

CAPÍTULO 1

MODELO LINEAL GENERAL

1.1. INTRODUCCIÓN

Este capítulo está dedicado a estudiar, desde el punto de vista econométrico, un modelo que es lineal en las variables originales o en transformaciones de las mismas y que tiene un número de variables explicativas superior a dos.

Ejemplo1.1: Como ilustración, tomaremos uno de los ejemplos más repetidos en la Econometría Aplicada: la función de producción Cobb-Douglas. En su forma más simple el modelo puede escribirse como:

$$Q_t = A K_t^\alpha L_t^\beta e^{u_t} \quad t = 1, \dots, T \quad (1.1)$$

en donde:

Q_t = cantidad producida en el periodo t .

K_t = servicios de capital utilizados en el periodo t .

L_t = servicios de trabajo utilizados en el periodo t .

A , α y β son parámetros y u_t es una variable que llamaremos perturbación aleatoria.

Tomando logaritmos, el modelo (1.1) puede escribirse como:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad (1.2)$$

en donde:

$$y_t = \log Q_t, \quad x_{2t} = \log K_t \quad \text{y} \quad x_{3t} = \log L_t$$

x_{1t} toma siempre el valor 1; es lo que llamaremos la constante del modelo.

Los T datos recogidos para las tres variables pueden referirse a la experiencia de una empresa en una serie de periodos, a la experiencia de un sector en diferentes

periodos, a todo un país en un periodo temporal o a varias empresas o sectores en un mismo periodo.

Nosotros, en este trabajo, vamos a utilizar los datos contenidos en el Cuadro 1.1. Son datos anuales de Estados Unidos para el periodo 1899-1922 y son los utilizados por Cobb y Douglas (1928) para estimar por primera vez este tipo de función de producción. Los datos corresponden a la producción (Y), el stock de capital (K) y el trabajo (L).

CUADRO 1.1. Datos para el ejemplo 1.1

OBS.	Y	K	L	LY	LK	LL
1899	100.0000	100.0000	100.0000	4.6052	4.6052	4.6052
1900	101.0000	107.0000	105.0000	4.6151	4.6728	4.6540
1901	112.0000	114.0000	110.0000	4.7185	4.7362	4.7005
1902	122.0000	122.0000	118.0000	4.8040	4.8040	4.7707
1903	124.0000	131.0000	123.0000	4.8203	4.8752	4.8122
1904	122.0000	138.0000	116.0000	4.8040	4.9273	4.7536
1905	143.0000	149.0000	125.0000	4.9628	5.0039	4.8283
1906	152.0000	163.0000	133.0000	5.0239	5.0938	4.8903
1907	151.0000	176.0000	138.0000	5.0173	5.1705	4.9273
1908	126.0000	185.0000	121.0000	4.8363	5.2204	4.7958
1909	155.0000	198.0000	140.0000	5.0434	5.2883	4.9416
1910	159.0000	208.0000	144.0000	5.0689	5.3375	4.9698
1911	153.0000	216.0000	145.0000	5.0304	5.3753	4.9767
1912	177.0000	226.0000	152.0000	5.1761	5.4205	5.0239
1913	184.0000	236.0000	154.0000	5.2149	5.4638	5.0370
1914	169.0000	244.0000	149.0000	5.1299	5.4972	5.0039
1915	189.0000	266.0000	154.0000	5.2417	5.5835	5.0370
1916	225.0000	298.0000	182.0000	5.4161	5.6971	5.2040
1917	227.0000	335.0000	196.0000	5.4250	5.8141	5.2781
1918	223.0000	366.0000	200.0000	5.4072	5.9026	5.2983
1919	218.0000	387.0000	193.0000	5.3845	5.9584	5.2627
1920	231.0000	407.0000	193.0000	5.4424	6.0088	5.2627
1921	179.0000	417.0000	147.0000	5.1874	6.0331	4.9904
1922	240.0000	431.0000	161.0000	5.4806	6.0661	5.0814

El objetivo de este capítulo es dar respuesta a cuestiones del siguiente tipo:

- ¿Qué procedimientos deben de utilizarse para, teniendo en cuenta los datos y toda la evidencia disponible, determinar si el modelo escrito en (1.2) es o no empíricamente razonable?
- ¿Podría aceptarse un modelo en el que se incorporara la hipótesis de rendimientos constantes?
- ¿Cuál es el mejor procedimiento para estimar los parámetros del modelo?

1.2. MODELO E HIPÓTESIS.

El modelo econométrico que vamos a estudiar esta constituido por las siguientes hipótesis:

Hipótesis 1: Para el periodo t-ésimo podemos escribir:

$$y_t = \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t \quad t = 1, 2, \dots, T.$$

En forma matricial el modelo puede escribirse como:

$$y = X \beta + u \quad (1.3)$$

en donde:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}; X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

Salvo que se indique lo contrario siempre supondremos que la primera columna de X tiene todos sus elementos igual a uno.

A la combinacion lineal de los regresores la llamaremos parte sistemática del modelo, mientras que u_t es la parte aleatoria o perturbación aleatoria. Esta distinción entre parte sistemática y parte aleatoria es muy útil porque pone de manifiesto que los factores relevantes que explican el comportamiento de la variable endógena quedan

recogidos en la parte sistemática mientras que la parte aleatoria recoge el efecto conjunto de aquellos factores que influyen sobre la variable dependiente pero no determinan una pauta identificable y estable que resulte útil para explicar el comportamiento de la variable dependiente.

Hipótesis relativas a la Parte Sistemática.

Hipótesis 2. Las variables explicativas no son estocásticas. Esta hipótesis hay que situarla en el marco de muestreo que siempre acompaña a la evaluación de todo modelo econométrico. Esta evaluación esta basada en lo que “ocurriría si” el cálculo de los estadísticos se repitiera muchas veces generando los datos en cada caso. Decir que las variables explicativas no son estocásticas equivale a decir que, en las sucesivas repeticiones, esos valores se mantendrán constantes. En este sentido, algunos autores hablan de X como si fuera una matriz de diseño.

Hipótesis 3. El rango de la matriz X es k . Este supuesto excluye la posibilidad de que una de las variables pueda expresarse como una combinación lineal de las $k-1$ restantes variables explicativas. En este caso decimos que no hay multicolinealidad exacta.

Hipótesis 4. Asintóticamente se tiene que:

$$\frac{X'X}{T} = \begin{bmatrix} \frac{\sum x_{1t}^2}{T} & \frac{\sum x_{1t}x_{2t}}{T} & \dots & \frac{\sum x_{1t}x_{kt}}{T} \\ \frac{\sum x_{1t}x_{2t}}{T} & \frac{\sum x_{2t}^2}{T} & \dots & \frac{\sum x_{2t}x_{kt}}{T} \\ \dots & \dots & \dots & \dots \\ \frac{\sum x_{kt}x_{1t}}{T} & \frac{\sum x_{kt}x_{2t}}{T} & \dots & \frac{\sum x_{kt}^2}{T} \end{bmatrix} \rightarrow Q \quad (1.4)$$

en donde Q es una matriz de constantes finitas. Esta hipótesis es muy importante porque lo que se esta suponiendo es que los promedios de cuadrados y de productos cruzados giran en torno a constantes. Aunque, como ya hemos dicho, las observaciones no son estocásticas el comportamiento que siguen es como si procedieran de un proceso estocástico estacionario. A veces, esta hipótesis se formula como: $X'X=O(T)$ queriendo decir que basta dividir los términos de la matriz por T para que los términos de la matriz tiendan a expresiones finitas bien definidas.

Hipótesis relativas a la Parte Aleatoria

Hipótesis 5. La esperanza matemática de cada perturbación aleatoria es cero

$$Eu_t = 0 \quad t = 1, 2, \dots, T$$

Esta hipótesis indica que aunque dentro de u_t se da cabida a muchos factores que influyen sobre y_t , el efecto conjunto de todos ellos no tiene un signo definido.

Hipótesis 6. No existe ninguna correlación entre perturbaciones correspondientes a periodos temporales diferentes:

$$Eu_t u_s = 0 \quad t \neq s \quad s, t = 1, 2, \dots, T$$

Si hubiera correlación entre perturbaciones correspondientes a diferentes periodos, entonces, dada la información de un periodo, podría utilizarse para predecir el comportamiento de la perturbación en otro periodo, identificándose así una pauta entre los factores recogidos en lo que se llama perturbación aleatoria.

Hipótesis 7. La varianza es la misma para cada una de las perturbaciones aleatorias. Es decir,

$$Eu_t^2 = \sigma^2 \quad \forall t$$

Esto quiere decir que la dispersión de los efectos de los factores recogidos en u_t es la misma sea cual sea el periodo. Si la dispersión fuera diferente entonces podría decirse que la perturbación aleatoria esconde algo relevante que puede ser utilizado explícitamente para dar cuenta del comportamiento de la variable dependiente.

Vectorialmente, la Hipótesis 5 puede escribirse como:

$$Eu = \begin{bmatrix} Eu_1 \\ Eu_2 \\ \vdots \\ Eu_T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \quad (1.5)$$

y las Hipótesis 6 y 7,

$$\begin{aligned} \text{Var}(u) = E u u' &= \begin{bmatrix} E u_1^2 & E u_1 u_2 & \dots & E u_1 u_T \\ E u_2 u_1 & E u_2^2 & \dots & E u_2 u_T \\ \vdots & \vdots & \ddots & \vdots \\ E u_T u_1 & E u_T u_2 & \dots & E u_T^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \\ &= \sigma^2 I_T \end{aligned} \quad (1.6)$$

Hipótesis 8. El vector de perturbaciones aleatorias sigue una distribución normal T-variante con vector de medias cero y matriz de varianzas y covarianzas igual a $\sigma^2 I_T$. Podemos escribir :

$$u \sim N[0, \sigma^2 I_T] \quad (1.7)$$

A partir de estas hipótesis puede derivarse la distribución de probabilidad de los elementos de y.

Resultado 1.1. El vector y, de T elementos, se distribuye como:

$$y \sim N[X\beta, \sigma^2 I_T] \quad (1.8)$$

Prueba: A partir de la Hipótesis 8 la distribución de probabilidad de u puede escribirse como:

$$f(u) = (2\pi)^{-\frac{T}{2}} |\sigma^2 I_T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} u' (\sigma^2 I_T)^{-1} u\right\}$$

La regla que permite obtener la función de distribución de una variable que es transformación de otra variable puede escribirse como:

$$f(y) = f(u) \left| \frac{\partial u}{\partial y} \right| \quad (1.9)$$

en donde $\left| \frac{\partial u}{\partial y} \right|$ es el Jacobiano que adopta la forma siguiente:

$$\frac{\partial u}{\partial y} = \begin{pmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \cdots & \frac{\partial u_1}{\partial y_T} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \cdots & \frac{\partial u_2}{\partial y_T} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_T}{\partial y_1} & \frac{\partial u_T}{\partial y_2} & \cdots & \frac{\partial u_T}{\partial y_T} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Por lo tanto, la función de probabilidad de y puede escribirse como:

$$f(y) = (2\pi)^{-\frac{T}{2}} |\sigma^2 I_T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)\right\} \quad (1.10)$$



Ejemplo 1.1. (Continuación). Veamos el significado de las hipótesis en el marco del modelo escrito en (1.2). Tal como aparece en (1.2), lo que se indica es que los únicos determinantes relevantes que explican la producción son el trabajo y el capital usando una relación con elasticidad constante. Cumpliéndose las Hipótesis 5, 6 y 7 se garantiza que los factores recogidos en la perturbación aleatoria no dan cuenta de nada relevante. También parece cumplirse la Hipótesis 3 ya que es difícil que exista una relación lineal exacta entre los dos inputs. Más difícil de sostener es la Hipótesis 4, ya que tanto el producto como los inputs parecen seguir una pauta creciente bien definida.

1.3. ESTIMADORES MÍNIMO CUADRÁTICO ORDINARIOS. (MCO)

El vector de estimadores MCO de β es el siguiente:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (1.11)$$

A este vector le corresponde la siguiente suma de cuadrados de los residuos:

$$S = \sum_1^T \hat{u}_t^2 = \hat{u}'\hat{u} \quad (1.12)$$

en donde \hat{u} es el vector de residuos MCO definido como:

$$\hat{u} = y - X\hat{\beta} \quad (1.13)$$

A continuación, vamos a demostrar que al vector de estimadores MCO le corresponde el menor valor de la suma de los cuadrados de los residuos. Previamente vamos a demostrar el siguiente resultado.

Resultado1.2. El vector de residuos MCO escrito en (1.13) es ortogonal con todas las columnas de la matriz X.

Prueba: La expresión (1.11) puede escribirse en forma de ecuaciones normales:

$$X'X\hat{\beta} = X'y \quad (1.14)$$

o, equivalentemente, como:

$$X'(y - X\hat{\beta}) = 0 \quad (1.15)$$

llegándose así al resultado. □

Resultado1.3. Sea β^* un vector de estimadores cualquiera de β y u^* el correspondiente vector de residuos. Entonces la suma de cuadrados de residuos de este vector de estimadores siempre supera la que corresponde al vector MCO escrita en (1.12).

Prueba: El vector de residuos correspondiente a β^* puede escribirse como:

$$u^* = y - X\beta^* = y - X\hat{\beta} + X(\hat{\beta} - \beta^*) = \hat{u} + X(\hat{\beta} - \beta^*) \quad (1.16)$$

Por lo tanto, la suma de cuadrados de estos residuos es la siguiente:

$$u^{*'}u^* = \hat{u}'\hat{u} + (\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) + 2\hat{u}'X(\hat{\beta} - \beta^*)$$

y teniendo en cuenta (1.15) se llega a:

$$u^{*'}u^* = \hat{u}'\hat{u} + (\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) \quad (1.17)$$

Esta es la llamada Identidad Fundamental. Al resultado se llega teniendo en cuenta que, por la Hipótesis 3, la matriz $X'X$ es definida positiva. □

Derivemos ahora las propiedades del vector de estimadores MCO escrito en (1.11).

Resultado1.4. El vector de estimadores MCO es insesgado.

$$E\hat{\beta} = \beta \quad (1.18)$$

Prueba: Sustituyendo (1.3) en (1.11) se obtiene:

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \quad (1.19)$$

y teniendo en cuenta las Hipótesis 2 y 5 el resultado sigue:

$$E\hat{\beta} = \beta + (X'X)^{-1}X'E u = \beta \quad \square$$

Resultado1.5. La matriz de varianzas y covarianzas del vector de estimadores MCO viene dada por:

$$\text{Var}(\hat{\beta}) = \begin{bmatrix} \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) & \dots & \text{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{Var}(\hat{\beta}_k) \end{bmatrix} = \sigma^2 (X'X)^{-1} \quad (1.20)$$

Prueba: La matriz de varianzas y covarianzas se define como:

$$\begin{aligned} \text{Var}(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = E \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \vdots \\ \hat{\beta}_k - \beta_k \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 - \beta_1 & \hat{\beta}_2 - \beta_2 & \dots & \hat{\beta}_k - \beta_k \end{bmatrix} = \\ &= \begin{bmatrix} E(\hat{\beta}_1 - \beta_1)^2 & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) & \dots & E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k) \\ E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_2 - \beta_2)^2 & \dots & E(\hat{\beta}_2 - \beta_2)(\hat{\beta}_k - \beta_k) \\ \vdots & \vdots & \ddots & \vdots \\ E(\hat{\beta}_k - \beta_k)(\hat{\beta}_1 - \beta_1) & E(\hat{\beta}_k - \beta_k)(\hat{\beta}_2 - \beta_2) & \dots & E(\hat{\beta}_k - \beta_k)^2 \end{bmatrix} \end{aligned}$$

Utilizando (1.19) podemos escribir:

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = E(X'X)^{-1} X'uu'X(X'X)^{-1}$$

y utilizando la Hipótesis 2 y el resultado escrito en (1.6) se llega a:

$$\text{Var}(\hat{\beta}) = (X'X)^{-1} X'Eu u'X(X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

llegándose así al resultado. □

Resultado 1.6. (Teorema de Gauss-Markov). El vector de estimadores MCO de $\beta, \hat{\beta}$, es el mejor estimador lineal insesgado en el siguiente sentido: cualquier otro estimador, β^* , que sea también combinación lineal de los elementos del vector y e insesgado tiene una matriz de varianzas y covarianzas que excede la de $\hat{\beta}$ en una matriz semidefinida positiva.

Prueba: Suponer que A es una matriz $k \times T$ de constantes y , sin pérdida de generalidad, el estimador β^* lo escribimos como:

$$\beta^* = [A + (X'X)^{-1}X']y \quad (1.21)$$

Por la propia definición, los estimadores son función lineal de los elementos de y .

Sustituyendo (1.3) en (1.21) se llega a:

$$\beta^* = [A + (X'X)^{-1}X'] [X\beta + u] = (AX + I)\beta + Au + (X'X)^{-1}X'u$$

Tomando esperanzas

$$E\beta^* = (AX + I)\beta$$

Para que sea insesgado se debe cumplir que:

$$AX = 0 \quad (1.22)$$

La matriz de varianzas y covarianzas es:

$$\text{Var}(\beta^*) = E [A + (X'X)^{-1}X'] uu' [A + (X'X)^{-1}X']' = \sigma^2 AA' + \sigma^2 (X'X)^{-1}$$

utilizando (1.22) en los productos cruzados. Podemos escribir:

$$\text{Var}(\beta^*) - \text{Var}(\hat{\beta}) = \sigma^2 AA'$$

y el resultado sigue por ser AA' semidefinida positiva. □

Resultado 1.7. Sea w un vector de k constantes. Entonces se tiene que:

$$\text{Var}(w'\hat{\beta}) \leq \text{Var}(w'\beta^*) \quad (1.23)$$

Prueba: Utilizando el Resultado 1.6 podemos escribir:

$$\text{Var}(w'\beta^*) = w' [\sigma^2 AA' + \sigma^2 (X'X)^{-1}] w \geq \sigma^2 w' (X'X)^{-1} w$$

cumpliendo el resultado. □

Hasta ahora, los resultados se han derivado sin utilizar la Hipótesis 8 que trata de la normalidad de las perturbaciones aleatorias. Si añadimos esta hipótesis llegamos al siguiente resultado.

Resultado 1.8. Suponer que se cumplen las ocho hipótesis comentadas en la sección anterior. Entonces la distribución de probabilidad de $\hat{\beta}$ viene dada por:

$$\hat{\beta} \sim N[\beta, \sigma^2 (X'X)^{-1}] \quad (1.24)$$

Prueba: A partir de (1.19) se ve que los elementos de $\hat{\beta}$ son igual a los correspondientes elementos de β más una combinación lineal de variables que son normales llegándose al resultado.

Los momentos primero y segundo han sido ya derivados en los Resultados 1.4 y 1.5, respectivamente. □

Ejemplo 1.1. (Continuación). A partir de los datos contenidos en el Cuadro 1.1 se suelen calcular algunos estadísticos que nos informan acerca del tipo de variables que estamos tratando.

En primer lugar, podemos dar cuenta de algunas características univariantes referidas a las variables en logaritmos.

CUADRO 1.2. Características Univariantes

Variable	y	x ₂	x ₃
Valor Máximo	5.48	6.06	5.29
Valor Mínimo	4.61	4.60	4.60
Media	5.07	5.35	4.96
Desviación Típica	0.27	0.50	0.20
Asimetría	-0.12	0.03	0.10
Apuntamiento	-1.06	-1.15	0.85
Coefici. de Variación	0.05	0.08	0.04

Otra información interesante se refiere a las correlaciones entre las variables.

CUADRO 1.3. Correlaciones entre las variables.

	y	x ₂	x ₃
y	1	0.95	0.96
x ₂		1	0.91
x ₃			1

Los estimadores MCO de β utilizando (1.11) son los siguientes:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} T & \sum x_{2t} & \sum x_{3t} \\ \sum x_{2t} & \sum x_{2t}^2 & \sum x_{2t}x_{3t} \\ \sum x_{3t} & \sum x_{3t}x_{2t} & \sum x_{3t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum x_{2t}y_t \\ \sum x_{3t}y_t \end{bmatrix} = \begin{bmatrix} 24 & 128.55 & 119.10 \\ & 693.45 & 639.92 \\ & & 592.02 \end{bmatrix}^{-1} \begin{bmatrix} 121.86 \\ 655.41 \\ 605.94 \end{bmatrix} = \begin{bmatrix} -0.1773 \\ 0.2305 \\ 0.8073 \end{bmatrix}$$

Notar que el valor de la elasticidad que corresponde al factor trabajo (0.8073) es bastante mayor que el correspondiente al capital (0.2305).

A continuación, vamos a derivar algunos resultados para aquellos casos en que el tamaño muestral tiende a hacerse grande. (Marco Asintótico). Comenzaremos presentando algunas definiciones que completan las presentadas en el Anexo 1.

Definición 1.1. Sea $\hat{\theta}$ un estimador de θ . Se dice que $\hat{\theta}$ es un estimador consistente asintóticamente normal de θ cuando: i) es consistente y ii) $\sqrt{T}(\hat{\theta} - \theta)$ converge en distribución a $N(0, \Sigma)$ en donde Σ es una matriz definida positiva con elementos finitos.

Definición 1.2. Sean $\hat{\theta}$ y θ^* dos estimadores consistentes y asintóticamente normales con matrices de varianzas y covarianzas de las correspondientes distribuciones asintóticas iguales a Σ y Ω , respectivamente. Entonces se dice que $\hat{\theta}$ es asintóticamente eficiente respecto a θ^* si $\Omega - \Sigma$ es semidefinida positiva.

Definición 1.3. Se dice que un estimador consistente y asintóticamente normal es asintóticamente eficiente si lo es con respecto a cualquier estimador consistente y asintóticamente normal que se considere.

Resultado 1.9. Una condición suficiente para que un estimador consistente y asintóticamente normal sea asintóticamente eficiente es que su matriz de varianzas y covarianzas asintótica sea igual al límite de la cota inferior de Crámer-Rao.

Prueba: Ver Theil (1971, Sección 8.4). □

Establecidas estas definiciones veamos ahora como se comportan los estimadores MCO de β en este marco asintótico.

Resultado 1.10. El vector de estimadores MCO de β es consistente pudiendose escribir:

$$\text{plim } \hat{\beta} = \beta \quad (1.25)$$

Prueba: El vector $\hat{\beta}$ sigue una distribución de probabilidad k-dimensional con vector de medias β y matriz de varianzas y covarianzas igual a $\sigma^2(X'X)^{-1}$. Utilizando la Hipótesis 4 se tiene que:

$$\lim_{T \rightarrow \infty} \text{Var}(\hat{\beta}) = \lim_{T \rightarrow \infty} \left[\frac{\sigma^2}{T} \left(\frac{X'X}{T} \right)^{-1} \right] = 0 \cdot Q^{-1} = 0$$

El sesgo cero y varianza que tiende a cero asintóticamente son condiciones suficientes para la convergencia en probabilidad tal como puede verse en el Anexo \square

Resultado 1.11. El estimador $\hat{\beta}$ es asintóticamente normal en el siguiente sentido:

$$\sqrt{T}(\hat{\beta} - \beta) \sim N(0, \sigma^2 Q^{-1}) \quad (1.26)$$

Prueba: Para cualquier tamaño muestral, la distribución de $\sqrt{T}(\hat{\beta} - \beta)$ es:

$$N\left(0, \sigma^2 \left(\frac{X'X}{T} \right)^{-1} \right)$$

A partir de aquí el resultado sigue utilizando la Hipótesis 4. \square

También puede demostrarse que $\hat{\beta}$ es asintóticamente eficiente. Nosotros presentaremos este resultado en el apartado dedicado a los estimadores máximo-verosímiles.

Comentadas las propiedades de los estimadores MCO de los parámetros de posición, a continuación, analizamos las propiedades del estimador MCO del parámetro de dispersión que es la varianza de la perturbación aleatoria para lo cuál derivamos, primero, las propiedades estocásticas del vector de residuos MCO escrito en (1.13).

Resultado 1.12. El vector de residuos MCO, \hat{u} , puede escribirse como:

$$\hat{u} = Mu \quad (1.27)$$

en donde $M = I - X(X'X)^{-1}X'$ es una matriz simétrica e idempotente de rango T-k.

Prueba: A partir de (1.13) se tiene:

$$\hat{u} = y - X(X'X)^{-1}X'y = My = M(X\beta + u) = Mu$$

por ser :

$$MX = (I - X(X'X)^{-1}X')X = X - X = 0$$



Resultado 1.13. La esperanza y matriz de varianzas y covarianzas de \hat{u} vienen dados, respectivamente, por:

$$E(\hat{u}) = 0 \quad (1.28)$$

y

$$\text{Var}(\hat{u}) = \sigma^2 M \quad (1.29)$$

Prueba:

$$E(\hat{u}) = M \cdot E(u) = 0$$

$$\text{Var}(\hat{u}) = E(\hat{u}\hat{u}') = E(Muu'M) = M \cdot E(uu') \cdot M = \sigma^2 M$$



Resultado 1.14. La esperanza de la suma de los cuadrados de los residuos MCO es igual a $(T-k)\sigma^2$, es decir:

$$E(\hat{u}'\hat{u}) = \sigma^2 (T-k) \quad (1.30)$$

Prueba:

$$\begin{aligned} E(\hat{u}'\hat{u}) &= E(u'MMu) = E(u'Mu) = E(\text{tr}(Mu'u)) = \text{tr}(M(E(uu'))) = \\ &= \sigma^2 \text{tr}M = \sigma^2 (T-k) \end{aligned}$$

por ser: $\text{tr}(M) = \text{rango}(M) = T-k$

Establecidas las propiedades del vector de residuos MCO, examinaremos, a continuación, las propiedades del estimador MCO de σ^2 definido como:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k} = \frac{\sum_{t=1}^T \hat{u}_t^2}{T-k} \quad (1.31)$$

Resultado 1.15. El estimador MCO de σ^2 escrito en (1.31) se distribuye como:

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{T-k} \chi^2(T-k) \quad (1.32)$$

en donde $\chi^2(T-k)$ es una variable chi-cuadrado central con $T-k$ grados de libertad.

Su esperanza matemática es:

$$E(\hat{\sigma}^2) = \sigma^2 \quad (1.33)$$

y su varianza:

$$\text{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{(T-k)^2} \text{Var}(\chi^2(T-k)) = \frac{2\sigma^4}{T-k} \quad (1.34)$$

Prueba: para demostrar (1.32) hay que tener en cuenta que $(T-k)\hat{\sigma}^2 = u'Mu$.

Por ser M una matriz idempotente de rango $(T-k)$, es posible escribir:

$$u'Mu = v_1^{*2} + v_2^{*2} + \dots + v_{T-k}^{*2} \quad (1.35)$$

en donde cada una de las variables v_j^* sigue una distribución Normal con media cero y varianza σ^2 ; además, todas ellas son independientes entre sí. Dividiendo ambos lados de (1.35) por σ^2 se tiene que:

$$\frac{u'Mu}{\sigma^2} = v_1^2 + v_2^2 + \dots + v_{T-k}^2$$

en donde ahora las v_j son variables normales tipificadas independientes entre sí.

Por lo tanto: $\frac{u'Mu}{\sigma^2} \sim \chi^2(T-k)$, y el resultado sigue.

Para obtener (1.33) hay que tener en cuenta que: $E(\chi^2(T-k)) = T-k$.

Por último, respecto a (1.34) basta recordar que: $\text{Var}(\chi^2(T-k))=2(T-k)$, por lo que: $\text{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{(T-k)^2} \text{Var}(\chi^2(T-k)) = \frac{2\sigma^4}{T-k}$ ☐

Resultado 1.16: El estimador $\hat{\sigma}^2$ de σ^2 es insesgado

Prueba: resultado obtenido en (1.33). ☐

Alternativamente, la insesgadez podría demostrarse utilizando el Resultado 1.14.

Resultado 1.17: $\hat{\beta}$ y $\hat{\sigma}^2$ son independientes.

Prueba: En el Resultado 1.15 hemos visto que $\hat{\sigma}^2$ depende de una forma cuadrática de u cuya matriz es M . Por otra parte, a partir de (1.19) podemos escribir:

$\hat{\beta} - \beta = (X'X)^{-1} X'u$ y como se cumple que $(X'X)^{-1} X'M = 0$, esto garantiza la independencia. ☐

Resultado 1.18: El estimador $\hat{\sigma}^2$ es un estimador consistente de σ^2 .

Prueba: $\hat{\sigma}^2$ es insesgado para cualquier tamaño muestral. Además, a partir de (1.34) se tiene que:

$$\lim_{T \rightarrow \infty} \text{Var}(\hat{\sigma}^2) = \lim_{T \rightarrow \infty} \frac{2\sigma^4}{T-k} = 0$$

Como ya hemos indicado para $\hat{\beta}$, la insesgadez y el límite cero de la varianza son condiciones suficientes para la consistencia del estimador. ☐

Se puede definir una matriz que sea la estimación de la matriz de varianzas y covarianzas de $\hat{\beta}$ escrita en (1.20) como $\hat{\sigma}^2 (X'X)^{-1}$. Por los Resultados 1.16 y 1.18 esta estimación será insesgada y consistente.

Resultado 1.19: El estimador $\hat{\sigma}^2$ es asintóticamente normal con:

$$\sqrt{T}(\hat{\sigma}^2 - \sigma^2) \sim N(0, 2\sigma^4) \quad (1.36)$$

Prueba: Ver Fomby, Hill y Johnson (1984) (Teorema 4.3.2). □

Ejemplo 1.1. (Continuación) . El estimador MCO de la varianza lo calculamos utilizando (1.31):

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T - k} = \frac{0,070982}{21} = 0,00338099$$

La estimación de la matriz de varianzas y covarianzas viene dada por:

$$\hat{\sigma}^2 (X'X)^{-1} = \begin{pmatrix} 0,1886 & 0,0199 & -0,0595 \\ & 0,0040 & -0,0083 \\ & & 0,0210 \end{pmatrix}$$

Suponer ahora que el modelo escrito en (1.3) lo particionamos como:

$$y = X_1\beta_1 + X_2\beta_2 + u \quad (1.37)$$

en donde β_1 tiene k_1 elementos y β_2 tiene k_2 elementos con $k_1 + k_2 = k$. Vamos a derivar la forma que adopta el estimador MCO de β_2 .

Resultado 1.20: El subvector de estimadores MCO, $\hat{\beta}_2$, puede escribirse como:

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y \quad (1.38)$$

con $M_1 = I_T - X_1(X_1'X_1)^{-1}X_1'$

Prueba: Primero, reescribimos las ecuaciones normales escritas en (1.14) particionadas conforme a (1.37):

$$X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'y$$

$$X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 = X_2'y$$

Se trata de dos bloques, el primero con k_1 relaciones y el segundo con k_2 relaciones. A partir del primer bloque se obtiene:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2$$

y sustituyendo a $\hat{\beta}_1$ en el segundo bloque:

$$X_2'X_1(X_1'X_1)^{-1}X_1'y - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2 + X_2'X_2\hat{\beta}_2 = X_2'y$$

que también puede escribirse como:

$$(X_2'M_1X_2)\hat{\beta}_2 = X_2'M_1y$$

Basta premultiplicar ambos lados de esta expresión por la inversa de $(X_2'M_1X_2)$ para llegar a (1.38). □

Lo que indica (1.38) es que $\hat{\beta}_2$ no es otra cosa que la regresión de M_1y sobre M_1X_2 , es decir, la regresión de aquella parte de y no explicada por X_1 sobre aquella parte de X_2 no explicada por X_1 . Es importante destacar que al mismo resultado se llegaría haciendo la regresión de y sobre M_1X_2 . Estos resultados explican que si una de las variables explicativas del modelo es una constante, el vector de estimadores MCO del resto de las $(k-1)$ variables explicativas obtenido utilizando las desviaciones con respecto a las medias de todas las variables sea el mismo que el que se obtenía utilizando variables no centradas.

1.4. ESTIMADORES MÁXIMO-VEROSÍMILES (MV).

Los estimadores MV de β y σ^2 son aquellos que maximizan el valor que toma la función de verosimilitud. Esta función es la misma escrita en (1.10) pero tomándola como función de β y σ^2 condicionada sobre unos valores dados de X e y . Para reflejar este cambio la escribiremos como:

$$L(\beta, \sigma^2/y, X) = (2\pi)^{-\frac{T}{2}} |\sigma^2 I_T|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta) \right\}$$

y su logarítmico:

$$l(\beta, \sigma^2/y, X) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \quad (1.39)$$

Veamos ahora la forma que adoptan el gradiente y la matriz hessiana de segundas derivadas.

Resultado 1.21. El gradiente de la función de verosimilitud definida en (1.39) toma la forma siguiente:

$$d(\beta, \sigma^2) = \begin{bmatrix} \frac{\partial l(.)}{\partial \beta} \\ \frac{\partial l(.)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}(X'y - X'X\beta) \\ -\frac{T}{2\sigma^2} + \frac{u'u}{2\sigma^4} \end{bmatrix} \quad (1.40) - (1.41)$$

La matriz hessiana de segundas derivadas es:

$$D(\beta, \sigma^2) = - \begin{bmatrix} \frac{1}{\sigma^2} X'X & \frac{1}{\sigma^4} X'(y - X\beta) \\ \cdot & -\frac{T}{2\sigma^4} + \frac{1}{\sigma^6} (y - X\beta)'(y - X\beta) \end{bmatrix} \quad (1.42)$$

Prueba: Teniendo en cuenta que :

$$(y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

y utilizando las reglas estandar de derivación vectorial se llega a:

$$\frac{\partial l(.)}{\partial \beta} = -\frac{1}{2\sigma^2}(-2X'y - 2X'X\beta)$$

llegándose así a (1.40). Para obtener (1.41) basta derivar (1.39) respecto a σ^2 .

$$\frac{\partial l(.)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{(y - X\beta)'(y - X\beta)}{2\sigma^4}$$

Respecto a las segundas derivadas, usando (1.40) se tiene que:

$$\frac{\partial^2 l(.)}{\partial \beta \partial \beta'} = -\frac{X'X}{\sigma^2}$$

La segunda derivada cruzada puede escribirse como:

$$\frac{\partial^2 l(.)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} (X'y + X'X\beta)$$

y, por último, la segunda derivada respecto a σ^2 se obtiene usando (1.41):

$$\frac{\partial l(.)}{(\partial \sigma^2)^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6} (y - X\beta)'(y - X\beta) \quad \square$$

Resultado 1.22. Los estimadores MV de β y σ^2 vienen dados por:

$$\tilde{\beta} = (X'X)^{-1} X'y \quad (1.43)$$

$$\tilde{\sigma}^2 = \frac{\tilde{u}'\tilde{u}}{T} \quad (1.44)$$

Prueba: La condición necesaria para maximizar la verosimilitud es que los elementos del gradiente sean cero. Es decir:

$$\begin{aligned} \frac{1}{\tilde{\sigma}^2} (X'y - X'X\tilde{\beta}) &= 0 \\ -\frac{T}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} (y - X\tilde{\beta})'(y - X\tilde{\beta}) &= 0 \end{aligned}$$

A partir del primer bloque de k relaciones se obtiene que:

$$X'y = X'X\tilde{\beta}$$

de donde se deriva (1.43).

La última relación puede escribirse como:

$$\frac{\tilde{u}'\tilde{u}}{\tilde{\sigma}^2} = T$$

llegándose a (1.44).



Hay que destacar que el estimador MV de β coincide con el estimador MCO coincidencia que no se da con los estimadores de σ^2 . Por lo tanto, las propiedades del estimador MV de β son las mismas que las ya comentadas para el estimador MCO.

Respecto a las propiedades de $\tilde{\sigma}^2$, hay que tener en cuenta que puede escribirse como:

$$\tilde{\sigma}^2 = \frac{T-k}{T} \hat{\sigma}^2 \quad (1.45)$$

A partir de esta expresión vamos a derivar las propiedades del estimador MV.

Resultado 1.23. El estimador MV de σ^2 se distribuye como:

$$\tilde{\sigma}^2 \sim \frac{\sigma^2}{T} \chi^2(T-k) \quad (1.46)$$

Su esperanza matemática es:

$$E \tilde{\sigma}^2 = \frac{T-k}{T} \sigma^2 \quad (1.47)$$

y su varianza:

$$\text{Var}(\tilde{\sigma}^2) = \frac{2(T-k)\sigma^4}{T^2} \quad (1.48)$$

Prueba: Teniendo en cuenta (1.45) basta utilizar el Resultado 1.15.



Teniendo en cuenta que el factor que multiplica a $\hat{\sigma}^2$ en (1.45) tiende a 1 conforme T tiende a infinito entonces las propiedades del estimador MV son las mismas que las del estimador MCO. Es decir, $\tilde{\sigma}^2$ es consistente y la distribución asintótica de $\sqrt{T}(\tilde{\sigma}^2 - \sigma^2)$ es igual a $N(0, 2\sigma^4)$.

Por último, vamos a demostrar que tanto los estimadores MCO como los MV son asintóticamente eficientes de acuerdo con lo establecido en el Resultado 1.9. Primeramente, tenemos que definir el límite de la Cota Inferior de Cramer-Rao.

Definición 1.4. La Cota Inferior de Cramer-Rao se define como:

$$I(\beta, \sigma^2)^{-1} = [-ED(\theta)]^{-1}$$

Asintóticamente, el concepto relevante es el límite de la expresión:

$$\lim \left(\frac{I(\beta, \sigma^2)}{T} \right)^{-1} \quad (1.49)$$

Resultado 1.24. La Matriz de Información en este caso viene dada por:

$$I(\beta, \sigma^2) = \frac{1}{\sigma^2} \begin{bmatrix} X'X & 0 \\ 0 & \frac{T}{2\sigma^2} \end{bmatrix} \quad (1.50)$$

y el límite escrito en (1.49):

$$\lim_{T \rightarrow \infty} \left(\frac{I(\beta, \sigma^2)}{T} \right)^{-1} = \begin{pmatrix} \sigma^2 Q^{-1} & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \quad (1.51)$$

Prueba: Teniendo en cuenta la matriz de segundas derivadas escrita en (1.42) el resultado (1.50) sigue teniendo en cuenta que:

$$\begin{aligned} E \frac{1}{\sigma^4} X'(y - X\beta) &= \frac{1}{\sigma^4} X'Eu = 0 \\ E \left[-\frac{T}{2\sigma^4} + \frac{(y - X\beta)'(y - X\beta)}{\sigma^6} \right] &= -\frac{T}{2\sigma^4} + \frac{Eu'u}{\sigma^6} = \\ &= -\frac{T}{2\sigma^4} + \frac{T\sigma^2}{\sigma^6} = \frac{T}{2\sigma^4} \end{aligned}$$

En lo que respecta a (1.51), basta tener en cuenta la Hipótesis 4 y las reglas estandar de convergencia. □

Examinando las distribuciones asintóticas derivadas para los estimadores MCO y MV se ve que todos ellos son asintóticamente eficientes.

1.5. ESTIMADORES DE LOS MÍNIMOS CUADRADOS RESTRINGIDOS (MCR). ERROR DE ESPECIFICACIÓN.

Los estimadores MCR de β , que llamaremos $\hat{\beta}_R$, son aquellos que minimizan la suma de cuadrados de los residuos cumpliendo una serie de restricciones especificadas a priori sobre los parámetros.

Escribimos las restricciones como:

$$R\beta = q \quad (1.52)$$

En donde R es una matriz $r.k$, y rango r , siendo r el número de restricciones; q es un vector de r constantes.

Resultado 1.25. El vector de estimadores MCR que minimiza la suma de cuadrados de los residuos cumpliendo al mismo tiempo las restricciones escritas en (1.52) es el siguiente:

$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R' \left[R(X'X)^{-1}R' \right]^{-1} (q - R\hat{\beta}) \quad (1.53)$$

Prueba: Como ya hemos indicado, el estimador MCR debe ser tal que:

$$\min_{\beta} (y - X\beta)'(y - X\beta) \quad \text{sujeto a} \quad R\beta = q$$

La función lagrangiana para este problema de optimización es:

$$L = (y - X\beta)'(y - X\beta) - 2\lambda'(R\beta - q)$$

En donde λ es un vector con r Multiplicadores de Lagrange. Las condiciones necesarias para la optimización se obtienen igualando a cero las derivadas con respecto a β y λ :

$$\frac{\partial L}{\partial \beta} = -2X'y + 2X'X\hat{\beta}_R - 2R'\hat{\lambda} = 0 \quad (1.54)$$

$$\frac{\partial L}{\partial \lambda'} = R\hat{\beta}_R = q \quad (1.55)$$

A partir de (1.54) se obtiene que:


$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R'\hat{\lambda} \quad (1.56)$$

Premultiplicando todos los términos de (1.56) por R y, teniendo en cuenta (1.55) se llega a:

$$q = R\hat{\beta} + R(X'X)^{-1}R'\hat{\lambda}$$

y despejando $\hat{\lambda}$:

$$\hat{\lambda} = [R(X'X)^{-1}R']^{-1}(q - R\hat{\beta})$$

Sustituyendo este estimador en (1.56) se llega al resultado escrito en (1.57) 

A la hora de evaluar las propiedades de este estimador hay que distinguir dos situaciones según las restricciones se cumplan realmente o no. Con este fin escribiremos:

$$R\hat{\beta} - q = \delta$$

Si las restricciones se cumplen entonces $\delta = 0$ y, si no se cumplen, $\delta \neq 0$.

Comencemos suponiendo que las restricciones se cumplen.

Resultado 1.26. En el caso en el que $\delta = 0$, el estimador MCR es insesgado.

Prueba: Teniendo en cuenta (1.19), podemos escribir:

$$R\hat{\beta} = R\beta + R(X'X)^{-1}X'u = q + R(X'X)^{-1}X'u$$

Por lo tanto,

$$q - R\hat{\beta} = -R(X'X)^{-1}X'u$$

Introduciendo este resultado y el escrito en (1.19) en (1.53) obtenemos:

$$\begin{aligned}\hat{\beta}_R &= \beta + (X'X)^{-1}X'u - \\ &- (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R(X'X)^{-1}X'u] = \\ &= \beta + [I - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R](X'X)^{-1}X'u\end{aligned}\quad (1.57)$$

Recordando la Hipótesis 2 y que R es una matriz de constantes se tiene que:

$$E\hat{\beta}_R = E\beta = \beta \quad \square$$

Resultado 1.27. La matriz de varianzas y covarianzas del estimador MCR viene dada por:

$$\text{Var}(\hat{\beta}_R) = \sigma^2(X'X)^{-1}[I - R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}] \quad (1.58)$$

Además, se cumple que:

$$\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_R) = C$$

en donde C es una matriz semidefinida positiva.

Prueba: Teniendo en cuenta (1.57) se tiene que:

$$\begin{aligned}\text{Var}(\hat{\beta}_R) &= E(\hat{\beta}_R - \beta)(\hat{\beta}_R - \beta)' = \\ &= E\{[I - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R](X'X)^{-1}X'u\end{aligned}$$

$$\begin{aligned}
& u'X(X'X)^{-1} \left\{ I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \right\}' = \\
& = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1} = \\
& = \text{Var}(\hat{\beta}) - C
\end{aligned}$$

en donde $C = \sigma^2(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}$

Esta matriz es semidefinida positiva porque $(X'X)^{-1}$ es definida positiva y R tiene rango fila completo. □

También en este caso sería posible demostrar un resultado similar al de Gauss-Markov. Es decir, el estimador MCR es el mejor estimador lineal e insesgado en el sentido de que es el estimador que tiene la menor varianza dentro de la clase de los estimadores que son lineales en y y q e insesgados.

En el caso en que las restricciones no sean correctas, es decir, cuando $\delta \neq 0$, entonces el estimador MCR es sesgado aunque mantiene la misma matriz de varianzas y covarianzas.

Resultado 1.28. En el caso en que $\delta \neq 0$, entonces el estimador MCR tiene la siguiente esperanza matemática:

$$E\hat{\beta}_R = \beta + (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}\delta \quad (1.59)$$

Prueba: Basta obtener la esperanza de (1.53) teniendo en cuenta que:

$$E\hat{\beta} = \beta$$

y $E(q - R\hat{\beta}) = q - R\beta = \delta$ □

En este caso ya no podemos decir que, en general, siempre el estimador MCR es mejor que el estimador MCO. Para compararlos habría que prestar atención al error cuadrático medio.

A continuación, vamos a derivar una serie de resultados que se cumplen siempre cualquiera que sea el valor δ .

Sea \hat{u}_R el vector de residuos MCR definidos como:

$$\hat{u}_R = y - X\hat{\beta}_R \quad (1.60)$$

Resultado 1.29. La suma de los cuadrados de los residuos MCR es siempre mayor que la de los residuos MCO pudiendose escribir:

$$\hat{u}_R' \hat{u}_R = \hat{u}' \hat{u} + (\hat{\beta} - \hat{\beta}_R)' X' X (\hat{\beta} - \hat{\beta}_R) \quad (1.61)$$

Prueba: Basta aplicar el Resultado 1.3 utilizando la expresión (1.17). □

Resultado 1.30. Sea w un vector de k constantes. Entonces se tiene que:

$$\text{Var}(w' \hat{\beta}_R) \leq \text{Var}(w' \hat{\beta}) \quad (1.62)$$

Prueba: Podemos escribir:

$$\begin{aligned} \text{Var}(w' \hat{\beta}) - \text{Var}(w' \hat{\beta}_R) &= w' \{ \text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_R) \} w = \\ &= w' C w \geq 0 \end{aligned}$$

por ser C semidefinida positiva tal como se ha demostrado en el Resultado 1.27 □

Cuando se imponen r restricciones sobre β lo que se está indicando es que los k parámetros no pertenecen a un espacio de k dimensiones sino a otro de $k-r$. Esto significa que los k parámetros pueden expresarse en función de otros $k-r$ parámetros que llamaremos libres, pudiendose escribir:

$$\beta = H\alpha \quad (1.63)$$

en donde H es una matriz $k \times (k-r)$ de constantes y α es el vector de $(k-r)$ parámetros libres.

Introduciendo esta expresión en el Modelo Lineal General se tiene:

$$y = X\beta + u = XH\alpha + u = Z\alpha + u$$

El estimador MCO de α es: $\hat{\alpha} = (Z'Z)^{-1}Z'y$; el estimador restringido de β viene dado por:

$$\hat{\beta}_R = H(Z'Z)^{-1}Z'y$$

Una aplicación directa de este resultado es la siguiente: Particionemos $R\beta = 0$ como:

$$R_1\beta_1 + R_2\beta_2 = 0 \quad (1.64)$$

en donde R_1 es r.r con rango máximo y la partición de β se corresponde con la partición de R .

Apartir de (1.64) se tiene:

$$\beta_1 = -R_1^{-1}R_2\beta_2$$

y sustituyendo en el modelo particionado

$$y = X_1\beta_1 + X_2\beta_2 + u = -X_1R_1^{-1}R_2\beta_2 + X_2\beta_2 + u = X^*\beta_2 + u$$

en donde: $X^* = -X_1R_1^{-1}R_2 + X_2$

El estimador MCR toma la forma siguiente.

$$\hat{\beta}_R = \begin{bmatrix} -R_1^{-1}R_2(X^{*'}X^*)^{-1}X^{*'}y \\ (X^{*'}X^*)^{-1}X^{*'}y \end{bmatrix} \quad (1.65)$$

En el caso en que las restricciones sólo afecten a β_1 entonces el estimador MCR es:

$$\hat{\beta}_R = \begin{bmatrix} 0 \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix} \quad (1.66)$$

Ejemplo 1.1. (Continuación). Se va a estimar el modelo imponiendo la hipótesis de rendimientos constantes que puede escribirse como:

$$\beta_2 + \beta_3 = 1$$

En este caso los elementos de la expresión (1.52) son:

$$R = (0 \ 1 \ 1) \quad \text{y} \quad q = 1$$

Teniendo en cuenta que: $(q - R\hat{\beta}) = (1 - 1.0378) = -0.0378$, aplicando (1.53) se obtiene:

$$\hat{\beta}_R = \begin{pmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{2R} \\ \hat{\beta}_{3R} \end{pmatrix} = \begin{pmatrix} .01454 \\ .25413 \\ .74587 \end{pmatrix}; \hat{\sigma}_R^2 = 0.0032565$$

La matriz de varianzas y covarianzas viene dada por:

$$\text{Var}(\hat{\beta}_R) = \begin{pmatrix} 0.000399 & 0.0006692 & -0.0006692 \\ & 0.0016994 & -0.0016994 \\ & & 0.0016994 \end{pmatrix}$$

Alternativamente, estos estimadores se pueden calcular utilizando (1.65). En este sentido hay que tener en cuenta que considerando la hipótesis de rendimientos constantes el modelo puede escribirse como:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + (1 - \beta_2) x_{3t} + u_t$$

o bien

$$y_t - x_{3t} = \beta_1 x_{1t} + \beta_2 (x_{2t} - x_{3t}) + u_t$$

Por lo que las estimaciones serán:

$$\begin{pmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{2R} \end{pmatrix} = \begin{pmatrix} T & \sum (x_{2t} - x_{3t}) \\ \sum (x_{2t} - x_{3t})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum (y_t - x_{3t}) \\ \sum (y_t - x_{3t})(x_{2t} - x_{3t}) \end{pmatrix} =$$

$$= \begin{pmatrix} 24 & 9.450255 \\ 5.637141 \end{pmatrix}^{-1} \begin{pmatrix} 2.750703 \\ 1.570109 \end{pmatrix} = \begin{pmatrix} .01454 \\ .25413 \end{pmatrix}$$

Vamos a terminar este capítulo haciendo referencia al concepto de error de especificación. El análisis del error de especificación se refiere a aquellas situaciones en las que se estudian las consecuencias que, sobre las propiedades de ciertos estadísticos, tiene el considerar un modelo diferente al que genera los datos.

El análisis del error de especificación abarca una amplia gama de casos dentro de la Econometría. Nosotros prestaremos atención a dos casos particulares: la omisión de variables relevantes y la inclusión de variables superfluas.

Considerar los dos modelos siguientes:

$$y = X_1\beta_1 + u_1 \quad (1.67)$$

$$y = X_1\beta_1 + X_2\beta_2 + u \quad (1.68)$$

En donde, como ya hemos dicho, X_1 tiene k_1 y X_2 tiene k_2 regresores. Nuestro interés es estimar β_1 .

Resultado 1.31. (Variables Relevantes Omitidas). Supongamos que el modelo que ha generado los datos es el escrito en (1.68) pero que, para estimar β_1 , utilizamos el modelo escrito en (1.67). Entonces el estimador resultante es sesgado.

Prueba: Se trata de un caso particular del Resultado 1.28. El estimador MCR en este caso es:

$$\hat{\beta}_{1R} = (X_1' X_1)^{-1} X_1' y = \beta_1 + (X_1' X_1)^{-1} X_1' X_2 \beta_2 + (X_1' X_1)^{-1} X_1' u$$

Por lo tanto, su esperanza es:

$$E\hat{\beta}_{1R} = \beta_1 + (X_1' X_1)^{-1} X_1' X_2 \beta_2 \quad (1.69)$$

siendo el sesgo:



$$E\hat{\beta}_{1R} - \beta_1 = (X_1' X_1)^{-1} X_1' X_2 \beta_2$$

El signo y tamaño del sesgo dependerá del tamaño y signo de los elementos de la matriz: $(X_1' X_1)^{-1} X_1' X_2$ y de los de β_2 . Hay que tener en cuenta que los elementos de esta matriz son las estimaciones de los coeficientes de la regresión de X_2 sobre X_1 . Cada columna son los elementos de la columna correspondiente en X_2 . Por ejemplo, si sólo se excluye una variable entonces esa matriz será un vector de k_1 elementos. El primer elemento de este vector será la estimación del coeficiente de la primera variable de X_1 en la regresión de la variable excluida sobre X_1 . Si β_2 y esa estimación son del mismo signo entonces el sesgo será positivo; si tienen diferente signo el sesgo será negativo.

Hay que destacar que aunque el estimador es sesgado, su matriz de varianzas y covarianzas es “menor” que la del estimador que no comete el error de especificación tal como puede verse en el Resultado 1.27. Por lo tanto, puede haber situaciones con poca evidencia en las que puede estar justificado el uso del estimador que omite variables consideradas como relevantes.

Resultado 1.32. (Inclusión de Variables Supérfluas). Suponer que el modelo que ha generado los datos es el escrito en (1.67) pero que para estimar β_1 utilizamos el modelo escrito en (1.68). El estimador resultante es insesgado pero tiene una varianza “mayor” que la que corresponde al estimador que no comete el error de especificación.

Prueba: Utilizando el Resultado 1.20 podemos escribir:

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 y$$

en donde:

$$M_2 = I - X_2 (X_2' X_2)^{-1} X_2'$$

A partir del modelo escrito en (1.67), podemos escribir:

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 (X_1 \beta_1 + u_1) = \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 u_1$$

de forma que:

$$E\hat{\beta}_1 = \beta_1 \quad \square$$

Por lo demostrado en el Resultado 1.27 se ve que la matriz de varianzas y covarianzas de $\hat{\beta}_1$ supera a la del estimador que no comete el error de especificación.

Ejemplo 1.1. (Continuación) Vamos a aproximar el error de especificación que se comete si tratamos de estimar el coeficiente del factor trabajo sin considerar la variable capital. Es decir, se trata de estimar:

$$y_t = \beta_1 x_{1t} + \beta_3 x_{3t} + u_t$$

Las estimaciones que se obtienen son:

$$\begin{pmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{3R} \end{pmatrix} = \begin{pmatrix} T & \sum x_{3t} \\ & \sum x_{3t}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum x_{3t} y_t \end{pmatrix} = \begin{pmatrix} -1.3313 \\ 1.2913 \end{pmatrix}$$

y la estimación de la matriz de varianzas y covarianzas:

$$V_{\hat{\beta}} \begin{pmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{3R} \end{pmatrix} = \begin{pmatrix} .14043 & -.028252 \\ & .005692 \end{pmatrix}; \sigma_R^2 = 0.005294$$

Destacar que la elasticidad estimada para el trabajo es muy superior a la obtenida anteriormente . Esto puede explicarse a partir de (1.69), ya que tanto la correlación entre X_1 y X_2 como el signo de β_2 son positivos.

EJERCICIOS

1.1) Sea X una variable que se distribuye normalmente con media μ y varianza σ^2 . Suponer que se han obtenido independientemente dos muestras aleatorias simples a partir de X , de tamaños T_1 y T_2 , y con medias muestrales \bar{X}_1 y \bar{X}_2 .

1). Un investigador pretende estimar μ y propone como estimadores alternativos

$$\hat{\mu} = \frac{\bar{X}_1 + \bar{X}_2}{2} \quad \text{y} \quad \tilde{\mu} = \frac{T_1 \bar{X}_1 + T_2 \bar{X}_2}{T_1 + T_2}$$

Comparar las propiedades de ambos estimadores.

2). En una etapa posterior pretende estimar μ^2 y propone los siguientes estimadores:

$$\bar{X}_1 \bullet \bar{X}_2, \quad \left(\frac{\bar{X}_1 + \bar{X}_2}{2} \right)^2, \quad \frac{\bar{X}_1^2 + \bar{X}_2^2}{2}$$

Comparar las propiedades de los tres estimadores.

1.2) Considerar una muestra aleatoria simple de T elementos obtenida a partir de una población con media μ y varianza σ^2 . Se pide:

$$\text{Var}(X_i - \bar{X}), \quad E(X_j \bar{X}), \quad E[X_i(X_j - \bar{X})], \quad E[\bar{X}(X_j - \bar{X})]$$

$$\text{Cov}[(X_i - \bar{X}), (X_j - \bar{X})]$$

1.3). Suponer que la variable y está relacionada con la variable x según el modelo:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad u_i \sim \text{iid } N(0, \sigma^2) \quad i=1, 2, \dots, n \quad (1)$$

Un investigador agrupa las observaciones en G grupos y toma el promedio de las dos variables en los grupos considerando el modelo

$$\bar{y}_g = \beta_0 + \beta_1 \bar{x}_g + \bar{u}_g, \quad \bar{u}_g = \frac{\sum_{j=1}^{n_g} u_{gj}}{n_g}, \quad n_1 + n_2 + \dots + n_G = n \quad (2)$$

en donde $\bar{y}_g, \bar{x}_g, \bar{u}_g$ son los promedios respectivos. Se pide:

1). Derivar las propiedades de \bar{u}_g .

2). Derivar las propiedades del estimador de los mínimos cuadrados ponderados (β^*)

de β_1 a partir del modelo (2).
$$\beta^* = \frac{\sum (\frac{y_i}{z_i})(\frac{x_i}{z_i})}{\sum (\frac{x_i}{z_i})^2}.$$

3). Comparar las propiedades del estimador definido en 2) con las del estimador MCO de β_1 definido a partir de (1).

1.4). Suponer que el proceso generador de datos es:

$$y_t = \alpha + \beta x_t + u_t$$

en donde u_t es una perturbación aleatoria que cumple las hipótesis ideales.

Obtener la esperanza y varianza de los siguientes estimadores:

$$\text{a). } \hat{\beta}_1 = \frac{\sum y_t}{\sum x_t} \qquad \text{c). } \hat{\beta}_2 = \frac{\sum y_t - T\alpha}{\sum x_t}$$

$$\text{b). } \hat{\beta}_3 = \frac{\sum \frac{y_t}{x_t}}{T} \qquad \text{d). } \hat{\beta}_4 = \frac{\sum \frac{y_t^*}{x_t^*}}{T}$$

Las variables con asterisco representan desviaciones con respecto a la media.

1.5). Supongamos el modelo:

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

en donde u cumple las hipótesis habituales y suponemos que las variables están medidas en desviaciones respecto a sus medias. Los parámetros β_1 y β_2 son estimados utilizando MCO y estimadores restringidos suponiendo: $\beta_1 = \beta_2 = \beta$. Se pide:

1). Demostrar que los estimadores restringidos equivalen a una media ponderada de los estimadores MCO.

2). Demostrar que el estimador insesgado más eficiente de β coincide con la media ponderada en caso de que se cumpla la restricción.

1.6). Partiendo del modelo con dos variables $y = \beta_1 x_1 + \beta_2 x_2 + u$, en el que suponemos que las variables están medidas en desviaciones respecto a sus medias, interpretar gráficamente:

- 1). El valor estimado de y y las estimaciones MCO de β_1 y β_2 .
- 2). Un caso en que la estimación de β_1 sea mayor que uno y otro en que dicha estimación sea menor que cero.
- 3). El coeficiente de correlación múltiple.
- 4). Los coeficientes de correlación parcial.
- 5). El valor del estadístico F para el contraste de significatividad de x_2 .

1.7). Para el modelo $y = X\beta + u$ se tiene la siguiente información:

$$X'X = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} ; \quad X'y = \begin{pmatrix} 2 \\ 1 \end{pmatrix} ; \quad \sigma^2 = 1$$

Se pide:

- 1). Obtener las estimaciones MCO de β y calcular la covarianza entre

$$\hat{\beta}_1 \text{ y } \hat{\beta}_2.$$

- 2). Calcular los estimadores restringidos de β_1 y β_2 con la restricción $\beta_2 = 0$. Calcular la covarianza entre $\hat{\beta}_{1R}$ y $\hat{\beta}_{2R}$. Obtener la diferencia entre la suma de cuadrados de los residuos restringidos y la de los residuos MCO. En caso de que $\hat{\beta}_1$ y $\hat{\beta}_{1R}$ no difieran ¿tiene alguna utilidad la información de que $\beta_2 = 0$?

- 3). Interpretar gráficamente la estimación de β dado que: $\beta_2 = 0$. Repetir lo anterior para $\beta_1 + \beta_2 = 0$.

- 4). Demostrar que los estimadores restringidos de β_1 y β_2 dado:

$\beta_1 + \beta_2 = 0$ se pueden obtener a partir de la regresión de y sobre $(x_{1t} - x_{2t})$.

- 5). Obtener la predicción puntual MCO y por intervalo de y_p y $E(y_p)$ con un nivel de confianza del 95% suponiendo que para el periodo extramuestral se sabe que $x_p = (2,1)$.

1.8) Suponer el siguiente modelo:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \quad u \sim N(0, \sigma^2 I_T)$$

con:

$$(X'X) = \frac{1}{3} \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix} \quad X'y = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \sigma^2 = 1$$

- 1). Escribir las ecuaciones normales, calcular los estimadores MCO, $\hat{\beta}$, y su matriz de varianzas y covarianzas
- 2). Se estima el modelo con la restricción $\beta_2 = 0$. Calcular los estimadores restringidos, $\hat{\beta}_R$, su matriz de varianzas y covarianzas y la diferencia entre la suma de los cuadrados de los residuos restringidos y la de los residuos MCO.
- 3). Suponer que se quiere estimar por intervalo la relación $\psi'\beta$ en donde: $\psi' = (1, 0)$. Adoptando un nivel de significación del 5% se pide determinar la amplitud de los intervalos correspondientes a los estimadores MCO y restringidos.

1.9). Suponer que el modelo que genera los datos es: $y_t = \beta x_t + u_t$ en donde la perturbación cumple las hipótesis ideales. Considerar los siguientes tres estimadores:

$$\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2}, \quad \hat{\beta}_2 = \frac{\sum y_t}{\sum x_t} \quad \text{y} \quad \hat{\beta}_3 = \frac{\sum x_t^* y_t^*}{\sum x_t^{*2}}$$

en donde las variables con asterisco son desviaciones con respecto a las medias.

- 1). Obtener el sesgo y la varianza de los tres estimadores.
- 2). Demostrar que $Cov(\hat{\beta}_1, \hat{\beta}_2) = Var(\hat{\beta}_1) > 0$ y que el coeficiente de correlación entre $\hat{\beta}_1$ y $\hat{\beta}_2$ es: $\rho_{12} = [Var(\hat{\beta}_1) / Var(\hat{\beta}_2)]^{\frac{1}{2}}$. Demostrar que si se define $\hat{\beta} = \alpha \hat{\beta}_1 + (1 - \alpha) \hat{\beta}_2$ el valor de α que minimiza la varianza de este estimador es 1.
- 3). Demostrar que $Cov(\hat{\beta}_2, \hat{\beta}_3) = 0$ y que la combinación de estos dos estimadores que minimiza la varianza es $\hat{\beta} = \rho_{12}^2 \hat{\beta}_2 + (1 - \rho_{12}^2) \hat{\beta}_3 = \hat{\beta}_1$.

