todos los cálculos restantes salvo los de SCR_K . El valor de SCR_K puede obtenerse por medio de una regresión de Y_i sobre X_{i2} , o sea estimando los coeficientes de

$$Y_i = \beta_1^{**} + \beta_2^{**} X_{i2} + e_i^{**}.$$

En este caso, tenemos

$$SCR_K = \hat{\beta}_2^{**} m_{Y2} = \left[\frac{m_{Y2}}{m_{22}} \right] m_{Y2} = \frac{(-3550)^2}{2250} = 5602.$$

El valor del estadístico de prueba correspondiente es, pues,

$$\frac{(SCR_Q - SCR_K)/(Q - K)}{SCE_Q/(n - Q)} = \frac{(6198 - 5602)/(4 - 2)}{102/(12 - 4)} = 23,4,$$

y nuevamente es mayor que el valor crítico 4.46. Por último, contrastaremos la hipótesis de que el coeficiente incorporado a la «variable» de interacción $X_{i2} X_{i3}$ es cero; es decir,

$$H_0: \beta_4 = 0,$$

$$H_A: \beta_4 \neq 0.$$

En las tablas, el valor de t para una contrastación con dos extremos a un nivel de significación del 5 % y ocho grados de libertad es 2,306. Por tanto, la región de aceptación del 5 % es

$$-2.306 \leq \frac{\hat{\beta}_4}{s_{\hat{\beta}_4}} \leq 2.306.$$

Como en nuestro caso tenemos

$$\frac{\hat{\beta}_k}{s_{\hat{\beta}_k}} = \frac{-0.446}{0.119} = -3.748,$$

tenemos que rechazar la hipótesis nula.

Otro modelo intrínsecamente lineal es el correspondiente a las ecuaciones de regresión que son lineales en términos de los logaritmos de las variables. Este tipo de modelos se conoce con el nombre de *modelos multiplicativos* y pueden describirse como

$$(11.52) \quad Y_i = \alpha X_{i2}^{\beta_2} X_{i3}^{\beta_3} \cdots X_{iK}^{\beta_K} 10^{\epsilon_i}.$$

Una característica importante de estos modelos es el hecho de que la elasticidad de $E(Y_i)$ respecto de cada una de las variables explicativas es constante, o sea que

$$\frac{\partial E(Y_i)}{\partial X_{ik}} \frac{X_{ik}}{E(Y_i)} = \beta_k \quad (k = 2, 3, \dots, K).$$

Un ejemplo de este tipo de modelo es la función de producción de Cobb-Douglas. Tomando logaritmos (en base 10) en los dos lados de la ecuación (11.52), obtenemos

$$(11.52a) \quad \log Y_i = \log \alpha + \beta_2 \log X_{i2} + \beta_3 \log X_{i3} + \cdots + \beta_K \log X_{iK} + \epsilon_i,$$

que puede escribirse en la forma

$$(11.52b) \quad Y_i^* = \alpha^* + \beta_2 X_{i2}^* + \beta_3 X_{i3}^* + \cdots + \beta_K X_{iK}^* + \epsilon_i,$$

en la que los símbolos con asterisco representan los logaritmos de los símbolos correspondientes sin asterisco. La ecuación (11.52b) es claramente una ecuación lineal ordinaria de regresión múltiple, y puede estimarse a través del procedimiento habitual. Sin embargo, es interesante señalar dos puntos respecto a ella. En primer lugar, la ecuación (11.52b) es lineal en α^* y las β , pero no con respecto a α . Por tanto, si se cumplen los supuestos del modelo clásico de regresión normal lineal, los estimadores minimocuadráticos ordinarios de α^* , $\beta_2, \beta_3, \dots, \beta_K$, poseerán las propiedades deseables. Como

$$\alpha^* = \log \alpha,$$

el estimador de α vendrá dado por

$$\hat{\alpha} = \text{antilog } \hat{\alpha}^*.$$

Aunque $\hat{\alpha}$ recibe todas las propiedades asintóticas deseables de $\hat{\alpha}^*$, las propiedades de esta última correspondientes a las muestras pequeñas, y en particular la insesgadez, no se transmiten a $\hat{\alpha}$. En segundo lugar, si suponemos que ϵ_i se distribuye normalmente con media cero y variancia σ^2 , podemos escribir

$$(11.52c) \quad Y_i = \alpha X_{i2}^{\beta_2} X_{i3}^{\beta_3} \cdots X_{iK}^{\beta_K} \eta_i$$

y suponer que el *logaritmo* (en base 10) de η_i se distribuye normalmente con media cero y variancia σ^2 . La distribución de η_i recibirá el nombre de distribución *normal logarítmica*. Sin embargo, si la ecuación de regresión se especifica en la forma

$$Y_i = \alpha X_{i2}^{\beta_2} X_{i3}^{\beta_3} \cdots X_{iK}^{\beta_K} + \epsilon_i,$$

no existe ninguna transformación de las variables que permita obtener una ecuación de regresión lineal en las β , de modo que la ecuación tiene que clasificarse como intrínsecamente no lineal.

Existen otros muchos modelos no lineales que pueden convertirse en lineales mediante una transformación adecuada. Con el fin de ilustrar el enfoque general, se han seleccionado algunos casos que presentamos a continuación. Una relación no lineal que puede aplicarse con frecuencia es la correspondiente a la *hipérbola*,

$$(11.53) \quad Y_i = \alpha + \beta \frac{1}{X_i} + \epsilon_i \quad (\beta < 0),$$

que describe una situación en la que el valor medio de Y_i tiende a un límite superior al aumentar X_i , tal como se ilustra en la figura 11-6. Por ejemplo,

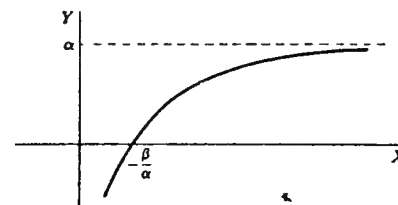


Figura 11-6

este modelo puede aplicarse al estudiar el consumo de alimentos como función de la renta. Obsérvese que

$$\frac{dE(Y_i)}{dX_i} = -\frac{\beta}{X_i^2}.$$

lo que indica que el cambio en $E(Y_i)$ correspondiente a una variación de X_i es inversamente proporcional al cuadrado de X_i . La ecuación (11.53) se puede convertir en una función lineal haciendo simplemente

$$\frac{1}{X_i} = Z_i.$$

De esta forma, obtenemos

$$(11.53a) \quad Y_i = \alpha + \beta Z_i + \varepsilon_i.$$

Una relación similar a la (11.53) es la representada por la *función semi-logarítmica*,

$$(11.54) \quad Y_i = \alpha + \beta \log X_i + \varepsilon_i \quad (\beta > 0).$$

Si la base de $\log X_i$ es $e = 2.71828...$, tenemos¹⁴

$$\frac{dE(Y_i)}{dX_i} = \frac{\beta}{X_i},$$

lo que indica que la variación en $E(Y_i)$ a consecuencia de una variación en X_i es inversamente proporcional al valor de X_i . La principal diferencia entre las expresiones (11.53) y (11.54) es que, en la segunda, el valor de $E(Y_i)$ puede aumentar sin límite. La ecuación (11.54) puede convertirse en una relación lineal haciendo

$$\log X_i = Z_i.$$

Una variación de (11.54) está representada por la función en la que el operador logarítmico aparece en la variable dependiente en lugar de en la explicativa, es decir

$$(11.55) \quad \log Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Si la base de $\log Y_i$ es e , esta relación puede escribirse de nuevo en la forma

$$(11.55a) \quad Y_i = e^{\alpha + \beta X_i + \varepsilon_i},$$

14. Obsérvese que los logaritmos en base 10 (logaritmos decimales) pueden convertirse en logaritmos en base e (logaritmos naturales) multiplicando los primeros por 2.3025850930. Para convertir los logaritmos naturales en logaritmos decimales basta multiplicarlos por 0.4342944819.

que recibe el nombre de *función exponencial*. Una característica importante de (11.55a) consiste en que el cociente entre dos valores consecutivos de $E(Y_i)$ correspondientes a valores de X_i que tienen igual separación entre ellos es una constante. Por ejemplo, si

$$X_{i+1} - X_i = 1,$$

$$\text{tendremos,} \quad \frac{E(Y_{i+1})}{E(Y_i)} = \frac{e^{\alpha + \beta X_{i+1}} E(e^{\varepsilon_{i+1}})}{e^{\alpha + \beta X_i} E(e^{\varepsilon_i})} = e^{\beta}.$$

La transformación

$$\log Y_i = Y_i^*$$

nos permite expresar (11.55) en forma de función lineal

$$(11.55b) \quad Y_i^* = \alpha + \beta X_i + \varepsilon_i.$$

En todos los modelos de regresión que acabamos de estudiar, y en muchos otros, el problema de la estimación de los parámetros de la regresión se simplifica si podemos reducirlos a una forma lineal. El único aspecto de la transformación que merece una cuidadosa atención es el relativo a la perturbación estocástica. En los casos que hemos presentado, esta perturbación estocástica se ha introducido siempre en la relación de forma que podríamos efectuar la transformación lineal sin ninguna dificultad, pero, sin embargo, la especificación del modelo implica la forma en que se introduce el término de perturbación no tiene que estar condicionada por razones de conveniencia matemática o de cálculo. Es importante tener presente que al efectuar este tipo de transformación estamos expresando formalmente nuestro conocimiento previo y nuestras ideas acerca de la relación que queremos representar en el modelo. Como la perturbación estocástica determina la distribución de la variable dependiente para cada conjunto de valores de las variables explicativas, su función en el modelo de regresión resulta crucial. Obviamente, tenemos que ser conscientes de las implicaciones de la especificación concreta que estamos planteando. Por ejemplo, en el caso de la función de producción de Cobb-Douglas presentada en (11.52), la forma particular en que se introduce el término de perturbación en la ecuación implica que la distribución de la producción correspondiente a cada conjunto de valores de los factores es normal logarítmica, y en consecuencia, asimétrica. Por tanto, si queremos insistir en esta especificación concreta, nuestra visión del mundo tiene que ser de este tipo.

Modelos intrínsecamente no lineales

Antes de pasar ahora a los *modelos intrínsecamente no lineales*, es decir, los modelos no lineales respecto a las variables y respecto a los parámetros. Naturalmente,

existe una gran variedad de este tipo de modelos, pero nos limitaremos a estudiar algunos casos interesantes. Consideremos la relación

$$(11.56) \quad Y_i = \alpha X_{i2}^{\beta_2} X_{i3}^{\beta_3} + \varepsilon_i$$

que es, básicamente, el modelo multiplicativo (11.52) *salvo* por el hecho de que la perturbación aparece aquí como un término aditivo y no multiplicativo. Una relación de este tipo puede describir, por ejemplo, la función de demanda de algún producto, siendo Y_i la cantidad demandada, X_{i2} el precio, y X_{i3} la renta. Como

$$\frac{\partial E(Y_i)}{\partial X_{i2}} \frac{X_{i2}}{E(Y_i)} = \beta_2$$

$$y \quad \frac{\partial E(Y_i)}{\partial X_{i3}} \frac{X_{i3}}{E(Y_i)} = \beta_3,$$

β_2 y β_3 representarán las elasticidades-precio, y renta, respectivamente. Como ya hemos señalado anteriormente, no existe ninguna transformación que nos permita convertir (11.56) en una relación lineal respecto a los parámetros. Sin embargo, si X_{i2} y X_{i3} son no estocásticas (o estocásticas, pero independientes de ε_i) y ε_i cumple todos los supuestos del modelo clásico de regresión normal, podemos utilizar el método de estimación de la máxima verosimilitud. La función de verosimilitud de una muestra de tamaño n es

$$(11.57) \quad L = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \alpha X_{i2}^{\beta_2} X_{i3}^{\beta_3})^2.$$

El paso siguiente consiste en derivar L respecto de α , β_2 , β_3 y σ^2 , e igualar a cero estas derivadas. Con ello obtenemos un sistema de cuatro ecuaciones no lineales con cuatro incógnitas. La solución algebraica de este sistema es difícil, pero no hay ninguna dificultad para obtener una solución mediante una calculadora electrónica ya que existen programas para la estimación no lineal. Estos programas se basan fundamentalmente en un enfoque sistemático de «prueba y error», es decir, la calculadora obtiene el valor de L correspondiente a un número de combinaciones distintas de los valores de los parámetros, hasta que se alcanza el valor máximo de L . Los valores de los parámetros correspondientes a este valor máximo de L son las estimaciones máximoverosímiles que pretendíamos obtener, y poseen las propiedades asintóticas deseables.

Otro interesante modelo intrínsecamente no lineal es el llamado *modelo logístico* representado por

$$(11.58) \quad Y_i = \frac{\gamma}{1 + e^{\alpha + \beta X_i}} + \varepsilon_i \quad (\gamma > 0, \beta < 0),$$

en la que $e = 2.71828...$ En este caso, la línea de regresión de la *probabilidad* viene dada por una «curva de crecimiento» logística, tal como se representa

en la figura 11-7. Obsérvese que los valores de $E(Y_i)$ quedan limitados al intervalo comprendido entre 0 y γ . Naturalmente, los valores de Y_i no tienen ninguna restricción. En este caso, podemos también obtener las estimaciones máximoverosímiles de los parámetros de (11.58) por medio de una calculadora. El problema de la estimación resulta aparentemente más simple si la especificación de la curva logística es

$$(11.59) \quad Y_i = \frac{\gamma}{1 + e^{\alpha + \beta X_i + \varepsilon_i}},$$

ya que, en este caso, podemos escribir

$$(11.59a) \quad \log \left(\frac{\gamma}{Y_i} - 1 \right) = \alpha + \beta X_i + \varepsilon_i.$$

La aparente simplificación se debe a que, en este caso, el único parámetro que aparece en forma no lineal en la ecuación de regresión es γ . Sin embargo, debe

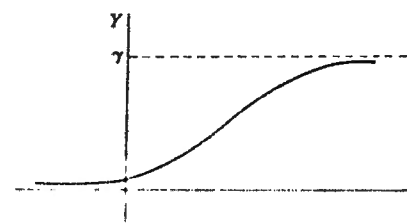


Figura 11-7

señalarse que el modelo (11.59) es extremadamente distinto del presentado en (11.58). Concretamente, en el modelo (11.58) los valores de Y_i correspondientes a un valor de X_i pueden estar comprendidos entre $-\infty$ y $+\infty$, mientras que en (11.59) los valores de Y_i quedan limitados al intervalo entre 0 y γ . Esto implica que, en dicho modelo, la variable dependiente Y_i no puede tener una distribución normal. En la sección 11.1, al considerar el caso de la variable dependiente cualitativa, nos encontramos con el mismo problema.

Como última ilustración de los modelos intrínsecamente no lineales, consideraremos la llamada función de producción CES (de elasticidad de sustitución constante), representada por

$$(11.60)$$

$$Q_i = \gamma [\delta K_i^{-\rho} + (1 - \delta) L_i^{-\rho}]^{-1/\rho} e^{\varepsilon_i}$$

$$(\gamma > 0; 1 > \delta > 0; \rho > 0; \rho \geq -1),$$

en la que Q_i = cantidad producida, K_i = cantidad de capital, L_i = cantidad de trabajo, y $e = 2,71828...$. El parámetro γ recibe el nombre de «parámetro de eficiencia», mientras que δ se denomina «parámetro de distribución», ν , «parámetro de los rendimientos de escala», y ρ , «parámetro de sustitución». Este tipo de función de producción ha obtenido gran popularidad, ya que constituye un caso general e incluye varias funciones de producción más especializadas. Este grado de generalidad lo consigue a través del «parámetro de sustitución» ρ , ya que cuando $\rho = 0$, la función de producción CES se reduce a la función de Cobb-Douglas, mientras que cuando $\rho \rightarrow \infty$ se convierte en la función de producción de «proporciones constantes». Tomando logaritmos (en base e) en los dos lados de la ecuación (11.60), obtenemos

$$(11.60a) \quad \log Q_i = \log \gamma - \frac{\nu}{\rho} \log [\delta K_i^{-\rho} + (1 - \delta)L_i^{-\rho}] + \varepsilon_i.$$

Si K_i y L_i son no estocásticas (o estocásticas pero independientes de ε_i), podemos establecer la función de verosimilitud en la forma habitual, obteniendo las estimaciones maximoverosímiles de α , δ , ν y ρ , con una calculadora.

Es posible obtener una estimación alternativa, y considerablemente más simple, de los parámetros de la función de producción CES sustituyendo la ecuación (11.60a) por su aproximación, que es lineal respecto de ρ . Utilizando la fórmula (11.47) de la expansión en serie de Taylor, podemos efectuar la expansión de $\log Q_i$ respecto de $\rho = 0$, y eliminar los términos en los que ρ aparece con exponentes mayores que la unidad, obteniendo¹⁵

$$(11.60b) \quad \log Q_i = \log \gamma + \nu \delta \log K_i + \nu(1 - \delta) \log L_i - \frac{1}{2} \nu \delta (1 - \delta) [\log K_i - \log L_i]^2 + \varepsilon_i.$$

Obsérvese que el lado derecho de (11.60b) se puede dividir en dos partes, una de las cuales corresponde a la función de producción de Cobb-Douglas, mientras que la otra representa una «corrección» para el caso en que ρ es distinta de cero. Esta última parte, representada por el término $-[\nu \delta (1 - \delta) / 2] [\log K_i - \log L_i]^2$, desaparecerá cuando $\rho = 0$. La estimación de los parámetros de (11.60b) es igual que la estimación en el caso en que existen restricciones no lineales en condiciones de identificación exacta. La versión «sin restricciones» de (11.60b) es

$$(11.60c) \quad \log Q_i = \beta_1 + \beta_2 \log K_i + \beta_3 \log L_i + \beta_4 [\log K_i - \log L_i]^2 + \varepsilon_i,$$

y representa un modelo de regresión intrínsecamente lineal. Si la estimación de β_4 no es significativamente distinta de cero, rechazaremos el modelo CES en favor del modelo de Cobb-Douglas. Los parámetros de la ecuación (11.60b) están relacionados con los de (11.60c) en la forma siguiente:

$$\gamma = \text{antilog } \beta_1,$$

$$\delta = \frac{\beta_2}{\beta_2 + \beta_3},$$

$$\nu = \beta_2 + \beta_3,$$

$$\rho = \frac{-2\beta_4(\beta_2 + \beta_3)}{\beta_2\beta_3}.$$

Por tanto, podemos utilizar las estimaciones minimocuadráticas ordinarias de los β para obtener estimaciones de los parámetros de (11.60b). Los errores standard estimados se pueden calcular utilizando la fórmula aproximada (11.40). Si (11.60b) es una aproximación razonable de (11.60a), y se cumplen los supuestos adecuados acerca de K_i , L_i y ε_i , las estimaciones de los parámetros de la función de producción obtenidos por este procedimiento serán casi asintóticamente eficientes.

EMPLO. Para ilustrar la estimación de la función de producción CES, utilizaremos los datos correspondientes a la producción y los factores de las 25 empresas

15. Todos los logaritmos son naturales. Si utilizásemos logaritmos decimales, tendríamos que multiplicar el término en el que aparece $[\log K_i - \log L_i]^2$ por 2,3026.

presentadas en el cuadro 11-4. El método de la máxima verosimilitud proporciona las estimaciones siguientes de los parámetros de la función de producción:

Cuadro 11-4

Empresa núm.	K_i	L_i	Q_i
1	8	23	106.00
2	9	14	81.08
3	4	38	72.80
4	2	97	57.34
5	6	11	66.79
6	6	43	98.23
7	3	93	82.68
8	6	49	99.77
9	8	36	110.00
10	8	43	118.93
11	4	61	95.05
12	8	31	112.83
13	3	57	64.54
14	6	97	137.22
15	4	93	86.17
16	2	72	56.25
17	3	61	81.10
18	3	97	65.23
19	9	89	149.56
20	3	25	65.43
21	1	81	36.06
22	4	11	56.92
23	2	64	49.59
24	3	10	43.21
25	6	71	121.24

$$\log \bar{\gamma} = 1.0564,$$

$$\delta = 0.4064,$$

$$\bar{\nu} = 0.8222,$$

$$\bar{\rho} = 0.6042.$$

Las estimaciones de los errores standard asintóticos de las estimaciones maximoverosimiles no se han calculado. Utilizando la forma aproximada de la función de producción CES presentada en (11.60b), obtenemos

$$\log \bar{\gamma} = 1.2371,$$

$$\delta = 0.4723,$$

(0.0291)

$$\bar{\nu} = 0.8245,$$

(0.0525)

$$\bar{\rho} = 0.4334,$$

(0.1899)

Las cifras entre paréntesis son las estimaciones de los errores standard asintóticos. Los resultados indican que la muestra no proporciona considerable evidencia contraria al modelo CES.

En general, los parámetros de los modelos intrínsecamente no lineales se pueden estimar formulando la función de verosimilitud y obteniendo las estimaciones maximoverosimiles. Si se cumplen los supuestos clásicos acerca del término de perturbación estocástico y las variables explicativas, las estimaciones resultantes poseerán todas las propiedades asintóticas deseables. Naturalmente, la aplicación del método de la máxima verosimilitud depende del supuesto de que ϵ_i se distribuye normalmente. Si no queremos efectuar este supuesto, podemos obtener estimaciones minimizando la suma de los cuadrados de las desviaciones de los valores observados respecto de los valores ajustados de Y_i , o sea por el método de los mínimos cuadrados. Como los parámetros a estimar aparecen en forma no lineal, este método se denomina normalmente *método de los mínimos cuadrados no lineales*. La principal diferencia entre este método y el de los mínimos cuadrados ordinarios (lineal) consiste en que, en el caso lineal, las estimaciones pueden expresarse en forma de funciones lineales de las perturbaciones, mientras que esto no es generalmente posible en el caso no lineal. Las estimaciones obtenidas por los mínimos cuadrados no lineales son exactamente iguales a las estimaciones maximoverosimiles siempre que la maximización de la función de verosimilitud es efectuada minimizando la suma de los cuadrados de las desviaciones de los valores observados respecto de los valores ajustados de Y_i . Puede mostrarse que, aunque no se efectúa el supuesto de normalidad, la distribución *asintótica* de las estimaciones minimocuadráticas no lineales resulta normal, y tiene la misma media y la misma variancia que la distribución asintótica de las estimaciones maximo-

verosímiles.¹⁶ Por tanto, el supuesto de normalidad de la perturbación estocástica no resulta crucial en todos los casos.

El aspecto más complejo de la estimación no lineal es el cálculo práctico de las estimaciones. Como hemos mencionado anteriormente, dichas estimaciones se obtienen a través de una calculadora, mediante un procedimiento de «prueba y error» sistemático. Esencialmente, la calculadora introduce distintos valores de los parámetros en la función de verosimilitud, obtiene el valor de la función y repite este proceso hasta que no es posible obtener ningún incremento adicional en el valor de la función de verosimilitud. Este procedimiento da buenos resultados si la función de verosimilitud sólo posee un máximo bien definido, pero si la función tiene más de un máximo es muy llana en su parte superior, pueden presentarse problemas. Estas posibilidades se representan gráficamente en la figura 11-8, en el caso en que sólo se intenta obtener el parámetro θ . En la figura 11-8 a) la función de verosimilitud posee un máximo bien definido y no existe ningún problema para obtener $\hat{\theta}$. En la figura 11-8 b), la función de verosimilitud tiene dos máximos, el menor de los cuales es de tipo local, mientras que el segundo es el máximo global de la función. En este caso, la dificultad se plantea debido a que, si la calculadora inicia el proceso para la obtención del valor máximo de θ en un punto próximo a $\hat{\theta}_1$, se detendrá en el punto $\hat{\theta}_1$ y lo presentará como la estimación maximoverosímil de θ , mientras que el valor correcto de esta estimación es realmente $\hat{\theta}_2$. En tal caso, no disponemos de muchas posibilidades, y lo único que podemos hacer es obligar a la calculadora a examinar todos los valores posibles de θ , aunque si el recorrido es infinito esta tarea resulta imposible. Sin embargo, en muchas ocasiones el recorrido de los valores admisibles de los parámetros es finito, en cuyo caso el problema de obtener el máximo global se hace posible. Un buen ejemplo de ello es el de la propensión marginal al consumo, que está limitada a los valores comprendidos entre 0 y 1, o el parámetro de autorregresión ρ , cuyo valor debe hallarse entre -1 y $+1$. El caso de las funciones de verosimilitud relativamente llanas, indicado en la figura 11-8 c) presenta otro problema distinto. En este caso, la función de verosimilitud

similitud es muy sensible a los cambios en los datos muestrales. Esto significa que incluso un pequeño error de medición o un error de redondeo puede desplazar el valor máximo de θ de forma considerable y, por tanto, la estimación resultante no proporciona un elevado grado de confianza.

Contrastaciones de linealidad

En todo lo anterior, el estudio de la estimación de los parámetros de los modelos no lineales se ha basado en el supuesto de que la forma funcional de la ecuación de regresión de la población se conoce (o se supone) a priori. De no ser así, podemos estar interesados en considerar la especificación de la forma funcional como una hipótesis contrastable y no como una hipótesis sostenida. En este contexto, la hipótesis más interesante es la de linealidad. Es decir, frecuentemente queremos contrastar la hipótesis de que la ecuación de regresión de la población es lineal respecto a las variables, contra alguna hipótesis alternativa. La hipótesis de linealidad se puede contrastar de varias formas distintas, pero, desgraciadamente, ninguna de estas contrastaciones está exenta de inconvenientes.

El caso más simple de las contrastaciones de linealidad es aquel en que la hipótesis alternativa consiste en que la ecuación de regresión es una función potencial de grado determinado. Esta contrastación se ha presentado ya en relación con las ecuaciones (11.46) y (11.46a). Su inconveniente es que, como alternativa al modelo lineal, nos tenemos que ceñir a una función potencial de grado específico. Obsérvese que la idea básica de esta contrastación radica en el hecho de que las funciones lineales son un caso especial de las funciones potenciales, en las que el grado es la unidad. Si los coeficientes correspondientes a los términos de grado superior de la variable explicativa son cero, las funciones potenciales se convierten en una regresión lineal simple. Esta idea puede utilizarse especificando otras formas funcionales que incluyen también la relación lineal como un caso especial. Una de estas especificaciones¹⁷ es

$$(11.61) \quad \frac{Y_i^\lambda - 1}{\lambda} = \alpha + \beta \left(\frac{X_i^\lambda - 1}{\lambda} \right) + \epsilon_i.$$

Examinemos esta función para algunos valores determinados de λ . En primer lugar, cuando $\lambda = 0$, las expresiones $(Y_i^\lambda - 1)/\lambda$ y $(X_i^\lambda - 1)/\lambda$ parecen indeterminadas. Sin embargo, obsérvese que todo número finito positivo Z puede escribirse en la forma

$$Z = e^{\log Z},$$

17. Esta especificación ha sido propuesta por G. E. P. Box y D. R. Cox, «An Analysis of Transformations», *Journal of the Royal Statistical Society, Serie B*, vol. 26, 1964, pp. 211-43.

16. Ver E. Malinvaud, *Statistical Methods of Econometrics*, Chicago, Rand McNally, 1966, pp. 290-94.

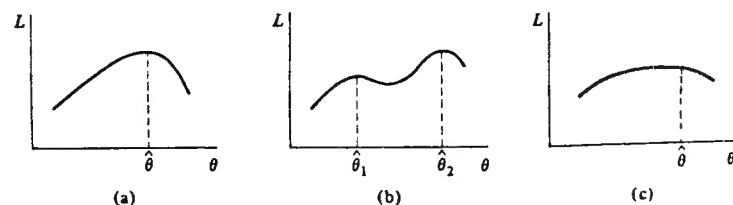


Figura 11-8

en la que la base del logaritmo es e , y $e^{\log Z}$ se puede desarrollar en la forma

$$e^{\log Z} = 1 + \log Z + \frac{1}{2!} (\log Z)^2 + \frac{1}{3!} (\log Z)^3 + \dots$$

Por tanto, vemos que

$$\begin{aligned} \frac{Y_i^\lambda - 1}{\lambda} &= \frac{1}{\lambda} \left[1 + \lambda \log Y_i + \frac{1}{2!} (\lambda \log Y_i)^2 + \dots - 1 \right] \\ &= \log Y_i + \frac{\lambda}{2!} (\log Y_i)^2 + \frac{\lambda^2}{3!} (\log Y_i)^3 + \dots \end{aligned}$$

Cuando $\lambda = 0$,

$$\frac{Y_i^\lambda - 1}{\lambda} = \log Y_i,$$

y, del mismo modo,

$$\frac{X_i^\lambda - 1}{\lambda} = \log X_i.$$

Esto significa que cuando $\lambda = 0$ (y X_i e Y_i son positivas) la ecuación de regresión (11.61) se reduce a

$$\log Y_i = \alpha + \beta \log X_i + \varepsilon_i.$$

Además, cuando $\lambda = 1$, obtenemos

$$(Y_i - 1) = \alpha + \beta(X_i - 1) + \varepsilon_i$$

$$Y_i = \alpha^* + \beta X_i + \varepsilon_i,$$

en la que

$$\alpha^* = \alpha - \beta + 1,$$

que es un modelo de regresión lineal simple.

En general, los distintos valores de λ en (11.61) producirán especificaciones funcionales distintas de la ecuación de regresión.¹⁸ Esto nos permite contrastar la hipótesis de linealidad respecto a la hipótesis alternativa de que la ecuación

18. Hay que prestar cierta atención a las restricciones sobre los valores de la variable dependiente. Según el supuesto de normalidad, el recorrido de ε_i se extiende entre $-\infty$ y $+\infty$, y, por tanto, el recorrido de $(Y_i^\lambda - 1)/\lambda$ debe ser el mismo. Sin embargo, esto puede no ser posible para algunos valores de λ (por ejemplo, $\lambda = 1/2$) en cuyo caso se puede exigir que la distribución sea *aproximadamente* normal. Este punto lo trata en primer lugar J. B. Ramsey.

de regresión es alguna función no lineal perteneciente a la familia de funciones definida por (11.61). Formalmente,

$$H_0: \lambda = 1,$$

$$H_A: \lambda \neq 1.$$

Para efectuar esta contrastación, necesitamos una estimación de λ y de su error standard. Evidentemente, podemos estimar λ junto a los demás parámetros de (11.61) por el método de la máxima verosimilitud. Al establecer la función de verosimilitud para $Y_1, Y_2, \dots, Y_n$, tenemos que obtener la distribución de las Y a partir de la distribución de las ε , que suponemos es normal. Según el teorema 18 de la sección 7.3 (el teorema del «cambio de variables»), tenemos

$$f(Y_i) = \left| \frac{d\varepsilon_i}{dY_i} \right| f(\varepsilon_i).$$

Pero

$$\varepsilon_i = \left(\frac{Y_i^\lambda - 1}{\lambda} \right) - \alpha - \beta \left(\frac{X_i^\lambda - 1}{\lambda} \right)$$

o sea, que

$$\frac{d\varepsilon_i}{dY_i} = Y_i^{\lambda-1}.$$

Por tanto, la función de verosimilitud correspondiente a $Y_1, Y_2, \dots, Y_n$ es

$$\begin{aligned} (11.62) \quad l &= (\lambda - 1) \sum_i \log Y_i - \frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 \\ &\quad - \frac{1}{2\sigma^2} \sum_i \left[\left(\frac{Y_i^\lambda - 1}{\lambda} \right) - \alpha - \beta \left(\frac{X_i^\lambda - 1}{\lambda} \right) \right]^2. \end{aligned}$$

Los valores de λ , α , β y σ^2 que maximizan esta función pueden obtenerse por medio de una calculadora electrónica, y sus errores standard respectivos pueden estimarse a partir de la matriz de información correspondiente. En las muestras grandes, las estimaciones maximoverosímiles se distribuirán normalmente, o por lo menos de forma aproximadamente normal.

La hipótesis de que la ecuación de regresión de la población es lineal respecto a las variables puede contrastarse también sin necesidad de especificar la forma o formas funcionales alternativas. Una manera de hacerlo consiste en considerar de forma más general algunas de las implicaciones de la linealidad. La más obvia de estas implicaciones es que la pendiente y la ordenada en el origen de la ecuación de regresión deben permanecer constantes para todos los valores de la variable explicativa. Podemos, pues, dividir las observaciones muestrales en varias submuestras, correspondiendo cada una de ellas a un intervalo distinto de los valores de la variable explicativa, sin que se produzcan superposiciones, estimar la pendiente y la ordenada en el origen de cada

submuestra, y contrastar si existen diferencias significativas de una submuestra a otra. En la figura 11-9 se representa gráficamente este procedimiento. Con el

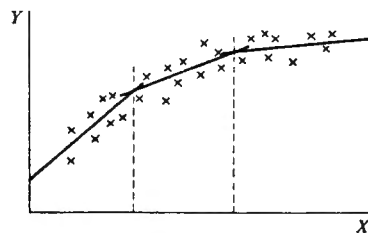


Figura 11-9

fin de indicar la forma en que se lleva a cabo esta contrastación, consideremos una muestra de n observaciones ordenadas de tal forma que

$$X_1 \leq X_2 \leq X_3 \leq \dots \leq X_n.$$

Supongamos que la muestra se divide en tres submuestras, asignando los valores de la variable explicativa en la forma siguiente:

Submuestra 1: $X_1, X_2, \dots, X_k,$

Submuestra 2: $X_{k+1}, X_{k+2}, \dots, X_m$

Submuestra 3: $X_{m+1}, X_{m+2}, \dots, X_n$

El modelo de regresión se puede representar en la forma

$$(11.63) \quad Y_i = \beta_1 + \beta_2 X_i + \gamma_1 Z_{i1} + \gamma_2 X_i Z_{i1} + \gamma_3 Z_{i2} + \gamma_4 X_i Z_{i2} + \epsilon_i,$$

en la que $Z_{i1} = 1$ si i pertenece a la submuestra 1,
 $= 0$ si no pertenece a esta submuestra;
 $Z_{i2} = 1$ si i pertenece a la submuestra 2,
 $= 0$ si no pertenece a esta submuestra.

Podemos contrastar la hipótesis de linealidad utilizando

$$H_0: \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$$

$$H_A: H_0 \text{ no es cierta.}$$

La contrastación de la F , presentada en (10.45), que es adecuada en este caso, presenta el inconveniente de que el resultado puede depender de la forma en que se ha efectuado la división de la muestra. Si el número de submuestras es demasiado pequeño, la no linealidad puede quedar disimulada, mientras que, si el número de submuestras es grande, se pierden muchos grados de libertad y, por tanto, disminuye la potencia de la contrastación. Como compromiso se ha sugerido que, en las aplicaciones prácticas, tres o cuatro muestras resultan suficientes.¹⁹

Otra de las contrastaciones de la linealidad que no se basa en la especificación de la forma funcional alternativa utiliza el diagrama de dispersión de los residuos en torno a la línea de regresión de la muestra. La idea consiste en utilizar el hecho de que, si se cumplen los supuestos del modelo clásico de regresión normal lineal, las perturbaciones se distribuyen aleatoriamente en torno a la línea de regresión de la población. Si esta regresión es no lineal, el diagrama de dispersión de las perturbaciones en torno a una línea recta no será aleatorio. En la figura 11-10 podemos ver que, debido a que la relación no es lineal, las desviaciones respecto a la línea recta tienden a ser negativas al principio, después positivas y nuevamente negativas. Esto sugiere que podemos contrastar la linealidad analizando si la secuencia de las desviaciones respecto de la línea de regresión se distribuye aleatoriamente.

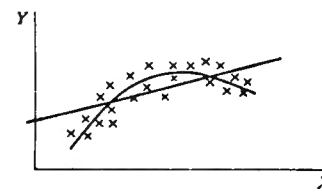


Figura 11-10

Si efectuamos una contrastación de este tipo nos encontramos con la obvia dificultad de que las desviaciones respecto de la línea de regresión de la po-

¹⁹ Malinvaud, *op. cit.*, p. 270.

blación (es decir, las perturbaciones) no son observables. Lo único que tenemos son las desviaciones respecto de la línea de regresión de la muestra (o sea, los residuos), pero los residuos no son independientes aunque las perturbaciones lo sean. Concretamente, para $i \neq j$, tenemos

$$\begin{aligned}
 (11.64) \quad \text{Cov}(e_i, e_j) &= E(y'_i - \beta x'_i)(y'_j - \beta x'_j) \\
 &= E(\beta x'_i + e'_i - \beta x'_j - e'_j)(\beta x'_j + e'_j - \beta x'_i) \\
 &= x'_i x'_j E(\beta - \beta)^2 - E(\beta - \beta) e'_i x'_j - E(\beta - \beta) e'_j x'_i + E(e'_i e'_j) \\
 &= -x'_i x'_j \left[\frac{\sigma^2}{\sum x_i'^2} \right] - \frac{\sigma^2}{n},
 \end{aligned}$$

que es claramente distinta de cero. La contrastación de Durbin-Watson presentada en la sección 8.2, y diseñada originalmente para contrastar la hipótesis de que las perturbaciones no están correlacionadas a lo largo del tiempo, respecto a la hipótesis de que siguen un esquema autorregresivo de primer orden, tiene en cuenta esta diferencia entre las e y las ε . Ordenando los residuos según los valores crecientes de la variable explicativa, en lugar de hacerlo según el período de tiempo, puede utilizarse dicha contrastación para investigar si las desviaciones respecto a la línea de regresión de la población son aleatorias. Sin embargo, la contrastación de Durbin-Watson no carece de inconvenientes. Si la ordenación de los residuos según el período de tiempo es similar a la ordenación según los valores crecientes de la variable explicativa, estaremos utilizando básicamente la misma contrastación para la linealidad que para la no existencia de autorregresión en las perturbaciones a lo largo del tiempo. Esta dificultad se debe al hecho de que, cuando las ordenaciones son similares, tanto la no linealidad como la autorregresión de las perturbaciones poseen las mismas implicaciones respecto a los residuos.

La construcción de la contrastación de Durbin-Watson se ha visto motivada por el hecho de que los residuos minimocuadráticos no son independientes entre sí, aun en el caso de que las perturbaciones lo sean, como se ha mostrado explícitamente en (11.64). Sin embargo, cabe señalar que la expresión correspondiente a la covarianza entre e_i y e_j ($i \neq j$) tiende a cero cuando el tamaño muestral tiende a infinito, de modo que, en las muestras grandes, los residuos serán casi independientes si las perturbaciones lo son. Por tanto, en las muestras grandes podemos utilizar cualquier contrastación de tipo general de la «aleatoriedad» sin correr el riesgo de cometer un error muy grande. Una de estas contrastaciones consiste en examinar los máximos y mínimos de una serie para ver si se pueden considerar debidos simplemente al azar. Podemos aplicar esta contrastación a la secuencia de los residuos ordenados según los valores

crecientes de la variable explicativa. Diremos que un residuo e_i es un «máximo» si

$$e_{i-1} < e_i > e_{i+1}$$

y un «mínimo» si

$$e_{i-1} > e_i < e_{i+1}.$$

Si dos o más residuos sucesivos tienen el mismo valor, y este valor es superior al de los residuos más próximos, consideraremos que constituyen conjuntamente un máximo, y lo mismo en el caso de los mínimos. Podemos designar a los máximos y los mínimos con el nombre genérico de «valores extremos». Puede mostrarse²⁰ que en una serie de n valores independientes el número total de valores extremos, p , cuando n es grande, se distribuye de forma aproximadamente normal con media

$$(11.65) \quad E(p) = \frac{2(n-2)}{3}$$

y variancia

$$(11.66) \quad \text{Var}(p) = \frac{16n-29}{90}.$$

La contrastación de la independencia implica, pues, contar el número de valores extremos en la serie de los residuos (ordenada según los valores crecientes de X) y comprobar si este número es significativamente distinto de $E(p)$. Es decir, la hipótesis de independencia (y, por tanto, probablemente la de linealidad) es

$$H_0: E(p) = \frac{2(n-2)}{3}$$

con una H_A de dos extremos. En las muestras grandes, el estadístico de prueba adecuado sería

$$(11.67) \quad \frac{p - 2(n-2)/3}{\sqrt{(16n-29)/90}} \sim N(0,1).$$

Esta contrastación no requiere el supuesto de que las perturbaciones se distribuyen normalmente. Su inconveniente, igual que en el caso de la contrastación

20. G. Udny Yule y M. G. Kendall, *An Introduction to the Theory of Statistics*. Londres, Charles Griffin, 1950, p. 638.

de Durbin-Watson, es que si la ordenación de los residuos según los valores crecientes de X es similar a la ordenación temporal, no podremos distinguir entre la no linealidad y la existencia de autorregresión en las perturbaciones a lo largo del tiempo.

Junto a los distintos inconvenientes que son específicos de cada una de ellas, todas las contrastaciones de la linealidad poseen un inconveniente común que reduce su utilidad: los datos muestrales no proporcionan información alguna acerca de la forma de la línea de regresión en el exterior del intervalo abarcado por los valores observados de la variable explicativa. Si la regresión de la población es aproximadamente lineal *dentro* de este intervalo, pero no lineal fuera de él, ninguna de las contrastaciones es capaz de detectar la no linealidad de la regresión de la población. En la figura 11-11 se representa gráficamente esta situación.

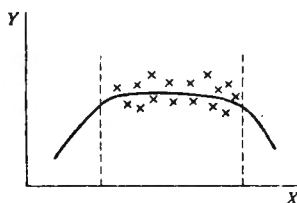


Figura 11-11

A partir de los datos muestrales de que se dispone, no es posible hacer nada para solucionar este problema. La única forma posible de resolver esta dificultad consiste en aumentar la muestra y ampliar el intervalo abarcado por los valores observados de la variable explicativa. En realidad, si a priori se sospecha la linealidad de la regresión de la población, éste sería ciertamente el procedimiento recomendado. En el pasado, la mayor parte del trabajo econométrico aplicado se ha efectuado únicamente en base a modelos lineales, principalmente por motivos de simplicidad de cálculo. Sin embargo, con la difusión de las calculadoras electrónicas esta razón ha perdido parte de su fuerza, y existen pocos motivos que justifiquen la no utilización de modelos no lineales cuando la teoría sugiere su adopción, o la no contrastación de la presencia de linealidad cuando resulta dudoso suponer su existencia.

11.4 Modelos con retardos distribuidos

Para explicar las variaciones a lo largo del tiempo en el valor medio de una variable dependiente podemos utilizar la ecuación de regresión simple

$$Y_t = \alpha + \beta X_t + \varepsilon_t,$$

en la que ε_t es una variable aleatoria con media cero (y que satisface los demás supuestos especificados), y X_t es no estocástica o, en caso de que sea estocástica, es independiente de ε_t . Al establecer la ecuación de regresión de esta forma, estamos suponiendo en realidad que el valor actual de Y depende del valor actual de X , pero *no* de ninguno de los valores previos de X . Una formulación más general, en la que el valor de Y depende del valor actual y de los valores previos de X , puede escribirse en la forma

$$(11.68) \quad Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_m X_{t-m} + \varepsilon_t.$$

Esta ecuación constituye la base para el análisis de esta sección. Por razones de simplicidad nos limitaremos al caso en que X es no estocástica, pero el análisis puede modificarse fácilmente para considerar el caso en que X es estocástica e independiente de ε .

La ecuación de regresión (11.68), unida a los supuestos relevantes acerca de la conducta de X y ε , se denomina *modelo con retardos distribuidos*, ya que la influencia de la variable explicativa sobre $E(Y_t)$ se distribuye a lo largo de un número de valores previos de X . Este número de valores, m , puede ser finito o infinito, pero, sin embargo, suponemos que la suma de las β es finita; es decir,

$$\sum_{i=0}^m \beta_i < \infty.$$

Este supuesto resulta necesario para eliminar la posibilidad de que aparezcan valores explosivos de $L(Y_t)$. Denominaremos *retardo medio* a la media ponderada de los retardos que aparecen en la ecuación, utilizando como ponderaciones los tamaños relativos de los coeficientes β respectivos. Formalmente,

$$\text{retardo medio} = \frac{\sum_i \beta_i}{\sum \beta_i},$$

en la que los sumatorios se extienden entre $i = 0$ e $i = m$. En principio, la ecuación (11.68) podría estimarse por el método de los mínimos cuadrados o por cualquier otro método que produzca estimaciones que posean algunas de las propiedades deseables de acuerdo con las especificaciones efectuadas para ε . Naturalmente, si m es grande, es posible que no dispongamos de una cantidad suficiente de observaciones para estimar todos los parámetros. Sin embargo, aun en el caso de que dispongamos de suficientes observaciones, es probable que nos encontremos con un elevado grado de multicolinealidad, lo cual producirá un efecto perjudicial en los errores standard de los coeficientes estimados. En realidad, los modelos con retardos distribuidos se han planteado y estimado en muy pocas ocasiones en una forma tan general como la especificada en (11.68). Frecuentemente, se han efectuado restricciones acerca de los coeficientes de regresión $\beta_0, \beta_1, \dots, \beta_m$, de modo que el número de los parámetros de la regresión ha quedado sustancialmente reducido. En las aplicacio-

nes prácticas, estas restricciones han tomado dos formas, consistiendo una de ellas en el requisito de que las β deben disminuir en progresión geométrica, y la segunda en el requisito de que las β deben aumentar en primer lugar, disminuyendo posteriormente.

Retardos geométricos

La forma más utilizada de los retardos distribuidos es sin duda la *distribución con retardos geométricos*, representada por la ecuación

$$(11.69) \quad Y_t = \alpha + \beta_0(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + \varepsilon_t,$$

en la que $0 \leq \lambda < 1$.

En este caso, el efecto de X sobre $E(Y_t)$ se extiende indefinidamente en el pasado (es decir, $m \rightarrow \infty$), pero los coeficientes disminuyen en una proporción constante y, por tanto, el efecto de los valores de X muy distantes en el pasado se hace insignificante. Este modelo se ha racionalizado de dos formas distintas, cada una de las cuales da como resultado la misma descripción de la ecuación de regresión de la población, pero posee implicaciones distintas acerca del comportamiento del término de perturbación de la regresión. La primera de estas racionalizaciones se conoce con el nombre de *modelo de expectativas adaptables*, y se basa en el razonamiento siguiente. Supongamos que se modifica un modelo de regresión simple, de forma que $E(Y_t)$ no es función de X_t , sino del valor «esperado» o «permanente» de X en el momento t , al que llamaremos X_t^* . Un ejemplo de ello sería la relación de demanda en la que la cantidad demandada es función del precio esperado, o una función de consumo en la que la variable explicativa fuese la «renta permanente». En este caso,

$$(11.70) \quad Y_t = \alpha + \beta X_t^* + \varepsilon_t,$$

en la que ε_t es, como en el caso anterior, una variable aleatoria con media cero. Como X_t^* no es directamente observable, tenemos que especificar cómo se determina. Para ello, postularemos que

$$(11.71) \quad X_t^* - X_{t-1}^* = (1 - \lambda)(X_t - X_{t-1}^*),$$

o, de forma equivalente,

$$(11.71a) \quad X_t^* = (1 - \lambda)X_t + \lambda X_{t-1}^*,$$

en la que

$$0 \leq \lambda < 1.$$

Esto presupone que el valor esperado o permanente de X en el momento t está representado por una media ponderada del valor actual de X y el valor esperado de X en el período anterior. Este proceso de formación de las expectativas se basa en la idea de que las expectativas del momento actual se obtienen modificando las expectativas previas según la experiencia del momento presente. Por ejemplo, en el caso de una función de demanda, X_t puede ser el precio al iniciarse el período actual, y X_{t-1}^* el precio medio que se esperaba rigiera durante el período anterior. O en el caso de la función de consumo, la renta «permanente» actual puede determinarse revisando el nivel de la renta «permanente» del período anterior en base a la experiencia de la renta en el período actual. Obsérvese que la ecuación (11.71a) puede escribirse también en la forma

$$(11.71b) \quad X_t^* = (1 - \lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots).$$

La equivalencia entre las ecuaciones (11.71a) y (11.71b) puede demostrarse por medio de la llamada «transformación de Koyck».²¹ Retardando un período la ecuación (11.71b), y multiplicando ambos lados por λ , obtenemos

$$\lambda X_{t-1}^* = (1 - \lambda)(\lambda X_{t-1} + \lambda^2 X_{t-2} + \dots).$$

Si restamos esta ecuación de (11.71 b), tenemos

$$X_t^* - \lambda X_{t-1}^* = (1 - \lambda)X_t,$$

que es igual a (11.71a). Por tanto, (11.71a) y (11.71b) son equivalentes. Sustituyendo X_t^* por su valor, según (11.71b), en la ecuación (11.70), obtenemos

$$(11.72) \quad Y_t = \alpha + \beta(1 - \lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + \varepsilon_t,$$

que representa el modelo con retardos geométricos (11.69) cuando $\beta_0 = \beta(1 - \lambda)$. En este caso, el valor del retardo medio es

$$\frac{\beta(1 - \lambda)(0 + \lambda + 2\lambda^2 + 3\lambda^3 + \dots)}{\beta(1 - \lambda)(1 + \lambda + \lambda^2 + \lambda^3 + \dots)} = \frac{\beta(1 - \lambda)[\lambda/(1 - \lambda)^2]}{\beta(1 - \lambda)[1/(1 - \lambda)]} = \frac{\lambda}{1 - \lambda}.$$

Evidentemente, la ecuación (11.72) resulta inconveniente desde el punto de vista de la estimación, debido al número infinito de regresores. Sin embargo, podemos simplificarla aplicando la transformación de Koyck. Si tomamos

21. L. M. Koyck, *Distributed Lags and Investment Analysis*. Amsterdam, North-Holland Publishing Company, 1954.

Evidentemente, como X_t es no estocástica, constituye una variable instrumental perfecta para sí misma, ya que satisface las condiciones necesarias y está «perfectamente correlacionada» consigo misma. La segunda variable instrumental satisface también las condiciones necesarias y es probable que esté correlacionada con Y_{t-1} . Las ecuaciones normales correspondientes a las estimaciones por variables instrumentales son, pues,

$$\begin{aligned}\sum_t Y_t &= \alpha_0^*(n-1) + \beta_0^* \sum_t X_t + \lambda^* \sum_t Y_{t-1}, \\ \sum_t X_t Y_t &= \alpha_0^* \sum_t X_t + \beta_0^* \sum_t X_t^2 + \lambda^* \sum_t X_t Y_{t-1}, \\ \sum_t X_{t-1} Y_t &= \alpha_0^* \sum_t X_{t-1} + \beta_0^* \sum_t X_t X_{t-1} + \lambda^* \sum_t X_{t-1} Y_{t-1},\end{aligned}$$

o, en notación matricial,

$$Z'Y = (Z'X)b^*,$$

en la que

$$Y = \begin{bmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & X_2 & X_1 \\ 1 & X_3 & X_2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_{n-1} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_2 & Y_1 \\ 1 & X_3 & Y_2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & Y_{n-1} \end{bmatrix}, \quad b^* = \begin{bmatrix} \alpha_0^* \\ \beta_0^* \\ \lambda^* \end{bmatrix}.$$

Nótese que la primera observación correspondiente a Y_t es Y_1 y, por tanto, los sumatorios se extienden entre $t = 2$ y $t = n$. La solución de las ecuaciones normales es

$$(11.86) \quad b^* = (Z'X)^{-1}(Z'Y).$$

La matriz de variancia y covariancias asintóticas de b^* puede estimarse de la forma siguiente:

$$(11.87) \quad \text{Est. asintótica Var-Cov}(b^*) = s^{*2}(Z'X)^{-1}(Z'AZ)(X'X)^{-1},$$

$$\text{en la que} \quad s^{*2} = \frac{1}{n-4} \sum_{t=2}^n (Y_t - \alpha_0^* - \beta_0^* X_t - \lambda^* Y_{t-1})^2$$

y A es una matriz $n \times m$ con $a_{tt} = 1$, $a_{t,t-1} = \lambda$, $t = 2, \dots, n$, $a_{t,t+1} = 1$, $t = 1, \dots, n-1$, y ceros en el resto.

A partir de las estimaciones consistentes de α_0 , β_0 , y λ podemos obtener fácilmente estimaciones consistentes de α y β , y sus variancias asintóticas pueden estimarse utilizando (11.40) junto a (11.87).

Los coeficientes de (11.73) pueden estimarse a través de un enfoque alternativo basado en el principio de la *máxima verosimilitud*. En este caso, los estimadores pueden obtenerse de la forma siguiente.²³ Podemos escribir la ecuación (11.73) en la forma

$$(Y_t - \varepsilon_t) = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda(Y_{t-1} - \varepsilon_{t-1}),$$

$$\text{o} \quad E(Y_t) = \alpha_0 + \beta_0 X_t + \lambda E(Y_{t-1}).$$

Sin embargo, como

$$E(Y_{t-1}) = \alpha_0 + \beta_0 X_{t-1} + \lambda E(Y_{t-2}),$$

$$E(Y_{t-2}) = \alpha_0 + \beta_0 X_{t-2} + \lambda E(Y_{t-3}),$$

$\vdots$

$$E(Y_1) = \alpha_0 + \beta_0 X_1 + \lambda E(Y_0),$$

también podemos escribir

$$(11.88) \quad E(Y_t) = \alpha_0(1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}) + \beta_0(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots + \lambda^{t-1} X_1) + \lambda^t E(Y_0).$$

Utilizando la fórmula de la suma de los términos de una progresión geométrica, y el hecho de que $\alpha_0 = \alpha(1 - \lambda)$, podemos reformular la ecuación (11.88) en la forma

$$(11.88a) \quad Y_t = \alpha + \beta_0 W_t^{(\lambda)} + (\theta_0 - \alpha)\lambda^t + \varepsilon_t,$$

$$\text{en la que} \quad W_t^{(\lambda)} = X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots + \lambda^{t-1} X_1$$

$$\text{y} \quad \theta_0 = E(Y_0).$$

El valor de $E(Y_0)$ es el valor medio inicial de Y , y puede considerarse como un parámetro. Obsérvese que, si conociésemos el valor de λ , la ecuación (11.88a) se convertiría en una ecuación lineal de regresión múltiple con dos variables explicativas, $W_t^{(\lambda)}$ y λ^t , y podría estimarse por el método de los mínimos cuadrados ordinarios. Naturalmente, el valor de λ no suele conocerse y

23. Estos estimadores se han obtenido en el apéndice del artículo de L. R. Klein, «The Estimation of Distributed Lags», *Econometrica*, vol. 26, octubre 1958, pp. 553-65, y se han desarrollado de forma operacional en P. J. Dhrymes, «Efficient Estimation of Distributed Lags with Autocorrelated Error Terms», *International Economic Review*, vol. 10, febrero 1969, pp. 47-67, y en A. Zellner y M. S. Geisel, «Analysis of Distributed Lag Models with Applications to Consumption Function Estimation», manuscrito, Universidad de Chicago (enero 1968), pendiente de publicación en *Econometrica*. La presentación efectuada aquí sigue estrechamente la de Zellner y Geisel.

debe estimarse junto a los valores de α , β_0 , y θ_0 . La función logarítmica de verosimilitud correspondiente a $Y_1, Y_2, \dots, Y_n$ es

$$(11.89) \quad L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [Y_i - \alpha - \beta_0 W_i^{(\lambda)} - (\theta_0 - \alpha)\lambda^i]^2.$$

La maximización de L respecto de α , β_0 , λ , y θ_0 equivale a minimizar

$$S^{(\lambda)} = \sum_{i=1}^n [Y_i - \alpha - \beta_0 W_i^{(\lambda)} - (\theta_0 - \alpha)\lambda^i]^2$$

respecto a los mismos parámetros. Como sabemos que $0 \leq \lambda < 1$, podemos calcular fácilmente los valores de α , β_0 , y θ_0 que minimizan esta expresión, y el valor de $S^{(\lambda)}$, correspondientes a distintos valores de λ comprendidos entre 0 y 0.95 ó 0.99, tomando entonces los valores de α , β_0 , θ_0 , y λ que producen el valor mínimo de $S^{(\lambda)}$. Estos valores serán las estimaciones maximoverosímiles de los parámetros respectivos. Si no tenemos más información acerca de X que los valores muestrales de $X_1, X_2, \dots, X_n$, las estimaciones maximoverosímiles de α , β_0 , y λ obtenidas por este procedimiento serán asintóticamente eficientes y sus variancias asintóticas pueden estimarse utilizando la matriz de información correspondiente.

EJEMPLO. El método anterior se ha utilizado para estimar los parámetros de un modelo de la función de consumo con observaciones trimestrales correspondientes a los Estados Unidos, desde 1947(I) a 1960(IV)²⁴. La función de consumo se obtiene a través de un modelo de expectativas adaptables y puede describirse en la forma

$$C_t = \beta Y_t^* + \varepsilon_t,$$

$$Y_t^* - Y_{t-1}^* = (1 - \lambda)(Y_t - Y_{t-1}^*),$$

en la que C = consumo real observado, Y^* = renta real «normal» o «permanente», e Y = renta real observada. Combinando las dos ecuaciones y eliminando Y^* , obtenemos

$$C_t = \beta(1 - \lambda)Y_t + \lambda C_t + \varepsilon_t - \lambda \varepsilon_{t-1}.$$

Los coeficientes se han estimado a partir de

$$C_t = \beta(1 - \lambda)W_t^{(\lambda)} + \theta_0 \lambda^t + \varepsilon_t,$$

en la que $W_t^{(\lambda)} = Y_t + \lambda Y_{t-1} + \dots + \lambda^{t-1} Y_1$

y $\theta_0 = E(Y_0)$.

24. Ver Zellner y Geisel, *op. cit.* Los datos sobre gastos personales de consumo y renta personal disponible, ambas series desestacionalizadas y divididas por el índice de precios, aparecen en Z. Griliches *et al.*, «Notes on Estimated Aggregate Quarterly Consumption Function», *Econometrica*, vol. 30, julio 1962, pp. 491-500.

Se ha calculado la suma de los cuadrados de los residuos (dividida por el número de observaciones) correspondiente a distintos valores de λ comprendidos entre 0 y 1, representando gráficamente los resultados en la figura 11-12. La curva tiene un mínimo local en el punto $\lambda = 0.45$, y un mínimo global en $\lambda = 0.963$. Este último punto es, pues, la estimación maximoverosímil de λ . Las estimaciones de β y θ_0 para este punto son

$$\hat{\beta} = 1.129, \quad \text{y} \quad \hat{\theta}_0 = 191.53$$

Como se supone que β representa la propensión marginal al consumo respecto a la renta «normal», los valores superiores a la unidad no resultan aceptables a priori y, por tanto, no podemos aceptar estos resultados. (Zellner y Geisel señalan que este resultado puede ser debido a las insuficiencias del modelo concreto y/o de los datos. Si se considera que tanto uno como otros son adecuados, debemos incorporar la restricción $0 < \beta < 1$ al procedimiento de estimación.)

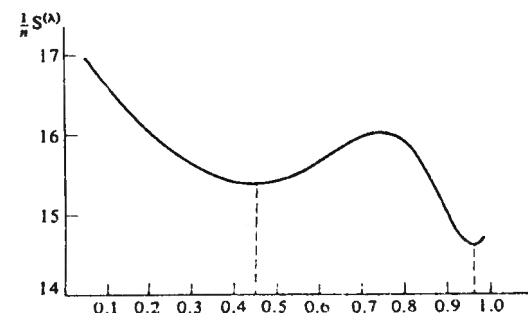


Figura 11-12

Podemos utilizar otro enfoque distinto para estimar los coeficientes de (11.73) observando que

$$(Y_t - \varepsilon_t) = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda(Y_{t-1} - \varepsilon_{t-1})$$

resulta formalmente similar al modelo de los «errores en las variables» estudiado en la sección 9.1. Es decir, podemos considerar ε_t como un «error» en el que hemos incurrido al medir $E(Y_t)$, y ε_{t-1} como un «error» en la medición de $E(Y_{t-1})$. Como suponemos que ε es homocedástica, la variancia de ε_t será igual a la de ε_{t-1} . Por tanto, podemos utilizar el método de la regresión ponderada asignando ponderaciones iguales a todos los errores. Concretamente, podemos minimizar

$$(11.90) \quad S = \sum_t \varepsilon_t^2 + \sum_t \varepsilon_{t-1}^2 = \sum_t [Y_t - \alpha(1 - \lambda) - \beta(1 - \lambda)X_t - \lambda \theta_{t-1}]^2 + \sum_t [Y_{t-1} - \theta_{t-1}]^2,$$

en la que $\theta_{t-1} = E(Y_{t-1})$,

respecto de α , β , λ , y cada una de las θ . Como el primero de los valores conocidos de Y es Y_1 , los sumatorios se extienden entre $t = 2$ y $t = n$. Las condiciones de primer grado para la minimización son

$$\sum_t (Y_t - \bar{\alpha}_0 - \beta_0 X_t - \lambda \theta_{t-1}) = 0,$$

$$\sum_t X_t (Y_t - \bar{\alpha}_0 - \beta_0 X_t - \lambda \theta_{t-1}) = 0,$$

$$\sum_t \theta_{t-1} (Y_t - \bar{\alpha}_0 - \beta_0 X_t - \lambda \theta_{t-1}) = 0,$$

$$\lambda (Y_t - \bar{\alpha}_0 - \beta_0 X_t - \lambda \theta_{t-1}) + (Y_{t-1} - \theta_{t-1}) = 0,$$

en las que $\bar{\alpha}_0 = \alpha(1 - \lambda)$,

$$\beta_0 = \beta(1 - \lambda),$$

$$t = 2, 3, \dots, n.$$

Según la primera condición, tenemos

$$\bar{\alpha}_0 = \bar{Y} - \beta_0 \bar{X} - \lambda \left[\frac{1}{n-1} \sum_t \theta_{t-1} \right].$$

Sustituyendo este resultado en las restantes condiciones de primer grado, podemos eliminar $\bar{\alpha}_0$, y expresar todas las variables en términos de desviaciones respecto de sus correspondientes medias muestrales. Una vez hecho esto, podemos expresar la última condición en la forma

$$\theta'_{t-1} = \frac{\lambda(y'_t - \beta_0 x'_t) + y'_{t-1}}{1 + \lambda^2},$$

en la que

$$\theta'_{t-1} = \theta_{t-1} - \frac{1}{n-1} \sum_t \theta_{t-1},$$

$$y'_t = Y_t - \bar{Y},$$

$$x'_t = X_t - \bar{X},$$

$$y'_{t-1} = Y_{t-1} - \frac{1}{n-1} \sum_t Y_{t-1}.$$

Sustituyendo $\bar{\theta}_{t-1}$ por su valor en la segunda y tercera condiciones de primer grado, y eliminando β_0 , obtenemos

$$(11.91) \quad \bar{\lambda}^2 \left[\frac{\sum x'_t y'_t \sum x'_t y'_{t-1}}{\sum x'^2_t} - \sum y'_t y'_{t-1} \right] + \bar{\lambda} \left[\frac{(\sum x'_t y'_{t-1})^2 - (\sum x'_t y'_t)^2}{\sum x'^2_t} + \sum y'^2_t - \sum y'^2_{t-1} \right] + \left[\sum y'_t y'_{t-1} - \frac{\sum x'_t y'_t \sum x'_t y'_{t-1}}{\sum x'^2_t} \right] = 0.$$

De entre las dos soluciones de (11.91) elijiremos el valor de λ que minimiza S , y tras obtener la solución de $\bar{\lambda}$ podemos calcular fácilmente los valores estimados de α , β , y las θ . Estas estimaciones son consistentes, pero no necesariamente asintóticamente eficientes.²⁵

Alteremos ahora la especificación del modelo de las expectativas adaptables eliminando el supuesto de independencia entre las ϵ indicado por (11.85) e introduciendo a cambio

$$(11.92) \quad \epsilon_t = \rho \epsilon_{t-1} + u_t,$$

$$\text{con} \quad E(\epsilon_{t-1} u_t) = 0,$$

$$u_t \sim N(0, \sigma_u^2),$$

$$\text{y} \quad \epsilon_0 \sim N\left(0, \frac{\sigma_u^2}{1 - \rho^2}\right).$$

Seríamos en primer lugar que, en el caso especial en que $\rho = \lambda$, el problema de la estimación de los coeficientes de (11.73) se simplifica extraordinariamente. La razón de ello es que, en este caso, la perturbación «transformada» η_t es igual a u_t , o sea que no es autorregresiva y no está correlacionada con Y_{t-1} . En estas circunstancias, la aplicación del método de los mínimos cuadrados ordinarios a (11.73) produce estimaciones que poseen todas las propiedades asintóticas deseables. Naturalmente, el supuesto de que las η son independientes entre sí resulta muy restrictivo y, lamentablemente, no podemos con-

25. El método de la regresión ponderada para la estimación de los parámetros de los modelos con retardos distribuidos fue propuesto y desarrollado por Klein (1958), *op. cit.* La relación entre este método y el de la máxima verosimilitud aparece en el apéndice del artículo. Ver también, T. Amemiya y W. A. Fuller, «A Comparative Study of Alternative Estimators in a Distributed Lag Model», *Econometrica*, vol. 35, julio-octubre 1967, pp. 509-29.

trastarlo mediante la contrastación de Durbin-Watson, ya que esta última no es aplicable a este caso. Teóricamente, la afirmación de que $\rho = \lambda$ es difícil de justificar.

Cuando $\lambda \neq \rho$, el problema de la estimación de los coeficientes de (11.73) bajo el supuesto (11.92) resulta más complicado. Como sucedía también en el caso en que las ϵ eran independientes, las estimaciones minimocuadráticas ordinarias son inconsistentes y tenemos que utilizar métodos de estimación alternativos. De nuevo, el procedimiento más sencillo para obtener estimaciones consistentes es el *método de las variables instrumentales*. Alternativamente, podemos utilizar el principio de la máxima verosimilitud obteniendo *estimadores maximoverosímiles* de los parámetros de (11.73) y (11.92), en cuyo caso escribiremos de nuevo la ecuación (11.73) en la forma

$$E(Y_t) = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda E(Y_{t-1})$$

$$\text{y} \quad \rho E(Y_{t-1}) = \alpha(1 - \lambda)\rho + \beta(1 - \lambda)\rho X_{t-1} + \lambda \rho E(Y_{t-2}).$$

Restando la segunda expresión de la primera, obtenemos

$$\begin{aligned} (11.93) \quad E(Y_t) - \rho E(Y_{t-1}) &= \alpha(1 - \lambda)(1 - \rho) + \beta(1 - \lambda)(X_t - \rho X_{t-1}) \\ &\quad + \lambda[E(Y_{t-1}) - \rho E(Y_{t-2})] \\ &= \alpha(1 - \lambda)(1 - \rho)(1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}) \\ &\quad + \beta(1 - \lambda)[(X_t - \rho X_{t-1}) + \lambda(X_{t-1} - \rho X_{t-2}) \\ &\quad + \dots + \lambda^{t-1}(X_1 - \rho X_0)] \\ &\quad + \lambda^t[E(Y_0) - \rho E(Y_{-1})]. \end{aligned}$$

Utilizando ahora (11.92) e introduciendo una notación abreviada, podemos escribir (11.93) en la forma

$$(11.93a) \quad (Y_t - \rho Y_{t-1}) = \alpha^* + \beta_0 W_t^{(\lambda, \rho)} + \delta_0 \lambda^t + u_t,$$

en la que

$$W_t^{(\lambda, \rho)} = (X_t - \rho X_{t-1}) + \lambda(X_{t-1} - \rho X_{t-2}) + \dots + \lambda^{t-2}(X_2 - \rho X_1) + \lambda^{t-1}X_1,$$

$$\alpha^* = \alpha(1 - \rho),$$

$$\beta_0 = \beta(1 - \lambda),$$

$$\delta_0 = E(Y_0) - \rho E(Y_{-1}) - \alpha^* - \frac{\beta_0 \rho X_0}{\lambda}.$$

En caso de conocer los valores de λ y ρ , la ecuación (11.93a) sería una ecuación lineal de regresión múltiple con dos variables explicativas, $W_t^{(\lambda, \rho)}$ y λ^t . Normalmente, no conoceremos estos valores y, en tal caso, podemos establecer la función de verosimilitud para $Y_2, Y_3, \dots, Y_n$ (condicionada a Y_1) en la forma

$$(11.94) \quad L = -\frac{n-1}{2} \log(2\pi\sigma_u^2) - \frac{1}{2\sigma_u^2} \sum_{t=2}^n [Y_t - \rho Y_{t-1} - \alpha^* - \beta_0 W_t^{(\lambda, \rho)} - \delta_0 \lambda^t]^2.$$

La maximización de L respecto de $\alpha^*, \beta_0, \delta_0, \lambda$, y ρ , equivale a minimizar

$$S^{(\lambda, \rho)} = \sum_{t=2}^n [Y_t - \rho Y_{t-1} - \alpha^* - \beta_0 W_t^{(\lambda, \rho)} - \delta_0 \lambda^t]^2$$

respecto a los mismos parámetros. Como sabemos que $0 \leq \lambda < 1$ y $-1 < \rho < 1$, podemos tomar distintos pares de valores λ y ρ y calcular las estimaciones de α^*, β_0 , y δ_0 correspondientes a cada uno de estos pares. De entre todos estos resultados elijiremos el conjunto de valores de los parámetros que produzcan el valor más pequeño de $S^{(\lambda, \rho)}$. Las estimaciones de α, β, λ , y ρ obtenidas de esta forma serán asintóticamente eficientes y sus variancias asintóticas pueden calcularse mediante la correspondiente matriz de información.

EJEMPLO Zellner y Geisel²⁶ han aplicado este método al modelo de consumo presentado en el ejemplo anterior, utilizando el mismo conjunto de observaciones. Los resultados obtenidos son

$$\hat{\beta} = 0.94, \\ (0.46)$$

$$\hat{\lambda} = 0.66, \\ (0.085)$$

$$\hat{\rho} = 0.69, \\ (0.076)$$

Las cifras entre paréntesis son los errores standard asintóticos estimados. Los resultados muestran que la estimación de λ es muy parecida a la de ρ , lo que indica que en este caso el supuesto de que $\lambda = \rho$ puede resultar bastante adecuado.

Pasemos ahora a la estimación de los parámetros de una relación con retardos distribuidos en forma geométrica, generada por un mecanismo de *ajuste*

26. Op. cit.

parcial (o de persistencia de hábitos). Este tipo de relación se presentó en la ecuación (11.78) en la forma

$$Y_t = \alpha\gamma + \beta\gamma X_t + (1 - \gamma)Y_{t-1} + \xi_t,$$

en la que ξ_t es una variable aleatoria que se distribuye normalmente con media cero y variancia σ^2 . La especificación del modelo de ajuste parcial no introduce ninguna otra restricción acerca de ξ_t , lo que hace que la estimación sea mucho más fácil que en el caso del modelo de las expectativas adaptables. Si podemos suponer que $E(\xi_t, \xi_s) = 0$ para todos los valores de $t \neq s$, podemos utilizar el método de los mínimos cuadrados ordinarios obteniendo estimaciones consistentes y asintóticamente eficientes de los parámetros de (11.78). Por otra parte, si ξ_t sigue un esquema autorregresivo de primer orden, es decir, si

$$(11.95) \quad \xi_t = \rho\xi_{t-1} + u_t,$$

en la que

$$E(\xi_{t-1}u_t) = 0,$$

$$u_t \sim N(0, \sigma_u^2),$$

y

$$\xi_0 \sim N\left(0, \frac{\sigma_u^2}{1 - \rho^2}\right),$$

la aplicación del método de los mínimos cuadrados ordinarios a la ecuación (11.78) produciría estimaciones inconsistentes. En este caso, podemos utilizar una transformación de (11.78) que elimina ξ_t y toma la forma

$$(11.96) \quad (Y_t - \rho Y_{t-1}) = \alpha\gamma(1 - \rho) + \beta\gamma(X_t - \rho X_{t-1}) + (1 - \gamma)(Y_{t-1} - \rho Y_{t-2}) + u_t.$$

Las estimaciones maximoverosímiles de α , β , γ , y ρ (condicionadas a Y_1) se pueden obtener maximizando

$$(11.97) \quad L = -\frac{n-1}{2} \log(2\pi\sigma_u^2) - \frac{1}{2\sigma_u^2} \sum_{t=2}^n [(Y_t - \rho Y_{t-1}) - \alpha\gamma(1 - \rho) - \beta\gamma(X_t - \rho X_{t-1}) - (1 - \gamma)(Y_{t-1} - \rho Y_{t-2})]^2$$

respecto a los parámetros desconocidos. Las estimaciones resultantes poseerán todas las propiedades asintóticas deseables.²⁷

27. Ver Malinvaud, *op. cit.*, p. 469. Para la contrastación de la hipótesis de que ξ_t no es autorregresiva, ver J. Durbin, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, vol. 38, mayo 1970, pp. 410-21.

Retardos de Pascal

En algunos casos, los modelos con retardos distribuidos en los que las ponderaciones decrecen geométricamente a partir del período actual pueden no ser totalmente adecuados. Por ejemplo, en los modelos que relacionan los gastos corrientes de capital con las consignaciones de capital recibidas en el período actual y en los períodos anteriores, es mucho más razonable suponer que las ponderaciones correspondientes a las consignaciones de lo período t , $t-1$, $t-2$, ..., crecerían en primer lugar, disminuyendo después, en lugar de disminuir desde el principio. Esta forma de distribución de los retardos puede denominarse *distribución de los retardos en forma de V invertida* y puede formularse mediante varios procedimientos distintos. Una posibilidad consiste en utilizar la llamada *distribución de retardos de Pascal*, que puede definirse de la forma siguiente: en primer lugar, escribiremos de nuevo (11.68) como

$$(11.68a) \quad Y_t = \alpha + \beta(w_0 X_t + w_1 X_{t-1} + w_2 X_{t-2} + \dots) + \varepsilon_t,$$

en la que ε_t es una perturbación que se distribuye normalmente con media cero y variancia σ^2 . Las ponderaciones correspondientes al modelo de los retardos de Pascal se expresan entonces en la forma

$$w_i = \binom{i+r-1}{i} (1-\lambda)^i \lambda^i = \frac{(i+r-1)!}{i!(r-1)!} (1-\lambda)^i \lambda^i \quad (i = 0, 1, 2, \dots),$$

en la que r es un número entero positivo, y λ es un parámetro que debemos estimar. La ecuación de regresión se convierte en

$$(11.98) \quad Y_t = \alpha + \beta(1-\lambda)[X_t + r\lambda X_{t-1} + \frac{r(r+1)}{2!}\lambda^2 X_{t-2} + \dots] + \varepsilon_t.$$

Obsérvese que cuando $r = 1$ tenemos $w_i = (1-\lambda)\lambda^i$, lo que significa que la distribución de Pascal se reduce a la distribución con retardos geométricos. Cuando los valores de r son mayores que la unidad, podemos obtener las

distribuciones con retardos en forma de V invertida. La figura 11-13 presenta la distribución de las ponderaciones para distintos valores de r cuando $\lambda = 0,6$.

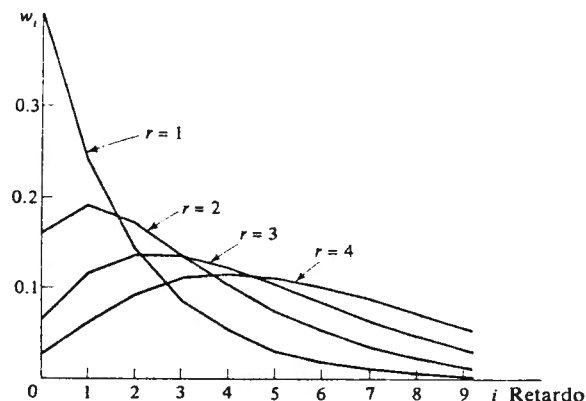


Figura 11-13

En el caso del modelo de los retardos de Pascal, la estimación de los parámetros resulta más complicada que en el modelo con retardos geométricos. Por ejemplo, consideremos el caso en que $r = 2$. La relación con retardos de Pascal es

$$(11.99) \quad Y_t = \alpha + \beta(1 - \lambda)^2(X_t + 2\lambda X_{t-1} + 3\lambda^2 X_{t-2} + \dots) + \varepsilon_t.$$

Podemos simplificar esta ecuación efectuando la transformación siguiente: retardemos (11.99) un período y multipliquemos ambos lados por -2λ , con lo que obtendremos

$$(11.99a) \quad -2\lambda Y_{t-1} = -2\alpha\lambda + \beta(1 - \lambda)^2(-2\lambda X_{t-1} - 4\lambda^2 X_{t-2} - 6\lambda^3 X_{t-3} - \dots) - 2\lambda\varepsilon_{t-1}.$$

A continuación, retardemos (11.99) dos períodos y multipliquemos por λ^2 . El resultado es

$$(11.99b) \quad \lambda^2 Y_{t-2} = \alpha\lambda^2 + \beta(1 - \lambda)^2(\lambda^2 X_{t-2} + 2\lambda^3 X_{t-3} + 3\lambda^4 X_{t-4} + \dots) + \lambda^2 \varepsilon_{t-2}.$$

Sumando ahora (11.99), (11.99a) y (11.99b), obtendremos

$$(11.100) \quad Y_t = \alpha(1 - \lambda)^2 + \beta(1 - \lambda)^2 X_t + 2\lambda Y_{t-1} - \lambda^2 Y_{t-2} + \eta_t,$$

en la que $\eta_t = \varepsilon_t - 2\lambda\varepsilon_{t-1} + \lambda^2\varepsilon_{t-2}$.

La estimación de (11.100) por el método de los mínimos cuadrados ordinarios produce estimaciones inconsistentes ya que η_t es autorregresiva. Para obtener estimaciones consistentes, podemos utilizar el *método de las variables instrumentales*. Como la ecuación (11.100) está superidentificada, la escribiremos de nuevo en la forma

$$(11.100a) \quad Y_t^* = \alpha_0 + \beta_0 X_t + \gamma_0 Y_{t-1}^* + \eta_t,$$

en la que

$$Y_t^* = Y_t - \lambda Y_{t-1},$$

$$Y_{t-1}^* = Y_{t-1} - \lambda Y_{t-2},$$

$$\alpha_0 = \alpha(1 - \lambda)^2,$$

$$\beta_0 = \beta(1 - \lambda)^2,$$

$$\gamma_0 = \lambda.$$

A continuación, utilizaremos como variables instrumentales para X_t e Y_{t-1}^* ,

$$Z_{1t} = X_t \quad \text{y} \quad Z_{2t} = X_{t-1},$$

y calcularemos las estimaciones de α , β y γ correspondientes a los distintos valores de λ entre 0 y 1. De entre todos estos resultados elegiremos el conjunto de valores de los parámetros en el cual γ_0 es igual, o aproximadamente igual, a λ .

La ecuación (11.100) podría estimarse también por el *método de la máxima verosimilitud*. Obsérvese que la podemos escribir en la forma

$$(Y_t - \varepsilon_t) = \alpha(1 - \lambda)^2 + \beta(1 - \lambda)^2 X_t + 2\lambda(Y_{t-1} - \varepsilon_{t-1}) - \lambda^2(Y_{t-2} - \varepsilon_{t-2}),$$

o

$$E(Y_t) = \alpha_0 + \beta_0 X_t + 2\lambda E(Y_{t-1}) - \lambda^2 E(Y_{t-2}).$$

Sin embargo, como

$$E(Y_{t-1}) = \alpha_0 + \beta_0 X_{t-1} + 2\lambda E(Y_{t-2}) - \lambda^2 E(Y_{t-3}),$$

$$E(Y_{t-2}) = \alpha_0 + \beta_0 X_{t-2} + 2\lambda E(Y_{t-3}) - \lambda^2 E(Y_{t-4}),$$

...

$$E(Y_1) = \alpha_0 + \beta_0 X_1 + 2\lambda E(Y_0) - \lambda^2 E(Y_{-1}),$$

podemos escribir también

$$(11.101) \quad E(Y_t) = \alpha_0(1 + 2\lambda + 3\lambda^2 + \dots + t\lambda^{t-1}) \\ + \beta_0(X_t + 2\lambda X_{t-1} + 3\lambda^2 X_{t-2} + \dots + t\lambda^{t-1} X_1) \\ + (t+1)\lambda^t E(Y_0) - t\lambda^{t+1} E(Y_{-1}).$$

Además,

$$\alpha_0(1 + 2\lambda + 3\lambda^2 + \dots + t\lambda^{t-1}) = \frac{\alpha_0}{1-\lambda} (1 + \lambda + \lambda^2 + \dots + \lambda^{t-1} - t\lambda^t) \\ = \frac{\alpha_0(1-\lambda^t)}{(1-\lambda)^2} - \frac{\alpha_0 t \lambda^t}{(1-\lambda)} \\ = \alpha(1-\lambda^t) - \alpha(1-\lambda)t\lambda^t.$$

Por tanto, podemos escribir la ecuación (11.101) en la forma

$$(11.101a) \quad Y_t = \alpha + \beta_0 W_{2t}^{(\lambda)} + \delta_0 \lambda^t + \phi_0 t \lambda^t + \varepsilon_t$$

en la que $W_{2t}^{(\lambda)} = X_t + 2\lambda X_{t-1} + 3\lambda^2 X_{t-2} + \dots + t\lambda^{t-1} X_1$,

$$\beta_0 = \beta(1-\lambda)^2,$$

$$\delta_0 = E(Y_0) - \alpha,$$

$$\phi_0 = E(Y_0) - \lambda E(Y_{-1}) - \alpha(1-\lambda).$$

En caso de conocer λ , la ecuación (11.101a) sería una ecuación de regresión lineal con tres regresores: $W_{2t}^{(\lambda)}$, λ^t y $t\lambda^t$. Naturalmente, la mayor parte de las veces no se conoce el valor de λ . Si suponemos que las ε son independientes entre sí, podemos obtener las estimaciones maximoverosímiles de los parámetros desconocidos minimizando

$$S^{(\lambda)} = \sum_{t=1}^n (Y_t - \alpha - \beta_0 W_{2t}^{(\lambda)} - \delta_0 \lambda^t - \phi_0 t \lambda^t)^2$$

respecto de α , β_0 , δ_0 y ϕ_0 para distintos valores de λ comprendidos entre 0 y 1. Las estimaciones maximoverosímiles de los parámetros serán los valores que produzcan el valor más pequeño de $S^{(\lambda)}$, y según los supuestos efectuados, las estimaciones de α , β y λ poseerán todas las propiedades asintóticas deseables. En el caso en que las ε siguen un esquema autorregresivo de primer orden, podemos transformar (11.100) en

$$E(Y_t) - \rho E(Y_{t-1}) = \alpha_0(1-\rho) + \beta_0(X_t - \rho X_{t-1}) \\ + 2\lambda[E(Y_{t-1}) - \rho E(Y_{t-2})] - \lambda^2[E(Y_{t-2}) - \rho E(Y_{t-3})] + \dots$$

y proceder a continuación como hemos hecho en (11.93) y (11.93a)

La forma de resolver el problema de la estimación de los parámetros en un modelo de retardos de Pascal en el que $r = 3, 4$, etc., sería similar al caso en que $r = 2$. Naturalmente, al aumentar este valor aumenta también el grado de complejidad de la estimación. Si el valor de r no se puede suponer a priori, tenemos que calcular el valor de la función de verosimilitud para cada valor de r y elegir después de entre estos valores el que produzca un máximo en la función de verosimilitud. Aunque en principio no existe ninguna dificultad en este procedimiento, su realización práctica es muy complicada y puede no ser posible.

Podemos ampliar la transformación de la relación de los retardos de Pascal presentada en la ecuación (11.100) para el caso en que $r = 2$, a todos los valores de r enteros y no negativos en la forma siguiente:

$$(11.102) \quad Y_t + \binom{r}{1}(-\lambda)^1 Y_{t-1} + \binom{r}{2}(-\lambda)^2 Y_{t-2} + \dots + \binom{r}{r}(-\lambda)^r Y_{t-r} \\ = \alpha(1-\lambda)^r + \beta(1-\lambda)^r X_t + \varepsilon_t + \binom{r}{1}(-\lambda)^1 \varepsilon_{t-1} + \dots + \binom{r}{r}(-\lambda)^r \varepsilon_{t-r}.$$

Obsérvese que en la ecuación (11.102) todos los coeficientes de los valores actuales y retardados de Y (y ε) presentan la restricción de ser funciones concretas de λ . Si generalizamos la forma de (11.102) eliminando esta restricción, obtenemos

$$(11.103) \quad Y_t + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \dots + \delta_r Y_{t-r} \\ = \alpha_0 + \beta_0 X_t + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \dots + \delta_r \varepsilon_{t-r}.$$

La ecuación (11.103) presenta el llamado *modelo con retardos distribuidos racionales* introducido por Jorgenson.²⁸ Este modelo podría estimarse de forma similar a la utilizada en el caso del modelo con retardos de Pascal.

La complejidad del problema de la estimación en el caso general de los modelos con retardos en forma de V invertida ha llegado a la búsqueda de formulaciones más simples. Podemos obtener una de estas simplificaciones si suponemos que las ponderaciones asignadas a los valores de $X_t, X_{t-1}, \dots$, alcanzan un máximo en uno de los $(h-1)$ períodos anteriores declinando des-

28. D. W. Jorgenson, «Rational Distributed Lag Functions», *Econometrica*, vol. 34, enero 1966, pp. 135-49.

(11.72) con un retardo de un período, multiplicamos ambos lados de la ecuación por λ y restamos el resultado de la ecuación original (11.72), obtenemos

$$(11.73) \quad Y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda Y_{t-1} + \eta_t$$

en la que

$$\eta_t = \varepsilon_t - \lambda \varepsilon_{t-1}.$$

En algunas ocasiones, el modelo de las expectativas adaptables se formula de forma algo distinta. En este caso, se mantiene la forma de la relación indicada en (11.70), pero el proceso de formación de las expectativas se postula en la forma

$$(11.74) \quad X_t^* - X_{t-1}^* = (1 - \lambda)(X_{t-1} - X_{t-1}^*),$$

en la que, nuevamente, $0 \leq \lambda < 1$.

La ecuación (11.74) se distingue de la (11.71) por el hecho de que X_t se sustituye por X_{t-1} . Este tipo de formulación sería el adecuado en el caso de que X_t no se conozca por adelantado, de forma que las expectativas tienen que revisarse comparando X_{t-1}^* con X_{t-1} que constituye la información más reciente que disponemos acerca de X . La ecuación (11.74) puede escribirse también en la forma

$$(11.74a) \quad X_t^* = (1 - \lambda)X_{t-1} + \lambda X_{t-1}^*,$$

lo que, a su vez, implica que

$$(11.74b) \quad X_t^* = (1 - \lambda)(X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \dots).$$

Sustituyendo X_t^* por su valor según la ecuación (11.74) en (11.70), y aplicando la transformación de Koyck a este resultado, obtenemos

$$(11.75) \quad Y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)X_{t-1} + \lambda Y_{t-1} + \eta_t,$$

en la que

$$\eta_t = \varepsilon_t - \lambda \varepsilon_{t-1}.$$

La única diferencia entre (11.73) y (11.75) es que en esta última ecuación se ha sustituido X_t por X_{t-1} , y este cambio no afecta para nada al problema de la estimación de los parámetros de la regresión.

Una racionalización alternativa de los retardos geométricos es la correspondiente a los modelos de *ajuste parcial* o de *persistencia de hábitos*. Supongamos que el nivel deseado de Y en el momento t , al que llamaremos Y_t^* , viene

ido por una función lineal de una variable explicativa X_t y un término de perturbación ε_{t1} ; es decir:

$$(11.76) \quad Y_t^* = \alpha + \beta X_t + \varepsilon_{t1}.$$

Por ejemplo, el nivel medio de existencias deseado por una empresa puede ser una función lineal de las ventas (recuérdese el «modelo de ajuste de inventarios» de la sección 11.2), o el nivel medio deseado de consumo puede ser función lineal de la riqueza. Los valores de Y^* no son directamente observables, pero suponemos que se intenta igualar el nivel real de Y a este nivel deseado y que este intento sólo consigue un éxito parcial en cada período. Los motivos por los cuales no se alcanza un ajuste completo de Y a Y^* en un solo período pueden ser muy distintos y pueden incluir limitaciones de carácter tecnológico, rigideces institucionales, persistencia de hábitos, etc...²² La relación entre el nivel real y el nivel deseado de Y puede especificarse en la forma siguiente:

$$(11.77) \quad Y_t - Y_{t-1} = \gamma(Y_t^* - Y_{t-1}) + \varepsilon_{t2}$$

en la que

$$0 < \gamma \leq 1,$$

y ε_{t2} es una perturbación aleatoria. El coeficiente γ se denomina «coeficiente de ajuste», ya que indica el porcentaje del ajuste de Y a Y^* . Despejando Y_t^* en la ecuación (11.77), obtenemos

$$(11.77a) \quad Y_t^* = \frac{1}{\gamma} Y_t + \frac{\gamma - 1}{\gamma} Y_{t-1} - \frac{1}{\gamma} \varepsilon_{t2}.$$

Sustituyendo Y_t^* por su valor según la ecuación (11.77a) en (11.76), obtenemos

$$(11.78) \quad Y_t = \alpha\gamma + \beta\gamma X_t + (1 - \gamma)Y_{t-1} + \xi_t,$$

en la que

$$\xi_t = \gamma\varepsilon_{t1} + \varepsilon_{t2}.$$

La ecuación (11.78) es formalmente igual a la ecuación (11.73) que representa el modelo de las expectativas adaptables presentado en (11.75), salvo por lo perturbación de (11.73) está generado por acontecimientos correspondientes al

22. Puede verse una explicación formulada con mayor precisión en un contexto concreto en Z. Griliches, «Distributed Lags: A Survey», *Econometrica*, vol. 35, enero 1967, p. 43.

período presente y al anterior. Obsérvese también que (11.78) describe un retardo geométrico de la forma

$$(11.78a) \quad Y_t = \alpha + \beta\gamma[X_t + (1 - \gamma)X_{t-1} + (1 - \gamma)^2X_{t-2} + \dots] + \zeta_t,$$

en la que $\zeta_t = \xi_t + (1 - \gamma)\xi_{t-1} + (1 - \gamma)^2\xi_{t-2} + \dots$

Como ocurría en el caso del modelo de las expectativas adaptables, el modelo de ajuste parcial se formula a veces de forma que el retardo geométrico empieza a partir del término X_{t-1} en lugar de X_t . Esta formulación se obtiene especificando el proceso de ajuste en la misma forma que (11.77), pero expresando la obtención del nivel deseado de Y en la forma

$$(11.79) \quad Y_t^* = \alpha + \beta X_{t-1} + \varepsilon_{t1}.$$

En esta formulación, el nivel deseado de Y en el período actual depende del nivel de X en el período precedente. Sustituyendo Y_t^* por su valor según la ecuación (11.77a) en (11.79), y reordenando los términos, obtenemos

$$(11.80) \quad Y_t = \alpha\gamma + \beta\gamma X_{t-1} + (1 - \gamma)Y_{t-1} + \xi_t$$

en la que $\xi_t = \gamma\varepsilon_{t1} + \varepsilon_{t2}$.

Nuevamente, la ecuación (11.80) es formalmente igual a la versión alternativa del modelo de las expectativas adaptables presentado en (11.75), salvo por lo que se refiere a la especificación del término de perturbación. Desde el punto de vista de la estimación de los parámetros de la regresión, la ecuación (11.80) presenta las mismas características que la (11.78).

Como nota de interés, señalemos que el modelo de las expectativas adaptables y el modelo del ajuste parcial pueden combinarse en un modelo con *retardos geométricos compuestos*. Si comparamos la ecuación (11.70) del modelo de las expectativas adaptables con la (11.76) del modelo de ajuste parcial, podemos ver que cada una de ellas representa una modificación inmediata del modelo de regresión simple

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

En el modelo de expectativas adaptables, sustituimos X_t por su valor «esperado» X_t^* , y en el modelo del ajuste parcial, sustituimos Y_t por su valor «deseado» Y_t^* . Combinando las dos especificaciones, tenemos

$$(11.81) \quad Y_t^* = \alpha + \beta X_t^* + \varepsilon_{t1}.$$

Esto significa que el valor medio deseado de Y es función lineal del valor esperado de X . Para completar el modelo tenemos que especificar la forma en que suponemos que se determinan los valores no observables X_t^* e Y_t^* . Suponemos que X_t^* se determina por medio de la ecuación (11.71b) del modelo de expectativas adaptables, mientras que Y_t^* se determina por (11.77a) del modelo del ajuste parcial. Efectuando las sustituciones adecuadas en la ecuación (11.81), obtenemos

$$(11.82) \quad Y_t = \alpha\gamma + \beta\gamma(1 - \lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + (1 - \gamma)Y_{t-1} + \xi_t,$$

en la que $\xi_t = \gamma\varepsilon_{t1} + \varepsilon_{t2}$.

Podemos simplificar la ecuación (11.82) utilizando la transformación de Koyck, de modo que

$$(11.83) \quad Y_t = \alpha\gamma(1 - \lambda) + \beta\gamma(1 - \lambda)X_t + [(1 - \gamma) + \lambda]Y_{t-1} - (1 - \gamma)\lambda Y_{t-2} + \omega_t,$$

en la que $\omega_t = \xi_t - \lambda\xi_{t-1}$.

Obsérvese que, por analogía con (11.78a), la ecuación (11.83) puede escribirse en la forma

$$(11.83a) \quad Y_t = \alpha + \beta\gamma[X_t^* + (1 - \gamma)X_{t-1}^* + (1 - \gamma)^2X_{t-2}^* + \dots] + \zeta_t,$$

en la que $\zeta_t = \xi_t + (1 - \gamma)\xi_{t-1} + (1 - \gamma)^2\xi_{t-2} + \dots$

De aquí que el modelo se denomine modelo con retardos geométricos compuestos: los coeficientes de regresión de (11.83a) siguen una progresión geométrica, mientras que cada uno de los valores X_t^* , X_{t-1}^* , X_{t-2}^* , ..., es una media ponderada de los valores previos de X , en la que las ponderaciones siguen también una progresión geométrica. Cuando $\gamma = 1$, el modelo se reduce al caso puro de expectativas adaptables, mientras que, si $\lambda = 0$, nos encontramos con el modelo puro de ajuste parcial. Si $\gamma = 1$ y además $\lambda = 0$, nos hallamos de nuevo en el caso del modelo de regresión simple.

Antes de pasar al problema de la estimación de los coeficientes de regresión de los modelos con retardos geométricos, deberíamos considerar la formulación de los modelos en los que los retardos distribuidos aparecen en más de una variable explicativa. Concretamente, consideremos un modelo con dos variables explicativas, X y Z , cada una de las cuales ejerce su efecto sobre $E(Y_t)$ a través de una estructura de retardos distribuidos en forma geométrica:

$$(11.84) \quad Y_t = \alpha + \beta_0(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + \delta_0(Z_t + \mu Z_{t-1} + \mu^2 Z_{t-2} + \dots) + \varepsilon_t,$$

en la que $0 \leq \lambda \leq 1$ y $0 \leq \mu < 1$.

Por ejemplo, la demanda de dinero puede depender de la renta «permanente», y de un tipo de interés «esperado». La ecuación (11.84) puede reducirse a una forma más manejable aplicando la transformación de Koyck dos veces consecutivas. En primer lugar, retardamos un período (11.84), multiplicando después ambos lados por λ y deduciendo la ecuación resultante de la (11.84) en su forma original. Con esto obtenemos

$$(11.84a) \quad Y_t = \alpha(1 - \lambda) + \lambda Y_{t-1} + \delta_0[Z_t + (\mu - \lambda)Z_{t-1} + \mu(\mu - \lambda)Z_{t-2} + \mu^2(\mu - \lambda)Z_{t-3} + \dots] + \beta_0 X_t + \varepsilon_t - \lambda \varepsilon_{t-1}.$$

A continuación, retardamos (11.84a) un período, multiplicamos ambos lados por μ y restamos la ecuación resultante de la (11.84a). El resultado es

$$(11.84b) \quad Y_t = \alpha(1 - \lambda)(1 - \mu) + (\lambda + \mu)Y_{t-1} - \lambda\mu Y_{t-2} + \beta_0 X_t - \beta_0\mu X_{t-1} + \delta_0 Z_t - \delta_0\lambda Z_{t-1} + \eta_t^*,$$

en la que $\eta_t^* = \varepsilon_t - (\lambda + \mu)\varepsilon_{t-1} + \lambda\mu\varepsilon_{t-2}$,

que es una ecuación de regresión con seis regresores (además del término constante). Como puede verse, podríamos manejar modelos con cualquier número de retardos distribuidos, mediante un procedimiento similar.

Consideremos ahora la estimación de los parámetros de un modelo con retardos distribuidos geométricamente,

$$Y_t = \alpha + \beta_0(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots) + \varepsilon_t,$$

en el que ε_t es una variable aleatoria normal con media cero y variancia σ^2 . Supongamos, además, que la relación ha sido generada por un mecanismo de *expectativas adaptables* como el descrito por las ecuaciones (11.70) y (11.71). La estimación de los parámetros de este modelo depende de si suponemos que las ε son mutuamente independientes o de si no efectuamos dicho supuesto. Consideraremos en primer lugar el caso en que las perturbaciones son independientes entre sí, es decir, suponemos que dichas perturbaciones se distribuyen normalmente y que

$$(11.85) \quad E(\varepsilon_t \varepsilon_s) = 0 \quad (t \neq s).$$

Evidentemente, el modelo con retardos distribuidos no resulta adecuado para la estimación en su forma original, ya que incorpora un número infinito de re-

gresores. Sin embargo, aplicando la transformación de Koyck, podemos escribir

$$Y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)X_t + \lambda Y_{t-1} + \eta_t.$$

$$Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + \eta_t,$$

en la que $\alpha_0 = \alpha(1 - \lambda)$,

$$\beta_0 = \beta(1 - \lambda),$$

y $\eta_t = \varepsilon_t - \lambda \varepsilon_{t-1}$.

Esta ecuación es la presentada anteriormente en (11.73). Su forma es relativamente simple, pero esta simplificación no se ha conseguido sin ningún coste. El inconveniente de la ecuación (11.73) es que el «nuevo» término de perturbación η_t está correlacionado con Y_{t-1} , que es ahora una de las variables explicativas. Concretamente,

$$\begin{aligned} E(\eta_t Y_{t-1}) &= E(\varepsilon_t - \lambda \varepsilon_{t-1})[\alpha + \beta(X_{t-1} + \lambda X_{t-2} + \dots) + \varepsilon_{t-1}] \\ &= -\lambda \sigma^2. \end{aligned}$$

Esto significa que las estimaciones minimocuadráticas ordinarias de los coeficientes de (11.73) son *inconsistentes* (véase la sección 8.3) y, por tanto, tenemos que acudir a otros métodos de estimación.

Cuando efectuamos el supuesto especificado en (11.85), podemos obtener estimaciones consistentes de los coeficientes de (11.73) por varios procedimientos distintos. El más sencillo es quizás el *método de las variables instrumentales*, expuesto ya previamente en relación con los modelos de «errores en las variables» de la sección 9.1. Como la ecuación (11.73) incluye dos variables explicativas, tenemos que hallar dos variables instrumentales, Z_1 y Z_2 . Estas variables deben cumplir las condiciones siguientes:

1. $\lim \text{prob } \Sigma_t(Z_{1t} - \bar{Z}_1) \eta_t / n = 0$ y $\lim \text{prob } \Sigma_t(Z_{2t} - \bar{Z}_2) \eta_t / n = 0$.
2. $\lim \text{prob } \Sigma_t(Z_{1t} - \bar{Z}_1) X_t / n$ y $\lim \text{prob } \Sigma_t(Z_{2t} - \bar{Z}_2) Y_{t-1} / n$ son valores finitos distintos de cero.

Una condición adicional, no necesaria para la consistencia pero que ayuda a reducir la variancia asintótica del estimador por variables instrumentales, es que estas últimas deben estar muy correlacionadas con los regresores respectivos. Se ha sugerido que deben usarse las variables instrumentales siguientes:

$$Z_{1t} = X_t, \text{ y } Z_{2t} = X_{t-1}.$$

Es decir, la variable instrumental para el primer regresor, X_t , será esta misma variable, y la variable instrumental para el segundo regresor, Y_{t-1} , será X_{t-1} .

pués en forma geométrica. En este caso, la relación de los retardos distribuidos puede formularse como

$$(11.104) \quad Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_{h-1} X_{t-h+1} + \beta_h (X_{t-h} + \lambda X_{t-h-1} + \lambda^2 X_{t-h-2} + \cdots) + \varepsilon_t.$$

Aplicando la transformación de Koyck, esto se convierte en

$$(11.104a) \quad Y_t = \alpha_0 + \beta_0 X_t + \beta_1^* X_{t-1} + \beta_2^* X_{t-2} + \cdots + \beta_h^* X_{t-h} + \lambda Y_{t-1} + \eta_t,$$

en la que

$$\begin{aligned} \alpha_0 &= \alpha(1 - \lambda), \\ \beta_1^* &= \beta_1 - \beta_0 \lambda, \\ \beta_2^* &= \beta_2 - \beta_1 \lambda, \\ &\vdots \\ \beta_h^* &= \beta_h - \beta_{h-1} \lambda, \\ \eta_t &= \varepsilon_t - \lambda \varepsilon_{t-1}. \end{aligned}$$

La estimación de la expresión (11.104a) puede efectuarse por el mismo procedimiento que la del modelo convencional con retardos geométricos que hemos estudiado previamente.

Retardos en forma de polinomio

Podemos efectuar una formulación distinta del modelo con retardos en forma de V invertida en el caso en que podemos suponer que las ponderaciones w de

$$Y_t = \alpha + \beta(w_0 X_t + w_1 X_{t-1} + \cdots + w_m X_{t-m}) + \varepsilon_t$$

siguen un polinomio de grado determinado. En tal caso, hablamos de *modelos con retardo en forma de polinomio*, y para formularlos tenemos que especificar el grado del polinomio y el número de períodos que debemos considerar antes de suponer que las ponderaciones tienen un valor cero. En el ejemplo de los gastos y consignaciones de capital, suponemos que los gastos corrientes de capital no se ven influidos por las consignaciones correspondientes a períodos anteriores a los últimos m períodos, ni por las que se efectuarán en el futuro. Por tanto, queremos ajustar un polinomio de grado p , por ejemplo, a las ponderaciones $w_{-1}, w_0, w_1, \dots, w_m, w_{m+1}$ de forma que

$$w_{-1} = 0 \quad \text{y} \quad w_{m+1} = 0$$

Las ponderaciones $w_{-2}, w_{-3}, \dots$, y las $w_{m+2}, w_{m+3}, \dots$, no se hallan sobre el polinomio, sino que se supone que son cero. El problema de la estimación resulta sencillo. Supongamos que el polinomio elegido es de cuarto grado. En este caso, para que las ponderaciones $w_{-1}, w_0, w_1, \dots, w_{m+1}$ se hallen sobre la curva de un polinomio de cuarto grado, especificamos

$$(11.105) \quad w_i = \lambda_0 + \lambda_1 i + \lambda_2 i^2 + \lambda_3 i^3 + \lambda_4 i^4, \quad (i = -1, 0, 1, 2, \dots, m, m+1).$$

El modelo con retardos en forma de polinomio se convierte, pues, en

$$(11.106) \quad Y_t = \alpha + \beta[\lambda_0 X_t + (\lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) X_{t-1} + (\lambda_0 + 2\lambda_1 + 2^2\lambda_2 + 2^3\lambda_3 + 2^4\lambda_4) X_{t-2} + \cdots + (\lambda_0 + m\lambda_1 + m^2\lambda_2 + m^3\lambda_3 + m^4\lambda_4) X_{t-m}] + \varepsilon_t,$$

que se puede concentrar en la forma

$$(11.106a) \quad Y_t = \alpha + \beta\lambda_0 Z_{t0} + \beta\lambda_1 Z_{t1} + \cdots + \beta\lambda_4 Z_{t4} + \varepsilon_t$$

en la que

$$\begin{aligned} Z_{t0} &= X_t + X_{t-1} + \cdots + X_{t-m}, \\ Z_{t1} &= X_{t-1} + 2X_{t-2} + \cdots + mX_{t-m}, \\ &\vdots \\ Z_{t4} &= X_{t-1} + 2^4 X_{t-2} + \cdots + m^4 X_{t-m}. \end{aligned}$$

La ecuación (11.106a) tiene en total siete parámetros, pero este número puede reducirse imponiendo las restricciones de que $w_{-1} = 0$ y $w_{m+1} = 0$, es decir, que

$$(11.107) \quad \lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 0,$$

$$(11.108) \quad \lambda_0 + (m+1)\lambda_1 + (m+1)^2\lambda_2 + (m+1)^3\lambda_3 + (m+1)^4\lambda_4 = 0.$$

En estas ecuaciones podemos despejar, por ejemplo, λ_0 y λ_1 , obteniendo

$$\lambda_0 = -(m+1)\lambda_2 - m(m+1)\lambda_3 - (m+1)(m^2 + m + 1)\lambda_4$$

y

$$\lambda_1 = -m\lambda_2 - (m^2 + m + 1)\lambda_3 - m(m^2 + 2m + 2)\lambda_4.$$

Sustituyendo estos resultados en la ecuación (11.104a) tenemos

$$(11.109) \quad Y_t = \alpha + \beta\lambda_2 W_{t2} + \beta\lambda_3 W_{t3} + \beta\lambda_4 W_{t4} + \varepsilon_t,$$

en la que

$$W_{t2} = -(m+1)Z_{t0} - mZ_{t1} + Z_{t2},$$

$$W_{t3} = -m(m+1)Z_{t0} - (m^2 + m + 1)Z_{t1} + Z_{t3},$$

$$W_{t4} = -(m+1)(m^2 + m + 1)Z_{t0} - m(m^2 + 2m + 2)Z_{t1} + Z_{t4}.$$

Si, por ejemplo, decidimos que m tome el valor 7, las W se definen como

$$\begin{aligned} W_{12} &= -8Z_{10} - 7Z_{11} + Z_{12} \\ &= -8X_t - 14X_{t-1} - 18X_{t-2} - 20X_{t-3} - 20X_{t-4} - 18X_{t-5} \\ &\quad - 14X_{t-6} - 8X_{t-7}, \\ W_{13} &= -56Z_{10} - 57Z_{11} + Z_{13} \\ &= -56X_t - 112X_{t-1} - 162X_{t-2} - 200X_{t-3} - 220X_{t-4} \\ &\quad - 216X_{t-5} - 182X_{t-6} - 112X_{t-7}, \\ W_{14} &= -456Z_{10} - 455Z_{11} + Z_{14} \\ &= -456X_t - 910X_{t-1} - 1350X_{t-2} - 1740X_{t-3} - 2020X_{t-4} \\ &\quad - 2106X_{t-5} - 1890X_{t-6} - 1240X_{t-7}. \end{aligned}$$

Obsérvese que el parámetro β sigue sin identificar y, en la práctica, generalmente se considera que su valor es la unidad. Los parámetros α , λ_2 , λ_3 y λ_4 de la ecuación (11.109) pueden estimarse por el método de los mínimos cuadrados ordinarios. Si ϵ_t cumple todos los supuestos del modelo clásico de regresión normal, las estimaciones resultantes poseerán toda las propiedades deseables. Si existe autocorrelación, podemos utilizar uno de los métodos de estimación presentados en la sección 8.2. Para obtener estimaciones de las ponderaciones $w_0, w_1, \dots, w_m$, podemos sustituir las λ por sus valores estimados en la ecuación (11.105)²⁹ y, si queremos que la suma de las ponderaciones sea la unidad, estableceremos una restricción adicional sobre las λ , haciendo que

$$\lambda_0(m+1) + \lambda_1 \sum i + \lambda_2 \sum i^2 + \lambda_3 \sum i^3 + \lambda_4 \sum i^4 = 1.$$

En este caso, la ecuación (11.109) quedaría aún más reducida, incluyendo sólo dos regresores en lugar de tres.

EJEMPLO. Almon³⁰ utilizó el modelo con retardos en forma de polinomio para estimar la relación entre los gastos corrientes de capital y las consignaciones de capital recibidas en el mismo período y en los anteriores, en las industrias de los

Estados Unidos. En todos los casos se utilizó un polinomio de cuarto grado, pero la longitud del retardo se consideró distinta para las distintas industrias. Los datos empleados para la estimación consistieron en observaciones trimestrales para los años 1953-1961. El modelo se especificó en la forma

$$Y_t = \alpha_1 S_{11} + \alpha_2 S_{12} + \alpha_3 S_{13} + \alpha_4 S_{14} + w_0 X_t + w_1 X_{t-1} + \dots + w_m X_{t-m} + \epsilon_t,$$

en la que Y representa los gastos de capital, las S representan variables ficticias estacionales, y las X representan las consignaciones de capital. El parámetro α_4 se especificó de forma que fuese igual a $(-\alpha_1 - \alpha_2 - \alpha_3)$. Las ponderaciones incluían la restricción de que $w_{-1} = 0$ y $w_{m+1} = 0$, pero no la de que sumasen la unidad. El resultado correspondiente a «todas las industrias manufactureras» fue el siguiente:

$$\begin{aligned} Y_t &= -283S_{11} + 13S_{12} - 50S_{13} + 320S_{14} + 0.048X_t + 0.099X_{t-1} + 0.141X_{t-2} \\ &\quad (0.023) \quad (0.016) \quad (0.013) \\ &\quad + 0.165X_{t-3} + 0.167X_{t-4} + 0.146X_{t-5} + 0.105X_{t-6} + 0.053X_{t-7} + \epsilon_t. \\ &\quad (0.023) \quad (0.023) \quad (0.013) \quad (0.016) \quad (0.024) \end{aligned}$$

Como puede verse, la longitud del retardo elegido en este caso fue de 7 periodos. Las ponderaciones suman 0,922, y la diferencia entre este valor y la unidad puede quedar casi totalmente explicada por las anulaciones. En la figura 11-14 se han representado gráficamente las ponderaciones estimadas.

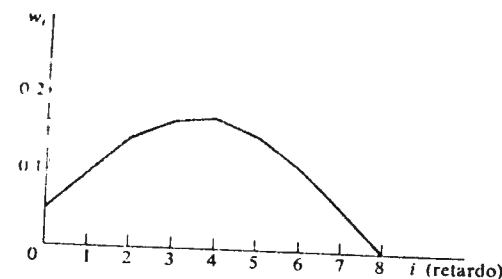


Figura 11-14

Uno de los problemas existentes en el uso del modelo con retardos en forma de polinomio es que tenemos que especificar a priori el grado del polinomio y la longitud del retardo, pero, sin embargo, los datos nos pueden prestar alguna ayuda por lo que respecta a esta última. Una posibilidad consiste en extender la longitud del retardo hasta que la contribución de las X adicionales a la suma de los cuadrados de la regresión deja de ser estadísticamente significativa. Pero, si las X están muy correlacionadas, este criterio puede no funcionar correctamente. Otra posibilidad consiste en elegir la longitud del re-

29. En S. Almon, «The Distributed Lag Between Capital Appropriations and Expenditures», *Econometrica*, vol. 33, enero 1965, pp. 178-96, se estiman los parámetros del modelo con la ayuda de la llamada «interpolación de Lagrange». Sin embargo, los resultados deberían ser los mismos que los obtenidos mediante la aplicación simple del método de los mínimos cuadrados ordinarios en la forma que se ha descrito aquí.

30. *Ibid.*

tardo que proporciona el mayor valor de R^2 , es decir, del coeficiente de determinación corregido según el «número de grados de libertad». Sin embargo, este método puede fallar también en algunos casos, puesto que las diferencias entre distintos valores de R^2 pueden ser muy pequeñas. No obstante, alguno de estos dos criterios, y otras consideraciones adicionales (por ejemplo, que todas las ponderaciones deben ser positivas) nos pueden ayudar a elegir el retardo «óptimo» para el problema que nos ocupa.

EJERCICIOS:

- 11-1. Tres vendedores empleados por una cierta empresa obtuvieron durante cinco días consecutivos los siguientes ingresos semanales:

Vendedor:	A	B	C
Ingresos semanales (dólares):	928	924	793

Se ha efectuado una contrastación del análisis de la variancia, para la hipótesis nula de que los ingresos medios de los tres vendedores son iguales, obteniendo un valor 5,00 para el estadístico F . Considérese este problema desde el punto de vista del análisis de regresión, con una ecuación de regresión expresada por

$$Y_{ig} = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

en la que $X_{i2} = 1$ si i corresponde a B,
 $= 0$ en caso contrario;
 $X_{i3} = 1$ si i corresponde a C,
 $= 0$ en caso contrario;
 $i = 1, 2, \dots, 5$;
 $g = 1, 2, 3$.

Obtener los estimadores minimocuadráticos de los coeficientes de la regresión y sus errores standard estimados. Calcular también el valor de R^2 .

- 11-2. Demostrar la equivalencia entre el modelo de regresión (11.8) y un modelo del «análisis de la variancia para dos variables» con interacciones.

- 11-3. Las estimaciones minimocuadráticas de un modelo de regresión son

$$Y_i = 10 + 5X_i + \epsilon_i.$$

(1) ($\sqrt{0.5}$)

La variable explicativa X_i es una variable binaria. Supóngase que, al comprobar los resultados, se descubre que, por error, se han efectuado los cálculos asignando a X_i los valores 0 y 2 en lugar de 0 y 1. Efectuar las correcciones adecuadas.

- 11-4. Los resultados de una encuesta realizada en una muestra de doce familias indican las rentas y gastos de consumo semanales presentados en el cuadro

11-5. Las familias cuya renta aparece señalada por medio de un asterisco (*) han indicado que su renta es mayor que en el año anterior. Contrastar la validez del modelo de respuesta asimétrica en este caso.

Cuadro 11-5

Familia núm.	Renta	Gastos de consumo
1	80	70
2	95*	76
3	105	91
4	115	100
5	125*	105
6	135*	113
7	145	122
8	155	120
9	165*	146
10	175*	135
11	185*	147
12	200*	155

- 11-5. Considérese el modelo siguiente de la función de consumo:

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 Z_t + \beta_4 Y_t Z_t + \epsilon_t,$$

en el que C = consumo, Y = renta, y $Z_t = 1$ en tiempo de guerra, y a 0 en los demás casos. Demostrar que las estimaciones minimocuadráticas de los coeficientes de regresión podrían haberse obtenido también a partir de dos regresiones distintas de C_t sobre Y_t , una de ellas estimada a partir de las observaciones correspondientes al período de paz y la otra con las observaciones del período de guerra.

- 11-6. Se ha especificado un modelo de la función de producción en la forma

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i,$$

en el que Y_i = log de la cantidad producida, X_{i2} = log de la cantidad de trabajo, y X_{i3} = log de la cantidad de capital. El subíndice i indica la empresa. En la muestra aparecen 23 observaciones y las matrices de los momentos (en términos de desviaciones respecto de las medias muestrales) son

$$(\bar{X}'\bar{X}) = \begin{bmatrix} 12 & 8 \\ 8 & 12 \end{bmatrix}, \quad (\bar{X}'\bar{Y}) = \begin{bmatrix} 10 \\ 8 \end{bmatrix}, \quad (\bar{Y}'\bar{Y}) = 10.$$

- a. Obtener las estimaciones minimocuadráticas de los coeficientes de regresión β_2 y β_3 , y sus errores standard estimados. Calcular el valor de R^2 .

- b. Efectuar la contrastación de la existencia de rendimientos constantes de escala (es decir, contrastar la hipótesis de que $\beta_2 + \beta_3 = 1$).
- c. Supóngase que queremos imponer a priori la restricción de que $\beta_1 + \beta_2 = 1$, y estimar la función de producción de esta forma. ¿Cuál es ahora la estimación minimocuadrática de β_2 y su error standard estimado? ¿Cuál es el valor de R^2 con esta especificación?

11-7. Considerar el modelo de «ajuste de inventarios»

$$Y_t = \alpha\gamma + \beta\gamma X_t + (1 - \gamma)Y_{t-1} + \varepsilon_t.$$

Indicar la fórmula del estimador minimocuadrático no lineal de α , $\hat{\alpha}$. Obtener la fórmula de la variancia estimada de $\hat{\alpha}$ utilizando

- a. La fórmula aproximada (11.40).
- b. La matriz de información maximoverosímil.

11-8. Los resultados de la regresión minimocuadrática ordinaria de Y sobre X_1 y X_2 son

$$Y_t = 10 + 12X_{1t} + 4X_{2t} + \varepsilon_t. \quad (1)$$

Además, el coeficiente de correlación muestral entre X_1 y X_2 es 0,5. Supongamos ahora que se sugiere que hubiese sido preferible especificar el modelo de forma que se efectuase una regresión de Y sobre una media ponderada de X_1 y X_2 , o sea, que la ecuación de regresión sería

$$Y_t = \beta_1 + \beta_2(w_1X_{1t} + w_2X_{2t}) + \varepsilon_t,$$

en la que $w_2 = 1 - w_1$. Obtener las estimaciones minimocuadráticas no lineales de β_2 y w_1 y sus errores standard estimados.

11-9. Utilizando los datos de consumo, renta y tenencias de activos líquidos en los Estados Unidos, presentados en el cuadro 11-6, estimar la función agregada de consumo

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 L_t + \varepsilon_t,$$

sujeta a la restricción de que

$$\frac{1}{2} \leq \beta_2 \leq 1.$$

11-10. Supóngase que la función agregada de consumo se especifica en la forma

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 L_t + \beta_4 Y_t L_t + \varepsilon_t.$$

Utilizando los datos del cuadro 11-6:

- a. Obtener las estimaciones minimocuadráticas de los coeficientes de regresión y sus errores standard estimados.

- b. Calcular el valor de R^2 y contrastar la existencia de relación entre la variable dependiente y las explicativas.
- c. Contrastar la hipótesis de que $\beta_4 = 0$.

11-11. Sea $Y_i = Q_i/L_i$, y $X_i = K_i/L_i$, siendo Q_i = cantidad producida, L_i = cantidad de trabajo, y K_i = cantidad de capital, y el subíndice i indica la empresa. Con los datos presentados en el cuadro 11-4, estimar la siguiente relación:

$$\frac{Y_i^{\lambda} - 1}{\lambda} = \alpha + \beta \left(\frac{X_i^{\lambda} - 1}{\lambda} \right) + \varepsilon_i.$$

Contrastar la hipótesis de que $\lambda = 0$.

11-12. Considerar el siguiente modelo con retardos distribuidos:

$$Y_t = \alpha + \beta(w_0X_t + w_1X_{t-1} + w_2X_{t-2} + w_3X_{t-3}) + \varepsilon_t.$$

Suponemos que las ponderaciones w_i ($i = 0,1,2,3$) se hallan sobre un polinomio de segundo grado y queremos que cumplan las restricciones siguientes:

$$w_0 + w_1 + w_2 + w_3 = 1,$$

$$w_{-1} = 0,$$

$$w_4 = 0.$$

Indíquese cómo se pueden estimar los coeficientes de este modelo utilizando la estimación por mínimos cuadrados ordinarios.

Cuadro 11-6*

	C_t	Y_t	L_t
1955 I	248.7	263.0	207.6
II	253.7	271.5	209.4
III	259.9	276.5	211.1
IV	261.8	281.4	213.2
1956 I	263.2	282.0	214.1
II	263.7	286.2	216.5
III	263.4	287.7	217.3
IV	266.9	291.0	217.3
1957 I	268.9	291.1	218.2
II	270.4	294.6	218.5
III	273.4	296.1	219.8
IV	272.1	293.3	219.5
1958 I	268.9	291.3	220.5
II	270.9	292.6	222.7
III	274.4	299.9	225.0
IV	278.7	302.1	229.4
1959 I	283.8	305.9	232.2
II	289.7	312.5	235.2
III	290.8	311.3	237.2
IV	292.8	313.2	237.7
1960 I	295.4	315.4	238.0
II	299.5	320.3	238.4
III	298.6	321.0	240.1
IV	299.6	320.1	243.3
1961 I	297.0	318.4	246.1
II	301.6	324.8	250.0

Fuente: Z. Griliches *et al.*, «Notes on Estimated Aggregate Quarterly Consumption Functions», *Econometrica*, vol. 30, julio 1962, p. 500.

* Todos los valores (en miles de millones de dólares) están desestacionalizados y se han dividido por el índice de precios correspondiente.

Capítulo 12

El modelo generalizado de regresión lineal y sus aplicaciones

El modelo clásico de regresión se basa en unos supuestos bastante restrictivos acerca del comportamiento del término de perturbación de la regresión, pero existe un modelo alternativo conocido con el nombre de «modelo generalizado de regresión lineal» bastante menos restrictivo en este sentido. En la sección 12.1 estudiaremos este modelo y sus implicaciones para la estimación de los coeficientes de regresión. En la sección 12.2 se presentará una interesante aplicación de este modelo al caso en que las observaciones proceden de un número de unidades de corte transversal a lo largo del tiempo. Como este tipo de observaciones es cada vez más abundante, el desarrollo de procedimientos de estimación adecuados tiene gran utilidad. El modelo generalizado de regresión lineal tiene también otra aplicación muy ingeniosa en la estimación de un conjunto de ecuaciones de regresión cuyas perturbaciones están correlacionadas. En la sección 12.3 analizaremos esta aplicación, que proporciona, además, un nexo de unión entre los modelos uniecuacionales y los modelos de ecuaciones simultáneas que se estudiarán en el capítulo 13.

12.1 El modelo generalizado de regresión lineal

El modelo clásico de regresión normal lineal se caracteriza por un número de supuestos acerca de la perturbación estocástica de la ecuación de regresión, que incluyen los de homoscedasticidad y no existencia de autocorrelación. De modo específico, se supone que el término de perturbación ϵ_i de la expresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_K X_{iK} + \epsilon_i$$

cumple los requisitos siguientes:

$$E(\epsilon_i^2) = \sigma^2 \text{ para todo } i,$$

$$E(\epsilon_i \epsilon_j) = 0 \text{ para todo } i \neq j.$$

Estos supuestos se han presentado en (10.4) y (10.5), y pueden describirse en notación matricial, como en (10.4a-10.5a), en la forma

$$E(\epsilon\epsilon') = \sigma^2 I_n,$$

en la que

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

y I_n es una matriz identidad de orden $(n \times n)$. Si no efectuamos estos dos supuestos, pero mantenemos todos los supuestos restantes del modelo clásico de regresión normal lineal, tenemos el llamado *modelo generalizado de regresión lineal* (algunos autores eliminan también el supuesto de normalidad de ϵ_i). La descripción completa de este modelo es

$$(12.1) \quad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_K X_{iK} + \epsilon_i.$$

$$(12.2) \quad \text{La distribución conjunta de } \epsilon_1, \epsilon_2, \dots, \epsilon_n \text{ es una normal multivariante,}$$

$$(12.3) \quad E(\epsilon_i) = 0, \quad i = 1, 2, \dots, n.$$

$$(12.4) \quad E(\epsilon_i \epsilon_j) = \sigma_{ij}, \quad i, j = 1, 2, \dots, n.$$

$$(12.5) \quad \text{Todas las variables explicativas son no estocásticas, y tales que, para todos los tamaños muestrales,}$$

$$\frac{1}{n} \sum_{i=1}^n (X_{ik} - \bar{X}_k)^2$$

es un número finito distinto de cero para todo $k = 2, 3, \dots, K$.

$$(12.6) \quad \text{El número de observaciones es mayor que el número de variables explicativas más uno, es decir, } n > K.$$

$$(12.7) \quad \text{No existe ninguna relación lineal exacta entre las variables explicativas.}$$

Obsérvese que, según (12.3) y (12.4), σ_{ii} es la variancia de ϵ_i , y σ_{ij} ($i \neq j$) es la covariancia entre ϵ_i y ϵ_j . Si utilizamos la notación matricial, podemos presentar (12.1) en la forma

$$(12.1a) \quad Y = X\beta + \epsilon,$$

en la que Y es el vector $(n \times 1)$ de los valores muestrales de Y , X es la matriz $(n \times K)$ de los valores muestrales de $X_{i1}, X_{i2}, \dots, X_{iK}$ (en la que $X_{i1} = 1$ para todo i), β es el vector $(K \times 1)$ de los coeficientes de regresión, y ϵ es el vector $(n \times 1)$ de los valores muestrales de ϵ . El supuesto (12.4) puede escribirse en la forma

$$(12.4a) \quad E(\epsilon\epsilon') = \Omega,$$

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}.$$

en la que

Este modelo se llama «generalizado» porque incluye a otros modelos como casos especiales. El modelo clásico de regresión normal lineal es uno de estos casos especiales, en el que Ω es una matriz diagonal con σ^2 en todos los elementos de la diagonal principal. Otro caso especial es el modelo heteroscedástico, en el que Ω es también diagonal, pero los elementos de la diagonal principal no son necesariamente iguales. En el modelo en que las perturbaciones siguen un esquema autorregresivo de primer orden, la matriz Ω se convierte en

$$\Omega = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Pasemos ahora al problema de la estimación de los parámetros del modelo generalizado de regresión lineal. De momento, supondremos que se conocen las variancias y covariancias de las perturbaciones (es decir, los elementos de la matriz Ω). En primer lugar, vemos que los *estimadores minimocuadráticos ordinarios* de los coeficientes de regresión se obtienen minimizando

$$\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_{i2} - \dots - \beta_K X_{iK})^2,$$

y, por tanto, son iguales a los estimadores minimocuadráticos del modelo clásico de regresión normal lineal; es decir,

$$\hat{\beta} = (X'X)^{-1}(X'Y).$$

Obsérvese que

$$E(\hat{\beta}) = E(X'X)^{-1}X'[X\beta + \epsilon] = \beta + E(X'X)^{-1}(X'\epsilon) = \beta$$

$$\text{y} \quad \lim \text{prob } \hat{\beta} = \beta + \lim \text{prob } \left[\frac{1}{n} X'X \right]^{-1} \left[\frac{1}{n} X'\epsilon \right] = \beta.$$

Esto indica que los estimadores minimocuadráticos ordinarios de β en el modelo generalizado de regresión lineal son insesgados y consistentes.

A continuación obtendremos los *estimadores lineales insesgados óptimos* (ELIO) de los coeficientes de regresión. Para obtener el ELIO de β_1 , haremos

$$\hat{\beta}_1 = \sum_{i=1}^n a_i Y_i$$

en la que $a_1, a_2, \dots, a_n$ son unas constantes que debemos calcular. En notación matricial escribiremos

$$\hat{\beta}_1 = Y'a,$$

en la que

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

La esperanza matemática de $\hat{\beta}_1$ es

$$E(\hat{\beta}_1) = E\left(\sum_i a_i Y_i\right) = \beta_1 \sum_i a_i + \beta_2 \sum_i a_i X_{i2} + \dots + \beta_K \sum_i a_i X_{iK}.$$

Según la condición de insesgadez de $\hat{\beta}_1$, es necesario que

$$\sum_i a_i = 1,$$

$$\sum_i a_i X_{i2} = 0,$$

$\vdots$

$$\sum_i a_i X_{iK} = 0,$$

o, en notación matricial,

$$X'a = \mathbf{1}_1,$$

en la que $\mathbf{1}_2$ es un vector $(K \times 1)$ definido en la forma

$$\mathbf{1}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

La variancia de $\hat{\beta}_1$ es

$$\text{Var}(\hat{\beta}_1) = E\left[\sum_i a_i Y_i - E\left(\sum_i a_i Y_i\right)\right]^2 = E\left(\sum_i a_i \epsilon_i\right)^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij},$$

o

$$\text{Var}(\hat{\beta}_1) = a' \Omega a.$$

Por tanto, tenemos que obtener los valores de $a_1, a_2, \dots, a_n$ que minimizan $\text{Var}(\hat{\beta}_1)$ sujeta a la condición de que $X'a = \mathbf{1}_1$. Utilizando el método de los multiplicadores de Lagrange, formamos

$$H = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij} - \lambda_1 \left(\sum_{i=1}^n a_i - 1 \right) - \lambda_2 \sum_{i=1}^n a_i X_{i2} - \dots - \lambda_K \sum_{i=1}^n a_i X_{iK}.$$

Derivando H respecto de $a_1, a_2, \dots, a_n, \lambda_1, \lambda_2, \dots, \lambda_K$, e igualando a ceros estas derivadas, obtenemos

$$2(a_1 \sigma_{11} + a_2 \sigma_{12} + \dots + a_n \sigma_{1n}) - \lambda_1 - \lambda_2 X_{12} - \dots - \lambda_K X_{1K} = 0,$$

$$2(a_1 \sigma_{21} + a_2 \sigma_{22} + \dots + a_n \sigma_{2n}) - \lambda_1 - \lambda_2 X_{22} - \dots - \lambda_K X_{2K} = 0,$$

$\vdots$

$$2(a_1 \sigma_{n1} + a_2 \sigma_{n2} + \dots + a_n \sigma_{nn}) - \lambda_1 - \lambda_2 X_{n2} - \dots - \lambda_K X_{nK} = 0,$$

$$-(\sum_i a_i - 1) = 0,$$

$$-\sum_i a_i X_{i2} = 0,$$

$\vdots$

$$-\sum_i a_i X_{iK} = 0.$$

En notación matricial podemos escribirlo en la forma

$$(12.8) \quad 2\Omega a - X\lambda = 0$$

$$(12.9) \quad X'a = \mathbf{1}_1,$$

en la que

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{bmatrix}$$

y $\mathbf{0}$ es un vector $(n \times 1)$ de ceros. A partir de (12.8) obtenemos

$$(12.8a) \quad \mathbf{a} = \frac{1}{2} \Omega^{-1} \mathbf{X} \lambda.$$

Sustituyendo $\mathbf{a}$ por su valor según esta expresión en (12.9), tenemos

$$(12.10) \quad \frac{1}{2} \mathbf{X}' \Omega^{-1} \mathbf{X} \lambda = \mathbf{u}_1,$$

o

$$(12.10a) \quad \lambda = 2(\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_1.$$

Sustituyendo λ por su valor según (12.10a), en (12.8a), obtenemos

$$(12.11) \quad \mathbf{a} = \Omega^{-1} \mathbf{X} (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_1.$$

y, por tanto,

$$(12.12) \quad \begin{aligned} \beta_1 &= \mathbf{Y}' \mathbf{a} = (\mathbf{Y}' \Omega^{-1} \mathbf{X}) (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_1 \\ &= \mathbf{u}_1' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} (\mathbf{X}' \Omega^{-1} \mathbf{Y}). \end{aligned}$$

Este es pues el ELIO de β_1 .¹ Además, como

$$\text{Var}(\beta_1) = \mathbf{a}' \Omega \mathbf{a},$$

podemos utilizar (12.11) para obtener

$$(12.13) \quad \begin{aligned} \text{Var}(\beta_1) &= [\mathbf{u}_1' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}' \Omega^{-1}] \Omega [\Omega^{-1} \mathbf{X} (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_1] \\ &= \mathbf{u}_1' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_1. \end{aligned}$$

procedimiento utilizado para obtener el ELIO de β_1 y su variancia puede utilizarse también con los demás coeficientes de regresión. Los resultados son

$$\beta_2 = \mathbf{u}_2' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} (\mathbf{X}' \Omega^{-1} \mathbf{Y}),$$

$$\text{Var}(\beta_2) = \mathbf{u}_2' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_2,$$

$\vdots$

$$\beta_K = \mathbf{u}_K' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} (\mathbf{X}' \Omega^{-1} \mathbf{Y}),$$

$$\text{Var}(\beta_K) = \mathbf{u}_K' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_K.$$

En general,

$$\beta_k = \mathbf{u}_k' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} (\mathbf{X}' \Omega^{-1} \mathbf{Y}),$$

$$\text{Var}(\beta_k) = \mathbf{u}_k' (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{u}_k, \quad (k = 1, 2, \dots, K),$$

en el que $\mathbf{u}_k$ es un vector $(K \times 1)$ en el que el elemento k -ésimo es igual a la unidad, y todos los demás elementos son cero. Estos resultados pueden resumirse escribiendo

$$(12.14) \quad \tilde{\beta} = (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} (\mathbf{X}' \Omega^{-1} \mathbf{Y}),$$

$$(12.15) \quad E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1},$$

en la que

$$\tilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}.$$

El estimador $\tilde{\beta}$ se denomina generalmente *estimador minimocuadrático generalizado de Aitken*.² Representando los elementos de la inversa de Ω mediante los supraíndices correspondientes, en la forma

$$\Omega^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} & \dots & \sigma^{1n} \\ \sigma^{21} & \sigma^{22} & \dots & \sigma^{2n} \\ \vdots & \vdots & & \vdots \\ \sigma^{n1} & \sigma^{n2} & & \sigma^{nn} \end{bmatrix},$$

1. La demostración de que este resultado corresponde a una variancia mínima y no máxima puede verse en J. Johnston, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, pp. 183-84.

2. A. C. Aitken, «On Least Squares and Linear Combination of Observations», *Proceedings of the Royal Society of Edinburgh*, vol. 55, 1934-1935, pp. 42-48.

podemos representar los elementos de $(X'\Omega^{-1}X)$ y $(X'\Omega^{-1}Y)$ y escribir

$$\tilde{\beta} = \begin{bmatrix} \sum_i \sum_j \sigma^{ij} X_{i1}^2 & \sum_i \sum_j \sigma^{ij} X_{i1} X_{j2} & \dots & \sum_i \sum_j \sigma^{ij} X_{i1} X_{jK} \\ \sum_i \sum_j \sigma^{ij} X_{i2} X_{j1} & \sum_i \sum_j \sigma^{ij} X_{i2}^2 & \dots & \sum_i \sum_j \sigma^{ij} X_{i2} X_{jK} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i \sum_j \sigma^{ij} X_{iK} X_{j1} & \sum_i \sum_j \sigma^{ij} X_{iK} X_{j2} & \dots & \sum_i \sum_j \sigma^{ij} X_{iK}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i \sum_j \sigma^{ij} X_{i1} Y_j \\ \sum_i \sum_j \sigma^{ij} X_{i2} Y_j \\ \vdots \\ \sum_i \sum_j \sigma^{ij} X_{iK} Y_j \end{bmatrix}$$

Además, definiendo

$$\tilde{\beta} = \begin{bmatrix} \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \vdots \\ \tilde{\beta}_K \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} (X_{12} - \tilde{X}_2) & (X_{13} - \tilde{X}_3) & \dots & (X_{1K} - \tilde{X}_K) \\ (X_{22} - \tilde{X}_2) & (X_{23} - \tilde{X}_3) & \dots & (X_{2K} - \tilde{X}_K) \\ \vdots & \vdots & \ddots & \vdots \\ (X_{n2} - \tilde{X}_2) & (X_{n3} - \tilde{X}_3) & \dots & (X_{nK} - \tilde{X}_K) \end{bmatrix},$$

$$\underline{Y} = \begin{bmatrix} Y_1 - \tilde{Y} \\ Y_2 - \tilde{Y} \\ \vdots \\ Y_n - \tilde{Y} \end{bmatrix},$$

donde $X_{i2} = \sum_j \sigma^{ij} X_{j2} / \sum_i \sigma^{ii}$, etc., entonces

como en el caso anterior, se puede mostrar que

$$(12.16) \quad \tilde{\beta} = (\underline{X}'\Omega^{-1}\underline{X})^{-1}(\underline{X}'\Omega^{-1}\underline{Y})$$

y

$$(12.17) \quad E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = (\underline{X}'\Omega^{-1}\underline{X})^{-1}.$$

Los estimadores *maximoverosímiles* de los coeficientes de regresión pueden obtenerse observando que, según los supuestos del modelo, la distribución conjunta de $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ viene dada por³

$$f(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = (2\pi)^{-(n/2)} |\Omega|^{-1/2} e^{-(1/2)\epsilon'\Omega^{-1}\epsilon},$$

3. Ver A. S. Goldberger, *Econometric Theory*. Nueva York, Wiley, 1964, p. 104. Si las ϵ de la muestra no son independientes de todas las ϵ fuera de la muestra, la distribución conjunta está condicionada a estas últimas.

La que Ω representa al determinante de la matriz Ω . La función *logarítmica* de verosimilitud para $Y_1, Y_2, \dots, Y_n$ viene dada por⁴

$$(12.18) \quad L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Omega| - \frac{1}{2} (Y - X\beta)' \Omega^{-1} (Y - X\beta) \\ = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Omega| \\ - \frac{1}{2} (Y'\Omega^{-1}Y - 2Y'\Omega^{-1}X\beta + \beta'X'\Omega^{-1}X\beta).$$

Derivando L respecto de β e igualando a cero el resultado, obtenemos

$$-\frac{1}{2} (-2X'\Omega^{-1}Y + 2X'\Omega^{-1}X\beta) = 0,$$

en la que 0 es un vector $(K \times 1)$ de ceros. En esta expresión podemos despejar

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y),$$

que es igual a la expresión del estimador minimocuadrático generalizado de Aitken indicado en (12.14) para β . Puede mostrarse que, en condiciones muy generales, dicho estimador posee las propiedades de consistencia, eficiencia asintótica y normalidad asintótica.⁵

Consideremos ahora algunos casos especiales. En primer lugar, si

$$\Omega = \sigma^2 I_n,$$

es decir, cuando el modelo generalizado se reduce al modelo *clásico*, tenemos

$$\Omega^{-1} = \frac{1}{\sigma^2} I_n,$$

4. Acerca del teorema de la transformación que permite pasar de las ϵ a las Y , véase *ibid.*, p. 106.

5. Ver A. Zellner, «An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias», *Journal of the American Statistical Association*, vol. 57, junio 1962, pp. 348-68.

12.2 Consideración conjunta de datos de series temporales y de sección transversal

El problema de las restricciones adecuadas acerca de Ω tiene un especial interés y significación en relación con las *observaciones conjuntas de series temporales y sección transversal*, como es el caso cuando se efectúan observaciones de varias economías domésticas (o Estados, países, etc.) a lo largo de varios períodos de tiempo. En tal caso, es probable que el comportamiento de las perturbaciones respecto a las unidades de sección o corte transversal (economías domésticas, Estados, países, etc.) sea distinto al comportamiento de las perturbaciones de una unidad determinada a lo largo del tiempo. En concreto, la relación entre las perturbaciones de dos economías domésticas en un momento determinado puede ser distinta de la relación entre las perturbaciones de una economía doméstica en dos períodos de tiempo distintos. Evidentemente, los distintos tipos de especificaciones previas acerca de las perturbaciones producirán distintos tipos de restricciones sobre Ω . En general, la ecuación de regresión para este tipo de datos puede escribirse en la forma

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \dots + \beta_K X_{it,K} + \epsilon_{it} \quad (i = 1, 2, \dots, N; t = 1, 2, \dots, T).$$

Es decir, los datos muestrales están compuestos por observaciones de N unidades de corte transversal durante T períodos de tiempo. En total, existen $n = N \times T$ observaciones. Suponemos que las variables explicativas y el término de perturbación de la regresión cumplen los supuestos del modelo generalizado de regresión lineal. En la mayoría de los casos (pero no necesariamente siempre) tendremos $X_{it,i} = 1$ para todo i y todo t . Utilizando la notación matricial, la ecuación de regresión puede escribirse en la forma

$$Y = X\beta + \epsilon,$$

en la que

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{NT} \end{bmatrix}, \quad X = \begin{bmatrix} X_{11,1} & X_{11,2} & \dots & X_{11,K} \\ X_{12,1} & X_{12,2} & \dots & X_{12,K} \\ \vdots & \vdots & & \vdots \\ X_{1T,1} & X_{1T,2} & \dots & X_{1T,K} \\ X_{21,1} & X_{21,2} & \dots & X_{21,K} \\ X_{22,1} & X_{22,2} & \dots & X_{22,K} \\ \vdots & \vdots & & \vdots \\ X_{NT,1} & X_{NT,2} & \dots & X_{NT,K} \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{NT} \end{bmatrix}.$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}.$$

Por tanto,

$$(12.21) \quad \Omega =$$

$$\begin{bmatrix} E(\epsilon_{11}^2) & E(\epsilon_{11}\epsilon_{12}) & \dots & E(\epsilon_{11}\epsilon_{1T}) & E(\epsilon_{11}\epsilon_{21}) & E(\epsilon_{11}\epsilon_{22}) & \dots & E(\epsilon_{11}\epsilon_{NT}) \\ E(\epsilon_{12}\epsilon_{11}) & E(\epsilon_{12}^2) & \dots & E(\epsilon_{12}\epsilon_{1T}) & E(\epsilon_{12}\epsilon_{21}) & E(\epsilon_{12}\epsilon_{22}) & \dots & E(\epsilon_{12}\epsilon_{NT}) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ E(\epsilon_{1T}\epsilon_{11}) & E(\epsilon_{1T}\epsilon_{12}) & \dots & E(\epsilon_{1T}^2) & E(\epsilon_{1T}\epsilon_{21}) & E(\epsilon_{1T}\epsilon_{22}) & \dots & E(\epsilon_{1T}\epsilon_{NT}) \\ E(\epsilon_{21}\epsilon_{11}) & E(\epsilon_{21}\epsilon_{12}) & \dots & E(\epsilon_{21}\epsilon_{1T}) & E(\epsilon_{21}^2) & E(\epsilon_{21}\epsilon_{22}) & \dots & E(\epsilon_{21}\epsilon_{NT}) \\ E(\epsilon_{22}\epsilon_{11}) & E(\epsilon_{22}\epsilon_{12}) & \dots & E(\epsilon_{22}\epsilon_{1T}) & E(\epsilon_{22}\epsilon_{21}) & E(\epsilon_{22}^2) & \dots & E(\epsilon_{22}\epsilon_{NT}) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ E(\epsilon_{NT}\epsilon_{11}) & E(\epsilon_{NT}\epsilon_{12}) & \dots & E(\epsilon_{NT}\epsilon_{1T}) & E(\epsilon_{NT}\epsilon_{21}) & E(\epsilon_{NT}\epsilon_{22}) & \dots & E(\epsilon_{NT}^2) \end{bmatrix}.$$

Esta especificación proporciona un marco general para el análisis de los distintos modelos que se ocupan de las observaciones conjuntas de series temporales y corte transversal.

Modelo heteroscedástico en sentido transversal y autorregresivo en sentido temporal

Una forma de especificar el comportamiento de las perturbaciones cuando utilizamos datos de corte transversal y de series temporales consiste en combinar los supuestos que se efectúan a menudo al utilizar cada uno de estos tipos de datos por separado. En el caso de las observaciones de corte transversal (por ejemplo, observaciones de las economías domésticas individuales en un punto, o período, de tiempo) se supone con frecuencia que las perturbaciones de la regresión son independientes entre sí pero heteroscedásticas.⁷ Por lo que respecta a los datos de las series temporales, generalmente se sospecha que las perturbaciones son autorregresivas, pero no necesariamente heteroscedásticas. Al utilizar observaciones conjuntas, de corte transversal y de series

7. Ver, por ejemplo, S. J. Prais y H. S. Houthakker, *The Analysis of Family Budgets*. Cambridge, Inglaterra, The University Press, 1955. Más adelante se eliminará el supuesto de independencia entre las unidades de corte transversal.

temporales, podemos combinar estos supuestos y adoptar un *modelo heterocedástico en sentido transversal y autorregresivo en sentido temporal*. Podemos caracterizar el modelo de la forma siguiente:

$$(12.22) \quad E(\varepsilon_{it}^2) = \sigma_i^2 \quad (\text{heteroscedasticidad})$$

$$(12.23) \quad E(\varepsilon_{it} \varepsilon_{jt}) = 0 \quad (i \neq j) \quad (\text{independencia en sentido transversal})$$

$$(12.24) \quad \varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + u_{it} \quad (\text{autorregresión})$$

siendo

$$u_{it} \sim N(0, \sigma_{ui}^2),$$

$$\varepsilon_{i0} \sim N\left(0, \frac{\sigma_{ui}^2}{1 - \rho_i^2}\right),$$

y

$$E(\varepsilon_{i,t-1} u_{it}) = 0 \quad \text{para todo } i, j.$$

Obsérvese que en este modelo el valor del parámetro ρ puede variar de una unidad de corte trasversal a otra. A partir de estas especificaciones se deduce que:

$$E(\varepsilon_{it} \varepsilon_{is}) = \rho_i^{t-s} \sigma_i^2 \quad (t \geq s),$$

$$E(\varepsilon_{it} \varepsilon_{jt}) = 0 \quad (i \neq j).$$

Efectuando la sustitución adecuada en (12.21) vemos que, en este modelo,

$$(12.25) \quad \Omega = \begin{bmatrix} \sigma_1^2 \mathbf{P}_1 & 0 & \dots & 0 \\ 0 & \sigma_2^2 \mathbf{P}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \mathbf{P}_N \end{bmatrix},$$

siendo

$$\mathbf{P}_i = \begin{bmatrix} 1 & \rho_i & \rho_i^2 & \dots & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & \dots & \rho_i^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \dots & 1 \end{bmatrix},$$

y cada $\mathbf{0}$ representa una matriz $(T \times T)$ de ceros.

Para obtener estimaciones consistentes de los elementos de (12.25), podemos proceder en la forma siguiente. En primer lugar, aplicaremos el método de los mínimos ordinarios a las $N \times T$ observaciones. Las estimaciones de

de los coeficientes de regresión son insesgadas y consistentes, y pueden utilizarse para calcular los residuos de la regresión e_{it} . A partir de estos residuos podemos obtener estimaciones de $\hat{\rho}_i$, a las que llamaremos $\hat{\rho}_i$, mediante

$$(12.26) \quad \hat{\rho}_i = \frac{\sum e_{it} e_{i,t-1}}{\sum e_{it}^2} \quad (i = 2, 3, \dots, T).$$

Como se ha indicado en la sección 8.2, $\hat{\rho}_i$ es un estimador consistente de ρ_i . A continuación, utilizaremos los valores de $\hat{\rho}_i$ para transformar las observaciones de acuerdo con (8.52); es decir, formaremos la expresión

$$(12.27) \quad Y_{it}^* = \beta_1 X_{it,1}^* + \beta_2 X_{it,2}^* + \dots + \beta_K X_{it,K}^* + u_{it}^*,$$

en la que

$$Y_{it}^* = Y_{it} - \hat{\rho}_i Y_{i,t-1},$$

$$X_{it,k}^* = X_{it,k} - \hat{\rho}_i X_{i,t-1,k} \quad (k = 1, 2, \dots, K),$$

$$u_{it}^* = e_{it} - \hat{\rho}_i e_{i,t-1},$$

$$t = 2, 3, \dots, T,$$

$$i = 1, 2, \dots, N.$$

Ahora tenemos que estimar σ_{ui}^2 a partir de observaciones que sean no autorregresivas, al menos asintóticamente. Para ello, podemos aplicar el método de los mínimos cuadrados ordinarios a la ecuación (12.27), para la cual disponemos de $N(T-1)$ observaciones. Los residuos de la regresión, a los que llamaremos $\hat{u}_{it}^*$, pueden utilizarse para estimar las variancias de u_{it} (o sea σ_{ui}^2) a través de

$$(12.28) \quad s_{ui}^2 = \frac{1}{T-K-1} \sum_{t=2}^T \hat{u}_{it}^{*2}.$$

Como

$$\sigma_{ui}^2 = \sigma_i^2 (1 - \rho_i^2),$$

podemos estimar σ_i^2 a través de

$$(12.29) \quad s_i^2 = \frac{s_{ui}^2}{1 - \hat{\rho}_i^2}.$$

Como $\hat{\rho}_i$ es un estimador consistente de ρ_i , y s_{ui}^2 es un estimador consistente de σ_{ui}^2 , s_i^2 será también un estimador consistente de σ_i^2 .

Una vez obtenidos los estimadores consistentes de ρ_i y σ_i^2 , hemos finalizado la tarea de obtener estimadores consistentes de los elementos de Ω . Sustituyendo $\hat{\Omega}$ por estos valores en la ecuación (12.19), y $\hat{\Omega}$ por Ω en (12.20), obtenemos las estimaciones que deseamos de los coeficientes de regresión y de las variancias. La solución de las ecuaciones (12.19) y (12.20) es muy laboriosa en términos de los cálculos, pero podemos utilizar un procedimiento ligero-

Se supone que el valor inicial de ϵ posee las propiedades siguientes:

$$\epsilon_{i0} \sim N\left(0, \frac{\phi_{ii}}{1 - \rho_i^2}\right),$$

$$E(\epsilon_{i0}\epsilon_{j0}) = \frac{\phi_{ij}}{1 - \rho_i\rho_j}.$$

Obsérvese que en este modelo las variancias de las u se representan por el símbolo Φ en lugar de por σ_u^2 , como hacíamos en el modelo anterior. En este caso, la matriz Ω es

$$(12.36) \quad \Omega = \begin{bmatrix} \sigma_{11}P_{11} & \sigma_{12}P_{12} & \dots & \sigma_{1N}P_{1N} \\ \sigma_{21}P_{21} & \sigma_{22}P_{22} & \dots & \sigma_{2N}P_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1}P_{N1} & \sigma_{N2}P_{N2} & \dots & \sigma_{NN}P_{NN} \end{bmatrix},$$

en la que

$$P_u = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \dots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & \dots & \rho_j^{T-2} \\ \rho_i^2 & \rho_i & 1 & \dots & \rho_j^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \dots & 1 \end{bmatrix}.$$

Para obtener estimaciones consistentes de los elementos de Ω , aplicamos en primer lugar el método de los mínimos cuadrados ordinarios a todas las observaciones originales y calculamos los correspondientes residuos e_{it} . Estos residuos se utilizan para calcular ρ_i aplicando la fórmula (12.26). Por medio de $\hat{\rho}_i$ transformados las variables y y formamos la ecuación (12.27). A continuación, se aplica de nuevo el método de los mínimos cuadrados ordinarios a esta ecuación y se calculan los residuos $\hat{u}_{it}^*$. Las variancias y covariancias de las ϵ (es decir, las σ_{it}) se pueden estimar a partir de

$$(12.37) \quad s_{ij} = \frac{\hat{\phi}_{ij}}{1 - \hat{\rho}_i\hat{\rho}_j},$$

en la que

$$\hat{\phi}_{ij} = \frac{1}{T - K - 1} \sum_{t=2}^T \hat{u}_{it}^* \hat{u}_{jt}^*.$$

Por este procedimiento obtenemos estimaciones consistentes de ρ_i y σ_{ii} , y, por tanto, de Ω . Esto nos permite utilizar las fórmulas (12.19) y (12.20) obtenien-

do así estimaciones asintóticamente eficientes de los coeficientes de regresión y de sus variancias.

Podemos simplificar este procedimiento aplicando las fórmulas de estimación de Aitken modificadas, (12.19) y (12.20), a las variables transformadas, es decir, utilizando un estimador de β

$$(12.38) \quad \hat{\beta} = (X^* \hat{\Phi}^{-1} X^*)^{-1} (X^* \hat{\Phi}^{-1} Y^*)$$

con

$$(12.39) \quad \text{Var-Cov asintóticas}(\hat{\beta}) = (X^* \hat{\Phi}^{-1} X^*)^{-1}.$$

En este caso, Y^* es el vector $N(T-1) \times 1$ de las observaciones transformadas de $Y_{it}^* = Y_{it} - \hat{\rho}_i Y_{i,t-1}$ y X^* es la matriz de orden $N(T-1) \times K$ de las observaciones transformadas $X_{it}^*, kX_{it,k} = -\hat{\rho}_i X_{i,t-1,k}$ ($t = 2, 3, \dots, T$; $i = 1, 2, \dots, N$; $k = 1, 2, \dots, K$). La matriz Φ es la matriz de variancias y covariancias estimadas de las u , y es de orden $N(T-1) \times N(T-1)$. Su expresión completa es

$$(12.40) \quad \hat{\Phi} = \begin{bmatrix} \hat{\phi}_{11}I_{T-1} & \hat{\phi}_{12}I_{T-1} & \dots & \hat{\phi}_{1N}I_{T-1} \\ \hat{\phi}_{21}I_{T-1} & \hat{\phi}_{22}I_{T-1} & \dots & \hat{\phi}_{2N}I_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\phi}_{N1}I_{T-1} & \hat{\phi}_{N2}I_{T-1} & \dots & \hat{\phi}_{NN}I_{T-1} \end{bmatrix},$$

en la que las Φ_i son las definidas en (12.37), e I_{T-1} es una matriz identidad de orden $(T-1) \times (T-1)$. En general, el valor del estimador de (12.38) será distinto del obtenido utilizando (12.19), pero las propiedades asintóticas de ambos estimadores son las mismas.

Modelo de los componentes de error

Otro de los enfoques en la especificación del comportamiento de las perturbaciones al combinar los datos de corte transversal y de series temporales es el adoptado por los proponentes del llamado *modelo de los componentes de error*. En este caso, el supuesto básico es el de que el término de perturbación de la regresión está formado por tres componentes independientes, uno de los cuales está asociado al tiempo, mientras que el segundo corresponde a las unidades de corte transversal y el tercero varía en las dos dimensiones. Concretamente,

$$(12.41) \quad \epsilon_{it} = u_i + v_t + w_{it} \quad (i = 1, 2, \dots, N; t = 1, 2, \dots, T),$$

en la que

$$u_i \sim N(0, \sigma_u^2),$$

$$v_t \sim N(0, \sigma_v^2),$$

$$w_{it} \sim N(0, \sigma_w^2),$$

y los componentes u_i , v_t , y w_{it} cumplen las condiciones siguientes:

$$E(u_i v_t) = E(u_i w_{it}) = E(v_t w_{it}) = 0,$$

$$E(u_i u_j) = 0 \quad (i \neq j),$$

$$E(v_t v_s) = 0 \quad (t \neq s),$$

$$E(w_{it} w_{js}) = E(w_{it} w_{jt}) = E(w_{it} w_{is}) = 0 \quad (i \neq j; t \neq s).$$

Obsérvese que esto implica que ε_{it} es *homoscedástico*, y que su variancia viene dada por

$$\text{Var}(\varepsilon_{it}) = \sigma^2 = \sigma_u^2 + \sigma_v^2 + \sigma_w^2.$$

El coeficiente de correlación entre ε_{it} y ε_{jt} ($i \neq j$), es decir, entre las perturbaciones correspondientes a dos unidades distintas de corte transversal en un momento de tiempo determinado, es

$$\frac{\text{Cov}(\varepsilon_{it}, \varepsilon_{jt})}{\sqrt{\text{Var}(\varepsilon_{it})\text{Var}(\varepsilon_{jt})}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2 + \sigma_w^2} \quad (i \neq j).$$

El coeficiente de correlación entre ε_{it} y ε_{is} ($i \neq s$), es decir, entre las perturbaciones correspondientes a una unidad de corte transversal en dos momentos distintos de tiempo, es

$$\frac{\text{Cov}(\varepsilon_{it}, \varepsilon_{is})}{\sqrt{\text{Var}(\varepsilon_{it})\text{Var}(\varepsilon_{is})}} = \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2 + \sigma_w^2} \quad (t \neq s).$$

Esta última característica del modelo de los componentes de error significa que, para cada unidad de corte transversal, la correlación entre las perturbaciones a lo largo del tiempo es la misma independientemente de lo distantes que se hallen entre sí. Esto presenta un marcado contraste con el supuesto habitual de autorregresión de primer orden, que implica que el grado de correlación disminuye geométricamente al aumentar el espacio de tiempo transcurrido. Por último, el coeficiente de correlación entre ε_{it} y ε_{jt} es

$$\frac{\text{Cov}(\varepsilon_{it}, \varepsilon_{jt})}{\sqrt{\text{Var}(\varepsilon_{it})\text{Var}(\varepsilon_{jt})}} = 0 \quad (i \neq j; t \neq s).$$

Sustituyendo estos resultados en (12.21), vemos que, en el modelo de los componentes de error, la matriz Ω es

(12.42)

$$\Omega = \begin{bmatrix} \sigma_u^2 \mathbf{A}_T & \sigma_v^2 \mathbf{I}_T & \dots & \sigma_w^2 \mathbf{I}_T \\ \sigma_v^2 \mathbf{I}_T & \sigma_u^2 \mathbf{A}_T & \dots & \sigma_w^2 \mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_w^2 \mathbf{I}_T & \sigma_v^2 \mathbf{I}_T & \dots & \sigma_u^2 \mathbf{A}_T \end{bmatrix},$$

en la que $\mathbf{A}_T$ es una matriz de orden $(T \times T)$ que se define en la forma

$$\mathbf{A}_T = \begin{bmatrix} \frac{\sigma^2}{\sigma_u^2} & 1 & \dots & 1 \\ 1 & \frac{\sigma^2}{\sigma_u^2} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & \frac{\sigma^2}{\sigma_u^2} \end{bmatrix}$$

e $\mathbf{I}_T$ es una matriz identidad de orden $T \times T$. Los elementos de la matriz pueden estimarse de la forma siguiente:⁹

$$(12.43) \quad \hat{\sigma}_w^2 = \frac{1}{(N-1)(T-1)} \sum_{i=1}^N \sum_{t=1}^T \left[e_{it} - \frac{1}{T} \sum_{t=1}^T e_{it} - \frac{1}{N} \sum_{i=1}^N e_{it} \right]^2,$$

$$\hat{\sigma}_u^2 = \frac{1}{T} \left\{ \frac{1}{(N-1)T} \sum_{i=1}^N \left[\sum_{t=1}^T e_{it} \right]^2 - \hat{\sigma}_w^2 \right\},$$

$$\hat{\sigma}_v^2 = \frac{1}{N} \left\{ \frac{1}{N(T-1)} \sum_{t=1}^T \left[\sum_{i=1}^N e_{it} \right]^2 - \hat{\sigma}_w^2 \right\},$$

en las que e_{it} representan los residuos obtenidos al aplicar el método de los mínimos cuadrados ordinarios a los datos originales. Utilizando estas estimaciones y las fórmulas (12.19) y (12.20), obtendremos unas estimaciones de los coeficientes de regresión con las mismas propiedades que el estimador minimocuadrático generalizado de Aitken.¹⁰

9. Ver T. D. Wallace y A. Hussain, «The Use of Error Components Models in Combining Cross-Section with Time Series Data», *Econometrica*, vol. 37, enero 1969, pp. 55-72.

10. Para una prueba de ello, véase *ibid.*

Modelo de la covariancia

Otro de los modelos que se utiliza corrientemente para operar con observaciones conjuntas de corte transversal y series temporales es el llamado *modelo de la covariancia*. Lo mencionamos al final de esta sección debido a que dicho modelo se estudia generalmente en el marco del modelo clásico de regresión, en lugar de considerarlo como un modelo generalizado de regresión. La idea subyacente en este modelo es el supuesto de que cada unidad de corte transversal y cada período de tiempo se caracterizan por una ordenada en el origen particular. Esta característica se incorpora a la ecuación de regresión mediante la introducción de variables binarias, con lo que la ecuación se convierte en

$$(12.44) \quad Y_{it} = \beta_1 + \beta_2 X_{it,2} + \cdots + \beta_K X_{it,K} + \gamma_2 Z_{2t} + \gamma_3 Z_{3t} + \cdots + \gamma_N Z_{Nt} \\ + \delta_2 W_{i2} + \delta_3 W_{i3} + \cdots + \delta_T W_{iT} + \epsilon_{it},$$

en la que $Z_{it} = 1$ en la i -ésima unidad de corte transversal,
 $= 0$ en los demás casos ($i = 2, 3, \dots, N$);
 $W_{it} = 1$ en el período de tiempo t ,
 $= 0$ en los demás casos ($t = 2, 3, \dots, T$).

Suponemos que la perturbación ϵ_{it} cumple los supuestos del modelo clásico de regresión normal lineal (aunque, naturalmente, podríamos considerar el caso en que ϵ_{it} es autorregresiva o heteroscedástica, y elegir el método de estimación adecuado). Obsérvese que, con la especificación anterior de la ecuación de regresión, tenemos

$$\begin{aligned} Y_{11} &= \beta_1 + \beta_2 X_{11,2} + \cdots + \beta_K X_{11,K} + \epsilon_{11}, \\ Y_{12} &= (\beta_1 + \delta_2) + \beta_2 X_{12,2} + \cdots + \beta_K X_{12,K} + \epsilon_{12}, \\ &\vdots \\ Y_{1T} &= (\beta_1 + \delta_T) + \beta_2 X_{1T,2} + \cdots + \beta_K X_{1T,K} + \epsilon_{1T}, \\ &\vdots \\ Y_{21} &= (\beta_1 + \gamma_2) + \beta_2 X_{21,2} + \cdots + \beta_K X_{21,K} + \epsilon_{21}, \\ &\vdots \\ Y_{22} &= (\beta_1 + \gamma_2 + \delta_2) + \beta_2 X_{22,2} + \cdots + \beta_K X_{22,K} + \epsilon_{22}, \\ &\vdots \\ Y_{NT} &= (\beta_1 + \gamma_N + \delta_T) + \beta_2 X_{NT,2} + \cdots + \beta_K X_{NT,K} + \epsilon_{NT}. \end{aligned}$$

La ecuación (12.44) contiene $K + (N-1) + (T-1)$ coeficientes que debemos estimar a partir de $N \times T$ observaciones. Si el modelo está correctamente especificado y se cumplen los supuestos clásicos, las estimaciones minimocuadráticas ordinarias de los coeficientes de regresión serán insesgadas y eficientes.

12.3 Regresiones aparentemente no relacionadas

Al estudiar los estimadores minimocuadráticos de los coeficientes de regresión bajo los supuestos del modelo clásico de regresión normal lineal, vimos que eran insesgados y eficientes. Este resultado se obtuvo en base al supuesto de que la especificación del modelo incluía *todo* lo que necesitamos saber acerca de la ecuación de regresión y de las variables que aparecen en ella. Si existe alguna otra información que no se ha tenido en cuenta, los resultados acerca de las propiedades de los estimadores minimocuadráticos no se pueden considerar válidos. Como ejemplo de una información adicional de este tipo, podemos tomar el conocimiento de que el término de perturbación de la regresión considerada puede estar correlacionado con el término de perturbación de otra ecuación de regresión (ver nota 2 del capítulo 7). En esta sección nos ocuparemos de este tipo de situaciones y examinaremos sus implicaciones para la estimación de los coeficientes de regresión.

Supongamos que queremos estimar los coeficientes de una o más de las M ecuaciones de regresión siguientes:

$$\begin{aligned} (12.45) \quad Y_{1t} &= \beta_{11} X_{1t,1} + \beta_{12} X_{1t,2} + \cdots + \beta_{1K_1} X_{1t,K_1} + \epsilon_{1t}, \\ Y_{2t} &= \beta_{21} X_{2t,1} + \beta_{22} X_{2t,2} + \cdots + \beta_{2K_2} X_{2t,K_2} + \epsilon_{2t}, \\ &\vdots \\ Y_{Mt} &= \beta_{M1} X_{Mt,1} + \beta_{M2} X_{Mt,2} + \cdots + \beta_{MK_M} X_{Mt,K_M} + \epsilon_{Mt}, \\ &t = 1, 2, \dots, T. \end{aligned}$$

Haciendo uso de la notación matricial, podemos escribir

$$\begin{aligned} (12.45a) \quad Y_1 &= X_1 \beta_1 + \epsilon_1, \\ Y_2 &= X_2 \beta_2 + \epsilon_2, \\ &\vdots \\ Y_M &= X_M \beta_M + \epsilon_M, \end{aligned}$$

o

$$(12.45b) \quad Y_m = X_m \beta_m + \epsilon_m \quad (m = 1, 2, \dots, M),$$

en la que Y_m es el vector $(T \times 1)$ de los valores muestrales de la variable dependiente, X_m es la matriz $(T \times K_m)$ de los valores muestrales de las variables explicativas, β_m es el vector $(K_m \times 1)$ de los coeficientes de regresión, y ϵ_m es el vector $(T \times 1)$ de los valores muestrales de las perturbaciones. Suponemos que ϵ_m se distribuye normalmente con media

$$(12.46) \quad E(\epsilon_{mt}) = 0 \quad (t = 1, 2, \dots, T)$$

y que su matriz de variancias y covariancias viene dada por

$$(12.47) \quad E(\epsilon_m \epsilon_m') = \sigma_{mm} \mathbf{I}_T,$$

en la que $\mathbf{I}_T$ es una matriz identidad de orden $(T \times T)$. Las variables explicativas se suponen no estocásticas y tales que $(\mathbf{X}'\mathbf{X})/T$ es no singular, y cuando $T \rightarrow \infty$ su límite es finito. Esto significa que se supone que todas las ecuaciones cumplen los supuestos del modelo clásico de regresión normal lineal. Supongamos, además, que no podemos eliminar la posibilidad de que las perturbaciones de las distintas ecuaciones estén *correlacionadas entre sí*. En este caso, tenemos

$$(12.48) \quad E(\epsilon_m \epsilon_p') = \sigma_{mp} \mathbf{I}_T \quad (m, p = 1, 2, \dots, M).$$

Así pues, σ_{mp} es la covariancia entre los términos de perturbación de la ecuación m y la ecuación p , que suponemos constante para todas las observaciones. Esta covariancia constituye el *único* lazo de unión entre las ecuaciones m y p , y al tratarse de una relación muy tenue, el sistema de M ecuaciones recibe el nombre de sistema de *ecuaciones de regresión aparentemente no relacionadas*. Como ejemplos de este tipo de regresiones podríamos considerar las funciones de demanda de distintos productos, o las funciones de producción de distintas industrias, con observaciones obtenidas a lo largo del tiempo (o a partir de varias unidades de corte transversal). Probablemente, el término de perturbación de la ecuación de demanda del bien A estará correlacionado con los términos de perturbación de las ecuaciones correspondientes a los bienes B , C , etc. Del mismo modo, la perturbación de la función de producción de una industria puede estar correlacionada con las de las funciones de producción de las demás industrias. Obsérvese que, si las ecuaciones de regresión (12.45) tienen una forma tal que los coeficientes de regresión de cada una de ellas son *iguales* a los de las demás, todo el sistema se reduciría a una sola ecuación y las observaciones se convertirían en datos conjuntos de series temporales y de corte transversal.

Estimación en el caso en que se conoce la matriz de variancias y covariancias

Pasemos ahora al problema de la estimación de los coeficientes de las ecuaciones de regresión aparentemente no relacionadas. Una posibilidad consiste en aplicar el método de los mínimos cuadrados ordinarios a cada ecuación por separado. En este caso, tendríamos

$$\hat{\beta}_m = (\mathbf{X}_m' \mathbf{X}_m)^{-1} (\mathbf{X}_m' \mathbf{Y}_m)$$

$$y \quad E(\hat{\beta}_m - \beta_m)(\hat{\beta}_m - \beta_m)' = \sigma_{mm} (\mathbf{X}_m' \mathbf{X}_m)^{-1}.$$

Obsérvese que $E(\hat{\beta}_m) = \beta_m + E(\mathbf{X}_m' \mathbf{X}_m)^{-1} (\mathbf{X}_m' \epsilon_m) = \beta_m$

$$y \quad \lim \text{prob} \quad \hat{\beta}_m = \beta_m + \lim \text{prob} \left[\frac{1}{T} \mathbf{X}_m' \mathbf{X}_m \right]^{-1} \left[\frac{1}{T} \mathbf{X}_m' \epsilon_m \right] = \beta_m,$$

lo que significa que los estimadores minimocuadráticos ordinarios de los coeficientes de regresión son insesgados y consistentes. Por tanto, el problema básico es el de la eficiencia. Al estimar cada ecuación por separado y de forma independiente, no utilizamos la información acerca de la correlación entre las perturbaciones y, por tanto, la eficiencia de los estimadores resulta dudosa.

Para tomar en consideración la correlación entre las perturbaciones de las distintas ecuaciones, tenemos que comprimir todo el modelo (12.45) en una sola ecuación. Esto se puede llevar a cabo observando que cada ecuación se puede escribir en la forma

$$(12.49) \quad \mathbf{Y}_m = \mathbf{X}_1^{(m)} \beta_1 + \mathbf{X}_2^{(m)} \beta_2 + \dots + \mathbf{X}_M^{(m)} \beta_M + \epsilon_m \quad (m = 1, 2, \dots, M),$$

en la que

$$\begin{aligned} \mathbf{X}_p^{(m)} &= \mathbf{X}_p \quad \text{si } m = p, \\ &= \mathbf{0} \quad \text{si } m \neq p, \end{aligned}$$

y $\mathbf{0}$ representa una matriz $(T \times K_p)$ de ceros. Redefiniendo las variables explicativas de esta forma, obtenemos un conjunto de ecuaciones de regresión en el que cada una de ellas contiene exactamente los mismos coeficientes de regresión que las demás. Por tanto, estamos formalmente en la misma situación que cuando tenemos observaciones conjuntas de series temporales y de corte transversal en una sola ecuación. Las ecuaciones (12.49) pueden representarse también en la forma

$$(12.49a) \quad \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^{(1)} & \mathbf{X}_2^{(1)} & \dots & \mathbf{X}_M^{(1)} \\ \mathbf{X}_1^{(2)} & \mathbf{X}_2^{(2)} & \dots & \mathbf{X}_M^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_1^{(M)} & \mathbf{X}_2^{(M)} & \dots & \mathbf{X}_M^{(M)} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix},$$

o

$$(12.49b) \quad \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix}.$$

Esto se puede escribir de forma más compacta como

$$(12.49c) \quad \mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

en la que $\mathbf{Y}$ es un vector $(MT \times 1)$, $\mathbf{X}$ es una matriz $(MT \times \sum_{m=1}^M K_m)$, β es una matriz $(\sum_{m=1}^M K_m \times 1)$, y ϵ es un vector $(MT \times 1)$. Obsérvese que, según los supuestos (12.46) a (12.48), la matriz de variancias y covariancias de ϵ es

$$(12.50) \quad \Omega = E(\epsilon\epsilon')$$

$$= \begin{bmatrix} E(\epsilon_1\epsilon_1') & E(\epsilon_1\epsilon_2') & \dots & E(\epsilon_1\epsilon_M') \\ E(\epsilon_2\epsilon_1') & E(\epsilon_2\epsilon_2') & \dots & E(\epsilon_2\epsilon_M') \\ \vdots & \vdots & \ddots & \vdots \\ E(\epsilon_M\epsilon_1') & E(\epsilon_M\epsilon_2') & \dots & E(\epsilon_M\epsilon_M') \end{bmatrix} = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \dots & \sigma_{1M}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \dots & \sigma_{2M}I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \dots & \sigma_{MM}I_T \end{bmatrix},$$

en la que I_T es una matriz identidad de orden $(T \times T)$ igual que en el caso anterior. En este caso, la información acerca de la correlación entre las perturbaciones de las distintas ecuaciones está contenida en la descripción de la matriz Ω y se tiene en cuenta en esta forma.

La ecuación (12.49c) más los supuestos acerca de X y ϵ se puede considerar como un modelo generalizado de regresión lineal (estudiado en la sección 12.1). En este modelo, el estimador lineal insesgado óptimo de β viene dado por la fórmula de los mínimos cuadrados generalizados de Aitken,

$$(12.51) \quad \tilde{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{Y}).$$

En el caso de las regresiones aparentemente no relacionadas, esto se convierte en

$$(12.51a) \quad \tilde{\beta} = \begin{bmatrix} \sigma^{11}(X_1'X_1) & \sigma^{12}(X_1'X_2) & \dots & \sigma^{1M}(X_1'X_M) \\ \sigma^{21}(X_2'X_1) & \sigma^{22}(X_2'X_2) & \dots & \sigma^{2M}(X_2'X_M) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1}(X_M'X_1) & \sigma^{M2}(X_M'X_2) & \dots & \sigma^{MM}(X_M'X_M) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{m=1}^M \sigma^{1m}(X_1'Y_m) \\ \sum_{m=1}^M \sigma^{2m}(X_2'Y_m) \\ \vdots \\ \sum_{m=1}^M \sigma^{Mm}(X_M'Y_m) \end{bmatrix},$$

en la que σ^{mp} representa el elemento correspondiente a la fila m y la columna p de la inversa de la matriz

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & \sigma_{MM} \end{bmatrix}.$$

Además, la matriz de variancias y covariancias de $\tilde{\beta}$ viene dada por

$$(12.52) \quad E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}.$$

Bajo el supuesto de normalidad, $\tilde{\beta}$ representa también el estimador maximo-verosímil de β . El estimador minimocuadrático ordinario de β es

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}),$$

que, en general, será distinto de β y, por tanto, la estimación minimocuadrática ordinaria de las regresiones aparentemente no relacionadas *no es eficiente*.

Por lo que respecta a la aplicación del estimador generalizado de Aitken a las regresiones aparentemente no relacionadas, se plantea el interesante problema de saber cuáles son las condiciones especiales en las que dicho estimador equivale al estimador minimocuadrático ordinario. Un caso obvio en el que sucede así es aquel en que las estimaciones están *realmente* no relacionadas; es decir, cuando

$$\sigma_{mp} = 0$$

para todo $m \neq p$. En este caso, $\tilde{\beta}$ se convierte en

$$\tilde{\beta} = \begin{bmatrix} \sigma^{11}(X_1'X_1) & 0 & \dots & 0 \\ 0 & \sigma^{22}(X_2'X_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^{MM}(X_M'X_M) \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}(X_1'Y_1) \\ \sigma^{22}(X_2'Y_2) \\ \vdots \\ \sigma^{MM}(X_M'Y_M) \end{bmatrix}$$

$$= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}),$$

que es la fórmula del estimador minimocuadrático. Los dos estimadores son también equivalentes en el caso en que todas las ecuaciones aparentemente no relacionadas incluyen *las mismas* variables explicativas; es decir, cuando

$$\mathbf{X}_m = \mathbf{X}_p$$

para todo m y p . A continuación, demostraremos la equivalencia entre la fórmula de los mínimos cuadrados ordinarios y la fórmula de Aitken en este caso, utilizando un conjunto de dos ecuaciones, pero esta demostración se puede ampliar a cualquier número de ecuaciones. (Además, la demostración se podría extender también al caso en que $\mathbf{X}_m$ y $\mathbf{X}_p$ no son exactamente iguales, sino que la primera es una combinación lineal de la segunda.) Cuando $M = 2$ y $\mathbf{X}_1 = \mathbf{X}_2$, el estimador generalizado de Aitken es

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sigma^{11}(X_1'X_1) & \sigma^{12}(X_1'X_1) \\ \sigma^{21}(X_1'X_1) & \sigma^{22}(X_1'X_1) \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11}(X_1'Y_1) + \sigma^{12}(X_1'Y_2) \\ \sigma^{21}(X_1'Y_1) + \sigma^{22}(X_1'Y_2) \end{bmatrix},$$

$$\begin{bmatrix} \sigma^{11}(X_1'X_1)\tilde{\beta}_1 + \sigma^{12}(X_1'X_1)\tilde{\beta}_2 \\ \sigma^{21}(X_1'X_1)\tilde{\beta}_1 + \sigma^{22}(X_1'X_1)\tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sigma^{11}(X_1'Y_1) + \sigma^{12}(X_1'Y_2) \\ \sigma^{21}(X_1'Y_1) + \sigma^{22}(X_1'Y_2) \end{bmatrix}.$$

Pero

$$\begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix},$$

o sea, que tenemos

$$\begin{aligned} \sigma_{22}(X_1'X_1)\tilde{\beta}_1 - \sigma_{12}(X_1'X_1)\tilde{\beta}_2 &= \sigma_{22}(X_1'Y_1) - \sigma_{12}(X_1'Y_2), \\ -\sigma_{12}(X_1'X_1)\tilde{\beta}_1 + \sigma_{11}(X_1'X_1)\tilde{\beta}_2 &= -\sigma_{12}(X_1'Y_1) + \sigma_{11}(X_1'Y_2). \end{aligned}$$

La solución para $\tilde{\beta}_1$ y $\tilde{\beta}_2$ es, pues,

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}(X_1'Y_1) \quad \text{y} \quad \tilde{\beta}_2 = (X_1'X_1)^{-1}(X_1'Y_2),$$

lo que coincide con los resultados de los mínimos cuadrados ordinarios.

Para ilustrar la diferencia entre los estimadores de Aitken y los estimadores mínimocuadráticos ordinarios de los coeficientes de las regresiones aparentemente no relacionadas, consideremos un sistema de dos ecuaciones. Supongamos que tenemos

$$\begin{aligned} (12.53) \quad Y_{1t} &= \beta_{11} + \beta_{12}X_{1t} + \epsilon_{1t}, \\ Y_{2t} &= \beta_{21} + \beta_{22}X_{2t} + \epsilon_{2t}, \end{aligned}$$

o

$$(12.53a) \quad Y_1 = X_1\beta_1 + \epsilon_1 \quad \text{y} \quad Y_2 = X_2\beta_2 + \epsilon_2,$$

$$\text{en las que} \quad X_r = \begin{bmatrix} 1 & X_{r1} \\ 1 & X_{r2} \\ \vdots & \vdots \\ 1 & X_{rT} \end{bmatrix}, \quad \beta_r = \begin{bmatrix} \beta_{r1} \\ \beta_{r2} \end{bmatrix} \quad (r = 1, 2).$$

En este caso,

$$\Omega^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22}I_T & -\sigma_{12}I_T \\ -\sigma_{12}I_T & \sigma_{11}I_T \end{bmatrix}.$$

El estimador generalizado de Aitken de los coeficientes de regresión es, pues,

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sigma_{22}(X_1'X_1) & -\sigma_{12}(X_1'X_2) \\ -\sigma_{12}(X_2'X_1) & \sigma_{11}(X_2'X_2) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{22}(X_1'Y_1) & -\sigma_{12}(X_1'Y_2) \\ -\sigma_{12}(X_2'Y_1) & \sigma_{11}(X_2'Y_2) \end{bmatrix}.$$

Esto da lugar al siguiente sistema de «ecuaciones normales minimocuadráticas generalizadas de Aitken»:

$$\begin{aligned} \sigma_{22} \sum_t (Y_{1t} - \beta_{11} - \beta_{12}X_{1t}) - \sigma_{12} \sum_t (Y_{2t} - \beta_{21} - \beta_{22}X_{2t}) &= 0, \\ \sigma_{22} \sum_t X_{1t}(Y_{1t} - \beta_{11} - \beta_{12}X_{1t}) - \sigma_{12} \sum_t X_{1t}(Y_{2t} - \beta_{21} - \beta_{22}X_{2t}) &= 0, \\ -\sigma_{12} \sum_t (Y_{1t} - \beta_{11} - \beta_{12}X_{1t}) + \sigma_{11} \sum_t (Y_{2t} - \beta_{21} - \beta_{22}X_{2t}) &= 0, \\ -\sigma_{12} \sum_t X_{2t}(Y_{1t} - \beta_{11} - \beta_{12}X_{1t}) + \sigma_{11} \sum_t X_{2t}(Y_{2t} - \beta_{21} - \beta_{22}X_{2t}) &= 0. \end{aligned}$$

En la primera y tercera ecuaciones podemos despejar $\tilde{\beta}_{11}$ y $\tilde{\beta}_{21}$ obteniendo

$$\begin{aligned} (12.54) \quad \tilde{\beta}_{11} &= \bar{Y}_1 - \tilde{\beta}_{12}\bar{X}_1, \\ \tilde{\beta}_{21} &= \bar{Y}_2 - \tilde{\beta}_{22}\bar{X}_2. \end{aligned}$$

Sustituyendo estos resultados en las ecuaciones restantes, y despejando $\tilde{\beta}_{12}$ y $\tilde{\beta}_{22}$, obtenemos

$$\begin{aligned} (12.55) \quad \tilde{\beta}_{12} &= \frac{\sigma_{11}m_{X_2X_2}(\sigma_{22}m_{X_1Y_1} - \sigma_{12}m_{X_1Y_2}) + \sigma_{12}m_{X_1X_2}(\sigma_{11}m_{X_2Y_2} - \sigma_{12}m_{X_2Y_1})}{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2X_2} - \sigma_{12}^2m_{X_1X_2}^2}, \\ \tilde{\beta}_{22} &= \frac{\sigma_{12}m_{X_1X_2}(\sigma_{22}m_{X_1Y_1} - \sigma_{12}m_{X_1Y_2}) + \sigma_{22}m_{X_1X_1}(\sigma_{11}m_{X_2Y_2} - \sigma_{12}m_{X_2Y_1})}{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2X_2} - \sigma_{12}^2m_{X_1X_2}^2}, \end{aligned}$$

$$\text{en las que} \quad m_{X_rX_s} = \sum_t (X_{rt} - \bar{X}_r)(X_{st} - \bar{X}_s),$$

$$m_{X_rY_s} = \sum_t (X_{rt} - \bar{X}_r)(Y_{st} - \bar{Y}_s),$$

$$r, s = 1, 2.$$

Obsérvese que, si ϵ_{1t} y ϵ_{2t} no están correlacionadas (es decir, si $\sigma_{12} = 0$), las fórmulas de $\tilde{\beta}_{12}$ y $\tilde{\beta}_{22}$ se convierten en

$$\begin{aligned} \tilde{\beta}_{12} &= \frac{\sigma_{11}\sigma_{22}m_{X_2X_2}m_{X_1Y_1}}{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2X_2}} = \frac{m_{X_1Y_1}}{m_{X_1X_1}}, \\ \tilde{\beta}_{22} &= \frac{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2Y_2}}{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2X_2}} = \frac{m_{X_2Y_2}}{m_{X_2X_2}}. \end{aligned}$$

que es el resultado correspondiente a los estimadores minimocuadráticos ordinarios. Del mismo modo, si $X_{1t} = X_{2t}$ (es decir, si las dos ecuaciones tienen la misma variable explicativa), obtenemos

$$\begin{aligned}\beta_{12} &= \frac{\sigma_{11}m_{X_1X_1}(\sigma_{22}m_{X_1Y_1} - \sigma_{12}m_{X_1Y_2}) + \sigma_{12}m_{X_1X_1}(\sigma_{11}m_{X_1Y_2} - \sigma_{12}m_{X_1Y_1})}{\sigma_{11}\sigma_{22}m_{X_1X_1}^2 - \sigma_{12}^2m_{X_1X_1}^2} \\ &= \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)m_{X_1X_1}m_{X_1Y_1}}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)m_{X_1X_1}^2} = \frac{m_{X_1Y_1}}{m_{X_1X_1}}, \\ \beta_{22} &= \frac{\sigma_{12}m_{X_1X_1}(\sigma_{22}m_{X_1Y_1} - \sigma_{12}m_{X_1Y_2}) + \sigma_{22}m_{X_1X_1}(\sigma_{11}m_{X_1Y_2} - \sigma_{12}m_{X_1Y_1})}{\sigma_{11}\sigma_{22}m_{X_1X_1}^2 - \sigma_{12}^2m_{X_1X_1}^2} \\ &= \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)m_{X_1X_1}m_{X_1Y_2}}{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)m_{X_1X_1}^2} = \frac{m_{X_1Y_2}}{m_{X_1X_1}},\end{aligned}$$

que corresponden nuevamente a las fórmulas minimocuadráticas ordinarias.

Como los estimadores minimocuadráticos generalizados son ELIO, su variancia es menor o igual que la de los estimadores minimocuadráticos ordinarios. Por tanto, resulta adecuado preguntarnos cuál es la diferencia entre las variancias y de qué depende. Obtendremos una respuesta en base al sistema de dos regresiones aparentemente no relacionadas que acabamos de presentar. Señalemos, en primer lugar, que la matriz de variancias y covariancias de los estimadores minimocuadráticos generalizados de β_{12} y β_{22} es

$$E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}.$$

En este caso,

$$(\tilde{\beta} - \beta) = \begin{bmatrix} \tilde{\beta}_{12} - \beta_{12} \\ \tilde{\beta}_{22} - \beta_{22} \end{bmatrix} \quad \text{y} \quad \mathbf{X} = \begin{bmatrix} (X_{11} - \bar{X}_1) & 0 \\ (X_{12} - \bar{X}_1) & 0 \\ \vdots & \vdots \\ (X_{1T} - \bar{X}_1) & 0 \\ 0 & (X_{21} - \bar{X}_2) \\ 0 & (X_{22} - \bar{X}_2) \\ \vdots & \vdots \\ 0 & (X_{2T} - \bar{X}_2) \end{bmatrix}$$

Esto significa que

$$\begin{bmatrix} \text{Var}(\tilde{\beta}_{12}) & \text{Cov}(\tilde{\beta}_{12}, \tilde{\beta}_{22}) \\ \text{Cov}(\tilde{\beta}_{12}, \tilde{\beta}_{22}) & \text{Var}(\tilde{\beta}_{22}) \end{bmatrix} = (\sigma_{11}\sigma_{22} - \sigma_{12}^2) \begin{bmatrix} \sigma_{22}m_{X_1X_1} & -\sigma_{12}m_{X_1X_2} \\ -\sigma_{12}m_{X_1X_2} & \sigma_{11}m_{X_2X_2} \end{bmatrix}^{-1}$$

En concreto, la variancia del estimador minimocuadrático generalizado de β_{12} es

$$(12.56) \quad \text{Var}(\tilde{\beta}_{12}) = \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\sigma_{11}m_{X_2X_2}}{\sigma_{11}\sigma_{22}m_{X_1X_1}m_{X_2X_2} - \sigma_{12}^2m_{X_1X_2}^2}.$$

Si representamos por

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

el coeficiente de correlación entre ε_{1t} y ε_{2t} , y por

$$r_{12} = \frac{m_{X_1X_2}}{\sqrt{m_{X_1X_1}m_{X_2X_2}}}$$

el coeficiente de correlación entre X_{1t} y X_{2t} en la muestra, podemos escribir

$$\text{Var}(\tilde{\beta}_{12}) = \frac{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)\sigma_{11}m_{X_2X_2}}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2r_{12}^2)m_{X_1X_1}m_{X_2X_2}} = \frac{\sigma_{11}}{m_{X_1X_1}} \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2r_{12}^2}.$$

Por otra parte, la variancia del estimador minimocuadrático ordinario de β_{12} es

$$\text{Var}(\beta_{12}) = \frac{\sigma_{11}}{m_{X_1X_1}}.$$

Por tanto, la razón entre las dos variancias es

$$(12.57) \quad \frac{\text{Var}(\tilde{\beta}_{12})}{\text{Var}(\beta_{12})} = \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2r_{12}^2} \leq 1.$$

Esta razón es función decreciente de ρ_{12}^2 y función creciente de r_{12}^2 . Esto significa que el aumento de eficiencia debido al uso del estimador minimocuadrático generalizado en lugar del estimador minimocuadrático ordinario es máximo cuando las perturbaciones de las dos ecuaciones están muy correlacionadas y, al mismo tiempo, las variables explicativas no están correlacionadas. El cuadro 12-1 presenta los valores de la razón $\text{Var}(\tilde{\beta}_{12})/\text{Var}(\beta_{12})$ correspondientes a distintos valores de ρ_{12}^2 y r_{12}^2 . Como puede verse claramente, en muchos casos el aumento de eficiencia resulta sustancial. Estos resultados podrían extenderse también a un sistema más complejo de ecuaciones aparentemente no relacionadas.

Cuadro 12-1

r_{12}^2	ρ_{12}^2					
	0	0.1	0.3	0.5	0.7	0.9
0	1.000	0.900	0.700	0.500	0.300	0.100
0.1	1.000	0.909	0.722	0.526	0.323	0.110
0.3	1.000	0.928	0.769	0.588	0.380	0.137
0.5	1.000	0.947	0.823	0.667	0.461	0.182
0.7	1.000	0.967	0.886	0.769	0.588	0.270
0.9	1.000	0.989	0.959	0.909	0.811	0.526
1.0	1.000	1.000	1.000	1.000	1.000	1.000

Estimación en el caso en que no se conoce la matriz de variancias y covariancias

En el análisis anterior hemos supuesto que se conocían los elementos de la matriz Ω (es decir, las variancias y covariancias de las perturbaciones). Sin embargo, si (como sucede generalmente) no se conocen, podemos sustituir Ω por un estimador consistente de dicha matriz. Como ya se ha señalado en la sección 12.1, el estimador de β que obtenemos en este caso posee las mismas propiedades asintóticas que el estimador minimocuadrático generalizado de Aitken. El problema consiste, pues, en obtener estimaciones consistentes de las variancias y covariancias de las perturbaciones de la regresión, y para ello podemos utilizar los residuos minimocuadráticos ordinarios (a los que llamaremos e_{mt}) como ha propuesto Zellner en su pionero artículo.¹¹ Utilizaremos

$$(12.58) \quad \hat{\Omega} = \begin{bmatrix} s_{11}I_T & s_{12}I_T & \dots & s_{1M}I_T \\ s_{21}I_T & s_{22}I_T & \dots & s_{2M}I_T \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1}I_T & s_{M2}I_T & \dots & s_{MM}I_T \end{bmatrix},$$

es decir, que

$$s_{mp} = \frac{1}{T - K_m} \sum_{t=1}^T e_{mt}e_{pt},$$

$$K_m \geq K_p,$$

$$m, p = 1, 2, \dots, M.$$

Sabemos que s_{mm} es un estimador insesgado y consistente de σ_{mm} , y se puede mostrar que s_{mp} ($m \neq p$) es un estimador consistente de σ_{mp} . (Como sólo nos interesa la consistencia, al calcular las estimaciones de σ_{mi} podríamos utilizar T en lugar de $(T - K_m)$, sin que ello afecte a las propiedades asintóticas del estimador de β .) El estimador de β resultante es

$$(12.59) \quad \tilde{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y)$$

con

$$(12.60) \quad \text{Var-Cov asintótica } (\tilde{\beta}) = (X'\hat{\Omega}^{-1}X)^{-1}$$

y se denomina *estimador de Aitken en dos etapas*, ya que su valor se calcula por un procedimiento en dos etapas. En primer lugar se obtienen las estimaciones minimocuadráticas ordinarias de cada ecuación, utilizando los residuos resultantes para estimar las variancias y covariancias de las perturbaciones y construir $\hat{\Omega}$. La segunda etapa consiste en sustituir Ω por $\hat{\Omega}$ en (12.59) y calcular los valores de los elementos de $\tilde{\beta}$. (Existe un programa para calculadoras, correspondiente a este método de estimación, llamado EFFST.) El estimador de Aitken en dos etapas es asintóticamente equivalente al estimador minimocuadrático generalizado de Aitken y, por tanto, al estimador maximo-verosímil de β . Por lo tanto, este estimador es asintóticamente eficiente y su distribución asintótica es normal. Por lo que respecta a las propiedades de dicho estimador en las muestras pequeñas, se dispone de algunos resultados experimentales y teóricos que indican que es insesgado y eficiente en relación al estimador minimocuadrático ordinario.¹²

Un procedimiento alternativo para estimar los elementos de Ω consiste en utilizar el método de la máxima verosimilitud. Según (12.18), la función de verosimilitud de Y en el contexto de las regresiones aparentemente no relacionadas es

$$(12.61) \quad L = -\frac{MT}{2} \log(2\pi) - \frac{1}{2} \log |\Omega| - \frac{1}{2}(Y - X\beta)'\Omega^{-1}(Y - X\beta).$$

12. Ver A. Zellner, «Estimators of Seemingly Unrelated Equations: Some Exact Finite Sample Results», *Journal of the American Statistical Association*, vol. 58, diciembre 1963, pp. 977-92; N. C. Kakwani, «The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators», *Journal of the American Statistical Association*, vol. 62, marzo 1967, pp. 141-42; J. Kmenta y R. F. Gilbert, «Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions», *Journal of the American Statistical Association*, vol. 63, diciembre 1968, pp. 1180-200.

11. Zellner, *op. cit.*

Podemos derivar L respecto a los elementos de β y Ω , igualar a cero estas derivadas y resolver las ecuaciones resultantes obteniendo los valores de los parámetros desconocidos. En el sistema de dos regresiones aparentemente no relacionadas que hemos presentado en (12.53), tenemos

$$L = -\frac{2T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_{11}\sigma_{22} - \sigma_{12}^2) - \frac{1}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} [\sigma_{22} \sum_i (Y_{1i} - \beta_{11} - \beta_{12}X_{1i})^2 + \sigma_{11} \sum_i (Y_{2i} - \beta_{21} - \beta_{22}X_{2i})^2 - 2\sigma_{12} \sum_i (Y_{1i} - \beta_{11} - \beta_{12}X_{1i})(Y_{2i} - \beta_{21} - \beta_{22}X_{2i})].$$

Las condiciones de primer grado para la maximización de L respecto a las β y las σ son

$$\begin{aligned} \hat{\sigma}_{22} \sum_i (Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i}) - \hat{\sigma}_{12} \sum_i (Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i}) &= 0, \\ \hat{\sigma}_{22} \sum_i X_{1i}(Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i}) - \hat{\sigma}_{12} \sum_i X_{1i}(Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i}) &= 0, \\ -\hat{\sigma}_{12} \sum_i (Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i}) + \hat{\sigma}_{11} \sum_i (Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i}) &= 0, \\ -\hat{\sigma}_{12} \sum_i X_{2i}(Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i}) + \hat{\sigma}_{11} \sum_i X_{2i}(Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i}) &= 0, \\ \hat{\sigma}_{11} &= \frac{1}{T} \sum_i (Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i})^2, \\ \hat{\sigma}_{22} &= \frac{1}{T} \sum_i (Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i})^2, \\ \hat{\sigma}_{12} &= \frac{1}{T} \sum_i (Y_{1i} - \hat{\beta}_{11} - \hat{\beta}_{12}X_{1i})(Y_{2i} - \hat{\beta}_{21} - \hat{\beta}_{22}X_{2i}). \end{aligned}$$

Obsérvese que las cuatro primeras ecuaciones son básicamente iguales que las «ecuaciones normales minimocuadráticas generalizadas de Aitken» que hemos obtenido anteriormente; la única diferencia es que las σ se han sustituido por $\hat{\sigma}$. En el caso del estimador de Aitken en dos etapas, las $\hat{\sigma}$ de las cuatro primeras ecuaciones se sustituyen por las estimaciones obtenidas a partir de los residuos minimocuadráticos ordinarios y, por tanto, las tres últimas ecuaciones no se utilizan. Sin embargo, en el caso del estimador maximoverosímil, las σ deben estimarse conjuntamente con las β utilizando las siete ecuaciones.

Las condiciones de primer grado para la maximización de L en un sistema más complejo de regresiones aparentemente no relacionadas pueden obtenerse mediante una simple generalización del caso de dos ecuaciones. En primer lugar, tenemos las ecuaciones correspondientes a las «ecuaciones normales minimocuadráticas generalizadas de Aitken» y, en segundo lugar, las expresiones correspondientes a las variancias y covariancias estimadas de las perturbaciones. Así pues, en el caso general los estimadores maximoverosímiles vienen dados por

$$(12.62) \quad \hat{\beta} = (X' \hat{\Omega}^{-1} X)^{-1} (X' \hat{\Omega}^{-1} Y),$$

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{11} I_T & \hat{\sigma}_{12} I_T & \dots & \hat{\sigma}_{1M} I_T \\ \hat{\sigma}_{21} I_T & \hat{\sigma}_{22} I_T & \dots & \hat{\sigma}_{2M} I_T \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{M1} I_T & \hat{\sigma}_{M2} I_T & \dots & \hat{\sigma}_{MM} I_T \end{bmatrix},$$

siendo

$$\hat{\sigma}_{mp} = \frac{1}{T} (Y_m - X_m \hat{\beta}_m)' (Y_p - X_p \hat{\beta}_p),$$

$m, p = 1, 2, \dots, M.$

La solución analítica de (12.62) resulta muy complicada, ya que las ecuaciones son no lineales respecto a las incógnitas, pero puede obtenerse con la ayuda de una calculadora electrónica. El estimador resultante $\hat{\beta}$ posee las mismas propiedades asintóticas que el estimador de Aitken en dos etapas, y existe también cierta evidencia de que las propiedades de estos dos estimadores en las muestras pequeñas son muy similares.¹³

EJEMPLO. Como ejemplo de regresiones aparentemente no relacionadas, consideremos el siguiente conjunto de funciones de inversión de empresas individuales:

$$I_i = \beta_{m1} + \beta_{m2} C_{i-1} + \beta_{m3} F_{i-1} + \epsilon_{mi}$$

en las que I = inversión bruta, C = cantidad de capital al final del período, y F = valor de las acciones en circulación al final del período. Las estimaciones si-

13. Ver Kmenta y Gilbert, *op. cit.*

guientes correspondientes a dos empresas, General Electric y Westinghouse, están basadas en los datos anuales del periodo 1935-1954¹⁴:

General Electric

$$\text{MCO} \quad I_t = -9.9563 + 0.1517C_{t-1} + 0.0266F_{t-1} + e_t,$$

(31.3742) (0.0257) (0.0156)

$$\text{EFFEST} \quad I_t = -27.7193 + 0.1390C_{t-1} + 0.0383F_{t-1} + e_t,$$

(29.3212) (0.0250) (0.0145)

$$\text{MV} \quad I_t = -30.7485 + 0.1359C_{t-1} + 0.0405F_{t-1} + e_t;$$

(29.6608) (0.0255) (0.0145)

Westinghouse

$$\text{OLS} \quad I_t = -0.5094 + 0.0924C_{t-1} + 0.0529F_{t-1} + e_t,$$

(8.0153) (0.0561) (0.0157)

$$\text{EFFEST} \quad I_t = -1.2520 + 0.0576C_{t-1} + 0.0640F_{t-1} + e_t,$$

(7.5452) (0.0530) (0.0145)

$$\text{MV} \quad I_t = -1.7016 + 0.0593C_{t-1} + 0.0557F_{t-1} + e_t,$$

(7.5149) (0.0529) (0.0144)

Como puede verse, parece existir un aumento de eficiencia, aunque no muy importante, al pasar de la estimación por mínimos cuadrados ordinarios a la de dos etapas de Aitken o a la de la máxima verosimilitud. La razón de que este aumento de eficiencia sea relativamente pequeño es, al menos en parte, el alto grado de correlación existente entre las variables explicativas de las dos ecuaciones.

Estimación de regresiones aparentemente no relacionadas con perturbaciones autorregresivas

Como último punto en relación con el problema de la estimación de regresiones aparentemente no relacionadas, consideraremos el caso en que las perturbaciones de cada ecuación no son independientes *a lo largo del tiempo*, sino que siguen un esquema autorregresivo de primer orden del tipo presentado en la

sección 8.2. En este caso, los supuestos (12.47) y (12.48) efectuados al iniciar esta sección deben sustituirse por

$$(12.63) \quad E(\epsilon_m \epsilon_m') = \sigma_{mm} \begin{bmatrix} 1 & \rho_m & \dots & \rho_m^{T-1} \\ \rho_m & 1 & \dots & \rho_m^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_m^{T-1} & \rho_m^{T-2} & \dots & 1 \end{bmatrix},$$

$$(12.64) \quad E(\epsilon_m \epsilon_p') = \sigma_{mp} \begin{bmatrix} 1 & \rho_p & \dots & \rho_p^{T-1} \\ \rho_p & 1 & \dots & \rho_p^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_p^{T-1} & \rho_p^{T-2} & \dots & 1 \end{bmatrix},$$

$$m, p = 1, 2, \dots, M,$$

en los que ρ_m es el coeficiente de autocorrelación en la ecuación m . En tal caso, podemos estimar ρ_m por separado para cada ecuación utilizando alguno de los métodos consistentes presentados en la sección 8.2. Supongamos que llamamos $\hat{\rho}_m$ a las estimaciones resultantes. Podemos utilizar estas estimaciones para transformar las observaciones individuales de modo que el sistema de regresiones aparentemente no relacionadas se convierte ahora en

$$(12.65) \quad (Y_{mt} - \hat{\rho}_m Y_{m,t-1}) = \beta_{m1}(X_{1t,1} - \hat{\rho}_m X_{1,t-1,1}) \\ + \beta_{m2}(X_{1t,2} - \hat{\rho}_m X_{1,t-1,2}) \\ + \dots + \beta_{mK_m}(X_{mt,K_m} - \hat{\rho}_m X_{m,t-1,K_m}) + u_{mt}, \\ m = 1, 2, \dots, M.$$

Las regresiones de (12.65) pueden estimarse en la forma habitual por el método de Aitken en dos etapas y las estimaciones resultantes de los coeficientes de regresión poseen las mismas propiedades asintóticas que las estimaciones minimocuadráticas generalizadas de Aitken.¹⁵

14. En *ibid.* Los resultados se basan en los datos que aparecen en J. C. G. Root y G. M. deWitt, «Investment Demand: An Empirical Contribution to the Aggregation Problem», *International Economic Review*, vol. 1, enero 1960, pp. 3-30.

15. Ver R. W. Parks, «Efficient Estimation of A System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated», *Journal of the American Statistical Association*, vol. 62, junio 1967, pp. 500-509; J. Kmenta y R. F. Gilbert, «Estimation of Seemingly Unrelated Regressions with Autorregressive Disturbances», *Journal of the American Statistical Association*, vol. 65, marzo 1970, pp. 186-97.

EJERCICIOS:

- 12-1.** Demostrar la validez de las ecuaciones (12.16) y (12.17).
- 12-2.** Considerar la estimación de una ecuación de regresión con observaciones procedentes de N economías domésticas en dos periodos de tiempo consecutivos, suponiendo que las perturbaciones de la regresión no están correlacionadas en sentido transversal pero son autorregresivas a lo largo del tiempo, con $\rho_i = \rho_j$ para todo i, j .
- Obtener los elementos de Ω (la matriz de variancias y covariancias de las perturbaciones).
 - Construir un procedimiento de estimación que proporcione un estimador consistente de Ω .
 - Diseñar una transformación de las observaciones que permita utilizar el método de los mínimos cuadrados ordinarios para obtener estimaciones asintóticamente eficientes de los coeficientes de regresión.
- 12-3.** Un modelo de regresión que se debe estimar a partir de observaciones conjuntas de series temporales y de corte transversal viene dado por

$$Y_{it} = \beta X_{it} + \varepsilon_{it} \quad (t = 1, 2, \dots, 21; i = 1, 2).$$

De modo alternativo, podemos escribir

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

en la que $Y_i \rightarrow (21 \times 1)$, $X_i \rightarrow (21 \times 1)$, $\varepsilon_i \rightarrow (21 \times 1)$, y β es un escalar. Supóngase que las X son no estocásticas y están limitadas, y que

$$E(\varepsilon_{it}) = 0,$$

$$E(\varepsilon_{it}\varepsilon_{is}) = 0 \quad (t \neq s),$$

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}.$$

Los datos muestrales presentan la forma siguiente:

$$\begin{array}{lll} X_1X_1 = 10, & X_1Y_1 = 10, & Y_1Y_1 = 13.90, \\ X_1X_2 = 8, & X_1Y_2 = 8, & Y_1Y_2 = 11.92, \\ X_2X_2 = 10, & X_2Y_1 = 8, & Y_2Y_2 = 12.30, \\ & X_2Y_2 = 8, & \end{array}$$

Obtener una estimación asintóticamente eficiente de β y de su error standard estimado.

- 12-4.** Considerar las dos ecuaciones de regresión siguientes:

$$\begin{aligned} Y_{1t} &= \beta_1 X_{1t} + \varepsilon_{1t} \\ Y_{2t} &= \beta_2 X_{2t} + \varepsilon_{2t} \end{aligned} \quad (t = 1, 2, \dots, 21).$$

Supóngase que las X son no estocásticas y están limitadas, y que

$$\varepsilon_{it} \sim N(0, \sigma_{it}) \quad (i = 1, 2),$$

$$E(\varepsilon_{it}\varepsilon_{is}) = 0 \quad (t \neq s),$$

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij} \quad (i, j = 1, 2).$$

Los resultados muestrales son

$$\begin{aligned} \sum X_{1t}^2 &= 10, & \sum X_{1t}Y_{1t} &= 10, & \sum Y_{1t}^2 &= 12.0, \\ \sum X_{2t}^2 &= 10, & \sum X_{1t}Y_{2t} &= 8, & \sum Y_{2t}^2 &= 12.4, \\ \sum X_{1t}X_{2t} &= 8, & \sum X_{2t}Y_{1t} &= 8, & \sum Y_{1t}Y_{2t} &= 10.0, \\ & & \sum X_{2t}Y_{2t} &= 8, & & \end{aligned}$$

Obtener estimaciones asintóticamente eficientes de los coeficientes de regresión y de sus errores standard estimados, utilizando:

- El método de Aitken en dos etapas, propuesto por Zellner.
- El método de la máxima verosimilitud.

- 12-5.** Se ha especificado un conjunto de tres ecuaciones de regresión aparentemente no relacionadas, en la forma

$$Y_{1t} = \alpha_1 + \beta_1 X_{1t} + \varepsilon_{1t},$$

$$Y_{12} = \alpha_2 + \beta_2 X_{12} + \varepsilon_{12},$$

$$Y_{13} = \alpha_3 + \beta_3 X_{13} + \varepsilon_{13},$$

en la que la matriz de variancias y covariancias de las perturbaciones se supone conocida. Considerar dos estimadores de β_1 , el primero obtenido aplicando la estimación minimocuadrática generalizada de Aitken a las tres ecuaciones, y el segundo aplicando este método a las dos primeras ecuaciones solamente, y examinar la eficiencia del primer estimador respecto al segundo.

- 12-6.** Considérese el siguiente conjunto de ecuaciones de demanda de distintos productos, para cuya estimación se dispone de datos de series temporales:

$$V_{it} = \alpha_i + \beta_i P_{it} + \gamma_i V_t + \varepsilon_{it}$$

en las que V_{it} = gasto en el producto i , P_{it} = precio del producto i , y V_t = gasto total en todos los productos. Como

$$\sum_i V_{it} = V_t,$$

existen las siguientes restricciones:

$$\sum_i \alpha_i = 0, \quad \sum_i \beta_i = 0, \quad \sum_i \gamma_i = 1, \quad \sum_i \epsilon_{it} = 0.$$

Describir un procedimiento de estimación que produzca estimaciones eficientes de los coeficientes de regresión.

Sistema de ecuaciones simultáneas

Los modelos económicos comprenden generalmente un *conjunto* de relaciones que tratan de explicar la conducta de ciertas variables. Por ejemplo, un modelo simple del mercado de un producto determinado puede estar formado por una función de oferta y una de demanda, y tratar de explicar el precio y la cantidad de equilibrio de la mercancía que se intercambia en el mercado. Del mismo modo, un modelo de la renta agregada puede explicar la determinación de los distintos componentes de la renta a través de relaciones especificadas en la forma adecuada. En estos modelos, el problema de la estimación de los parámetros presenta unas características especiales que no aparecen en los modelos compuestos por una sola relación. En particular, cuando una relación forma parte de un sistema, algunos regresores suelen ser estocásticos y están correlacionados con la perturbación de la regresión. En este caso, los estimadores minimocuadráticos ordinarios de los coeficientes de regresión son inconsistentes, y es preciso construir otros métodos que proporcionen estimaciones consistentes.

Este capítulo se ocupa del problema de la estimación de ecuaciones que pertenecen a un sistema de relaciones y del análisis e interpretación de tales sistemas. La sección 13.1 contiene una descripción general de los sistemas de ecuaciones simultáneas e introduce algunos conceptos básicos. La sección 13.2 se ocupa del problema de la identificación, que es de importancia crucial para la estimación, y el análisis representa una extensión lógica del problema de la identificación en los modelos uniecuacionales, estudiado en la sección 11.2, a los modelos multiecuacionales de este capítulo. En la sección 13.3 se presentan varios métodos de estimación de una sola ecuación que pertenece a un sistema de ecuaciones, mientras que los métodos presentados en la sección 13.4 están pensados para obtener estimaciones de *todas* las ecuaciones del sistema. La sección 13.5 se ocupa de la comparación entre los distintos métodos de estimación y de algunos problemas especiales. Por último, la sección 13.6 presenta un análisis e interpretación de los sistemas económicos dinámicos.

13.1 Descripción de los sistemas de ecuaciones simultáneas

El requisito básico que deben satisfacer los modelos económicos es que el número de variables cuyos valores se pretende explicar sea igual al número de relaciones independientes en el modelo (es decir, al número de piezas de información relevante distintas), ya que, de lo contrario, los valores de estas variables no serían determinados. Además de las variables cuyos valores se pretende explicar, los modelos pueden contener (y generalmente contienen) variables cuyos valores no se ven afectados de forma inmediata por el mecanismo presentado en el modelo. La importancia de estas variables radica en su función como factores explicativos. Esto nos lleva a efectuar una distinción entre las variables cuyos valores deben ser explicados por el modelo y las variables que contribuyen a proporcionar esta explicación; las primeras se denominan *endógenas* y las segundas, *predeterminadas*. Las variables predeterminadas pueden subdividirse en *exógenas* y *endógenas retardadas*. Los valores de las variables exógenas se determinan por completo al margen del sistema que se esté considerando, mientras que los valores de las variables endógenas retardadas están representados por los valores previos de las variables endógenas del modelo. Los modelos sin variables endógenas retardadas son relativamente corrientes, pero es muy poco frecuente encontrar modelos sin ninguna variable predeterminada. Por ejemplo, un modelo del mercado de un producto determinado puede estar formado por una relación de demanda y una de oferta, en las que la cantidad intercambiada y el precio de equilibrio del período sean las variables endógenas, y los factores que originan desplazamientos sistemáticos de las funciones de oferta y demanda, las variables predeterminadas. Se dice que un modelo constituye un sistema de ecuaciones simultáneas si, para obtener el valor de una o más de las variables endógenas del modelo, se necesitan todas las relaciones que lo componen. Esto implica que una o más de las relaciones incluyen más de una variable endógena.

La definición de sistema de ecuaciones simultáneas se puede interpretar de forma más rigurosa cuando los modelos económicos se formulan específicamente como un conjunto de relaciones estocásticas bien definidas, es decir, cuando se convierten en lo que generalmente se denominan *modelos econométricos*. Normalmente, la teoría económica nos dice qué relaciones forman el modelo, las variables que deben incluirse en cada una de las relaciones y los signos de algunas de las derivadas parciales. Pero, por regla general, la teoría económica apenas nos informa acerca de la forma funcional de las relaciones, los retardos temporales existentes y los valores de los parámetros. Además, las relaciones son de tipo determinístico, de modo que no se considera la existencia de perturbaciones estocásticas. Para convertir un modelo económico en una proposición contrastable es preciso especificar la forma funcional de las relaciones, los períodos de tiempo a los que se refieren las variables y la naturaleza estocástica del sistema. El resultado consiste en un modelo econ-

métrico que es susceptible de estimación o contrastación. Este modelo representa un resumen del conocimiento previo del investigador acerca del fenómeno en cuestión. En el estado actual de la ciencia económica, estos conocimientos previos proceden en parte de la teoría económica, y en parte de conjeturas y razonamientos *ad hoc*. Por ejemplo, consideremos el siguiente modelo simplificado de oferta y demanda de un producto:

$$\begin{aligned} \text{(demanda)} \quad Q &= f_1(P, Y) & \frac{\partial f_1}{\partial P} &\leq 0, & \frac{\partial f_1}{\partial Y} &\geq 0, & \frac{\partial f_2}{\partial P} &\geq 0, \\ \text{(oferta),} \quad Q &= f_2(P) \end{aligned}$$

en el que Q = cantidad de equilibrio comprada y vendida en el mercado, p = precio de equilibrio, e Y = renta de los consumidores. Las variables Q y p son endógenas, mientras que Y es exógena. Obsérvese que para obtener los valores de las dos variables endógenas necesitamos conocer las dos relaciones, de modo que se trata de un sistema de ecuaciones simultáneas. Estas relaciones se podrían representar a través de un modelo econométrico del tipo siguiente:

$$\begin{aligned} (13.1) \quad \text{(demanda),} \quad Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{1t} \\ \text{(oferta),} \quad Q_t &= \beta_1 + \beta_2 P_t + \varepsilon_{2t} \\ \alpha_2 &\leq 0, & \alpha_3 &\geq 0, & \beta_2 &\geq 0, \end{aligned}$$

en el que las α y β son parámetros, las ε son perturbaciones aleatorias, y t representa un período de tiempo determinado. Cada término de perturbación está caracterizado por los supuestos del modelo clásico de regresión normal lineal. Las variancias y la covariancia de las perturbaciones son

$$\begin{aligned} E(\varepsilon_{1t}^2) &= \sigma_{11}, \\ E(\varepsilon_{2t}^2) &= \sigma_{22}, \\ E(\varepsilon_{1t}\varepsilon_{2t}) &= \sigma_{12}. \end{aligned}$$

Las ecuaciones (13.1) se denominan *forma estructural* del modelo y se obtienen a partir de la teoría económica. Si en las ecuaciones estructurales despejamos las variables endógenas, obtenemos

$$\begin{aligned} (13.2) \quad Q_t &= \left(\frac{\alpha_2 \beta_1 - \alpha_1 \beta_2}{\alpha_2 - \beta_2} \right) - \left(\frac{\alpha_3 \beta_2}{\alpha_2 - \beta_2} \right) Y_t + \left(\frac{-\beta_2 \varepsilon_{1t} + \alpha_2 \varepsilon_{2t}}{\alpha_2 - \beta_2} \right) \\ P_t &= \left(\frac{-\alpha_1 + \beta_1}{\alpha_2 - \beta_2} \right) - \left(\frac{\alpha_3}{\alpha_2 - \beta_2} \right) Y_t + \left(\frac{-\varepsilon_{1t} + \varepsilon_{2t}}{\alpha_2 - \beta_2} \right). \end{aligned}$$

Las ecuaciones (13.2) constituyen la *forma reducida* del modelo. Las ecuaciones de la forma reducida muestran explícitamente que las variables endógenas *dependen conjuntamente* de las variables predeterminadas y las perturbaciones del sistema. En el caso de (13.2) podemos ver que los valores de Q_t y P_t están totalmente determinados por Y_t , ϵ_{1t} y ϵ_{2t} . Por otra parte, se considera que el valor de Y_t se determina fuera del mercado considerado y sin ninguna influencia procedente de P_t o Q_t . Si Y_t es aleatoria, se supone que se distribuye independientemente de ϵ_{1t} y ϵ_{2t} .

Desde el punto de vista de la inferencia estadística, la única característica significativa de los sistemas de ecuaciones simultáneas (y la única que requiere consideración especial) es la presencia de variables endógenas entre las variables explicativas de algunas o todas las ecuaciones estructurales. Esta característica causa problemas porque, en general, las variables endógenas están correlacionadas con el término de perturbación de la ecuación en la que aparecen. Consideremos, por ejemplo, el modelo de oferta y demanda de (13.1). La variable endógena P_t aparece como variable explicativa en las dos ecuaciones, pero, según (13.2), podemos ver que

$$E(P_t \epsilon_{1t}) = \frac{-\sigma_{11} + \sigma_{12}}{\alpha_2 - \beta_2}$$

y

$$E(P_t \epsilon_{2t}) = \frac{-\sigma_{12} + \sigma_{22}}{\alpha_2 - \beta_2},$$

lo que indica que dicha variable está correlacionada con las dos perturbaciones. Como ya se ha señalado en la sección 8.3, la existencia de correlación entre una variable explicativa y el término de perturbación hace que los estimadores minimocuadráticos ordinarios de los coeficientes de regresión sean inconsistentes. Resulta, pues, que la característica fundamental de las variables predeterminadas en un sistema de ecuaciones es el hecho de que no están correlacionadas con las perturbaciones del mismo período.

En general, la forma estructural de un sistema de ecuaciones simultáneas puede expresarse en la forma siguiente:

$$\begin{aligned} (13.3) \quad & \beta_{11}y_{1t} + \beta_{12}y_{2t} + \cdots + \beta_{1G}y_{Gt} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \cdots + \gamma_{1K}x_{Kt} = u_{1t} \\ & \beta_{21}y_{1t} + \beta_{22}y_{2t} + \cdots + \beta_{2G}y_{Gt} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + \cdots + \gamma_{2K}x_{Kt} = u_{2t} \\ & \vdots \\ & \beta_{G1}y_{1t} + \beta_{G2}y_{2t} + \cdots + \beta_{GG}y_{Gt} + \gamma_{G1}x_{1t} + \gamma_{G2}x_{2t} + \cdots + \gamma_{GK}x_{Kt} = u_{Gt}. \end{aligned}$$

en las que las y son variables endógenas, las x son variables predeterminadas, las u son las perturbaciones estocásticas, y $t = 1, 2, \dots, T$. Las β y γ reciben el nombre de coeficientes estructurales. En el sistema existen G variables en

las que G y K variables predeterminadas, pero, en general, no todas estarán presentes en cada ecuación, ya que puede saberse que algunas de las β y las γ son cero. Además, en cada ecuación se considera que una de las β es la unidad, lo que indica que una de las variables endógenas aparece como variable «dependiente» cuando la ecuación se escribe en forma de una ecuación de regresión corriente. Debe señalarse también que algunas de las ecuaciones pueden ser en realidad identidades, lo que significa que se conocen todos sus coeficientes y no contienen perturbaciones estocásticas. El sistema completo de ecuaciones puede escribirse en forma matricial como

$$(13.3a) \quad By_t + \Gamma x_t = u_t,$$

siendo

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Gt} \end{bmatrix}, \quad x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Kt} \end{bmatrix}, \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Gt} \end{bmatrix},$$

$$(G \times 1) \quad (K \times 1) \quad (G \times 1)$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GG} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{G1} & \gamma_{G2} & \cdots & \gamma_{GK} \end{bmatrix},$$

$$(G \times G) \quad (G \times K)$$

Si en alguna de las ecuaciones figura una constante, una de las x será igual a la unidad para todo $t = 1, 2, \dots, T$. Por lo que respecta a las perturbaciones estocásticas, se especifica que cada perturbación cumple los supuestos del modelo clásico de regresión normal lineal, es decir, que

$$\begin{aligned} u_{gt} &\sim N(0, \sigma_{gg}), & g &= 1, 2, \dots, G; \\ E(u_{gt}u_{gs}) &= 0, & t, s &= 1, 2, \dots, T; \\ & & t &\neq s. \end{aligned}$$

Sin embargo, no se elimina la posibilidad de que las perturbaciones de distintas ecuaciones estén correlacionadas, o sea, que

$$E(u_{gt}u_{ht}) = \sigma_{gh} \quad (g, h = 1, 2, \dots, G).$$

En notación matricial estos supuestos se convierten en

$$(13.4) \quad u_t \sim N(0, \Phi),$$

$$(13.5) \quad E(u_t u_t') = 0,$$

siendo

$$\Phi = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1G} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2G} \\ \vdots & \vdots & & \vdots \\ \sigma_{G1} & \sigma_{G2} & \dots & \sigma_{GG} \end{bmatrix} \\ (G \times G)$$

La matriz Φ se conoce con el nombre de matriz de variancias y covariancias de las perturbaciones estructurales. Si en el sistema existe alguna identidad, Φ se refiere únicamente a las ecuaciones que no son identidades y, por tanto, su dimensión debe reducirse en la forma adecuada.

La forma reducida del sistema se obtiene resolviendo las ecuaciones de la forma estructural y despejando los valores de las variables endógenas, o sea expresando las y en términos de x y u . El resultado puede escribirse en la forma

$$(13.6) \quad \begin{aligned} y_{1t} &= \pi_{11}x_{1t} + \pi_{12}x_{2t} + \dots + \pi_{1K}x_{Kt} + v_{1t}, \\ y_{2t} &= \pi_{21}x_{1t} + \pi_{22}x_{2t} + \dots + \pi_{2K}x_{Kt} + v_{2t}, \\ &\vdots \\ y_{Gt} &= \pi_{G1}x_{1t} + \pi_{G2}x_{2t} + \dots + \pi_{GK}x_{Kt} + v_{Gt}. \end{aligned}$$

Las π representan los coeficientes de la forma reducida y las v son las perturbaciones de la forma reducida. En general, cada una de estas perturbaciones es función lineal de *todos* los términos de perturbación estructurales. Utilizando la notación matricial, podemos escribir (13.6) como

$$(13.6a) \quad y_t = \Pi x_t + v_t, \\ \text{en la que} \quad \Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2K} \\ \vdots & \vdots & & \vdots \\ \pi_{G1} & \pi_{G2} & \dots & \pi_{GK} \end{bmatrix}, \quad v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Gt} \end{bmatrix} \\ (G \times K) \quad (G \times 1)$$

La relación entre la forma estructural y la forma reducida puede obtenerse de forma explícita despejando y_t en (13.3a). En este caso obtenemos

$$(13.6b) \quad y_t = -B^{-1}\Gamma x_t + B^{-1}u_t.$$

Si comparamos este resultado con la forma reducida (13.6a), podemos ver que

$$(13.7) \quad \Pi = -B^{-1}\Gamma$$

y

$$(13.8) \quad v_t = B^{-1}u_t.$$

La matriz de variancias y covariancias de las perturbaciones de la forma reducida, Ψ , es

$$(13.9) \quad \Psi = E(v_t v_t') = E[B^{-1}u_t u_t'(B^{-1})'] = B^{-1}\Phi(B^{-1})'.$$

EJEMPLO 1. El modelo de oferta y demanda presentado en (13.1) es

$$Q_t - \alpha_1 - \alpha_2 P_t - \alpha_3 Y_t = \varepsilon_{1t},$$

$$Q_t - \beta_1 - \beta_2 P_t = \varepsilon_{2t}.$$

Según (13.3a), lo podemos escribir en la forma

$$\begin{bmatrix} 1 & -\alpha_2 \\ 1 & -\beta_2 \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} + \begin{bmatrix} -\alpha_1 & -\alpha_3 \\ -\beta_1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

La forma reducida del sistema es

$$\begin{bmatrix} Q_t \\ P_t \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix},$$

en la que

$$\begin{aligned} \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} &= -B^{-1}\Gamma = \begin{bmatrix} 1 & -\alpha_2 \\ 1 & -\beta_2 \end{bmatrix}^{-1} \begin{bmatrix} -\alpha_1 & -\alpha_3 \\ -\beta_1 & 0 \end{bmatrix} \\ &= -\begin{bmatrix} 1 & -\alpha_2 \\ \alpha_2 - \beta_2 & 1 \end{bmatrix} \begin{bmatrix} -\beta_2 & \alpha_2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -\alpha_1 & -\alpha_3 \\ -\beta_1 & 0 \end{bmatrix} \\ &= -\begin{bmatrix} 1 & -\alpha_2 \\ \alpha_2 - \beta_2 & 1 \end{bmatrix} \begin{bmatrix} (\alpha_1\beta_2 - \alpha_2\beta_1) & \alpha_3\beta_2 \\ (\alpha_1 - \beta_1) & \alpha_3 \end{bmatrix} \end{aligned}$$

y

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = B^{-1} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_2 \\ \alpha_2 - \beta_2 & 1 \end{bmatrix} \begin{bmatrix} -\beta_2 & \alpha_2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_2 \\ \alpha_2 - \beta_2 & 1 \end{bmatrix} \begin{bmatrix} -\beta_2\varepsilon_{1t} + \alpha_2\varepsilon_{2t} \\ -\varepsilon_{1t} + \varepsilon_{2t} \end{bmatrix}.$$

Naturalmente, estos resultados son iguales que los presentados en (13.2).

EJEMPLO 2. A continuación, presentamos un modelo muy simplificado de la economía:

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t} \quad (\text{consumo}),$$

$$I_t = \beta_0 + \beta_1 r_t + \beta_2 I_{t-1} + u_{2t} \quad (\text{inversión}),$$

$$r_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 M_t + u_{3t} \quad (\text{mercado de dinero}),$$

$$Y_t = C_t + I_t + G_t \quad (\text{identidad de la renta}),$$

en el que C = consumo, Y = renta, I = inversión, r = tipo de interés, M = oferta monetaria y G = gasto público. Las variables C_t , I_t , Y_t y r_t son endógenas y las

demás son predeterminadas. Podemos describir el modelo en la forma

$$\begin{bmatrix} 1 & 0 & -\alpha_1 & 0 \\ 0 & 1 & 0 & -\beta_1 \\ 0 & 0 & -\gamma_1 & 1 \\ -1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} C_t \\ I_t \\ Y_t \\ r_t \end{bmatrix} + \begin{bmatrix} -\alpha_0 & -\alpha_2 & 0 & 0 & 0 \\ -\beta_0 & 0 & -\beta_2 & 0 & 0 \\ -\gamma_0 & 0 & 0 & -\gamma_2 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ C_{t-1} \\ I_{t-1} \\ M_t \\ G_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ 0 \end{bmatrix}$$

La forma reducida de este sistema es

$$\begin{bmatrix} C_t \\ I_t \\ Y_t \\ r_t \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \alpha_0(1 - \beta_1\gamma_1) + \alpha_1(\beta_0 + \beta_1\gamma_0) & \alpha_1\alpha_2 & \alpha_1\beta_2 & \alpha_1\beta_1\gamma_2 & \alpha_1 \\ \alpha_0\beta_1\gamma_1 + (1 - \alpha_1)(\beta_0 + \beta_1\gamma_0) & \alpha_2\beta_1\gamma_1 & (1 - \alpha_1)\beta_2 & (1 - \alpha_1)\beta_1\gamma_2 & \beta_1\gamma_1 \\ \alpha_0 + \beta_0 + \beta_1\gamma_0 & \alpha_2 & \beta_2 & \beta_1\gamma_2 & 1 \\ (\alpha_0 + \beta_0)\gamma_1 + (1 - \alpha_1)\gamma_0 & \alpha_2\gamma_1 & \beta_2\gamma_1 & (1 - \alpha_1)\gamma_2 & \gamma_1 \end{bmatrix} \begin{bmatrix} 1 \\ C_{t-1} \\ I_{t-1} \\ M_t \\ G_t \end{bmatrix} + \frac{1}{\Delta} \begin{bmatrix} 1 - \beta_1\gamma_1 & \alpha_1 & \alpha_1\beta_1 \\ \beta_1\gamma_1 & (1 - \alpha_1) & (1 - \alpha_1)\beta_1 \\ 1 & 1 & \beta_1 \\ \gamma_1 & \gamma_1 & (1 - \alpha_1) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

en la que $\Delta = 1 - \alpha_1 - \beta_1\gamma_1$.

Tipos de modelos estructurales

La posición de los ceros que aparecen en la matriz B indica las variables endógenas que no aparecen en las distintas ecuaciones estructurales, y podemos hacer uso de ello para distinguir entre varios tipos de estructuras.

a) Si B es *diagonal*, es decir, si

$$B = \begin{bmatrix} \beta_{11} & 0 & \dots & 0 \\ 0 & \beta_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_{GG} \end{bmatrix},$$

en cada ecuación sólo aparece una variable endógena, lo que significa que las ecuaciones no son simultáneas, sino *aparentemente no relacionadas*. Este caso ya se ha estudiado en la sección 12.3.

b) Si B es *diagonal en bloques*, es decir, si

$$B = \begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_R \end{bmatrix},$$

en la que $B_1, B_2, \dots, B_R$ son matrices cuadradas, y los 0 representan matrices de ceros con las dimensiones adecuadas, cada bloque contiene su propio conjunto de variables endógenas. En este caso, tenemos R sistemas de ecuaciones simultáneas y cada bloque constituye un sistema distinto ya que al obtener las soluciones de la forma reducida sólo utilizamos las ecuaciones estructurales de un mismo grupo. Esto puede verse claramente si particionamos las ecuaciones estructurales en la forma siguiente:

$$\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_R \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ \vdots \\ y_t^{(R)} \end{bmatrix} + \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_R \end{bmatrix} x_t = \begin{bmatrix} u_t^{(1)} \\ u_t^{(2)} \\ \vdots \\ u_t^{(R)} \end{bmatrix}$$

en la que $y_t^{(r)}$ es el vector $(G_r \times 1)$ de las variables endógenas que aparecen en el bloque r , $u_t^{(r)}$ es el vector $(G_r \times 1)$ de las perturbaciones estructurales correspondientes, y Γ_r es la matriz $(G_r \times K)$ de los coeficientes γ en las ecuaciones estructurales del bloque r ($r = 1, 2, \dots, R$). La solución de la forma reducida para $y_t^{(r)}$ es

$$y_t^{(r)} = -B_r^{-1}\Gamma_r x_t + B_r^{-1}u_t^{(r)} \quad (r = 1, 2, \dots, R).$$

En este resultado no se efectúa ninguna referencia a las ecuaciones que no pertenecen al bloque r . Así pues, cuando la matriz B es diagonal en bloques decimos que se trata de una *estructura no integrada*. Si la matriz de variancias de las perturbaciones estructurales es también diagonal en bloques de la misma forma que la matriz B, a efectos de la estimación cada bloque puede considerarse como un sistema separado. Sin embargo, si la matriz de variancias y covariancias de las perturbaciones estructurales no es diagonal en bloques, los bloques sólo están «aparentemente» no relacionados. Podríamos añadir que existe otro tipo más estricto de no integración, que se conoce con el nombre de *no integración dinámica*, y corresponde al caso en que las ecuaciones estructurales de cada bloque no incluyen variables endógenas corrientes ni retardadas de ningún otro bloque. Este tipo de no integración implica que tanto el valor actual como la trayectoria a través del tiempo de una variable endógena quedan determinadas por completo por las ecuaciones del bloque en que aparece la variable en cuestión. En la sección 13.6 nos ocuparemos de nuevo de este caso.

c) Si B es *triangular*, es decir, si

$$B = \begin{bmatrix} \beta_{11} & 0 & \dots & 0 \\ \beta_{21} & \beta_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \dots & \beta_{GG} \end{bmatrix},$$

el sistema se denomina *recursivo*. En este caso, la solución de la variable endógena que aparece en el lugar g incluye solamente las g primeras ecuaciones estructurales, lo que significa que sólo se necesitarán todas las G ecuaciones estructurales para la solución correspondiente a la última variable endógena y_{Gt} . La primera ecuación estructural incluye solamente una variable endógena, de modo que coincide con la primera ecuación de la forma reducida.

b) Si B es *triangular en bloques*, es decir, si

$$B = \begin{bmatrix} B_{11} & 0 & \dots & 0 \\ B_{21} & B_{22} & \dots & 0 \\ \vdots & & & \\ B_{R1} & B_{R2} & \dots & B_{RR} \end{bmatrix}$$

en la que las B son matrices de dimensiones determinadas, el sistema se denomina *recursivo en bloques* y posee las mismas características que el sistema recursivo que acabamos de presentar, salvo por el hecho de que tendremos que hablar de bloques de ecuaciones en lugar de ecuaciones individuales.

e) Si B no es diagonal ni triangular (ni con bloques ni sin ellos), hablaremos de una *estructura integrada*. Este tipo de estructuras son las que han recibido mayor atención por parte de los econométricos y constituyen el objeto fundamental de nuestro estudio de los sistemas de ecuaciones simultáneas. Los sistemas de ecuaciones que se caracterizan por una estructura integrada se llaman a veces *sistemas generales interdependientes*.

13.2 El problema de la identificación

Debido a la presencia de variables endógenas entre las variables explicativas de las ecuaciones simultáneas, los estimadores minimocuadráticos ordinarios de los coeficientes estructurales no son consistentes, al menos en general. Sin embargo, en las ecuaciones de la forma reducida, las variables explicativas están constituidas por las variables predeterminadas del sistema, de modo que los estimadores minimocuadráticos ordinarios de los coeficientes de la forma reducida son consistentes. Esto sugiere la posibilidad de intentar estimar los coeficientes estructurales a través de la forma reducida, y el problema radica en saber si podemos obtener estimaciones de los coeficientes estructurales a través de las estimaciones consistentes de los coeficientes de la forma reducida. Obviamente, esto será posible siempre que los coeficientes estructurales (las β y γ) puedan expresarse en términos de los coeficientes de la forma reducida (π). Por tanto, se trata de un problema de *identificación* como el estudiado en la sección 11.2 en el contexto de los modelos uniecuacionales, pero referido aquí

a los sistemas de ecuaciones. Las ecuaciones de la forma reducida presentadas en (13.6a) en la forma

$$y_t = \Pi x_t + v_t$$

representan la versión sin restricciones de dichas ecuaciones, mientras que la forma (13.6b) representada por

$$y_t = -B^{-1}\Gamma x_t + B^{-1}u_t$$

indica la versión con restricciones. Como ocurría en el caso de los modelos uniecuacionales, cuando existe una correspondencia unívoca entre los parámetros con restricciones y sin restricciones (en el sentido de que existe una solución única para los parámetros con restricciones en términos de los coeficientes sin restricciones) nos encontramos con la *identificación exacta*. Por otra parte, cuando el número de coeficientes sin restricciones es mayor que el número de parámetros con ellas, no existe una solución única y nos hallamos en el caso de la *superidentificación*. Por último, si el número de coeficientes sin restricciones es insuficiente para obtener la solución, nos encontramos con la *subidentificación*. Se dice que una ecuación está identificada en los casos en que está exactamente identificada o superidentificada.

Podemos ilustrar el problema de la identificación refiriéndonos al modelo de oferta y demanda (13.1):

$$\text{(demanda),} \quad Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{1t}$$

$$\text{(oferta),} \quad Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{2t}$$

y a su forma reducida (13.2):

$$Q_t = \pi_{11} + \pi_{12} Y_t + v_{1t},$$

$$P_t = \pi_{21} + \pi_{22} Y_t + v_{2t}.$$

El procedimiento más sencillo para descubrir si es posible expresar las α y β en términos de las π consiste en sustituir Q_t y P_t por sus valores según la forma reducida (13.2) en la forma estructural (13.1). Esto da como resultado

$$\text{(demanda),} \quad (\pi_{11} + \pi_{12} Y_t + v_{1t}) = \alpha_1 + \alpha_2(\pi_{21} + \pi_{22} Y_t + v_{2t}) + \alpha_3 Y_t + \varepsilon_{1t}$$

$$\text{(oferta),} \quad (\pi_{11} + \pi_{12} Y_t + v_{1t}) = \beta_1 + \beta_2(\pi_{21} + \pi_{22} Y_t + v_{2t}) + \varepsilon_{2t}$$

Según las definiciones de v_t y v'_t , efectuadas anteriormente, las perturbaciones estocásticas de cada ecuación se eliminan y las expresiones se reducen a

$$(\text{demanda}), \quad \pi_{11} + \pi_{12}Y_t = (\alpha_1 + \alpha_2\pi_{21}) + (\alpha_2\pi_{22} + \alpha_3)Y_t$$

$$(\text{oferta}), \quad \pi_{11} + \pi_{12}Y_t = (\beta_1 + \beta_2\pi_{21}) + \beta_2\pi_{22}Y_t$$

La ecuación de demanda implica las igualdades siguientes

$$\pi_{11} = \alpha_1 + \alpha_2\pi_{21} \quad \text{y} \quad \pi_{12} = \alpha_2\pi_{22} + \alpha_3.$$

Como sólo tenemos dos igualdades, no podemos obtener las soluciones para las tres incógnitas α_1 , α_2 y α_3 . Las igualdades correspondientes a la ecuación de oferta son

$$\pi_{11} = \beta_1 + \beta_2\pi_{21} \quad \text{y} \quad \pi_{12} = \beta_2\pi_{22}.$$

De esto se sigue que

$$\beta_1 = \pi_{11} - \frac{\pi_{12}\pi_{21}}{\pi_{22}},$$

$$\beta_2 = \frac{\pi_{12}}{\pi_{22}}.$$

Por tanto, la ecuación de demanda está subidentificada y la ecuación de oferta está exactamente identificada.

Evidentemente, sería deseable disponer de alguna regla general para conocer la identificabilidad de una ecuación estructural determinada, y podemos obtener una regla de este tipo en la forma siguiente. Las ecuaciones estructurales son

$$B\mathbf{y}_t + \Gamma\mathbf{x}_t = \mathbf{u}_t,$$

mientras que las ecuaciones de la forma reducida son

$$\mathbf{y}_t = \Pi\mathbf{x}_t + \mathbf{v}_t.$$

Sustituyendo $\mathbf{y}_t$ por su valor según la expresión de la forma reducida en la forma estructural, obtenemos

$$B\Pi\mathbf{x}_t + B\mathbf{v}_t + \Gamma\mathbf{x}_t = \mathbf{u}_t.$$

Pero como

$$\mathbf{v}_t = B^{-1}\mathbf{u}_t,$$

podemos escribir

$$B\Pi\mathbf{x}_t = -\Gamma\mathbf{x}_t$$

$$B\Pi = -\Gamma.$$

o
...

Esta es la relación que se utiliza para conocer la identificabilidad de las ecuaciones de oferta y demanda. Vamos a tratar ahora de utilizar esta relación para obtener un criterio general de identificación para cada ecuación estructural. Escribiendo totalmente las matrices tenemos

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2G} \\ \vdots & \vdots & & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GG} \end{bmatrix} \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \vdots & \vdots & & \vdots \\ \pi_{G1} & \pi_{G2} & \cdots & \pi_{GK} \end{bmatrix} = - \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma_{G1} & \gamma_{G2} & \cdots & \gamma_{GK} \end{bmatrix}.$$

Para una de las ecuaciones del sistema, por ejemplo, la g , esto se convierte en

$$(13.10) \quad [\beta_{g1} \quad \beta_{g2} \quad \cdots \quad \beta_{gG}] \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \vdots & \vdots & & \vdots \\ \pi_{G1} & \pi_{G2} & \cdots & \pi_{GK} \end{bmatrix} = -[\gamma_{g1} \quad \gamma_{g2} \quad \cdots \quad \gamma_{gK}]$$

o

$$(13.10a) \quad \beta_g \Pi = -\gamma_g,$$

en la que

$$\beta_g = [\beta_{g1} \quad \beta_{g2} \quad \cdots \quad \beta_{gG}],$$

$$\gamma_g = [\gamma_{g1} \quad \gamma_{g2} \quad \cdots \quad \gamma_{gK}].$$

Si algunas variables endógenas y algunas predeterminadas no aparecen en la ecuación g , algunas de las β y algunas de las γ serán cero.

Sea G^A = número de variables endógenas que aparecen en la ecuación g (es decir, número de elementos distintos de cero en β_g);

$$G^{AA} = G - G^A;$$

K^* = número de variables predeterminadas que aparecen en la ecuación g (es decir, número de elementos distintos de cero en γ_g);

$$K^{**} = K - K^*.$$

Sin perder generalidad, podemos suponer que los elementos de β_g y γ_g están ordenados de forma que, en primer lugar, aparecen los elementos distintos de cero, y a continuación los ceros. En este caso, podemos particionar β_g y γ_g en la forma

$$(13.11) \quad \beta_g = [\beta_A \quad 0_{AA}],$$

$$\gamma_g = [\gamma_A \quad 0_{**}],$$

$$\begin{aligned}\text{siendo} \quad \beta_{\Delta} &= [\beta_{\Delta 1} \quad \beta_{\Delta 2} \quad \dots \quad \beta_{\Delta G^{\Delta}}] \rightarrow 1 \times G^{\Delta}, \\ 0_{\Delta\Delta} &= [0 \quad 0 \quad \dots \quad 0] \rightarrow 1 \times G^{\Delta\Delta}, \\ \gamma_{*} &= [\gamma_{\Delta 1} \quad \gamma_{\Delta 2} \quad \dots \quad \gamma_{\Delta K^{**}}] \rightarrow 1 \times K^{**}, \\ 0_{**} &= [0 \quad 0 \quad \dots \quad 0] \rightarrow 1 \times K^{**}.\end{aligned}$$

La matriz Π se puede particionar de forma similar:

$$(13.12) \quad \Pi = \begin{bmatrix} \Pi_{\Delta*} & \Pi_{\Delta**} \\ \Pi_{\Delta\Delta*} & \Pi_{\Delta\Delta**} \end{bmatrix},$$

$$\begin{aligned}\text{siendo} \quad \Pi_{\Delta*} &\rightarrow (G^{\Delta} \times K^{*}), \\ \Pi_{\Delta**} &\rightarrow (G^{\Delta} \times K^{**}), \\ \Pi_{\Delta\Delta*} &\rightarrow (G^{\Delta\Delta} \times K^{*}), \\ \Pi_{\Delta\Delta**} &\rightarrow (G^{\Delta\Delta} \times K^{**}).\end{aligned}$$

Utilizando (13.11) y (13.12) podemos escribir de nuevo (13.10a) en la forma

$$(13.13) \quad [\beta_{\Delta} \quad 0_{\Delta\Delta}] \begin{bmatrix} \Pi_{\Delta*} & \Pi_{\Delta**} \\ \Pi_{\Delta\Delta*} & \Pi_{\Delta\Delta**} \end{bmatrix} = -[\gamma_{*} \quad 0_{**}].$$

A partir de (13.13) podemos obtener las igualdades siguientes:

$$(13.14) \quad \begin{aligned} \beta_{\Delta} \Pi_{\Delta*} &= -\gamma_{*}, \\ (1 \times G^{\Delta})(G^{\Delta} \times K^{*}) & \quad (1 \times K^{*}) \end{aligned}$$

$$(13.15) \quad \begin{aligned} \beta_{\Delta} \Pi_{\Delta**} &= 0_{**}, \\ (1 \times G^{\Delta})(G^{\Delta} \times K^{**}) & \quad (1 \times K^{**}) \end{aligned}$$

En cada una de las actuaciones estructurales una de las β es igual a la unidad y, por tanto, las igualdades (13.14) y (13.15) incluyen $(G^{\Delta} - 1)$ β desconocidas, y K^{*} valores desconocidos de γ . La igualdad (13.15) resulta particularmente importante ya que en ella no aparece ninguna de las γ , y si podemos resolver β_{Δ} a partir de ella, podremos también resolver fácilmente (13.14) obteniendo los valores de γ^{*} . Ahora bien, la igualdad (13.15) contiene en total K^{**} ecuaciones, una para cada elemento del vector de orden $(1 \times K^{**})$. Evidentemente, si queremos obtener una solución para los $(G^{\Delta} - 1)$ elementos desconocidos de β_{Δ} necesitamos, como mínimo, $(G^{\Delta} - 1)$ ecuaciones. Lo que significa que es necesario que

$$(13.16) \quad K^{**} \geq G^{\Delta} - 1.$$

Esto se conoce con el nombre de *condición de orden* para la identificabilidad. Esta condición dice en realidad que una condición necesaria para la identificación de una ecuación estructural es que el número de variables predetermi-

nadas excluidas de la ecuación sea al menos igual al número de variables endógenas incluidas en dicha ecuación menos una. Obsérvese que se trata solamente de una condición *necesaria* pero no suficiente para la identificación, ya que puede ocurrir que las K^{**} ecuaciones de (13.15) no sean independientes. Es decir, puede suceder que las ecuaciones de (13.15) contengan menos de $G^{\Delta} - 1$ piezas distintas de información acerca de la relación entre las β y las π . Por tanto, la condición necesaria y suficiente para la identificación es que el número de ecuaciones independientes en (13.15) sea $G^{\Delta} - 1$. Esto ocurrirá si, y solamente si, el orden del mayor determinante distinto de cero que pueda formarse con todas las submatrices cuadradas de $\Pi_{\Delta**}$ es $G^{\Delta} - 1$, es decir, si

$$(13.17) \quad \text{rango}(\Pi_{\Delta**}) = G^{\Delta} - 1.$$

Esto se conoce con el nombre de *condición del rango* para la identificabilidad.

Un procedimiento sencillo para calcular el rango de $\Pi_{\Delta**}$ consiste en particionar las matrices de los coeficientes estructurales en la forma siguiente:

$$B = \begin{bmatrix} \beta_{\Delta} & 0_{\Delta\Delta} \\ B_{\Delta} & B_{\Delta\Delta} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_{*} & 0_{**} \\ \Gamma_{*} & \Gamma_{**} \end{bmatrix},$$

en las que β_{Δ} , γ_{*} , $0_{\Delta\Delta}$, y 0_{**} son los vectores fila definidos en (13.11) y

$$B_{\Delta} \rightarrow (G - 1) \times G^{\Delta},$$

$$B_{\Delta\Delta} \rightarrow (G - 1) \times G^{\Delta\Delta},$$

$$\Gamma_{*} \rightarrow (G - 1) \times K^{*},$$

$$\Gamma_{**} \rightarrow (G - 1) \times K^{**}.$$

Obsérvese que $B_{\Delta\Delta}$ y Γ_{**} son las matrices de los coeficientes estructurales de las variables omitidas en la ecuación g pero incluidas en las restantes ecuaciones estructurales. Si ahora formamos una nueva matriz Δ que se define como

$$\Delta = [B_{\Delta\Delta} \quad \Gamma_{**}],$$

tendremos

$$(13.17a) \quad \text{rango}(\Pi_{\Delta**}) = \text{rango}(\Delta) - G^{\Delta\Delta}.$$

Esto puede demostrarse de la forma siguiente. Definamos Δ_{*} como

$$\Delta_{*} = \begin{bmatrix} 0_{**} & 0_{\Delta\Delta} \\ \Gamma_{**} & B_{\Delta\Delta} \end{bmatrix}.$$

Evidentemente, el rango de Δ_* es igual al de Δ , ya que el rango de una matriz no varía si le añadimos una fila de ceros o si intercambiamos dos o más columnas. Ahora bien, Δ_* puede escribirse en la forma

$$\Delta_* = \begin{bmatrix} \beta_{\Delta} & 0_{\Delta\Delta} \\ B_{\Delta} & B_{\Delta\Delta} \end{bmatrix} \begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta\Delta\Delta} \\ -\Pi_{\Delta\Delta**} & I_{\Delta\Delta} \end{bmatrix},$$

en la que $0_{\Delta\Delta\Delta}$ es una matriz de ceros de orden $G^{\Delta} \times G^{\Delta\Delta}$, e $I_{\Delta\Delta}$ es una matriz identidad de orden $G^{\Delta\Delta} \times G^{\Delta\Delta}$. Para verlo, podemos efectuar la multiplicación indicada en la igualdad anterior, obteniendo

$$\Delta_* = \begin{bmatrix} -\beta_{\Delta}\Pi_{\Delta**} & 0_{\Delta\Delta} \\ (-B_{\Delta}\Pi_{\Delta**} - B_{\Delta\Delta}\Pi_{\Delta\Delta**}) & B_{\Delta\Delta} \end{bmatrix}.$$

Pero, según (13.15),

$$-\beta_{\Delta}\Pi_{\Delta**} = 0_{**},$$

y según la igualdad $B\Pi = -\Gamma$, se sigue que

$$-B_{\Delta}\Pi_{\Delta**} - B_{\Delta\Delta}\Pi_{\Delta\Delta**} = \Gamma_{**}.$$

Utilizando el teorema que dice¹ que, si multiplicamos una matriz A por una matriz no singular, el producto tiene el mismo rango que A, podemos escribir

$$\begin{aligned} \text{rango}(\Delta_*) &= \text{rango}(B^{-1}\Delta_*) \\ &= \text{rango} \begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta\Delta\Delta} \\ -\Pi_{\Delta\Delta**} & I_{\Delta\Delta} \end{bmatrix} \\ &= \text{rango} \begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta\Delta\Delta} \\ -\Pi_{\Delta\Delta**} & I_{\Delta\Delta} \end{bmatrix} \begin{bmatrix} \Gamma_{**} & 0_{**\Delta\Delta} \\ \Pi_{\Delta\Delta**} & I_{\Delta\Delta} \end{bmatrix} \\ &= \text{rango} \begin{bmatrix} -\Pi_{\Delta**} & 0_{\Delta\Delta\Delta} \\ 0_{\Delta\Delta**} & I_{\Delta\Delta} \end{bmatrix} \\ &= \text{rango}(\Pi_{\Delta**}) + G^{\Delta\Delta}, \end{aligned}$$

en las que $0_{**\Delta\Delta}$ y $0_{\Delta\Delta**}$ son matrices de ceros de orden $K^{**} \times G^{\Delta\Delta}$ y $G^{\Delta\Delta} \times K^{**}$, respectivamente, e I_{**} es una matriz identidad de orden $K^{**} \times K^{**}$. Con esto finaliza la demostración de (13.17a). Evidentemente, resulta mucho más fácil calcular el rango de $\Pi_{\Delta**}$ a partir de (13.17a) que a partir de la solución directa de $\Pi_{\Delta**}$ en términos de los coeficientes estructurales.

La condición de orden y la del rango nos permiten establecer la siguiente regla general para conocer la identificabilidad de una ecuación estructural.

1. Si $K^{**} > G^{\Delta} - 1$, y $\text{rango}(\Pi_{\Delta**}) = G^{\Delta} - 1$, la ecuación está superidentificada.
2. Si $K^{**} = G^{\Delta} - 1$, y $\text{rango}(\Pi_{\Delta**}) = G^{\Delta} - 1$, la ecuación está exactamente identificada.
3. Si $K^{**} \geq G^{\Delta} - 1$, y $\text{rango}(\Pi_{\Delta**}) < G^{\Delta} - 1$, la ecuación está subidentificada.
4. Si $K^{**} < G^{\Delta} - 1$, la ecuación está subidentificada.

EJEMPLO 1. El modelo de oferta y demanda presentado en (13.1) es

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \epsilon_{1t} \quad (\text{demanda}).$$

$$Q_t = \beta_1 + \beta_2 P_t + \epsilon_{2t} \quad (\text{oferta}).$$

Consideremos en primer lugar la identificabilidad de la ecuación de demanda. En esta ecuación hay dos variables endógenas incluidas, Q_t y P_t , de modo que

$$G^{\Delta} - 1 = 1.$$

Como en la ecuación de demanda no hay variables predeterminadas excluidas, tenemos

$$K^{**} = 0.$$

Por tanto,

$$K^{**} < G^{\Delta} - 1,$$

lo que significa que no se satisface la condición de orden y, en consecuencia, la ecuación de demanda está *subidentificada*.

Pasando ahora a la ecuación de oferta, vemos que incluye también las dos variables endógenas Q_t y P_t , es decir, que

$$G^{\Delta} - 1 = 1,$$

pero, en este caso, la variable predeterminada Y_t , que aparece en el sistema no aparece en la ecuación de oferta. Por tanto, tenemos

$$K^{**} = 1,$$

o sea, que se satisface la condición de orden. En cuanto a la condición del rango, tenemos

$$\text{rango}(\Delta) = \text{rango}[-\alpha_3] = 1.$$

Así, pues, según (13.17a),

$$\text{rango}(\Pi_{\Delta**}) = 1 - 0 = 1,$$

1. Ver, por ejemplo, J. Johnston, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, p. 92.

y se cumple también la condición del rango. Obsérvese que, en la ecuación de oferta de este modelo, la igualdad (13.15) es

$$[1 \ \beta_2] - \begin{bmatrix} \pi_{QY} \\ \pi_{PY} \end{bmatrix} = 0,$$

o sea, que consiste en una ecuación y una incógnita, y, por tanto, la función de oferta está *exactamente* identificada.

EJEMPLO 2. El modelo agregado de la economía que se ha presentado anteriormente es:

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_t && \text{(consumo),} \\ I_t &= \beta_0 + \beta_1 r_t + \beta_2 I_{t-1} + u_{2t} && \text{(inversión),} \\ r_t &= \gamma_0 + \gamma_1 Y_t + \gamma_2 M_t + u_{3t} && \text{(mercado de dinero),} \\ Y_t &= C_t + I_t + G_t && \text{(identidad de la renta).} \end{aligned}$$

En cada una de las tres primeras ecuaciones estructurales tenemos

$$G^\Delta - 1 = 1,$$

$$K^{**} = 3,$$

de modo que la condición de orden se cumple en los tres casos. En cuanto a la condición del rango, en la *función de consumo* el rango de Δ es

$$\text{rango}(\Delta) = \text{rango} \begin{bmatrix} 1 & -\beta_1 & -\beta_2 & 0 & 0 \\ 0 & 1 & 0 & -\gamma_2 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} = 3.$$

Por tanto, $\text{rango}(\Pi_{\Delta^{**}}) = 3 - 2 = 1$,

o sea, que se cumple la condición del rango. Esto significa que la ecuación de consumo está *identificada*. Obsérvese que, en el caso de esta función de consumo, la igualdad (13.15) se convierte en

$$[1 - \alpha_1] \begin{bmatrix} \pi_{CI-1} & \pi_{CM} & \pi_{CG} \\ \pi_{YI-1} & \pi_{YM} & \pi_{YG} \end{bmatrix} = [0 \ 0 \ 0].$$

Es decir, tenemos tres ecuaciones para obtener una incógnita y, por tanto, la función de consumo está *superidentificada*. La obtención de la condición del rango en las dos ecuaciones estructurales restantes sería similar.

Las condiciones de orden y del rango para la identificabilidad se han presentado en términos de los parámetros de la población, suponiendo que ninguno de estos parámetros estructurales es cero. Sin embargo, no sabemos cuáles son los valores de estos parámetros y, en realidad, podemos querer contrastar la hipótesis de que *son cero*. Si esta hipótesis no se rechaza en todos los casos, algunas variables se considerarían irrelevantes y, por tanto, no deberían incluirse en la cuenta. Por ejemplo, para saber si una ecuación estructural determinada satisface la condición de orden para la identificabilidad, contamos el número de variables predeterminadas que se excluyen de la ecuación,

pero si algunas de ellas son irrelevantes no las tendríamos que considerar. Esto significa que una ecuación estructural que, en principio, parece cumplir la condición de orden, puede no cumplirla en realidad una vez desechadas las variables irrelevantes. Por tanto, cabe considerar la identificabilidad de una ecuación como una hipótesis a contrastar y no como una especificación previa. En la literatura puede verse una exposición del procedimiento de contrastación sugerido para dicho caso.²

Identificación según las restricciones sobre la matriz de variancias y covariancias de las perturbaciones

El análisis anterior de las condiciones de identificabilidad se ha limitado a la especificación de la ecuación estructural sin hacer ninguna referencia a la matriz de variancias y covariancias de las perturbaciones, y se ha mostrado que se puede identificar una ecuación estructural especificando que algunas de las variables que aparecen en el sistema no aparecen en la ecuación que se considera. Como la omisión de una variable en una ecuación lineal equivale a especificar que el correspondiente coeficiente β o γ es igual a cero, puede decirse que la identificación de una ecuación se consigue estableciendo *restricciones cero* sobre algunos de los coeficientes. (Naturalmente, podría conseguirse también la identificación estableciendo restricciones *distintas de cero* sobre los coeficientes estructurales, por ejemplo, especificando que algunos coeficientes tienen un valor determinado, no necesariamente cero, o especificando la razón o razones entre los coeficientes de una ecuación lineal.) Pero en este caso, debe ser posible también conseguir la identificación estableciendo restricciones previas sobre algunos elementos de la matriz de variancias y covariancias de las perturbaciones de la regresión. Ilustraremos esta posibilidad refiriéndonos al modelo de oferta y demanda (13.1). Al estudiar la identificabilidad de las ecuaciones estructurales se llegó a la conclusión de que, dadas las restricciones cero sobre los coeficientes estructurales, la ecuación de demanda está subidentificada y la ecuación de oferta está exactamente identificada. Introduzcamos ahora una restricción adicional especificando que los términos de perturbación de las dos ecuaciones son independientes entre sí, es decir, que

$$E(e_{1t}e_{2t}) = 0$$

2. R. L. Basman. «On Finite Sample Distributions of Generalized Classical Linear Identifiability Tests Statistics», *Journal of the American Statistical Association*, vol. 55, diciembre 1960, pp. 650-59.

para todo t . Los resultados anteriores, obtenidos a partir de las restricciones cero sobre los coeficientes estructurales han indicado las siguientes relaciones entre los coeficientes estructurales y los de la forma reducida:

$$\pi_{11} = \alpha_1 + \alpha_2\pi_{21}, \quad \pi_{11} = \beta_1 + \beta_2\pi_{21},$$

$$\pi_{12} = \alpha_2\pi_{22} + \alpha_3, \quad \pi_{12} = \beta_2\pi_{22}.$$

Como se ha señalado anteriormente, a partir de estas ecuaciones podemos obtener los valores de las β pero no los de las α . Sin embargo, ahora podemos añadir otra relación que se deriva de la restricción del valor cero de la covariancia entre las dos perturbaciones. Señalemos en primer lugar que las perturbaciones de la forma reducida vienen dadas por

$$v_{1t} = \frac{-\beta_2 e_{1t} + \alpha_2 e_{2t}}{\alpha_2 - \beta_2},$$

$$v_{2t} = \frac{-e_{1t} + e_{2t}}{\alpha_2 - \beta_2}.$$

Por tanto, podemos escribir

$$e_{1t} = v_{1t} - \alpha_2 v_{2t},$$

$$e_{2t} = v_{1t} - \beta_2 v_{2t}.$$

Así pues, la restricción

$$E(e_{1t}e_{2t}) = 0$$

implica que

$$E(v_{1t} - \alpha_2 v_{2t})(v_{1t} - \beta_2 v_{2t}) = 0$$

o

$$(13.18) \quad \psi_{11} - \alpha_2\psi_{12} - \beta_2\psi_{12} + \alpha_2\beta_2\psi_{22} = 0,$$

en la que las Ψ representan los elementos de la matriz de variancias y covariancias de las perturbaciones de la forma reducida. Como sabemos que

$$\beta_2 = \frac{\pi_{12}}{\pi_{22}},$$

podemos despejar α_2 en la ecuación (13.18) obteniendo

$$\alpha_2 = \frac{\pi_{12}\psi_{12} - \pi_{22}\psi_{11}}{\pi_{12}\psi_{22} - \pi_{22}\psi_{12}}.$$

Además, como

$$\alpha_1 = \pi_{11} - \alpha_2\pi_{21}$$

$$\alpha_3 = \pi_{12} - \alpha_2\pi_{22},$$

y

podemos obtener α_1 y α_3 . Por tanto, al añadir la restricción sobre la covariancia entre las perturbaciones estructurales, la ecuación de demanda que estaba subidentificada se convierte en identificada.

Subidentificación

El problema de la identificación es importante a causa de sus implicaciones para la estimación y la contrastación de hipótesis. Si una ecuación estructural está identificada, podemos obtener estimadores consistentes de sus coeficientes estructurales, pero en el caso de subidentificación esto no es posible. Naturalmente, podemos estimar los coeficientes de una ecuación estructural subidentificada por el método de los mínimos cuadrados ordinarios, pero las estimaciones resultantes son inconsistentes. Podemos ilustrar el hecho de que la estimación consistente de los coeficientes de una ecuación estructural subidentificada no tiene éxito, utilizando al modelo de oferta y demanda (13.1). En este caso, vimos que, cuando no se dispone de ninguna información previa sobre la matriz de variancias y covariancias de las perturbaciones, la ecuación de demanda de este sistema está subidentificada. Esta ecuación se expresó en la forma

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + e_{1t}.$$

Supongamos que tratamos de estimar los coeficientes de esta ecuación por el método de las variables instrumentales, que produce estimaciones consistentes. Este método se presentó en relación con la estimación de modelos de regresión con errores en las variables, en la sección 9.1, y en la sección 11.4 en el caso de los modelos con retardos distribuidos. El método consiste en emparejar cada variable explicativa de la ecuación que queremos estimar con una variable instrumental que se supone no correlacionada con la perturbación de la regresión, pero correlacionada con la variable explicativa con la que está emparejada. En el caso de la ecuación de demanda, las variables explicativas son Y_t y P_t , y como la primera no está correlacionada con e_{1t} , según los supuestos iniciales, puede servir como su propia variable instrumental, de modo que sólo tenemos que encontrar una variable instrumental para P_t . Si llamamos Z_t a esta variable, las «ecuaciones normales de las variables instrumentales» son

$$\sum_t Q_t = \alpha_1^* T + \alpha_2^* \sum_t P_t + \alpha_3^* \sum_t Y_t,$$

$$\sum_t Q_t Z_t = \alpha_1^* \sum_t Z_t + \alpha_2^* \sum_t P_t Z_t + \alpha_3^* \sum_t Y_t Z_t,$$

$$\sum_t Q_t Y_t = \alpha_1^* \sum_t Y_t + \alpha_2^* \sum_t P_t Y_t + \alpha_3^* \sum_t Y_t^2,$$

en las que $\alpha_1^\dagger$, $\alpha_2^\dagger$ y $\alpha_3^\dagger$ son los estimadores por variables instrumentales de los coeficientes respectivos. Por ejemplo, la solución para $\alpha_2^\dagger$ es

$$\alpha_2^\dagger = \frac{m_{QZ}m_{YY} - m_{QY}m_{YZ}}{m_{PZ}m_{YY} - m_{PY}m_{YZ}},$$

en la que

$$m_{QZ} = \sum_i (Q_i - \bar{Q})(Z_i - \bar{Z}),$$

$$m_{YY} = \sum_i (Y_i - \bar{Y})^2,$$

etc. Según la solución de la forma reducida, tenemos

$$(Q_i - \bar{Q}) = \left(\frac{-\alpha_3\beta_2}{\alpha_2 - \beta_2} \right) (Y_i - \bar{Y}) + (v_{1i} - \bar{v}_1),$$

$$(P_i - \bar{P}) = \left(\frac{-\alpha_3}{\alpha_2 - \beta_2} \right) (Y_i - \bar{Y}) + (v_{2i} - \bar{v}_2),$$

y, por tanto, se sigue que

$$\lim \text{prob } \frac{1}{T} m_{QZ} = \left(\frac{-\alpha_3\beta_2}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{1}{T} m_{YZ},$$

$$\lim \text{prob } \frac{1}{T} m_{QY} = \left(\frac{-\alpha_3\beta_2}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{1}{T} m_{YY},$$

$$\lim \text{prob } \frac{1}{T} m_{PZ} = \left(\frac{-\alpha_3}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{1}{T} m_{YZ},$$

$$\lim \text{prob } \frac{1}{T} m_{PY} = \left(\frac{-\alpha_3}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{1}{T} m_{YY}.$$

Por tanto,

$$\lim \text{prob } \alpha_2^\dagger =$$

$$= \frac{\left(\frac{-\alpha_3\beta_2}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{m_{YZ}}{T} \cdot \lim \text{prob } \frac{m_{YY}}{T} - \left(\frac{-\alpha_3\beta_2}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{m_{YY}}{T} \cdot \lim \text{prob } \frac{m_{YZ}}{T}}{\left(\frac{-\alpha_3}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{m_{YZ}}{T} \cdot \lim \text{prob } \frac{m_{YY}}{T} - \left(\frac{-\alpha_3}{\alpha_2 - \beta_2} \right) \lim \text{prob } \frac{m_{YY}}{T} \cdot \lim \text{prob } \frac{m_{YZ}}{T}} = 0.$$

es decir, el $\lim \text{prob } \alpha_2^\dagger$ es indeterminado, y para $\alpha_1^\dagger$ y $\alpha_3^\dagger$ obtendremos resultados similares. En consecuencia, el método de las variables instrumentales fracasa como habíamos pronosticado.

La incapacidad para obtener estimaciones consistentes de los coeficientes estructurales de una ecuación subidentificada se puede explicar también de forma más intuitiva. En el modelo de oferta y demanda de (13.1) hemos visto

que la identificación de la ecuación de oferta se debe a la presencia de la variable exógena renta en la función de demanda, ya que, de no ser así, las dos ecuaciones del modelo estarían subidentificadas. En este caso, todas las observaciones del precio y la cantidad se esparcirían en torno al punto de intersección del precio medio y la cantidad media. Un aumento en el número de observaciones de la muestra aumentaría únicamente la densidad de la nube de puntos sin proporcionar mayor información acerca de las dos líneas que se pretenden estimar. Sin embargo, cuando la variable renta se halla presente, la función de demanda se desplaza al variar la renta y, de esta forma, observamos los puntos que se esparcen en torno a la función de oferta. Cuanto mayor sea el número de puntos que observemos, más precisas serán las estimaciones de los coeficientes de la función de oferta. Como en esta última ecuación no aparece ninguna variable exógena, no podemos localizar la función de demanda de forma similar, pero, sin embargo, si las perturbaciones de las dos funciones son independientes entre sí, su covarianza es cero y esta información adicional nos permite identificar también los parámetros de la función de demanda. Obsérvese que, como los parámetros de las ecuaciones subidentificadas no se pueden estimar de forma consistente, las hipótesis sobre sus valores no pueden ser refutadas por las observaciones muestrales. Esto significa que, desde el punto de vista científico, la teoría subyacente es incompleta.

13.3 Métodos de estimación de ecuaciones aisladas

En la sección 13.1 se puso de relieve que la aplicación del método de los mínimos cuadrados ordinarios para la estimación de las ecuaciones estructurales de un sistema de ecuaciones simultáneas produce, en general, estimaciones inconsistentes y, por tanto, tenemos que desarrollar otros métodos para obtener estimaciones consistentes de estos coeficientes estructurales. Como ya se ha señalado, ello no es posible cuando la ecuación estructural está subidentificada.

Nuestro objetivo serán, pues, las ecuaciones estructurales identificadas para las que disponemos de métodos consistentes de estimación. Estos métodos para la estimación de ecuaciones estructurales pueden agruparse en dos categorías: 1) métodos diseñados para estimar una sola ecuación estructural haciendo referencia al resto del sistema solamente en forma limitada y 2) métodos en los que se estiman simultáneamente todas las ecuaciones del sistema. En esta sección nos ocuparemos de los métodos pertenecientes a la primera categoría.

Estimación de una ecuación exactamente identificada

Consideremos el problema de la estimación de una ecuación estructural *exactamente identificada* que pertenece a un sistema general interdependiente de

ecuaciones simultáneas sin ninguna restricción sobre la matriz de variancias y covariancias de los términos de perturbación. En este caso, existe una solución única para los coeficientes estructurales en términos de los coeficientes de la forma reducida. Las ecuaciones de la forma reducida se han expresado en (13.6a) como

$$y_t = \Pi x_t + v_t,$$

en la que y_t es el vector ($G \times 1$) de las variables endógenas, Π es la matriz ($G \times K$) de los coeficientes de la forma reducida, x_t es el vector ($K \times 1$) de las variables predeterminadas, y v_t es el vector ($G \times 1$) de las perturbaciones de la forma reducida. Como estas últimas son combinaciones lineales de las perturbaciones estructurales, cumplen todos los supuestos del modelo clásico de regresión normal lineal. Las variables predeterminadas, que constituyen las variables explicativas de las ecuaciones de la forma reducida, son no estocásticas o, en caso de ser estocásticas, son independientes de las perturbaciones en el mismo período, y supondremos también que sus variancias y covariancias son finitas cuando $T \rightarrow \infty$, y que no existe ninguna relación lineal exacta entre ellas. Si las variables predeterminadas incluyen variables endógenas retardadas, supondremos que los valores iniciales de estas variables son fijos. Bajo estos supuestos, la aplicación del método de los mínimos cuadrados ordinarios (sin restricciones) a cada ecuación de la forma reducida produce estimaciones consistentes de las π . Si todas las ecuaciones estructurales están exactamente identificadas, estas estimaciones serán equivalentes a las estimaciones maximo-verosímiles, y además de ser consistentes poseerán también las propiedades de eficiencia asintótica y normalidad asintótica. Las estimaciones minimocuadráticas ordinarias de los coeficientes de la forma reducida pueden utilizarse para obtener las correspondientes estimaciones de los coeficientes estructurales en la forma especificada en (13.14) y (13.15). Estas últimas reciben el nombre de *estimaciones minimocuadráticas indirectas*, y aparecen generalmente como funciones no lineales de las estimaciones de la forma reducida, teniendo además sus mismas propiedades asintóticas.

EJEMPLO. Para ilustrar la utilización del método de los mínimos cuadrados indirectos, usaremos el modelo de oferta y demanda presentado en (13.1):

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{1t} \quad (\text{demanda}).$$

$$Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{2t} \quad (\text{oferta}).$$

en el que Q_t y P_t son variables endógenas, e Y_t es exógena. Si no existe ninguna restricción sobre la matriz de variancias y covariancias de las perturbaciones...

don de demanda está subidentificada, pero la de oferta está exactamente identificada. Las ecuaciones de la forma reducida son

$$Q_t = \pi_{11} + \pi_{12} Y_t + v_{1t},$$

$$P_t = \pi_{21} + \pi_{22} Y_t + v_{2t}.$$

Sustituyendo Q_t y P_t por sus valores, según la forma reducida, en la ecuación de oferta, obtenemos

$$\pi_{11} + \pi_{12} Y_t + v_{1t} = \beta_1 + \beta_2(\pi_{21} + \pi_{22} Y_t + v_{2t}) + \varepsilon_{2t}.$$

Por tanto,

$$\pi_{11} = \beta_1 + \beta_2 \pi_{21},$$

$$\pi_{12} = \beta_2 \pi_{22},$$

o

$$\beta_2 = \frac{\pi_{12}}{\pi_{22}},$$

$$\beta_1 = \pi_{11} - \frac{\pi_{12}\pi_{21}}{\pi_{22}}.$$

Los estimadores minimocuadráticos ordinarios de los coeficientes de la forma reducida son

$$\hat{\pi}_{12} = \frac{m_{QY}}{m_{YY}},$$

$$\hat{\pi}_{11} = \bar{Q} - \left(\frac{m_{QY}}{m_{YY}} \right) \bar{Y},$$

$$\hat{\pi}_{22} = \frac{m_{PY}}{m_{YY}},$$

$$\hat{\pi}_{21} = \bar{P} - \left(\frac{m_{PY}}{m_{YY}} \right) \bar{Y},$$

siendo

$$m_{QY} = \sum_t (Q_t - \bar{Q})(Y_t - \bar{Y}),$$

$$m_{PY} = \sum_t (P_t - \bar{P})(Y_t - \bar{Y}),$$

$$m_{YY} = \sum_t (Y_t - \bar{Y})^2.$$

En consecuencia, los estimadores minimocuadráticos indirectos de β_1 y β_2 son

$$\hat{\beta}_2 = \frac{m_{QY}}{m_{PY}},$$

$$\hat{\beta}_1 = \bar{Q} - \left(\frac{m_{QY}}{m_{PY}} \right) \bar{P}.$$

El problema de la estimación de los coeficientes de una ecuación estructural puede resolverse también a través del *método de las variables instrumentales*. Supongamos que la ecuación estructural que queremos estimar es la primera ecuación del sistema, es decir,

$$(13.19) \quad \beta_{11}y_{1t} + \beta_{12}y_{2t} + \dots + \beta_{1G}y_{Gt} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \dots + \gamma_{1K}x_{Kt} = u_{1t}.$$

Esto no implica ninguna pérdida de generalidad, ya que las ecuaciones estructurales pueden escribirse en el orden que deseamos. Supongamos, además, que β_{11} es igual a la unidad y que las variables endógenas y predeterminadas incluidas son $y_{1t}, y_{2t}, \dots, y_{Gt}, x_{1t}, x_{2t}, \dots, x_{K*t}$. En este caso, podemos escribir (13.19) en la forma

$$(13.19a) \quad y_{1t} = -\beta_{12}y_{2t} - \beta_{13}y_{3t} - \dots - \beta_{1G}y_{Gt} - \gamma_{11}x_{1t} - \gamma_{12}x_{2t} - \dots - \gamma_{1K*}x_{K*t} + u_{1t}.$$

En notación matricial, la ecuación (13.19a) puede expresarse como

$$(13.20) \quad y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1,$$

en la que

$$y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{21} & y_{31} & \dots & y_{G1} \\ y_{22} & y_{32} & \dots & y_{G2} \\ \vdots & \vdots & & \vdots \\ y_{2T} & y_{3T} & \dots & y_{GT} \end{bmatrix}, \quad X_1 = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{K*1} \\ x_{12} & x_{22} & \dots & x_{K*2} \\ \vdots & \vdots & & \vdots \\ x_{1T} & x_{2T} & \dots & x_{K*T} \end{bmatrix},$$

$$(T \times 1) \quad (T \times (G^A - 1)) \quad (T \times K^*)$$

$$\beta_1 = \begin{bmatrix} -\beta_{12} \\ -\beta_{13} \\ \vdots \\ -\beta_{1G} \end{bmatrix}, \quad \gamma_1 = \begin{bmatrix} -\gamma_{11} \\ -\gamma_{12} \\ \vdots \\ -\gamma_{1K*} \end{bmatrix}.$$

$$(\overline{G^A - 1} \times 1) \quad (K^* \times 1)$$

De forma alternativa, podemos escribir (13.20) como

$$(13.20a) \quad y_1 = Z_1\alpha_1 + u_1,$$

$$\text{en la que} \quad Z_1 = [Y_1 \quad X_1] \quad y \quad \alpha_1 = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}.$$

El estimador por variables instrumentales de los coeficientes estructurales es

$$(13.21) \quad \alpha^{\dagger} = (W_1'Z_1)^{-1}(W_1'y_1),$$

y su matriz de variancias y covariancias asintóticas es

$$(13.22) \quad \text{Var-Cov asintóticas } (\alpha^{\dagger}) = s_{11}(W_1'Z_1)^{-1}(W_1'W_1)(Z_1'W_1)^{-1},$$

en la que W_1 es la matriz de orden $T \times (G^A - 1 + K^*)$ de los valores observados de las variables instrumentales elegidas (las fórmulas del estimador por variables instrumentales y de su variancia se han presentado en (11.86) y (11.87) de la sección 11.4). El parámetro σ_{11} puede estimarse de forma consistente mediante

$$(13.23) \quad s_{11} = \frac{(y_1 - Z_1\alpha_1^{\dagger})(y_1 - Z_1\alpha_1^{\dagger})'}{(T - G^A + 1 - K^*)}$$

(La consistencia del estimador no variará si, en lugar de $(T - G^A + 1 - K^*)$, utilizamos cualquier denominador D a condición de que $\lim \text{prob } (D/T) = 1$. Por lo que respecta a la elección de las variables instrumentales adecuadas, el problema se limita a las $(G^A - 1)$ variables endógenas incluidas, ya que las variables predeterminadas que aparecen en la ecuación pueden servir como sus propias variables instrumentales. Los candidatos naturales para la función de variables instrumentales «externas» son las variables predeterminadas que aparecen en el sistema pero no en la ecuación que queremos estimar. El número de estas variables es K^{**} , que en el caso de una ecuación estructural *exactamente identificada* es igual a $(G^A - 1)$. Llamemos X_1 a la matriz de los valores observados de las variables predeterminadas excluidas, es decir,

$$(13.24) \quad X_1 = \begin{bmatrix} x_{K^*+1,1} & x_{K^*+2,1} & \dots & x_{K1} \\ x_{K^*+1,2} & x_{K^*+2,2} & \dots & x_{K2} \\ \vdots & \vdots & & \vdots \\ x_{K^*+1,T} & x_{K^*+2,T} & \dots & x_{KT} \end{bmatrix}.$$

El orden en que aparecen estas variables no tiene ninguna importancia. Las variables instrumentales de la ecuación estructural están representadas, pues, por

$$(13.25) \quad W_1 = [X_1 \quad X_1].$$

Por tanto,

$$W_1'Z_1 = \begin{bmatrix} X_1' \\ X_1' \end{bmatrix} [Y_1 \quad X_1] = \begin{bmatrix} X_1'Y_1 & X_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix},$$

$$y \quad W_1'y_1 = \begin{bmatrix} X_1'y_1 \\ X_1'y_1 \end{bmatrix}.$$

En consecuencia, en el caso de identificación exacta, el estimador por variables instrumentales de los coeficientes estructurales es

$$(13.26) \quad \alpha_1^\dagger = \begin{bmatrix} \beta_1^\dagger \\ \gamma_1^\dagger \end{bmatrix} = \begin{bmatrix} X_1' Y_1 & X_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} X_1' y_1 \\ X_1' y_1 \end{bmatrix}.$$

Además, según (13.22) y (13.23), definiremos un estimador consistente de la matriz de variancias y covariancias de $\alpha_1^\dagger$ en la forma

$$(13.27) \quad \text{Est. Var-Cov}(\alpha_1^\dagger) = s_{11} \begin{bmatrix} X_1' Y_1 & X_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} X_1' X_1 & X_1' X_1 \\ X_1' X_1 & X_1' X_1 \end{bmatrix} \begin{bmatrix} Y_1' X_1 & Y_1' X_1 \\ X_1' X_1 & X_1' X_1 \end{bmatrix}^{-1},$$

en el que s_{11} se obtiene igual que en (13.23).

El estimador por variables instrumentales de la expresión (13.26) es, en realidad, *exactamente igual* al estimador minimocuadrático indirecto que se ha presentado anteriormente. No vamos a efectuar una demostración general de esta proposición, pero demostraremos su validez en el caso de la ecuación de oferta del modelo de oferta y demanda presentado en (13.1). Esta ecuación,

$$Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{2t},$$

está exactamente identificada y la variable predeterminada excluida es Y_t . En términos de la notación utilizada en la estimación por medio de variables instrumentales, tenemos

$$Z_1 = [P \quad 1],$$

$$W_1 = [Y \quad 1],$$

$$y_1 = [Q],$$

siendo

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_T \end{bmatrix}, \quad P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_T \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}, \quad 1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

(T × 1) (T × 1) (T × 1) (T × 1)

Por tanto,

$$W_1' Z_1 = \begin{bmatrix} Y' \\ 1' \end{bmatrix} [P \quad 1] = \begin{bmatrix} \sum_t P_t Y_t & \sum_t Y_t \\ \sum_t P_t & T \end{bmatrix}.$$

$$W_1' y_1 = \begin{bmatrix} Y' \\ 1' \end{bmatrix} [Q] = \begin{bmatrix} \sum_t Q_t Y_t \\ \sum_t Q_t \end{bmatrix}.$$

Así pues, según (13.21), tenemos

$$(13.28) \quad \begin{bmatrix} \beta_2^\dagger \\ \beta_1^\dagger \end{bmatrix} = \frac{1}{T \sum P_t Y_t - (\sum P_t)(\sum Y_t)} \begin{bmatrix} T & -\sum_t Y_t \\ -\sum_t P_t & \sum_t P_t Y_t \end{bmatrix} \begin{bmatrix} \sum_t Q_t Y_t \\ \sum_t Q_t \end{bmatrix}$$

$$= \begin{bmatrix} \frac{m_{QY}}{m_{PY}} \\ \bar{Q} - \left(\frac{m_{QY}}{m_{PY}} \right) \bar{P} \end{bmatrix}.$$

Obviamente, estas fórmulas son exactamente iguales que las fórmulas de la estimación por mínimos cuadrados indirectos.

La equivalencia entre estas dos formas de estimación nos permite utilizar la fórmula de las variables instrumentales para la matriz de variancias y covariancias asintóticas de los coeficientes estimados. Esto resulta de gran utilidad, ya que, de lo contrario, la obtención de la matriz de variancias y covariancias asintóticas del estimador minimocuadrático indirecto es muy compleja. Por ejemplo, consideremos la estimación de la matriz de variancias y covariancias de los coeficientes estimados de la ecuación de oferta presentada en (13.1). Efectuando las sustituciones adecuadas en (13.27), obtenemos

$$(13.29) \quad \text{Est. Var-Cov} \begin{bmatrix} \beta_2^\dagger \\ \beta_1^\dagger \end{bmatrix}$$

$$= s_{22} \begin{bmatrix} \sum_t P_t Y_t & \sum_t Y_t \\ \sum_t P_t & T \end{bmatrix}^{-1} \begin{bmatrix} \sum_t Y_t^2 & \sum_t Y_t \\ \sum_t Y_t & T \end{bmatrix} \begin{bmatrix} \sum_t P_t Y_t & \sum_t P_t \\ \sum_t Y_t & T \end{bmatrix}^{-1}$$

$$= \frac{s_{22}}{(T m_{PY})^2} \begin{bmatrix} T^2 m_{YY} & -T(\sum P_t) m_{YY} \\ -T(\sum P_t) m_{YY} & (\sum P_t)^2 m_{YY} + T m_{PY}^2 \end{bmatrix}$$

$$= s_{22} \begin{bmatrix} \left(\frac{m_{YY}}{m_{PY}^2} \right) & \left(\frac{-\bar{P} m_{YY}}{m_{PY}^2} \right) \\ \left(\frac{-\bar{P} m_{YY}}{m_{PY}^2} \right) & \left(\frac{\bar{P}^2 m_{YY}}{m_{PY}^2} + \frac{1}{T} \right) \end{bmatrix},$$

en la que

$$(13.30) \quad s_{22} = \frac{1}{T-2} \sum_t (Q_t - \beta_1^* - \beta_2^* P_t)^2$$

$$= \frac{1}{T-2} \left(\frac{m_{QQ}m_{PY}^2 - 2m_{QY}m_{QP}m_{PY} + m_{PP}m_{QY}^2}{m_{PY}^2} \right).$$

EJEMPLO. La ecuación de oferta del modelo (13.1) puede estimarse a partir de los datos anuales correspondientes al mercado de carne en los Estados Unidos en el período 1919-1941.³

Q_t = consumo de carne per cápita (en libras);

P_t = precios de la carne al por menor (índice 1935-1939 = 100);

Y_t = renta real disponible per cápita (en dólares).

Las medias muestrales de estas variables son

$$\bar{Q} = 166.1913,$$

$$\bar{P} = 92.3391,$$

$$\bar{Y} = 495.5652;$$

y las sumas de los cuadrados y productos cruzados de las desviaciones respecto de las medias muestrales son

$$m_{QQ} = 1,369.53826, \quad m_{PP} = 1,581.49478,$$

$$m_{QP} = -352.55217, \quad m_{PY} = 8,354.59130,$$

$$m_{QY} = 3,671.91304, \quad m_{YY} = 83,433.65217.$$

Las estimaciones de los coeficientes estructurales de la ecuación de oferta por el método de los mínimos cuadrados indirectos son las siguientes:

$$\hat{\beta}_2 = \frac{3,671.91304}{8,354.59130} = 0.43951,$$

$$\hat{\beta}_1 = 166.1913 - 0.43951 \times 92.3391 = 125.60749.$$

La estimación de la variancia del término de perturbación estructural, calculada según (13.30) es

$$s_{22} = \frac{1,984.9311}{21} = 94.5205.$$

3. Este ejemplo está tomado de G. Tintner, *Econometrics*. Nueva York, Wiley, 1963, p. 169.

Las variancias estimadas de las estimaciones minimocuadráticas indirectas se obtienen a partir de (13.29) en la forma siguiente:

$$\text{Est. Var}(\hat{\beta}_2) = 94.5205 \left(\frac{83,433.65217}{8,354.59130^2} \right) = 0.11298318,$$

$$\text{Est. Var}(\hat{\beta}_1) = 94.5205 \left(\frac{92.3391^2 \times 83,433.65217}{8,354.59130^2} + \frac{1}{23} \right) = 10.23547072.$$

Por tanto, el resultado final de la estimación por el método de los mínimos cuadrados indirectos es

$$Q_t = 125.60749 + 0.43951 P_t + \varepsilon_{2t}.$$

(3.19929) (0.33613)

Con fines de comparación, presentamos también las estimaciones minimocuadráticas ordinarias (inconsistentes) de la ecuación de oferta:

$$Q_t = 145.60676 - 0.22292 P_t + \varepsilon_{2t}.$$

La estimación de una ecuación estructural *exactamente identificada* en la que la identificación se logra a través de *restricciones sobre la matriz de variancias y covariancias* de las perturbaciones estructurales puede examinarse también en el contexto del modelo de oferta y demanda presentado en (13.1). La ecuación de demanda

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{1t}$$

está subidentificada cuando no se establece ninguna restricción la matriz de variancias y covariancias de las perturbaciones, pero, como se ha visto en la sección 13.2, si suponemos que ε_{1t} y ε_{2t} no están correlacionadas, se convierte en exactamente identificada. Las relaciones de identificación entre los coeficientes estructurales y los parámetros de la forma reducida son las siguientes. En primer lugar, sustituyendo Q_t y P_t por sus valores según las expresiones de la forma reducida, en la ecuación de demanda, obtenemos

$$(13.31) \quad \pi_{11} = \alpha_1 + \alpha_2 \pi_{21},$$

$$(13.32) \quad \pi_{12} = \alpha_2 \pi_{22} + \alpha_3.$$

A continuación, según (13.18), sabemos que

$$\psi_{11} - \alpha_2 \psi_{12} - \beta_2 \psi_{12} + \alpha_2 \beta_2 \psi_{22} = 0,$$

en la que las Ψ son los elementos de la matriz de variancias y covariancias de las perturbaciones de la forma reducida, y β_2 es un coeficiente estructural de la ecuación de oferta. Según los resultados correspondientes a la ecuación de oferta sabemos que

$$\beta_2 = \frac{\pi_{12}}{\pi_{22}},$$

y, por tanto, podemos escribir

$$(13.33) \quad \alpha_2 = \frac{\pi_{12}\psi_{12} - \pi_{21}\psi_{11}}{\pi_{12}\psi_{22} - \pi_{22}\psi_{12}}.$$

Las ecuaciones (13.31) (13.32) y (13.33) nos permiten expresar los coeficientes estructurales α_1 , α_2 , y α_3 en términos de las π y ψ de las ecuaciones de la forma reducida. Los parámetros de la forma reducida pueden estimarse aplicando las fórmulas de los mínimos cuadrados ordinarios, y a partir de las estimaciones de los parámetros de la forma reducida podemos obtener las estimaciones correspondientes de las α . Como en este caso las dos ecuaciones están exactamente identificadas, dichas estimaciones son máximoverosímiles y la matriz de sus variancias y covariancias asintóticas puede obtenerse a través de la matriz de información de la correspondiente función de verosimilitud. Obsérvese que la función de verosimilitud para las observaciones muestrales de Q_t y P_t en el modelo de oferta y demanda es

$$L = -T \log(2\pi) - \frac{T}{2} \log(\sigma_{11}\sigma_{22}) + T \log |\alpha_2 - \beta_2| - \frac{\sum_i (Q_i - \alpha_1 - \alpha_2 P_i - \alpha_3 Y_i)^2}{2\sigma_{11}} - \frac{\sum_i (Q_i - \beta_1 - \beta_2 P_i)^2}{2\sigma_{22}}.$$

EJEMPLO. Estimaremos la ecuación de demanda del modelo (13.1), con el supuesto de covariancia nula entre ε_{1t} y ε_{2t} , a partir de los datos correspondientes al mercado de carne de los Estados Unidos que hemos presentado en el ejemplo anterior. En primer lugar, obsérvese que los estimadores minimocuadráticos ordinarios de los coeficientes de la forma reducida son

$$\begin{aligned} \hat{\pi}_{12} &= \frac{m_{QY}}{m_{YY}}, \\ \hat{\pi}_{11} &= \bar{Q} - \hat{\pi}_{12} \bar{Y}, \\ \hat{\pi}_{22} &= \frac{m_{PY}}{m_{YY}}, \\ \hat{\pi}_{21} &= \bar{P} - \hat{\pi}_{22} \bar{Y}. \end{aligned}$$

Las variancias y covariancias de las perturbaciones de la forma reducida se estiman de la forma siguiente:

$$\begin{aligned} \hat{\psi}_{11} &= \frac{m_{QQ} - \hat{\pi}_{12}m_{QY}}{T} = \frac{m_{QQ}m_{YY} - m_{QY}^2}{Tm_{YY}}, \\ \hat{\psi}_{22} &= \frac{m_{PP} - \hat{\pi}_{22}m_{PY}}{T} = \frac{m_{PP}m_{YY} - m_{PY}^2}{Tm_{YY}}, \\ \hat{\psi}_{12} &= \frac{m_{QP} - \hat{\pi}_{12}m_{PY} - \hat{\pi}_{22}m_{QY} + \hat{\pi}_{12}\hat{\pi}_{22}m_{YY}}{T} = \frac{m_{QP}m_{YY} - m_{QY}m_{PY}}{Tm_{YY}}. \end{aligned}$$

Sustituyendo estos resultados en (13.33) y efectuando algunas simplificaciones, obtenemos

$$\begin{aligned} \hat{\alpha}_2 &= \frac{m_{QY}m_{QP} - m_{QO}m_{PY}}{m_{QY}m_{PP} - m_{QP}m_{PY}} \\ &= \frac{3,671.91304 \times (-352.55217) - 1,369.53826 \times 8,354.59130}{3,671.91304 \times 1,581.49478 - (-352.55217) \times 8,354.59130} = -1.4551. \end{aligned}$$

Además, según (13.31), tenemos

$$\hat{\alpha}_1 = \hat{\pi}_{11} - \hat{\alpha}_2 \hat{\pi}_{21} = 206.5374,$$

y, según (13.32),

$$\hat{\alpha}_3 = \hat{\pi}_{12} - \hat{\alpha}_2 \hat{\pi}_{22} = 0.1897.$$

Por tanto, la ecuación de demanda estimada es

$$Q_t = 206.5374 - 1.4551P_t + 0.1897Y_t + \hat{\varepsilon}_{2t}.$$

Según los supuestos del modelo, estas estimaciones son consistentes y asintóticamente eficientes. Para poder compararlasy, presentamos también las estimaciones minimocuadráticas ordinarias que no son consistentes:

$$Q_t = 185.8452 - 0.9739P_t + 0.1418Y_t + \varepsilon_{1t}.$$

Estimación por mínimos cuadrados en dos etapas

Para estimar una ecuación estructural *superidentificada* que pertenece a un sistema general interdependiente de ecuaciones simultáneas pueden utilizarse varios métodos distintos que proporcionan estimaciones consistentes. Probablemente, el más conocido de ellos es el método de los *mínimos cuadrados en dos etapas*. Supongamos que la ecuación estructural superidentificada es la primera ecuación del sistema:

$$(13.34) \quad y_1 = Y_1\beta_1 + X_1Y_1 + u_1,$$

en el que y_1 es el vector ($T \times 1$) de la variable endógena cuyo coeficiente en la primera ecuación es la unidad, Y_1 es la matriz de orden $T \times (G^* - 1)$ de las demás variables endógenas de la primera ecuación, X_1 es la matriz ($T \times K^*$) de las variables predeterminadas de la primera ecuación, y u_1 es el vector ($T \times 1$) de las perturbaciones de esta ecuación.

Una forma alternativa de escribir (13.34) es

$$(13.34 \text{ a}) \quad y_1 = Z_1\alpha_1 + u_1,$$

en la que $Z_1 = [Y_1 \ X_1]$ y $\alpha_1 = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}$.

Podemos particionar la matriz Y_1 en la forma

$$Y_1 = [y_2 \ y_3 \ \dots \ y_{G^A}],$$

en la que cada una de las y es un vector de orden $T \times 1$. Las ecuaciones de la forma reducida para estas variables son

$$y_2 = X\pi_2 + v_2,$$

$$y_3 = X\pi_3 + v_3,$$

$$\vdots$$

$$y_{G^A} = X\pi_{G^A} + v_{G^A},$$

en las que X es la matriz $(T \times K)$ de todas las variables predeterminadas del sistema, cada una de las π representa un vector $(K \times 1)$ de los coeficientes de la forma reducida correspondientes, y cada una de las V es un vector $(T \times 1)$ de las correspondientes perturbaciones de la forma reducida. Definamos ahora

$$V_1 = [v_2 \ v_3 \ \dots \ v_{G^A}]$$

$$y \quad Y_1 - V_1 = [X\pi_2 \ X\pi_3 \ \dots \ X\pi_{G^A}].$$

Por tanto, podemos escribir (13.34) en la forma

$$(13.35) \quad y_1 = (Y_1 - V_1)\beta_1 + X_1\gamma_1 + (u_1 + V_1\beta_1).$$

Como $(Y_1 - V_1)$ sólo depende de X y no incluye ninguna perturbación, no está correlacionada con $(u_1 + V_1\beta_1)$ y, por tanto, aplicando el método de los mínimos cuadrados ordinarios a (13.35), obtendríamos estimaciones consistentes de β_1 y γ_1 . La dificultad consiste en que V_1 y, por tanto $(Y_1 - V_1)$, no es observable, pero, sin embargo, podemos sustituir V_1 por los correspondientes residuos minimocuadráticos de la forma reducida y utilizar

$$\hat{Y}_1 - \hat{V}_1 = \hat{Y}_1 = [X\hat{\pi}_2 \ X\hat{\pi}_3 \ \dots \ X\hat{\pi}_{G^A}].$$

Puede verse claramente que

$$\lim \text{prob } (Y_1 - \hat{V}_1) = [X\pi_2 \ X\pi_3 \ \dots \ X\pi_{G^A}] = Y_1 - V_1,$$

de modo que $(Y_1 - V_1)$ y $(u_1 + V_1\beta_1)$ no están correlacionadas asintóticamente. Por tanto, si aplicamos el método de los mínimos cuadrados ordinarios a

$$(13.36) \quad y_1 = \hat{Y}_1\beta_1 + X_1\gamma_1 + u_1^*,$$

$$\text{en la que} \quad u_1^* = u_1 + \hat{V}_1\beta_1.$$

obtendremos estimaciones consistentes de β_1 y γ_1 . Estas estimaciones son las denominadas estimaciones «minimocuadráticas en dos etapas», ya que podemos considerar que el procedimiento de estimación consiste en dos aplicaciones sucesivas del método de los mínimos cuadrados ordinarios. En la primera etapa se estiman las ecuaciones de la forma reducida para $y_2, y_3, \dots, y_{G^A}$ y se calculan los valores ajustados de estas variables, mientras que en la segunda etapa se aplica el método de los mínimos cuadrados a las ecuaciones (13.36) que utilizan como variables explicativas estos valores ajustados.

Por analogía con (13.34a), la ecuación (13.36) podría escribirse también en la forma

$$(13.36a) \quad y_1 = Z_1\alpha_1 + u_1^*,$$

$$\text{en la que} \quad Z_1 = [\hat{Y}_1 \ X_1].$$

Aplicando el método de los mínimos cuadrados a (13.36a), se obtiene

$$(13.37) \quad \tilde{\alpha}_1 = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\gamma}_1 \end{bmatrix} = (\hat{Z}'\hat{Z})^{-1}(\hat{Z}'y_1) = \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1'y_1 \\ X_1'y_1 \end{bmatrix}.$$

(En lo que sigue utilizaremos el hecho de que los residuos minimocuadráticos ordinarios son ortogonales respecto al valor ajustado de la variable dependiente y a todas las variables explicativas, lo que implica que $V_1'Y_1 = 0$ y $V_1'X_1 = 0$). Como

$$\hat{Y}_1'\hat{Y}_1 = (Y_1 - \hat{V}_1)'(Y_1 - \hat{V}_1) = Y_1'Y_1 - \hat{V}_1'Y_1 - Y_1'\hat{V}_1 + \hat{V}_1'\hat{V}_1$$

$$= Y_1'Y_1 + \hat{V}_1'(Y_1 + \hat{V}_1) - (\hat{V}_1 + \hat{V}_1)'\hat{V}_1 + \hat{V}_1'\hat{V}_1$$

$$= Y_1'Y_1 - \hat{V}_1'\hat{V}_1 - \hat{V}_1'\hat{V}_1 + \hat{V}_1'\hat{V}_1 = Y_1'Y_1 - \hat{V}_1'\hat{V}_1,$$

$$\hat{Y}_1'X_1 = (Y_1 - \hat{V}_1)'X_1 = Y_1'X_1,$$

$$X_1'\hat{Y}_1 = X_1'(Y_1 - \hat{V}_1) = X_1'Y_1,$$

$$y \quad \hat{Y}_1'y_1 = (Y_1 - \hat{V}_1)'y_1 = Y_1'y_1 - \hat{V}_1'y_1,$$

la ecuación (13.37) puede escribirse en la forma

$$(13.38) \quad \tilde{\alpha}_1 = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\gamma}_1 \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - \hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'y_1 - \hat{V}_1'y_1 \\ X_1'y_1 \end{bmatrix}.$$

Esta es la forma en que se presenta generalmente la fórmula del estimador minimocuadrático en dos etapas, ya que en ella se puede ver claramente la diferencia entre este estimador y el estimador minimocuadrático ordinario (inconsistente) que viene dado por

$$(13.39) \quad \hat{\alpha}_1 = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'y_1 \\ X_1'y_1 \end{bmatrix}.$$

El estimador minimocuadrático en dos etapas puede presentarse también como un *estimador por variables instrumentales*, en el que $\hat{Y}_1$ sirve como instrumento para Y_1 , y X_1 se utiliza como su propio instrumento. Esto puede mostrarse en la forma siguiente. Según (13.21), el estimador por variables instrumentales de los coeficientes de la primera ecuación estructural es

$$\alpha_1' = (W_1' Z_1)^{-1} (W_1' y_1),$$

en el que W_1 es la matriz de las variables instrumentales. Haciendo

$$W_1 = [\hat{Y}_1 \quad X_1]$$

obtenemos

$$W_1' Z_1 = \begin{bmatrix} \hat{Y}_1' \\ X_1' \end{bmatrix} [Y_1 \quad X_1] = \begin{bmatrix} \hat{Y}_1' Y_1 & \hat{Y}_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix},$$

y

$$W_1' y_1 = \begin{bmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{bmatrix}.$$

Pero

$$\hat{Y}_1' Y_1 = (Y_1 - \hat{v}_1)' Y_1 = Y_1' Y_1 - \hat{v}_1' (Y_1 + \hat{v}_1) = Y_1' Y_1 - \hat{v}_1' \hat{v}_1,$$

y, en base a resultados anteriores,

$$\hat{Y}_1' X_1 = Y_1' X_1,$$

$$\hat{Y}_1' y_1 = Y_1' y_1 - \hat{v}_1' y_1.$$

Por tanto,

$$(13.40) \quad \alpha_1' = \begin{bmatrix} Y_1' Y_1 - \hat{v}_1' \hat{v}_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1' y_1 - \hat{v}_1' y_1 \\ X_1' y_1 \end{bmatrix},$$

que es precisamente la misma fórmula que la de $\tilde{\alpha}_1$ presentada en (13.38). Debido a la equivalencia entre α_1' y $\tilde{\alpha}_1$ podemos utilizar la fórmula (13.22) para calcular la matriz de variancias y covariancias asintóticas del estimador minimocuadrático en dos etapas. En este caso, obtenemos

$$(13.41) \quad \begin{aligned} \text{Var-Cov asintóticas}(\tilde{\alpha}_1) &= s_{11} (W_1' Z_1)^{-1} (W_1' W_1) (Z_1' W_1)^{-1} = s_{11} (W_1' Z_1)^{-1} \\ &= s_{11} \begin{bmatrix} Y_1' Y_1 - \hat{v}_1' \hat{v}_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1}. \end{aligned}$$

en la que $\sigma_{11} = \text{Var}(u_{1t})$. Según (13.23), un estimador consistente de σ_{11} es

$$(13.42) \quad s_{11} = \frac{(y_1 - Y_1 \tilde{\beta}_1 - X_1 \tilde{\gamma}_1)' (Y_1 - Y_1 \tilde{\beta}_1 - X_1 \tilde{\gamma}_1)}{T - G^* + 1 - K^*}.$$

El estimador minimocuadrático en dos etapas es consistente, pero en general no es asintóticamente eficiente, porque no tiene en cuenta la correlación entre los términos de perturbación estructurales de las distintas ecuaciones. Sin embargo, para su construcción no es necesario conocer completamente todo el sistema, y lo único que se requiere es una enumeración de todas las variables predeterminadas y de sus valores muestrales.

Si se aplica el método de los mínimos cuadrados en dos etapas a una ecuación exactamente identificada, las estimaciones resultantes son iguales a las obtenidas por el método de los mínimos cuadrados indirectos. Esto puede demostrarse de la forma siguiente. El estimador minimocuadrático en dos etapas puede expresarse como

$$\tilde{\alpha}_1 = (W_1' Z_1)^{-1} (W_1' y_1),$$

en el que

$$W_1 = [\hat{Y}_1 \quad X_1] \quad \text{y} \quad Z_1 = [Y_1 \quad X_1].$$

Ahora podemos escribir

$$(13.43) \quad \hat{Y}_1 = X_1 A,$$

en la que X_1 es la matriz ($T \times K^{**}$) de las variables predeterminadas excluidas definida en (13.24), y A es la matriz ($K^{**} \times K^*$) que se define como

$$A = (X_1' X_1)^{-1} (X_1' Y_1).$$

Naturalmente, la ecuación a estimar tiene que estar exactamente identificada, es decir, es preciso que $K^{**} = G^* - 1$, ya que, de lo contrario, la dimensión de $(X_1' A)$ no sería igual que la de Y_1 . Puede establecerse la validez de (13.43) sustituyendo A por su valor y premultiplicando ambos lados por X_1' :

$$X_1' \hat{Y}_1 = X_1' X_1 (X_1' X_1)^{-1} (X_1' Y_1) = X_1' Y_1 = X_1' (\hat{Y}_1 + \hat{v}_1) = X_1' \hat{Y}_1.$$

Sustituyendo Y_1 por su valor según (13.43) en la fórmula de $\tilde{\alpha}_1$ obtenemos

$$(13.44) \quad \begin{aligned} \tilde{\alpha}_1 &= (W_1' Z_1)^{-1} (W_1' y_1) = \begin{bmatrix} A' X_1' Y_1 & A' X_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} A' X_1' y_1 \\ X_1' y_1 \end{bmatrix} \\ &= \left\{ \begin{bmatrix} A' & 0 \\ 0 & I_{K^*} \end{bmatrix} \begin{bmatrix} X_1' Y_1 & X_1' X_1 \end{bmatrix} \right\}^{-1} \begin{bmatrix} A' & 0 \\ 0 & I_{K^*} \end{bmatrix} \begin{bmatrix} X_1' y_1 \\ X_1' y_1 \end{bmatrix} \\ &= \begin{bmatrix} X_1' Y_1 & X_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} X_1' y_1 \\ X_1' y_1 \end{bmatrix}. \end{aligned}$$

Esta expresión de α_1 es exactamente igual a la del estimador minimocuadrático indirecto presentado en (13.26).

EJEMPLO. A continuación se presenta un modelo simplificado para explicar las variaciones en los precios y consumo de alimentos:

$$(13.45) \quad \begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 D_t + u_{1t} & (\text{demanda}), \\ Q_t &= \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + u_{2t} & (\text{oferta}). \end{aligned}$$

En este modelo, tenemos Q_t = consumo de alimentos per cápita, P_t = razón entre los precios de los alimentos y los precios generales de consumo, D_t = renta disponible a precios constantes, F_t = razón entre los precios recibidos por los agricultores por sus productos y los precios generales de consumo, referidos ambos al año anterior, y A_t = tiempo en años. Las variables Q_t y P_t son endógenas, mientras que D_t , F_t y A_t son predeterminadas. La ecuación de demanda está supridentificada y la de oferta exactamente identificada. En lugar de estimar este modelo a partir de los datos reales, *simularemos* las observaciones muestrales correspondientes a los precios y las cantidades estableciendo previamente los valores de los parámetros, especificando los valores muestrales de las variables predeterminadas y obteniendo los valores de las perturbaciones de forma aleatoria a partir de una población normal. Utilizando esta muestra artificial, podemos probar distintos métodos y comparar las estimaciones de los parámetros con los valores verdaderos. Concretamente, supongamos que el modelo verdadero es

$$(13.45a) \quad \begin{aligned} Q_t &= 96,5 - 0,25P_t + 0,30D_t + u_{1t} & (\text{demanda}), \\ Q_t &= 62,5 + 0,15P_t + 0,20F_t + 0,36A_t + u_{2t} & (\text{oferta}). \end{aligned}$$

La verdadera forma reducida de este modelo es

$$(13.46) \quad \begin{aligned} Q_t &= 75,25 + 0,1125D_t + 0,125F_t + 0,225A_t + v_{1t}, \\ P_t &= 85,00 + 0,75D_t - 0,50F_t - 0,90A_t + v_{2t}. \end{aligned}$$

Supongamos ahora que obtenemos dos grupos de 20 valores de desviaciones aleatorias $N(0,1)$ a partir de la tabla del Apéndice D. Llamemos ε_{1t} a los valores del primer grupo, y ε_{2t} a los del segundo ($t = 1, 2, \dots, 20$). A continuación, se obtienen los valores de las perturbaciones de la forma reducida según

$$\begin{aligned} v_{1t} &= 2\varepsilon_{1t}, \\ v_{2t} &= -0,5 v_{1t} + \varepsilon_{2t}. \end{aligned}$$

Esto implica la siguiente matriz de variancias y covariancias de las perturbaciones de la forma reducida:

$$\Psi = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}.$$

como en este modelo

$$\begin{aligned} u_{1t} &= v_{1t} + 0,25v_{2t}, \\ u_{2t} &= v_{1t} - 0,15v_{2t}, \end{aligned}$$

la matriz de variancias y covariancias de las perturbaciones estructurales es

$$\Phi = \begin{bmatrix} 3,125 & 3,725 \\ 3,725 & 4,645 \end{bmatrix}.$$

Naturalmente, la elección de los valores de los elementos de Ψ es arbitraria. Los valores muestrales de D_t y F_t se han tomado de un artículo de Girschick y Haavelmo⁴, y representan los valores reales correspondientes a los Estados Unidos durante los años 1922-1941, expresados en términos de números índices en los que el promedio para 1935-1939 = 100. Los valores de P_t y Q_t se calculan según (13.46). Los datos muestrales resultantes se presentan en el cuadro 13-1. Las estimaciones minimocuadráticas ordinarias de las ecuaciones de la forma reducida son

$$\begin{aligned} Q_t &= 72,2778 + 0,1126D_t + 0,1646F_t + 0,1648A_t + \hat{\varepsilon}_{1t}, & R^2 &= 0,7076, \\ P_t &= 87,3149 + 0,7020D_t - 0,5206F_t - 0,5209A_t + \hat{\varepsilon}_{2t}, & R^2 &= 0,8719. \end{aligned}$$

Podemos comparar estas estimaciones con las verdaderas ecuaciones de la forma reducida expresadas en (13.45). Aplicando el método de los mínimos cuadrados en dos etapas a las ecuaciones estructurales de (13.45) obtenemos

$$\begin{aligned} Q_t &= 94,6333 - 0,2436 P_t + 0,3140 D_t + u_{1t} & (\text{demanda}), \\ & (7,9208) \quad (0,0965) \quad (0,0469) \\ Q_t &= 49,5324 + 0,2401 P_t + 0,2556 F_t + 0,2529 A_t + u_{2t} & (\text{oferta}), \\ & (12,0105) \quad (0,0999) \quad (0,0472) \quad (0,0996) \end{aligned}$$

Las cifras entre paréntesis son los errores standard estimados. Estos resultados pueden compararse con las verdaderas ecuaciones estructurales (13.45a). También es interesante presentar las estimaciones minimocuadráticas ordinarias (inconsistentes) de las dos ecuaciones estructurales:

$$\begin{aligned} Q_t &= 99,8954 - 0,3163 P_t + 0,3346 D_t + u_{1t} & (\text{demanda}), \\ & R^2 = 0,7638. \\ Q_t &= 58,2754 + 0,1604 P_t + 0,2481 F_t + 0,2483 A_t + u_{2t} & (\text{oferta}), \\ & R^2 = 0,6548. \end{aligned}$$

4. M. A. Girschick y T. Haavelmo, «Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations», *Econometrica*, vol. 15, abril 1947, pp. 79-110; reproducido en W. C. Hood y T. C. Koopmans (eds.), *Studies in Econometric Method*. Nueva York, Wiley, 1953.

Cuadro 13.1

Q_i	P_i	D_i	F_i	A_i
98.485	100.323	87.4	98.0	1
99.187	104.264	97.6	99.1	2
102.163	103.435	96.7	99.1	3
101.504	104.506	98.2	98.1	4
104.240	98.001	99.8	110.8	5
103.243	99.456	100.5	108.2	6
103.993	101.066	103.2	105.6	7
99.900	104.763	107.8	109.8	8
100.350	96.446	96.6	108.7	9
102.820	91.228	88.9	100.6	10
95.435	93.085	75.1	81.0	11
92.424	98.801	76.9	68.6	12
94.535	102.908	84.6	70.9	13
98.757	98.756	90.6	81.4	14
105.797	95.119	103.1	102.3	15
100.225	98.451	105.1	105.0	16
103.522	86.498	96.4	110.5	17
99.929	104.016	104.4	92.5	18
105.223	105.769	110.7	89.3	19
106.232	113.490	127.1	93.0	20

Estimadores de clase k

La fórmula del estimador minimocuadrático en dos etapas presentada en (13.31) se puede generalizar para abarcar una clase completa de estimadores distintos. Estos estimadores, llamados *estimadores de clase k* , se definen como

$$(13.47) \quad \tilde{\alpha}_1^{(k)} = \begin{bmatrix} \tilde{\beta}_1^{(k)} \\ \tilde{\gamma}_1^{(k)} \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'y_1 - k\hat{V}_1'y_1 \\ X_1'y_1 \end{bmatrix}.$$

Esta fórmula se distingue solamente de la (13.38) por el hecho de que V_1V_1' y V_1y_1 están multiplicadas por el escalar k . El valor de este escalar se puede establecer a priori o se puede calcular a partir de las observaciones muestrales, según alguna regla. Hasta ahora hemos encontrado ya dos estimadores de clase k , el estimador minimocuadrático en dos etapas en el que $k = 1$, y el estimador minimocuadrático ordinario [expresado en (13.39)] en el que $k = 0$. El hecho de que el estimador minimocuadrático ordinario pertenezca a la

clase k indica que no todos los estimadores de esta clase son consistentes. Para descubrir los valores de k que producen estimaciones consistentes, señalemos en primer lugar que la ecuación (13.47) se puede expresar formalmente como la fórmula de un *estimador por variables instrumentales*. Esto puede mostrarse en la forma siguiente. Según (13.21), el estimador por variables instrumentales de la primera ecuación estructural se define como

$$\text{en la que} \quad \alpha_1^* = (W_1'Z_1)^{-1}(W_1'y_1), \\ Z_1 = [Y_1 \quad X_1],$$

igual que en el caso anterior. Si elegimos como variables instrumentales

$$W_1 = [(Y_1 - k\hat{V}_1) \quad X_1],$$

tenemos

$$\alpha_1^* = \begin{bmatrix} (Y_1 - k\hat{V}_1)'Y_1 & (Y_1 - k\hat{V}_1)'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - k\hat{V}_1)'y_1 \\ X_1'y_1 \end{bmatrix}.$$

Sin embargo,

$$(Y_1 - k\hat{V}_1)'Y_1 = Y_1'Y_1 - k\hat{V}_1'(\hat{V}_1 + V_1) = Y_1'Y_1 - k\hat{V}_1'\hat{V}_1$$

$$(Y_1 - k\hat{V}_1)'X_1 = Y_1'X_1 - k\hat{V}_1'X_1 = Y_1'X_1,$$

$$(Y_1 - k\hat{V}_1)'y_1 = Y_1'y_1 - k\hat{V}_1'y_1;$$

y, por tanto,

$$(13.48) \quad \alpha_1^* = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'y_1 - k\hat{V}_1'y_1 \\ X_1'y_1 \end{bmatrix},$$

que es exactamente igual que $\tilde{\alpha}_1^{(k)}$ en (13.47). Sin embargo, W_1 no se puede considerar como un conjunto de variables instrumentales *a menos que* sepamos que estas variables no están asintóticamente correlacionadas con la perturbación estructural u_1 , es decir, a menos que

$$\lim \text{prob} \frac{1}{T} W_1'u_1 = 0.$$

Es decir, es preciso que

$$(13.49) \quad \lim \text{prob} \frac{1}{T} (Y_1 - k\hat{V}_1)'u_1 = 0,$$

$$(13.50) \quad \lim \text{prob} \frac{1}{T} X_1'u_1 = 0.$$

La condición (13.50) se satisface siempre, ya que, por definición, las variables predeterminadas no están correlacionadas con las perturbaciones estructurales. Por lo que respecta a (13.49), tenemos

(13.49a)

$$\begin{aligned}\lim \text{prob } \frac{1}{T} (Y_1 - k\hat{V}_1)'u_1 &= \lim \text{prob } \frac{1}{T} Y_1' u_1 - \lim \text{prob } k \cdot \lim \text{prob } \hat{V}_1' u_1 \\ &= \lim \text{prob } \frac{1}{T} (Y_1 + \hat{V}_1)' u_1 - \lim \text{prob } k \cdot \lim \text{prob } \hat{V}_1' u_1 \\ &= \lim \text{prob } \frac{1}{T} \hat{V}_1' u_1 - \lim \text{prob } k \cdot \lim \text{prob } \hat{V}_1' u_1.\end{aligned}$$

Como $\lim \text{prob } (V_1' u_1)/T$ es un número finito distinto de cero, podemos ver que

$$\lim \text{prob } \frac{1}{T} (Y_1 - k\hat{V}_1)' u_1 = 0$$

si, y solamente si,

(13.51)

$$\lim \text{prob } k = 1.$$

En el caso de la estimación por mínimos cuadrados en dos etapas, esta condición se cumple automáticamente, ya que k es *siempre* igual a la unidad. Además, puede mostrarse que la matriz de variancias y covariancias asintóticas de todos los estimadores de clase k en los que

(13.52)

$$\lim \text{prob } \sqrt{T}(k-1) = 0$$

se reduce a (13.41).⁵ Lo que significa que todos estos estimadores deben tener la misma matriz de variancias y covariancias asintóticas que el estimador mínimocuadrático en dos etapas. Esta matriz se estima generalmente de la forma siguiente:

$$(13.53) \quad \text{Est. Var-Cov}(\tilde{a}^{(k)}) = s_{11} \begin{bmatrix} Y_1' Y_1 - k\hat{V}_1' \hat{V}_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1}$$

5. Ver H. Theil, *Economic Forecasts and Policy*, 2.^a ed., Amsterdam, North-Holland Publishing Company, 1961, p. 232.

en la que

$$(13.54) \quad s_{11} = \frac{(y_1 - Z_1 \tilde{a}_1^{(k)})'(y_1 - Z_1 \tilde{a}_1^{(k)})}{T - G^A + 1 - K^*}.$$

Naturalmente, esto equivale asintóticamente a (13.41).

Estimación maximoverosímil con información limitada

Otro de los estimadores para una sola ecuación es el denominado *estimador maximoverosímil con información limitada*, que pertenece también a la familia de la clase k y es consistente. Este estimador se obtiene maximizando la función de verosimilitud para las observaciones de las variables endógenas incluidas en la ecuación que pretendemos estimar. La frase «información limitada» significa que, al establecer la función de verosimilitud, nos limitamos a las variables endógenas que aparecen en la ecuación estudiada y prescindimos de las restricciones de identificación sobre las demás ecuaciones estructurales. Supongamos que esta ecuación es nuevamente la primera del sistema presentado en (13.34):

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + u_1,$$

en la que

$$Y_1 = [y_2 \ y_3 \ \dots \ y_{G^A}].$$

Por tanto, las variables endógenas incluidas son $y_{1t}, y_{2t}, \dots, y_{G^A t}$. Las ecuaciones de la forma reducida para estas variables son

$$y_{1t} = \pi_{11} x_{1t} + \pi_{12} x_{2t} + \dots + \pi_{1K} x_{Kt} + v_{1t},$$

$$y_{2t} = \pi_{21} x_{1t} + \pi_{22} x_{2t} + \dots + \pi_{2K} x_{Kt} + v_{2t},$$

⋮

$$y_{G^A t} = \pi_{G^A 1} x_{1t} + \pi_{G^A 2} x_{2t} + \dots + \pi_{G^A K} x_{Kt} + v_{G^A t}.$$

Como $v_{1t}, v_{2t}, \dots, v_{G^A t}$ son perturbaciones estocásticas que se distribuyen normalmente, su distribución conjunta es⁶

$$f(v_{1t}, v_{2t}, \dots, v_{G^A t}) = (2\pi)^{-(G^A/2)} |\Psi_\Delta|^{-1/2} e^{-(1/2) v_\Delta' \Psi_\Delta^{-1} v_\Delta}$$

en la que

$$v_\Delta = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{G^A t} \end{bmatrix} \quad y \quad \Psi_\Delta = E(v_\Delta v_\Delta') = \begin{bmatrix} \psi_{11} & \psi_{12} & \dots & \psi_{1G^A} \\ \psi_{21} & \psi_{22} & \dots & \psi_{2G^A} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{G^A 1} & \psi_{G^A 2} & \dots & \psi_{G^A G^A} \end{bmatrix}.$$

6. A. S. Goldberger, *Econometric Theory*, Nueva York, Wiley, 1964, p. 104. Si las x son estocásticas, la distribución conjunta de las v se debe considerar condicionada a los valores muestrales de las x .

Según los supuestos clásicos acerca de las perturbaciones, la distribución conjunta de $v_{1t}, v_{2t}, \dots, v_{Gt}$ para todas las observaciones muestrales es

$$(13.55) \quad f(v_{11}, v_{12}, \dots, v_{1T}, v_{21}, v_{22}, \dots, v_{G^{\Delta}T}) = (2\pi)^{-(TG^{\Delta}/2)} |\Psi_{\Delta}|^{-(T/2)} e^{-\frac{1}{2} v_{\Delta}' \Psi_{\Delta}^{-1} v_{\Delta}}$$

Queremos obtener la función de verosimilitud para las observaciones muestrales de las G^{Δ} variables endógenas incluidas en la primera ecuación estructural. Esta es la llamada *función de verosimilitud con información limitada* y puede obtenerse a partir de (13.55), expresando los elementos de $v_{\Delta t}$ en términos de las y . Además, sabemos que estas últimas están linealmente relacionadas a través de la ecuación estructural que estamos estimando. Esto se tiene en cuenta maximizando la función de verosimilitud con información limitada, *sujeta* a las restricciones de identificación de esta ecuación. Estas restricciones, que proceden de (13.10), son

$$(13.56) \quad \begin{aligned} \pi_{11} &= -\beta_{12}\pi_{21} - \beta_{13}\pi_{31} - \dots - \beta_{1G^{\Delta}}\pi_{G^{\Delta}1} - \gamma_{11}, \\ \pi_{12} &= -\beta_{12}\pi_{22} - \beta_{13}\pi_{32} - \dots - \beta_{1G^{\Delta}}\pi_{G^{\Delta}2} - \gamma_{12}, \\ &\vdots \\ \pi_{1K^*} &= -\beta_{12}\pi_{2K^*} - \beta_{13}\pi_{3K^*} - \dots - \beta_{1G^{\Delta}}\pi_{G^{\Delta}K^*} - \gamma_{1K^*}, \\ \pi_{1, K^*+1} &= -\beta_{12}\pi_{2, K^*+1} - \beta_{13}\pi_{3, K^*+1} - \dots - \beta_{1G^{\Delta}}\pi_{G^{\Delta}, K^*+1}, \\ &\vdots \\ \pi_{1K} &= -\beta_{12}\pi_{2K} - \beta_{13}\pi_{3K} - \dots - \beta_{1G^{\Delta}}\pi_{G^{\Delta}K}. \end{aligned}$$

Podemos introducir estas restricciones en la función de verosimilitud con información limitada escribiendo

$$\begin{aligned} v_{1t} &= y_{1t} - \pi_{11}x_{1t} - \dots - \pi_{1K}x_{Kt}, \\ v_{2t} &= y_{2t} - \pi_{21}x_{1t} - \dots - \pi_{2K}x_{Kt}, \\ &\vdots \\ v_{G^{\Delta}t} &= y_{G^{\Delta}t} - \pi_{G^{\Delta}1}x_{1t} - \dots - \pi_{G^{\Delta}K}x_{Kt}, \end{aligned}$$

y sustituyendo después $\pi_{11}, \pi_{12}, \dots, \pi_{1K}$ por las expresiones del lado derecho de (13.56). A continuación, la función resultante puede maximizarse respecto a los parámetros desconocidos, que incluyen los coeficientes estructurales.

La obtención del estimador máximoverosímil con información limitada a través de la maximización de la función de verosimilitud con restricciones es muy complicada, pero, sin embargo, pueden obtenerse también los mismos resultados utilizando el llamado *principio de la razón mínima de variancias*. La equivalencia entre estos dos enfoques es ampliamente conocida y no la va-

mos a demostrar aquí.⁷ Las estimaciones por el método de la razón mínima de variancias pueden obtenerse de la forma siguiente. La primera ecuación estructural

$$y_1 = X_1\beta_1 + X_1\gamma_1 + u_1$$

puede escribirse en la forma

$$(13.57) \quad \tilde{y}_1 = X_1\gamma_1 + u_1,$$

en la que

$$\tilde{y}_1 = y_1 - X_1\beta_1.$$

La variable «compuesta» $\tilde{y}_1$ representa una combinación lineal de las variables endógenas incluidas en la primera ecuación estructural. Si se pudiesen observar sus valores, se podría estimar γ_1 por el método de los mínimos cuadrados ordinarios. El resultado sería

$$\hat{\gamma}_1 = (X_1'X_1)^{-1}X_1'\tilde{y}_1,$$

y la suma de los cuadrados de los residuos sería

$$SCE_1 = (\tilde{y}_1 - X_1\hat{\gamma}_1)'(\tilde{y}_1 - X_1\hat{\gamma}_1) = \tilde{y}_1'\tilde{y}_1 - \tilde{y}_1'X_1(X_1'X_1)^{-1}X_1'\tilde{y}_1.$$

Obsérvese que en (13.57) las variables explicativas son las variables predeterminadas incluidas en la primera ecuación estructural. Si se ampliase este conjunto para incluir *todas* las variables predeterminadas del sistema tendríamos

$$(13.58) \quad \tilde{y}_1 = X\gamma + u_1,$$

en la que X es la matriz de orden $(T \times K)$ de todas las variables predeterminadas y

$$\gamma = \begin{bmatrix} \gamma_1 \\ 0_{**} \end{bmatrix} \quad (K \times 1)$$

Si aplicásemos el método de los mínimos cuadrados ordinarios a (13.58), ignorando el hecho de que el verdadero valor de algunos de los coeficientes γ es cero, obtendríamos

$$\hat{\gamma} = (X'X)^{-1}(X'\tilde{y}_1),$$

y la suma de los cuadrados de los residuos sería

$$SCE = (\tilde{y}_1 - X\hat{\gamma})'(\tilde{y}_1 - X\hat{\gamma}) = \tilde{y}_1'\tilde{y}_1 - \tilde{y}_1'X(X'X)^{-1}X'\tilde{y}_1.$$

7. Ver T. C. Koopmans y W. C. Hood, «The Estimation of Simultaneous Linear Economic Relationships», en Hood y Koopmans, *op. cit.*, pp. 166-77.

Como la adición de nuevas variables explicativas no puede aumentar nunca la suma de los cuadrados de los residuos, la razón

$$(13.59) \quad \ell = \frac{SCE_1}{SCE} = \frac{\tilde{y}_1' \tilde{y}_1 - \tilde{y}_1' X_1 (X_1' X_1)^{-1} X_1' \tilde{y}_1}{\tilde{y}_1' \tilde{y}_1 - \tilde{y}_1' X (X' X)^{-1} X' \tilde{y}_1}$$

no puede ser nunca menor que la unidad.

El problema consiste ahora en estimar los elementos de β_1 . Escribamos

$$(13.60) \quad \tilde{y}_1 = Y_{1\Delta} \beta_{1\Delta},$$

en la que $Y_{1\Delta} = [y_1 \ y_2 \ \dots \ y_{G^A}]$ y $\beta_{1\Delta} = \begin{bmatrix} 1 \\ \beta_{12} \\ \vdots \\ \beta_{1G^A} \end{bmatrix}$.

Por tanto, podemos escribir (13.59) en la forma

$$(13.59a) \quad \ell = \frac{\beta_{1\Delta}' W_{1\Delta} \beta_{1\Delta}}{\beta_{1\Delta}' W_1 \beta_{1\Delta}}$$

en la que $W_{1\Delta} = Y_{1\Delta}' Y_{1\Delta} - (Y_{1\Delta}' X_1) (X_1' X_1)^{-1} X_1' Y_{1\Delta},$
 $W_1 = Y_1' Y_1 - (Y_1' X) (X' X)^{-1} X' Y_1.$

El estimador de $\beta_{1\Delta}$ de la razón mínima de variancias viene dado por los valores de los coeficientes β que producen el valor mínimo de ℓ . Por tanto, tenemos que minimizar ℓ respecto de $\beta_{1\Delta}$. Tomando la primera derivada de ℓ , obtenemos

$$(13.61) \quad \frac{\partial \ell}{\partial \beta_{1\Delta}} = \frac{2(W_{1\Delta} \beta_{1\Delta})(\beta_{1\Delta}' W_1 \beta_{1\Delta}) - 2(\beta_{1\Delta}' W_{1\Delta} \beta_{1\Delta})(W_1 \beta_{1\Delta})}{(\beta_{1\Delta}' W_1 \beta_{1\Delta})^2}.$$

Igualando a cero esta derivada, tenemos

$$(13.62) \quad W_{1\Delta} \tilde{\beta}_{1\Delta} - \left[\frac{\beta_{1\Delta}' W_{1\Delta} \beta_{1\Delta}}{\beta_{1\Delta}' W_1 \beta_{1\Delta}} \right] W_1 \tilde{\beta}_{1\Delta} = 0$$

o

$$(13.62a) \quad (W_{1\Delta} - \ell W_1) \tilde{\beta}_{1\Delta} = 0.$$

Ahora bien, cuando $\tilde{\beta}_{1\Delta} \neq 0$, debemos tener

$$(13.63) \quad \text{Det}(W_{1\Delta} - \ell W_1) = 0;$$

es decir, el determinante de la matriz $(W_{1\Delta} - \ell W_1)$, cuya dimensión es $(G^A \times G^A)$, debe ser igual a cero. Como podemos obtener todos los elementos de $W_{1\Delta}$ y W_1 a partir de las observaciones muestrales, la ecuación (13.63) se convierte en un polinomio de grado G^A en la variable ℓ . Como hemos señalado, queremos que ℓ esté lo más próxima posible a la unidad y, por tanto, la solución de (13.63) que nos interesa es la más pequeña, a la que llamaremos ℓ_1 . Utilizando ℓ_1 en lugar de ℓ en (13.62a) podemos obtener una solución para los $(G^A - 1)$ elementos desconocidos de $\tilde{\beta}_{1\Delta}$. En este caso, los elementos de γ_1 se estiman a través de

$$(13.64) \quad \tilde{y}_1 = (X_1' X_1)^{-1} (X_1' \tilde{y}_1) = (X_1' X_1)^{-1} (X_1' Y_{1\Delta}) \tilde{\beta}_{1\Delta}.$$

Estas son las estimaciones maximoverosímiles con información limitada.

Al introducir este tipo de estimador, se mencionó que pertenece a la familia de los estimadores de clase k , y ahora podemos añadir que, en este estimador,

$$k = \ell_1$$

y que

$$\lim \text{prob } \ell_1 = 1.$$

$$\lim \text{prob } \sqrt{T}(\ell_1 - 1) = 0.$$

La demostración de estas proposiciones puede verse en otras obras.⁸ Sus implicaciones son que el estimador maximoverosímil con información limitada es consistente y posee la misma matriz de variancias y covariancias asintóticas que el estimador minimocuadrático en dos etapas. Una de las diferencias entre estos dos estimadores es que el de la máxima verosimilitud con información limitada es *invariante* respecto a la elección de la variable endógena cuyo coeficiente estructural se iguala a la unidad, mientras que el estimador minimocuadrático en dos etapas no es invariante en este sentido. Consideremos la primera ecuación estructural escrita en la forma

$$\beta_{11} y_{1t} + \beta_{12} y_{2t} + \dots + \beta_{1G^A} y_{G^A t} + \gamma_{11} x_{1t} + \dots + \gamma_{1K} x_{Kt} = u_{1t}.$$

Al obtener el estimador minimocuadrático en dos etapas y el de la máxima verosimilitud con información limitada hemos considerado que el valor de β_{11} es igual a la unidad. Esto significa que podemos escribir la ecuación estructural en la forma

$$y_{1t} = -\beta_{12} y_{2t} - \dots - \beta_{1G^A} y_{G^A t} - \gamma_{11} x_{1t} - \dots - \gamma_{1K} x_{Kt} + u_{1t}.$$

En nuestra notación, la estimación de β_{12} correspondiente a los mínimos cuadrados en dos etapas es β_{12} , y la de la máxima verosimilitud con información limitada es β_{12} . Supongamos ahora que cambiamos la especificación y conside-

⁸ Ver Goldberger, *op. cit.*, pp. 341-44.

ramos que, en lugar de β_{11} , β_{12} es igual a la unidad. En este caso, escribiríamos la ecuación estructural en la forma

$$Y_{2t} = -\beta_{11}Y_{1t} - \dots - \beta_{1G}Y_{Gt} - \gamma_{11}X_{1t} - \dots - \gamma_{1K}X_{Kt} + u_{1t}.$$

La estimación minimocuadrática en dos etapas de β_{11} es β_{11} y la de la máxima verosimilitud con información limitada es β_{11} . Por tanto,

$$\tilde{\beta}_{11} \neq \frac{1}{\tilde{\beta}_{12}},$$

pero

$$\tilde{\beta}_{11} = \frac{1}{\tilde{\beta}_{12}}.$$

Esto es lo que queremos decir al hablar de la «invariancia» del estimador maximoverosimil con información limitada. En este sentido, dicho estimador es único entre todos los miembros de la familia de estimadores de clase k^9 . La invariancia puede resultar útil en todos los casos en los que la teoría económica no ofrece ninguna ayuda para la elección de la variable endógena cuyo coeficiente estructural debe ser la unidad. Por último, puede señalarse que, cuando la ecuación estructural está exactamente identificada, el método de la máxima verosimilitud con información limitada produce el mismo resultado que el método de los mínimos cuadrados indirectos.

EJEMPLO. Podemos ilustrar la construcción de las estimaciones maximoverosimiles con información limitada en el caso del modelo del mercado de alimentos de (13.45). Las dos ecuaciones del modelo son

$$\begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 D_t + u_{1t} && \text{(demanda),} \\ Q_t &= \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + u_{2t} && \text{(oferta).} \end{aligned}$$

Las observaciones muestrales ($T = 20$) se han presentado en el cuadro 13-1. Para la ecuación de demanda, que está superidentificada, tenemos

$$Y_{1A} = \begin{bmatrix} Q_1 & P_1 \\ Q_2 & P_2 \\ \vdots & \vdots \\ Q_{20} & P_{20} \end{bmatrix}, \quad X_1 = \begin{bmatrix} 1 & D_1 \\ 1 & D_2 \\ \vdots & \vdots \\ 1 & D_{20} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & D_1 & F_1 & A_1 \\ 1 & D_2 & F_2 & A_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & D_{20} & F_{20} & A_{20} \end{bmatrix},$$

$$\beta_{1A} = \begin{bmatrix} 1 \\ -\alpha_2 \end{bmatrix}, \quad \gamma_1 = \begin{bmatrix} \alpha_1 \\ \alpha_3 \end{bmatrix}.$$

Por tanto,

$$W_{1*} = \begin{bmatrix} \sum Q_t^2 & \sum Q_t P_t \\ \sum Q_t P_t & \sum P_t^2 \end{bmatrix} - \begin{bmatrix} \sum Q_t & \sum Q_t D_t \\ \sum P_t & \sum P_t D_t \end{bmatrix} \begin{bmatrix} T & \sum D_t \\ \sum D_t & \sum D_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Q_t & \sum P_t \\ \sum Q_t D_t & \sum P_t D_t \end{bmatrix}$$

y

$$W_1 = \begin{bmatrix} \sum Q_t^2 & \sum Q_t P_t \\ \sum Q_t P_t & \sum P_t^2 \end{bmatrix} - \begin{bmatrix} \sum Q_t & \sum Q_t D_t & \sum Q_t F_t & \sum Q_t A_t \\ \sum P_t & \sum P_t D_t & \sum P_t F_t & \sum P_t A_t \end{bmatrix} \times \begin{bmatrix} T & \sum D_t & \sum F_t & \sum A_t \\ \sum D_t & \sum D_t^2 & \sum D_t F_t & \sum D_t A_t \\ \sum F_t & \sum F_t D_t & \sum F_t^2 & \sum F_t A_t \\ \sum A_t & \sum A_t D_t & \sum A_t F_t & \sum A_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Q_t & \sum P_t \\ \sum Q_t D_t & \sum P_t D_t \\ \sum Q_t F_t & \sum P_t F_t \\ \sum Q_t A_t & \sum P_t A_t \end{bmatrix}.$$

En consecuencia, el determinante de $(W_{1*} - W_1)$ es un polinomio de segundo grado en la variable t . Sus dos soluciones son

$$t_1 = 1.1739 \quad \text{y} \quad t_2 = 12.0031.$$

Por tanto, (13.62a) se convierte en

$$(13.65) \quad [W_{1*} - 1.1739 W_1] \begin{bmatrix} 1 \\ -\tilde{\alpha}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Las dos ecuaciones de (13.65) producen el mismo valor para $\tilde{\alpha}_2$:

$$\tilde{\alpha}_2 = -0.2295.$$

A partir de (13.64), la solución para $\tilde{Y}_1$ es

$$\tilde{Y}_1 = \begin{bmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \end{bmatrix} = \begin{bmatrix} T & \sum D_t \\ \sum D_t & \sum D_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Q_t & \sum P_t \\ \sum Q_t D_t & \sum P_t D_t \end{bmatrix} \begin{bmatrix} 1 \\ 0.2295 \end{bmatrix} = \begin{bmatrix} 93.6192 \\ 0.3100 \end{bmatrix}.$$

Los errores standard estimados de estas estimaciones se calculan según (13.53) y (13.54). Los resultados para las dos ecuaciones del modelo son

$$\begin{aligned} Q_t &= 93.6192 - 0.2295 P_t + 0.3100 D_t + u_{1t} && \text{(demanda),} \\ &\quad (8.0312) \quad (0.0980) \quad (0.0472) \quad (0.0996) \\ Q_t &= 49.5324 + 0.2401 P_t + 0.2556 F_t + 0.2529 A_t + u_{2t} && \text{(oferta).} \\ &\quad (12.0105) \quad (0.0999) \quad (0.0474) \end{aligned}$$

Podemos comparar la ecuación de demanda estimada con la ecuación de demanda verdadera de (13.45a) y con el resultado obtenido al utilizar el método de los mínimos cuadrados en dos etapas en el ejemplo anterior. Los resultados para la ecuación de oferta, que está exactamente identificada, son exactamente iguales a los obtenidos por el método de los mínimos cuadrados en dos etapas

9. Ver A. S. Goldberger, «An Instrumental Variable Interpretation of k -Class Estimation», *Indian Economic Journal*, vol. 13, núm. 3, 1965, pp. 424-31.

13.4 Métodos de estimación de sistemas

Los métodos de estimación para una sola ecuación producen estimaciones que son consistentes, pero, en general, no son asintóticamente eficientes. La razón de ello es que no consideran la correlación entre las perturbaciones de distintas ecuaciones. (Una explicación alternativa de la falta de eficiencia asintótica es que los estimadores para una sola ecuación no tienen en cuenta las restricciones previas sobre las demás ecuaciones del modelo.) Esta situación es parecida al caso en que se estima una ecuación de regresión que pertenece a un conjunto de regresiones aparentemente no relacionadas por el método de los mínimos cuadrados ordinarios (ver sección 12.3). Si no tenemos en cuenta la correlación entre las perturbaciones de las distintas ecuaciones estructurales, no estamos utilizando toda la información disponible sobre cada ecuación y, por tanto, no conseguimos la eficiencia asintótica. Al igual que ocurría en el caso de las regresiones aparentemente no relacionadas, podemos superar esta deficiencia estimando simultáneamente todas las ecuaciones del sistema. Para ello, podemos utilizar alguno de los llamados *métodos de estimación de sistemas* presentados en esta sección.

Estimación por mínimos cuadrados en tres etapas

El más sencillo de los métodos de estimación de sistemas es el de los *mínimos cuadrados en tres etapas*, que consiste en una aplicación inmediata de la estimación generalizada de Aitken al sistema de ecuaciones estructurales escrito, según (13.36), en la forma

$$(13.66) \quad \begin{aligned} y_1 &= \hat{Y}_1\beta_1 + X_1Y_1 + u_1^*, \\ y_2 &= \hat{Y}_2\beta_2 + X_2Y_2 + u_2^*, \\ &\vdots \\ y_G &= \hat{Y}_G\beta_G + X_GY_G + u_G^*. \end{aligned}$$

en la que

$$\begin{aligned} y &\rightarrow T \times 1, \\ \hat{Y} &\rightarrow T \times (G - 1), \\ X &\rightarrow T \times K_g, \\ \beta &\rightarrow (G - 1) \times 1, \\ Y &\rightarrow K_g \times 1, \\ u^* &\rightarrow T \times 1. \end{aligned}$$

obsérvese que G_g = número de variables endógenas incluidas en la ecuación g ;

K_g = número de variables predeterminadas incluidas en la ecuación g ;

$$g = 1, 2, \dots, G.$$

Si en el sistema existe alguna identidad, al escribir (13.66) prescindimos de ella. De forma alternativa, podemos expresar el sistema en la forma

$$(13.66a) \quad \begin{aligned} y_1 &= \hat{Z}_1\alpha_1 + u_1^*, \\ y_2 &= \hat{Z}_2\alpha_2 + u_2^*, \\ &\vdots \\ y_G &= \hat{Z}_G\alpha_G + u_G^*, \end{aligned} \quad \text{en la que} \quad \begin{aligned} \hat{Z} &= [\hat{Y} \quad X], \\ \alpha &= \begin{bmatrix} \beta \\ \gamma \end{bmatrix}. \end{aligned}$$

Podemos escribir (13.66a) en una forma más compacta, como

$$(13.66b) \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 & 0 & \dots & 0 \\ 0 & \hat{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{Z}_G \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_G \end{bmatrix} + \begin{bmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_G^* \end{bmatrix},$$

$$\text{o (13.66c)} \quad y = \hat{Z}\alpha + u^*$$

$$\text{en la que} \quad y \rightarrow GT \times 1,$$

$$\hat{Z} \rightarrow GT \times \sum_{g=1}^G (G_g - 1 + K_g),$$

$$\alpha \rightarrow \sum_{g=1}^G (G_g - 1 + K_g) \times 1,$$

$$u^* \rightarrow GT \times 1.$$

Si describimos el sistema de ecuaciones en la forma (13.66c), podemos aplicar la fórmula de los mínimos cuadrados generalizados de Aitken presentada en (12.50). De esta forma, obtenemos

$$(13.67) \quad \bar{\alpha} = (\hat{Z}'\Omega^{-1}\hat{Z})^{-1}(\hat{Z}'\Omega^{-1}y)$$

en la que

$$\Omega = E(u^*u^{*'}).$$

El estimador de α por mínimos cuadrados en tres etapas se obtiene sustituyendo Ω cuyos valores se desconocen, por $\hat{\Omega}$, que se define como

$$(13.68) \quad \hat{\Omega} = \begin{bmatrix} \hat{\psi}_{11}I_T & \hat{\psi}_{12}I_T & \dots & \hat{\psi}_{1G}I_T \\ \hat{\psi}_{21}I_T & \hat{\psi}_{22}I_T & \dots & \hat{\psi}_{2G}I_T \\ \vdots & \vdots & & \vdots \\ \hat{\psi}_{G1}I_T & \hat{\psi}_{G2}I_T & \dots & \hat{\psi}_{GG}I_T \end{bmatrix},$$

$$(13.69) \quad \hat{\psi}_{gh} = \frac{(y_g - Y_g\tilde{\beta}_g - X_g\tilde{\gamma}_g)'(y_h - Y_h\tilde{\beta}_h - X_h\tilde{\gamma}_h)}{(T - G_g + 1 - K_g)},$$

$$G_g + K_g \geq G_h + K_h,$$

$$g, h = 1, 2, \dots, G,$$

y $\tilde{\beta}$ y $\tilde{\gamma}$ representan las estimaciones minimocuadráticas en dos etapas de los coeficientes respectivos. (Algunos autores prefieren utilizar T en lugar de $(T - G_g + 1 - K_g)$ en el denominador de $\hat{\psi}_{gh}$, lo que no altera para nada las propiedades asintóticas del estimador minimocuadrático en tres etapas. Obsérvese, sin embargo, que, si el número de ecuaciones es mayor que el número de observaciones, $\hat{\Omega}$ será singular.) El estimador minimocuadrático en tres etapas de los coeficientes estructurales es, pues,

$$(13.70) \quad \hat{\alpha} = (\hat{Z}'\hat{\Omega}^{-1}\hat{Z})^{-1}(\hat{Z}'\hat{\Omega}^{-1}y) \\ = \begin{bmatrix} \hat{\psi}^{11}(\hat{Z}'_1\hat{Z}_1) & \hat{\psi}^{12}(\hat{Z}'_1\hat{Z}_2) & \dots & \hat{\psi}^{1G}(\hat{Z}'_1\hat{Z}_G) \\ \hat{\psi}^{21}(\hat{Z}'_2\hat{Z}_1) & \hat{\psi}^{22}(\hat{Z}'_2\hat{Z}_2) & \dots & \hat{\psi}^{2G}(\hat{Z}'_2\hat{Z}_G) \\ \vdots & \vdots & & \vdots \\ \hat{\psi}^{G1}(\hat{Z}'_G\hat{Z}_1) & \hat{\psi}^{G2}(\hat{Z}'_G\hat{Z}_2) & \dots & \hat{\psi}^{GG}(\hat{Z}'_G\hat{Z}_G) \end{bmatrix}^{-1} \begin{bmatrix} \sum_g \hat{\psi}^{1g}(\hat{Z}'_1y_g) \\ \sum_g \hat{\psi}^{2g}(\hat{Z}'_2y_g) \\ \vdots \\ \sum_g \hat{\psi}^{Gg}(\hat{Z}'_Gy_g) \end{bmatrix},$$

en el que $\hat{\psi}^{gh}$ representa el correspondiente elemento de la inversa de $\hat{\Omega}$. Al igual que en el caso del estimador de Aitken en dos etapas de la sección 12.3, la matriz de variancias y covariancias del estimador minimocuadrático en tres etapas puede estimarse mediante

$$(13.71) \quad \text{Est. Var-Cov}(\hat{\alpha}) = (\hat{Z}'\hat{\Omega}^{-1}\hat{Z})^{-1}.$$

Puede presentarse también una forma alternativa para la obtención de la fórmula de la estimación por mínimos cuadrados en tres etapas. Las ecuaciones estructurales del sistema pueden escribirse como

$$(13.72) \quad \begin{aligned} y_1 &= Z_1\alpha_1 + u_1, & \vdots \\ y_2 &= Z_2\alpha_2 + u_2, & y_G = Z_G\alpha_G + u_G. \end{aligned}$$

en las que

$$Z_g = [Y_g \quad X_g],$$

$$\alpha_g = \begin{bmatrix} \beta_g \\ \gamma_g \end{bmatrix}$$

Premultiplicado cada ecuación de (13.72) por X' obtenemos

$$(13.73) \quad \begin{aligned} X'y_1 &= X'Z_1\alpha_1 + X'u_1, \\ X'y_2 &= X'Z_2\alpha_2 + X'u_2, \\ &\vdots \\ X'y_G &= X'Z_G\alpha_G + X'u_G. \end{aligned}$$

Esto puede escribirse en la forma

$$(13.73a) \quad \begin{bmatrix} X'y_1 \\ X'y_2 \\ \vdots \\ X'y_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & 0 & \dots & 0 \\ 0 & X'Z_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X'Z_G \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_G \end{bmatrix} + \begin{bmatrix} X'u_1 \\ X'u_2 \\ \vdots \\ X'u_G \end{bmatrix},$$

$$(13.73b) \quad \bar{y} = \bar{Z}\alpha + \bar{u},$$

en la que

$$\bar{y} \rightarrow KG \times 1,$$

$$\bar{Z} \rightarrow KG \times \sum_{g=1}^G (G_g - 1 + K_g),$$

$$\alpha \rightarrow \sum_{g=1}^G (G_g - 1 + K_g) \times 1,$$

$$\bar{u} \rightarrow KG \times 1.$$

Señalemos que

$$E(\bar{u}\bar{u}') = \begin{bmatrix} \sigma_{11}(X'X) & \sigma_{12}(X'X) & \dots & \sigma_{1G}(X'X) \\ \sigma_{21}(X'X) & \sigma_{22}(X'X) & \dots & \sigma_{2G}(X'X) \\ \vdots & \vdots & & \vdots \\ \sigma_{G1}(X'X) & \sigma_{G2}(X'X) & \dots & \sigma_{GG}(X'X) \end{bmatrix},$$

en la que las σ representan las variancias y covariancias de las perturbaciones estructurales. Si aplicamos la fórmula de los mínimos cuadrados generalizados de Aitken a (13.73b), obtenemos

$$(13.74) \quad \hat{\alpha} = (\bar{Z}'\bar{\Theta}^{-1}\bar{Z})^{-1}(\bar{Z}'\bar{\Theta}^{-1}\bar{y}),$$

en la que $\Theta = E(\tilde{u}\tilde{u}')$.

Pueden obtenerse estimaciones consistentes de las variancias y covariancias de las perturbaciones estructurales utilizando los residuos minimocuadráticos en dos etapas. Es decir, podemos obtener estimaciones consistentes de σ_{ϵ}^2 utilizando

$$(13.75) \quad s_{gh} = \frac{(y_g - Y_g\tilde{\beta}_g - X_g\tilde{\gamma}_g)'(y_h - Y_h\tilde{\beta}_h - X_h\tilde{\gamma}_h)}{(T - G_g + 1 - K_g)},$$

$$G_g + K_g \geq G_h + K_h,$$

$$g, h = 1, 2, \dots, G.$$

Por tanto, podemos estimar α a través de

(13.76)

$$\hat{\alpha} = (Z'\Theta^{-1}Z)^{-1}(Z'\Theta^{-1}\tilde{y})$$

$$= \begin{bmatrix} s^{11}Z_1'X(X'X)^{-1}X'Z_1 & s^{12}Z_1'X(X'X)^{-1}X'Z_2 & \dots & s^{1G}Z_1'X(X'X)^{-1}X'Z_G \\ s^{21}Z_2'X(X'X)^{-1}X'Z_1 & s^{22}Z_2'X(X'X)^{-1}X'Z_2 & \dots & s^{2G}Z_2'X(X'X)^{-1}X'Z_G \\ \vdots & \vdots & \ddots & \vdots \\ s^{G1}Z_G'X(X'X)^{-1}X'Z_1 & s^{G2}Z_G'X(X'X)^{-1}X'Z_2 & \dots & s^{GG}Z_G'X(X'X)^{-1}X'Z_G \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_g s^{1g}Z_1'X(X'X)^{-1}X'y_g \\ \sum_g s^{2g}Z_2'X(X'X)^{-1}X'y_g \\ \vdots \\ \sum_g s^{Gg}Z_G'X(X'X)^{-1}X'y_g \end{bmatrix},$$

en la que s^{gh} representa el correspondiente elemento de la inversa de Θ . La expresión (13.76) equivale a la (13.70). Además,

$$(13.77) \quad \text{Est. Var-Cov}(\hat{\alpha}) = (Z'\Theta^{-1}Z)^{-1},$$

que es equivalente a (13.71). Según los supuestos efectuados al principio de este capítulo, las estimaciones minimocuadráticas en tres etapas de los coeficientes estructurales son consistentes y asintóticamente eficientes. (Si las variables predeterminadas del sistema incluyen variables endógenas retardadas, se supone que sus valores iniciales son fijos.) En los modelos en que se pueda suponer que las perturbaciones estructurales de las distintas ecuaciones no están correlacionadas, Θ será diagonal y las estimaciones minimocuadráticas

en tres etapas serán iguales a las de dos etapas. Señalemos también que, al igual que el estimador en dos etapas, el estimador minimocuadrático en tres etapas *no es invariante* respecto a la elección de la variable endógena cuyo coeficiente estructural se toma como unidad. Un último punto que resulta importante señalar es que, al aplicar el método de los mínimos cuadrados en tres etapas, la omisión de las ecuaciones exactamente identificadas no afectará las estimaciones de los coeficientes de las ecuaciones restantes.¹⁰ Esto significa que las ecuaciones exactamente identificadas no añaden ninguna información relevante para la estimación de las ecuaciones superidentificadas del sistema, pero, sin embargo, lo contrario no es cierto.

EJEMPLO. Siguiendo con el ejemplo del mercado de alimentos, utilizaremos los resultados de la aplicación del método de los mínimos cuadrados en dos etapas para estimar la matriz de variancias y covariancias de las perturbaciones estructurales. Para (13.75), esto es

$$\tilde{\Phi} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} 3.8664 & 4.3574 \\ 4.3574 & 6.0396 \end{bmatrix}.$$

Las estimaciones minimocuadráticas en tres etapas de los coeficientes estructurales son, pues,

$$Q_t = 94.6333 - 0.2436 P_t + 0.3140 D_t + u_{1t} \quad (\text{demanda}).$$

$$(7.9208) \quad (0.0965) \quad (0.0440)$$

$$Q_t = 52.1176 + 0.2289 P_t + 0.2290 F_t + 0.3579 A_t + u_{2t} \quad (\text{oferta}).$$

$$(11.8934) \quad (0.0997) \quad (0.0469) \quad (0.0729)$$

La ecuación de oferta está exactamente identificada y, por tanto, no añade ninguna información para la estimación de la ecuación superidentificada de demanda. Por tanto, las estimaciones minimocuadráticas en tres etapas de la ecuación de demanda son iguales a las obtenidas por el método de los mínimos cuadrados en dos etapas. Sin embargo, los resultados para la ecuación de oferta son distintos en los dos métodos.

Puede señalarse que los residuos de las ecuaciones estimadas por los mínimos cuadrados en tres etapas pueden utilizarse para obtener nuevas estimaciones de las variancias y covariancias de las perturbaciones estructurales, que pueden sustituir a las estimaciones anteriores en la fórmula de los mínimos cuadrados en tres etapas, dando lugar así a nuevas estimaciones de los coeficientes estructurales. Este proceso podría repetirse hasta que no se produjera ningún cambio en los coeficientes estructurales estimados. Las estimaciones resultantes, denominadas *estimaciones iterativas por mínimos cuadrados en*

10. Ver A. Zellner y H. Theil, «Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations», *Econometrica*, vol. 30, enero 1962, pp. 63-68.

tres etapas, poseen las mismas propiedades asintóticas que las estimaciones ordinarias por mínimos cuadrados en tres etapas.¹¹ En el ejemplo del mercado de alimentos, los coeficientes estimados convergen en el décimo lugar decimal tras efectuar nueve iteraciones. Las estimaciones finales de la matriz de variancias y covariancias de las perturbaciones estructurales son

$$\hat{\Phi} = \begin{bmatrix} 3.8664 & 5.0907 \\ 5.0907 & 6.9564 \end{bmatrix}.$$

Como resultado, los puntos de convergencia en la iteración de los mínimos cuadrados en tres etapas son

$$Q_t = 94.6333 - 0.2436P_t + 0.3140D_t + \bar{u}_{1t} \quad (\text{demanda}),$$

(7.9208) (0.0965) (0.0469)

$$Q_t = 52.5527 + 0.2271P_t + 0.2245F_t + 0.3756A_t + \bar{u}_{2t} \quad (\text{oferta}).$$

(12.7408) (0.1069) (0.0465) (0.0717)

Estimación de la máxima verosimilitud con información completa

Otro de los métodos de esta sección diseñado para la estimación de los coeficientes estructurales de un sistema de ecuaciones simultáneas es el *método de la máxima verosimilitud con información completa*, y se basa en la aplicación usual del principio de la máxima verosimilitud a todas las ecuaciones estocásticas del sistema simultáneamente. Consideremos el sistema completo

$$By_t + \Gamma x_t = u_t$$

especificado en (13.3a) y la correspondiente matriz de variancias y covariancias de las perturbaciones

$$\Phi = E(u_t u_t').$$

Según los supuestos especificados previamente, la distribución conjunta de los elementos de u_t es

$$f(u_t) = (2\pi)^{-(G/2)} |\Phi|^{-(1/2)} e^{-(u_t' \Phi^{-1} u_t)/2}.$$

La transformación de probabilidades a partir de las u_t inobservables a las y_t observables es

$$f(y_t | x_t) = f(u_t) \left| \frac{\partial u_t}{\partial y_t} \right| = f(u_t) \left| \frac{\partial (By_t + \Gamma x_t)}{\partial y_t} \right| = f(u_t) |B|,$$

11. Ver A. Madansky, «On the Efficiency of Three-Stage Least Squares Estimations», *Econometrica*, vol. 32, enero-abril 1964, p. 55.

en la que $|B|$ es el Jacobiano, representado por el valor absoluto del determinante de la matriz B . En este caso, la función logarítmica de verosimilitud para las T observaciones de y_t condicionada a los valores de x_t es

$$(13.78) \quad L = -\frac{GT}{2} \log(2\pi) - \frac{T}{2} \log |\Phi| + T \log |B| - \frac{1}{2} \sum_{t=1}^T (By_t + \Gamma x_t)' \Phi^{-1} (By_t + \Gamma x_t).$$

Los estimadores maximoverosímiles de B , Γ y Φ se obtienen maximizando (13.78) respecto a estos parámetros. Estos estimadores son consistentes, asintóticamente eficientes, y su distribución asintótica es normal. Esto significa que poseen las mismas propiedades asintóticas y la misma matriz de variancias y covariancias asintóticas que los estimadores minimocuadráticos en tres etapas. En los estimadores maximoverosímiles con información completa, esta última matriz puede estimarse por el procedimiento habitual a través de la matriz de información apropiada. Los estimadores maximoverosímiles con información completa son equivalentes a los llamados *estimadores de la mínima variancia residual generalizada*, que se obtienen minimizando el determinante de la matriz de variancias y covariancias de los residuos de la forma reducida.¹²

EJEMPLO. Para ilustrar la construcción de la función logarítmica de verosimilitud (13.78) utilizaremos el modelo del mercado de alimentos (13.45):

$$\begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 D_t + u_{1t} & (\text{demanda}), \\ Q_t &= \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + u_{2t} & (\text{oferta}). \end{aligned}$$

$$\text{En este caso,} \quad y_t' = [Q_t \ P_t], \quad x_t' = [1 \ D_t \ F_t \ A_t]$$

y

$$B = \begin{bmatrix} 1 & -\alpha_2 \\ 1 & -\beta_2 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} -\alpha_1 & -\alpha_3 & 0 & 0 \\ -\beta_1 & 0 & -\beta_3 & -\beta_4 \end{bmatrix}, \quad \Phi = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

Por tanto,

$$\begin{aligned} |\Phi| &= \sigma_{11}\sigma_{22} - \sigma_{12}^2, \\ |B| &= -\beta_2 + \alpha_2, \end{aligned}$$

12. Ver Goldberger (1964), *op. cit.*, pp. 352-56. H. Wold y E. J. Mosback presentan una interpretación ligeramente distinta de los sistemas interdependientes y un método de estimación iterativo llamado de «punto fijo» en *Interdependent Systems, Structure and Estimation*. Amsterdam, North-Holland Publishing Company, 1970.

y

$$(\mathbf{B}y_t + \mathbf{\Gamma}x_t)' \Phi^{-1} (\mathbf{B}y_t + \mathbf{\Gamma}x_t)$$

$$= [u_{1t} \ u_{2t}] \frac{1}{|\Phi|} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \frac{1}{|\Phi|} (\sigma_{22}u_{1t}^2 + \sigma_{11}u_{2t}^2 - 2\sigma_{12}u_{1t}u_{2t}),$$

en la que

$$u_{1t} = Q_t - \alpha_1 - \alpha_2 P_t - \alpha_3 D_t,$$

$$u_{2t} = Q_t - \beta_1 - \beta_2 P_t - \beta_3 F_t - \beta_4 A_t.$$

Por tanto, la función logarítmica de verosimilitud (13.78) se convierte en

$$(13.79) \quad L = -\frac{GT}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_{11}\sigma_{22} - \sigma_{12}^2) + T \log|\alpha_2 - \beta_2| \\ - \frac{1}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} \left\{ \sigma_{22} \sum_{t=1}^T (Q_t - \alpha_1 - \alpha_2 P_t - \alpha_3 D_t)^2 \right. \\ + \sigma_{11} \sum_{t=1}^T (Q_t - \beta_1 - \beta_2 P_t - \beta_3 F_t - \beta_4 A_t)^2 \\ \left. - 2\sigma_{12} \sum_{t=1}^T (Q_t - \alpha_1 - \alpha_2 P_t - \alpha_3 D_t)(Q_t - \beta_1 - \beta_2 P_t - \beta_3 F_t - \beta_4 A_t) \right\}$$

en la que $|\alpha_2 - \beta_2|$ representa el valor absoluto de $(\alpha_2 - \beta_2)$.

En el análisis anterior no hemos considerado la posibilidad de que el sistema incluya identidades además de las ecuaciones estocásticas. De ser así, la matriz de variancias y covariancias de las perturbaciones estructurales será singular y, por tanto, no tendrá inversa. Uno de los procedimientos para resolver este problema consiste en eliminar todas las identidades (y las correspondientes variables endógenas) por sustitución, antes de establecer la función de verosimilitud. Consideremos, por ejemplo, el modelo siguiente:

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t} \quad (\text{consumo}),$$

$$I_t = \beta_0 + \beta_1 r_t + \beta_2 I_{t-1} + u_{2t} \quad (\text{inversión}),$$

$$r_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 M_t + u_{3t} \quad (\text{mercado de dinero}),$$

$$Y_t = C_t + I_t + G_t \quad (\text{identidad de la renta}),$$

en el que C = consumo, Y = renta, I = inversión, r = tipo de interés, M = oferta monetaria y G = gasto público. Las variables C_t , Y_t , I_t y r_t , son endógenas y las demás, predeterminadas. Podemos eliminar la identidad de la renta sustituyendo Y_t por su valor según esta identidad en el resto del sistema. En este caso, el sistema se convierte en

$$C_t = \left(\frac{\alpha_0}{1 - \alpha_1} \right) + \left(\frac{\alpha_1}{1 - \alpha_1} \right) I_t + \left(\frac{\alpha_1}{1 - \alpha_1} \right) G_t + \left(\frac{\alpha_2}{1 - \alpha_1} \right) C_{t-1} + \left(\frac{1}{1 - \alpha_1} \right) u_{1t},$$

$$I_t = \beta_0 + \beta_1 r_t + \beta_2 I_{t-1} + u_{2t},$$

$$r_t = \gamma_0 + \gamma_1 C_t + \gamma_1 I_t + \gamma_1 G_t + \gamma_2 M_t + u_{3t},$$

que incluye únicamente ecuaciones estocásticas. Naturalmente, la matriz de variancias y covariancias de las perturbaciones estructurales, Φ , debe sustituirse por Φ^* , que se define como

$$\Phi^* = \begin{bmatrix} \frac{\sigma_{11}}{(1 - \alpha_1)^2} & \frac{\sigma_{12}}{(1 - \alpha_1)} & \frac{\sigma_{13}}{(1 - \alpha_1)} \\ \frac{\sigma_{12}}{(1 - \alpha_1)} & \sigma_{22} & \sigma_{23} \\ \frac{\sigma_{13}}{(1 - \alpha_1)} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

El inconveniente de este enfoque se debe a que introduce restricciones distintas de cero en las ecuaciones estructurales y, por tanto, complica la obtención de las estimaciones máximoverosímiles. Existe también un procedimiento alternativo que se ocupa de las identidades sin efectuar una sustitución directa.¹³

EJEMPLO. Las estimaciones máximoverosímiles con información completa de las ecuaciones estructurales del modelo del mercado de alimentos son

$$Q_t = 93.6192 - 0.2295 P_t + 0.3100 D_t + u_{1t} \quad (\text{demanda}),$$

$$(8.0312) \quad (0.0980) \quad (0.0457)$$

$$Q_t = 51.9445 + 0.2373 P_t + 0.2208 F_t + 0.3697 A_t + u_{2t} \quad (\text{oferta}).$$

$$(12.7460) \quad (0.1078) \quad (0.0474) \quad (0.0765)$$

Obsérvese que las estimaciones de los coeficientes de la ecuación de demanda son idénticas a las de la máxima verosimilitud con información limitada presentadas anteriormente. Esto es, nuevamente, resultado del hecho de que la ecuación de oferta está exactamente identificada y, por tanto, no proporciona ninguna información sobre la ecuación de demanda.

13.5 Comparación de los métodos alternativos de estimación y problemas especiales

Al estudiar los métodos de estimación de las ecuaciones estructurales de los sistemas generales interdependientes, hemos utilizado un modelo del mercado de alimentos en el que conocíamos los verdaderos valores de los parámetros. Por tanto, puede ser interesante comparar los resultados obtenidos para este modelo al utilizar los distintos métodos de estimación. No obstante, al efectuar la comparación, debemos tener presente que sólo disponemos de una muestra

13. Puede verse una descripción detallada en W. L. Ruble, «Improving the Computation of Simultaneous Stochastic Linear Equations Estimates», *Agricultural Economics Report No. 116*, Department of Agricultural Economics, Michigan State University, East Lansing, 1968, pp. 163-71.

en la que basar las comparaciones y, por tanto, la comparación entre los distintos estimadores se parece a una comparación entre distintas armas basada en un solo disparo de cada una de ellas. En el cuadro 13-2 se presenta un resumen de los resultados que incluye también las estimaciones MCO. Estas últimas son inconsistentes, pero, sin embargo, se utilizan a veces en el trabajo empírico debido principalmente a que resultan fáciles de calcular. En

Cuadro 13.2

Método *	Demanda			Oferta			
	Constante	P_t	D_t	Constante	P_t	F_t	A_t
Valores verdaderos de los coeficientes	96.5	-0.25	0.30	62.5	0.15	0.20	0.36
ESTIMACIONES							
MCO	99.8954 (7.5194)	-0.3163 (0.0907)	0.3346 (0.0454)	58.2754 (11.4629)	0.1604 (0.0949)	0.2481 (0.0462)	0.2483 (0.0975)
MC2E	94.6333 (7.9208)	-0.2436 (0.0965)	0.3140 (0.0469)	49.5324 (12.0105)	0.2401 (0.0999)	0.2556 (0.0472)	0.2529 (0.0996)
MVIL	93.6192 (8.0312)	-0.2295 (0.0980)	0.3100 (0.0474)	49.5324 (12.0105)	0.2401 (0.0999)	0.2556 (0.0472)	0.2529 (0.0996)
MC3E	94.6333 (7.9208)	-0.2436 (0.0965)	0.3140 (0.0469)	52.1176 (11.8934)	0.2289 (0.0997)	0.2290 (0.0440)	0.3579 (0.0729)
MC3EI	94.6333 (7.9208)	-0.2436 (0.0965)	0.3140 (0.0469)	55.5527 (12.7408)	0.2271 (0.1069)	0.2245 (0.0465)	0.3756 (0.0717)
MVIC	93.6192 (8.0312)	-0.2295 (0.0980)	0.3100 (0.0474)	51.9445 (12.7460)	0.2373 (0.1078)	0.2208 (0.0457)	0.3697 (0.0765)
ERRORES MUESTRALES							
MCO	3.3954	-0.0663	0.0346	-4.2246	0.0104	0.0481	-0.1117
MC2E	-1.8667	0.0064	0.0140	-12.9676	0.0901	0.0556	-0.1071
MVIL	-2.8808	0.0205	0.0100	-12.9676	0.0901	0.0556	-0.1071
MC3E	-1.8667	0.0064	0.0140	-10.3824	0.0789	0.0290	-0.0021
MC3EI	-1.8667	0.0064	0.0140	-6.9473	0.0771	0.0245	0.0156
MVIC	-2.8808	0.0205	0.0100	-10.5555	0.0873	0.0208	0.0097

* Notación: MCO mínimos cuadrados ordinarios
 MC2E mínimos cuadrados en dos etapas
 MVIL máxima verosimilitud con información limitada
 MC3E mínimos cuadrados en tres etapas
 MC3EI mínimos cuadrados en tres etapas iterativas
 MVIC máxima verosimilitud con información completa

nuestro ejemplo concreto, su actuación en la estimación de la ecuación de oferta se compara bastante bien con la de los estimadores MC2E y MVIL. Por lo que respecta a los métodos consistentes, podemos distinguir entre los resultados correspondientes a la ecuación de demanda, superidentificada, y los correspondientes a la ecuación de oferta exactamente identificada. En el caso de la *ecuación de demanda*, obtenemos resultados idénticos con los MC2E, MC3E y MC3EI por una parte, y con MVIL y MVIC por otra. Las razones de ello ya se han indicado en el texto. (Sin embargo, obsérvese que si el modelo hubiese contenido *más de una* ecuación superidentificada, además de la exactamente identificada, los resultados de los métodos de estimación de sistemas hubiesen diferido de los correspondientes para ecuaciones aisladas.) Por tanto, en este caso no existe ninguna mejora al pasar de los métodos para una sola ecuación (MC2E o MVIL) a los métodos de estimación de sistemas (MC3E, MC3EI o MVIC). Además, los resultados de MC2E y MVIL no resultan ser muy distintos. Por lo que respecta a la *ecuación de oferta*, obtenemos resultados idénticos con el método de los MC2E y el de MVIL. Comparando los métodos de estimación de sistemas con los de una sola ecuación, vemos que los primeros funcionan claramente mejor que los últimos, pero, sin embargo, no existe respuesta obvia a la pregunta de *cuál* de los métodos de estimación de sistemas funciona mejor. Naturalmente, en términos de facilidad de cálculo, el preferible es el de MC3E. Al evaluar el funcionamiento general de los métodos consistentes, debemos observar que *en todos los casos* los intervalos de confianza del 95 % calculados en la forma convencional abarcan los verdaderos valores de los coeficientes. Este resultado sería el esperado en una muestra grande, ya que los errores standard estimados están justificados en términos asintóticos, pero resulta consolador ver que ocurre también en una muestra de sólo 20 observaciones.

Todos los métodos de estimación de ecuaciones simultáneas estudiados en esta sección poseen algunas propiedades asintóticas deseables. Estas propiedades se hacen efectivas en las muestras grandes, pero, como las muestras de que disponemos son normalmente pequeñas, sería más interesante conocer las *propiedades en las muestras pequeñas* de estos estimadores. Lamentablemente, nuestros conocimientos en este campo no son muy amplios. Los resultados correspondientes al ejemplo del mercado de alimentos, presentados en el cuadro 13-2, se refieren a un solo conjunto de valores de los distintos estimadores, pero no proporcionan información sobre las características de sus distribuciones muestrales. Idealmente, nos gustaría conocer la descripción matemática de las distribuciones de cada uno de estos estimadores, pero sólo se dispone de este conocimiento en un pequeño número de casos especiales.¹⁴ A partir de

14. Ver, especialmente, R. L. Basman, «A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Over-Identified Cases», *Journal of the American Statistical Association*, vol. 56, septiembre 1961, pp. 619-36, y «A Note on the Exact Finite Sample Frequency Functions of Generalized Classical

ello podemos llegar a la conclusión de que la distribución del estimador MC2E en las muestras pequeñas tiene una media finita, pero que su variancia es, al menos en algunos casos, infinita. La forma de la distribución resulta ser similar a la de la distribución normal, pero su cima es más alta y los extremos son más gruesos que en el caso de la distribución normal. La mayor parte de la evidencia acerca de las propiedades de los estimadores de ecuaciones simultáneas en las muestras pequeñas procede de experimentos muestrales (o de Monte Carlo) similares a los descritos en el capítulo 2 y en la parte final de la sección 8.2 (por ejemplo, el caso de los alimentos podría constituir la primera muestra de un experimento de Monte Carlo). Como en otras obras se puede obtener una visión panorámica de estos experimentos,¹⁵ sólo vamos a presentar los resultados principales. En su mayor parte, los experimentos han presentado algunas de las características básicas de las distribuciones muestrales en lugar de sus descripciones completas, y generalmente estas características han sido la media, la variancia y la media de los errores al cuadrado. Naturalmente, cuando un estimador no tiene una variancia finita, el valor de la variancia o la media del error al cuadrado de una distribución muestral experimental no tienen ningún sentido, ya que tienden a aumentar sin límite al aumentar el número de muestras. Sin embargo, estos valores pueden servir de base para comparar estimadores alternativos respecto a las mismas muestras cuyo número es fijo. En todos los experimentos, las variables predeterminadas incluyen solamente variables exógenas, pero puede pensarse que la presencia de variables endógenas retardadas no cambiaría los puntos esenciales de los resultados obtenidos. Los resultados de los distintos experimentos no parecen coincidir por completo, pero algunas características comunes se han confirmado repetidamente. En concreto, el estimador MCO tiende a presentar un mayor sesgo, pero una variancia más pequeña que el MC2E o el MVIL. Si se utiliza la media del error al cuadrado como criterio de comparación entre los distintos estimadores de una sola ecuación, los resultados son algo confusos. En algunos casos, el estimador MCO funciona mejor que el MC2E o el MVIL, pero en otros casos funciona peor y a veces aproximadamente igual. Comparando MC2E con MVIL, el primer estimador presenta generalmente una media de error al cuadrado menor que el segundo, especialmente cuando las variables exógenas están muy intercorrelacionadas. Para los estimadores de sistemas se dispone de muchos menos resultados experimentales que para los estimadores de una sola ecuación. A partir de la evidencia disponible, parece que la ventaja asintótica del primer tipo de estimadores sobre los del segundo tipo persiste también en las muestras pequeñas, pero no de una forma tan pronunciada. Sin

Linear Estimators in a Leading Three-Equation Case», *Journal of the American Statistical Association*, vol. 58, marzo 1963, pp. 161-71.

15. Ver Johnston, *op. cit.*, capítulo 10, y C. F. Christ, *Econometric Models and Methods*. Nueva York, Wiley, 1966, pp. 474-81.

embargo, el método de MVIC parece mucho más sensitivo a los errores de especificación (como, por ejemplo, la omisión de una variable explicativa relevante) que el de MC2E.

En muchos casos, al valorar los distintos métodos de estimación no sólo nos interesan las propiedades de los estimadores, sino también el funcionamiento de los distintos métodos en la *contrastación de hipótesis*. En este contexto, el valor más importante es el de la razón entre el coeficiente estimado y su error standard estimado. Este estadístico de prueba se denomina generalmente «razón de *t*» a pesar del hecho de que la distribución *t* no es necesariamente la apropiada. Cuando se utilizan métodos consistentes de estimación de ecuaciones simultáneas, la distribución asintótica de este estadístico es normal. En las muestras pequeñas, las regiones de aceptación o los intervalos de confianza deseados se calculan generalmente a través de las tablas de la distribución *t*. Evidentemente, el procedimiento no es exactamente válido, ya que el estadístico de prueba *no* tiene una distribución *t* y, por tanto, el problema consiste en ver si esta distribución puede utilizarse como una aproximación razonable de la distribución verdadera, de modo que los resultados de las contrastaciones y de la estimación de intervalos no presenten mucha distorsión. La evidencia disponible a partir de estudios de Monte Carlo sugiere que, en muchos casos (pero no siempre), la distorsión es razonablemente pequeña. Por ejemplo, Cragg, que ha efectuado gran número de experimentos, señala lo siguiente acerca de este problema:

«Generalmente, utilizando los errores standard de los métodos consistentes se obtienen resultados válidos, pero no siempre es así. Los errores standard de los MCD no han sido útiles para efectuar inferencias acerca de los verdaderos valores de los coeficientes»¹⁶.

Summers ha llegado también a una conclusión muy similar,¹⁷ lo que sugiere que, al menos desde el punto de vista de la contrastación de hipótesis y la estimación de intervalos, el método de MCO es inferior a los métodos de estimación consistentes.

Otro importante criterio para enjuiciar los distintos estimadores es su capacidad para efectuar *pronósticos*. En los sistemas de ecuaciones simultáneas, pueden efectuarse predicciones de los valores de las variables endógenas a través de las ecuaciones de la forma reducida. Supongamos que x_0 es el vector de orden $K \times 1$ de los valores de las variables predeterminadas para el período

16. J. G. Cragg, «On the Relative Small-Sample Properties of Several Structural-Equation Estimators», *Econometrica*, vol. 35, enero 1967, p. 109. El término «MCD» hace referencia al estimador MCO.

17. R. M. Summers, «A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators», *Econometrica*, vol. 33, enero 1965, pp. 1-41.

do del pronóstico. En este caso, el pronóstico de los valores de las G variables endógenas es

$$\hat{y}_0 = \hat{\Pi}x_0,$$

en el que $\hat{\Pi}$ es la matriz de orden $G \times K$ de los coeficientes de la forma reducida estimados a partir de las observaciones muestrales. Pueden obtenerse estimaciones consistentes de $\hat{\Pi}$ aplicando directamente el método de los MCO a las ecuaciones de la forma reducida, o utilizando estimaciones consistentes de los coeficientes estructurales, por ejemplo $\hat{B}$ y $\hat{\Gamma}$, y obteniendo

$$(13.80) \quad \hat{\Pi} = -\hat{B}\hat{\Gamma}$$

según (13.7). Las estimaciones de los coeficientes de la forma reducida obtenidas mediante la aplicación directa de los MCO se denominan «sin restricciones», y las del segundo procedimiento se llaman «derivadas». Los coeficientes sin restricciones de la forma reducida no coincidirán con los coeficientes derivados, a menos que todas las ecuaciones estructurales estén exactamente identificadas. En general, el estimador minimocuadrático sin restricciones de $\hat{\Pi}$ no incorpora las restricciones previas impuestas por la relación

$$\Pi = -B^{-1}\Gamma,$$

por tanto, será asintóticamente menos eficiente que el estimador derivado $\hat{\Pi}$. Además, como los métodos de MC3E y MVIC producen estimaciones asintóticamente eficientes de B y Γ , y esta propiedad se transmite a todas las funciones univariadas de B y Γ , el estimador derivado de $\hat{\Pi}$ basado en las estimaciones por MC3E o MVIC de los coeficientes estructurales es también asintóticamente eficiente. Esto implica que dicho estimador derivado produce una menor variancia asintótica del error del pronóstico que los estimadores derivados a partir de MC2E o MVIL. En base a los experimentos de Monte Carlo, existe cierta evidencia de que estos resultados aparecen también en las muestras pequeñas, aunque esta ordenación debe considerarse muy tentativa.¹⁸

Sistemas recursivos

En los apartados anteriores nos hemos ocupado de la estimación de los sistemas generales interdependientes de ecuaciones, y ahora vamos a considerar brevemente el problema de la estimación en el caso en que el sistema de ecuaciones es *recursivo*. Como se ha señalado en la sección 13.1, el sistema

$$By_t + \Gamma x_t = u_t$$

se denomina *recursivo* si la matriz B es triangular. Un ejemplo muy conocido de sistema recursivo es el llamado «modelo de la telaraña» que puede expresarse de la forma siguiente:

$$(13.81) \quad \begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_{t-1} + u_{1t} && \text{(oferta),} \\ P_t &= \beta_1 + \beta_2 Q_t + \beta_3 Y_t + u_{2t} && \text{(demanda),} \end{aligned}$$

en el que Q = cantidad de equilibrio, P = precio de equilibrio, e Y = (renta exógena). En este caso,

$$B = \begin{bmatrix} 1 & 0 \\ -\beta_2 & 1 \end{bmatrix} \quad \text{y} \quad \Gamma = \begin{bmatrix} -\alpha_1 & -\alpha_2 & 0 \\ -\beta_1 & 0 & -\beta_3 \end{bmatrix}.$$

Obsérvese que la ecuación de oferta sólo incluye una variable endógena y , por tanto, la utilización de los mínimos cuadrados ordinarios produciría estimaciones consistentes. Sin embargo, estas estimaciones no serían asintóticamente eficientes porque no tienen en cuenta la correlación entre las perturbaciones de las dos ecuaciones. La ecuación de demanda se podría estimar de forma consistente por MCI. Alternativamente, ambas ecuaciones podrían estimarse por uno de los métodos de estimación de sistemas, en cuyo caso las estimaciones no sólo serían consistentes, sino también asintóticamente eficientes. Una característica importante de los modelos recursivos como el (13.81) es que el cálculo de las estimaciones MVIC se simplifica en gran medida por el hecho de que el jacobiano (es decir, el determinante de B) es igual a la unidad (esto presupone que todos los elementos de la diagonal principal de B son unos; de no ser así, podemos dividir cada ecuación estructural por el correspondiente coeficiente β_{kk}).

Si el sistema

$$By_t + \Gamma x_t = u_t$$

es *diagonalmente recursivo*, es decir, si B es triangular y la matriz de variancias y covariancias de las perturbaciones estructurales, Φ , es diagonal, el problema de la estimación se hace realmente simple. La razón es que las variables endógenas, que son los factores explicativos en las ecuaciones estructurales, no están correlacionadas con el término de perturbación de la ecuación en la que aparecen. Por ejemplo, en la ecuación de demanda de (13.81), las variables explicativas incluyen la variable endógena Q_t . Sin embargo, si u_{1t} y u_{2t} son independientes y no autorregresivas, tenemos

$$E(Q_t u_{2t}) = E[(\alpha_1 + \alpha_2 P_{t-1} + u_{1t}) u_{2t}] = 0.$$

Esto significa que la aplicación del método MCO a cada una de las ecuaciones estructurales producirá estimaciones consistentes y asintóticamente eficientes (si en la ecuación no aparecen variables endógenas retardadas, las estimacio-

18. Ibid., pp. 31-32.

nes MCO son también *mejoradas*. Debe ponerse de relieve que esto sólo ocurre cuando el sistema es *diagonalmente* recursivo y no en los demás casos. Por último, señalemos que si B es triangular en bloques y la matriz Φ es diagonal en bloques, las ecuaciones correspondientes a cada uno de los bloques diagonales de Φ pueden considerarse como pertenecientes a sistemas distintos, lo que simplifica también de forma considerable la estimación.

Perturbaciones autorregresivas

Como último punto en relación con el problema de la estimación de los coeficientes de las ecuaciones estructurales en los sistemas de ecuaciones simultáneas, consideraremos el caso en que las perturbaciones estructurales no son independientes a lo largo del tiempo. En particular, supongamos que las perturbaciones estructurales siguen un esquema *autorregresivo* de primer orden del tipo considerado en la sección 8.2. En este caso, tenemos

$$u_{1t} = \rho_1 u_{1,t-1} + \varepsilon_{1t},$$

$$u_{2t} = \rho_2 u_{2,t-1} + \varepsilon_{2t},$$

$$\vdots$$

$$u_{Gt} = \rho_G u_{G,t-1} + \varepsilon_{Gt}.$$

Las ε no se suponen independientes entre las distintas ecuaciones. Consideremos la estimación de la primera ecuación estructural del sistema tal como se presenta en (13.34):

$$y_1 = Y_1 \beta_1 + X_1 \gamma_1 + u_1.$$

La observación t -ésima viene representada por

$$(13.82) \quad y_{1t} = -\beta_{12} y_{2t} - \dots - \beta_{1G} y_{Gt} - \gamma_{11} x_{1t} - \dots - \gamma_{1K} x_{Kt} + u_{1t}.$$

De momento, nos limitaremos al caso en que *las variables predeterminadas del sistema no incluyen variables endógenas retardadas*. En este caso, las estimaciones MCO de los coeficientes de la forma reducida son consistentes. Por tanto, la aplicación directa del método de MC2E produce estimaciones consistentes de los coeficientes estructurales, pero las estimaciones de la matriz de variancias y covariancias son inconsistentes. Sin embargo, podemos modificar fácilmente el método de MC2E para tener en cuenta la naturaleza autorregresiva de u_{1t} . En primer lugar, aplicaremos la transformación presentada en (8.52) a (13.82), obteniendo

$$(13.83) \quad y_{1t} - \rho_1 y_{1,t-1} = -\beta_{12}(y_{2t} - \rho_1 y_{2,t-1}) - \beta_{13}(y_{3t} - \rho_1 y_{3,t-1}) - \dots - \beta_{1G}(y_{Gt} - \rho_1 y_{G,t-1}) - \gamma_{11}(x_{1t} - \rho_1 x_{1,t-1}) - \gamma_{12}(x_{2t} - \rho_1 x_{2,t-1}) - \dots - \gamma_{1K}(x_{Kt} - \rho_1 x_{K,t-1}) + \varepsilon_{1t}.$$

Podemos aplicar una transformación similar a cada una de las ecuaciones estructurales. De este modo, se sustituyen las u autorregresivas por las ε no autorregresivas. A continuación, puede resolverse el sistema resultante obteniendo los valores presentes en las variables endógenas en función de los valores retardados de las propias variables endógenas y de los valores presentes y retardados de las exógenas. Las ecuaciones obtenidas de esta forma pueden llamarse «ecuaciones aumentadas de la forma reducida», ya que contienen un conjunto incrementado de variables explicativas en comparación con las ecuaciones ordinarias de la forma reducida. Por lo que respecta a (13.83), las ecuaciones aumentadas de la forma reducida que resultan relevantes son las de y_{2t} , y_{3t} , ..., y_{Gt} . Los coeficientes de estas ecuaciones pueden estimarse aplicando el método de los mínimos cuadrados ordinarios (sin restricciones). Llamemos $\hat{y}_{2t}$, $\hat{y}_{3t}$, ..., $\hat{y}_{Gt}$ a los valores ajustados resultantes de las y , y formemos

$$(13.84) \quad y_{1t} - \rho_1 y_{1,t-1} = -\beta_{12}(\hat{y}_{2t} - \rho_1 \hat{y}_{2,t-1}) - \beta_{13}(\hat{y}_{3t} - \rho_1 \hat{y}_{3,t-1}) - \dots - \beta_{1G}(\hat{y}_{Gt} - \rho_1 \hat{y}_{G,t-1}) - \gamma_{11}(x_{1t} - \rho_1 x_{1,t-1}) - \gamma_{12}(x_{2t} - \rho_1 x_{2,t-1}) - \dots - \gamma_{1K}(x_{Kt} - \rho_1 x_{K,t-1}) + \varepsilon_{1t}^*$$

o

$$(13.84a) \quad y_1^* = \hat{Y}_1^* \beta_1 + X_1^* \gamma_1 + \varepsilon_1^*,$$

en la que

$$y_1^* = \begin{bmatrix} y_{12} - \rho_1 y_{11} \\ y_{13} - \rho_1 y_{12} \\ \vdots \\ y_{1T} - \rho_1 y_{1,T-1} \end{bmatrix}, \quad \hat{Y}_1^* = \begin{bmatrix} (\hat{y}_{22} - \rho_1 \hat{y}_{21}) & (\hat{y}_{32} - \rho_1 \hat{y}_{31}) & \dots & (\hat{y}_{G2} - \rho_1 \hat{y}_{G1}) \\ (\hat{y}_{23} - \rho_1 \hat{y}_{22}) & (\hat{y}_{33} - \rho_1 \hat{y}_{32}) & \dots & (\hat{y}_{G3} - \rho_1 \hat{y}_{G2}) \\ \vdots & \vdots & \ddots & \vdots \\ (\hat{y}_{2T} - \rho_1 \hat{y}_{2,T-1}) & (\hat{y}_{3T} - \rho_1 \hat{y}_{3,T-1}) & \dots & (\hat{y}_{GT} - \rho_1 \hat{y}_{G,T-1}) \end{bmatrix},$$

y

$$X_1^* = \begin{bmatrix} (x_{12} - \rho_1 x_{11}) & (x_{22} - \rho_1 x_{21}) & \dots & (x_{K2} - \rho_1 x_{K1}) \\ (x_{13} - \rho_1 x_{12}) & (x_{23} - \rho_1 x_{22}) & \dots & (x_{K3} - \rho_1 x_{K2}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{1T} - \rho_1 x_{1,T-1}) & (x_{2T} - \rho_1 x_{2,T-1}) & \dots & (x_{KT} - \rho_1 x_{K,T-1}) \end{bmatrix}.$$

Los coeficientes ρ_1 , β_1 y γ_1 pueden estimarse aplicando el método de los mínimos cuadrados con restricciones a (13.84) tal como se ha presentado en la sección 8.2.

El método de MC3E puede modificarse también fácilmente para tener en cuenta el carácter autorregresivo de las perturbaciones estructurales. Si aplicamos el método modificado de MC2E a cada ecuación estructural, obtenemos estimaciones consistentes de ρ_1 , ρ_2 , ..., ρ_G . Estas estimaciones pueden utilizarse para transformar las variables de forma que podamos escribir el sistema como

$$(13.85) \quad \begin{aligned} y_1^{**} &= \hat{Y}_1^{**} \beta_1 + X_1^{**} \gamma_1 + \epsilon_1^{**}, \\ y_2^{**} &= \hat{Y}_2^{**} \beta_2 + X_2^{**} \gamma_2 + \epsilon_2^{**}, \\ &\vdots \\ y_G^{**} &= \hat{Y}_G^{**} \beta_G + X_G^{**} \gamma_G + \epsilon_G^{**}, \end{aligned}$$

en el que el doble asterisco indica que se han sustituido las ρ por sus estimaciones consistentes, mientras que la notación restante es la misma que en (13.84a). Los coeficientes de (13.85) se pueden estimar por el método de estimación de Aitken en dos etapas, como en el caso del método ordinario de MC3E. También sería posible establecer una función de distribución conjunta para las ϵ y utilizarla para obtener las estimaciones MVIC para el caso autorregresivo, pero ello resultaría muy complicado.

Todo este análisis se ha limitado al caso en que las variables predeterminadas del sistema de ecuaciones simultáneas no incluyen ninguna variable endógena retardada. Si no es así, cuando existen perturbaciones autorregresivas la estimación resulta muy difícil. El origen de esta dificultad es el hecho de que, en estas circunstancias, las variables predeterminadas están correlacionadas con las perturbaciones estructurales, y una consecuencia de ello es que las estimaciones MCO de los coeficientes de la forma reducida son inconsistentes, lo que causa serias complicaciones. Una forma de resolver esta dificultad podría ser la utilización del método de las variables instrumentales, y Fisher ha propuesto otro enfoque distinto, similar al adoptado en el contexto de los modelos con retardos distribuidos.¹⁹

19. Ver F. M. Fisher, *The Identification Problem in Econometrics*. Nueva York, McGraw-Hill, 1966, pp. 168-75. El enfoque de las variables instrumentales se presenta en R. C. Fair, «The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors», *Econometrica*, vol. 58, mayo 1970, pp. 507-16.

13.6 Análisis de los modelos econométricos dinámicos²⁰

Los modelos de ecuaciones simultáneas representan un conjunto de relaciones entre las variables endógenas y las predeterminadas del sistema. Si todas las variables predeterminadas son puramente exógenas, el modelo especifica la forma en que las variables exógenas, junto a las perturbaciones estocásticas, generan los valores de las variables endógenas en un momento (o período) determinado de tiempo. Esto se ilustró de forma detallada en el ejemplo del mercado de alimentos de la sección 13.2. Sin embargo, si las variables predeterminadas incluyen también variables endógenas retardadas, el modelo especifica no sólo la forma en que las variables predeterminadas (y las perturbaciones) generan los valores presentes de las variables endógenas, sino también la forma en que las trayectorias temporales de las variables exógenas y de las perturbaciones determinan la trayectoria temporal de las variables endógenas. Esto se halla implícito en la estructura de los modelos, pero resulta interesante formular esta dependencia dinámica de modo explícito. Otros problemas de interés son la estabilidad del sistema y la influencia de los valores previos de las variables exógenas sobre los valores presentes de las endógenas. El propósito de esta sección es examinar estas cuestiones, ya que sus respuestas constituyen una parte natural e importante de la tarea de los econométricos. El procedimiento consistirá en ilustrar los problemas y su solución mediante un modelo específico, ya que, de esta forma, los conceptos especiales del análisis dinámico adquieren un sentido concreto. Sin embargo, los procedimientos básicos pueden adaptarse también fácilmente a otros modelos.

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t} && \text{(consumo),} \\ I_t &= \beta_0 + \beta_1 r_t + \beta_2 I_{t-1} + u_{2t} && \text{(inversión),} \\ r_t &= \gamma_0 + \gamma_1 Y_t + \gamma_2 M_t + u_{3t} && \text{(mercado de dinero),} \\ Y_t &= C_t + I_t + G_t && \text{(identidad de la renta),} \end{aligned}$$

en el que C = consumo, Y = renta, I = inversión, r = tipo de interés, M = oferta monetaria, y G = gasto público. C , Y , I y r son endógenas, y M y G son exógenas. Estas dos últimas pueden considerarse como variables

20. En esta exposición presentaremos únicamente los elementos más básicos del análisis dinámico aplicable a los sistemas lineales simples. En la actualidad, se dispone de métodos considerablemente más sofisticados que incluyen simulación, análisis de medidas de política económica y multiplicadores dinámicos no lineales. Ver, por ejemplo, L. R. Klein, M. K. Evans y M. Hartley, *Econometric Gaming*. Nueva York, Macmillan, 1969; Gary Fromm y Paul Taubman, *Policy Simulations with an Econometric Model*. Washington, Brookings Institution, 1968; o M. K. Evans y L. R. Klein, *The Wharton Econometric Forecasting Model*. Filadelfia, University of Pennsylvania Press, 1967.

de política económica que pueden ser, y son, manejadas a voluntad por los responsables de la política económica. La forma reducida de este sistema es

$$(13.86) \quad C_t = \pi_{11} + \pi_{12}C_{t-1} + \pi_{13}I_{t-1} + \pi_{14}M_t + \pi_{15}G_t + v_{1t},$$

$$(13.87) \quad I_t = \pi_{21} + \pi_{22}C_{t-1} + \pi_{23}I_{t-1} + \pi_{24}M_t + \pi_{25}G_t + v_{2t},$$

$$(13.88) \quad Y_t = \pi_{31} + \pi_{32}C_{t-1} + \pi_{33}I_{t-1} + \pi_{34}M_t + \pi_{35}G_t + v_{3t},$$

$$(13.89) \quad r_t = \pi_{41} + \pi_{42}C_{t-1} + \pi_{43}I_{t-1} + \pi_{44}M_t + \pi_{45}G_t + v_{4t}.$$

Naturalmente, los coeficientes de la forma reducida pueden expresarse en términos de los coeficientes estructurales. En el lenguaje del análisis dinámico, los coeficientes de la forma reducida se denominan *multiplicadores instantáneos*, ya que miden la respuesta inmediata de las variables endógenas a los cambios en las variables predeterminadas. Por ejemplo, π_{15} mide el cambio en el valor medio del consumo presente a consecuencia de un cambio de una unidad en el gasto público presente, dado el valor de la oferta monetaria en este mismo período y los niveles de consumo e inversión del período anterior.

Las ecuaciones de la forma reducida tienen su principal utilidad en las previsiones a corto plazo. Podemos pronosticar los valores de las variables endógenas en el próximo período en base al conocimiento de los valores actuales de I y C y de los valores futuros de las variables de política económica. Sin embargo, las ecuaciones de la forma reducida no resultan útiles para determinar cómo opera el sistema bajo el impacto continuo de las variables exógenas. Por ejemplo, la ecuación del consumo en forma reducida muestra que el valor de dicha variable en el período actual depende de los valores actuales de G y M , pero también de los valores previos del consumo y la inversión. Esto no es muy revelador, ya que transfiere parte de la responsabilidad en la explicación del nivel presente de consumo a su propia historia reciente. Para comprender el mecanismo de la determinación del consumo y de las demás variables endógenas, necesitamos una solución que explique el nivel de estas variables *sin* referirse a su pasado inmediato. Esta solución puede obtenerse manipulando las ecuaciones de la forma reducida de modo que se eliminen todas las variables endógenas retardadas. Esta eliminación puede efectuarse en dos etapas. En primer lugar, se reformulan las ecuaciones de la forma reducida de manera que cada ecuación incluya solamente una variable endógena, del momento presente o retardada. A continuación se eliminan las variables endógenas retardadas tal como se pretende.

Tomemos la forma reducida de la ecuación de consumo. La única variable endógena retardada «exterior» es I_{t-1} , pero, según (13.87), tenemos

$$(13.90) \quad I_t - \pi_{23}I_{t-1} = \pi_{21} + \pi_{22}C_{t-1} + \pi_{24}M_t + \pi_{25}G_t + v_{2t}$$

o

$$(13.90a) \quad I_{t-1} - \pi_{23}I_{t-2} = \pi_{21} + \pi_{22}C_{t-2} + \pi_{24}M_{t-1} + \pi_{25}G_{t-1} + v_{2,t-1}.$$

Además, retardando (13.86) un período y multiplicando por $-\pi_{23}$ obtenemos

$$(13.91) \quad -\pi_{23}C_{t-1} = -\pi_{11}\pi_{23} - \pi_{12}\pi_{23}C_{t-2} - \pi_{13}\pi_{23}I_{t-2} - \pi_{14}\pi_{23}M_{t-1} \\ - \pi_{15}\pi_{23}G_{t-1} - \pi_{23}v_{1,t-1}.$$

Sumando (13.86) y (13.91), se obtiene

$$(13.92) \quad C_t = \pi_{11}(1 - \pi_{23}) + (\pi_{12} + \pi_{23})C_{t-1} - \pi_{12}\pi_{23}C_{t-2} \\ + \pi_{13}(I_{t-1} - \pi_{23}I_{t-2}) + \pi_{14}M_t - \pi_{14}\pi_{23}M_{t-1} \\ + \pi_{15}G_t - \pi_{15}\pi_{23}G_{t-1} + v_{1t} - \pi_{23}v_{1,t-1}.$$

Sustituyendo $(I_{t-1} - \pi_{23}I_{t-2})$ por su valor según (13.90a) en (13.92) tenemos

$$(13.93) \quad C_t = [\pi_{11}(1 - \pi_{23}) + \pi_{13}\pi_{21}] + (\pi_{12} + \pi_{23})C_{t-1} \\ + (\pi_{13}\pi_{22} - \pi_{12}\pi_{23})C_{t-2} + \pi_{14}M_t + (\pi_{13}\pi_{24} - \pi_{14}\pi_{23})M_{t-1} \\ + \pi_{15}G_t + (\pi_{13}\pi_{25} - \pi_{15}\pi_{23})G_{t-1} + v_{1t} - \pi_{23}v_{1,t-1} \\ + \pi_{13}v_{2,t-1}.$$

o, en notación obvia,

$$(13.93a) \quad C_t = \delta + \mu_1 C_{t-1} + \mu_2 C_{t-2} + \kappa_0 M_t + \kappa_1 M_{t-1} + v_0 G_t + v_1 G_{t-1} + \varepsilon_t.$$

La ecuación (13.93), o su equivalente (13.93a), se denomina *ecuación dinámica fundamental* del consumo, y pueden obtenerse ecuaciones similares para las demás variables endógenas del sistema. Puede mostrarse que, en un sistema general interdependiente lineal, todas las variables satisfacen la misma ecuación final autorregresiva²¹ y, por tanto, nos referiremos únicamente al consumo. La ecuación dinámica fundamental tiene especial importancia para la determinación de la estabilidad del sistema y volveremos a ella más adelante. Nuestro objetivo es ahora expresar el consumo presente en términos de los valores presentes y pasados de las variables exógenas y de las perturbaciones, lo que puede hacerse a través de una serie de sustituciones sucesivas. En primer lugar, obsérvese que según (13.93a), cuando $t = 1$, la expresión para C_t es

$$(13.94) \quad C_1 = \delta + \mu_1 C_0 + \mu_2 C_{-1} + \kappa_0 M_1 + \kappa_1 M_0 + v_0 G_1 + v_1 G_0 + \varepsilon_1.$$

Los valores de C_0 y C_{-1} se han obtenido antes del principio de la serie y se denominan *condiciones iniciales*. Para los fines del análisis dinámico, suponemos que sus valores están dados. Los valores de M_0 y G_0 no tienen ningún

21. Ver A. S. Goldberger, *Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model*. Amsterdam, North-Holland Publishing Company, 1959, pp. 106-108.

interés especial y pueden combinarse con el término constante, de modo que (13.94) puede escribirse en la forma

$$(13.94a) \quad C_1 = \eta + \mu_1 C_0 + \mu_2 C_{-1} + \kappa_0 M_1 + \nu_0 G_1 + \varepsilon_1.$$

A continuación, cuando $t = 2$, la expresión para C_t es

$$(13.95) \quad C_2 = \delta + \mu_1 C_1 + \mu_2 C_0 + \kappa_0 M_2 + \kappa_1 M_1 + \nu_0 G_2 + \nu_1 G_1 + \varepsilon_2.$$

Como el valor de C_0 se considera dado, en (13.95) sólo tenemos que eliminar C_1 , sustituyéndolo por su valor según (13.94a), con lo cual obtenemos

$$(13.96) \quad C_2 = (\delta + \mu_1 \eta) + (\mu_1^2 + \mu_2) C_0 + \mu_1 \mu_2 C_{-1} + \kappa_0 M_2 + (\kappa_1 + \kappa_0 \mu_1) M_1 + \nu_0 G_2 + (\nu_1 + \nu_0 \mu_1) G_1 + \varepsilon_2 + \mu_1 \varepsilon_1.$$

Por el mismo procedimiento podemos obtener las soluciones para C_3 , C_4 , etc. El resultado es una expresión general para C_t en la forma

$$(13.97) \quad C_t = \eta_t + \eta_1 C_0 + \eta_2 C_{-1} + \zeta_0 M_t + \zeta_1 M_{t-1} + \dots + \zeta_{t-1} M_1 + \xi_0 G_t + \xi_1 G_{t-1} + \dots + \xi_{t-1} G_1 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_{t-1} \varepsilon_1.$$

Así pues, para unas condiciones iniciales determinadas, C_t se expresa únicamente en términos de los valores presentes y pasados de las variables endógenas y de las perturbaciones. (Si el sistema es estable, para valores grandes de t , η_t se hace constante.) Podrían obtenerse también expresiones similares para las demás variables endógenas, y estas ecuaciones se denominan *forma final* del sistema de ecuaciones. Estas expresiones muestran la forma en que las trayectorias temporales de las variables exógenas determinan la trayectoria de cada variable endógena, y los coeficientes de cada variable exógena son los llamados *multiplicadores dinámicos*. Obsérvese que las ecuaciones del tipo de la (13.97) pueden utilizarse también para resolver cuestiones acerca de la influencia de alguna medida de política económica del pasado sobre el valor presente de las variables endógenas, o de la amplitud de las medidas necesarias para la consecución de determinados cambios en las variables endógenas durante cualquier número de períodos futuros.

Condiciones de estabilidad y análisis dinámico

Pasemos ahora a considerar el problema de la estabilidad de los sistemas. En general, decimos que un sistema es estable si, cuando los valores de las variables exógenas se mantienen constantes a lo largo del tiempo, los valores medios

de las variables endógenas toman ciertos niveles constantes (naturalmente, los valores de las variables endógenas fluctuarán a consecuencia de los efectos de las perturbaciones estocásticas, pero sólo consideraremos casos en los que la probabilidad de que un sistema que sería estable se convierta en inestable a causa de las perturbaciones aleatorias es muy pequeña). Esto significa que un sistema se considera inestable si los valores medios de las variables endógenas toman valores cada vez mayores o menores sin límite, o presentan un movimiento oscilatorio regular, cuando las variables exógenas toman valores constantes. Un procedimiento para conocer si un sistema es estable o no consiste en estudiar las ecuaciones de la forma final. Evidentemente, para que el sistema tienda a unos valores constantes cuando los niveles de las variables exógenas no varían, la suma de cada conjunto de multiplicadores dinámicos tiene que ser finita. Es decir, la condición de estabilidad es que las sumas

$$\sum_{i=0}^{\infty} \zeta_i \quad \text{y} \quad \sum_{i=0}^{\infty} \xi_i$$

sean finitas. Estas sumas representan los *multiplicadores a largo plazo* o de *equilibrio* para el consumo, respecto a la oferta monetaria y al gasto público. Otro procedimiento para saber si el sistema es estable o no consiste en examinar la ecuación dinámica fundamental. Si las variables exógenas se mantienen constantes (y prescindimos de las perturbaciones), esta ecuación se convierte en una ecuación lineal no homogénea en diferencias finitas del tipo corriente, y puede resolverse por los procedimientos adecuados.²² En nuestro caso, según (13.93a), tenemos

$$(13.98) \quad C_t - \mu_1 C_{t-1} - \mu_2 C_{t-2} = \text{constante}.$$

La ecuación característica para (13.98) es

$$(13.99) \quad \lambda^2 - \mu_1 \lambda - \mu_2 = 0$$

$$\text{y sus raíces son} \quad \lambda_{1,2} = \frac{\mu_1 \pm \sqrt{\mu_1^2 + 4\mu_2}}{2}.$$

El sistema es estable si, y solamente si, el valor absoluto (o módulo) de la mayor de las raíces de (13.99) es menor que la unidad. Si es así, podemos calcular el valor medio de equilibrio del consumo a partir de (13.93a) en la forma

$$(13.100) \quad C_E = \frac{\delta + \kappa_0 M_t + \kappa_1 M_{t-1} + \nu_0 G_t + \nu_1 G_{t-1}}{1 - \mu_1 - \mu_2}.$$

22. Ver, por ejemplo, R. G. D. Allen, *Mathematical Economics*. Londres, Macmillan, 1955, pp. 176-95.

La condición de estabilidad es de gran importancia desde el punto de vista económico. Además, es importante tener presente que en el proceso de estimación se supone en realidad la existencia de estabilidad o, en el peor de los casos, de un movimiento oscilatorio regular. El supuesto de que las variables predeterminadas del sistema tienen variancias finitas cuando $t \rightarrow \infty$ se aplica también a las variables endógenas retardadas, y este supuesto no se cumpliría si dichas variables endógenas aumentasen o disminuyesen sin ningún límite. De no efectuar estos supuestos existirían dificultades para demostrar las propiedades asintóticas de los estimadores.

La solución dinámica indicada por las ecuaciones de la forma final permite responder a las preguntas acerca de la influencia de cambios hipotéticos en las variables exógenas del sistema. Si lo que nos interesa son los efectos de cambios que tuvieron lugar realmente durante el período observado en los valores de las variables exógenas, los podemos apreciar de forma más conveniente utilizando primeras diferencias. En nuestro modelo, a partir de (13.97) tenemos

$$(13.101) \quad C_t - C_{t-1} = (\eta_t - \eta_{t-1}) + \zeta_0(M_t - M_{t-1}) + \zeta_1(M_{t-1} - M_{t-2}) \\ + \dots + \zeta_{t-2}(M_2 - M_1) + \zeta_{t-1}M_1 + \xi_0(G_t - G_{t-1}) \\ + \xi_1(G_{t-1} - G_{t-2}) + \dots + \xi_{t-2}(G_2 - G_1) \\ + \xi_{t-1}G_1 + \omega_t.$$

El primero de los términos del lado derecho indica la influencia de las condiciones iniciales y de la dinámica del sistema en ausencia de cambios en las variables exógenas. En los sistemas estables, el valor de este término disminuirá a medida que el sistema se mueve desde su posición inicial, y tenderá hacia el valor cero. Los términos $\zeta_{t-1}M_1$ y $\xi_{t-1}G_1$ miden la influencia de los valores iniciales de las variables exógenas, y el término ω_t resume la influencia de las perturbaciones aleatorias. Sustituyendo los símbolos por los valores reales de las variables exógenas correspondientes al período investigado, en (13.101), podemos calcular la parte debida a cada medida de política económica en el cambio observado en el nivel de consumo.

La exposición anterior se ha efectuado en términos de los parámetros de la población que, naturalmente, son desconocidos y, por tanto, deben sustituirse por estimaciones muestrales. En consecuencia, sería deseable presentar el error estándar junto a cada estimación. Existen fórmulas para la estimación de los errores estándar de los multiplicadores instantáneos de las ecuaciones de la forma reducida, pero no para los multiplicadores dinámicos de las ecuaciones de la forma final. Por lo que respecta a las raíces características, que son cruciales para la determinación de la condición de estabilidad de un sistema, Theil y Boot han obtenido una fórmula para la variancia asintótica de

la raíz mayor.²³ Sin embargo, desde el punto de vista de la inferencia estadística, el análisis dinámico todavía no se ha desarrollado del todo en la econometría.

EJEMPLO. El modelo I de Klein, correspondiente a la economía de los Estados Unidos para el período 1921-1941, se ha estimado de la forma siguiente²⁴.

$$\begin{aligned} \text{Consumo:} & C_t = 16,555 + 0,017 P_t + 0,216 P_{t-1} + 0,810 W_t + \tilde{u}_{1t}; \\ \text{Inversión:} & I_t = 20,278 + 0,150 P_t + 0,616 P_{t-1} - 0,158 K_{t-1} + \tilde{u}_{2t}; \\ \text{Salarios privados:} & W_t^* = 1,500 + 0,439 E_t + 0,147 E_{t-1} + 0,130 A_t + \tilde{u}_{3t}; \\ \text{Producto:} & Y_t + T_t = C_t + I_t + G_t; \\ \text{Renta:} & Y_t = P_t + W_t; \\ \text{Capital:} & K_t = K_{t-1} + I_t; \\ \text{Salarios:} & W_t = W_t^* + W_t^{**}; \\ \text{Producto privado:} & E_t = Y_t + T_t - W_t^{**}. \end{aligned}$$

Variables endógenas:

C	consumo	Y	renta nacional
I	inversión	K	cantidad de capital a fin de año
W*	importe de los salarios privados	W	importe total de los salarios
P	beneficios	E	producto privado

Variables exógenas:

W**	importe de los salarios en el sector público
T	impuestos indirectos
A	tiempo en años (1931 = 1)

Todas las variables menos el tiempo se miden en miles de millones de dólares constantes (1934). Nos interesa especialmente la variable renta, y la ecuación dinámica fundamental estimada para esta variable es

$$\begin{aligned} Y_t &= 1.726 Y_{t-1} + 1.029 Y_{t-2} - 0.183 Y_{t-3} \\ &= 4.880 + 1.773 G_t - 1.493 G_{t-1} + 0.154 A_t - 0.294 A_{t-1} + 0.162 A_{t-2} \\ &\quad - 1.254 T_t + 0.673 T_{t-1} + 0.213 T_{t-2} + 0.183 T_{t-3} + 0.663 W_t^{**} \\ &\quad - 1.443 W_{t-1}^{**} + 1.029 W_{t-2}^{**} - 0.183 W_{t-3}^{**} + \tilde{\omega}_t. \end{aligned}$$

23. H. Theil y J. C. G. Boot, «The Final Form of Econometric Equation Systems», *Review of the International Statistical Institute*, vol. 30, 1962, pp. 136-52.

24. Este modelo se ha presentado en L. R. Klein, *Economic Fluctuations in the United States 1921-1941*. Nueva York, Wiley, 1950, pp. 65-68. Las estimaciones de los coeficientes son estimaciones minimocuadráticas en dos etapas presentadas en A. S. Goldberger, *op. cit.* (1964), p. 365.

714 ELEMENTOS DE ECONOMETRIA

La ecuación característica

$$\lambda^3 - 1.726\lambda^2 + 1.029\lambda - 0.183 = 0$$

tiene las raíces

$$\lambda_1 = 0.310 \quad y \quad \lambda_{2,3} = 0.708 \pm 0.298i.$$

El módulo de las raíces complejas es $\sqrt{0.708^2 + 0.298^2} = 0.768$, que es menor que la unidad. Por tanto, según las estimaciones el sistema resulta estable. El cuadro 13.3 presenta los multiplicadores dinámicos del gasto público respecto a la renta (Y) o al producto nacional neto ($Y + T$). Por último, nos interesa saber la forma en que el gasto público afectó a los cambios en el producto nacional neto durante el período 1921-1941. El cuadro 13-4 presenta los cambios reales en ($Y + T$) y los cambios estimados que se consideraran debidos a los cambios en el gasto público, en el mismo período y en los anteriores. Esta última columna se calculó, según (13.101), en la forma

$$\xi_0(G_t - G_{t-1}) + \xi_1(G_{t-1} - G_{t-2}) + \dots + \xi_{t-2}(G_2 - G_1).$$

El cuadro 13-4 puede utilizarse para examinar la medida de los efectos de la política fiscal anticíclica durante el período estudiado.

Cuadro 13.3 Multiplicadores dinámicos de G sobre ($Y + T$)

Retardo	Multiplicador
0	1.773
1	1.567
2	0.881
3	0.232
4	-0.219
5	-0.456
6	-0.519
7	-0.466
8	-0.355
9	-0.227
10	-0.112
11	-0.025
12	0.031
13	0.058
14	0.064
15	0.057
16	0.042
17	0.027
18	0.013
19	0.002
20	-0.004
Total	2.071

Cuadro 13.4 Cambios en ($Y + T$) a consecuencia de los cambios actuales y pasados de G

Año	$\Delta(Y + T)$	Debidos a los cambios de G
1922	4.7	-0.9
1923	7.1	-1.4
1924	0.1	0.5
1925	4.0	0.8
1926	3.1	0.8
1927	0.7	2.4
1928	0.2	2.4
1929	2.8	1.8
1930	-5.6	3.0
1931	-7.2	4.2
1932	-8.6	1.6
1933	1.1	-1.8
1934	5.0	-1.4
1935	4.8	-0.5
1936	9.6	-0.8
1937	1.6	0.4
1938	-3.1	3.9
1939	8.7	5.7
1940	6.4	5.6
1941	13.2	15.2

Comentarios finales

Para concluir, podemos llamar la atención del lector sobre algunas de las principales lagunas que todavía existen. La mayor parte de los estudios sobre la estimación de los modelos de ecuaciones simultáneas han utilizado los supuestos básicos del modelo clásico de regresión lineal y existe una clara necesidad de estudiar el problema de la estimación en los casos en que estos supuestos no se cumplen. Además del problema de la autorregresión en las perturbaciones, que se ha mencionado al final de la sección 13.5, se necesitan estimadores para los modelos de ecuaciones simultáneas con ecuaciones no lineales, y con ecuaciones que contengan retardos distribuidos. Pero el más importante de los problemas es quizás el relativo a los errores de medición y especificación. Lamentablemente, estos problemas aún no se han resuelto de forma totalmente satisfactoria ni siquiera en el contexto de los modelos uniecuacionales y, por tanto, menos aún en el marco relativamente complejo de las ecuaciones simultáneas.

EJERCICIOS:

13-1. Un modelo keynesiano muy simplificado puede expresarse en la forma

$$C_t = \alpha + \beta Y_t + e_t,$$

$$Y_t = C_t + I_t,$$

en el que Y = renta, C = consumo, e I = inversión exógena. Mostrar que los métodos de estimación por MCI, MC2E, MVIL, y MVIC producen resultados idénticos.

13-2. Considerar el siguiente modelo de ecuaciones simultáneas:

$$Y_{1t} = \beta_{12}Y_{2t} + \gamma_{11} + u_{1t},$$

$$Y_{2t} = \gamma_{21} + \gamma_{22}X_t + u_{2t},$$

en el que X es una variable exógena no estocástica. Se supone que las perturbaciones tienen media cero y variancias constantes, y no son autorregresivas a lo largo del tiempo. Además, u_{1t} y u_{2t} son independientes entre sí. Comparar las variancias del estimador minimocuadrático ordinario y del estimador minimocuadrático indirecto de β_{12} .

13-3. El modelo de Tintner para el mercado de carne en los Estados Unidos es

$$Y_{1t} = a_0 + a_1Y_{2t} + a_2X_{1t} + u_{1t} \quad (\text{demanda}),$$

$$Y_{1t} = b_0 + b_1Y_{2t} + b_2X_{2t} + b_3X_{3t} + u_{2t} \quad (\text{oferta}),$$

en el que u_{1t} y u_{2t} son los términos de perturbación, Y_{1t} e Y_{2t} son las variables endógenas, y X_{1t} , X_{2t} , y X_{3t} son las variables predeterminadas. Estas variables se definen de la forma siguiente:

- Y_1 consumo de carne per cápita;
- Y_2 índice de precios de la carne;
- X_1 renta disponible per cápita;
- X_2 índice de costes de tratamiento de la carne;
- X_3 índice de costes de producción agrícola.

Se supone que los vectores de orden $T \times 1$ de las perturbaciones, u_1 y u_2 , tienen la siguiente matriz de covariancias:

$$E \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} u_1' & u_2' \end{bmatrix} = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T \end{bmatrix}.$$

- a. Estudiar la identificación de las dos ecuaciones.
- b. Supóngase que se sabe a priori que $b_2/b_3 = K$, siendo K un número conocido. Estudiar la identificación del modelo cuando se añade esta especificación.
- c. Supóngase a cambio que se añade el término a_3X_{4t} a la ecuación de demanda, siendo X_4 un índice de precios de las grasas y proteínas no animales. ¿En qué forma altera este nuevo término la identificación del modelo?

13-4. Supóngase el siguiente modelo agregado de la economía:

$$C_t = \alpha_1 + \alpha_2 D_t + u_{1t},$$

$$I_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_{t-1} + u_{2t},$$

$$Y_t = D_t + T_t,$$

$$Y_t = C_t + I_t + G_t,$$

en el que C = consumo, D = renta nacional neta, I = inversión, Y = producto nacional neto, T = impuestos indirectos, y G = gastos autónomos. Las variables C , D , I , e Y son endógenas, y las T y G son exógenas.

- a. Examinar las condiciones de identificación de las dos primeras ecuaciones del modelo.
- b. Obtener estimaciones consistentes de las ecuaciones de consumo e inversión utilizando los datos del cuadro 13-5, correspondientes a la economía de los Estados Unidos durante el periodo 1921-1941.

Cuadro 13.5 *

	1920	C	D	Y	I
1920	—	—	—	47.1	—
1921	41.9	40.6	48.3	—0.2	—
1922	45.0	49.1	53.0	1.9	—
1923	49.2	55.4	60.1	5.2	—
1924	50.6	56.4	60.2	3.0	—
1925	52.6	58.7	64.2	5.1	—
1926	55.1	60.3	67.3	5.6	—
1927	56.2	61.3	68.0	4.2	—
1928	57.3	64.0	68.2	3.0	—
1929	57.8	67.0	71.0	5.1	—
1930	55.0	57.7	65.4	1.0	—
1931	50.9	50.7	58.2	-3.4	—
1932	45.6	41.3	49.6	-6.2	—
1933	46.5	45.3	50.7	-5.1	—
1934	48.7	48.9	55.7	-3.0	—
1935	51.3	53.3	60.5	-1.3	—
1936	57.7	61.8	70.1	2.1	—
1937	58.7	65.0	71.7	2.0	—
1938	57.5	61.2	68.6	-1.9	—
1939	61.6	68.4	77.3	1.3	—
1940	65.0	74.1	83.7	3.3	—
1941	69.7	85.3	96.9	4.9	—

Fuente: L. R. Klein, *Economic Fluctuations in the United States 1921-1941*. Nueva York, Wiley, 1950, p. 135.

* Todas las variables medidas en miles de millones de dólares de 1934.

13-5. Utilizando los resultados del ejercicio 13-4,

- a. Obtener la ecuación dinámica fundamental y utilizarla para determinar si el sistema es estable o no.
- b. Calcular los multiplicadores dinámicos de Y respecto a G .
- c. Obtener el valor del multiplicador a largo plazo (de equilibrio) de Y con respecto a G .

13-6. Considerar el siguiente modelo dinámico del mercado de cierto producto:

$$S_t = \alpha_0 + \alpha_1 P_t + u_{1t},$$

$$D_t = \beta_0 + \beta_1 P_t + u_{2t},$$

$$P_t - P_{t-1} = m(Q_{t-1} - Q_{t-2}) + u_{3t},$$

$$S_t - D_t = Q_t - Q_{t-1},$$

en el que S_t = producción, D_t = ventas, P_t = precio, y Q_t = nivel de existencias al final del período. Obsérvese que

$$\alpha_1 > 0,$$

$$\beta_1 < 0,$$

$$\alpha_0 < \beta_0,$$

$$m < 0.$$

- a. Determinar la identificabilidad de las tres primeras ecuaciones.
- b. Mostrar que la trayectoria temporal de Q_t está representada por una ecuación homogénea en diferencias finitas de segundo orden, y la de P_t por una ecuación no homogénea en diferencias finitas de primer orden.
- c. Obtener las condiciones para la existencia de equilibrio en este mercado.
- d. Obtener el valor de m para el que la trayectoria temporal de P_t consiste en una serie de oscilaciones regulares (prescindiendo de las perturbaciones estocásticas).

APENDICE

A. Algebra de sumatorios

Las expresiones que incluyen sumas se utilizan ampliamente en estadística y, generalmente, se abrevian mediante la llamada «notación Σ ». La definición básica de esta notación señala que, si m y n son enteros y $m \leq n$, tenemos

$$\sum_{i=m}^n x_i = x_m + x_{m+1} + x_{m+2} + \cdots + x_n.$$

Esta notación se utiliza muy a menudo para las sumas de las observaciones de una variable o variables determinadas, con el subíndice indicando la numeración de las observaciones. Al utilizar esta notación, es útil conocer las reglas siguientes:

$$(A.1) \quad \sum_{i=1}^n kx_i = k \sum_{i=1}^n x_i.$$

Demostración:

$$\sum_{i=1}^n kx_i = kx_1 + kx_2 + \cdots + kx_n = k(x_1 + x_2 + \cdots + x_n) = k \sum_{i=1}^n x_i.$$

$$(A.2) \quad \sum_{i=1}^n k = nk.$$

Demostración:

$$\sum_{i=1}^n k = k + k + \cdots + k = nk.$$

$$(A.3) \quad \sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i.$$

Demostración:

$$\begin{aligned} \sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n) \\ &= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i. \end{aligned}$$

Tomando la suma de los primeros n números enteros positivos y las de sus potencias se obtienen fórmulas de gran utilidad. Los resultados, presentados sin demostración, son

$$(A.4) \quad \sum_{i=1}^n i = 1 + 2 + \cdots + n = \frac{1}{2}[n(n+1)] = \frac{1}{2}[(n+1)^2 - (n+1)],$$

$$(A.5) \quad \sum_{i=1}^n i^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{1}{6}[n(n+1)(2n+1)] \\ = \frac{1}{6}[(n+1)^3 - \frac{1}{2}(n+1)^2 + \frac{1}{2}(n+1)],$$

$$(A.6) \quad \sum_{i=1}^n i^3 = 1^3 + 2^3 + \cdots + n^3 = \frac{1}{4}[n^2(n+1)^2] \\ = \frac{1}{4}[(n+1)^4 - 2(n+1)^3 + (n+1)^2],$$

$$(A.7) \quad \sum_{i=1}^n i^4 = 1^4 + 2^4 + \cdots + n^4 = \frac{1}{30}[n(n+1)(2n+1)(3n^2+3n-1)] \\ = \frac{1}{30}[(n+1)^5 - \frac{1}{2}(n+1)^4 + \frac{1}{3}(n+1)^3 - \frac{1}{6}(n+1)].$$

La notación Σ puede extenderse también a sumas múltiples. Por ejemplo, el doble sumatorio se define como

$$\sum_{i=1}^n \sum_{j=1}^m x_{ij} = \sum_{i=1}^n (x_{i1} + x_{i2} + \cdots + x_{im}) \\ = (x_{11} + x_{21} + \cdots + x_{n1}) + (x_{12} + x_{22} + \cdots + x_{n2}) \\ + \cdots + (x_{1m} + x_{2m} + \cdots + x_{nm}).$$

En los dobles sumatorios, tienen especial interés las reglas siguientes:

$$(A.8) \quad \sum_{i=1}^n \sum_{j=1}^m (x_{ij} + y_{ij}) = \sum_{i=1}^n \sum_{j=1}^m x_{ij} + \sum_{i=1}^n \sum_{j=1}^m y_{ij}.$$

Demostración:

$$\sum_{i=1}^n \sum_{j=1}^m (x_{ij} + y_{ij}) = \sum_{j=1}^m (x_{1j} + y_{1j}) + \sum_{j=1}^m (x_{2j} + y_{2j}) + \cdots + \sum_{j=1}^m (x_{nj} + y_{nj}) \\ = \sum_{j=1}^m (x_{1j} + x_{2j} + \cdots + x_{nj}) + \sum_{j=1}^m (y_{1j} + y_{2j} + \cdots + y_{nj}) \\ = \sum_{i=1}^n \sum_{j=1}^m x_{ij} + \sum_{i=1}^n \sum_{j=1}^m y_{ij}. \\ (A.9) \quad \sum_{i=1}^n \sum_{j=1}^m x_i = m \sum_{i=1}^n x_i.$$

Demostración:

$$\sum_{i=1}^n \sum_{j=1}^m x_i = \sum_{j=1}^m \left[\sum_{i=1}^n x_i \right] = m \sum_{i=1}^n x_i.$$

$$(A.10) \quad \sum_{i=1}^n \sum_{j=1}^m x_i y_j = \left[\sum_{i=1}^n x_i \right] \left[\sum_{j=1}^m y_j \right].$$

Demostración:

$$\sum_{i=1}^n \sum_{j=1}^m x_i y_j = \sum_{i=1}^n (x_i y_1 + x_i y_2 + \cdots + x_i y_m) \\ = y_1 \sum_{i=1}^n x_i + y_2 \sum_{i=1}^n x_i + \cdots + y_m \sum_{i=1}^n x_i = \left[\sum_{i=1}^n x_i \right] \left[\sum_{j=1}^m y_j \right].$$

$$(A.11) \quad \sum_{i=1}^n \sum_{j=1}^m x_i y_{ij} = \sum_{i=1}^n x_i \sum_{j=1}^m y_{ij}.$$

Demostración:

$$\sum_{i=1}^n \sum_{j=1}^m x_i y_{ij} = \sum_{i=1}^n (x_i y_{i1} + x_i y_{i2} + \cdots + x_i y_{im}) = \sum_{i=1}^n x_i (y_{i1} + y_{i2} + \cdots + y_{im}) \\ = \sum_{i=1}^n x_i \sum_{j=1}^m y_{ij}.$$

$$(A.12) \quad \left[\sum_{i=1}^n x_i \right]^2 = \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i x_j = \sum_{i=1}^n x_i^2 + 2 \sum_{i < j} x_i x_j.$$

Demostración:

$$\left[\sum_{i=1}^n x_i \right]^2 = (x_1 + x_2 + \cdots + x_n)^2 \\ = x_1^2 + x_2^2 + \cdots + x_n^2 + 2(x_1 x_2 + x_1 x_3 + \cdots + x_1 x_n \\ + x_2 x_3 + x_2 x_4 + \cdots + x_{n-1} x_n) = \sum_{i=1}^n x_i^2 + 2 \sum_{i < j} x_i x_j.$$

Un caso especial de (A.12) es

$$(A.13) \quad \sum_{i=1}^n (x_i - \bar{x})^2 = -2 \sum_{i < j} (x_i - \bar{x})(x_j - \bar{x}).$$

Demostración: Como

$$\sum_{i=1}^n (x_i - \bar{x}) = 0,$$

tenemos

$$\left[\sum_{i=1}^n (x_i - \bar{x}) \right]^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + 2 \sum_{i < j} (x_i - \bar{x})(x_j - \bar{x}) = 0,$$

$$\text{o} \quad \sum_{i=1}^n (x_i - \bar{x})^2 = -2 \sum_{i < j} (x_i - \bar{x})(x_j - \bar{x}).$$

B. Elementos de álgebra matricial

Definiciones

Una matriz ($M \times N$) se define como una ordenación rectangular de números reales en M filas y N columnas en la forma

$$(B.1) \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix}.$$

Los números a_{ij} ($i = 1, 2, \dots, M$; $j = 1, 2, \dots, N$) son los *elementos* de A ; generalmente se utiliza el término a_{ij} para indicar el elemento representativo de A . El número de filas y columnas, que en el caso de (B.1) se ha representado por $M \times N$, se denomina *orden* o *dimensión* de la matriz. Una matriz de orden 1×1 se denomina *escalar*, las de orden $M \times 1$ se denominan *vectores columna* y las de orden $1 \times N$, *vectores fila*. Las matrices con el mismo número de filas y columnas se llaman *matrices cuadradas*. La suma de los elementos de la diagonal principal de una matriz cuadrada se llama *traza* de la matriz. Una *matriz diagonal* es una matriz cuadrada en la que todos los elementos que no pertenecen a la diagonal principal son ceros; es decir,

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{MM} \end{bmatrix}.$$

Una *matriz unidad* o *identidad* es una matriz diagonal en la que todos los elementos de la diagonal principal son unos; es decir,

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

En ocasiones se utiliza un subíndice para indicar el orden de la matriz identidad a la que se refiere, por ejemplo, I_{ij} . Por último, una *matriz de ceros* es cualquier matriz cuyos elementos son todos ceros.

Operaciones básicas con matrices

(B.2) **Igualdad de matrices** $A = B$ si, y solamente si, A y B son del mismo orden y $a_{ij} = b_{ij}$ para todo i, j .

(B.3) **Suma de matrices** $A + B = C$ si, y solamente si, A , B , y C son del mismo orden y $a_{ij} + b_{ij} = c_{ij}$ para todo i, j .

EJEMPLO

$$\begin{bmatrix} 5 & 0 \\ 2 & -4 \end{bmatrix} + \begin{bmatrix} -1 & 3 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 5 & 1 \end{bmatrix}.$$

(B.4) **Multipliación escalar.** Si k es un escalar, $kA = [ka_{ij}]$.

Esto significa que para multiplicar una matriz A por un escalar tenemos que multiplicar cada elemento de A por este escalar.

(B.5) **Multipliación de matrices.** Si A es de orden $M \times N$ y B es de orden $N \times P$, el producto de las dos matrices se define como

$$AB = C.$$

siendo C una matriz de orden $M \times P$ cuyo elemento correspondiente a la fila i y a la columna j viene dado por

$$c_{ij} = \sum_{k=1}^N a_{ik}b_{kj}.$$

Es decir, el elemento c_{11} se obtiene multiplicando los elementos de la primera fila de A por los elementos de la primera columna de B , y sumando todos estos productos; el elemento c_{12} se obtiene efectuando la misma operación con los elementos de la primera fila de A y la segunda columna de B , y así sucesivamente.

EJEMPLO 1

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) \\ (a_{31}b_{11} + a_{32}b_{21}) & (a_{31}b_{12} + a_{32}b_{22}) \end{bmatrix}$$

EJEMPLO 2

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} (1 \times 7 + 2 \times 8 + 3 \times 9) & (1 \times 10 + 2 \times 11 + 3 \times 12) \\ (4 \times 7 + 5 \times 8 + 6 \times 9) & (4 \times 10 + 5 \times 11 + 6 \times 12) \end{bmatrix} \\ = \begin{bmatrix} 50 & 68 \\ 122 & 167 \end{bmatrix}.$$

Obsérvese que, en general,

$$AB \neq BA.$$

Esta desigualdad implica que debe distinguirse entre premultiplicación y postmultiplicación por una matriz.

Transposición

(B.6) Si A es una matriz ($M \times N$), la transpuesta de A , que se representa por A' , es una matriz ($N \times M$) que se obtiene intercambiando las filas por las columnas en la matriz A ; es decir,

$$A' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{M1} \\ a_{12} & a_{22} & \dots & a_{M2} \\ \vdots & \vdots & & \vdots \\ a_{1N} & a_{2N} & \dots & a_{MN} \end{bmatrix}.$$

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 6 \\ 4 & 7 \end{bmatrix}, \quad A' = \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix}.$$

Las matrices transpuestas cumplen los siguientes teoremas:

$$(B.7) \quad (A')' = A,$$

$$(B.8) \quad (A + B)' = A' + B',$$

$$(B.9) \quad (AB)' = B'A'.$$

Las demostraciones resultan inmediatas a partir de las definiciones de las operaciones con matrices.

(B.10) Si A es una matriz cuadrada y $A = A'$, decimos que A es una matriz simétrica.

Una expresión muy conocida en la que se utiliza un vector transpuesto recibe el nombre de *forma cuadrática*. Supongamos que x es un vector ($M \times 1$) distinto de cero, y A es una matriz cuadrada simétrica, de orden $M \times M$. En este caso, la forma cuadrática de x se define como

$$(B.11) \quad x'Ax = \sum_{i=1}^M \sum_{j=1}^M a_{ij}x_i x_j.$$

EJEMPLO. Si

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad y \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

tenemos

$$\begin{aligned} x'Ax &= [x_1 \ x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1 \ x_2] \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_1x_2 + a_{22}x_2^2 = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2. \end{aligned}$$

Se dice que una matriz simétrica A es *positiva definida* si $x'Ax > 0$, y *positiva semidefinida* si $x'Ax \geq 0$. Del mismo modo, decimos que A es *negativa definida* si $x'Ax < 0$, y *negativa semidefinida* si $x'Ax \leq 0$. Obsérvese que todas las matrices de variancia y covariancias son positivas definidas.

Matrices particionadas

En muchas ocasiones resulta conveniente *particionar* una matriz en varias submatrices. En este caso, las submatrices se tratan como elementos escalares, salvo por el hecho de que debemos asegurarnos de que se respetan las reglas para la multiplicación de matrices. En particular, si

$$A = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \quad y \quad B = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

en las que las dimensiones de las submatrices son

$$\begin{aligned} P_{11} &\rightarrow (M_1 \times N_1), & Q_{11} &\rightarrow (N_1 \times R_1), \\ P_{12} &\rightarrow (M_1 \times N_2), & Q_{12} &\rightarrow (N_1 \times R_2), \\ P_{21} &\rightarrow (M_2 \times N_1), & Q_{21} &\rightarrow (N_2 \times R_1), \\ P_{22} &\rightarrow (M_2 \times N_2), & Q_{22} &\rightarrow (N_2 \times R_2), \end{aligned}$$

tenemos

$$(B.12) \quad A' = \begin{bmatrix} P'_{11} & P'_{21} \\ P'_{12} & P'_{22} \end{bmatrix}$$

y

$$(B.13) \quad AB = \begin{bmatrix} P_{11}Q_{11} + P_{12}Q_{21} & P_{11}Q_{12} + P_{12}Q_{22} \\ P_{21}Q_{11} + P_{22}Q_{21} & P_{21}Q_{12} + P_{22}Q_{22} \end{bmatrix}.$$

Determinantes

Los determinantes son escalares cuyo valor se obtiene a partir de una matriz cuadrada, siguiendo unas reglas determinadas. El determinante de la matriz A se representa por $\det A$, o $|A|$. Las reglas para la obtención de los determinantes son las siguientes. En una matriz (2×2)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

el determinante viene dado por

$$(B.14) \quad \det A = a_{11}a_{22} - a_{12}a_{21}.$$

Obsérvese que el valor de $\det A$ se ha obtenido multiplicando determinados elementos de A por otros elementos y dotando de signo positivo o negativo al producto resultante. Utilizando líneas continuas para representar los productos con signo

positivo y líneas de puntos para los productos con signo negativo, podemos representar gráficamente la regla para el cálculo del determinante de una matriz (2×2) en la forma de la figura B-1 a).

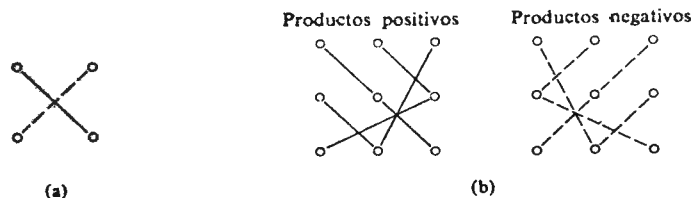


Figura B-1

En la matriz (3×3)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

el determinante es

$$(B.15) \quad \det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{21}a_{12},$$

y se representa gráficamente en la figura B-1 b).

La representación gráfica de las reglas de cálculo de los determinantes de las matrices de orden superior resulta confusa y se utiliza muy raramente. En su lugar, utilizaremos una regla general para calcular los determinantes de las matrices de cualquier orden. Esta regla incluye el uso de los *menores* y *cofactores*. Supongamos que en una matriz A de orden $M \times M$ eliminamos la fila i -ésima y la columna j -ésima, es decir, la fila y columna correspondientes al elemento a_{ij} . El determinante de la matriz $(M-1) \times (M-1)$ resultante se denomina *menor* de a_{ij} , y el *cofactor* de a_{ij} es simplemente este menor multiplicado por $(-1)^{i+j}$. La forma habitual para indicar el cofactor de a_{ij} es A_{ij} . La regla general para obtener el determinante de A consiste en tomar cualquier fila o cualquier columna de dicha matriz, multiplicar cada elemento por su cofactor correspondiente, y sumar los resultados. Por tanto, el determinante de una matriz A de orden $M \times M$ puede calcularse en la forma

$$(B.16) \quad \det A = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1M}A_{1M}.$$

En la fórmula (B.16) hemos utilizado la primera fila como base para los cálculos pero cualquier otra fila o cualquier columna también hubiese servido. La importancia de la «fórmula de los cofactores» (B.16) es que nos permite reducir un determinante de cualquier orden a una combinación lineal de determinantes de orden inferior. Naturalmente, esta fórmula se aplica también a los casos $M = 2$ y $M = 3$.

los que hemos dado fórmulas especiales en (B.14) y (B.15). En el caso de $M = 2$, tenemos

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

$$A_{11} = a_{22},$$

$$A_{12} = -a_{21},$$

o sea, que, según (B.16),

$$\det A = a_{11}(a_{22}) + a_{12}(-a_{21}) = a_{11}a_{22} - a_{12}a_{21},$$

que es la misma respuesta de la fórmula (B.14). Además, cuando $M = 3$ tenemos

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

$$A_{11} = \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} = a_{22}a_{33} - a_{23}a_{32},$$

$$A_{12} = -\det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} = -a_{21}a_{33} + a_{23}a_{31},$$

$$A_{13} = \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = a_{21}a_{32} - a_{22}a_{31},$$

o sea, que, según (B.16),

$$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(-a_{21}a_{33} + a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31},$$

que es la misma expresión de la fórmula (B.15).

Algunas de las propiedades útiles e interesantes de los determinantes son las siguientes

- (B.17) Si se intercambian dos filas o dos columnas cualesquiera de A , se invierte el signo de $\det A$.
- (B.18) Si se multiplican todos los elementos de una fila o una columna de A por un escalar, $\det A$ queda multiplicado por el mismo escalar.
- (B.19) El valor de $\det A$ queda inalterado si se suma a cualquier fila (o columna) de A un múltiplo de cualquier otra fila (o columna).
- (B.20) El valor del determinante de una matriz que tiene dos filas (o dos columnas) iguales es cero.
- (B.21) $\det A = \det A'$.

$$(B.22) \quad \det AB = (\det A)(\det B).$$

Los determinantes se utilizan también para expresar algunas propiedades de las matrices. En particular, se dice que una matriz A es *singular* cuando $\det A = 0$, y *no singular* cuando $\det A \neq 0$. Además, el orden del mayor determinante distinto de cero contenido en una matriz $(M \times N)$ se llama *rango* de la matriz.

EJEMPLO 1. Si

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix},$$

tenemos

$$\det A = 1 \times 5 \times 10 + 2 \times 6 \times 7 + 3 \times 8 \times 4 - 3 \times 5 \times 7 - 6 \times 8 \times 1 - 10 \times 4 \times 2 = -3.$$

Podríamos obtener el mismo resultado utilizando la «fórmula de los cofactores» (B.16):

$$\begin{aligned} \det A &= 1 \times \det \begin{bmatrix} 5 & 6 \\ 8 & 10 \end{bmatrix} + 2 \times (-1) \det \begin{bmatrix} 4 & 6 \\ 7 & 10 \end{bmatrix} + 3 \times \det \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} \\ &= 1 \times 2 + 2 \times 2 + 3 \times (-3) = -3. \end{aligned}$$

Obsérvese también que

$$A' = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} \quad \text{y} \quad \det A' = -3.$$

Supongamos que efectuamos los cambios siguientes en la matriz A , a) multiplicamos la primera fila por 4 y restamos este producto de la segunda fila, b) multiplicamos la primera fila por 7 y la restamos de la tercera. La matriz resultante, A^* , es

$$A^* = \begin{bmatrix} 1 & 2 & 3 \\ (4-4) & (5-8) & (6-12) \\ (7-7) & (8-14) & (10-21) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix}$$

$$\text{y} \quad \det A^* = 1 \det \begin{bmatrix} -3 & -6 \\ -6 & -11 \end{bmatrix} = -3 = \det A.$$

EJEMPLO 2. Sean

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}.$$

En este caso

$$\det A = 1 \times 4 - 2 \times 3 = -2,$$

$$\det B = 5 \times 8 - 6 \times 7 = -2.$$

Definamos ahora

$$C = AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

y observemos que

$$\det C = 950 - 946 = 4 = (\det A)(\det B)$$

Inversión de matrices

La *inversa* de una matriz cuadrada A es una matriz tal que, si se premultiplica o postmultiplica por A , da como resultado la matriz identidad. La matriz inversa de A se representa por A^{-1} y, por tanto, podemos definir

$$(B.23) \quad B = A^{-1} \text{ si, y solamente si, } BA = AB = I.$$

No todas las matrices tienen inversa, y el teorema siguiente establece las condiciones en las que la inversa de una matriz existe:

$$(B.24) \quad \text{La matriz } A \text{ tiene inversa si, y solamente si, } \det A \neq 0, \text{ es decir, si y solamente si } A \text{ es una matriz no singular.}$$

Para calcular la inversa de una matriz debemos formar *adjunta*. La matriz adjunta de una matriz cuadrada A se forma sustituyendo cada elemento de A por su cofactor correspondiente y transponiendo después este resultado. Si representamos por adjunta A esta matriz adjunta, tenemos

$$(B.25) \quad \text{adj } A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1M} \\ A_{21} & A_{22} & \dots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \dots & A_{MM} \end{bmatrix}' = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{M1} \\ A_{12} & A_{22} & \dots & A_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1M} & A_{2M} & \dots & A_{MM} \end{bmatrix}.$$

A continuación, la inversa de A se obtiene de la forma siguiente:

$$(B.26) \quad A^{-1} = \frac{1}{\det A} \text{adj } A.$$

Puede mostrarse¹ que

$$(B.27) \quad (AB)^{-1} = B^{-1}A^{-1},$$

$$(B.28) \quad (A^{-1})' = (A')^{-1}.$$

Obsérvese que el resultado de (B.28) implica que si A es simétrica (y no singular), A^{-1} también es simétrica.

EJEMPLO 1. Si

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

¹ Véase, por ejemplo, J. Johnston, *Econometric Methods* (Nueva York: McGraw-Hill, 1961), pp. 85-86.

tenemos

$$A_{11} = a_{22},$$

$$A_{12} = -a_{21},$$

$$A_{21} = -a_{12},$$

$$A_{22} = a_{11},$$

y

$$\text{adj } A = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}' = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Por tanto,

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

EJEMPLO 2. Si

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix},$$

tenemos

$$A_{11} = 2, \quad A_{21} = 4, \quad A_{31} = -3,$$

$$A_{12} = 2, \quad A_{22} = -11, \quad A_{32} = 6,$$

$$A_{13} = -3, \quad A_{23} = 6, \quad A_{33} = -3,$$

y

$$\det A = -3.$$

Por tanto,

$$\text{adj } A = \begin{bmatrix} 2 & 2 & -3 \\ 4 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}' = \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix},$$

y

$$A^{-1} = \frac{1}{-3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} & -\frac{4}{3} & 1 \\ -\frac{2}{3} & \frac{11}{3} & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

Para comprobar este resultado, multipliquemos

$$\begin{aligned} A^{-1}A &= \begin{bmatrix} -\frac{2}{3} & -\frac{4}{3} & 1 \\ -\frac{2}{3} & \frac{11}{3} & -2 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \\ &= \begin{bmatrix} (-\frac{2}{3} - \frac{16}{3} + 7) & (-\frac{8}{3} - \frac{20}{3} + 8) & (-\frac{2}{3} - \frac{20}{3} + 10) \\ (-\frac{2}{3} + \frac{44}{3} - 14) & (-\frac{8}{3} + \frac{55}{3} - 16) & (-\frac{2}{3} + \frac{22}{3} - 2) \\ (1 - 8 + 7) & (2 - 10 + 8) & (3 - 12 + 10) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Inversión particionada

En ocasiones podemos estar interesados en obtener la inversa de una matriz en forma particionada. Supongamos que se particiona A en la forma

$$A = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

en la que las dimensiones de las submatrices son

$$P_{11} \rightarrow M_1 \times M_1,$$

$$P_{12} \rightarrow M_1 \times M_2,$$

$$P_{21} \rightarrow M_2 \times M_1,$$

$$P_{22} \rightarrow M_2 \times M_2.$$

En este caso¹

$$A^{-1} = \begin{bmatrix} P_{11}^{-1} + P_{11}^{-1}P_{12}Q_{22}^{-1}P_{21}P_{11}^{-1} & -P_{11}^{-1}P_{12}Q_{22}^{-1} \\ -Q_{22}^{-1}P_{21}P_{11}^{-1} & Q_{22}^{-1} \end{bmatrix}$$

siendo

$$Q_{22} = P_{22} - P_{21}P_{11}^{-1}P_{12}.$$

Obsérvese que la inversa de una matriz diagonal en bloques es también diagonal en bloques. De forma específica,

$$(B.29) \quad \text{Si } A = \begin{bmatrix} P_{11} & O \\ O & P_{22} \end{bmatrix} \quad \text{tenemos } A^{-1} = \begin{bmatrix} P_{11}^{-1} & O \\ O & P_{22}^{-1} \end{bmatrix}$$

Derivadas en notación matricial

Sea $y = f(x_1, x_2, \dots, x_M)$ un escalar, y x un vector columna definido en la forma

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}.$$

¹ Véase A. S. Goldberger, *Econometric Theory* (Nueva York: Wiley, 1964), p. 27.

En este caso, la primera derivada parcial de y respecto a cada elemento x se define como

$$(B.30) \quad \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_M} \end{bmatrix},$$

y la segunda derivada parcial como

$$(B.31) \quad \frac{\partial^2 y}{\partial \mathbf{x}^2} = \begin{bmatrix} \frac{\partial^2 y}{\partial x_1^2} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 y}{\partial x_1 \partial x_M} \\ \frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2^2} & \dots & \frac{\partial^2 y}{\partial x_2 \partial x_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 y}{\partial x_M \partial x_1} & \frac{\partial^2 y}{\partial x_M \partial x_2} & \dots & \frac{\partial^2 y}{\partial x_M^2} \end{bmatrix}.$$

Estas reglas básicas pueden aplicarse a algunos casos que aparecen muy a menudo. En primer lugar, si

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

en el que a_i ($i = 1, 2, \dots, M$) son constantes, tenemos

$$(B.32) \quad \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}.$$

Esto puede demostrarse fácilmente señalando que

$$\mathbf{a}'\mathbf{x} = \sum_{i=1}^M a_i x_i,$$

y utilizando la regla (B.30). En segundo lugar, si $\mathbf{A}$ es una matriz simétrica de orden $M \times M$ cuyo elemento general es la constante a_{ij} ,

$$(B.33) \quad \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

Esto se sigue del hecho de que

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^M \sum_{j=1}^M a_{ij} x_i x_j,$$

y de la aplicación de la regla de derivación presentada en (B.30). Por último, si $\mathbf{A}$ y $\mathbf{B}$ son dos matrices simétricas de orden $M \times M$ y sus elementos son constantes, se cumple que

$$(B.34) \quad \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \frac{2(\mathbf{A}\mathbf{x})(\mathbf{x}'\mathbf{B}\mathbf{x}) - 2(\mathbf{x}'\mathbf{A}\mathbf{x})(\mathbf{B}\mathbf{x})}{(\mathbf{x}'\mathbf{B}\mathbf{x})^2}.$$

Esto puede demostrarse de la misma forma que (B.33).

C. Diseño del método de cálculo para la estimación minimocuadrática

C-1. Modelo de regresión lineal simple

El esquema de cálculo que se presenta a continuación está construido para la obtención, con gran precisión de cálculo, de las estimaciones minimocuadráticas simples de los parámetros de regresión del modelo

$$Y_i = \alpha + \beta X_i + e_i.$$

El proceso ha sido calculado para llegar a un alto grado de precisión de cómputo.

Cálculos básicos

En primer lugar, se obtienen las sumas de los cuadrados y productos cruzados de X e Y . Estas sumas se llaman generalmente «momentos muestrales respecto al origen» y pueden presentarse en la forma siguiente

	1	X	Y
1	n	$\sum X_i$	$\sum Y_i$
X		$\sum X_i^2$	$\sum X_i Y_i$
Y			$\sum Y_i^2$

Como prueba, se construyen n observaciones para la variable artificial S que se define como

$$S_i = X_i + Y_i.$$

y debemos tener

$$\sum S_i^2 = \sum X_i^2 + \sum Y_i^2 + 2 \sum X_i Y_i.$$

A continuación, se calculan los valores de los «momentos muestrales aumentados respecto a las medias» que se definen en la forma siguiente:

$$M_{xx} = n \sum X_i^2 - (\sum X_i)^2,$$

$$M_{xy} = n \sum X_i Y_i - (\sum X_i)(\sum Y_i),$$

$$M_{yy} = n \sum Y_i^2 - (\sum Y_i)^2,$$

$$M_{ss} = n \sum S_i^2 - (\sum S_i)^2.$$

Estos valores pueden ordenarse en la forma

$$\begin{array}{c|cc} & X & Y \\ \hline X & M_{XX} & M_{XY} \\ Y & & M_{YY} \end{array}$$

Para comprobar los cálculos, se observa si se cumple la igualdad:

$$M_{SS} = M_{XX} + M_{YY} + 2M_{XY}.$$

Estimaciones de los coeficientes de regresión

Las estimaciones minimocuadráticas de los coeficientes de regresión se calculan de la forma siguiente:

$$\hat{\beta} = \frac{M_{XY}}{M_{XX}},$$

$$\hat{\alpha} = \frac{(\sum Y_i - \hat{\beta} \sum X_i)}{n}.$$

Además,

$$nSCR = \hat{\beta} M_{XY},$$

$$nSTC = M_{YY},$$

$$nSCE = nSTC - nSCR,$$

con lo que obtenemos

$$R^2 = \frac{nSCR}{nSTC},$$

$$ns^2 = \frac{nSCE}{n-2},$$

$$s_{\hat{\beta}}^2 = \frac{ns^2}{M_{XX}},$$

$$s_{\hat{\alpha}}^2 = \frac{(ns^2)(\sum X_i^2)}{nM_{XX}}.$$

El resultado final es

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i, \quad R^2 = \dots$$

(s_â) (s_{β̂})

C-2. Modelo de regresión lineal múltiple

A continuación se presenta un esquema de cálculo para el modelo

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + e_i.$$

Sin embargo, el procedimiento puede modificarse fácilmente para su aplicación a los modelos con menor cantidad de variables explicativas. El diseño de cálculo pretende lograr un grado de precisión razonable, compatible con la eficiencia en los cálculos.

Cálculo de los momentos aumentados

Sean

$$M_{rr} = n \sum_i X_{ir} X_{is} - (\sum_i X_{ir})(\sum_i X_{is}),$$

$$M_{rY} = n \sum_i X_{ir} Y_i - (\sum_i X_{ir})(\sum_i Y_i),$$

$$M_{YY} = n \sum_i Y_i^2 - (\sum_i Y_i)^2,$$

$$M_{SS} = n \sum_i S_i^2 - (\sum_i S_i)^2,$$

$$r, s = 2, 3, 4, \text{ y } 5;$$

$$S_i = X_{i2} + X_{i3} + X_{i4} + X_{i5} + Y_i.$$

Estos momentos pueden presentarse en la forma

	X_2	X_3	X_4	X_5	Y
X_2	M_{22}	M_{23}	M_{24}	M_{25}	M_{2Y}
X_3		M_{33}	M_{34}	M_{35}	M_{3Y}
X_4			M_{44}	M_{45}	M_{4Y}
X_5				M_{55}	M_{5Y}
Y					M_{YY}

Si todas las operaciones son correctas, debe cumplirse que

$$M_{SS} = M_{22} + M_{33} + M_{44} + M_{55} + M_{YY}$$

$$+ 2(M_{23} + M_{24} + M_{25} + M_{2Y} + M_{34} + M_{35} + M_{3Y} + M_{45} + M_{4Y} + M_{5Y}).$$

Cambio de escala en las variables

Si todos los momentos son aproximadamente de la misma magnitud, conseguiremos una mayor precisión en los resultados. Por tanto, variaremos las unidades en las que se miden las variables, multiplicando cada una de ellas por 10 elevado a la potencia necesaria para que todos los valores de los momentos diagonales estén entre 0,1 y 10,0. Naturalmente, esto producirá también un ajuste en los momentos que no pertenecen a la diagonal principal. Para muchos fines prácticos es suficiente enumerar los seis primeros decimales de los momentos en las nuevas unidades. Las variables cambiadas de escala y sus momentos se representarán utilizando asteriscos.

EJEMPLO. Supongamos

$$\begin{aligned} M_{22} &= 6560.3575, & M_{YX} &= 0.0604, \\ M_{33} &= 320.12, & M_{2Y} &= 76.2344, \\ M_{23} &= 21.066, & M_{3Y} &= 1.177. \end{aligned}$$

En este caso, efectuaremos las siguientes transformaciones:

$$X_{12}^* = 10^{-2} X_{12},$$

$$X_{13}^* = 10^{-1} X_{13},$$

$$Y_1^* = 10^1 Y_1.$$

En consecuencia,

$$\begin{aligned} M_{22}^* &= 0.656036 & M_{YX}^* &= 6.040000 \\ M_{33}^* &= 3.201200 & M_{2Y}^* &= 7.623440 \\ M_{23}^* &= 0.021066 & M_{3Y}^* &= 1.177000 \end{aligned}$$

Solución básica para las ecuaciones normales minimocuadráticas

Para resolver las ecuaciones normales minimocuadráticas utilizamos el método de Gauss-Doolittle, que constituye un procedimiento sistemático para resolver ecuaciones lineales simétricas³. En primer lugar, presentamos los momentos aumentados cambiados de escala en la forma siguiente:

	Columna									
Fila	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) R_1 :	M_{22}^*	M_{23}^*	M_{24}^*	M_{25}^*	M_{2Y}^*	1	0	0	0	S_1
(2) R_2 :	M_{32}^*	M_{33}^*	M_{34}^*	M_{35}^*	M_{3Y}^*	0	1	0	0	S_2
(3) R_3 :	M_{42}^*	M_{43}^*	M_{44}^*	M_{45}^*	M_{4Y}^*	0	0	1	0	S_3
(4) R_4 :	M_{52}^*	M_{53}^*	M_{54}^*	M_{55}^*	M_{5Y}^*	0	0	0	1	S_4

La columna (10) es una columna de comprobación y sus elementos son los totales de las filas. Las filas que deben seguir a estas cuatro se obtienen de la forma siguiente:

$$(5) A_{11} = R_{11} \text{ (es decir, se reproduce la fila (1))}.$$

$$(6) B_{11} = A_{11}/A_{11} \text{ (es decir, se divide la fila (4) por su primer término)}.$$

$$(7) A_{21} = R_{21} - A_{12}B_{11}.$$

$$(8) B_{21} = A_{21}/A_{22}.$$

$$(9) A_{31} = R_{31} - A_{13}B_{11} - A_{23}B_{21}.$$

$$(10) B_{31} = A_{31}/A_{33}.$$

$$(11) A_{41} = R_{41} - A_{14}B_{11} - A_{24}B_{21} - A_{34}B_{31}.$$

$$(12) B_{41} = A_{41}/A_{44}.$$

El cuadro incluye ahora las filas siguientes:

Fila	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(5) A_1 :	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	1	0	0	0	S_4
(6) B_1 :	1	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	0	0	0	S_5
(7) A_2 :	0	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	1	0	0	S_6
(8) B_2 :	0	1	B_{23}	B_{24}	B_{25}	B_{26}	B_{27}	0	0	S_7
(9) A_3 :	0	0	A_{33}	A_{34}	A_{35}	A_{36}	A_{37}	1	0	S_8
(10) B_3 :	0	0	1	B_{34}	B_{35}	B_{36}	B_{37}	B_{38}	0	S_9
(11) A_4 :	0	0	0	A_{44}	A_{45}	A_{46}	A_{47}	A_{48}	1	S_{10}
(12) B_4 :	0	0	0	1	B_{45}	B_{46}	B_{47}	B_{48}	B_{49}	S_{11}

A continuación, las reglas son las siguientes

$$(13) C_{41} = B_{41}.$$

$$(14) C_{31} = B_{31} - B_{34}C_{41}.$$

$$(15) C_{21} = B_{21} - B_{23}C_{31} - B_{24}C_{41}.$$

$$(16) C_{11} = B_{11} - B_{12}C_{21} - B_{13}C_{31} - B_{14}C_{41}.$$

Estas filas se añaden en orden inverso:

Fila	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(16) C_1 :	1	0	0	0	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	S_{15}
(15) C_2 :	0	1	0	0	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	S_{14}
(14) C_3 :	0	0	1	0	C_{35}	C_{36}	C_{37}	C_{38}	C_{39}	S_{13}
(13) C_4 :	0	0	0	1	C_{45}	C_{46}	C_{47}	C_{48}	C_{49}	S_{12}

Para completar el cuadro no es necesario calcular los valores que deben ser iguales a 0 ó 1. Compruébese cada fila a medida que se obtiene por medio de la columna de comprobación. Es decir, la cifra de la columna (10) de cada fila debe ser igual a la suma de todos los demás valores que aparecen en esta fila y, de no ser así, se ha cometido algún error en las operaciones. Naturalmente, se pueden aceptar los errores de redondeo (en el quinto o sexto decimal). Si el modelo sólo tiene tres variables explicativas en lugar de cuatro, las columnas (4) y (9), y las filas (4), (11), (12), y (13) no son aplicables, y las filas (14), (15), y (16) deben reducirse en la forma adecuada.

³ Véase P. S. Dwyer, *Linear Computations* (Nueva York: Wiley, 1951).

Resultados

Las estimaciones minimocuadráticas de los coeficientes de regresión aparecen en las últimas cuatro filas del cuadro. Concretamente,

$$\beta_2^* = C_{15},$$

$$\beta_3^* = C_{25},$$

$$\beta_4^* = C_{35},$$

$$\beta_5^* = C_{45},$$

$$y \quad \beta_1^* = Y - \beta_2^* X_2^* - \beta_3^* X_3^* - \beta_4^* X_4^* - \beta_5^* X_5^*.$$

Además,

$$nSCR^* = \sum_{r=2}^5 \beta_r^* M_{ry}^*,$$

$$nSTC = M_{yy}^*,$$

$$nSCE = nSTC^* - nSCR^*,$$

y, por tanto,

$$R^2 = \frac{nSCR^*}{nSTC^*}$$

$$y \quad ns^{*2} = \frac{nSCE^*}{n-5}.$$

Las variancias estimadas de los coeficientes de regresión se obtienen de la forma siguiente:

$$s_{\beta_2^*}^{*2} = (ns^{*2}) C_{16},$$

$$s_{\beta_3^*}^{*2} = (ns^{*2}) C_{27},$$

$$s_{\beta_4^*}^{*2} = (ns^{*2}) C_{38},$$

$$s_{\beta_5^*}^{*2} = (ns^{*2}) C_{49}.$$

La variancia estimada de β_1^* puede calcularse a través de la fórmula (10.32) del texto. El estadístico F para la contrastación de la hipótesis de que no existe relación entre la variable dependiente y las explicativas puede calcularse a través de

$$\frac{nSCR^*/4}{ns^{*2}} \sim F_{4, n-5}.$$

Para volver a las variables originales, sin asteriscos, sustituimos Y_i^* y las X_i^* por sus valores en

$$Y_i^* = \beta_1^* + \beta_2^* X_{i2}^* + \beta_3^* X_{i3}^* + \beta_4^* X_{i4}^* + \beta_5^* X_{i5}^* + e_i^*$$

para obtener

$$(10^{a_y}) Y_i = \beta_1^* + \beta_2^* (10^{a_2}) X_{i2} + \beta_3^* (10^{a_3}) X_{i3} + \beta_4^* (10^{a_4}) X_{i4} + \beta_5^* (10^{a_5}) X_{i5} + e_i^*,$$

en la que $a_y, a_2, \dots, a_5$ representan los exponentes utilizados en la transformación de las variables originales. Así, pues,

$$\beta_1 = (10^{-a_y}) \beta_1^*,$$

$$\beta_2 = (10^{a_2 - a_y}) \beta_2^*,$$

$$\beta_3 = (10^{a_3 - a_y}) \beta_3^*,$$

$$\beta_4 = (10^{a_4 - a_y}) \beta_4^*,$$

$$\beta_5 = (10^{a_5 - a_y}) \beta_5^*.$$

Obsérvese que los errores standard estimados de los coeficientes de regresión se multiplican por el mismo factor que los propios coeficientes, y que R^2 y el estadístico F permanecen inalterados.

LECTURAS ADICIONALES SUGERIDAS

Lecturas adicionales sugeridas

Las obras se presentan generalmente por orden de importancia y de relación con el material presentado en el texto de cada capítulo.

Capítulo 1. Introducción a la inferencia estadística

- W. A. WALLIS y H. V. ROBERTS, *The Nature of Statistics*. Nueva York, The Free Press, 1956.
 G. U. YULE y M. G. KENDALL, *An Introduction to the Theory of Statistics*. Londres, Charles Griffin, 1950, cap. 16.
 IAN HOCKING, *Logic of Statistical Inference*. Nueva York, Cambridge University Press, 1965.
 H. JEFFREYS, *Theory of Probability*. Nueva York, Oxford University Press, 1961, cap. 1.
 C. L. LASTRUCCI, *The Scientific Approach*. Cambridge, Mass., Schenkman, 1963.

Capítulo 2. Obtención experimental de distribuciones muestrales

- W. J. DIXON y F. J. MASSEY, *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill, 1957, cap. 4.
 J. M. HAMMERSLEY y D. C. HANDSCOMB, *Monte Carlo Method*. Nueva York, Wiley, 1964.

Capítulo 3. Probabilidad y distribuciones de probabilidad

- S. H. HYMAN, *Probability Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1967, caps. 1-4.
 F. MOSTELLER, R. E. K. ROURKE y G. B. THOMAS, *Probability with Statistical Applications*. Reading, Mass., Addison-Wesley, 1961, caps. 1-5.
 J. E. FREUND, *Mathematical Statistics*. Englewood Cliffs, N. J., Prentice-Hall, 1962, caps. 1-6.

- A. M. MOOD y F. A. GRAYBILL, *Introduction to the Theory of Statistics*. 2.ª ed. Nueva York, McGraw-Hill, 1963, caps. 1-5.
 E. PARZEN, *Modern Probability Theory and Its Applications*. Nueva York, Wiley, 1960.

Capítulo 4. Obtención teórica de distribuciones muestrales

- S. S. WILKS, *Elementary Statistical Analysis*. Princeton, N. J., Princeton University Press, 1958, cap. 9.
 S. H. HYMAN, *Probability Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1967, caps. 5 y 6.
 W. A. WALLIS y H. V. ROBERTS, *Statistics: A New Approach*. Londres, Methuen, 1957, cap. 11.
 C. R. RAO, *Linear Statistical Inference and Its Applications*. Nueva York, Wiley, 1965, cap. 3.
 M. G. KENDALL y A. STUART, *The Advanced Theory of Statistics*. Vol. I, 2.ª ed., Nueva York, Macfay, 1963, caps. 11-13.

Capítulo 5. Contrastación de hipótesis

- W. A. WALLIS y H. V. ROBERTS, *Statistics: A New Approach*. Londres, Methuen, 1957, caps. 12 y 13.
 J. E. FREUND, *Mathematical Statistics*. Englewood Cliffs, N. J., Prentice-Hall, 1962, caps. 11 y 12.
 A. M. MOOD y F. A. GRAYBILL, *Introduction to the Theory of Statistics*. 2.ª ed. Nueva York, McGraw-Hill, 1963, cap. 12.
 R. L. ANDERSON y T. A. BANCROFT, *Statistical Theory in Research*. Nueva York, McGraw-Hill, 1952, caps. 11 y 12.
 E. L. LEHMANN, *Testing Statistical Hypotheses*. Nueva York, Wiley, 1959.

Capítulo 6. Estimación

- S. H. HYMAN, *Probability Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1967, cap. 7.
 B. W. LINDGREN, *Statistical Theory*. Nueva York, Macmillan, 1968, cap. 5.
 A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 3.
 P. J. DHRIMES, *Econometrics*. Nueva York, Harper & Row, 1970, cap. 3.
 R. DEUTSCH, *Estimation Theory*. Englewood Cliffs, N. J., Prentice-Hall, 1965.

Capítulo 7. Regresión simple

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 1.
 Trad. al castellano: *Métodos de Econometría*. Barcelona, Ed. Vicens-Vives, 1972.

- A. S. GOLDBERGER, *Topics in Regression Analysis*. Nueva York, Macmillan, 1968, caps. 1 y 2.
- R. L. ANDERSON y T. A. BANCROFT, *Statistical Theory in Research*. Nueva York, McGraw-Hill, 1952, cap. 13.
- A. M. MOOD y F. A. GRAYBILL, *Introduction to the Theory of Statistics*. 2.^a ed. Nueva York, McGraw-Hill, 1963, cap. 13.
- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 3.

Capítulo 8. Violaciones de los supuestos básicos

8.1 Heteroscedasticidad

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 8, sec. 2.
- E. J. KANE, *Economic Statistics and Econometrics*. Nueva York, Harper & Row, 1968, cap. 14.
- H. C. RUTEMILLER y D. A. BOWERS, «Estimation in a Heteroskedastic Regression Model,» *Journal of the American Statistical Association*, vol. 63, junio 1968, pp. 552-57.
- S. M. GOLDFELD y R. F. QUANDT, «Some Tests for Homoscedasticity,» *Journal of the American Statistical Association*, vol. 60, septiembre 1965, pp. 539-47.

8.2 Perturbaciones autorregresivas

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 7.
- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 13.
- C. F. CHRIST, *Econometric Models and Methods*. Nueva York, Wiley, 1966, cap. 9, sec. 13, y cap. 10, sec. 4.

8.3 Variable explicativa estocástica

- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 6.

Capítulo 9. Estimación con datos deficientes

9.1 Errores de medición

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 6.
- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 10.
- M. G. KENDALL y A. STUART, *The Advanced Theory of Statistics*. Vol. II Londres, Charles Griffin, 1961, cap. 29.

- M. HALPERIN, «Fitting of Straight Lines and Prediction When Both Variables Are Subject to Error,» *Journal of the American Statistical Association*, vol. 56, septiembre 1961, pp. 657-69.

9.2 Estimación a partir de datos agrupados

- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 8, sec. 2.
- J. S. CRAMER, «Efficient Grouping, Regression and Correlation in Engel Curve Analysis,» *Journal of the American Statistical Association*, vol. 59, marzo 1964, pp. 233-50.

9.3 Estimación faltando algunas observaciones

- A. A. AFIFI y R. M. ELASHOFF, «Missing Observations in Multivariate Statistics II. Point Estimation in Simple Linear Regression,» *Journal of the American Statistical Association*, vol. 62, marzo 1967, pp. 10-29.
- A. A. AFIFI y R. M. ELASHOFF, «Missing Observations in Multivariate Statistics III. Large Sample Analysis of Simple Linear Regression,» *Journal of the American Statistical Association*, vol. 64, marzo 1969, pp. 337-58.
- A. A. AFIFI y R. M. ELASHOFF, «Missing Observations in Multivariate Statistics IV. A Note on Simple Linear Regression,» *Journal of the American Statistical Association*, vol. 64, marzo 1969, pp. 359-65.

Capítulo 10. Regresión múltiple

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, caps. 2 y 4.
- K. A. FOX, *Intermediate Economic Statistics*. Nueva York, Wiley, 1968, caps. 7 y 13.
- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 4.
- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, caps. 6 y 8, sec. 5.
- J. B. RAMSEY, «Tests for Specification Errors in Classical Linear Least-squares Regression Analysis,» *Journal of the Royal Statistical Society, serie B*, vol. 31, 1969, pp. 350-71.

Capítulo 11. Formulación y estimación de modelos especiales

11.1 Modelos con variables binarias

- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 8, sec. 4.

- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 5, secs. 2 y 3.
- H. SCHEFFE, *The Analysis of Variance*. Nueva York, Wiley, 1959.

11.2 Modelos con restricciones en los coeficientes

- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 5, sec. 6.
- L. G. JUDGE y T. TAKAYAMA, «Inequality Restrictions in Regression Analysis», *Journal of the American Statistical Association*, vol. 61, marzo 1966, pp. 166-81.
- H. THEIL y A. S. GOLDBERGER, «On Pure and Mixed Statistical Estimation in Econometrics», *International Econometric Review*, vol. 2, enero 1961, pp. 65-78.

11.3 Modelos no lineales

- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 9.
- G. E. P. BOX y D. R. COX, «An Analysis of Transformations», *Journal of the Royal Statistical Society, series B*, vol. 26, 1964, pp. 211-43.

11.4 Modelos con retardos distribuidos

- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, cap. 15.
- Z. GRILICHES, «Distributed Lags: A Survey», *Econometrica*, vol. 35, enero 1967, pp. 16-49.
- A. ZELLNER y M. S. GEISEL, «Analysis of Distributed Lag Models with Applications to Consumption Function Estimation», trabajo presentado a la Reunión Europea de la Econometric Society, Amsterdam, septiembre 1968 (próximamente en *Econometrica*).

Capítulo 12. El modelo generalizado de regresión lineal y sus aplicaciones

12.1 El modelo generalizado de regresión lineal

- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 5, sec. 4.
- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, cap. 7, sec. 3.

12.2 Consideración conjunta de datos de series temporales y de sección transversal

- T. D. WALLACE y A. HUSSAIN, «The Use of Error Components Models in Combining Cross-Section with Time-Series Data», *Econometrica*, vol. 37, enero 1969, pp. 55-72.
- P. BALESTRA y M. NERLOVE, «Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas», *Econometrica*, vol. 34, julio 1966, pp. 585-612.

12.3 Regresiones aparentemente no relacionadas

- A. ZELLNER, «An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias», *Journal of the American Statistical Association*, vol. 57, junio 1962, pp. 348-68.
- L. G. TELSER, «Iterative Estimation of a Set of Linear Regression Equations», *Journal of the American Statistical Association*, vol. 59, septiembre 1964, pp. 842-62.
- J. KMENTA y R. F. GILBERT, «Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions», *Journal of the American Statistical Association*, vol. 63, diciembre 1968, pp. 1180-200.
- R. W. PARKS, «Efficient Estimation of a System of Regression Equations When Distributions Are Both Serially and Contemporaneously Correlated», *Journal of the American Statistical Association*, vol. 62, junio 1967, pp. 500-509.

Capítulo 13. Sistemas de ecuaciones simultáneas

- E. MALINVAUD, *Statistical Methods of Econometrics*. Chicago, Rand McNally, 1966, caps. 16-18.
- C. F. CHRIST, *Econometric Models and Methods*. Nueva York, Wiley, 1966, caps. 8 y 9.
- F. M. FISHER, *The Identification Problem in Econometrics*. Nueva York, McGraw-Hill, 1966.
- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 7.
- J. JOHNSTON, *Econometric Methods*. Nueva York, McGraw-Hill, 1963, caps. 9 y 10.
- P. R. FISK, *Stochastically Dependent Equations*. Nueva York, Hafner, 1967.
- P. J. DHRYMES, *Econometrics*. Nueva York, Harper & Row, 1970, caps. 4 y 6-8.

13.6 Análisis de los modelos econométricos dinámicos

- A. S. GOLDBERGER, *Impact Multipliers and Dynamic Properties of the Klein-*

Goldberger Model. Amsterdam, North-Holland, 1959.

- I. ADELMAN y F. L. ADELMAN, «The Dynamic Properties of the Klein-Goldberger Model.» *Econometrica*, vol. 27, octubre 1959, pp. 597-625.
- L. R. KLEIN, M. K. EVANS y M. HARTLEY, *Econometric Gaming*. Nueva York, Macmillan, 1969.
- M. K. EVANS y L. R. KLEIN, *The Wharton Econometric Forecasting Model*. Filadelfia, University of Pennsylvania Press, 1967.
- G. FROMM y P. TAUBMAN, *Policy Simulations with an Econometric Model*. Washington, The Brookings Institution, 1968.
- H. THEIL y J. C. G. BOOT, «The Final Form of Econometric Equation Systems.» *Review of the International Statistical Institute*, vol. 30, 1962, pp. 136-52.

Apéndices

B. Elementos de álgebra matricial

- J. JOHNSTON, *Métodos de Econometría*. Barcelona, Vicens-Vives, cap. 3.
- A. S. GOLDBERGER, *Econometric Theory*. Nueva York, Wiley, 1964, cap. 2.
- G. Hadley, *Linear Algebra*. Reading, Mass., Addison-Wesley, 1961.

C. Diseño del método de cálculo para la estimación minimocuadrática

- R. L. ANDERSON y R. A. BANCROFT, *Statistical Theory in Research*. Nueva York, McGraw-Hill, 1952, cap. 15.
- L. R. KLEIN, *Textbook of Econometrics*. Evanston, Ill., Row, Peterson, 1953, cap. 4.
- P. S. DWYER, *Linear Computations*. Nueva York, Wiley, 1951.

MÉTODOS MATEMÁTICOS DEL ANÁLISIS ECONÓMICO MODERNO. *Hukukane Nikaido*

ECONOMÍA PÚBLICA. *L. Johansen*

PROGRAMACIÓN MATEMÁTICA Y CÁLCULO ECONÓMICO. *Josep M.ª Vergara*

ANÁLISIS DE KEYNES Y DE LA ECONOMÍA KEYNESIANA. *Axel Leijonhufvud*

ELEMENTOS DE ECONOMETRÍA. *Jan Kmenta*

MÉTODOS ECONÓMETRICOS. *J. Johnston*

HISTORIA ECONÓMICA (de los orígenes a 1789). *J. Imbert*

HISTORIA ECONÓMICA MUNDIAL (moderna y contemporánea). *J. A. Trousard y C. Gerard*

HISTORIA ECONÓMICA DE ESPAÑA. *J. Vicens Vives*

INTRODUCCIÓN A LA ECONOMÍA POSITIVA. *R. G. Lipsey*

PRACTICAS DE ECONOMÍA POSITIVA. *J. A. Stillwell y R. G. Lipsey*

ECONOMÍA REGIONAL. *H. W. Richardson*

INTRODUCCIÓN A LA MICROECONOMÍA CONTEMPORÁNEA. *V. Ch. Walsh*

IMPUESTOS E INFLACIÓN. *A. Morag*

GESTIÓN DE LAS EMPRESAS AGRARIAS Y

DESARROLLO RURAL. *C. Couffin*

DOSCIENTOS PROBLEMAS DE ESTADÍSTICA DESCRIPTIVA. *E. Casa Aruta*

PERSPECTIVA DEMOGRÁFICA. *L. Henry*

EL HAMBRE, LA GUERRA Y EL CONTROL DE LA NATALIDAD (ensayo sobre el malthusianismo y las teorías marxistas). *A. Sauvy*

ESTUDIOS DE ECONOMÍA ANTIGUA DE LA PENINSULA IBERICA. *M. Tarradell*

TEORÍAS MODERNAS DEL CRECIMIENTO ECONÓMICO. *Henry Y. Wan Jr.*

PRINCIPIOS DE LA TEORÍA DE LA DECISIÓN. *Dennis V. Lindley*

LA ECONOMÍA DE LA POBREZA. *Th. S. Balogh*

TIPO DE CAMBIO Y RELACIONES DE INTERCAMBIO. *S. A. Orze*

ECONOMÍA POSITIVA Y OBJETIVOS DE LA POLÍTICA ECONÓMICA. *T. W. Hutchison*

DESARROLLO Y REFORMAS DE LOS PAÍSES SOCIALISTAS. *J. Wilczynski*