PUBLICACIONES DE 4er CURSO

Grado: Economía

Asignatura: ECONOMETRÍA III

Tema 2: Estimación
Apartado 2.3: Modelo Lineal General. Estimación Restringida

Grupos:

Profesores: Antonio Aznar y M. Teresa Aparicio

Departamento de
ANÁLISIS ECONÓMICO

Curso Académico 2014/15

Facultad de
Economía y Empresa
Universidad Zaragoza

Apartado 2.3
Modelo Lineal General

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

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

$$Q_t = A\ K_t^{\alpha}\ L_t^{\beta}\ e^{u_t} \qquad t = 1, \dots, T \qquad (2.3.1)$$

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 (3.2.1) puede escribirse como:

$$y_t = \beta_1 x_{1t} + \beta_2\, x_{2t} + \beta_3\, x_{3t} + u_t \qquad (2.3.2)$$

en donde:

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

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, 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 2.3.1. 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 2.3.1. Datos para el ejemplo 1.1

OBS.	Y	K	L	LY	LK	LL
1899	100.00	100.00	100.00	4.6052	4.6052	4.6052
1900	101.00	107.00	105.00	4.6151	4.6728	4.6540
1901	112.00	114.00	110.00	4.7185	4.7362	4.7005
1902	122.00	122.00	118.00	4.8040	4.8040	4.7707
1903	124.00	131.00	123.00	4.8203	4.8752	4.8122
1904	122.00	138.00	116.0	4.8040	4.9273	4.7536
1905	143.0	149.00	125.00	4.9628	5.0039	4.8283
1906	152.00	163.00	133.00	5.0239	5.0938	4.8903
1907	151.00	176.00	138.00	5.0173	5.1705	4.9273
1908	126.00	185.00	121.00	4.8363	5.2204	4.7958
1909	155.00	198.00	140.00	5.0434	5.2883	4.9416
1910	159.00	208.00	144.00	5.0689	5.3375	4.9698
1911	153.00	216.00	145.00	5.0304	5.3753	4.9767
1912	177.00	226.00	152.00	5.1761	5.4205	5.0239
1913	184.00	236.00	154.00	5.2149	5.4638	5.0370
1914	169.00	244.00	149.00	5.1299	5.4972	5.0039
1915	189.00	266.00	154.00	5.2417	5.5835	5.0370
1916	225.00	298.00	182.00	5.4161	5.6971	5.2040
1917	227.00	335.00	196.00	5.4250	5.8141	5.2781
1918	223.00	366.00	200.00	5.4072	5.9026	5.2983
1919	218.00	387.00	193.00	5.3845	5.9584	5.2627
1920	231.00	407.00	193.00	5.4424	6.0088	5.2627
1921	179.00	417.00	147.00	5.1874	6.0331	4.9904

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 (3.2.2) 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?

Modelo e Hipótesis

El modelo econométrico que vamos a estudiar lo escribimos así:

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

En forma matricial, el modelo puede escribirse como:

$$y = X\beta + u \tag{2.3.3}$$

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 combinación 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 dicha variable. Las hipótesis que acompañan al modelo son las siguientes

Hipótesis 1. Exogeneidad Estricta

Se dice que x_j es una variable estrictamente exógena cuando la esperanza de u_t es independiente de cualquier valor de x, x_{js}, e igual a cero para todo t y s. Esta hipótesis también se escribe cómo

$$E\left(u_t / x_{js}\right) = 0 \qquad \text{s, t=1,2,…,T; j=2,3,..,k}$$

Las implicaciones más importantes de esta hipótesis son las siguientes,

1) $Eu_t = 0 \quad t = 1,2,\ldots,T$

Este resultado es consecuencia de la Ley Total de Expectativas (LTE) que toma la forma siguiente,

$$\text{LTE: } E\left[E\left(u_t / x_{js}\right)\right] = Eu_t$$

2) x_t y u_t son ortogonales porque cumplen $E\left(u_t x_{js}\right) = 0$. A este resultado se llega teniendo en cuenta que, por ser independientes, se cumple que $E\left(u_t x_{jt}\right) = Eu_t Ex_{jt}$.

La condición de ortogonalidad implica que la covarianza entre las dos variables es cero.

$$Cov(u_t, x_{js}) = E(u_t x_{js}) - Eu_t Ex_{js} = E(u_t x_{js}) = 0$$

En algunas ocasiones se establece que las variables explicativas no son estocásticas. Cuando se supone que los regresores no son estocásticos la hipótesis de exogeneidad estricta se cumple automáticamente.

Hipótesis relativas a la Parte Sistemática.

Hipótesis 2. 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 3. Asintóticamente se tiene que:

$$\frac{X'X}{T} = \begin{bmatrix} \frac{\sum x_{1t}^2}{T} & \frac{\sum x_{1t}x_{2t}}{T} & \dots & \frac{\sum x_{1t}x_{kt}}{T} \\ \frac{\sum x_{1t}x_{2t}}{T} & \frac{\sum x_{2t}^2}{T} & \dots & \frac{\sum x_{2t}x_{kt}}{T} \\ \dots & \dots & \dots & \dots \\ \frac{\sum x_{kt}x_{1t}}{T} & \frac{\sum x_{kt}x_{2t}}{T} & \dots & \frac{\sum x_{kt}^2}{T} \end{bmatrix} \rightarrow Q \qquad (2.3.4)$$

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. Este tipo de comportamiento corresponde a 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 dichos términos tiendan a expresiones finitas bien definidas.

<u>Hipótesis relativas a la Parte Aleatoria</u>

<u>Hipótesis 4</u>. Hipótesis de no Autocorrelación. No existe ninguna correlación entre perturbaciones correspondientes a periodos temporales diferentes:

$$E\, u_t u_s = 0 \qquad t \neq s \qquad s, t = 1, 2, \ldots\ldots, T$$

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.

<u>Hipótesis 5</u>. Hipótesis de Homoscedasticidad. La varianza es la misma para cada una de las perturbaciones aleatorias. Es decir,

$$E\, u_t^2 = \sigma^2 \qquad \forall t$$

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

Vectorialmente, la Hipótesis 1 puede escribirse como:

$$Eu = \begin{bmatrix} Eu_1 \\ Eu_2 \\ \vdots \\ Eu_T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \tag{2.3.5}$$

y las Hipótesis 4 y 5,

$$Var(u) = Euu' = \begin{bmatrix} Eu_1^2 & Eu_1u_2 & \dots & Eu_1u_T \\ Eu_2u_1 & Eu_2^2 & \dots & Eu_2u_T \\ \vdots & \vdots & \ddots & \vdots \\ Eu_Tu_1 & Eu_Tu_2 & \dots & Eu_T^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 I_T \tag{2.3.6}$$

<u>Hipótesis 6</u>. 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 :

$$u \sim N[0,\ \sigma^2 I_T] \tag{2.3.7}$$

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

<u>Resultado 2.3.1</u>. El vector y, de T elementos, se distribuye como:

$$y \sim N[X\beta,\ \sigma^2 I_T] \tag{2.3.8}$$

<u>Prueba</u>:A partir de la Hipótesis 6 la distribución de probabilidad de u puede escribirse como:

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

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

$$f(y) = f(u) \left|\frac{\partial u}{\partial y}\right| \tag{2.3.9}$$

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

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

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

$$f(y) = (2\pi)^{-\frac{T}{2}} \left|\sigma^2 I_T\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\} \quad (2.3.10)$$

<u>Ejemplo1.1</u>. (Continuación). Veamos el significado de las hipótesis en el marco del modelo escrito en (2.3.2). Tal como aparece en (2.3.2), 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 3, 4 y 5 se garantiza que los factores recogidos en la perturbación aleatoria no dan cuenta de nada relevante. También parece cumplirse la Hipótesis 2 ya que es difícil que exista una relación lineal exacta entre los dos inputs. Más difícil de sostener es la Hipótesis 3, ya que tanto el producto como los inputs parecen seguir una pauta creciente bien definida.

ESTIMADORES MÍNIMO CUADRÁTICO ORDINARIOS. (MCO)

El vector de estimadores MCO de β es el siguiente:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.3.11)$$

Se llama así porque minimiza la suma de cuadrados de los residuos

$$S = \sum_{1}^{T} \hat{u}_t^2 = \hat{u}'\hat{u} \tag{2.3.12}$$

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

$$\hat{u} = y - X\hat{\beta} \tag{2.3.13}$$

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.

Resultado 2.3.2. El vector de residuos MCO escrito en (2.3.13) es ortogonal con todas las colúmnas de la matriz X.

Prueba: La expresión (2.3.11) puede escribirse en forma de ecuaciones normales:

$$X'X\hat{\beta} = X'y \tag{2.3.14}$$

o, equivalentemente, como:

$$X'\left(y - X\hat{\beta}\right) = 0 \tag{2.3.15}$$

llegándose así al resultado.

Resultado 2.3.3. 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 (2.3.12).

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

$$u^* = y - X\beta^* = y - X\hat{\beta} + X(\hat{\beta} - \beta^*) = \hat{u} + X(\hat{\beta} - \beta^*) \tag{2.3.16}$$

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

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

y teniendo en cuenta (2.3.15) se llega a:

$$u^{*\prime}u^* = \hat{u}'\hat{u} + (\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) \tag{2.3.17}$$

Esta es la llamada Identidad Fundamental. Al resultado se llega teniendo en cuenta que, por la Hipótesis 2, la matriz $X'X$ es

Derivemos ahora las propiedades del vector de estimadores MCO escrito en (2.3.11).

Resultado 2.3.4. El vector de estimadores MCO es insesgado.

$$E\hat{\beta} = \beta \tag{2.3.18}$$

Prueba: Sustituyendo (2.3.3) en (2.3.11) se obtiene:

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \tag{2.3.19}$$

y teniendo en cuenta las Hipótesis 1 el resultado sigue:

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

Resultado 2.3.5. La matriz de varianzas y covarianzas del vector de estimadores MCO viene dada por:

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

(2.3.20)

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

$$Var\left(\hat{\beta}\right) = E\left(\hat{\beta} - E\hat{\beta}\right)\left(\hat{\beta}\text{-E}\hat{\beta}\right)' = E\left(\hat{\beta} - \beta\right)\left(\hat{\beta}\text{-}\beta\right)' =$$

$$E\begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \vdots \\ \hat{\beta}_k - \beta_k \end{bmatrix}\begin{bmatrix} \hat{\beta}_1 - \beta_1 & \hat{\beta}_2 - \beta_2 & \dots & \hat{\beta}_k - \beta_k \end{bmatrix} =$$

$$=$$

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

Utilizando (2.3.19) podemos escribir:

y utilizando la Hipótesis 1 y el resultado escrito en (2.3.6) se llega a:

$$\text{Var}\left(\hat{\beta}\right)=(X'X)^{-1}X'Euu'X(X'X)^{-1}=\sigma^2(X'X)^{-1}$$

llegándose así al resultado.

Resultado 2.3.6. (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:

$$\beta^* = [A + (X'X)^{-1}X']\, y \tag{2.3.21}$$

Por la propia definición, los estimadores son función lineal de los ele de y. Sustituyendo (2.3.3) en (2.3.21) 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:

$$AX = 0 \tag{2.3.22}$$

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 (2.3.22) 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.

Hasta ahora, los resultados se han derivado sin utilizar la Hipótesis 6 que trata de la normalidad de las perturbaciones aleatorias. Si añadimos esta hipótesis llegamos al siguiente resultado

Resultado 2.3.7. Suponer que se cumplen las seis hipótesis comentadas en la sección anterior. Entonces la distribución de probabilidad de $\hat{\beta}$ viene dada por:

$$\hat{\beta} \sim N[\beta, \sigma^2 (X'X)^{-1}] \tag{2.3.23}$$

Prueba: A partir de (2.3.19) 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.
Veamos ahora las propiedades de los estimadores MCO en el marco asintótico.

Resultado 2.3.8. El vector de estimadores MCO de β es consistente pudiendose escribir:

$$\text{plim } \hat{\beta} = \beta \tag{2.3.24}$$

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 3 se tiene que:

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

El sesgo cero y varianza que tiende a cero asintóticamente son condiciones suficientes para la convergencia en probabilidad.
Resultado 2.3.9. El estimador $\hat{\beta}$ es asintóticamente normal en el siguiente sentido:

$$\sqrt{T}(\hat{\beta} - \beta) \sim N(0, \sigma^2 Q^{-1}) \tag{2.3.25}$$

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

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

A partir de aquí el resultado sigue utilizando la Hipótesis 3.
También puede demostrarse que $\hat{\beta}$ es asintóticamente eficiente.

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 (2.3.13).

Resultado 2.3.10. El vector de residuos MCO, $\hat{u}$, puede escribirse como:

$$\hat{u} = Mu \tag{2.3.26}$$

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

Prueba: A partir de (2.3.13) 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 2.3.11. La esperanza y matriz de varianzas y covarianzas de $\hat{u}$ vienen dados, respectivamente, por:

$$E(\hat{u})=0 \tag{2.3.27}$$

y

$$Var(\hat{u})=\sigma^2 M \tag{2.3.28}$$

Prueba:

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

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

Resultado 2.3.12. La esperanza de la suma de los cuadrados de los residuos MCO es igual a $(T\text{-}k)\sigma^2$, es decir:

$$E(\hat{u}'\hat{u}) = \sigma^2 (T\text{-}k) \tag{2.3.29}$$

Prueba:

$$E(\hat{u}'\hat{u}) = E(u'MMu) = E(u'Mu) = E(tr(Muu')) = \sigma^2 trM = \sigma^2 (T\text{-}k)$$

por ser: tr(M) = 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:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k} = \frac{\sum_{t=1}^{T} \hat{u}_t^2}{T-k} \qquad (2.3.30)$$

Se puede demostrar fácilmente que este estimador se puede escribir como:

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{T-k}\chi^2(T-k) \qquad (2.3.31)$$

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

<u>Resultado 2.3.13</u>. Los momentos del estimador MCO de σ^2 escrito en (2.3.30) son

$$E(\hat{\sigma}^2) = \sigma^2 \qquad (2.3.32)$$

$$Var(\hat{\sigma}^2) = \frac{2\sigma^4}{T-k} \qquad (2.3.33)$$

<u>Prueba:</u> Para obtener (2.3.32) hay que tener en cuenta que: $E(\chi^2(T\text{-}k)) = T\text{-}k$.
Por último, respecto a (2.3.33) basta recordar que: $Var(\chi^2(T\text{-}k))=2(T\text{-}k)$, por lo que:

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

<u>Resultado 2.3.14</u>: El estimador $\hat{\sigma}^2$ de σ^2 es insesgado

Resultado 2.3.15: 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 (2.3.33) se tiene que:

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

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

Resultado 2.3.16: El estimador $\hat{\sigma}^2$ es asintóticamente normal con:

$$\sqrt{T}(\hat{\sigma}^2 - \sigma^2) \sim N(0, 2\sigma^4) \qquad (2.3.34)$$

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

Suponer ahora que el modelo lo escribimos como:

$$y = X_1\beta_1 + X_2\beta_2 + u \qquad (2.3.35)$$

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

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y \qquad (2.3.36)$$

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

Prueba: Primero, reescribimos las ecuaciones normales como

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

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:

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

$$(X_2' M_1 X_2)\hat{\beta}_2 = X_2' M_1 y$$

Basta premultiplicar ambos lados de esta expresión por la inversa de $(X_2' M_1 X_2)$ para llegar a (2.3.36).

Lo que indica (2.3.36) 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.

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 (2.3.10) 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:

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

y su logarítmo:

$$l(\beta, \sigma^2 / y\ , X) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \quad (2.3$$

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

<u>Resultado 2.3.18</u>. El gradiente de la función de verosimilitud definida en (2.3.37) toma la forma siguiente:

$$d(\beta, \sigma^2) = \begin{bmatrix} \dfrac{\partial l(.)}{\partial \beta} \\ \dfrac{\partial l(.)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}(X'y - X'X\beta) \\ -\dfrac{T}{2\sigma^2} + \dfrac{u'u}{2\sigma^4} \end{bmatrix} \qquad (2.3.38) - (2.3.39)$$

La matriz hessiana de segundas derivadas es:

$$H\left(\beta,\sigma^2\right)=-\begin{bmatrix} \frac{1}{\sigma^2}X'X & \frac{1}{\sigma^4}X'(y-X\beta) \\ . & -\frac{T}{2\sigma^4}+\frac{1}{\sigma^6}(y-X\beta)'\ \left(\text{y-X}\beta\right) \end{bmatrix} (2.3$$

Prueba: Teniendo en cuenta que :

$$\left(\mathrm{y}-\mathrm{X}\beta\right)'\left(\mathrm{y}-\mathrm{X}\beta\right)=\mathrm{y}'\mathrm{y}-2\mathrm{y}'\mathrm{X}\beta+\beta'\mathrm{X}'\mathrm{X}\beta$$

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 (2.3.38). Para obtener (2.3.39) basta derivar (2.3.37) respecto a σ^2.

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

Respecto a las segundas derivadas, usando (2.3.38) se tiene que:

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

La segunda derivada cruzada puede escribirse como:

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

y, por último, la segunda derivada respecto a σ^2 se obtiene usando (2.3.39):

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

<u>Resultado 2.3.19</u>. Los estimadores MV de β y σ^2 vienen dados por:

$$\tilde{\beta}=(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} \qquad (2.3.41)$$

$$\tilde{\sigma}^2=\frac{\tilde{\mathrm{u}}'\tilde{\mathrm{u}}}{\ } \qquad (2.3.42)$$

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

$$\frac{1}{\tilde{\sigma}^2}(X'y - X'X\tilde{\beta}) = 0$$

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

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

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

de donde se deriva (2.3.41).

La última relación puede escribirse como:

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

llegándose a (2.3.42).

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:

$$\tilde{\sigma}^2 = \frac{T-k}{T}\hat{\sigma}^2 \qquad (2.3.43)$$

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

Resultado 2.3.20. El estimador MV de σ^2 se distribuye como:

$$\tilde{\sigma}^2 \sim \frac{\sigma^2}{T}\chi^2(T-k) \qquad (2.3.44)$$

Su esperanza matemática es:

$$E\tilde{\sigma}^2 = \frac{T-k}{T}\sigma^2 \qquad (2.3.45)$$

y su varianza:

$$\left(\tilde{\sigma}^2\right) \quad 2(T-k)\sigma^4$$

Prueba: Teniendo en cuenta (2.3.43) basta utilizar el Resultado 2.3.13.

Teniendo en cuenta que el factor que multiplica a $\hat{\sigma}^2$ en (2.3.43) tiende a 1 conforme T tiende a infinito entonces las propiedades asintóticas 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 en el sentido definido en el apartado anterior. Primeramente, tenemos que definir el límite de la Cota Inferior de Cramer-Rao.

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

$$\mathrm{I}\left(\beta,\sigma^2\right)^{-1} = \left[-EH\left(\theta\right)\right]^{-1}$$

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

$$\lim\left(\frac{\mathrm{I}\left(\beta,\sigma^2\right)}{T}\right)^{-1} \tag{2.3.47}$$

Resultado 2.3.21. La Matriz de Información en este caso viene dada por:

$$\mathrm{I}\left(\beta,\sigma^2\right) = \frac{1}{\sigma^2}\begin{bmatrix} X'X & 0 \\ 0 & \dfrac{T}{2\sigma^2} \end{bmatrix} \tag{2.3.48}$$

y el límite escrito en (2.3.47):

$$\lim_{T\to\infty}\left(\frac{\mathrm{I}\left(\beta,\sigma^2\right)}{T}\right)^{-1} = \begin{pmatrix} \sigma^2 Q^{-1} & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \tag{2.3.49}$$

Prueba: Teniendo en cuenta la matriz de segundas derivadas escrita en (2.3.40) el resultado (2.3.48) sigue teniendo en cuenta que:

$$E\frac{1}{\sigma^4}X'\left(y - X\beta\right) = \frac{1}{\sigma^4}X'Eu = 0$$

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

$$=-\frac{T}{2\sigma^4}+\frac{T\sigma^2}{\sigma^6}=\frac{T}{2\sigma^4}$$

En lo que respecta a (2.3.49), 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.

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:

$$R\beta = q \tag{2.3.50}$$

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 2.3.22. El vector de estimadores MCR que minimiza la suma de cuadrados de los residuos cumpliendo, al mismo tiempo, las restricciones escritas en (2.3.50) es el siguiente:

$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left(q-R\hat{\beta}\right) \tag{2.3.51}$$

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

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

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 :

$$\frac{\partial L}{\partial \beta} = -2X'y + 2X'X\hat{\beta}_R - 2R'\hat{\lambda} = 0 \qquad (2.3.52)$$

$$\frac{\partial L}{\partial \lambda'} = R\hat{\beta}_R = q \qquad (2.3.53)$$

A partir de (2.3.52) se obtiene que:

$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R'\hat{\lambda} \qquad (2.3.54)$$

Premultiplicando todos los términos de (2.3.54) por R y, teniendo en cuenta (2.3.53) se llega a:

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

y despejando $\hat{\lambda}$:

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

Sustituyendo este estimador en (2.3.54) se llega al resultado escrito en (2.3.51).

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:

$$R\beta - q = \delta$$

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

Comencemos suponiendo que las restricciones se cumplen.

Resultado 2.3.23. En el caso en el que $\delta = 0$, el estimador MCR es insesgado.

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

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

Por lo tanto,

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

$$= \delta - R(X'X)^{-1}X'u \qquad (2.3.55)$$

Introduciendo este resultado en (2.3.51), se obtiene

$$\hat{\beta}_R = \hat{\beta} + A\left(q - R\hat{\beta}\right) =$$

$$\beta + (X'X)^{-1}X'u - A\delta - AR(X'X)^{-1}X'u$$

O bien

$$\hat{\beta}_R = \beta - A\delta + \left[I - AR\right]\left(X'X\right)^{-1}X'u$$

Por lo tanto,

$$E\hat{\beta}_R = \beta - A\delta$$

El estimador restringido será insesgado si $\delta = 0$ y será sesgado si es diferente de cero.

Resultado 2.3.24. La matriz de varianzas y covarianzas del estimador MCR viene dada por:

$$\mathrm{Var}(\hat{\beta}_R) = \sigma^2 (X'X)^{-1}\left[I - R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}\right] \qquad (2.3.5$$

Además, se cumple que:

$$\mathrm{Var}\left(\hat{\beta}\right) - \mathrm{Var}\left(\hat{\beta}_R\right) = C$$

en donde C es una matriz semidefinida positiva.

Prueba: Teniendo en cuenta (2.3.55) se tiene que:

$$\mathrm{Var}\left(\hat{\beta}_R\right) = E\left(\hat{\beta}_R - \beta\right)\left(\hat{\beta}_R - \beta\right)' =$$

$$E\{I - AR\}(X'X)^{-1}X'u\ u'X\left(X'X\right)^{-1}\ \{\text{I-A}R\}' =$$

$$\sigma^2\left(I - AR\right)\left(X'X\right)^{-1}\left(I - AR\right)' =$$

$$\sigma^2\left\{\left(X'X\right)^{-1} - AR\left(X'X\right)^{-1} - \left(X'X\right)^{-1}R'A' + AR\left(X'X\right)^{-1}R'$$

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

$$\mathrm{Var}(\hat{\beta}) - C$$

en donde

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

teniendo en cuenta que

$$AR\left(X'X\right)^{-1} = \left(X'X\right)^{-1}R'A' = AR\left(X'X\right)^{-1}R'A' = \frac{1}{\sigma^2}C$$

La matriz C es semidefinida positiva porque $(X'X)^{-1}$ es definida positiva y R tiene rango fila completo. $\square$

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.

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 de δ.

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

$$\hat{u}_R = y - X\hat{\beta}_R \qquad (2.3.58)$$

Resultado 2.3.25. La suma de los cuadrados de los residuos MCR es siempre mayor que la de los residuos MCO pudiendose escribir:

$$\hat{u}_R'\hat{u}_R = \hat{u}'\hat{u} + (\hat{\beta} - \hat{\beta}_R)'X'X(\hat{\beta} - \hat{\beta}_R) \qquad (2.3.59)$$

Prueba: Basta aplicar el Resultado 2.3.3 utilizando la expresión (2.3.17).

Resultado 2.3.26. Sea w un vector de k constantes. Entonces se tiene que:

$$\mathrm{Var}(w'\hat{\beta}_R) \leq \mathrm{Var}(w'\hat{\beta}) \qquad (2.3.60)$$

Prueba: Podemos escribir:

$$\mathrm{Var}(w'\hat{\beta}) - \mathrm{Var}(w'\hat{\beta}_R) = w'\{\mathrm{Var}(\hat{\beta}) - \mathrm{Var}(\hat{\beta}_R)\}w = \\ = w'Cw \geq 0$$

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

Vamos a terminar este apartado haciendo referencia al concepto de <u>error de especificación</u>. 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 supérfluas. En el Tema 3 se ampliará el tratamiento de estos puntos.
Considerar los dos modelos siguientes:

$$y = X_1\beta_1 + u_1 \qquad (2.3.60)$$

$$y = X_1\beta_1 + X_2\beta_2 + u \qquad (2.3.61)$$

En donde X_1 tiene k_1 y X_2 tiene k_2 regresores. Nuestro interes es estimar β_1.

<u>Resultado 2.3.27</u>. (Variables Relevantes Omitidas). Supongamos que el modelo que ha generado los datos es el escrito en (2.3.61) pero que, para estimar β_1, utilizamos el modelo escrito en (2.3.60). Entonces el estimador resultante es sesgado.

<u>Prueba</u>: 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:

$$E\hat{\beta}_{1R} = \beta_1 + (X'_1 X_1)^{-1} X'_1 X_2\beta_2 \qquad (2.3.62)$$

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 la 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

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 2.3.24. 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 2.3.28. (Inclusión de Variables Supérfluas). Suponer que el modelo que ha generado los datos es el escrito en (2.3.60) pero que para estimar β_1 utilizamos el modelo escrito en (2.3.61). 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 2.3.17 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 (2.3.60), 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:

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

Por lo demostrado en el Resultado 2.3.24 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.

Referencias

Aznar, A. (2012): "Curso de Econometría" Copy Center Digital. Zaragoza.