

Modelo de regresión lineal múltiple:

especificación, estimación y contraste

Tomás del Barrio Castro
Miquel Clar López
Jordi Suriñach Caralt

PID_00160616



Universitat Oberta
de Catalunya

www.uoc.edu

Índice

Introducción	5
Objetivos	7
1. Introducción a la econometría	9
1.1. ¿Qué es la econometría?	9
1.2. Variables, relaciones y parámetros.....	10
1.3. La modelización econométrica. Fases de la investigación econométrica.....	12
1.4. Objetivos de la modelización econométrica	13
2. El modelo de regresión lineal múltiple estándar	15
2.1. Especificación.....	15
2.2. Hipótesis básicas del modelo de regresión lineal múltiple estándar	19
2.2.1. Hipótesis generales del modelo	19
2.2.2. Hipótesis sobre el término de perturbación.....	20
2.2.3. Hipótesis sobre las variables explicativas del modelo.....	23
2.2.4. Hipótesis sobre los parámetros del modelo.....	23
2.3. Estimación para mínimos cuadrados ordinarios (MCO)	23
2.3.1. Descripción del método de estimación.....	24
2.3.2. Propiedades de los estimadores MCO de los β_j	28
2.4. Análisis de los residuos y estimación de σ_u^2	32
2.4.1. Propiedades de los residuos	33
2.4.2. Estimación de la varianza del término de perturbación.....	35
2.5. Estimación por máxima verosimilitud	38
2.6. Medidas de la bondad del ajuste	40
2.7. Significación de los parámetros del modelo	43
2.7.1 Significación económica.....	44
2.7.2 Significación estadística.....	45
2.8. El modelo de regresión lineal múltiple en desviaciones respecto a la media	49
2.9. Predicción.....	50
2.9.1. Predicción puntual.....	50
2.9.2. Predicción por intervalo	51
3. El modelo de regresión con restricciones lineales	53
3.1. Contrastación de restricciones lineales	53
3.1.1. Formulación matricial de las restricciones lineales.....	54
3.1.2. Metodología para contrastar restricciones lineales: estadístico de prueba	55
3.1.3. Un método alternativo para contrastar restricciones lineales	58

3.2. Estimación restringida por mínimos cuadrados (MCR).....	63
3.2.1. El estimador de mínimos cuadrados restringidos	63
3.2.2. Propiedades del estimador restringido	64
3.2.3. Un ejemplo aclaratorio	65
3.3. Análisis de la permanencia estructural. Contraste de Chow	70
3.3.1. Introducción	70
3.2.2. Contraste de Chow de permanencia estructural.....	71
3.3.3. Un caso particular: tamaño insuficiente de una de las submuestras	73
3.3.4. Limitaciones del contraste de Chow	73
Glosario	77
Bibliografía	79

Introducción

Este módulo didáctico está formado por los tres apartados siguientes:

1) El primero es una introducción que nos permitirá ponernos en contacto con los aspectos relacionados con la econometría. En concreto, veremos cuáles son los **fundamentos de la econometría**, sus objetivos, qué tipo de problemas nos permite solucionar, y pondremos de manifiesto las relaciones que tiene con otros ámbitos de la economía.

2) En el segundo apartado introduciremos el **modelo de regresión múltiple (MRLM)**, que será la base de todos los aspectos que trataremos a lo largo de esta asignatura y de su continuación. En concreto, veremos los contenidos siguientes:

- a) La formulación del MRLM.
- b) Las hipótesis básicas relativas al comportamiento de las diferentes partes que lo integran: las variables (endógena y explicativas), los parámetros y el término de perturbación.
- c) La estimación de los parámetros desconocidos del modelo por los métodos de mínimos cuadrados ordinarios (MCO) y máxima verosimilitud (MV).
- d) Las propiedades de los estimadores cuando se cumplen las hipótesis básicas.
- e) Las medidas que nos permitirán cuantificar la bondad del ajuste y evaluar el modelo.
- f) La manera de obtener predicciones una vez que el modelo ya se ha formulado, estimado y validado.

3) El tercer y último apartado de este módulo lo dedicaremos a estudiar los aspectos relacionados con la **contrastación de restricciones lineales**. Es en este apartado, pues, donde presentaremos las herramientas que necesitaremos para contrastar hipótesis que se puedan formular sobre el comportamiento de los parámetros. En concreto, veremos los puntos siguientes:

- a) Los tipos de restricciones lineales que podremos contrastar.
- b) La manera de formularlos matricialmente.
- c) Los estadísticos de prueba adecuados para comprobar si las hipótesis formuladas sobre los parámetros de la población en forma de restricciones lineales se pueden considerar ciertas en el ámbito de la población o no.

- d) Los contrastes de significación individual y global de parámetros que se estudian en el segundo apartado, que no son más que casos particulares de los contrastes de restricciones lineales.
- e) Los estimadores restringidos (aquellos que hacemos que cumplan las restricciones lineales planteadas) y sus propiedades.
- f) La manera de contrastar la permanencia estructural del modelo.

Objetivos

Una vez trabajados los contenidos de este módulo didáctico, los estudiantes tenéis que ser capaces de:

1. Conocer las hipótesis básicas que debe cumplir el modelo de regresión múltiple que denominaremos *modelo estándar*.
2. Obtener los estimadores de mínimos cuadrados ordinarios y de máxima verosimilitud de los parámetros desconocidos del modelo de regresión múltiple, y conocer las propiedades que tienen cuando se cumplen las hipótesis básicas.
3. Cuantificar la bondad del ajuste del modelo.
4. Determinar cuál de las variables exógenas contribuye más a explicar el comportamiento de la variable endógena, y contrastar la significación individual de un parámetro y la global del modelo.
5. Obtener la predicción puntual y por intervalo de la variable endógena.
6. Expresar restricciones lineales matricialmente.
7. Poder contrastar cualquier restricción lineal homogénea de igualdad mediante métodos distintos.
8. Obtener los estimadores restringidos y conocer sus propiedades tanto si partimos de la hipótesis de que las restricciones lineales planteadas son ciertas como si partimos de que no lo son.
9. Saber cómo contrastar la permanencia estructural de los parámetros del modelo.

1. Introducción a la econometría

El objetivo de este apartado es presentar las bases sobre las cuales se asienta la metodología econométrica. Por ello, en primer lugar, después de una pequeña introducción sobre cuestiones como, por ejemplo, el origen de la econometría y qué es la econometría, presentamos toda una serie de conceptos básicos y, cuando nos hayamos familiarizado con ellos, abordaremos el cuerpo central del módulo: la modelización econométrica. Con esta finalidad, explicamos las diferentes etapas que hay que seguir en todo estudio econométrico. Para acabar, presentamos los objetivos que se pueden alcanzar con un estudio de este tipo.

1.1. ¿Qué es la econometría?

El nacimiento de la econometría, del mismo modo que otras disciplinas del ámbito de la ciencia económica, se produce ante la necesidad de resolver toda una serie de problemas con la información económica existente. Esta aparición se basa en el desarrollo de determinadas técnicas que facilitan el análisis cuantificado de las relaciones económicas.

El objetivo que persigue la econometría es encontrar y cuantificar las relaciones económicas empleando técnicas basadas en los métodos inferenciales de la estadística.

En concreto, en el nacimiento de la econometría confluyeron muchos factores, entre los cuales destaca el planteamiento distinto que existía, durante las décadas de los años veinte y treinta, del estudio de los ciclos económicos. En estos años apareció un grupo de economistas preocupados por la colaboración entre matemáticos, estadísticos y economistas. Creían que la introducción de los métodos matemáticos en la investigación, en las ciencias sociales en general y en la economía en particular, permitiría avanzar en su desarrollo. Además, criticaban la no-consideración de la teoría económica en los modelos explicativos de los ciclos económicos. Las aportaciones de estos autores constituyeron el antecedente del análisis económico propuesto por la Comisión Cowles y por T. Haavelmo (1944), basado en el enfoque probabilístico que hay en las interrelaciones económicas.

Algunos de los economistas...

... preocupados por la colaboración entre matemáticos, estadísticos y economistas eran personajes como R. Frisch, J. Tinbergen, K. Pearson y J. Slutskij, entre otros.

En la literatura podemos encontrar definiciones distintas de la econometría. Entre ellas, podemos destacar las dos siguientes:

¿Cómo se puede definir la econometría?

1) "Análisis cuantitativo de los fenómenos económicos reales, basado en el desarrollo simultáneo de la teoría y la observación que se relacionan mediante los métodos de inferencia adecuados."

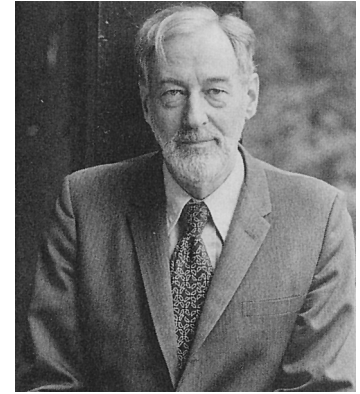
P.A. Samuelson; T.C. Koopmans; M.H. Stone (1954).

2) “Rama de la economía relacionada con la estimación empírica de las relaciones económicas. Emplea la teoría económica, incorporada en un modelo econométrico, hechos, resumidos en datos, y estadística teórica, adecuada por las técnicas econométricas para medir y contrastar empíricamente determinadas relaciones entre variables económicas, dando contenido empírico al razonamiento económico.”

M.D. Intriligator (1978).

Atendiendo a todo lo que hemos dicho con anterioridad, podemos señalar que los rasgos más relevantes que caracterizan la econometría son los siguientes:

- a) La econometría es la rama de la ciencia económica que se ocupa del análisis cuantitativo de los fenómenos económicos.
- b) La econometría está relacionada con otras disciplinas como, por ejemplo, la teoría económica, la estadística y las matemáticas.
- c) La econometría se basa en un enfoque probabilístico de la realidad.



T.C. Koopmans, autor del manual *Econometric Theory* (1992)

1.2. Variables, relaciones y parámetros

En todo modelo se distinguen dos tipos de variables: la variable endógena (también llamada *variable dependiente* o *variable que se debe explicar*) y las variables explicativas (o *variables independientes* o *variables exógenas*).

La **variable endógena** es aquella cuyo comportamiento queremos conocer y explicar. Las **variables explicativas** son aquellas que, de acuerdo con los postulados de la teoría económica, permiten explicar el comportamiento de la variable endógena. Dependiendo del número de variables explicativas que se introducen, una o más, el modelo se denominará **modelo simple** o **modelo múltiple**.

Así, por ejemplo, siguiendo el modelo keynesiano consumo-renta, el consumo de los individuos depende de (se explica por) la renta de cada uno de ellos. Por tanto, en este modelo, la variable endógena (aquello que queremos conocer y explicar) es el consumo de los individuos, y la variable explicativa (aquello que permite explicarlo) es la renta de los individuos. Se trata, pues, de un modelo simple. Si partimos de la hipótesis de que, además de la renta de los individuos, el número de hijos también permite explicar las pautas de comportamiento del consumo, tendremos dos variables explicativas y, por tanto, estaremos ante un modelo múltiple.

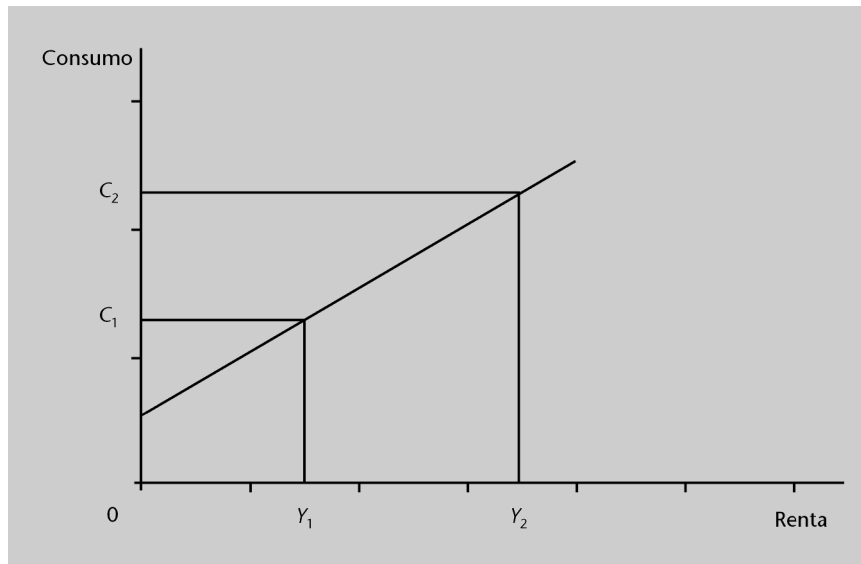
Consultad el modelo keynesiano consumo-renta en la asignatura *Macroeconomía*.



Ejemplo ilustrativo de un modelo simple y de un modelo múltiple

Con el fin de ilustrar un modelo simple utilizaremos la función keynesiana de consumo. Ésta estipula que el consumo C_i de las unidades domésticas depende de su renta Y_i . Supongamos que en nuestro modelo simple la dependencia es lineal:

$$C_i = \alpha + \beta Y_i.$$



La función anterior, tal y como se puede ver, es determinista. Nos dice, por ejemplo, que para un nivel de renta Y_1 el consumo será C_1 y que para un nivel de renta Y_2 el consumo será C_2 .

Si tenemos en cuenta que el número de hijos N_i también permite explicar el consumo, entonces tendremos un modelo múltiple, que puede ser el siguiente:

$$C_i = \alpha + \beta Y_i + \gamma N_i.$$

Entre la variable endógena y las variables explicativas* existe, de acuerdo con lo que hemos comentado en el párrafo anterior, una relación de causalidad que se caracteriza por el hecho de ser unidireccional: los comportamientos de las variables explicativas causan (determinan, explican) el de la variable endógena. Precisamente la existencia de esta relación de causalidad es la que permite formular un modelo. No obstante, esta relación que se establece entre las variables del modelo puede ser de muchos tipos: lineal, cuadrática, exponencial, logarítmica, etc. En consecuencia, en el momento de especificar el modelo hay que determinar (también de acuerdo con los postulados de la teoría económica) la forma funcional que adopta la relación entre la variable endógena y las explicativas. De todos modos, en el ámbito del modelo que estudiaremos supondremos que la relación es lineal y que, si no lo es, se puede linealizar mediante una transformación adecuada.

Adicionalmente, en todo modelo aparecen lo que denominaremos *parámetros*. Los **parámetros**, que están asociados a cada variable explicativa, cuantifican la relación existente entre la variable endógena y cada una de las variables explicativas. Son, por tanto, lo que se desconoce y se debe estimar.

* De ahora en adelante supondremos que trabajamos con un modelo múltiple.

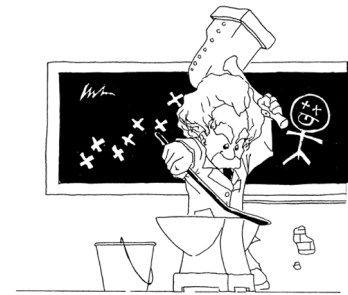
La relación entre...

... la variable endógena y las variables explicativas no es lineal o linealizable en todos los casos. A veces, nos encontraremos ante modelos no lineales, que también se podrían estudiar, pero que quedan fuera de los objetivos de este material didáctico.

1.3. La modelización econométrica.

Fases de la investigación econométrica

La econometría es la rama de la economía que tiene que ver con la estimación empírica, con la cuantificación de las relaciones económicas: a partir de los postulados que establece la teoría económica se especifica un modelo econométrico, el cual, a partir de un conjunto de información estadística (datos), se estima empleando técnicas estadísticas y econométricas con el fin de medir y contrastar empíricamente determinadas relaciones entre variables económicas.

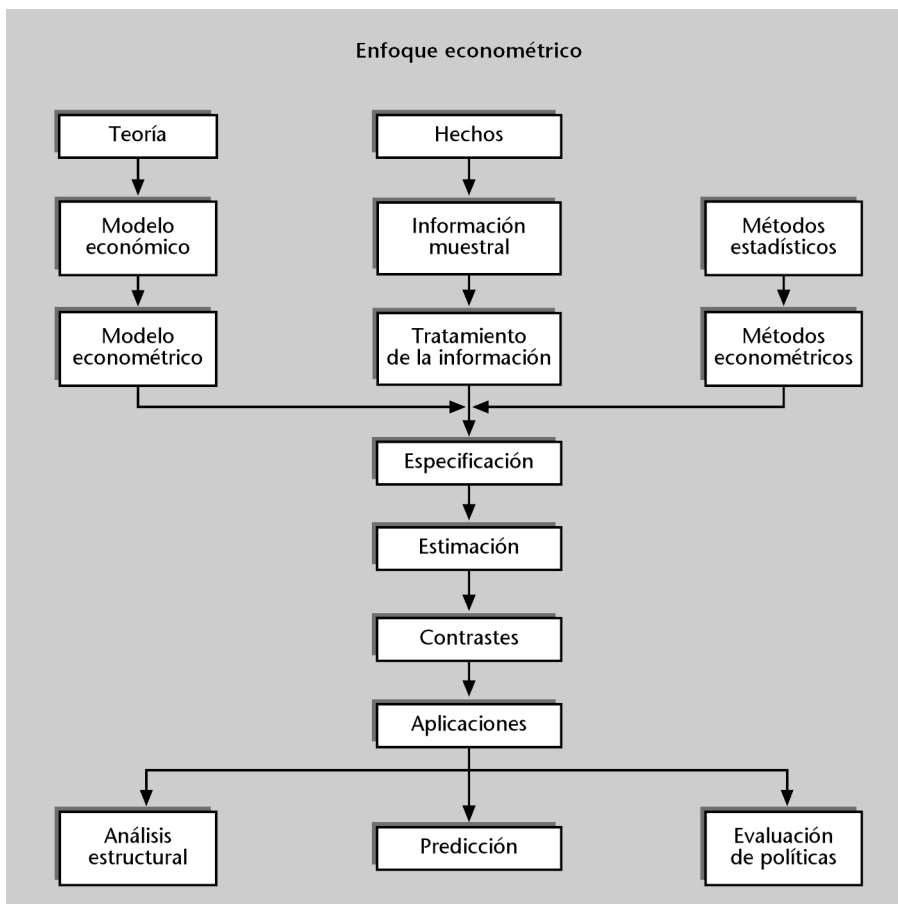


La modelización econométrica presenta tres fases: la especificación, la estimación y el contraste.

Las técnicas econométricas...

... no sólo están limitadas al mundo economicoempresarial. Por el contrario, también son susceptibles de aplicación no sólo a otros campos de las ciencias sociales (como la sociología, la historia, etc.), sino también a otros ámbitos (como la educación, la sanidad, el medio ambiente, etc.).

En el cuadro siguiente presentamos gráficamente en qué consiste el enfoque econométrico. Como podemos observar, de acuerdo con lo que hemos dicho anteriormente, hay dos pilares básicos que constituyen las materias primas en cualquier estudio econométrico: la teoría y los hechos.



Fuente: M.D. Intrilligator (1978, pág. 3)

1) El primer pilar, la **teoría**, permite derivar un modelo (el modelo económico), que sintetiza la incógnita relevante sobre el fenómeno (la variable endógena) objeto de análisis y del cual deriva el modelo econométrico que permite medirlo y contrastarlo empíricamente.

Ejemplo de modelo económico

Un ejemplo de aplicación de un modelo sería tener el objetivo de explicar el comportamiento de la variable endógena “cantidad demandada de un producto”, Q_i . La teoría económica propondría el conjunto de variables que la explican a partir de las hipótesis de competencia perfecta, oligopolio, etc. A partir, pues, de la teoría económica se podría proponer un modelo económico concreto que sintetizaría la teoría:

$$Q_i = f(P_i, \bar{P}_i),$$


donde P_i sería el precio del bien y \bar{P}_i , el precio de los productos sustitutivos. A partir de este modelo se podría especificar un modelo econométrico.

2) El otro pilar básico, los **hechos** (sucesos del mundo real referidos al fenómeno que se investiga), se concreta en una serie de datos que pueden ser de corte transversal, si hacen referencia a distintos individuos en el mismo instante de tiempo, o de serie temporal, si se observan durante un periodo de tiempo determinado.

Para garantizar la calidad de los datos es necesario, a veces, someterlos a un tratamiento previo (deflación, enlace, interpolación de datos ausentes, obtención de la tendencia de la serie, etc.). Saber de qué información estadística se dispone (de qué variables se tiene información) también condiciona el modelo que se pueda proponer como idóneo.


Una vez que se especifica el modelo y se dispone de la información estadística convenientemente tratada, se llega a la etapa siguiente del trabajo econométrico: la **etapa de estimación**. En esta etapa se requiere utilizar toda una serie de técnicas econométricas que, en general, puede decirse que son extensiones de los métodos clásicos de estimación empleados en cursos básicos de estadística. El *output* de esta etapa de estimación permite medir y contrastar las relaciones sugeridas por la teoría económica.

1.4. Objetivos de la modelización econométrica

En la literatura econométrica existe un consenso bastante generalizado en fijar como objetivos de la modelización econométrica el análisis estructural, la predicción y la evaluación de políticas. De todos modos, hay que tener presente que estos tres objetivos no son excluyentes entre sí; por el contrario, un estudio econométrico puede tener uno o dos de ellos o incluso los tres. Observad en qué consiste cada uno de estos objetivos: 


1) El **análisis estructural** consiste en medir cuantitativamente las relaciones económicas entre las variables incluidas en el modelo. Asimismo, facilita la comparación de teorías rivales sobre un mismo fenómeno.


2) La **predicción** no es más que obtener los valores que determinadas variables tomarán fuera de la muestra. Así pues, las predicciones son (pueden ser) de gran utilidad para emprender determinadas acciones.

 Consultad cómo se puede especificar un modelo econométrico en el subapartado 2.1 de este módulo didáctico.

Ejemplos de tipos de datos

Las observaciones (datos) correspondientes a las ventas de un conjunto de empresas referidas a un mismo periodo (año, trimestre, etc.) constituyen un conjunto de datos de corte transversal. Por otro lado, las ventas de una empresa realizadas, por ejemplo, desde 1960 hasta 1997 constituyen una serie temporal.

 Recordad las estimaciones del modelo de regresión simple, que se han estudiado en la asignatura *Fundamentos de Estadística*.

 Consultad los objetivos de la modelización econométrica en el cuadro del subapartado 1.3 de este módulo didáctico.

3) La **evaluación de políticas** es el uso de los modelos econométricos para seleccionar entre políticas alternativas. En concreto, una posibilidad a menudo empleada por los *policymakers* (directores de la política económica de una empresa, de un estado, etc.) consiste en simular políticas alternativas y hacer predicciones condicionadas a futuros valores de las variables explicativas relevantes en cada alternativa.

Por último, no queremos finalizar este apartado sin poner de manifiesto la potencialidad del uso de los métodos econométricos en el mundo empresarial. Esto se debe a que los empresarios tienen que tomar sus decisiones en un entorno de incertidumbre. Sin duda, cualquier método que permita reducirla favorecerá esta tarea o, como mínimo, permitirá llevarla a término con más elementos de juicio.


Pensad, por ejemplo, en un empresario que tiene dos alternativas: aumentar su producción en el próximo ejercicio o mantenerla en los niveles actuales. Evidentemente, su experiencia puede resultar un elemento que se debe considerar, pero basar su decisión final únicamente en este factor puede ser insuficiente. Por lo tanto, si realiza un análisis que le permita conocer el comportamiento de su variable de interés (las ventas), podrá tomar una decisión con más argumentos. Para hacerlo, puede especificar un modelo y plantear distintas hipótesis sobre el comportamiento que experimentarán las ventas de su producto en el próximo ejercicio.

2. El modelo de regresión lineal múltiple estándar

En este apartado estudiaremos la especificación de un modelo de regresión lineal múltiple* estándar. Presentaremos las hipótesis básicas y, a partir de éstas, analizaremos los métodos adecuados de estimación de los parámetros del modelo. A continuación, se realizarán los contrastes estadísticos de significación de los parámetros y, finalmente, se estudiará la manera de llevar a cabo predicciones una vez que ya se ha estimado el modelo.

* A partir de aquí abreviaremos modelo de regresión lineal múltiple usando la sigla MRLM.

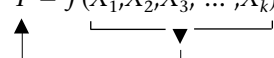
2.1. Especificación

En este apartado analizaremos un modelo de regresión que, tal como indica el título, presenta dos características importantes: 

1) Se trata de un **modelo de regresión múltiple**, lo cual supone que el comportamiento de una determinada variable, que denominaremos *variable endógena*, *variable dependiente* o *variable que se debe explicar* y que representaremos con la letra Y , es causado y, por tanto, puede ser explicado adecuadamente, por un conjunto de k variables que denominaremos *explicativas* (*independientes* o *exógenas*) y que, en general, representaremos mediante la letra X . Es decir:

$$Y = f(X_1, X_2, X_3, \dots, X_k). \quad (2.1)$$

Es importante destacar que existe una relación de causalidad unidireccional entre las variables explicativas y la endógena, y no al revés; es decir, la variable Y es una función de las variables X_1, \dots, X_k .

$$Y = f(X_1, X_2, X_3, \dots, X_k). \quad (2.2)$$


2) La segunda característica del modelo de regresión que estudiaremos se refiere a la **linealidad**. Esto quiere decir que la relación que hipotéticamente existe entre la variable endógena y las k explicativas es de tipo lineal; por lo tanto, podemos expresar la variable dependiente como combinación lineal de las variables explicativas. Aunque no es estrictamente necesario, normalmente especificaremos el MRLM incluyendo en las variables explicativas un término independiente. Así, a menudo se considera que la variable X_1 es una constante igual a la unidad:

$$X_1 = 1.$$

Como norma general supondremos siempre que $X_1 = 1$.

Por lo tanto, el MRLM queda del modo siguiente:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k, \quad (2.3)$$

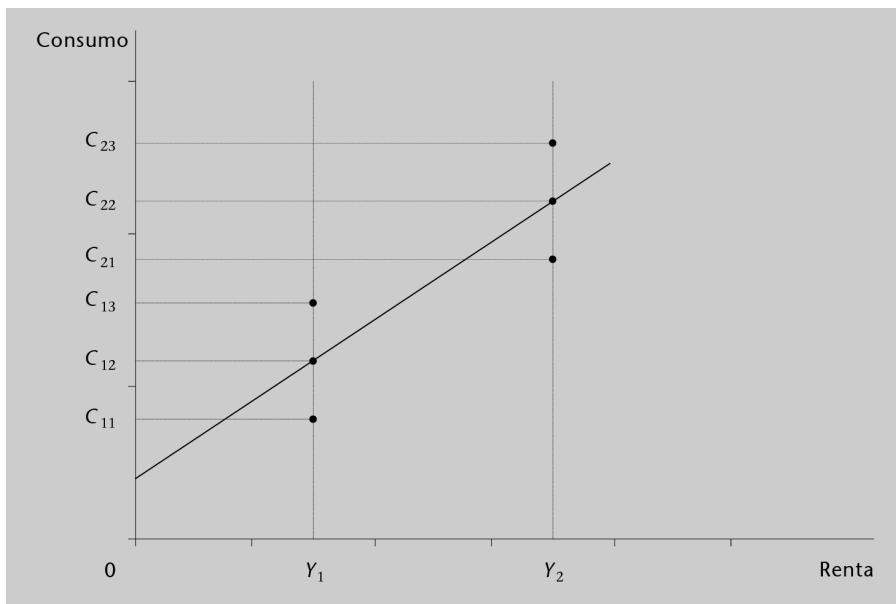
donde $\beta_1, \beta_2, \dots, \beta_k$ son los parámetros que supondremos constantes para el conjunto de la muestra de datos analizada. Estos parámetros nos permitirán medir la intensidad que tienen de media los efectos de las variaciones de las variables explicativas sobre la variable endógena. Esta interpretación de los parámetros β_j , $j = 1, \dots, k$ es inmediata si se toman derivadas parciales de la variable endógena respecto a cada una de las variables explicativas:

$$\frac{\partial Y}{\partial X_j} = \beta_j \quad \forall j = 2, \dots, k. \quad (2.4)$$

La relación entre la variable endógena Y y las variables explicativas X , tal como ya se ha visto hasta ahora, es determinista, es decir, no es aleatoria. No obstante, en la realidad, no se cumple casi nunca que las relaciones entre las variables económicas sean de este tipo, sino que las relaciones de dependencia tienen un cierto grado de aleatoriedad. Recordemos la función keynesiana de consumo, que estipula que el consumo de las unidades domésticas depende de su renta:

$$C_i = \alpha + \beta Y_i.$$

La función anterior es determinista, pero si preguntásemos a un conjunto de agentes económicos sobre sus niveles de renta y los recursos que destinan al consumo, nos hallaríamos con una situación como la del gráfico siguiente:



Relación funcional real renta-consumo

El gráfico nos dice que hay otros condicionantes en la decisión de consumo de los agentes económicos que no quedan reflejados en la renta. Esto se ve en el hecho de que los puntos tienen una determinada incertidumbre: no se encuentran exactamente sobre la recta, sino que se sitúan en algún punto más o menos próximo a la recta.

En consecuencia, es necesario incluir algún término en el modelo de regresión que capte esta aleatoriedad, ya que un modelo determinista no puede explicar totalmente el comportamiento de la variable endógena. Así, en un modelo de regresión se introduce el término de perturbación para recoger:

a) Todas las demás variables que explican el comportamiento de la variable endógena pero que no han quedado explicitadas como regresores*. Muchas

* Los regresores son las variables explicativas que se utilizan para especificar el modelo.

Observad la linealidad en los ejemplos de modelo simple y modelo múltiple en el subapartado 1.2 de este módulo didáctico.

de estas variables pueden ser pequeños factores de los cuales no se dispone de datos, y se supone que su efecto conjunto sobre la variable endógena es nulo.

b) Un segundo factor contenido en el término de perturbación es el mismo comportamiento aleatorio que hay en la conducta humana en particular, y en las relaciones económicas y sociales en general.

c) El tercer factor lo constituyen los errores de medida en las variables incluidas en el modelo y los errores en la ecuación*.

* Los errores en la ecuación se pueden deber a una mala especificación del modelo.

Pues bien, este término que incorporaremos se conoce con el nombre de **término de perturbación** y lo representaremos con la letra u .

Por lo tanto, cuando introducimos este término, el MRLM queda de la manera siguiente:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u. \tag{2.5}$$

\longleftarrow
 Parte determinista

\downarrow
 Parte aleatoria

Como ya hemos explicado, el término de perturbación incorpora el efecto conjunto de otras variables o comportamientos no explicitados en el modelo, para los cuales su efecto individual no resulta relevante. Por tanto, el término de perturbación no es realmente observable*, por lo cual tendremos que establecer una serie de hipótesis sobre su comportamiento.

* Si el término de perturbación fuese observable, se trataría como una variable explicativa del modelo.

Nuestro objetivo es asignar valores numéricos a los parámetros $\beta_1, \beta_2, \beta_3, \dots, \beta_k$. Es decir, pretendemos estimar el modelo de manera que, al combinar los valores de las estimaciones de los parámetros y los valores observados de las variables explicativas, obtengamos unos valores de la variable endógena tan cercanos a los valores reales observados de Y como sea posible.

Para poder hacer la estimación, se debe disponer de un conjunto de observaciones para cada una de las variables observables implicadas en el modelo, es decir, para la variable endógena Y y para las k variables explicativas. Denominaremos a las observaciones **valores muestrales**. Cuando trabajamos con datos de corte transversal, empleamos el subíndice i , mientras que, cuando trabajamos con datos de serie temporal, utilizamos el subíndice t . Así pues, podemos expresar el modelo del modo siguiente:

– Corte transversal:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, 2, 3, \dots, N. \tag{2.6}$$

– Serie temporal:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + u_t \quad t = 1, 2, 3, \dots, T.$$

Las ecuaciones anteriores expresan la relación de dependencia entre la variable endógena y las variables explicativas para la *i*-ésima y la *t*-ésima observaciones, respectivamente. Si no se dice lo contrario, a lo largo de este módulo supondremos que, por defecto, estamos trabajando con datos de corte transversal. Si desarrollamos esta ecuación para los *N* individuos de la muestra, tendremos:

$$\begin{array}{rcl}
 Y_1 & = & \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \dots + \beta_j X_{j1} + \dots + \beta_k X_{k1} + u_1 \\
 Y_2 & = & \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \dots + \beta_j X_{j2} + \dots + \beta_k X_{k2} + u_2 \\
 Y_3 & = & \beta_1 + \beta_2 X_{23} + \beta_3 X_{33} + \dots + \beta_j X_{j3} + \dots + \beta_k X_{k3} + u_3 \\
 \vdots & & \vdots & & \vdots & & \vdots \\
 Y_i & = & \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_j X_{ji} + \dots + \beta_k X_{ki} + u_i & (2.7) \\
 \vdots & & \vdots & & \vdots & & \vdots \\
 Y_N & = & \beta_1 + \beta_2 X_{2N} + \beta_3 X_{3N} + \dots + \beta_j X_{jN} + \dots + \beta_k X_{kN} + u_N
 \end{array}$$

donde, por ejemplo, Y_1 representa el valor de la variable endógena para la primera observación; Y_2 , el valor de la variable endógena para la segunda observación, etc. Por lo tanto, en general, Y_i es el valor de la variable endógena para la *i*-ésima observación.

X_{ji} representa el valor de la variable *j*-ésima ($j = 2, 3, \dots, k$) para la observación *i*-ésima ($i = 1, 2, 3, \dots, N$). Así, por ejemplo, X_{21} es el valor de la segunda variable explicativa en la primera observación, X_{46} es el valor que toma la cuarta variable explicativa para la sexta observación, etc.

Puesto que trabajar con el sistema anterior es bastante pesado, ya que tenemos tantas ecuaciones como observaciones, lo expresaremos habitualmente de manera matricial. Por lo tanto, la expresión 2.7 puede escribirse de la manera siguiente:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \tag{2.8}$$

donde \mathbf{Y} es un vector de dimensión *N*, que contiene las observaciones de la variable endógena:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_i \\ \vdots \\ Y_N \end{bmatrix} \begin{array}{l} 1.^a \text{ observación} \\ 2.^a \text{ observación} \\ 3.^a \text{ observación} \\ \\ i\text{-ésima observación} \\ \\ N\text{-ésima observación.} \end{array} \tag{2.9}$$

\mathbf{X} es una matriz de dimensión $N \times k$ que contiene los *N* valores que se han observado para las *k* variables explicativas:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{j1} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{j2} & \dots & X_{k2} \\ 1 & X_{23} & X_{33} & \dots & X_{j3} & \dots & X_{k3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & X_{2i} & X_{3i} & \dots & X_{ji} & \dots & X_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & X_{2N} & X_{3N} & \dots & X_{jN} & \dots & X_{kN} \end{bmatrix}. \tag{2.10}$$

En la matriz anterior tenemos en cada columna las observaciones de la variable explicativa correspondiente. Por ejemplo, en la primera columna está el término independiente; en la segunda columna, los valores de la variable X_2 , etc. Por filas, tenemos las observaciones de cada individuo, correspondientes a las k variables explicativas.

Observad que las dimensiones de las matrices son las adecuadas para realizar las operaciones.

\mathbf{B} es un vector de dimensión k , que contiene los k parámetros del modelo de regresión:

$$\mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix} \begin{array}{l} 1.^{\text{er}} \text{ parámetro} \\ 2.^{\text{a}} \text{ parámetro} \\ 3.^{\text{er}} \text{ parámetro} \\ \vdots \\ j\text{-ésimo parámetro} \\ \vdots \\ k\text{-ésimo parámetro.} \end{array} \quad (2.11)$$

\mathbf{U} es un vector de dimensión N , que contiene los N términos de perturbación, asociados a cada una de las ecuaciones:

$$\mathbf{U} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix} \begin{array}{l} 1.^{\text{a}} \text{ observación} \\ 2.^{\text{a}} \text{ observación} \\ 3.^{\text{a}} \text{ observación} \\ \vdots \\ i\text{-ésima observación} \\ \vdots \\ N\text{-ésima observación.} \end{array} \quad (2.12)$$

Hasta ahora, hemos especificado o formulado el modelo. A continuación, estudiaremos las hipótesis básicas del MRLM.

2.2. Hipótesis básicas del modelo de regresión lineal múltiple estándar

En el MRLM que ya hemos formulado, es necesario que hagamos un conjunto de hipótesis básicas para poder determinar las propiedades de los estimadores surgidos como resultado de la aplicación de métodos distintos de estimación y el tipo de contraste que hay que realizar para saber la significación de los parámetros. Estudiaremos los cuatro grupos de hipótesis siguientes: las hipótesis generales del MRLM, las hipótesis sobre el término de perturbación, las hipótesis sobre las variables explicativas del modelo, y las hipótesis sobre los parámetros del modelo.

Consultad la formulación del MRLM en el subapartado 2.1 de este módulo didáctico.

2.2.1. Hipótesis generales del modelo

Este conjunto de hipótesis se refiere al conjunto del modelo y, de hecho, ya se ha mencionado. Lo vemos a continuación:

Consultad la introducción de las hipótesis generales del modelo en el subapartado 2.1 de este módulo didáctico.

1) El MRLM es estocástico, y la relación de dependencia entre la variable endógena y las variables explicativas es lineal. Estas hipótesis pueden parecernos cier-

tamente restrictivas, pero, como veremos a continuación, no lo son. Hay numerosas relaciones no lineales entre variables que se pueden transformar, sin excesiva dificultad, en lineales. Simplemente hay que aplicar unas transformaciones sencillas a las variables implicadas en el modelo. Como ejemplo podemos considerar la función de producción de Cobb-Douglas, en la cual sólo hay que tomar logaritmos neperianos para obtener una expresión lineal:

$$Q_i = AL_i^{\beta_1}K_i^{\beta_2} \rightarrow \ln(Q_i) = \ln(A) + \beta_1\ln(L_i) + \beta_2\ln(K_i).$$

2) Supondremos que disponemos de información estadística suficientemente amplia sobre el conjunto de variables del modelo. Es un requisito mínimo que el número de observaciones sea mayor o igual que el número de parámetros que hay que estimar. Es decir, los grados de libertad del modelo deben ser iguales a cero o mayores ($N - k \geq 0$). Sin embargo, tal como veremos más adelante, es deseable que se pueda tener un número relativamente elevado de observaciones, con el fin de poder disponer de los grados de libertad suficientes para garantizar la fiabilidad de los resultados obtenidos en los procesos de estimación y de contraste estadístico.

El modelo de Cobb-Douglas...

... es intrínsecamente lineal, ya que se ha podido linealizar. Otros modelos no lineales son, además, intrínsecamente no lineales si no se pueden linealizar. Un ejemplo es $y_i = \alpha + \beta_1 e^{\beta_2 x_i}$. Puesto que presentan un mayor grado de complejidad, este tipo de modelos queda fuera del análisis de este módulo didáctico.

Consultad en el subapartado 2.3.2 una razón para la conveniencia de tener un número relativamente elevado de observaciones.

2.2.2. Hipótesis sobre el término de perturbación

El conjunto de hipótesis que formularemos a continuación hace referencia al comportamiento del término de perturbación, que, como ya hemos dicho antes, es la fuente de aleatoriedad del modelo y el término que incluye todas aquellas variables o aspectos que puntualmente han tenido influencia en el comportamiento de la variable endógena. De todas maneras, este término por sí solo no tiene ningún poder explicativo sobre la evolución de la variable endógena.

Consultad el término de perturbación en el subapartado 2.1 de este módulo didáctico.

A continuación, presentamos las hipótesis sobre el término de perturbación:

1) Supondremos que la esperanza matemática de los términos de perturbación es cero, es decir:

$$E[u_i] = 0 \quad \forall i = 1, \dots, N,$$

Recordad los conceptos de esperanza matemática y de varianza de una variable aleatoria, tratados en la asignatura *Fundamentos de Estadística*.

o, en notación matricial:

$$E[\mathbf{U}] = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} E[u_1] \\ E[u_2] \\ E[u_3] \\ \vdots \\ E[u_i] \\ \vdots \\ E[u_N] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}_{N \times 1}. \tag{2.13}$$

Lo que se supone con esta hipótesis es que, por término medio, el efecto conjunto de los factores incluidos en el término de perturbación es nulo. Es decir, que los efectos puntuales de las variables que no se consideran relevantes se compensan entre sí.

Como veremos más adelante, la hipótesis anterior se cumplirá siempre que el modelo esté especificado correctamente, en el sentido de que todas las variables relevantes, a la hora de explicar el comportamiento de la variable endógena, se han incorporado a la matriz X .

2) Supondremos que el término de perturbación tiene varianza constante para todas las observaciones. Esta propiedad se llama **homoscedasticidad**.

$$\text{VAR}[u_i] = \sigma_u^2 \quad \forall i = 1, \dots, N. \quad (2.14)$$

El incumplimiento de esta hipótesis, es decir, el hecho de que no todos los términos de perturbación tengan la misma varianza, se conoce con el nombre de **heteroscedasticidad**.

3) Además de las hipótesis que hemos hecho sobre la esperanza y la varianza, también se establecen hipótesis sobre la covarianza entre cada dos términos de perturbación. Recordemos que, según el concepto de covarianza, tenemos lo siguiente:

$$\text{COV}[u_i, u_j] = E[(u_i - E[u_i])(u_j - E[u_j])].$$

Por lo tanto, dado que hemos supuesto que la esperanza matemática del término de perturbación es $E[u] = 0$, a partir de la expresión anterior obtenemos la expresión de la covarianza e imponemos que valga cero:

$$\text{COV}[u_i, u_j] = E[u_i u_j] = 0 \quad \forall i, j = 1, \dots, N.$$

Así pues, se supone que no hay autocorrelación entre los distintos términos de perturbación, es decir, se da por supuesto que los términos de perturbación son independientes entre sí.


$$\text{COV}[u_i, u_j] = 0 \quad \forall i, j = 1, \dots, N. \quad (2.15)$$

Si el término de perturbación del modelo cumple las propiedades de homoscedasticidad y de ausencia de autocorrelación, se dice que es **esférico**. Cuando el término de perturbación presenta heteroscedasticidad o está autorrelacionado, o ambas cosas a la vez, se dice que es **no esférico**.

Con referencia a la corrección del modelo, consultad las hipótesis sobre las variables explicativas en el subapartado 2.2.3 de este módulo didáctico.

Nota

Para que exista heteroscedasticidad basta con que uno de los términos de perturbación tenga una varianza distinta de la del resto.

Por otro lado, cuando el término de perturbación del modelo cumple las propiedades de homoscedasticidad y de ausencia de autocorrelación, decimos que su matriz de varianzas y covarianzas es **escalar**. La forma general de la matriz de varianzas y covarianzas es la siguiente: 

$$\text{VAR}[\mathbf{U}] = \begin{bmatrix} \text{VAR}[u_1] & \text{COV}[u_1, u_2] & \text{COV}[u_1, u_3] & \dots & \text{COV}[u_1, u_N] \\ \text{COV}[u_2, u_1] & \text{VAR}[u_2] & \text{COV}[u_2, u_3] & \dots & \text{COV}[u_2, u_N] \\ \text{COV}[u_3, u_1] & \text{COV}[u_3, u_2] & \text{VAR}[u_3] & \dots & \text{COV}[u_3, u_N] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{COV}[u_N, u_1] & \text{COV}[u_N, u_2] & \text{COV}[u_N, u_3] & \dots & \text{VAR}[u_N] \end{bmatrix}.$$

Y, teniendo en cuenta las fórmulas 2.14 y 2.15, obtenemos la matriz de varianzas y covarianzas siguiente:

$$\text{VAR}[\mathbf{U}] = \begin{bmatrix} \sigma_u^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{I}_N,$$

donde \mathbf{I}_N indica la matriz identidad de dimensión $N \times N$.

Matriz de varianzas y covarianzas

La matriz de varianzas y covarianzas del término de perturbación tiene una serie de características que son comunes a todas las matrices de varianzas y covarianzas:

- Se trata de una matriz cuadrada, en nuestro caso de dimensión $N \times N$.
- Es simétrica, ya que $\text{COV}[u_i, u_j] = \text{COV}[u_j, u_i]$.
- Es una matriz definida positiva, ya que los valores de los elementos de su diagonal son varianzas (y, por tanto, positivos) y se puede comprobar que los menores de la matriz también son definidos positivos.

Una manera alternativa de resumir las propiedades de homoscedasticidad y de ausencia de autocorrelación del término de perturbación es la siguiente:

$$E[u_i u_j] = \begin{cases} \sigma_u^2 & \forall i = j \quad i, j = 1, \dots, N. \\ 0 & \forall i \neq j \quad i, j = 1, \dots, N. \end{cases} \quad (2.16)$$

4) La última hipótesis que formularemos respecto al comportamiento del término de perturbación es que cada uno de los componentes se distribuyen según una ley normal.



Ley de distribución normal

Recordad la representación gráfica de la función de densidad de una distribución normal con esperanza 0 y varianza σ^2 , $N(0, \sigma^2)$.

Para finalizar con las hipótesis relativas al término de perturbación, podemos ver que todas son susceptibles de ser resumidas en notación matricial en la expresión siguiente:

$$\mathbf{U} \sim N(\mathbf{0}_{N \times 1}, \sigma_u^2 \mathbf{I}_N),$$

y para cada componente del vector del término de perturbación:

$$u_i \sim N(0, \sigma_u^2) \quad \forall i = 1, \dots, N.$$

2.2.3. Hipótesis sobre las variables explicativas del modelo

En este subapartado presentamos las cinco hipótesis que hacen referencia al comportamiento de las variables explicativas del modelo, que son las siguientes:

1) Supondremos que las variables explicativas son fijas o deterministas*. Esto significa, como ya hemos mencionado antes, que la única fuente de aleatoriedad que tiene el modelo es el término de perturbación, que da carácter estocástico al modelo.

* El hecho de decir que las variables explicativas son fijas o deterministas es equivalente a decir que son variables no aleatorias.

2) Supondremos que las variables explicativas están incorrelacionadas con el término de perturbación, es decir:

$$E[X_{ji}u_i] = 0 \quad \forall i, j; \quad i = 1, \dots, N; \quad j = 1, 2, \dots, k. \quad (2.17)$$

Esta hipótesis sólo es una consecuencia del hecho de que se supone que las variables explicativas son fijas.

3) Supondremos que no hay ninguna relación lineal exacta entre las variables explicativas del modelo, es decir, que no hay ninguna combinación lineal exacta entre las columnas de la matriz X. Cuando esto suceda, diremos que no hay multicolinealidad perfecta entre las variables explicativas.

Para que no haya multicolinealidad perfecta es necesario que la matriz X tenga rango pleno o máximo, es decir, que $\rho(X) = k$.

Denotamos el rango de una matriz X por $\rho(X)$.

4) Además, también supondremos que las variables explicativas se han medido sin error.

5) Finalmente, supondremos que en el modelo no se han excluido variables relevantes y que tampoco se han incluido variables irrelevantes. Es decir, supondremos que todas las variables relevantes, en el momento de explicar el comportamiento de la variable endógena, se encuentran en la matriz X.

2.2.4. Hipótesis sobre los parámetros del modelo

En cuanto a los parámetros de la población, la única hipótesis que haremos sobre su comportamiento es la que se conoce con el nombre de **permanencia estructural**, que consiste en decir que los parámetros de interés en el modelo (β_j) son constantes para toda la muestra.

2.3. Estimación para mínimos cuadrados ordinarios (MCO)

En este subapartado veremos, en primer lugar, cómo podemos encontrar los estimadores de un modelo por el método de los mínimos cuadrados ordinarios y, a continuación, detallaremos sus propiedades.

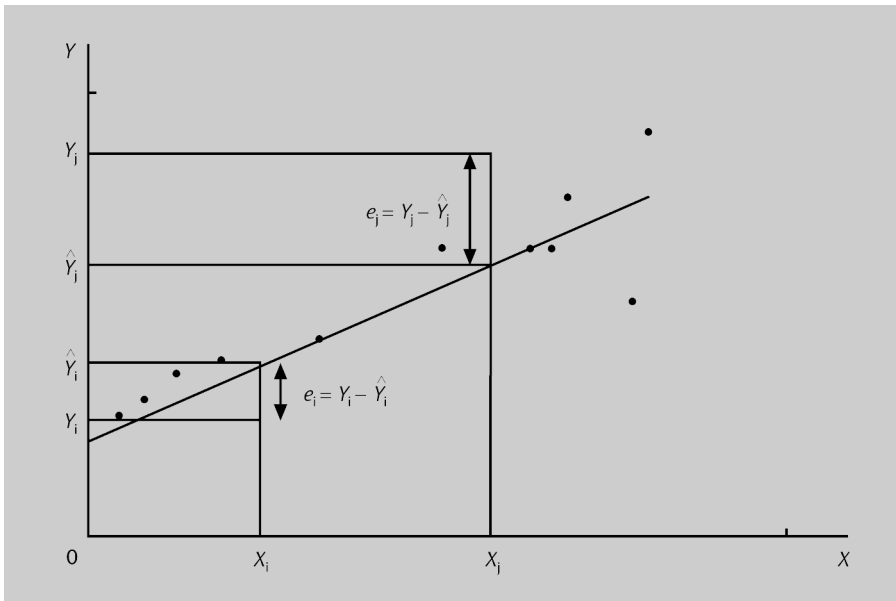
2.3.1. Descripción del método de estimación

Una vez que hemos formulado el modelo de regresión y hemos establecido las hipótesis sobre el comportamiento de sus términos, el paso siguiente consiste en estimarlo. Es decir, asignar valores numéricos a los parámetros β_j desconocidos, a partir de la información muestral disponible de las variables observables del modelo. Con esta finalidad disponemos de varios métodos de estimación, de los cuales sólo estudiaremos dos:

- El **método de mínimos cuadrados ordinarios (MCO)**, que presentaremos a continuación.
- El **método de máxima verosimilitud (MV)**, que presentaremos en otro apartado de este módulo, más adelante.

Antes de iniciar la descripción del método, veremos gráficamente la idea en la que se basa este procedimiento de estimación. Para hacerlo, emplearemos el **modelo de regresión lineal simple (MRLS)**:

$$Y_i = \alpha + \beta X_i + u_i \quad \forall i = 1, \dots, N.$$



En el gráfico anterior observamos la nube de puntos que resulta de representar en los ejes cartesianos las observaciones correspondientes a las variables endógena y explicativa. Como ya hemos dicho, nuestro objetivo es asignar valores numéricos a los parámetros desconocidos, en este caso α y β , y de esta manera poder cuantificar la relación de dependencia que existe entre ambas variables. Determinar estos valores es equivalente a determinar una recta que pase a través de la nube de puntos. Sin embargo, no nos interesa cualquier recta, sino que queremos encontrar aquella recta que se ajuste mejor a la nube de puntos, es decir, a los datos disponibles. !

Consultad el método de estimación por máxima verosimilitud en el subapartado 2.5 de este módulo didáctico. !

Consultad el modelo de regresión lineal simple en la asignatura *Fundamentos de Estadística*. !

Hemos definido los objetivos para los cuales se especifica un modelo en el subapartado 2.1 de este módulo didáctico. !

Pues bien, lo que nos interesa es que la diferencia señalada en el gráfico $e_i = Y_i - \hat{Y}_i$ entre la variable endógena observada (Y_i) y la estimada (\hat{Y}_i) sea mínima para cada una de las observaciones. A esta diferencia, e_i , la denominaremos **error** o **residuo**, y la hemos definido del modo siguiente:

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}) \quad \forall i = 1, \dots, N, \quad (2.18)$$


y en notación matricial diremos:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}.$$

Lo que nos interesa al hacer la estimación es que el ajuste minimice estos errores, es decir, que los coeficientes estimados resulten de tal manera que la diferencia entre el valor observado y el estimado de la variable endógena sea tan pequeña como sea posible.

De esta manera, en principio, podríamos pensar que un buen criterio para asignar los valores a los β_j sería minimizar la suma de los errores cometidos, es decir, $\min \sum_{i=1}^N e_i$. En cambio, este criterio no nos resulta útil, ya que se compensan los errores positivos y los errores negativos, tal como se puede ver en el gráfico anterior. Una alternativa puede consistir en minimizar la suma de errores cometidos en valor absoluto, $\min \sum_{i=1}^N |e_i|$, pero esto dificulta el proceso de estimación porque el operador valor absoluto no es diferenciable; por tanto, hay que emplear otra alternativa, que consiste en minimizar la suma de cuadrados de los errores:

$$\min \sum_{i=1}^N e_i^2 = \min \sum_{i=1}^N (Y_i - \hat{Y}_i)^2. \quad (2.19)$$

Así evitamos el problema de la compensación de signos positivos y negativos, y tenemos una función derivable. Por lo tanto, los **estimadores de mínimos cuadrados ordinarios** (MCO) son aquellos que minimizan la suma de cuadrados de las diferencias entre los valores reales (observados) y los estimados (los que se encuentran a partir del ajuste) de la variable endógena, es decir, que hacen mínima la suma de cuadrados de los residuos, que denominaremos **suma de cuadrados de los errores** y que representaremos con la sigla *SCE*. Podemos expresar la *SCE* del modo siguiente: 


$$SCE = \sum_{i=1}^N e_i^2 = e_1^2 + e_2^2 + e_3^2 + \dots + e_i^2 + \dots + e_N^2.$$

Y también la podemos escribir en notación matricial de las maneras que presentamos a continuación:

$$\bullet \quad SCE = \mathbf{e}'\mathbf{e} = [e_1 \ e_2 \ e_3 \ \dots \ e_N] \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_N \end{bmatrix}$$

En el MRLM...

... hay un término que no es observable; se trata del término de perturbación u . Un modo de obtener una aproximación de su valor es ver la diferencia entre Y y $X\hat{B}$, que llamamos e .

 Recordad las propiedades de la transposición de matrices, estudiadas en la asignatura *Fundamentos de Matemáticas*.

- $$\begin{aligned}
 SCE &= \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) = \\
 &= (\mathbf{Y}' - \hat{\mathbf{B}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}.
 \end{aligned}
 \tag{2.20}$$

Puesto que la SCE es una escalar, cada uno de los sumandos también lo es; por lo tanto, vemos que podemos escribir la expresión 2.20 de la manera siguiente:

$$SCE = \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}.$$

Pues bien, la expresión anterior es la que minimizaremos, por lo cual calcularemos la derivada parcial respecto a $\hat{\mathbf{B}}$ y la igualaremos a cero:

$$\frac{\partial SCE}{\partial \hat{\mathbf{B}}} = \frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\mathbf{B}}} = \frac{\partial (\mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}})}{\partial \hat{\mathbf{B}}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = 0$$

A partir de la última ecuación, podemos realizar las operaciones siguientes:

$$2\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = 2\mathbf{X}'\mathbf{Y} \Rightarrow \mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = \mathbf{X}'\mathbf{Y} \Rightarrow (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

De aquí se deduce la expresión del estimador MCO del MRLM que usaremos repetidamente.

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{2.21}$$

Para comprobar que en la expresión anterior hemos obtenido un mínimo (y no un máximo), tendremos que calcular la segunda derivada de la expresión de la SCE respecto a $\hat{\mathbf{B}}$.

$$\begin{aligned}
 \frac{\partial^2 SCE}{\partial \hat{\mathbf{B}} \partial \hat{\mathbf{B}}'} &= \frac{\partial^2 \mathbf{e}'\mathbf{e}}{\partial \hat{\mathbf{B}} \partial \hat{\mathbf{B}}'} = \frac{\partial^2 (\mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}})}{\partial \hat{\mathbf{B}} \partial \hat{\mathbf{B}}'} = \\
 &= \frac{\partial (-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\mathbf{B}})}{\partial \hat{\mathbf{B}}} = 2\mathbf{X}'\mathbf{X}.
 \end{aligned}$$

Puesto que el resultado de esta operación es una matriz definida positiva, efectivamente tenemos un mínimo.

Si volvemos a la expresión del estimador MCO, observamos que hallamos $\hat{\mathbf{B}}$ a partir de lo que se conoce con el nombre de **sistema de ecuaciones normales** $\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = \mathbf{X}'\mathbf{Y}$.

Analizamos, a continuación, los elementos de este sistema:

- La matriz $\mathbf{X}'\mathbf{X}$ tiene los elementos siguientes:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & X_{23} & \dots & X_{2N} \\ X_{31} & X_{32} & X_{33} & \dots & X_{3N} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \dots & X_{kN} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ 1 & X_{23} & X_{33} & \dots & X_{k3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{2N} & X_{3N} & \dots & X_{kN} \end{bmatrix} =$$

La derivación respecto a vectores...

... sigue unos criterios similares a los de la derivación respecto a una única variable. De hecho, se tiene que derivar componente a componente.

Recordad que...

... para constatar que una expresión es un mínimo se debe comprobar que la matriz de derivadas segundas es definida positiva.

$$= \begin{bmatrix} N & \sum_{i=1}^N X_{2i} & \sum_{i=1}^N X_{3i} & \dots & \sum_{i=1}^N X_{ki} \\ \sum_{i=1}^N X_{2i} & \sum_{i=1}^N X_{2i}^2 & \sum_{i=1}^N X_{2i}X_{3i} & \dots & \sum_{i=1}^N X_{2i}X_{ki} \\ \sum_{i=1}^N X_{3i} & \sum_{i=1}^N X_{3i}X_{2i} & \sum_{i=1}^N X_{3i}^2 & \dots & \sum_{i=1}^N X_{3i}X_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N X_{ki} & \sum_{i=1}^N X_{ki}X_{2i} & \sum_{i=1}^N X_{ki}X_{3i} & \dots & \sum_{i=1}^N X_{ki}^2 \end{bmatrix}.$$

Como puede verse, la matriz X'X es cuadrada y simétrica.

- La matriz X'Y se escribe así:

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & X_{23} & \dots & X_{2N} \\ X_{31} & X_{32} & X_{33} & \dots & X_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \dots & X_{kN} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N Y_i \\ \sum_{i=1}^N X_{2i}Y_i \\ \sum_{i=1}^N X_{3i}Y_i \\ \vdots \\ \sum_{i=1}^N X_{ki}Y_i \end{bmatrix}.$$

Si volvemos a la expresión de los estimadores por MCO, $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, el único problema con el que nos podríamos encontrar a la hora de calcularlos es que la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ no existiese. Esta situación se dará cuando el determinante de la matriz X'X sea cero, lo cual puede suceder, por ejemplo, cuando exista una combinación lineal exacta entre las columnas de la matriz X y, por tanto, cuando el rango de las matrices X y X'X sea inferior a k ($\rho(\mathbf{X}'\mathbf{X}) < k$). Es decir, habrá multicolinealidad perfecta, lo cual supone que no se cumple una de las hipótesis básicas del MRLM.

Ejemplo de estimación de un modelo concreto

Explicaremos las ventas (Y_i) de diez empresas en función de sus gastos respectivos en publicidad (X_{2i}) y del precio medio de sus productos (X_{3i}). Especificamos el modelo siguiente:

Como puede verse, optamos por incluir un término independiente en el modelo.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \forall i = 1, \dots, 10.$$

Los datos que utilizaremos están en la tabla siguiente:

Y_i	X_{2i}	X_{3i}
16	0	13
18	1	12
10	4	19
13	4	17
18	2	13
38	8	10
25	5	11
41	10	14
11	3	17
21	4	16

La matriz X de las variables explicativas y el vector Y de la variable endógena son, pues, las siguientes:

$$X = \begin{bmatrix} 1 & 0 & 13 \\ 1 & 1 & 12 \\ 1 & 4 & 19 \\ 1 & 4 & 17 \\ 1 & 2 & 13 \\ 1 & 8 & 10 \\ 1 & 5 & 11 \\ 1 & 10 & 14 \\ 1 & 3 & 17 \\ 1 & 4 & 16 \end{bmatrix}; \quad Y = \begin{bmatrix} 16 \\ 18 \\ 10 \\ 13 \\ 18 \\ 38 \\ 25 \\ 41 \\ 11 \\ 21 \end{bmatrix}.$$

Después de hacer la transposición de la matriz X , podemos calcular los productos $X'X$ y $X'Y$ y llegamos a los resultados siguientes:

$$X'X = \begin{bmatrix} 10 & 41 & 142 \\ 41 & 251 & 572 \\ 142 & 572 & 2.094 \end{bmatrix}; \quad X'Y = \begin{bmatrix} 211 \\ 1.102 \\ 2.821 \end{bmatrix}.$$

Para calcular $(X'X)^{-1}$ es necesario calcular el determinante y la matriz de adjuntos de la matriz $X'X$, ya que:

$$(X'X)^{-1} = \frac{1}{|X'X|} \text{Adj}(X'X) = \frac{1}{63.290} \begin{bmatrix} 198.410 & -4.630 & -12.190 \\ -4.630 & 776 & 102 \\ -12.190 & 102 & 829 \end{bmatrix}.$$

Finalmente, obtenemos el vector de estimadores de mínimos cuadrados ordinarios a partir de su definición:

$$\begin{aligned} \hat{B} &= (X'X)^{-1}X'Y = \frac{1}{|X'X|} \text{Adj}(X'X)X'Y = \\ &= \frac{1}{63.290} \begin{bmatrix} 198.410 & -4.630 & -12.190 \\ -4.630 & 776 & 102 \\ -12.190 & 102 & 829 \end{bmatrix} \begin{bmatrix} 211 \\ 1.102 \\ 2.821 \end{bmatrix} = \begin{bmatrix} 37,5140 \\ 2,6223 \\ -1,9131 \end{bmatrix}. \end{aligned}$$

Por lo tanto, nuestro modelo queda estimado de la manera siguiente:

$$\hat{Y}_i = 37,5140 + 2,6223X_{2i} - 1,9131X_{3i} \quad \forall i = 1, \dots, 10.$$

2.3.2. Propiedades de los estimadores MCO de los β_j

Una vez que hemos visto cómo se obtienen los estimadores MCO de los parámetros β_j y la manera de calcularlos, estudiaremos sus propiedades que, como veremos, son: linealidad, ausencia de sesgo, eficiencia y consistencia.

Linealidad

Respecto a esta primera propiedad, se debe decir que la podemos comprobar directamente a partir de la expresión de los estimadores, que son una combinación lineal de los componentes del vector Y .

Asimismo, si sustituimos Y por $XB + U$ en la expresión de los estimadores, tendremos:

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{B} + \mathbf{U}) = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}.\end{aligned}\quad (2.22)$$

Como se puede observar, los estimadores MCO quedan expresados como una combinación lineal del término de perturbación, de las variables explicativas y de los parámetros de la población. Puesto que el vector de estimadores MCO es una combinación lineal del término de perturbación y éste es la fuente de aleatoriedad del modelo, los estimadores también serán un vector de variables aleatorias. Por lo tanto, si el término de perturbación se distribuye según una ley normal, el **vector de estimadores MCO** también se distribuirá según una ley normal:

$$\hat{\mathbf{B}} \sim N(E[\hat{\mathbf{B}}], \text{VAR}[\hat{\mathbf{B}}]), \quad (2.23)$$

y cada uno de los estimadores individualmente:

$$\hat{\beta}_j \sim N(E[\hat{\beta}_j], \text{VAR}[\hat{\beta}_j]) \quad \forall j = 1, \dots, k.$$

Por lo tanto, estamos interesados en tener perfectamente descrito el comportamiento de los estimadores MCO, en conocer cuál es su valor esperado y su **matriz de varianzas y covarianzas**.

Ausencia de sesgo

Recordemos que el **sesgo de un estimador** es la diferencia entre su esperanza matemática y el valor del parámetro que estima, es decir:

$$\text{Sesgo}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Un estimador no tiene **sesgo** cuando su valor esperado coincide con el parámetro de la población que se quiere estimar.

Así, para ver si los estimadores MCO no tienen sesgo calculamos el valor esperado:

$$\begin{aligned}E[\hat{\mathbf{B}}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = E[\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}] = \\ &= E[\mathbf{B}] + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}] = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{U}] = \mathbf{B}.\end{aligned}\quad (2.24)$$

Entonces el estimador MCO de \mathbf{B} no tiene sesgo.

Eficiencia

Esta tercera propiedad de los estimadores MCO se refiere a la varianza de los estimadores.

Para llegar al resultado...

... que nos muestra que el estimador MCO no tiene sesgo, nos hemos basado en el hecho de que los parámetros de la población y los elementos de la matriz X son cantidades fijas y de que la esperanza del término de perturbación es cero.

Enunciaremos esta propiedad diciendo que los estimadores MCO son **estimadores eficientes** o **estimadores óptimos**. Lo que queremos decir es que, en el conjunto de todos los estimadores lineales sin sesgo, los estimadores MCO son los que tienen una varianza más pequeña.

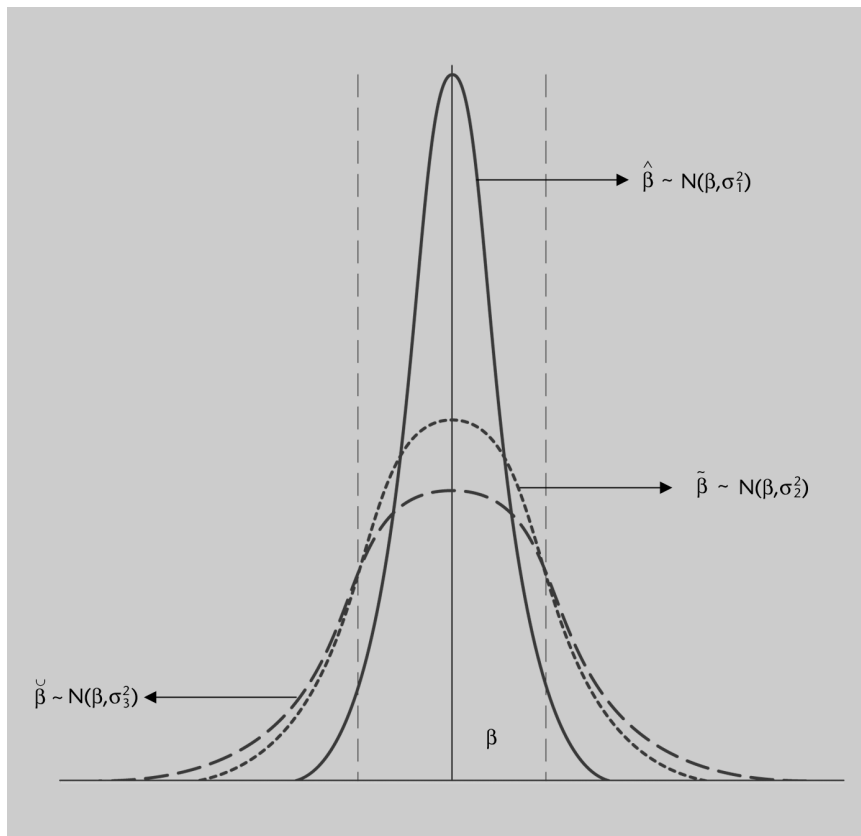
Por lo tanto, los estimadores por mínimos cuadrados ordinarios son los que tienen una dispersión más baja en torno a su valor esperado (es decir, el valor de la población).

Es importante que los estimadores MCO cumplan la propiedad que estamos considerando, ya que la ausencia de sesgo solamente no garantiza que el valor numérico de un estimador MCO esté muy cerca o demasiado alejado del verdadero valor del coeficiente, sino que sólo dice que, por término medio, coincide con este valor.

Comparación de estimadores lineales y sin sesgo

Supongamos que tenemos tres estimadores, todos lineales y sin sesgo, $\hat{\beta}$, $\tilde{\beta}$ y $\check{\beta}$, que se distribuyen según una ley normal, todos con el mismo valor esperado, que coincide con el valor de la población β (porque no tienen sesgo).

La única diferencia entre los tres estimadores es que tienen varianza distinta $-\sigma_1^2$, σ_2^2 i σ_3^2 , respectivamente-, de modo que $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$. Esta situación puede observarse en el gráfico siguiente:



Como puede verse, de los tres estimadores, el que preferimos es el primero ($\hat{\beta}$), porque puesto que tiene la varianza menor (menos dispersión en torno a su valor esperado) es el que tiene la probabilidad más elevada de tomar valores dentro del intervalo señalado.

Esta propiedad, la eficiencia, garantiza que no exista otro estimador lineal sin sesgo que tenga una varianza menor que la varianza que tiene el estimador

MCO. El **teorema de Gauss-Markov** demuestra que la $\text{VAR}[\hat{\beta}_j]$ es la menor de todo el conjunto de estimadores lineales y no sesgados.

A continuación veremos cuál es la expresión concreta de la matriz de varianzas y covarianzas de los estimadores:

$$\begin{aligned}\text{VAR}[\hat{\mathbf{B}}] &= E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})'] = \\ &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}]'\} = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}\mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{U}\mathbf{U}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_u^2\mathbf{I}_N\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\ &= \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}\quad (2.25)$$

Debemos tener presente que para llegar a esta expresión de la matriz de varianzas y covarianzas del vector de estimadores MCO, hemos utilizado las hipótesis básicas siguientes:

- 1) Los elementos de la matriz \mathbf{X} son fijos.
- 2) La matriz de varianzas y covarianzas del término de perturbación es escalar, es decir:

$$E[\mathbf{U}\mathbf{U}'] = \sigma_u^2\mathbf{I}_N.$$

De la expresión 2.25 se deduce que la varianza de cada estimador es:

$$\text{VAR}[\hat{\beta}_j] = \sigma_u^2 a_{jj} \quad \forall j = 1, \dots, k,$$

donde a_{jj} es el elemento j -ésimo de la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

Consistencia

Un estimador es consistente cuando, al aumentar el tamaño de la muestra, el estimador se aproxima más al verdadero valor de la población.

Nosotros estudiaremos la consistencia en términos del error cuadrático medio (ECM). Por lo tanto, para comprobar que los estimadores MCO son consistentes tendremos que calcular el $\lim_{N \rightarrow \infty} \text{ECM}[\hat{\mathbf{B}}]$ y ver si es igual a cero.

Sabemos que el error cuadrático medio puede expresarse como:

$$\text{ECM}[\hat{\mathbf{B}}] = \text{VAR}[\hat{\mathbf{B}}] + (\text{Sesgo}[\hat{\mathbf{B}}])^2,$$

y si sabemos que los estimadores MCO no tienen sesgo, entonces tendremos:

$$\text{ECM}[\hat{\mathbf{B}}] = \text{VAR}[\hat{\mathbf{B}}] = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}.\quad (2.26)$$

Finalmente, para ver si el ECM tiende a cero a medida que se incrementa el tamaño muestral, sólo hará falta ver si las varianzas de los estimadores $\hat{\mathbf{B}}$ tien-

Lecturas complementarias

Podéis encontrar la demostración del teorema de Gauss-Markov en la mayoría de los manuales de econometría; por ejemplo, podéis consultar las obras siguientes:

J. Johnston (1987). *Métodos de econometría* (trad. J. Sánchez Fernández). Barcelona: Vicens-Vives.
A. Novales (1993). *Econometría* (2.ª ed.). Madrid: McGraw-Hill.

den a cero. Si multiplicamos y dividimos por N , podemos escribir la expresión siguiente:

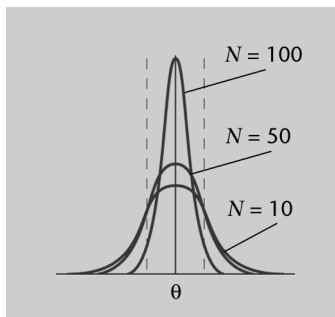
$$\lim_{N \rightarrow \infty} \text{ECM}[\hat{\mathbf{B}}] = \lim_{N \rightarrow \infty} \text{VAR}[\hat{\mathbf{B}}] = \lim_{N \rightarrow \infty} \left[\frac{\sigma_u^2}{N} \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \right] = 0. \quad (2.27)$$

Para obtener el resultado anterior, suponemos que $\left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1}$ existe cuando $N \rightarrow \infty$ y que es un número finito.

Comparación de estimadores sin sesgo al variar el tamaño de la muestra

Decimos que un estimador $\hat{\theta}$ es consistente si el límite de su error cuadrático medio cuando N tiende a infinito es igual a cero: $\lim_{N \rightarrow \infty} \text{ECM}[\hat{\theta}] = 0$.

En el gráfico representamos casos de estimadores sin sesgo; por ello se distribuyen en torno al valor central θ .



En el gráfico puede observarse...

... cómo, al aumentar el tamaño muestral (N), la varianza de $\hat{\theta}$ se hace más pequeña.

Como resumen de las propiedades que hemos analizado del estimador $\hat{\mathbf{B}}_{\text{MCO}}$ podemos decir que es lineal, sin sesgo, óptimo y consistente. Además, puesto que $\hat{\mathbf{B}}$ es un vector de variables aleatorias que se distribuye según una norma multivariante de orden k , podemos resumir sus propiedades de la manera siguiente:

$$\hat{\mathbf{B}} \sim N[\mathbf{B}, \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}]. \quad (2.28)$$

Y para cada uno de los estimadores $\hat{\beta}_j$, individualmente:

$$\hat{\beta}_j \sim N[\hat{\beta}_j, \sigma_u^2 a_{jj}] \quad \forall j = 1, \dots, k.$$

2.4. Análisis de los residuos y estimación de σ_u^2

Antes de comenzar el análisis de los residuos, o errores, del ajuste MCO es necesario que justifiquemos por qué queremos realizar este análisis.

Como hemos visto en un apartado anterior, el MRLM se basa, entre otros aspectos, en las hipótesis establecidas sobre el término de perturbación. Por lo tanto, hay que comprobar en primer lugar si este término las cumple. El

Recordad el concepto de residuo, que hemos visto en el subapartado 2.3.1 de este módulo didáctico.

problema es que el término de perturbación no es observable. Así, si no empleamos una estimación, nos resultará muy difícil comprobar el cumplimiento de sus hipótesis básicas.

Consultad las hipótesis establecidas sobre el término de perturbación en el subapartado 2.2.2 de este módulo didáctico.

Los estimadores del término de perturbación que utilizaremos son los **residuos**. Se usarán para comprobar si las hipótesis básicas se cumplen. A continuación, analizaremos las propiedades de los residuos MCO.

2.4.1. Propiedades de los residuos

Los residuos presentan las propiedades siguientes:

1) El vector de residuos se puede expresar como una combinación lineal de los elementos del vector de la variable endógena:

$$e = Y - \hat{Y} = Y - X\hat{B} = Y - X(X'X)^{-1}X'Y = [I_N - X(X'X)^{-1}X']Y = MY. \quad (2.29)$$

La matriz M tiene las propiedades siguientes:

- a) Es una matriz cuadrada de dimensión $N \times N$: $M = I_N - X(X'X)^{-1}X'$.
- b) Es una matriz simétrica, es decir, $m_{ij} = m_{ji} \quad \forall i \neq j \quad i = 1, \dots, N \quad j = 1, \dots, N$.
- c) Es una matriz idempotente, es decir, que $MM = MM' = M'M = M'M' = M$.

Comprobación de la propiedad de idempotencia

Podemos comprobar que la matriz M es idempotente:

$$\begin{aligned} MM &= [I_N - X(X'X)^{-1}X'] [I_N - X(X'X)^{-1}X'] = \\ &= I_N - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' = \\ &= I_N - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X' = I_N - X(X'X)^{-1}X' = M. \end{aligned}$$

d) La matriz M es ortogonal a la matriz de variables explicativas, es decir, $MX = \mathbf{0}_{N \times k}$ y $X'M' = \mathbf{0}_{k \times N}$. Efectivamente,

$$MX = [I_N - X(X'X)^{-1}X']X = X - X(X'X)^{-1}X'X = X - X = \mathbf{0}_{N \times k}.$$

e) La traza de la matriz M es igual al número de grados de libertad del modelo: $\text{tr}[M] = N - k$.

Comprobación de la propiedad de la traza

Podemos comprobar que la traza de M coincide con los grados de libertad del modelo:

$$\begin{aligned} \text{tr}[M] &= \text{tr}[I_N - X(X'X)^{-1}X'] = \text{tr}[I_N] - \text{tr}[X(X'X)^{-1}X'] = \text{tr}[I_N] - \text{tr}[X'X(X'X)^{-1}] = \\ &= \text{tr}[I_N] - \text{tr}[I_k] = N - k. \end{aligned}$$

Para obtener este resultado hay que tener presente la propiedad del operador traza, que permite que los elementos de un producto matricial se puedan permutar circularmente sin que el resultado de la traza se altere.

f) Se trata de una matriz singular, es decir, su determinante es igual a cero y, por tanto, su inversa no está definida.

2) También podemos expresar el vector de residuos MCO como una combinación lineal de los términos de perturbación. Si partimos de 2.29, podemos escribir:

$$\mathbf{e} = \mathbf{M}\mathbf{Y} = \mathbf{M}(\mathbf{X}\mathbf{B} + \mathbf{U}) = \mathbf{M}\mathbf{X}\mathbf{B} + \mathbf{M}\mathbf{U} = \mathbf{M}\mathbf{U}, \quad (2.30)$$

lo cual es cierto, porque la matriz \mathbf{M} es ortogonal a la matriz \mathbf{X} .

3) A partir de las dos propiedades anteriores de los residuos MCO, podemos deducir una tercera: el vector de residuos es ortogonal a la matriz \mathbf{X} , es decir, $\mathbf{X}'\mathbf{e} = \mathbf{e}'\mathbf{X} = \mathbf{0}_{k \times 1}$:

$$\mathbf{X}'\mathbf{e} = \mathbf{X}'\mathbf{M}\mathbf{Y} = \mathbf{X}'\mathbf{M}\mathbf{U} = \mathbf{0}_{k \times 1}, \quad (2.31)$$

lo cual es cierto, porque la matriz \mathbf{M} es ortogonal a la matriz \mathbf{X} . En forma matricial nos quedará del modo siguiente:

$$\mathbf{X}'\mathbf{e} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2N} \\ \vdots & \vdots & & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kN} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N e_i \\ \sum_{i=1}^N X_{2i}e_i \\ \vdots \\ \sum_{i=1}^N X_{ki}e_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (2.32)$$

4) La media muestral de los residuos es igual a cero, siempre que en el modelo haya un término independiente. Esto es cierto porque de la primera fila de la matriz que sale en 2.32 podemos escribir lo siguiente:

$$\sum_{i=1}^N e_i = 0 \Rightarrow \bar{e} = \frac{1}{N} \sum_{i=1}^N e_i = 0. \quad (2.33)$$

5) Puesto que el vector de residuos se puede expresar como una combinación lineal de los términos de perturbación ($\mathbf{e} = \mathbf{M}\mathbf{U}$), y puesto que el término de perturbación es la fuente de aleatoriedad del modelo, que se distribuye según una ley normal, entonces el vector de residuos también es una variable aleatoria que se distribuye según una ley normal. Para poder tener perfectamente definida la ley de probabilidad que rige el comportamiento de los residuos, hay que conocer su valor esperado y su matriz de varianzas y covarianzas, que definen la ley normal:

$$\mathbf{e} \sim N(\mathbf{E}[\mathbf{e}], \text{VAR}[\mathbf{e}]),$$

lo cual implica, de manera individual, lo siguiente:

$$e_i \sim N(\mathbf{E}[e_i], \text{VAR}[e_i]) \quad \forall i = 1, \dots, N.$$

Por lo tanto, a continuación calcularemos el valor esperado y la matriz de varianzas y covarianzas del vector de residuos MCO. Comenzaremos por el valor esperado:

$$E[\mathbf{e}] = E[\mathbf{MU}] = \mathbf{ME}[\mathbf{U}] = \mathbf{0}_{N \times 1}.$$

En lo que respecta a la matriz de varianzas y covarianzas, veamos en primer lugar qué forma tiene:

$$\text{VAR}[\mathbf{e}] = \begin{bmatrix} \text{VAR}[e_1] & \text{COV}[e_1, e_2] & \dots & \text{COV}[e_1, e_N] \\ \text{COV}[e_2, e_1] & \text{VAR}[e_2] & \dots & \text{COV}[e_2, e_N] \\ \vdots & \vdots & \ddots & \vdots \\ \text{COV}[e_N, e_1] & \text{COV}[e_N, e_2] & \dots & \text{VAR}[e_N] \end{bmatrix}.$$

Buscamos ahora cuál es concretamente la matriz de varianzas y covarianzas de los residuos:

$$\begin{aligned} E[\mathbf{ee}'] &= E[(\mathbf{MU})(\mathbf{MU})'] = E[\mathbf{MUU}'\mathbf{M}'] = \mathbf{M} E[\mathbf{UU}'] \mathbf{M}' = \\ &= \mathbf{M}\sigma_u^2 \mathbf{I}_N \mathbf{M}' = \sigma_u^2 \mathbf{M}\mathbf{M}' = \sigma_u^2 \mathbf{M}. \end{aligned} \tag{2.34}$$

Respecto a la distribución de los residuos, podemos resumir lo que hemos visto hasta ahora en este apartado de la manera siguiente:


$$\mathbf{e} \sim N[\mathbf{0}_N, \sigma_u^2 \mathbf{M}], \tag{2.35}$$

o en particular:

$$e_i \sim N[0, \sigma_u^2 m_{ii}] \quad \forall i = 1, \dots, N.$$

2.4.2. Estimación de la varianza del término de perturbación


Una vez que conocemos las propiedades de los residuos MCO, podemos obtener un estimador para la varianza del término de perturbación (σ_u^2), el único parámetro desconocido del MRLM que aún no hemos explicado cómo debe estimarse.

Antes de deducir la expresión del estimador, veremos los motivos por los cuales consideramos importante obtener una estimación del parámetro σ_u^2 . Son los siguientes: 

1) Es necesario obtener una estimación de este parámetro para poder estimar la matriz de varianzas y covarianzas de los estimadores MCO de \mathbf{B} . Efectivamente, el parámetro que analizamos ahora sale en la expresión $\text{VAR}[\hat{\mathbf{B}}] = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$, y, si no tuviésemos una estimación, sólo podríamos calcular $(\mathbf{X}'\mathbf{X})^{-1}$, por lo cual no conocemos $\text{VAR}[\hat{\mathbf{B}}]$.

Para obtener el valor esperado...

... del vector de residuos MCO hemos utilizado la hipótesis que establece que los elementos de la matriz \mathbf{X} son cantidades fijas; por lo tanto, los de la matriz \mathbf{M} también lo son.

 Recordad la expresión de la matriz de varianzas y covarianzas que hemos utilizado al introducir las propiedades de homoscedasticidad y de ausencia de autocorrelación en el subapartado 2.2.2 de este módulo didáctico.

Para llegar a la expresión 2.34...

... hemos supuesto que la matriz de varianzas y covarianzas del término de perturbación es escalar, que los elementos de la matriz \mathbf{X} son fijos y que, por tanto, también lo son los de la matriz \mathbf{M} . Además, también hemos empleado las propiedades de simetría y de idempotencia de la matriz \mathbf{M} .

Esfericidad de los residuos

En la expresión para la distribución de los residuos se comprueba el hecho de que los residuos sólo son esféricos si la matriz \mathbf{M} es igual a la identidad. Por lo tanto, si se utilizan los residuos para comprobar la esfericidad del término de perturbación, se incurrirá en un sesgo sistemático.

2) Para poder realizar los contrastes de hipótesis sobre los parámetros individuales o bien sobre combinaciones lineales de éstos, necesitaremos, como podremos ver más adelante, conocer la matriz de varianzas y covarianzas de los estimadores y, en consecuencia, necesitaremos conocer la varianza del término de perturbación.

El problema que se nos plantea es cómo obtener una estimación de σ_u^2 , ya que los términos de perturbación no son observables. Esto lo solucionaremos utilizando los residuos MCO. De hecho, ya hemos visto antes que éstos son los estimadores de los términos de perturbación.

Puesto que podemos expresar el vector de residuos como combinación lineal de los elementos del vector de términos de perturbación, $\mathbf{e} = \mathbf{MU}$, el valor esperado de la SCE es el siguiente:

$$E[SCE] = E[\mathbf{e}'\mathbf{e}] = E[(\mathbf{MU})'(\mathbf{MU})] = E[\mathbf{U}'\mathbf{M}'\mathbf{MU}] = E[\mathbf{U}'\mathbf{MU}]. \quad (2.36)$$

Ya que $\mathbf{U}'\mathbf{MU}$ es un escalar, puede escribirse que $E[\mathbf{U}'\mathbf{MU}] = E[\text{tr}[\mathbf{U}'\mathbf{MU}]]$. Adicionalmente, si permutamos circularmente los elementos del producto de vectores y matrices, obtenemos que $E[\text{tr}[\mathbf{U}'\mathbf{MU}]] = E[\text{tr}[\mathbf{U}\mathbf{U}'\mathbf{M}]] = E[\text{tr}[\mathbf{M}\mathbf{U}\mathbf{U}']]$. Por otra parte, el operador esperanza matemática tiene la propiedad de que el valor esperado de la traza de una matriz es igual a la traza del valor esperado de esta matriz. Por lo tanto, se cumple que $E[\text{tr}[\mathbf{M}\mathbf{U}\mathbf{U}']] = \text{tr}[E[\mathbf{M}\mathbf{U}\mathbf{U}']]$.

Finalmente, pues, la expresión 2.36 resulta:

$$\begin{aligned} E[\mathbf{U}'\mathbf{MU}] &= \text{tr}\{E[\mathbf{M}\mathbf{U}\mathbf{U}']\} = \text{tr}\{\mathbf{M}E[\mathbf{U}\mathbf{U}']\} = \text{tr}[\mathbf{M}\sigma_u^2\mathbf{I}_N] = \\ &= \sigma_u^2\text{tr}[\mathbf{M}\mathbf{I}_N] = \sigma_u^2\text{tr}[\mathbf{M}] = \sigma_u^2(N - k). \end{aligned} \quad (2.37)$$

A partir de este resultado, podemos definir el estimador sin sesgo de la varianza del término de perturbación de la manera siguiente:

$$\hat{\sigma}_u^2 = \frac{\mathbf{e}'\mathbf{e}}{N - k}. \quad (2.38)$$

En consecuencia, ya podemos estimar la matriz de varianzas y covarianzas de los estimadores MCO de \mathbf{B} :

Matriz de varianzas y covarianzas	Matriz de varianzas y covarianzas estimada
$\text{VAR}[\hat{\mathbf{B}}] = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$	$\hat{\text{VAR}}[\hat{\mathbf{B}}] = \hat{\sigma}_u^2(\mathbf{X}'\mathbf{X})^{-1}$ donde: $\hat{\sigma}_u^2 = \frac{\mathbf{e}'\mathbf{e}}{N - k}$

Podemos comprobar fácilmente...

... que $\hat{\sigma}_u^2$ es efectivamente un estimador de la varianza del término de perturbación sin sesgo si calculamos su valor esperado:

$$\begin{aligned} E[\hat{\sigma}_u^2] &= E\left[\frac{\mathbf{e}'\mathbf{e}}{N - k}\right] = \\ &= \frac{1}{N - k} E[\mathbf{e}'\mathbf{e}] = \\ &= \frac{1}{N - k} \sigma_u^2(N - k) = \sigma_u^2. \end{aligned}$$

En cuanto a las varianzas o la precisión con que estimamos los parámetros, nos encontramos ante una situación similar:

Varianzas	Varianzas estimadas
$\text{VAR}[\hat{\beta}_j] = \sigma_u^2 a_{jj} \quad \forall j = 1, \dots, k$	$\widehat{\text{VAR}}[\hat{\beta}_j] = \hat{\sigma}_u^2 a_{jj} \quad \forall j = 1, \dots, k$

Acabamos de ver que el estimador sin sesgo de la varianza del término de perturbación está determinado por $\hat{\sigma}_u^2 = \frac{\mathbf{e}'\mathbf{e}}{N - k}$. En la práctica, para obtenerlo necesitamos calcular previamente la *SCE* ($\mathbf{e}'\mathbf{e}$).

La aproximación más directa consiste en obtener los residuos para todas las observaciones, elevarlos al cuadrado y sumarlos, pero este cálculo es muy laborioso, así que ahora estudiaremos dos maneras alternativas más simples de calcular la *SCE*. La primera es muy útil para facilitar el cálculo, mientras que la conveniencia de la segunda se verá en los apartados posteriores.

Las dos alternativas que proponemos para calcular la *SCE* son las que presentamos a continuación:

- $SCE = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} \quad (2.39)$

- $SCE = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}} \quad (2.40)$

En primer lugar, mostraremos que $\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$, teniendo en cuenta que esta expresión es un escalar:

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) = (\mathbf{Y}' - \hat{\mathbf{B}}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) = \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} = \\ &= \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}. \end{aligned}$$

Asimismo, también se cumple que $\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$, ya que utilizando $\mathbf{X}'\mathbf{e} = \mathbf{0}_{k \times 1}$ se tiene lo siguiente:

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'(\mathbf{X}\hat{\mathbf{B}} + \mathbf{e}) = \\ &= \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}}. \end{aligned}$$

Ejemplo de estimación de la varianza del término de perturbación

Consideremos la estimación del modelo concreto ya conocido y calculemos el estimador de la varianza del término de perturbación ($\hat{\sigma}_u^2$) y la matriz de varianzas y covarianzas de los estimadores de los parámetros de la población.

Consultad el ejemplo de estimación de un modelo concreto propuesto en el subapartado 2.3.1 de este módulo didáctico.



- Cálculo del estimador de la varianza del término de perturbación:

$$Y'Y = [16 \ 18 \ 10 \ 13 \ 18 \ 38 \ 25 \ 41 \ 11 \ 21] \begin{bmatrix} 16 \\ 18 \\ 10 \\ 13 \\ 18 \\ 38 \\ 25 \\ 41 \\ 11 \\ 21 \end{bmatrix} = 5.485.$$

$$\begin{aligned} e'e &= Y'Y - \hat{B}'X'Y = 5.485 - [37,51 \ 2,62 \ -1,91] \begin{bmatrix} 211 \\ 1.102 \\ 2.821 \end{bmatrix} = \\ &= 5.485 - 5.408,48 = 76,52. \end{aligned}$$

- Cálculo de la matriz de varianzas de los estimadores de los parámetros de la población:

$$\begin{aligned} \widehat{\text{VAR}}[\hat{B}] &= \hat{\sigma}_u^2(X'X)^{-1} = 10,93 \begin{bmatrix} 3,1349 & -0,0732 & -0,1926 \\ -0,0732 & 0,0123 & 0,0016 \\ -0,1926 & 0,0016 & 0,0131 \end{bmatrix} = \\ &= \begin{bmatrix} 34,27 & -0,80 & -2,11 \\ -0,80 & 0,13 & 0,02 \\ -2,11 & 0,02 & 0,14 \end{bmatrix}. \end{aligned}$$

2.5. Estimación por máxima verosimilitud

Hasta ahora hemos visto el método de estimación de mínimos cuadrados ordinarios, que consiste en asignar los parámetros desconocidos de manera que la SCE sea lo menor posible (es decir, hemos minimizado $SCE = e'e = \sum_{i=1}^N e_i^2$), y que la única hipótesis necesaria para garantizar la existencia de estimadores sea que la matriz $X'X$ debe ser invertible. A continuación presentaremos el método de máxima verosimilitud (MV).

El **método de máxima verosimilitud** es un método de estimación que propone como estimador del parámetro el valor que maximiza la probabilidad de obtener las observaciones muestrales disponibles.

De ahora en adelante escribiremos **método de máxima verosimilitud** de modo abreviado con la sigla **MV**.

La diferencia con el método de estimación de MCO es que la estimación por máxima verosimilitud se basa en las hipótesis que se establecen para la distribución de las variables aleatorias que aparecen en el modelo.

El modelo que utilizamos es $Y = XB + U$, donde suponemos que el componente aleatorio es el término de perturbación. Suponemos que el término de perturbación se distribuye según una ley normal; por lo tanto, cada

uno de los términos de perturbación tendrá la **función de densidad** siguiente:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left\{-\frac{1}{2\sigma_u^2} (u_i)^2\right\} \quad \forall i = 1, \dots, N. \quad (2.41)$$

En la anterior expresión hemos tenido en cuenta que se cumplen las hipótesis básicas siguientes:

$$E[u_i] = 0 \quad \text{y} \quad \text{VAR}[u_i] = \sigma_u^2 \quad \forall i = 1, \dots, N.$$

En cuanto a la **función de densidad conjunta** del vector columna \mathbf{U} , si se tienen en cuenta las hipótesis básicas de homoscedasticidad y de ausencia de autocorrelación, se obtendrá la expresión siguiente:


$$\begin{aligned} f(\mathbf{U}) &= \prod_{i=1}^N f(u_i) = (2\pi\sigma_u^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma_u^2} \sum_{i=1}^N u_i^2\right\} = \\ &= (2\pi)^{-N/2} (\sigma_u^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma_u^2} \mathbf{U}'\mathbf{U}\right\}. \end{aligned} \quad (2.42)$$

Si el vector columna \mathbf{U} se distribuye según una ley normal multivariante de orden N , el vector columna de la variable endógena \mathbf{Y} también se distribuirá según una ley normal multivariante, ya que es una combinación lineal de los términos de perturbación ($\mathbf{Y} = \mathbf{XB} + \mathbf{U}$).

$$f^*(\mathbf{Y}) = f(\mathbf{U}) = (2\pi)^{-N/2} (\sigma_u^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma_u^2} \mathbf{U}'\mathbf{U}\right\}. \quad (2.43)$$

Para que la anterior función de densidad sea una **función de verosimilitud**, hay que expresar el vector \mathbf{U} en función del vector \mathbf{Y} , es decir, $\mathbf{U} = \mathbf{Y} - \mathbf{XB}$:

$$L(\mathbf{Y}; \mathbf{B}, \sigma_u^2) = (2\pi)^{-N/2} (\sigma_u^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma_u^2} (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})\right\}. \quad (2.44)$$

Como ya hemos comentado antes, el MV consiste en elegir como estimadores de \mathbf{B} y σ_u^2 a los valores que maximizan la función de verosimilitud que acaba de escribirse. Por lo tanto, habrá que calcular las derivadas parciales de la función de verosimilitud mencionada con respecto a \mathbf{B} y σ_u^2 e igualarlas a cero. Puesto que la función de verosimilitud $L(\mathbf{Y}; \mathbf{B}, \sigma_u^2)$ es complicada, y también lo es el cálculo de sus derivadas parciales, en la práctica lo que se hace es maximizar el logaritmo de la función de verosimilitud. 

$$\ln\{L(\mathbf{Y}; \mathbf{B}, \sigma_u^2)\} = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma_u^2) - \frac{1}{2\sigma_u^2} [(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})]. \quad (2.45)$$

Por lo tanto, para obtener los estimadores MV:

$$\begin{aligned} \frac{\partial \ln L}{\partial \mathbf{B}} &= -\frac{1}{2\hat{\sigma}_u^2} [-2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})] = 0, \\ \frac{\partial \ln L}{\partial \hat{\sigma}_u^2} &= -\frac{N}{2\hat{\sigma}_u^2} + \frac{1}{2\hat{\sigma}_u^4} [(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})] = 0. \end{aligned}$$

Tomar logaritmos neperianos...

... para hallar los valores que maximizan la función de verosimilitud es una transformación monótona que no altera los resultados de la maximización. Este procedimiento matemático es válido porque no estamos interesados en conocer el valor del máximo, sino en saber en que punto se alcanza.

Y aislando se obtienen los resultados siguientes:

$$\begin{aligned} (\mathbf{X}'\mathbf{X})\hat{\mathbf{B}} &= \mathbf{X}'\mathbf{Y} \Rightarrow \hat{\mathbf{B}}_{MV} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \\ \hat{\sigma}_{u_{MV}}^2 &= \frac{(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})}{N} = \frac{\mathbf{e}'\mathbf{e}}{N}. \end{aligned} \quad (2.46)$$

Como podemos ver:

a) El estimador máximo verosímil de \mathbf{B} coincide con el de MCO; por tanto, tendrá las mismas propiedades, es decir, será un estimador lineal, sin sesgo, óptimo y consistente, en términos de ECM.

b) Por el contrario, el estimador MV de σ_u^2 difiere del de MCO y, tal como veremos a continuación, es un estimador que tiene sesgo:

$$E[\hat{\sigma}_{u_{MV}}^2] = E\left[\frac{\mathbf{e}'\mathbf{e}}{N}\right] = \frac{1}{N} E[\mathbf{e}'\mathbf{e}] = \frac{1}{N} \sigma_u^2(N - k) = \frac{(N - k)}{N} \sigma_u^2 \neq \sigma_u^2. \quad (2.47)$$

El estimador MV es, sin embargo, asintóticamente no sesgado, ya que, a medida que aumenta el tamaño muestral, su sesgo tiende a cero:

$$\lim_{N \rightarrow \infty} \frac{(N - k)}{N} \sigma_u^2 = \sigma_u^2.$$

2.6. Medidas de la bondad del ajuste

Hasta ahora hemos visto una serie de expresiones que nos permiten, a partir de la información disponible, estimar los $k + 1$ parámetros de la población del modelo, es decir, los estimadores MCO y MV de \mathbf{B} y σ_u^2 , y las propiedades de estos estimadores en el caso de que se cumplan las hipótesis básicas del modelo. Sin embargo, de momento, aún no tenemos ninguna medida que nos permita calibrar la calidad del ajuste realizado.

Para definir medidas de bondad del ajuste, empezaremos a trabajar a partir de la SCE, que, como ya hemos visto antes, se puede expresar de cualquiera de las formas siguientes:

$$\mathbf{e}'\mathbf{e} = \sum_{i=1}^N e_i^2 = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}}. \quad (2.48)$$

Por lo que respecta a la identidad $\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}}$, decimos que la suma de cuadrados de los errores es igual a la suma de cuadrados de los valores de la variable endógena menos la suma de cuadrados de los valores estimados por el modelo de la variable endógena; es decir, $\sum_{i=1}^N e_i^2 = \sum_{i=1}^N Y_i^2 - \sum_{i=1}^N \hat{Y}_i^2$.

Consultad dos expresiones equivalentes de la SCE en el apartado 2.4.2 de este módulo didáctico.



Si aislamos la suma de cuadrados de los valores de la variable endógena, obtenemos:

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \mathbf{e}'\mathbf{e} \rightarrow \sum_{i=1}^N Y_i^2 = \sum_{i=1}^N \hat{Y}_i^2 + \sum_{i=1}^N e_i^2, \quad (2.49)$$

y restando a ambos miembros de la igualdad la cantidad $N\bar{Y}^2$, queda:

$$\mathbf{Y}'\mathbf{Y} - N\bar{Y}^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2 + \mathbf{e}'\mathbf{e}. \quad (2.50)$$


El primer miembro de la igualdad anterior se conoce con el nombre de **suma de cuadrados totales (SCT)** y no es otra cosa que la suma de cuadrados de las desviaciones de la variable endógena respecto a su media aritmética. Concretamente:

$$\begin{aligned} \sum_{i=1}^N (Y_i - \bar{Y})^2 &= \sum_{i=1}^N (Y_i^2 + \bar{Y}^2 - 2\bar{Y}Y_i) = \sum_{i=1}^N Y_i^2 + N\bar{Y}^2 - 2\bar{Y} \sum_{i=1}^N Y_i = \\ &= \mathbf{Y}'\mathbf{Y} + N\bar{Y}^2 - 2N\bar{Y}^2 = \mathbf{Y}'\mathbf{Y} - N\bar{Y}^2. \end{aligned}$$

Nota


En el desarrollo de la SCT hemos empleado la igualdad siguiente:

$$N\bar{Y} = \sum_{i=1}^N Y_i$$

Como podemos ver, la SCT corresponde exactamente al numerador de la varianza muestral de la variable endógena y, por tanto, la podemos emplear como medida de la variabilidad o dispersión que presenta la variable endógena. Puesto que nuestro objetivo fundamental consiste en explicar las fluctuaciones de la variable endógena, parece interesante que veamos en qué partes se puede descomponer la SCT. 

Respecto a los primeros sumandos del segundo miembro de la igualdad de 2.50, en el caso de que el modelo tenga término independiente podemos decir que se denomina **suma de cuadrados de la regresión (SCR)**. Se corresponde con la suma de las desviaciones de los valores estimados de la variable endógena respecto a su media aritmética:

$$\begin{aligned} \sum_{i=1}^N (\hat{Y}_i - \bar{\hat{Y}})^2 &= \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^N (\hat{Y}_i^2 + \bar{Y}^2 - 2\bar{Y}\hat{Y}_i) = \\ &= \sum_{i=1}^N \hat{Y}_i^2 + N\bar{Y}^2 - 2\bar{Y} \sum_{i=1}^N \hat{Y}_i = \\ &= \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + N\bar{Y}^2 - 2N\bar{Y}^2 = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2. \end{aligned}$$

Para obtener este resultado, hemos supuesto que $\sum_{i=1}^N \hat{Y}_i = \sum_{i=1}^N Y_i$, o, equivalentemente, $\bar{\hat{Y}} = \bar{Y}$. Esta condición sólo se garantiza cuando en el modelo hay término independiente, ya que en otro caso no se debe cumplir obligatoriamente $\sum_{i=1}^N e_i = 0$. 

Si volvemos a la interpretación de la SCR, vemos que incorpora la parte de las fluctuaciones de la variable endógena que el modelo es capaz de explicar.

Por lo tanto, cuando en el modelo haya término independiente, podremos escribir lo siguiente:

$$\begin{aligned}\sum_{i=1}^N (Y_i - \bar{Y})^2 &= \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^N e_i^2. \\ \mathbf{Y}'\mathbf{Y} - N\bar{Y}^2 &= \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2 + \mathbf{e}'\mathbf{e}. \\ SCT &= SCR + SCE.\end{aligned}\quad (2.51)$$

La expresión 2.51 nos dice que la variabilidad total de la variable endógena (SCT) puede descomponerse en dos partes: la parte que podemos explicar mediante el modelo especificado (SCR) y la parte que no podemos explicar (SCE). Es importante subrayar que esta identidad sólo es cierta si hay un término independiente en el modelo.

Dos maneras alternativas de expresar la SCR son las siguientes:

$$SCR = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2 = \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} - N\bar{Y}^2 = \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} - N\bar{Y}^2. \quad (2.52)$$

Dos maneras equivalentes de expresar la SCR

La equivalencia $\hat{\mathbf{Y}}'\hat{\mathbf{Y}} - N\bar{Y}^2 = \hat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{B}} - N\bar{Y}^2$ es evidente.

Para mostrar la segunda equivalencia, empezamos por una de las expresiones que hemos visto para la SQE , $\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$. Si en esta expresión restamos y sumamos la misma cantidad $N\bar{Y}^2$ al segundo miembro, obtenemos lo siguiente:

$$\mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} = \underbrace{\mathbf{Y}'\mathbf{Y} - N\bar{Y}^2}_{SCT} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} + N\bar{Y}^2 = \underbrace{\mathbf{Y}'\mathbf{Y} - N\bar{Y}^2}_{SCT} - \underbrace{(\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} - N\bar{Y}^2)}_{SCR},$$

y si reagrupamos los términos, vemos que el último sumando corresponde a la SCR .

La primera medida de bondad del ajuste que veremos es el **coeficiente de determinación R^2** (llamado *R cuadrado*). A partir de lo que hemos comentado anteriormente, la definiremos así:

$$R^2 = 1 - \frac{SCE}{SCT}. \quad (2.53)$$

Además, si en el modelo hay un término independiente, se cumple la identidad $SCT = SCR + SCE$, y será cierta la expresión alternativa siguiente:

$$R^2 = \frac{SCR}{SCT}.$$

El coeficiente de determinación indica qué proporción de la variabilidad total de la variable endógena queda explicada por la regresión, es decir, indica la parte de la variación total de la variable endógena que pueden captar las variables explicativas incluidas en el modelo.

Consultad la existencia de término independiente en el modelo en el subapartado 2.1 de este módulo didáctico.


El rango de variación de R^2 del modelo con término independiente es el siguiente:

$$0 \leq R^2 \leq 1.$$

La interpretación del coeficiente de determinación es inmediata; tan sólo hay que tener presente que, cuando la SCE vale cero, el coeficiente de determinación vale uno ($R^2 = 1$). Por lo tanto, cuanto más se aproxime el coeficiente a la unidad, mejor será el ajuste, y al revés.

Rango de variación de \bar{R}^2

Se cumple que $R^2 \leq 1$ porque la SCT y la SCE son cantidades positivas y, por tanto, su cociente también lo es. Además, $SCT > SCE$. Cuando hay término independiente, se cumple la identidad $SCT = SCR + SCE$ y, por tanto, el valor mínimo de R^2 es cero; si no hay término independiente, no puede afirmarse que $SCE \leq SCT$.

Una cuestión importante para la utilización práctica del coeficiente R^2 es que no se puede utilizar como criterio para seleccionar **modelos anidados**, ya que si lo usamos con esta finalidad, siempre elegiremos el modelo con más variables explicativas. Este hecho es cierto, porque el coeficiente R^2 aumenta a medida que se incorporan variables explicativas, ya que la SCE disminuye. 

Se considera que dos modelos están anidados si el conjunto de variables explicativas de uno de los modelos es un subconjunto del otro.

Por lo tanto, cuando queremos hacer comparaciones entre modelos que permanecen en el nido o, en general, entre modelos con un número diferente de grados de libertad, emplearemos otra medida, que llamaremos **coeficiente de determinación corregido**, \bar{R}^2 , el cual tiene en cuenta el número de regresores de cada modelo:

$$\bar{R}^2 = 1 - \frac{N-1}{N-k} (1 - R^2). \quad (2.54)$$

Ejemplo de cálculo de R^2 y \bar{R}^2

Podemos volver a nuestro modelo de ejemplo de siempre y calcular su R^2 y su \bar{R}^2 , pero antes calcularemos la SCT y la SCR :


$$SCT = Y'Y - N\bar{Y}^2 = Y'Y - N\left(\frac{\sum Y_i}{N}\right)^2 = 5.485 - 10\left(\frac{211}{10}\right)^2 = 1.032,90.$$

$$SCR = \hat{B}'X'Y - N\bar{Y}^2 = \hat{B}'X'Y - N\left(\frac{\sum Y_i}{N}\right)^2 = 5.408,48 - 10\left(\frac{211}{10}\right)^2 = 956,38.$$

O también $SCE = SCT - SCR = 1.032,90 - 956,38 = 76,52$.

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{76,52}{1.032,90} = 0,926 = 92,6\%.$$

$$\bar{R}^2 = 1 - \frac{N-1}{N-k} (1 - R^2) = 1 - \frac{10-1}{10-3} (1 - 0,926) = 0,905 = 90,5\%.$$

Consultad el ejemplo de estimación de un modelo concreto en el subapartado 2.3.1 de este módulo didáctico. 

2.7. Significación de los parámetros del modelo

En este apartado veremos dos aspectos relacionados con las estimaciones \hat{B} que nos ayudan tanto a estudiar las relaciones económicas que se producen entre las variables del modelo como a detectar si el modelo está bien especi-

ficado. Nos referimos a la significación económica y estadística de los parámetros del modelo.

2.7.1 Significación económica

Una vez que disponemos de una estimación de los parámetros de la población del modelo, tenemos que interpretarlos. Como sabemos, si el modelo está especificado en niveles, es decir, con los valores observados, el parámetro refleja el efecto que tiene, de media, una variación unitaria de la variable explicativa sobre la variable endógena:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_j X_{ji} + \dots + \beta_k X_{ki} + u_i \quad \forall i = 1, \dots, N.$$

$$\frac{\partial Y_i}{\partial X_{ji}} = \beta_j \quad \forall j = 2, \dots, k. \quad (2.55)$$

Si, por el contrario, el modelo se encuentra especificado en logaritmos neperianos, los parámetros se interpretan como una elasticidad, tal como se ve a continuación:


$$Y_i = X_{1i}^{\beta_1} X_{2i}^{\beta_2} \dots X_{ji}^{\beta_j} \dots X_{ki}^{\beta_k} \exp\{u_i\} \quad \forall i = 1, \dots, N.$$


$$\ln Y_i = \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \dots + \beta_j \ln X_{ji} + \dots + \beta_k \ln X_{ki} + u_i \quad \forall i = 1, \dots, N.$$

$$\frac{\partial \ln Y_i}{\partial \ln X_{ji}} = \frac{\partial Y_i}{\partial X_{ji}} \frac{X_{ji}}{Y_i} = \beta_j.$$

Por lo tanto, según qué tipo de especificación utilicemos, la interpretación de los parámetros puede variar. Hay que comprobar que las estimaciones obtenidas son coherentes respecto a la teoría económica en lo que concierne al signo y a la magnitud. Si no lo son, se tendrá que revisar la base de datos utilizada o el modelo.

En ciertas ocasiones, es interesante determinar cuál es el regresor que tiene más influencia sobre el comportamiento de la variable endógena. Si el modelo está especificado en niveles*, los parámetros dependerán de las unidades de medida de los regresores. Por lo tanto, identificar una variable como la más influyente sobre el comportamiento de la variable endógena en función de la magnitud del parámetro estimado puede ser erróneo, excepto si todas las variables están medidas por las mismas unidades de medida.

Cuando queremos saber cuál es el regresor con más influencia sobre la variable endógena y las variables no están expresadas en las mismas unidades de medida, o no es posible transformarlas adecuadamente, tendremos que calcular la contribución de cada una o bien calcular los coeficientes beta. La variable con más influencia es la que tiene una contribución mayor o un coeficiente beta más elevado. 

 Consultad la interpretación de los parámetros en la especificación del MRLM en el subapartado 2.1 de este módulo didáctico.

Ejemplo de coherencia

Supongamos un modelo explicativo de la demanda de un bien normal en el cual el signo esperado por el parámetro asociado al precio es negativo. En el caso de que obtuviésemos una estimación positiva de este parámetro, tendríamos que revisar la base de datos y el modelo utilizado.

* En la especificación en niveles, los valores observados se usan directamente sin sufrir ninguna transformación.

La contribución media de una variable X_j que tiene media \bar{X}_j , la definiremos del modo siguiente:

$$\text{Contribución media de } X_j = \hat{\beta}_j \bar{X}_j \quad (2.56)$$

El coeficiente beta asociado a una variable se define así:

$$\hat{\beta}_j^* = \hat{\beta}_j \frac{S_{X_j}}{S_Y} \quad \forall j = 1, \dots, k, \quad (2.57)$$

donde S_{X_j} y S_Y son las desviaciones típicas de X_j e Y , respectivamente.

Alternativa de obtención de los coeficientes beta

Un modo alternativo de obtener los coeficientes beta consiste en transformar previamente el modelo, dividiendo todas las variables por su desviación típica correspondiente, y estimar el modelo transformado.

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_j X_{ji} + \dots + \beta_k X_{ki} + u_i \quad \forall i = 1, \dots, N,$$

$$\frac{Y_i}{S_Y} = \beta_1^* \frac{X_{1i}}{S_{X_1}} + \beta_2^* \frac{X_{2i}}{S_{X_2}} + \beta_3^* \frac{X_{3i}}{S_{X_3}} + \dots + \beta_j^* \frac{X_{ji}}{S_{X_j}} + \dots + \beta_k^* \frac{X_{ki}}{S_{X_k}} + u_i^* \quad \forall i = 1, \dots, N.$$

De hecho, el cálculo de los coeficientes beta no es más que un cambio de escala de las variables del modelo. Este cambio es útil en casos aplicados.

2.7.2 Significación estadística

Aparte de comprobar si la magnitud y los signos de las estimaciones de los parámetros son coherentes, nos interesa analizar, mediante una serie de contrastes, la significación estadística individual y conjunta de los parámetros del modelo. Observad que siempre consideramos que tenemos un término independiente incluido en el modelo.

Concretamente, queremos contrastar las hipótesis siguientes:

Significación individual	Significación conjunta
$H_0 : \beta_j = 0$ $H_A : \beta_j \neq 0$	$H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$ $H_A : \text{No } H_0$

Si tras haber realizado el contraste tenemos que rechazar la hipótesis nula, esto significará que, en la población, el parámetro β_j es significativamente distinto de cero y, por tanto, la variable que se le asocia es estadísticamente relevante a la hora de explicar la evolución de la variable endógena. En caso contrario, el parámetro β_j no es significativamente distinto de cero y, en consecuencia, la variable explicativa que lo acompaña no tiene un efecto estadísticamente significativo.

Recordad los contenidos sobre los contrastes de hipótesis de la asignatura *Fundamentos de Estadística*.

Acerca de las hipótesis...


... que queremos contrastar señalamos dos aspectos:

a. Denotamos por H_0 la hipótesis nula y por H_A la hipótesis alternativa.

b. Notamos que en el contraste de significación conjunta no se especifica ninguna condición sobre el parámetro del término independiente.

En el contraste de significación conjunta, o contraste de significación global, valoramos conjuntamente la significación de los parámetros del modelo, excepto el asociado al término independiente. Por lo tanto, contrastamos la relevancia conjunta de todas las variables.

Consultad los estadísticos de prueba utilizados en los subapartados 3.1.2 y 3.1.3 de este módulo didáctico.

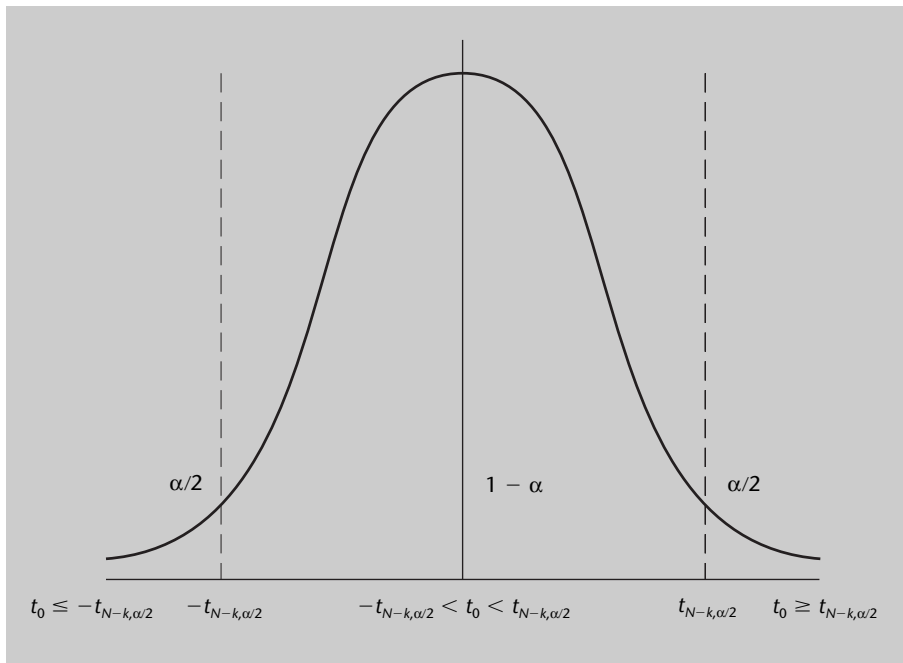
Los estadísticos de prueba que utilizaremos para realizar el contraste de las hipótesis anteriores sólo son casos particulares de los estadísticos que se formularán para contrastar restricciones lineales exactas de igualdad. 

Consideramos la realización de cada contraste de significación:

1) Para realizar el contraste de significación individual de un parámetro, utilizamos el estadístico de prueba siguiente:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}_u^2 a_{jj}}} \sim t_{N-k}, \quad (2.58)$$

donde $\sqrt{\hat{\sigma}_u^2 a_{jj}}$ es el error estimado de $\hat{\beta}_j$, y a_{jj} es el elemento j -ésimo de la diagonal principal de la matriz $(X'X)^{-1}$. Este estadístico se distribuye bajo la hipótesis nula según una t de Student con tantos grados de libertad como los que presenta el modelo, es decir, $N - k$ grados de libertad.



En el gráfico,...

... y también en la formulación de la hipótesis alternativa, se puede ver que siempre comparamos el valor absoluto del estadístico de prueba con el valor de las tablas $t_{N-k, \alpha/2}$, ya que se trata de un contraste estadístico de dos colas o bidireccional.

La regla de decisión que usaremos para establecer si un parámetro es significativamente distinto de cero o no es la siguiente:

a) Si $|t_0| \geq t_{N-k, \alpha/2}$, el estadístico de prueba se halla en la región crítica, y en ese caso rechazaremos la hipótesis nula. Por lo tanto, de acuerdo con la evidencia empírica, el parámetro β_j es significativamente distinto de cero.

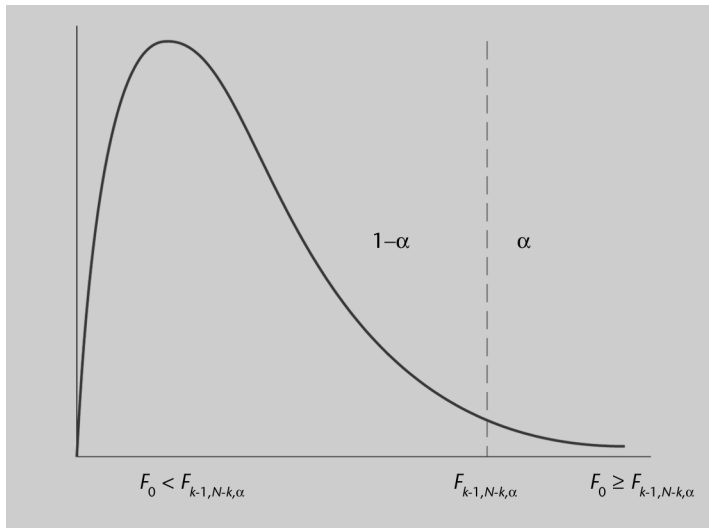
b) Si $|t_0| < t_{N-k,\alpha/2}$, el estadístico de prueba se halla en la región de no-rechazo de la hipótesis nula; en consecuencia, de acuerdo con la evidencia empírica, el parámetro de la población β_j es igual a cero.

2) Por otra parte, para realizar el contraste de significación global utilizaremos el estadístico de prueba siguiente:

$$F_0 = \frac{\frac{SCR}{k-1}}{\frac{SCE}{N-k}} \sim F_{k-1,N-k} \quad (2.59)$$

Este estadístico se distribuye bajo la hipótesis nula según una distribución F de Snedecor con $k-1$ grados de libertad en el numerador y $N-k$ grados de libertad en el denominador.

Consultad la distribución F de Snedecor en la asignatura *Fundamentos de Estadística*.



En este gráfico...

...comparamos el estadístico de prueba con el valor de las tablas $F_0 < F_{k-1,N-k,\alpha}$, pero no hay que tomar valor absoluto, ya que el estadístico es siempre positivo.

La regla de decisión utilizada para contrastar la significación global del modelo es la siguiente:

a) Si $F_0 \geq F_{k-1,N-k,\alpha}$, el estadístico de prueba se halla en la región crítica, y en ese caso rechazaremos la hipótesis nula. Por lo tanto, según la evidencia empírica, consideraremos que el modelo es globalmente significativo (es decir, que por lo menos uno de los parámetros es significativamente distinto de cero).

b) Por el contrario, si $F_0 < F_{k-1,N-k,\alpha}$, el estadístico de prueba se halla en la región de aceptación de la hipótesis nula y entonces no la rechazaremos. En consecuencia, según la evidencia empírica, podremos afirmar que el modelo no es globalmente significativo.

Teniendo en cuenta 2.53, una expresión alternativa del estadístico es la siguiente:

$$F_0 = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{N-k}} \quad (2.60)$$

De este resultado podemos deducir que no existirán contradicciones entre el contraste de significación global de parámetros y el coeficiente de determinación. Por lo tanto, si R^2 es próximo a la unidad, el test de significación global conducirá a rechazar la hipótesis nula.

Ejemplo de contrastes de significación individual y global

Una vez que hemos visto cómo hay que contrastar la significación individual y global de los parámetros de un MRLM, podemos aplicar estos contrastes a nuestro ejemplo. Utilizaremos algunos de los resultados que ya habíamos obtenido anteriormente.

Consultad el ejemplo de estimación de un modelo concreto y el de estimación de la varianza del término de perturbación en los subapartados 2.3.1. y 2.4.2 de este módulo didáctico.

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 37,51 \\ 2,62 \\ -1,91 \end{bmatrix}.$$

$$\text{VAR}[\hat{\mathbf{B}}] = \hat{\sigma}_v^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 34,27 & -0,80 & -2,11 \\ -0,80 & 0,13 & 0,02 \\ -2,11 & 0,02 & 0,14 \end{bmatrix}.$$

$$\widehat{\text{VAR}}[\hat{\beta}_1] = 34,27; \quad \widehat{\text{VAR}}[\hat{\beta}_2] = 0,13; \quad \widehat{\text{VAR}}[\hat{\beta}_3] = 0,14.$$

$$t_{\beta_1} = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{VAR}}[\hat{\beta}_1]}} = \frac{37,51}{\sqrt{34,27}} = 6,41 > t_{7,0,975} = 2,365 \quad \text{rechazamos } H_0.$$

$$t_{\beta_2} = \frac{\hat{\beta}_2}{\sqrt{\widehat{\text{VAR}}[\hat{\beta}_2]}} = \frac{2,62}{\sqrt{0,13}} = 7,16 > t_{7,0,975} = 2,365 \quad \text{rechazamos } H_0.$$

$$t_{\beta_3} = \frac{\hat{\beta}_3}{\sqrt{\widehat{\text{VAR}}[\hat{\beta}_3]}} = \frac{-1,91}{\sqrt{0,14}} = -5,10; \quad |-5,10| > t_{7,0,975} = 2,365 \quad \text{rechazamos } H_0.$$

$$F_0 = \frac{\frac{SCR}{k-1}}{\frac{SCE}{N-k}} = \frac{956,38}{3-1} = 43,75 > F_{2,7,0,95} = 4,74, \quad \text{rechazamos } H_0.$$

Todos estos resultados que hemos calculado los puede proporcionar de manera casi automática cualquier sistema informático de tratamiento estadístico y econométrico.

Resultados de la estimación				
Variable dependiente: Y				
Número de observaciones: 10				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	37.513983	5.853835	6.408	0.0004
X ₂	2.622278	0.366091	7.163	0.0002
X ₃	-1.913051	0.378387	-5.056	0.0015
R-squared	0.92592	Mean of dependent var		21.100
Adjusted R-squared	0.90476	S.D. of dependent var		10.713
S.E. of regression	0.30605	Sum of squared resid		76.520
		F-statistic		43.747
Durbin-Watson stat	2.24354	Prob(F-statistic)		0.0000

En la tabla os hemos señalado en negrita los resultados que ya conocíais.

En la primera parte del cuadro, bajo el título COEFICIENTES, se muestran las estimaciones MQO de los parámetros β_j ($\hat{\beta}_j$). A continuación, y debajo del título ERROR STD., podéis encontrar los errores estándar estimados para cada $\hat{\beta}_j$. En la columna siguiente, y bajo el título ESTADÍSTICO T, se presentan los valores de los estadísticos de prueba para realizar los contrastes de significación individual de cada parámetro, y en la última

columna, bajo el título SIGNIFICACIÓN, se encuentra la probabilidad de que una variable que se distribuye según una t de Student tome un valor superior, en valor absoluto, que el valor obtenido por el estadístico de prueba.


En la segunda parte, y bajo estas columnas, al lado de los títulos “R-squared”, “Adjusted R-squared” y “S.E. of regression”, están los valores del coeficiente de determinación, el coeficiente de determinación corregido y el estimador de la desviación estándar del término de perturbación, es decir, $\hat{\sigma}_u = \sqrt{\hat{\sigma}_u^2}$.

En la columna siguiente a esta y junto a los títulos “F-statistic” y “Prob(F-statistic)”, podéis encontrar el valor del estadístico de prueba para contrastar la significación global del modelo y la probabilidad que deja este valor a la derecha de la distribución que teóricamente sigue bajo la hipótesis nula. Finalmente, bajo el título “Sum of squared resid”, encontramos el valor de la SSE.

2.8. El modelo de regresión lineal múltiple en desviaciones respecto a la media

El modelo de regresión lineal múltiple en desviaciones respecto a la media no es otra cosa que una especificación particular del MRLM, tal como lo hemos visto hasta ahora, donde las variables se han transformado con un cambio de origen. Concretamente, las variables están expresadas en desviaciones respecto a sus valores medios, es decir:

$$\begin{aligned}\tilde{Y}_i &= Y_i - \bar{Y}. \\ \tilde{X}_{ji} &= X_{ji} - \bar{X}_j \quad \forall j = 1, \dots, k.\end{aligned}$$

La utilidad de esta especificación particular del MRLM radica tanto en la simplificación de los cálculos como en la de algunas demostraciones. Por otra parte, esta especificación es más compacta, y sobre todo muy útil, cuando en el MRLM el término independiente no tiene una interpretación de interés. 

El modelo en desviaciones se expresa de la manera siguiente:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{B}_2 + \mathbf{U},$$

donde:

$$\tilde{\mathbf{Y}} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_N - \bar{Y} \end{bmatrix}; \quad \tilde{\mathbf{X}} = \begin{bmatrix} X_{21} - \bar{X}_2 & X_{31} - \bar{X}_3 & \dots & X_{k1} - \bar{X}_k \\ X_{22} - \bar{X}_2 & X_{32} - \bar{X}_3 & \dots & X_{k2} - \bar{X}_k \\ \vdots & \vdots & & \vdots \\ X_{2N} - \bar{X}_2 & X_{3N} - \bar{X}_3 & \dots & X_{kN} - \bar{X}_k \end{bmatrix};$$

$$\mathbf{B}_2 = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}; \quad \mathbf{U} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Observad que en el modelo en desviaciones no hay término independiente. Por ello, en la matriz de las variables explicativas no aparece la columna de unos. Esta matriz tiene, por tanto, dimensión $N \times (k - 1)$.


Consultad la especificación del MRLM en el subapartado 2.1 de este módulo didáctico.

2.9. Predicción

Como ya hemos visto en el apartado 1.4, una vez que hemos especificado, estimado y validado un modelo, lo podemos utilizar con objetivos muy diferentes. Uno de ellos es la predicción.

Cuando trabajemos con datos de serie temporal, podemos estar interesados en predecir el comportamiento en el futuro de la variable endógena, y cuando trabajemos con datos de corte transversal, podemos estar interesados en predecir el comportamiento de un individuo o unidad que no se haya incluido en la muestra.

Cuando hacemos predicciones, podemos optar por intentar predecir el valor puntual que tomará la variable endógena o bien por determinar un intervalo de posibles valores entre los cuales esté el valor observado o esperado de esta variable. El primer caso se denomina **predicción puntual** y el segundo, **predicción por intervalo** del valor observado o del valor esperado.

En ambos casos, cuando queramos determinar el valor o cuando queramos encontrar un intervalo de valores posibles para la variable endógena de una observación fuera de la muestra, deberemos suponer que las hipótesis que hemos formulado sobre X , B y U se mantendrán también para las observaciones fuera de la muestra. Conviene resaltar que es fundamental suponer que se cumple la hipótesis de permanencia estructural del modelo. Si la hipótesis de permanencia estructural no se mantiene para las observaciones fuera de la muestra, en ese caso, como veremos, no tiene sentido formular una predicción. 

En primer lugar, veremos la predicción puntual y, a continuación, pasaremos a la predicción por intervalo.

2.9.1. Predicción puntual

Supongamos que la variable endógena ajustada para una determinada observación i es igual a:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki}. \quad (2.61)$$

Si queremos estimar el valor de la variable endógena, para una observación $N + h$, podemos usar la expresión siguiente:

$$\hat{Y}_{N+h} = \hat{\beta}_1 + \hat{\beta}_2 X_{2(N+h)} + \hat{\beta}_3 X_{3(N+h)} + \dots + \hat{\beta}_k X_{k(N+h)}, \quad (2.62)$$

Hay que tener en cuenta que...

... para poder hacer una predicción, se deben conocer los valores de las variables explicativas para la observación $N + h$.

y, utilizando la notación matricial, podemos escribir:

$$\hat{Y}_{N+h} = \mathbf{X}'_{N+h} \hat{\mathbf{B}} = [1 \quad X_{2(N+h)} \quad \dots \quad X_{k(N+h)}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}.$$

2.9.2. Predicción por intervalo

Podemos diferenciar la predicción por intervalo sobre Y_{N+h} de la predicción por intervalo sobre su valor esperado $E[Y_{N+h}]$. La diferencia es la siguiente:

1) En primer lugar, para obtener el intervalo del valor esperado de la variable endógena para la observación $N + h$, $E[Y_{N+h}]$, utilizaremos la expresión siguiente:

$$P\left(|E[Y_{N+h}] - \hat{Y}_{N+h}| < t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}}\right) = 1 - \alpha, \quad (2.63)$$

donde $t_{N-k, \alpha/2}$, es el valor de las tablas de una t de Student con $N - k$ grados de libertad. Lo que nos dice la expresión 2.63 es que, una vez que se ha elegido un nivel de significación igual a α , la probabilidad de que $E[Y_{N+h}]$ se encuentre en el intervalo delimitado por el extremo superior dado por $\hat{Y}_{N+h} + t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}}$ y por el extremo inferior dado por $\hat{Y}_{N+h} - t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}}$ es $(1 - \alpha)$.

2) En segundo lugar, para obtener la predicción por intervalo del valor observado de la variable endógena en la observación $N + h$, Y_{N+h} , utilizaremos la expresión siguiente:

$$P\left(|Y_{N+h} - \hat{Y}_{N+h}| < t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \{1 + \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}\}}\right) = 1 - \alpha. \quad (2.64)$$

Análogamente al caso anterior, con la expresión 2.64 decimos que, una vez que se ha elegido el nivel de significación igual a α , la probabilidad de que Y_{N+h} se halle en el intervalo de extremo superior igual a $\hat{Y}_{N+h} + t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \{1 + \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}\}}$ y de extremo inferior igual a $\hat{Y}_{N+h} - t_{N-k, \alpha/2} \sqrt{\hat{\sigma}_u^2 \{1 + \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}\}}$ es $1 - \alpha$.

Ejemplo de predicción puntual y de predicción por intervalo

En nuestro ejemplo, si dispusiésemos de los valores que tiene el gasto publicitario (X_2) y el precio medio de los productos (X_3) para una empresa adicional, podríamos calcular sus ventas previstas.

X_{2N+h}	X_{3N+h}
12	15

$$\mathbf{X}_{N+h} = \begin{bmatrix} 1 \\ 12 \\ 15 \end{bmatrix}.$$

Consultad el ejemplo de estimación de un modelo concreto en el subapartado 2.3.1 de este módulo didáctico.

Podemos estar interesados en disponer de una estimación puntual de sus ventas o de un intervalo de posibles valores entre los cuales estuviesen comprendidas las ventas de la empresa o de su valor esperado. Pues bien, para ello únicamente debemos aplicar las expresiones 2.62, 2.63 y 2.64. A continuación hacemos los tres cálculos:

1) En cuanto a la predicción puntual, usaremos la expresión 2.62:

$$\hat{Y}_{N+h} = \mathbf{X}'_{N+h} \hat{\mathbf{B}} = [1 \ 12 \ 15] \begin{bmatrix} 37,51 \\ 2,62 \\ -1,91 \end{bmatrix} = 40,29.$$

2) Con referencia al intervalo por el valor esperado de las ventas, utilizaremos 2.63. Para calcular sus extremos, tan sólo debemos determinar el nivel de significación, que en este caso fijaremos en el 5%. Entonces:

• El extremo superior valdrá:

$$\begin{aligned} & \hat{Y}_{N+h} + t_{7;0,975} \sqrt{\hat{\sigma}_u^2 \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}} = \\ & = 40,29 + 2,365 \sqrt{10,93 [1 \ 12 \ 15] \begin{bmatrix} 3,1349 & -0,0732 & -0,1926 \\ -0,0732 & 0,0123 & 0,0016 \\ -0,1926 & 0,0016 & 0,0131 \end{bmatrix} \begin{bmatrix} 1 \\ 12 \\ 15 \end{bmatrix}} = 47,68. \end{aligned}$$

• El extremo inferior valdrá:

$$\begin{aligned} & \hat{Y}_{N+h} - t_{7;0,975} \sqrt{\hat{\sigma}_u^2 \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}} = \\ & = 40,29 - 2,365 \sqrt{10,93 [1 \ 12 \ 15] \begin{bmatrix} 3,1349 & -0,0732 & -0,1926 \\ -0,0732 & 0,0123 & 0,0016 \\ -0,1926 & 0,0016 & 0,0131 \end{bmatrix} \begin{bmatrix} 1 \\ 12 \\ 15 \end{bmatrix}} = 32,89. \end{aligned}$$

3) Para acabar, con relación al intervalo del valor observado de las ventas, aplicaremos la expresión 2.64. Entonces:

El extremo superior será:

$$\begin{aligned} & \hat{Y}_{N+h} + t_{7;0,975} \sqrt{\hat{\sigma}_u^2 \{1 + \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}\}} = \\ & = 40,29 + 2,365 \sqrt{10,93 \left\{ 1 + [1 \ 12 \ 15] \begin{bmatrix} 3,1349 & -0,0732 & -0,1926 \\ -0,0732 & 0,0123 & 0,0016 \\ -0,1926 & 0,0016 & 0,0131 \end{bmatrix} \begin{bmatrix} 1 \\ 12 \\ 15 \end{bmatrix} \right\}} = 51,05. \end{aligned}$$

• El extremo inferior será:

$$\begin{aligned} & \hat{Y}_{N+h} - t_{7;0,975} \sqrt{\hat{\sigma}_u^2 \{1 + \mathbf{X}'_{N+h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_{N+h}\}} = \\ & = 40,29 - 2,365 \sqrt{10,93 \left\{ 1 + [1 \ 12 \ 15] \begin{bmatrix} 3,1349 & -0,0732 & -0,1926 \\ -0,0732 & 0,0123 & 0,0016 \\ -0,1926 & 0,0016 & 0,0131 \end{bmatrix} \begin{bmatrix} 1 \\ 12 \\ 15 \end{bmatrix} \right\}} = 29,52. \end{aligned}$$

Al examinar los extremos de ambos intervalos podemos ver que la amplitud del intervalo asociado al valor observado es mucho mayor que la del intervalo asociado al valor esperado. Por lo tanto, tendremos más precisión sobre el valor esperado que sobre el valor observado cuando queramos determinar un intervalo de posibles valores entre los cuales se encuentren comprendidos.

3. El modelo de regresión con restricciones lineales

En este apartado entenderemos cómo deben contrastarse ciertas hipótesis sobre los parámetros de un modelo. Nos centraremos únicamente en las hipótesis más habituales, que son las restricciones lineales. Adicionalmente, analizaremos cómo estimar de una manera más eficiente los modelos en los cuales haya estas restricciones. Para acabar, estudiaremos un caso particular de restricciones sobre los parámetros, como aquél en que se contrasta una de las hipótesis básicas del MRLM estándar: la permanencia estructural.


3.1. Contrastación de restricciones lineales

Tal como hemos estudiado en un apartado anterior, cuando queremos realizar un análisis estructural del comportamiento de una determinada variable, especificamos un modelo de regresión donde esta variable está causada (determinada) por un conjunto de variables explicativas y , a partir de la información muestral disponible sobre las variables observables del modelo, estimamos cuál es el efecto de estas variables explicativas sobre la variable endógena.

Consultad la especificación del MRLM estándar en el apartado 2 de este módulo didáctico.

Además, también podemos contrastar si el parámetro que acompaña a una determinada variable explicativa o si el modelo globalmente considerado son o no son estadísticamente significativos mediante los estadísticos correspondientes que hemos estudiado en el apartado anterior.

Consultad los estadísticos de prueba en el subapartado 2.7. de este módulo didáctico.

No obstante, es posible que estemos interesados en contrastar determinadas hipótesis alternativas que la teoría económica postula y que se manifiestan en restricciones que deberían cumplir los parámetros del modelo. 

Ejemplos de restricciones en los parámetros del modelo

Si para un determinado sector industrial especificamos una función de producción de tipo Cobb-Douglas:

$$\ln Q_i = \ln A + \beta_1 \ln K_i + \beta_2 \ln L_i + u_i \quad i = 1, \dots, N, \quad (3.1)$$


podemos estar interesados en saber si en este sector industrial hay rendimientos crecientes, decrecientes o constantes a escala.

Estas tres situaciones están determinadas por la condición siguiente: que la suma de la elasticidad producción-capital (β_1) y la elasticidad producción-trabajo (β_2) sea mayor, menor o igual a la unidad, respectivamente. Por lo tanto, una vez que se ha estimado el modelo 3.1, es interesante contrastar la hipótesis sobre la suma de estas elasticidades.

Todos los tipos de contrastes en los cuales se estudia el comportamiento de determinadas combinaciones lineales entre los parámetros del modelo se conocen como **contrastos de restricciones lineales**.

Así, generalizando, supongamos que hemos especificado el modelo de regresión siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, \dots, N; \quad (3.2)$$

en este caso podemos estar interesados, una vez estimado el modelo, en contrastar tres tipos de restricciones: 

a) Las **restricciones lineales exactas de igualdad**. Son aquellas restricciones en las que una determinada combinación lineal entre varios parámetros del modelo de regresión es igual a una constante:

$$a\beta_j + b\beta_l = c.$$

Si c es cero, se trata de una **restricción lineal de igualdad homogénea** (por ejemplo, $\beta_3 + \beta_4 + 2\beta_5 = 0$). En cambio, si c es una constante cualquiera, distinta de cero, se trata de una **restricción lineal de igualdad no homogénea** (por ejemplo, $3\beta_2 + \beta_3 + 4\beta_5 = 4$).

Ejemplo de restricciones lineales exactas de igualdad

En la función de producción de Cobb-Douglas 3.1, el hecho de contrastar si hay rendimientos constantes a escala es una restricción lineal no homogénea, ya que se especificaría del modo siguiente: $\beta_1 + \beta_2 = 1$.

b) Las **restricciones lineales exactas de desigualdad**. En este caso, se quiere estudiar si una combinación lineal concreta entre los parámetros del modelo de regresión cumple una determinada desigualdad.

Ejemplo de restricciones lineales exactas de igualdad

Un contraste de este tipo es la restricción lineal $0 < \beta_j < 1$. Otro ejemplo sería contrastar, en el modelo de la función de producción Cobb-Douglas 3.1, si hay rendimientos crecientes o decrecientes a escala, que se especificarían de la manera $\beta_1 + \beta_2 > 1$ o $\beta_1 + \beta_2 < 1$, respectivamente.

c) Las **restricciones estocásticas**. En este caso, a diferencia de los anteriores, se supone que la relación que existe entre los parámetros del modelo está sujeta a un error aleatorio.

De estos tres tipos de restricciones lineales que acabamos de presentar, únicamente estudiaremos los primeros, es decir, obtendremos un procedimiento que permita contrastar hipótesis que postulan que una determinada combinación lineal entre los parámetros del modelo es igual a una constante, tanto si es cero como diferente de cero.

3.1.1. Formulación matricial de las restricciones lineales

Antes de estudiar cómo se contrastan las restricciones lineales, conviene saber expresar matricialmente cualquier restricción lineal de igualdad entre los

Los contrastes de significación individual...

... de un parámetro (del tipo $\beta_j = 0$) y de significación global de la regresión (del tipo $\beta_2 = \beta_3 = \dots = \beta_k = 0$), que ya hemos estudiado, no son más que casos particulares de los contrastes de restricciones lineales de este tipo, como veremos más adelante.

Ejemplo de restricciones estocásticas

Encontramos este tipo de restricciones, por ejemplo, en un caso como el siguiente:
 $a\beta_j + b\beta_l = c + v$,
 donde v representa el error aleatorio al que están sujetas las relaciones entre los parámetros del modelo.

parámetros de la población del modelo. Para hacerlo, definimos una matriz y un vector:

1) La matriz \mathbf{R} , de dimensión $q \times k$, donde q representa el número de restricciones lineales que queremos contrastar y k , como es habitual, el número de parámetros del modelo. En consecuencia, la matriz \mathbf{R} tendrá tantas columnas como parámetros tenga el modelo y tantas filas como restricciones lineales contenga la hipótesis nula que contrastemos.

Cada fila de la matriz \mathbf{R} está formada por los coeficientes fijos asociados a todos los parámetros del vector \mathbf{B} en cada una de las restricciones. Por lo tanto, $E[\mathbf{R}] = \mathbf{R}$.

2) El vector \mathbf{r} , de dimensión q , que está formado por los términos independientes de las restricciones lineales. Por lo tanto, del mismo modo que la matriz \mathbf{R} , el vector \mathbf{r} tendrá tantas filas como restricciones. Los elementos de este vector siempre serán fijos, por lo cual $E[\mathbf{r}] = \mathbf{r}$.

Adicionalmente, se debe cumplir que $q \leq k$ y también que $\rho(\mathbf{R}) = q$, es decir que la matriz \mathbf{R} tenga rango máximo. Este segundo supuesto garantiza que las q restricciones lineales sean linealmente independientes entre sí.

Una vez definidos la matriz \mathbf{R} y el vector \mathbf{r} , para escribir matricialmente una restricción lineal o un conjunto de ellas, tendremos que emplear la expresión $\mathbf{RB} = \mathbf{r}$.

Ejemplo de formulación de la matriz \mathbf{R} y el vector \mathbf{r}

Supongamos que en el modelo de regresión $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$, $i = 1, \dots, N$ queremos contrastar la restricción lineal $3\beta_2 - 2\beta_3 = 4$.

En este caso, \mathbf{R} y \mathbf{r} están definidos por $\mathbf{R} = (0 \ 3 \ 0 \ 0 \ -2)$ y $\mathbf{r} = (4)$. La dimensión de la matriz \mathbf{R} es 1×5 y la de \mathbf{r} es 1, dado que contrastamos una única restricción lineal y el modelo tiene cinco parámetros.

Adicionalmente, podemos comprobar que, haciendo $\mathbf{RB} = \mathbf{r}$, obtenemos exactamente la restricción lineal formulada:

$$\mathbf{RB} = \mathbf{r} \Rightarrow 0\beta_1 + 3\beta_2 + 0\beta_3 + 0\beta_4 - 2\beta_5 = 4 \Rightarrow 3\beta_2 - 2\beta_3 = 4.$$

Supongamos ahora que en el mismo modelo se desea contrastar simultáneamente las tres restricciones lineales siguientes: $\beta_2 = 0$, $\beta_3 - \beta_4 + \beta_5 = -1$ y $\beta_5 = -3,1$

En este caso, \mathbf{R} y \mathbf{r} son:

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} 0 \\ -1 \\ -3,1 \end{bmatrix},$$

donde hay tres filas, dado que contrastamos tres restricciones lineales simultáneamente.

Restricciones linealmente independientes

Garantizar que las restricciones lineales son linealmente independientes es garantizar que en el conjunto de restricciones lineales que se quieren contrastar no puede haber dos restricciones como por ejemplo $\beta_2 - 2\beta_3 = 4$ y $3\beta_2 - 6\beta_3 = 12$, donde la segunda restricción es igual a la primera multiplicada por tres.

3.1.2. Metodología para contrastar restricciones lineales: estadístico de prueba

Una vez que conocemos cómo hay que expresar las restricciones lineales en el contexto de un MRLM, debemos estudiar cómo se hacen los contrastes de

hipótesis sobre los parámetros del modelo. Seguiremos los pasos siguientes:

1) Como hipótesis nula, se plantea que las restricciones lineales son ciertas y, como alternativa, que los parámetros de la población no satisfacen estas restricciones:


$$H_0: \mathbf{RB} = \mathbf{r} \quad H_A: \mathbf{RB} \neq \mathbf{r}.$$

2) Una vez definida la hipótesis nula, el paso siguiente es encontrar un estadístico de prueba que permita rechazarla o no rechazarla. Para hacerlo, partimos del MRLM sobre cuyos parámetros deseamos contrastar alguna o algunas restricciones lineales:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U},$$

y lo estimamos por MCO. Se puede demostrar que el estadístico de prueba para contrastar las restricciones lineales exactas de igualdad sobre los parámetros es el siguiente:

$$F_0 = \frac{(\hat{\mathbf{RB}} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} (\hat{\mathbf{RB}} - \mathbf{r})}{\frac{\mathbf{e}'\mathbf{e}}{N - k}} \sim F_{q, N-k}. \quad (3.3)$$

A partir del valor numérico que tome el estadístico de prueba es posible determinar si la diferencia entre $\hat{\mathbf{RB}}$ y \mathbf{r} es estadísticamente significativa o no lo es. En este sentido, la regla de decisión es la siguiente: 

- Si el valor del estadístico de prueba F_0 es mayor que el valor crítico de la F de Snedecor con q y $N - k$ grados de libertad para un nivel de significación, α , prefijado, se rechaza la hipótesis nula. Por lo tanto, las restricciones lineales no son ciertas en el ámbito de la población.
- Por el contrario, si F_0 es inferior al valor crítico de tablas, no se puede rechazar la hipótesis nula, lo cual significa que las restricciones lineales son ciertas en el ámbito de la población.

Ejemplo de contraste de significación individual de un parámetro

Teniendo en cuenta el MRLM siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, \dots, N, \quad (3.4)$$


queremos contrastar la hipótesis nula $H_0: \beta_j = 0$. El objetivo de este ejemplo es aplicar la metodología estudiada para llevar a cabo el contraste citado. Como ya hemos adelantado en otro apartado, el estadístico de prueba para este modelo está determinado por:

$$t_{\beta_j} = \frac{\hat{\beta}_j - 0}{\sqrt{\widehat{\text{VAR}}[\hat{\beta}_j]}} \sim t_{N-k, \alpha/2}.$$

Lectura complementaria

La demostración sobre el estadístico de prueba utilizado para contrastar las restricciones lineales exactas de igualdad sobre los parámetros se puede encontrar, por ejemplo, en la obra siguiente:

J. Johnston (1987). *Métodos de econometría* (trad. J. Sánchez Fernández). Barcelona: Vicens-Vives.

 Consultad el estadístico de prueba utilizado en el contraste de significación individual 2.58 de un parámetro del modelo en el subapartado 2.7.2. de este módulo didáctico.

En este ejemplo veremos que también podemos llegar a este estadístico trabajando con la expresión 3.3, obtenida para cualquier contraste de restricciones lineales. La matriz \mathbf{R} y el vector \mathbf{r} que permitirán reproducir la hipótesis nula que queremos contrastar son los siguientes:

$$\mathbf{R} = (0 \dots 0 \overset{j}{\mathbf{1}} 0 \dots 0) \quad \mathbf{r} = (0).$$

Los componentes del estadístico de prueba 3.3 para el caso analizado son los siguientes:

- q , que representa el número de restricciones lineales que contrastamos; en este caso es 1.

$$\bullet \quad \mathbf{R}\hat{\mathbf{B}} - \mathbf{r} = [0 \dots 0 \ 1 \ 0 \dots 0] \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{j-1} \\ \hat{\beta}_j \\ \hat{\beta}_{j+1} \\ \vdots \\ \hat{\beta}_k \end{bmatrix} - 0 = \hat{\beta}_j.$$

- En $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ será:

$$\begin{aligned} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' &= [0 \dots 0 \ 1 \ 0 \dots 0] \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1k} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & a_{jj} & \dots & a_{jk} \\ \vdots & & \vdots & & \vdots \\ a_{k1} & \dots & a_{kj} & \dots & a_{kk} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \\ &= [0 \dots 0 \ 1 \ 0 \dots 0] \begin{bmatrix} a_{1j} \\ \vdots \\ a_{jj} \\ \vdots \\ a_{kj} \end{bmatrix} = a_{jj}. \end{aligned}$$

En consecuencia, el estadístico de prueba 3.3 queda como se ve a continuación:

$$F_0 = \frac{\frac{\hat{\beta}_j(a_{jj})^{-1}\hat{\beta}_j}{1}}{\frac{\mathbf{e}'\mathbf{e}}{N-k}} = \frac{\hat{\beta}_j^2}{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}} \sim F_{1, N-k}. \quad (3.5)$$

Por otra parte, como sabemos, una F de Snedecor con un grado de libertad en el numerador y m grados de libertad en el denominador equivale al cuadrado

Consultad el estadístico de prueba utilizado para realizar el contraste de significación individual de un parámetro del modelo en el subapartado 2.7.2 de este módulo didáctico.

de una t de Student con m grados de libertad. Por lo tanto, llegamos al estadístico 2.58, que hemos visto en el apartado anterior:

$$F_0 = \frac{\hat{\beta}_j^2}{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}} = \left(\frac{\hat{\beta}_j}{\sqrt{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}}} \right)^2 \sim t_{N-k}^2 \Rightarrow \frac{\hat{\beta}_j}{\sqrt{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}}} \sim t_{N-k}.$$

Actividad

3.1. Demostrad que, si la hipótesis nula que queremos contrastar en el modelo que acabamos de ver es $H_0: \beta_j = \beta_j^0$, utilizando el estadístico 3.3 se llega también al estadístico siguiente:

$$F_0 = \frac{(\hat{\beta}_j - \beta_j^0)^2}{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}} = \left(\frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}}} \right)^2 \sim t_{N-k}^2 \Rightarrow \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\frac{\mathbf{e}'\mathbf{e}}{N-k} a_{jj}}} \sim t_{N-k}.$$

3.1.3. Un método alternativo para contrastar restricciones lineales

Un método alternativo al anterior para contrastar restricciones lineales de este tipo consiste en comparar la suma de cuadrados de los errores del modelo restringido (SCE_R o bien $\mathbf{e}'_R \mathbf{e}_R$) con la del modelo no restringido o ampliado (SCE_A o $\mathbf{e}'_A \mathbf{e}_A$).

La expresión del estadístico de prueba en este caso es la siguiente:

$$F_0 = \frac{\frac{SCE_R - SCE_A}{s}}{\frac{SCE_A}{N-k}} = \frac{SCE_R - SCE_A}{\frac{s}{\hat{\sigma}_u^2}}, \quad (3.6)$$

donde s representa la diferencia entre los grados de libertad del modelo restringido y los grados de libertad del modelo ampliado, y coincide con el número de restricciones lineales que contrastamos (q):

$$s = [N - (k - q)] - [N - k] = q.$$

El **modelo restringido** es el que incorpora la hipótesis nula, es decir, el modelo resultante de considerar la restricción lineal como cierta, mientras que el **modelo no restringido** es el modelo general que no incorpora la restricción lineal (la hipótesis nula).

Consideramos los dos contrastes de significación del modelo por separado: 

1) Contraste de significación individual de un parámetro del modelo

Es posible llevar a cabo este contraste a partir del estadístico 3.6. Por eso suponemos el modelo de regresión siguiente, que llamaremos *modelo ampliado*:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, \dots, N. \quad (3.7)$$

Si incorporamos la hipótesis nula $H_0: \beta_j = 0$, el modelo que incorpora la restricción, que llamamos *modelo restringido*, es:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_{j-1} X_{(j-1)i} + \beta_{j+1} X_{(j+1)i} + \dots + \beta_k X_{ki} + v_i \quad i = 1, \dots, N. \quad (3.8)$$

Una vez especificados los modelos restringido y ampliado, se calculan las sumas de cuadrados de los errores de ambos modelos, SCE_R y SCE_A , y se construye el estadístico de prueba teniendo en cuenta los grados de libertad adecuados:

$$F_0 = \frac{\frac{SCE_R - SCE_A}{1}}{\frac{SCE_A}{N - k}} \sim F_{1, N-k},$$

donde s es igual a 1, dado que contrastamos una única restricción lineal. Observad que s coincide con la diferencia entre los grados de libertad de las dos χ^2 :

$$\frac{SCE_R}{\sigma_u^2} \sim \chi_{N-(k-1)}^2 \quad \frac{SCE_A}{\sigma_u^2} \sim \chi_{N-k}^2 \quad \frac{SCE_R - SCE_A}{\sigma_u^2} \sim \chi_{N-(k-1)-(N-k)}^2 = \chi_1^2.$$

2) Contraste de significación global del modelo

El contraste de significación conjunta se puede plantear también desde la óptica de un contraste de restricciones lineales a partir del estadístico de prueba 3.6. El modelo de regresión ampliado* es:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, \dots, N, \quad (3.9)$$

mientras que el modelo restringido, es decir, bajo la hipótesis nula $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$, es:

$$Y_i = \beta_1 + v_i \quad i = 1, \dots, N. \quad (3.10)$$

Si estimamos el modelo restringido por MCO, obtenemos el estimador por mínimos cuadrados restringidos (MCR) siguiente:

$$\hat{\mathbf{B}}_{\text{MCR}} = (\mathbf{X}_R' \mathbf{X}_R)^{-1} \mathbf{X}_R' \mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N Y_i \\ N \end{bmatrix} = [\bar{Y}];$$

$$\hat{\mathbf{B}}_{\text{MCR}} = \hat{\beta}_1 = [\bar{Y}],$$

Los términos de perturbación...

... de los modelos ampliado y restringido no son iguales porque los modelos son diferentes.

* El modelo de regresión ampliado es el modelo libre de restricciones.

Observad que el estimador resultante es un vector de dimensión 1, es decir, un escalar.

donde X_R es la matriz de variables explicativas asociada al modelo restringido. Entonces:

$$e_{i_r} = Y_i - \hat{Y}_{i_r} = Y_i - \hat{\beta}_{1_r} = Y_i - \bar{Y}. \tag{3.11}$$


Por lo tanto, la *SCE* del modelo restringido está determinada por:

$$SCE_R = \mathbf{e}_R' \mathbf{e}_R = \sum_{i=1}^N e_{i_r}^2 = \sum_{i=1}^N (Y_i - \bar{Y})^2 = SCT. \tag{3.12}$$

A partir del resultado anterior, y teniendo en cuenta que, puesto que en el modelo hay término independiente, es posible descomponer la variabilidad total de la variable endógena de la forma $SCT = SCE + SCR$, se puede comprobar que la *SCR* es cero.

SCT y SCR

Recordad que la variabilidad total de la variable endógena es la *SCT*, y que la *SCR* es la parte de la variabilidad de la variable endógena explicada por los regresores.

Otros resultados de interés asociados a este ejemplo de modelo restringido son los siguientes: 

$$\text{VAR}[\hat{\mathbf{B}}_{\text{MCR}}] = \text{VAR}[\hat{\beta}_{1\text{MCR}}] = \sigma_v^2 (\mathbf{X}_R' \mathbf{X}_R)^{-1} = \frac{\sigma_v^2}{N}$$


$$\hat{\sigma}_{vR}^2 = \frac{\mathbf{e}_R' \mathbf{e}_R}{N - k} = \frac{SCE_R}{N - k} = \frac{SCT}{N - k}, \text{ donde } k = 1;$$

$$R_R^2 = 1 - \frac{SCE_R}{SCT} = 1 - \frac{SCT}{SCT} = 0.$$

Por otra parte, si denominamos SCE_A y SCR_A a la suma de cuadrados de los errores y de la regresión (explicada) del modelo sin restringir (modelo 3.9), el estadístico de prueba 3.6, teniendo en cuenta que s representa el número de restricciones lineales consideradas en la hipótesis nula, queda como vemos a continuación:

$$F_0 = \frac{\frac{SCE_R - SCE_A}{s}}{\frac{SCE_A}{N - k}} = \frac{\frac{SCT - SCE_A}{k - 1}}{\frac{SCE_A}{N - k}} = \frac{\frac{SCR_A}{k - 1}}{\frac{SCE_A}{N - k}} \sim F_{k-1, N-k}. \tag{3.13}$$

y, como podemos ver, 3.13 es la misma expresión para el estadístico de prueba que la fórmula 2.59, definida en el apartado anterior.

 Consultad el estadístico de prueba definido para hacer el contraste de significación global en el subapartado 2.7.2 de este módulo didáctico.

Ejemplos de contrastes de significación individual y global

A continuación ilustraremos cómo debe llevarse a cabo el contraste de restricciones lineales empleando el estadístico 3.6.

Empezamos por hacer el contraste con el primer método que hemos estudiado. En el cuadro se presenta la estimación MCO del modelo siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \quad i = 1, \dots, N. \tag{3.14}$$

Resultados de la estimación del modelo 3.14

Variable dependiente: Y				
Número de observaciones: 100				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
Constant	2.2697394	0.3007269	7.5475106	0.0000
X ₂	4.4181150	0.3390604	13.030464	0.0000
X ₃	6.3137433	0.3458716	18.254588	0.0000
X ₄	9.9596302	0.3670705	27.132743	0.0000
R-squared	0.933400	Mean of dependent var	12.18422	
Adjusted R-squared	0.931318	S.D. of dependent var	3.749500	
S.E. of regression	0.982639	Sum of squared resid	92.69562	
Log likelihood	-138.1014	F-statistic	448.4770	
Durbin-Watson stat	1.693090	Prob(F-statistic)	0.000000	

A partir de los resultados anteriores, podemos observar lo siguiente:

1) Vemos que todos los coeficientes de las variables son significativos, puesto que el estadístico t_0 permite rechazar la hipótesis nula de no-significación individual en todos los casos: la probabilidad que dejan en los extremos los valores obtenidos por los estadísticos de prueba del contraste de significación individual es, en todos los casos, 0,0000, claramente inferior al nivel de significación habitual del 5%, lo cual significa que el estadístico de prueba se halla en la zona de rechazo de la hipótesis nula.

2) Del mismo modo podemos comprobar que el modelo es conjuntamente significativo, dado que el valor obtenido por el estadístico de prueba F_0 (448.4770) se encuentra en la zona de rechazo de la hipótesis nula, por el hecho de que la probabilidad que existe por encima del valor citado es inferior al nivel de significación habitual del 5%.

A continuación ejemplificamos cómo también podemos hacer este contraste a partir del estadístico 3.6. En el cuadro anterior se presenta la estimación del modelo ampliado, y en el cuadro siguiente, la del modelo restringido:

$$Y_i = C + v_i \quad i = 1, \dots, N, \quad (3.15)$$

donde $\beta_2 = \beta_3 = \beta_4 = 0$.

Resultados de la estimación del modelo 3.15

Variable dependiente: Y				
Número de observaciones: 100				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
Constant	12.184217	0.3749500	32.495581	0.0000
R-squared	0.000000	Mean of dependent var	12.18422	
Adjusted R-squared	0.000000	S.D. of dependent var	3.749500	
S.E. of regression	3.749500	Sum of squared resid	1391.816	
Log likelihood	-273.5536	Durbin-Watson stat	1.721521	

Algunos aspectos que hay que destacar del cuadro de los resultados de la estimación del modelo 3.15 son los siguientes:

- El valor estimado por el término independiente coincide con la media muestral de la variable endógena, lo cual significa que el modelo estimado sólo refleja el valor medio de la variable y no su variabilidad.
- La desviación estándar estimada del término de perturbación coincide numéricamente con la desviación estándar de la variable endógena. Este hecho pone de manifiesto nuevamente que el modelo no explica la variabilidad de la variable endógena.
- Como consecuencia de lo que hemos comentado en los párrafos anteriores, el valor del coeficiente de determinación, R^2 , es cero, puesto que la SCE coincide con la SCT y, por lo tanto, la SCR es nula.

Volviendo al contraste de significación global del modelo, el estadístico de prueba 3.6 calculado a partir de la SCE de los modelos ampliado (cuadro de resultados de la estimación del modelo 3.14) y restringido (cuadro de resultados de la estimación del modelo 3.15) tiene el valor siguiente:

$$F_0 = \frac{\frac{1.391,816 - 92,69562}{3}}{\frac{92,69562}{100 - 4}} = 448,477.$$

En consecuencia, dado que el valor en tablas de una F de Snedecor con 3 y 96 grados de libertad es (aproximadamente) 2,7, se rechaza la hipótesis nula, lo cual significa que el modelo es globalmente significativo.

Para acabar, supongamos que estamos interesados en estudiar si en el modelo 3.14 se cumple la restricción lineal siguiente:

$$H_0: 2\beta_2 = \beta_4. \quad (3.16)$$

Para contrastar la restricción lineal 3.16, en primer lugar debemos definir el modelo restringido correspondiente a este caso. Por eso tenemos que incorporar la restricción en el modelo 3.14:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + (2\beta_2) X_{4i} + v_i \quad i = 1, \dots, N.$$

Entonces, realizando las operaciones algebraicas correspondientes, se llega al modelo restringido siguiente:

$$Y_i = \beta_1 + \beta_2(X_{2i} + 2X_{4i}) + \beta_3 X_{3i} + u_i = \beta_1 + \beta_2 Z_i + \beta_3 X_{3i} + v_i \quad i = 1, \dots, N. \quad (3.17)$$

Estimando 3.17 por el método de mínimos cuadrados ordinarios, se obtienen los resultados presentados en el cuadro de resultados de la estimación del modelo 3.17 siguientes:

Resultados de la estimación del modelo 3.17				
Variable dependiente: Y				
Número de observaciones: 100				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	2.1606118	0.2927975	7.3792026	0.0000
Z	4.8509642	0.1614785	30.040937	0.0000
X ₃	6.3408917	0.3473184	18.256710	0.0000
R-squared	0.931942	Mean of dependent var	12.18422	
Adjusted R-squared	0.930539	S.D. of dependent var	3.749500	
S.E. of regression	0.988200	Sum. of squared resid	94.72422	
Log likelihood	-139.1838	F-statistic	664.1274	
Durbin-Watson stat	1.661722	Prob(F-statistic)	0.000000	

Calculamos el estadístico 3.6 a partir de las SCE de los modelos ampliado y restringido:

$$F_0 = \frac{\frac{94,72422 - 92,69562}{1}}{\frac{92,69562}{100 - 4}} = 2,1,$$

y, dado que el valor crítico de una F de Snedecor con 1 y 96 grados de libertad es, para un nivel de significación del 5%, 3,94, no se puede rechazar la hipótesis nula, y, por tanto, debemos considerar que la restricción lineal es cierta en el ámbito de la población.

3.2. Estimación restringida por mínimos cuadrados (MCR)

Cuando después de hacer un contraste sobre unas determinadas restricciones lineales se concluye que no existe ninguna evidencia para rechazar la hipótesis nula, es interesante estimar el modelo teniendo en cuenta estas restricciones.

El método de estimación que permite estimar el modelo con restricciones es conocido con el nombre de **mínimos cuadrados restringidos**.

Aquí presentamos este método de estimación y algunos resultados interesantes asociados a él.

3.2.1. El estimador de mínimos cuadrados restringidos

El método de estimación de MCR, a diferencia del método de MCO, tiene en consideración las restricciones lineales; consiste en minimizar la suma

Consultad las SCE del modelo ampliado y del modelo restringido en los cuadros de estimación de resultados 3.14 y 3.17, respectivamente, de este módulo didáctico.

Abreviaremos método de mínimos cuadrados restringidos con la sigla **MCR**.


de los errores al cuadrado (como los MCO), pero imponiendo las restricciones lineales:

$$\text{Minimizar } e'e \text{ sujeto a } \mathbf{RB} - \mathbf{r} = 0.$$


Fruto de esta minimización, podemos obtener el estimador MCR de los k parámetros β_j del modelo, que será el siguiente:

$$\hat{\mathbf{B}}_{\text{MCR}} = \hat{\mathbf{B}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{B}}). \quad (3.18)$$

Analizando 3.18, podemos afirmar que el estimador MCR no es más que una corrección del estimador MCO, puesto que se obtiene como la suma del estimador MCO más una expresión. El tamaño de esta corrección está determinado por la diferencia, mayor o menor, que exista entre la estimación de las restricciones por MCO y las restricciones lineales que se contrastan.

Así, en el caso de que el estimador MCO satisfaga las restricciones lineales, la discrepancia $\mathbf{r} - \mathbf{R}\hat{\mathbf{B}}$ es cero, de modo que los estimadores MCR coinciden con los estimadores MCO. 

3.2.2. Propiedades del estimador restringido

El estimador MCR presenta las tres propiedades siguientes: 

- Se puede demostrar que se trata de un estimador sin sesgo sólo si las restricciones lineales son ciertas. Además, el sesgo depende del desacierto de las restricciones lineales: cuanto mayor es la diferencia $\mathbf{r} - \mathbf{RB}$, mayor es el sesgo.
- Una segunda propiedad del estimador MCR es que siempre satisface las restricciones que imponamos en el proceso de la minimización de la SCE. Esto ocurre tanto si las restricciones son ciertas como si son falsas.
- La matriz de varianzas y covarianzas de los estimadores MCR, $\hat{\mathbf{B}}_{\text{MCR}}$, es la que presentamos a continuación:

$$\text{VAR}[\hat{\mathbf{B}}_{\text{MCR}}] = \text{VAR}[\hat{\mathbf{B}}] - \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}. \quad (3.19)$$

La expresión 3.19 que acabamos de ver es la matriz de varianzas y covarianzas de los estimadores MCR siempre que se cumplan las hipótesis básicas del modelo y, en particular, siempre que el término de perturbación tenga la matriz de varianzas y covarianzas igual a $\sigma_u^2\mathbf{I}_N$.


Recordad que la matriz de varianzas y covarianzas del estimador MCO está determinada por:

$$\text{VAR}[\hat{\mathbf{B}}] = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Lecturas complementarias

La demostración que lleva a la expresión 3.18 del estimador MCR de los k parámetros β_j se puede encontrar, por ejemplo, en las obras siguientes:

J. Johnston (1987). *Métodos de econometría* (trad. J. Sánchez Fernández). Barcelona: Vicens-Vives.
A. Novales (1993). *Econometría* (2a ed.). Madrid: McGraw-Hill.


 Consultad la matriz de varianzas y covarianzas del estimador MCO en el subapartado 2.3.2 de este módulo didáctico.

En la lectura recomendada que aparece en el margen se demuestra que la matriz:

$$\sigma_u^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

es semidefinida positiva; por lo tanto, los estimadores MCR presentan siempre varianzas menores, o iguales en el caso más extremo, que las varianzas de los estimadores mínimos cuadrados, independientemente de si las restricciones son ciertas o son falsas:

$$\text{VAR}[\hat{\beta}_{j\text{MCR}}] \leq \text{VAR}[\hat{\beta}_j] \quad j = 1, 2, \dots, k.$$

Teniendo en cuenta todo lo que hemos dicho anteriormente, la elección entre estimar por MCO o por MCR puede ser difícil (si las restricciones consideradas son erróneas). Si bien en principio, en términos de no-sesgo, conviene trabajar con los estimadores MCO –ya que estos estimadores garantizan que no haya sesgo, y en los MCR únicamente se garantiza si las restricciones impuestas son ciertas–, si se pretende trabajar en términos de ECM, puede ser mejor estimar por MCR, porque los estimadores presentan varianzas menores que los MCO. 

$$\text{ECM}[\hat{\mathbf{B}}] = [\text{sesgo}(\hat{\mathbf{B}})]^2 + \text{VAR}[\hat{\mathbf{B}}].$$


3.2.3. Un ejemplo aclaratorio

En este subapartado, presentamos un ejemplo sencillo y meramente académico con el objetivo de aclarar los diferentes conceptos que hemos tratado hasta ahora. Así, supongamos que para estimar el modelo siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad i = 1, \dots, N, \quad (3.20)$$

disponemos de cinco observaciones correspondientes a cada una de las variables observables implicadas en el modelo. Las observaciones son las siguientes:

$$X = \begin{bmatrix} 1 & 1 & 7 \\ 1 & 5 & 2 \\ 1 & 3 & 6 \\ 1 & 2 & 1 \\ 1 & 4 & 3 \end{bmatrix} \quad Y = \begin{bmatrix} -28 \\ 17 \\ -13 \\ 7 \\ 6 \end{bmatrix}.$$

A partir de las observaciones, podemos obtener los resultados siguientes: 

1) Estimar por MCO el modelo de regresión.


$$\begin{aligned} \hat{\mathbf{B}} &= (X'X)^{-1}X'Y = \begin{bmatrix} 5 & 15 & 19 \\ 15 & 55 & 49 \\ 19 & 49 & 99 \end{bmatrix}^{-1} \begin{bmatrix} -11 \\ 56 \\ -215 \end{bmatrix} = \\ &= \begin{bmatrix} 2,98 & -0,54 & -0,3 \\ -0,54 & 0,13 & 0,04 \\ -0,3 & 0,04 & 0,05 \end{bmatrix} \begin{bmatrix} -11 \\ 56 \\ -215 \end{bmatrix} = \begin{bmatrix} 2,1 \\ 4,9 \\ -5 \end{bmatrix} \end{aligned}$$

Lectura recomendada

Para ver que las varianzas de los estimadores MCR son siempre menores, o como máximo iguales, que las de los estimadores MCO, consultad la obra siguiente:
G.G. Judge; R.C. Hill; W.E. Griffiths; H. Lütkepohl; T.C. Lee (1988). *Introduction to the Theory and Practice of Econometrics* (2.ª ed.). Nueva York: John Wiley & Sons.

Trabajar con el estimador MCR...

... resulta más interesante, lógicamente, en el caso de que las restricciones lineales sean ciertas, dado que no tiene sesgo.

 Respecto a los puntos 1, 2 y 3, únicamente se presentan aquí los resultados, dado que son contenidos desarrollados en los subapartados 2.3.1 y 2.4.2 de este módulo didáctico.

2) Obtener el estimador no sesgado de la varianza del término de perturbación, el coeficiente de determinación y el coeficiente de determinación corregido por los grados de libertad del modelo.

- El estimador no sesgado de la varianza del término de perturbación es el siguiente:

$$\hat{\sigma}_u^2 = \frac{\mathbf{e}'\mathbf{e}}{N - k} = \frac{0,7}{5 - 3} = 0,35.$$

- El coeficiente de determinación estará determinado por:

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{Y}'\mathbf{Y} - T\bar{Y}^2} = 1 - \frac{0,7}{1.302,8} = 0,9995.$$

El resultado pone de manifiesto que el ajuste obtenido es suficientemente bueno, dado que las variables exógenas consideradas en la regresión explican el 99,95% de la variabilidad total de la variable endógena.

- Por otra parte, el coeficiente de determinación corregido por los grados de libertad del modelo es el siguiente:

$$\bar{R}^2 = 1 - \left[\frac{N - 1}{N - k} (1 - R^2) \right] = 1 - \left[\frac{5 - 1}{5 - 3} (1 - 0,9995) \right] = 0,9989.$$

3) Obtener la matriz de varianzas y covarianzas de los estimadores MCO.

$$\begin{aligned} \text{VAR}(\hat{\mathbf{B}}) &= \hat{\sigma}_{u_{MCO}}^2 (\mathbf{X}'\mathbf{X})^{-1} = 0,35 \begin{bmatrix} 2,98 & -0,54 & -0,30 \\ -0,54 & 0,131 & 0,04 \\ -0,30 & 0,04 & 0,05 \end{bmatrix} \\ &= \begin{bmatrix} 1,04 & -0,19 & -0,11 \\ -0,19 & 0,046 & 0,014 \\ -0,11 & 0,014 & 0,017 \end{bmatrix} \end{aligned}$$

4) Contrastar la restricción lineal: $\beta_3 = -\beta_2$.

La restricción lineal puede escribirse así: $\beta_2 + \beta_3 = 0$. Para llevar a cabo este contraste se dispone de diferentes alternativas:

a) Utilizar el estadístico 3.3. En este caso tenemos lo siguiente:

$$\mathbf{R} = [0, 1, 1], \quad \mathbf{r} = [0], \quad \mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}.$$

A continuación, calculamos cada uno de los componentes del estadístico de prueba 3.3:

$$\mathbf{RB} - \mathbf{r} = [0 \ 1 \ 1] \begin{bmatrix} 2,1 \\ 4,9 \\ -5 \end{bmatrix} - 0 = 0(2,1) + 1(4,9) + 1(-5) = 4,9 - 5 = -0,1.$$

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = [0 \ 1 \ 1] \begin{bmatrix} 2,98 & -0,54 & -0,30 \\ -0,54 & 0,131 & 0,04 \\ -0,30 & 0,04 & 0,05 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 0,259.$$

Por lo tanto, $(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} = 3,864$.

- El número de restricciones es $q = 1$.

Por tanto, el estadístico 3.3 queda de la siguiente manera:

$$F_0 = \frac{(-0,1)(3,864)(-0,1)}{0,35} = 0,11.$$

Comparando este valor con 18,51 –que es el valor en tablas de la F de Snedecor con un grado de libertad en el numerador y dos ($5 - 3 = 2$) grados de libertad en el denominador, con un nivel de significación, α , del 5%– podemos concluir que no puede rechazarse la hipótesis nula.

b) Una segunda posibilidad consiste en comparar las SCE del modelo restringido y del modelo no restringido, o ampliado, empleando el estadístico 3.6. De este estadístico de prueba, conocemos los componentes siguientes:

- La SCE_A se corresponde con la SCE asociada al modelo estimado por MCO. Por lo tanto, es 0,7.
- El número de restricciones, q o s , es 1.

Así pues, sólo falta conocer la SCE del modelo restringido. Para calcularla, en primer lugar debemos introducir la restricción lineal $\beta_2 + \beta_3 = 0$ en el modelo. El modelo restringido queda del modo siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} - \beta_2 X_{3i} + u_i \Rightarrow Y_i = \beta_1^* + \beta_2^* Z_i + u_i,$$

donde $Z_i = X_{2i} - X_{3i}$, $\forall i = 1, \dots, N$.

A continuación, se estima el modelo restringido por MCO:

$$\hat{\mathbf{B}}_{\text{MCR}}^* = (\mathbf{X}_R' \mathbf{X}_R)^{-1} \mathbf{X}_R' \mathbf{Y}$$

siendo las matrices correspondientes las siguientes:

$$\mathbf{X}_R = \begin{bmatrix} 1 & X_{R,21} \\ 1 & X_{R,22} \\ \vdots & \vdots \\ 1 & X_{R,25} \end{bmatrix} = \begin{bmatrix} 1 & X_{21} - X_{31} \\ 1 & X_{22} - X_{32} \\ \vdots & \vdots \\ 1 & X_{25} - X_{35} \end{bmatrix} = \begin{bmatrix} 1 & -6 \\ 1 & 3 \\ 1 & -3 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}; \quad \mathbf{Y} = \begin{bmatrix} -28 \\ 17 \\ -13 \\ 7 \\ 6 \end{bmatrix}.$$

Por lo tanto,

$$\hat{\mathbf{B}}_{\text{MQR}}^* = (\mathbf{X}_R' \mathbf{X}_R)^{-1} \mathbf{X}_R' \mathbf{Y} = \begin{bmatrix} 0,21 & 0,015 \\ 0,02 & 0,019 \end{bmatrix} \begin{bmatrix} -11 \\ 271 \end{bmatrix} = \begin{bmatrix} 1,7727 \\ 4,9659 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{1\text{MQR}}^* \\ \hat{\beta}_{2\text{MQR}}^* \end{bmatrix}$$

Finalmente, se calcula la SCE del modelo restringido:

$$\mathbf{e}_R' \mathbf{e}_R = \mathbf{Y}' \mathbf{Y} - \hat{\mathbf{B}}_{\text{MQR}}^* \mathbf{X}_R' \mathbf{Y} = 1.327 - [1,7727 \quad 4,9659] \begin{bmatrix} -11 \\ 271 \end{bmatrix} = 0,7386.$$

Si introdujéramos

$$\beta_2 = -\beta_3, \dots$$

... el modelo restringido sería $Y_i = \beta_1^* + \beta_3^*(X_{3i} - X_{2i}) + u_i$, $i = 1, \dots, N$. Sin embargo, en cualquier caso se obtendrían los mismos resultados.

Para acabar, el estadístico de prueba 3.7 queda como vemos a continuación:

$$F_0 = \frac{\frac{\mathbf{e}'_R \mathbf{e}_R - \mathbf{e}'_A \mathbf{e}_A}{q}}{\frac{\mathbf{e}'_A \mathbf{e}_A}{N - k}} = \frac{0,7386 - 0,7}{\frac{1}{0,7}} = 0,11,$$

que coincide con lo que se había obtenido con la primera estrategia.

c) Un tercer estadístico que permitiría (aunque teóricamente no se había enunciado hasta ahora) llevar a cabo el contraste de restricciones lineales es:

$$F_0 = \frac{\frac{(\hat{\mathbf{B}}_{\text{MCR}} - \hat{\mathbf{B}})' \mathbf{X}' \mathbf{X} (\hat{\mathbf{B}}_{\text{MCR}} - \hat{\mathbf{B}})}{q}}{\frac{\mathbf{e}' \mathbf{e}}{N - k}} = \frac{0,038}{0,35} = 0,11,$$

$$\text{donde } \hat{\mathbf{B}}_{\text{MQR}} = \begin{bmatrix} \hat{\beta}_{1\text{MQR}}^* \\ \hat{\beta}_{2\text{MQR}}^* \\ -\hat{\beta}_{2\text{MQR}}^* \end{bmatrix} = \begin{bmatrix} 1,7727 \\ 4,9659 \\ -4,9659 \end{bmatrix}$$

que, lógicamente, coincide con los resultados anteriores.

d) Como se quiere contrastar una única restricción lineal, una cuarta manera de hacerlo es mediante el test de la t de Student. En este caso, primero hay que hallar la distribución de probabilidad del estimador MCO de la restricción:

$$\hat{\beta}_2 + \hat{\beta}_3 \sim N[E[\hat{\beta}_2 + \hat{\beta}_3], \text{VAR}[\hat{\beta}_2 + \hat{\beta}_3]].$$

$$E[\hat{\beta}_2 + \hat{\beta}_3] = E[\hat{\beta}_2] + E[\hat{\beta}_3] = \beta_2 + \beta_3 = 0.$$

$$\begin{aligned} \text{VAR}[\hat{\beta}_2 + \hat{\beta}_3] &= \text{VAR}[\hat{\beta}_2] + \text{VAR}[\hat{\beta}_3] + 2\text{CÓV}[\hat{\beta}_2, \hat{\beta}_3] = \\ &= 0,046 + 0,017 + 2 \cdot 0,014 = 0,0906. \end{aligned}$$

El estadístico de prueba es el siguiente:

$$t_0 = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{\sqrt{\text{VAR}[\hat{\beta}_2 + \hat{\beta}_3]}} = \frac{(4,9 - 5) - 0}{\sqrt{0,0906}} = -0,3322.$$

El valor del estadístico de prueba es inferior al valor crítico y, por tanto, no se puede rechazar la hipótesis nula.

5) Contrastar la significación global del modelo.

La hipótesis nula en este caso es $H_0: \beta_2 = \beta_3 = 0$, y se presentan tres estrategias, que señalamos a continuación:

a) Utilizar el estadístico de prueba siguiente:

$$F_0 = \frac{\frac{SCR}{N - k}}{\frac{SCE}{N - k}} = \frac{R^2}{1 - R^2} \sim F_{k-1, N-k}$$

Si $\beta_2 - 3\beta_3 = 0$ fuese...

... la restricción lineal que estuviésemos interesados en contrastar, tendríamos lo siguiente:

$$E[\hat{\beta}_2 - 3\hat{\beta}_3] = \beta_2 - 3\beta_3 = 0$$

$$\begin{aligned} \text{VAR}[\hat{\beta}_2 - 3\hat{\beta}_3] &= \text{VAR}[\hat{\beta}_2] + \\ &+ 3^2 \text{VAR}[\hat{\beta}_3] - 6\text{CÓV}[\hat{\beta}_2, \hat{\beta}_3]. \end{aligned}$$

En el ejemplo,...

... pese a que el valor del estadístico de prueba es inferior al valor crítico, observad que $t_0^2 \cong F_0$. Entonces tenemos que $(-0,3322)^2 \cong 0,11$.

Hemos definido el estadístico de prueba que utilizamos en esta estrategia en el subapartado 2.7.2 de este módulo didáctico.

que en este ejemplo vale aproximadamente 1.860 si tomamos una $R^2 = 0,99946$. El valor obtenido por el estadístico de prueba se debe comparar con el valor de una F de Snedecor con dos grados de libertad en el numerador y el denominador. El valor crítico es 19,00 al 5% de significación. Por lo tanto, rechazamos la hipótesis nula, y el modelo resulta adecuado en el ámbito de la población.

b) Comparar las SCE del modelo restringido y del modelo ampliado empleando el estadístico 3.7. La información sobre el modelo ampliado la hemos tenido que calcular previamente. En el ejemplo que consideramos tenemos:

$$s = 2; N = 5; k = 3; SCE_A = 0,7.$$

Sólo falta por conocer la SCE del modelo que incorpora la hipótesis nula, es decir, del modelo restringido. En este caso, el modelo es el siguiente:

$$Y_i = \beta_1 + v_i \quad i = 1, \dots, N.$$

Además, teniendo en cuenta las características peculiares de este modelo, sabemos que:

$$\hat{\beta}_1 = \bar{Y} = -2,2; SCE_R = SCT = 1.302,8 \Rightarrow SCR_R = 0.$$

En consecuencia, el estadístico de prueba vale:

$$F_0 = \frac{\frac{1.302,8 - 0,7}{2}}{\frac{0,7}{2}} = 1.860.$$

c) Una tercera manera de llevar a cabo el contraste de significación global del modelo es considerarlo como un contraste de restricciones lineales: $H_0: \mathbf{RB} = \mathbf{r}$, empleando el estadístico 3.3 que hemos visto anteriormente. Los componentes de este estadístico son los siguientes:

- $q = 2; \mathbf{e}'\mathbf{e} = \mathbf{e}_A'\mathbf{e}_A = 0,7, N - k = 5 - 3 = 2;$
- $\mathbf{RB} - \mathbf{r} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2,1 \\ 4,9 \\ -5 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4,9 \\ -5,0 \end{bmatrix};$
- $[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} = \left\{ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2,98 & -0,54 & -0,30 \\ -0,54 & 0,131 & 0,04 \\ -0,30 & 0,04 & 0,049 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}^{-1} = \begin{bmatrix} 10 & -8 \\ -8 & 26,8 \end{bmatrix}.$

Por lo tanto, el estadístico de prueba queda como vemos a continuación:

$$F_0 = \frac{\begin{bmatrix} 4,9 & -5 \end{bmatrix} \begin{bmatrix} 10 & -8 \\ -8 & 26,8 \end{bmatrix} \begin{bmatrix} 4,9 \\ -5 \end{bmatrix}}{2} = 1.860.$$

Actividades

3.2. Teniendo en cuenta el modelo de regresión siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i,$$

¿cómo contrastarías la restricción $\beta_j + \beta_h = 0$? Proponed tres métodos distintos para realizar este contraste y comentadlos.

3.3. Comprobad que, si se estima directamente por MCR aplicando la expresión 3.18, suponiendo que $\beta_2 = \beta_3 = 0$, se obtendrán unos estimadores restringidos que cumplirán la restricción impuesta, aunque ésta, en realidad, es falsa.

3.4. Contrastad la restricción lineal $\beta_2 + \beta_3 = 15$ mediante las estrategias que hemos explicado antes y comprobad que el estadístico F_0 vale 2.517 y el de la t_0 , -50,17.

3.3. Análisis de la permanencia estructural. Contraste de Chow

En este subapartado estudiamos una aplicación del contraste de restricciones lineales al caso en que se contrasta la constancia del valor de la población de los parámetros β del modelo de regresión, a lo largo de toda la muestra analizada.

3.3.1. Introducción

Como hemos dicho en un subapartado anterior, una de las hipótesis básicas del modelo de regresión estándar es la hipótesis de permanencia estructural de los parámetros del modelo.

¿Qué significa el incumplimiento de esta hipótesis? La respuesta depende del tipo de datos con los que tratemos:

- Si trabajamos con datos de serie temporal, que haya un cambio de estructura dentro del periodo muestral significa que se han producido uno o más cambios en la estructura interna del proceso generador de los datos.
- Por otra parte, el hecho de que trabajemos con datos de corte transversal significa que la muestra contiene dos o más grupos de individuos que presentan comportamientos distintos entre ellos.

En la figura de la izquierda que presentamos a continuación mostramos un ejemplo de incumplimiento de la hipótesis de permanencia estructural. En $t = 100$ se produce un cambio en la estructura: los parámetros β_j del modelo (ordenada en el origen y pendiente en este ejemplo) presentan unos valores claramente distintos para las observaciones correspondientes a periodos anteriores a t (crecientes) y para las observaciones posteriores a t (decrecientes).

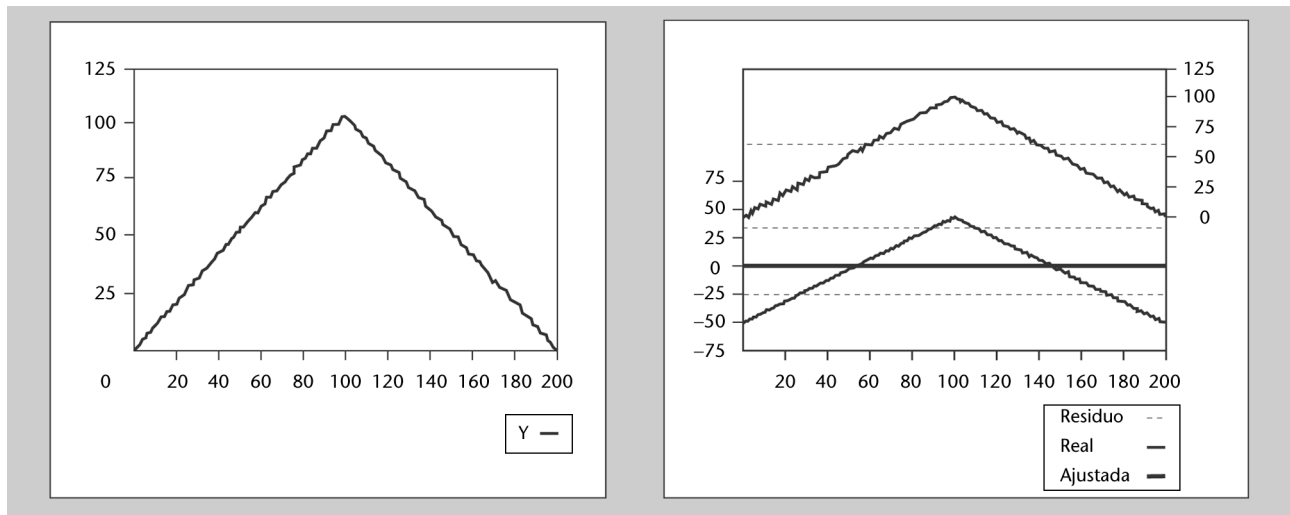
Consultad la hipótesis de permanencia estructural de los parámetros en el subapartado 2.2.4 de este módulo didáctico.

Los cambios en la estructura interna...

... del proceso generador de los datos los pueden haber provocado, por ejemplo, los cambios en el contexto socioeconómico.

Un comportamiento diferente...

... entre grupos de individuos puede ser, por ejemplo, que el salario que una determinada empresa paga a sus trabajadores sea diferente dependiendo de si el empleado está afiliado a un sindicato o no lo está.



Por otra parte, en la figura de la derecha presentamos la serie real y ajustada en el caso de que hubiésemos supuesto que se cumpliera la hipótesis de permanencia estructural. Como podemos ver, si no se tiene en cuenta el cambio estructural que ponen de manifiesto las observaciones, el ajuste no es bueno y los estimadores de \mathbf{B} que se obtienen son aproximadamente una media ponderada de los estimadores que se obtendrían si, considerando la existencia de cambio estructural, se hubiesen ajustado dos regresiones distintas, una para cada una de las dos submuestras en que puede dividirse la muestra total (la primera desde 1 hasta 100 y la segunda desde 101 hasta 200).

Las consecuencias que esto implica son graves; por una parte, las predicciones son poco fiables y, por otra, los contrastes de significación son erróneos, dado que, si los estimadores son ineficientes, sus varianzas están infladas. En este caso, aumenta la probabilidad de no rechazar la hipótesis nula de significación individual de un parámetro para el periodo muestral total, por lo cual puede considerarse como no relevante, aunque en realidad lo pueda resultar para cada una de las submuestras por separado.

Por lo tanto, si después de realizar la representación gráfica de la nube de puntos se detecta un posible incumplimiento de la hipótesis de permanencia estructural, antes de estimar un único modelo de regresión para todo el tamaño muestral, hay que contrastar el cumplimiento de esta hipótesis. De modo que, si después de realizar el contraste se concluye que existe cambio estructural, es decir, que dos o más grupos de observaciones (tanto si son de corte transversal como de serie temporal) tienen una estructura diferente entre sí, para estimar correctamente el modelo tendremos que especificar una regresión distinta para cada una de las submuestras detectadas, en lugar de especificar una única regresión para todo el tamaño muestral.

La matriz de varianzas y covarianzas...

... de los estimadores MCO bajo las hipótesis del MRLM es la siguiente:

$$\text{VAR}[\hat{\mathbf{B}}] = \sigma_{\varepsilon}^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Por lo tanto, si se ajusta una única recta en presencia de cambio estructural, tal como podemos ver en la figura de la derecha del gráfico anterior, aumenta la *SCE*, lo cual implica un incremento en la varianza del término de perturbación que se traslada a las varianzas de los estimadores.

Consultad la introducción de variables ficticias en el modelo como solución alternativa a un posible incumplimiento de la hipótesis de permanencia estructural en el subapartado 1.2 del módulo "Variables exógenas cualitativas" de esta asignatura.

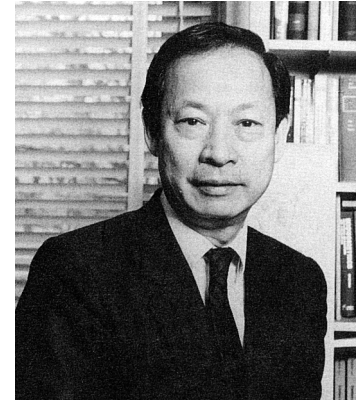
3.2.2. Contraste de Chow de permanencia estructural

El contraste de permanencia estructural que presentamos a continuación es el que se conoce como **contraste de Chow**.


Las hipótesis nula y alternativa que formula este test son las siguientes:

$$H_0: \text{permanencia estructural}; H_A: \text{no-permanencia estructural. (3.21)}$$

Este test consiste en comparar las *SCE* de la regresión para todo el tamaño muestral con las *SCE* de las regresiones para cada una de las submuestras fijadas.



El economista G.C. Chow

Por lo tanto, y suponiendo que únicamente existen dos submuestras posibles (como es el caso de la figura de la izquierda del gráfico anterior), el contraste de Chow se desarrollaría como describimos a continuación: 

1) Se descompone el periodo muestral en las dos submuestras posibles detectadas. En este ejemplo supondremos que la primera submuestra engloba a las T_1 primeras observaciones, y la segunda, las T_2 últimas (de la $T_1 + 1$ a la última).

2) Acto seguido, se especifican las tres regresiones siguientes:

a) Una para todo el tamaño muestral: $Y_t = XB_T + U_t \quad t = 1, 2, \dots, T.$

b) Una segunda para la primera submuestra: $Y_i = XB_1 + U_i \quad i = 1, 2, \dots, T_1.$

c) Para acabar, una tercera para la segunda submuestra: $Y_j = XB_2 + U_j \quad j = T_1 + 1, \dots, T.$

3) A continuación, se estiman por MCO cada una de las tres regresiones anteriores y se calculan las *SCE* asociadas a cada una. Así, se obtienen las *SCE* asociadas a todo el periodo muestral (SCE_T), las asociadas a la primera submuestra (SCE_1) y las asociadas a la segunda submuestra (SCE_2).

Una vez que hemos llegado a este punto, se construye el estadístico de prueba, que se define de la manera siguiente:

$$F_0 = \frac{\frac{SCE_T - (SCE_1 - SCE_2)}{k}}{\frac{SCE_1 + SCE_2}{T - 2k}} \quad (3.22)$$

y que se distribuye según una *F* de Snedecor con k y $T - 2k$ grados de libertad en el numerador y el denominador, respectivamente.

Sobre los grados de libertad del modelo

Los grados de libertad se calculan a partir de la suma y/o resta de los grados de libertad de sus componentes. Por ejemplo, el numerador tiene k grados de libertad porque:

$$\frac{SCE_T - (SCE_1 - SCE_2)}{\sigma_u^2} \sim \chi_{(T-k) - (T_1-k + T_2-k)}^2 = \chi_k^2.$$

Si el valor del estadístico de prueba, F_0 , es superior al valor crítico en tablas de la F de Snedecor, se rechaza la hipótesis nula; por lo tanto, se tienen indicios de que en el ámbito de la población existe cambio estructural.

Por el contrario, si el estadístico de prueba no supera el valor crítico, se concluye que en el ámbito de la población existe permanencia estructural.

Finalmente, hay que señalar que en el supuesto de que existan tres submuestras posibles diferentes, el estadístico de prueba del contraste de Chow está determinado por la expresión siguiente:


$$F_0 = \frac{\frac{SCE_T - (SCE_1 + SCE_2 + SCE_3)}{2k}}{\frac{SCE_1 + SCE_2 + SCE_3}{T - 3k}}$$

SCE_T es la suma de cuadrados de los errores surgidos de la estimación con las T observaciones, mientras que SCE_1 , SCE_2 y SCE_3 son las sumas de cuadrados de los errores asociados a las tres regresiones, trabajando con las tres submuestras en que se ha dividido T .

3.3.3. Un caso particular: tamaño insuficiente de una de las submuestras

A veces, cuando se debe contrastar la hipótesis de permanencia estructural de dos submuestras, nos podemos encontrar en el caso de que una tenga un número insuficiente de observaciones. En ese caso, no es posible aplicar el test de Chow tal como se ha definido hasta ahora, dado que no habrá suficientes observaciones para estimar el modelo de regresión asociado a esta submuestra.

3.3.4. Limitaciones del contraste de Chow

Podemos señalar, como veremos a continuación, cuatro inconvenientes o limitaciones del test de Chow para detectar el posible incumplimiento de la hipótesis de permanencia estructural: 

1) Se puede detectar un cambio estructural espurio, que tiene lugar cuando se detecta la presencia de cambio estructural como consecuencia de una especificación errónea del modelo en su parte determinista. Se dan casos particulares de esto cuando se especifica una relación lineal, pero la relación no lo es, o cuando se especifica un modelo con omisión de variables explicativas relevantes.

El problema de insuficiencia...

... de observaciones también se plantea si la diferencia entre el número de observaciones de la submuestra y el número de parámetros del modelo es reducida, dado que la SCE obtenida será imprecisa.

La expresión utilizada es:

$$F_0 = \frac{\frac{SCE_T - SCE_1}{T_2}}{\frac{SCE_1}{T_1 - k}} \sim F_{T_2, T_1 - k}$$

2) Es sensible a la existencia de heteroscedasticidad en el término de perturbación. Por lo tanto, si se incumple la hipótesis de homoscedasticidad, será necesario corregirlo antes de analizar la permanencia estructural.

El problema de la heteroscedasticidad se estudia en el módulo "Incumplimiento de las hipótesis básicas en el término de error".

3) Pierde potencia a medida que el punto donde se produce el cambio estructural se acerca a uno de los extremos de la muestra: cuanto menor sea una de las submuestras, menor será la potencia del contraste de Chow. Recordad que la potencia de un contraste es la probabilidad de aceptar la hipótesis alternativa si es cierta. Es el complementario del error tipo II.

Para tener más detalles sobre los conceptos de error tipo I, tipo II y potencia de un contraste, consultad la asignatura *Fundamentos de Estadística*.

4) A veces, no se dispone de información suficiente para conocer el punto exacto donde se produce el posible cambio estructural. Este inconveniente, debido a la razón que acabamos de explicar, se da principalmente cuando se trabaja con datos de serie temporal.

Ejemplo de aplicación del test de Chow

A partir de la observación del cuadro de resultados de la estimación MCO del modelo 3.23, para todo el tamaño muestral, para la primera submuestra y para la segunda submuestra, se quiere estudiar si se detecta cambio estructural en la observación 71, asociado al modelo siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i. \tag{3.23}$$

Para aplicar el test de Chow 3.22, hay que estimar el modelo por MCO con todo el tamaño muestral y con cada una de las dos submuestras (la primera hasta la observación 70, y la segunda desde la 71 hasta la última, la 100).

Los resultados obtenidos pueden verse en los cuadros que presentamos más adelante, y son: $SCE_T = 37.083,88$, $SCE_1 = 18,51$, $SCE_2 = 56,12$. A partir de los resultados anteriores se construye el estadístico de prueba siguiente:

$$F_0 = \frac{\frac{37.083,88 - (18,51 + 56,12)}{2}}{\frac{18,51 + 56,12}{96}} = 23.803,59.$$

Por lo tanto, se rechaza la hipótesis nula de permanencia estructural, ya que el valor en tablas de Snedecor es aproximadamente 3,1.

Estimación MCO del modelo 3.23 para todo el tamaño muestral

Variable dependiente: Y				
Número de observaciones: 100				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	-0.2072400	4.4601625	-0.0464647	0.9630
X ₂	7.1390007	1.1426618	6.2476937	0.0000
R-squared	0.284847	Mean of dependent var	24.86850	
Adjusted R-squared	0.277550	S.D. of dependent var	22.88630	
S.E. of regression	19.45268	Sum of squared resid	37083.88	
Log likelihood	-437.6822	F-statistic	39.03368	
Durbin-Watson stat	0.307508	Prob(F-statistic)	0.000000	

Estimación MCO del modelo 3.23 para la primera submuestra

Variable dependiente: Y				
Número de observaciones: 70				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	1.9967329	0.1396661	14.296475	0.0000
X ₂	3.0269722	0.0371059	81.576620	0.0000
R-squared	0.989885	Mean of dependent var	12.19114	
Adjusted R-squared	0.989736	S.D. of dependent var	5.150637	
S.E. of regression	0.521810	Sum of squared resid	18.51540	
Log likelihood	-52.77947	F-statistic	6654.745	
Durbin-Watson stat	2.452634	Prob(F-statistic)	0.000000	

Estimación MCO del modelo 3.23 para la segunda submuestra

Variable dependiente: Y				
Número de observaciones: 30				
VARIABLES	COEFICIENTES	ERROR STD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	8.4067762	0.6384511	13.167456	0.0000
X ₂	11.959019	0.1516330	78.868197	0.0000
R-squared	0.995519	Mean of dependent var	54.44900	
Adjusted R-squared	0.995359	S.D. of dependent var	20.78107	
S.E. of regression	1.415760	Sum of squared resid	56.12254	
Log likelihood	-51.96326	F-statistic	6220.192	
Durbin-Watson stat	2.142763	Prob(F-statistic)	0.000000	

Glosario

análisis estructural

Aplicación de la modelización econométrica que consiste en medir cuantitativamente las relaciones económicas entre las variables incluidas en el modelo.

autocorrelación

Existencia de correlación entre términos de perturbación diferentes.

cambio estructural

Fenómeno producido cuando no se cumple la hipótesis básica de permanencia estructural y, por tanto, las relaciones que se producen entre las variables presentes en el modelo no se mantienen constantes a lo largo de toda la muestra.

coeficientes beta

Coefficientes que permiten determinar qué variable hace una aportación mayor a la explicación del comportamiento de la variable endógena, cuando los regresores están expresados en unidades de medida diferentes.

coeficiente R^2

Coefficiente que mide el porcentaje de la variabilidad total de la variable endógena explicado por el modelo. Está fijado entre $0 \leq R^2 \leq 1$, si hay término independiente y entre $-\infty \leq R^2 \leq 1$, si no lo hay.

coeficiente \bar{R}^2

Coefficiente que mide el porcentaje de la variabilidad total de la variable endógena explicado por el modelo, pero que, a diferencia del coeficiente R^2 , tiene en cuenta la pérdida de grados de libertad.

contraste de Chow

Método destinado a contrastar la hipótesis básica de permanencia estructural de los parámetros del modelo de regresión entre dos o más submuestras.

contraste de significación global

Método para contrastar la existencia de, por lo menos, una diferencia significativa entre los parámetros de la población y el valor 0 para todos estos parámetros.

contraste de significación individual de un parámetro

Método para contrastar la existencia de una diferencia significativa entre uno de los parámetros de la población y el valor 0.

error

Ved *residuo*.

especificación de un modelo

Fase inicial del trabajo econométrico. Consiste en escribir la variable que se pretende explicar a partir de un conjunto de parámetros y variables.

estimación del modelo

Etapas de la investigación econométrica que consiste en asignar valores numéricos a los parámetros desconocidos del modelo.

estimadores MCO de \mathbf{B} y σ_u^2

Estimadores obtenidos mediante el criterio de minimizar la suma de cuadrados de los errores. Si se cumplen los supuestos básicos del MRLM, los estimadores MCO de \mathbf{B} serán no sesgados, de varianza mínima y consistentes, mientras que el estimador MCO de σ_u^2 será no sesgado.

estimadores MV de \mathbf{B} y σ_u^2

Estimadores que se obtienen a partir del criterio de maximizar el logaritmo neperiano de la función de verosimilitud. Los estimadores MV (de máxima verosimilitud) de \mathbf{B} coinciden con los MCO si se cumplen todas las hipótesis básicas del MRLM, mientras que el estimador MV de σ_u^2 no coincide.

evaluación de políticas

Aplicación de la modelización econométrica que consiste en simular distintos escenarios alternativos con el propósito de tomar decisiones en un entorno de menos incertidumbre.

forma funcional

Tipo de relación entre la variable endógena y las variables explicativas. Puede ser lineal, cuadrática, exponencial, logarítmica, etc.

heteroscedasticidad

Incumplimiento de la propiedad de homoscedasticidad.

homoscedasticidad

Propiedad que cumple el término de perturbación aleatoria si posee la misma varianza para todos los individuos considerados.

matriz R

Matriz de restricciones, formada por los coeficientes fijos asociados a todos los parámetros del vector B en cada una de las restricciones consideradas. Es de dimensión $q \times k$, donde q es el número de restricciones lineales que hay que contrastar y k , el de parámetros del modelo.

mínimos cuadrados restringidos

Método de estimación que se basa en el hecho de minimizar la suma de cuadrados de los errores, sujeta a una restricción lineal impuesta. En consecuencia, las estimaciones siempre cumplirán las restricciones impuestas, tanto si éstas son ciertas como si no lo son.

modelo en desviaciones

Modelo de regresión lineal planteado de modo que supone que las observaciones muestrales de las variables se han transformado previamente tomando las diferencias de cada observación respecto a la media muestral de la variables que mide.

modelo intrínsecamente lineal

Modelo que permite una expresión equivalente en la cual la variable dependiente (o una transformación de ésta) se escribe como combinación lineal de las variables explicativas más un término de perturbación.

modelo intrínsecamente no lineal

Modelo que no permite la transformación descrita en la definición de *modelo intrínsecamente lineal*.

MRLM

Modelo de regresión lineal múltiple.

MRLS

Modelo de regresión lineal simple.

multicolinealidad

Existencia de correlación muestral elevada entre las observaciones de las variables explicativas.

parámetros

Valores desconocidos que cuantifican la relación que existe entre la variable endógena y cada una de las variables explicativas. Un objetivo de la econometría consiste en estimar estos valores.

permanencia estructural

Hipótesis básica del modelo de regresión mediante la cual se supone que las relaciones que existen entre las variables presentes en el modelo se mantienen constantes a lo largo de toda la muestra.

perturbación esférica

Existencia de homoscedasticidad en el modelo e independencia entre los términos de perturbación de los individuos considerados.

perturbación no esférica

Incumplimiento de los requerimientos de la *perturbación esférica*.

predicción

Nombre que recibe una de las aplicaciones de la modelización econométrica, que consiste en obtener los valores que en el futuro (o para otros individuos fuera de la muestra) puede tomar la variable endógena.

predicción por intervalo del valor esperado

Estimación del intervalo de pertenencia del verdadero valor de la esperanza matemática de la variable dependiente para unos valores fijos de las variables explicativas.

predicción por intervalo del valor observado

Estimación del intervalo de pertenencia del verdadero valor que tomará la variable dependiente para unos valores fijos de las variables explicativas.

predicción puntual

Estimación del valor que tomará la variable dependiente para valores fijados de las variables explicativas.

residuo

Diferencia entre el valor observado de la variable dependiente y el valor ajustado por el modelo una vez que se han estimado sus parámetros.

restricción estocástica

Restricción en la cual la combinación de parámetros del modelo está sujeta a un error aleatorio.

restricción homogénea

Restricción en la cual la combinación de parámetros del modelo es igual a cero.

restricción lineal exacta de desigualdad

Restricción en la cual una determinada combinación lineal entre distintos parámetros del modelo econométrico de regresión cumple una determinada desigualdad.

restricción lineal exacta de igualdad

Restricción en la cual una determinada combinación lineal entre distintos parámetros del modelo de regresión es igual a una constante.

SCE

Ved *suma de cuadrados de los errores*.

SCR

Ved *suma de cuadrados de la regresión*.

SCT

Ved *suma de cuadrados totales*.

suma de cuadrados de la regresión

Parte de la variabilidad total de la variable endógena que el modelo no puede explicar.

suma de cuadrados de los errores

Parte de la variabilidad total de la variable endógena que el modelo no puede explicar.

suma de cuadrados totales

Medida de la variabilidad total de la variable endógena.

término de perturbación

Variable aleatoria del modelo. El término de perturbación refleja el comportamiento de la variable endógena que no refleja la parte determinista del modelo especificado, así como también los errores cometidos al medir las variables explicativas incluidas en el modelo.

variable endógena

Variable del modelo econométrico que estamos interesados en conocer y cuyo comportamiento deseamos explicar.

variable explicativa

Variable que, de acuerdo con los postulados de la teoría económica, permite explicar el comportamiento de la variable endógena. Dependiendo de si consideramos una o más de una, llamaremos al modelo *simple* o *múltiple*, respectivamente.

vector \mathbf{r}

Vector de dimensión q , formado por los términos independientes de las restricciones lineales.

Bibliografía

Gujarati, D. (1990). *Econometría* (2.^a ed.). Bogotá: McGraw-Hill.

En el momento de estudiar el apartado 2 de este módulo didáctico es recomendable consultar los capítulos 2, 3, 4, 5 y 7 de este libro. Para el estudio del apartado 3, consultad los capítulos 5, 6 y 7.

Haavelmo, T. (1944). "The Statistical Implications of Assistent of Simultaneous Equation". *Econometrica* (núm. 11, pág. 1-12).

Intrilligator, M. D. (1978). *Econometric Models, Techniques and Applications* (cap. 1). Amsterdam: North-Holland.

Johnston, J. (1987). *Métodos de econometría* (trad. J. Sánchez Fernández). Barcelona: Vicens-Vives.

Consultad el capítulo 1 para los temas trabajados en el apartado 1 de este módulo didáctico. Los capítulos 2, 3, 4, 5 y 7 son útiles para el estudio del apartado 2. Finalmente, para el estudio del apartado 3 podéis consultar las páginas 50-56 y 245-270.

Judge, G. G.; Hill, R. C.; Griffiths, W. E.; Lütkepohl, H.; Lee, T. C. (1988). *Introduction to the Theory and Practice of Econometrics* (2.^a ed., cap. 6). Nueva York: John Willey & Sons.

Maddala, G. S. (1992). *Econometría*. México: McGraw-Hill.

Los capítulos 7 y 8 de este libro van bien para la materia trabajada en los apartados 2 y 3 de este módulo didáctico.

Novalés, A. (1993). *Econometría* (2.^a ed.). Madrid: McGraw-Hill.

Con relación al apartado 1 de este módulo didáctico, es interesante consultar el capítulo "Introducción" de este libro. Los capítulos 3 y 4 son adecuados como bibliografía del apartado 2 de este módulo. También es recomendable consultar el capítulo 4 para estudiar el apartado 3.

Pulido, A. (1983). *Modelos econométricos*. Madrid: Pirámide.

Podéis consultar los capítulos 1, 2, 3 y 4 para el estudio de los contenidos del apartado 1 de este módulo didáctico. Para los contenidos del apartado 2, podéis consultar los capítulos 5, 6 y 7. Para los contenidos del apartado 3, podéis consultar los capítulos 6, 7 y 8.

Uriel, E. y otros (1990). *Econometría. El modelo lineal*. Madrid: AC.

Podéis consultar el capítulo 1 de este libro en el momento de estudiar los contenidos del apartado 1 de este módulo didáctico. Los contenidos del apartado 2 pueden consultarse en los capítulos 2, 3, 4 y 5. También podéis consultar los contenidos del apartado 3 en los capítulos 4, 5 y 6 del libro.