# 3 Regresión lineal múltiple: estimación y propiedades

**Ezequiel Uriel**
**Universidad de Valencia**
**Versión 09-2013**



## 3.1 El modelo de regresión lineal múltiple

El modelo de regresión lineal simple no es adecuado para modelizar muchos fenómenos económicos, ya que para explicar una variable económica se requiere en general tener en cuenta más de un factor. Veamos algunos ejemplos.

En la función keynesiana clásica el consumo se hace depender de la renta disponible como única variable relevante:

$$cons = \beta_1 + \beta_2 renta + u \tag{3-1}$$

Sin embargo, hay otros factores que pueden considerarse relevantes en el comportamiento del consumidor. Uno de esos factores podrían ser la riqueza. Con la inclusión de este factor se tendrá un modelo con dos variables explicativas:

$$cons = \beta_1 + \beta_2 inc + \beta_3 riqueza + u \tag{3-2}$$

En el análisis de la producción se utilizan a menudo las funciones potenciales, que con una especificación adecuada pueden ser transformadas (tomando logaritmos naturales, en este caso) en modelos lineales en los parámetros. Utilizando un solo input (*trabajo*), un modelo para explicar el *output* se especifica del siguiente modo:

$$\ln(output) = \beta_1 + \beta_2 \ln(trabajo) + u \tag{3-3}$$

El modelo anterior es claramente insuficiente para el análisis económico. Sería mejor utilizar el conocido modelo de Cobb-Douglas, en el que se consideran dos inputs primarios (*trabajo* y *capital*):

$$\ln(output) = \beta_1 + \beta_2 \ln(trabajo) + \beta_3 \ln(capital) + u \qquad (3\text{-}4)$$

De acuerdo con la teoría microeconómica, los costes totales (*costot*) se expresan como una función de la cantidad producida (*cantprod*). Una primera aproximación para explicar el coste total podría ser un modelo con un único regresor:

$$costot = \beta_1 + \beta_2 cantprod + u \qquad (3\text{-}5)$$

Sin embargo, es muy restrictivo considerar que, como sería el caso del modelo anterior, el coste marginal permanece constante, independientemente de la cantidad producida. En la teoría económica se propone, una función cúbica, lo que conduce al siguiente modelo econométrico:

$$costot = \beta_1 + \beta_2 cantprod + \beta_3 cantprod^2 + \beta_4 cantprod^3 + u \qquad (3\text{-}6)$$

En este caso, a diferencia de los anteriores, en el modelo sólo hay una variable explicativa, pero que da lugar a tres regresores.

Los salarios se determinan por diferentes factores. Un modelo relativamente simple para explicar los salarios en función de los años de educación y de los años de experiencia es el siguiente:

$$salarios = \beta_1 + \beta_2 educ + \beta_3 exper + u \qquad (3\text{-}7)$$

De todos modos, otros factores importantes para explicar los salarios pueden ser variables cuantitativas tales como el tiempo de formación y la edad, o variables cualitativas, como el sexo, la rama de actividad, etc.

Por último, para explicar los gastos en consumo de pescado los factores relevantes pueden ser su precio, el precio de un producto sustitutivo como la carne, y la renta disponible. Es decir:

$$gastopescado = \beta_1 + \beta_2 preciopescado + \beta_3 preciocarne + \beta_4 renta + u \qquad (3\text{-}8)$$

Por lo tanto, los ejemplos anteriores han puesto de relieve la necesidad de utilizar modelos de regresión múltiple. El tratamiento econométrico del modelo de regresión lineal simple se hizo utilizando álgebra ordinaria. El tratamiento de un modelo econométrico de dos variables explicativas mediante el uso de álgebra ordinaria es tedioso y engorroso; por otra parte, un modelo con tres variables explicativas es prácticamente intratable con esta herramienta. Por esta razón, el modelo de regresión se va a presentar utilizando álgebra matricial.

### 3.1.1 Modelo de regresión poblacional y función de regresión poblacional

En el modelo de regresión lineal múltiple, el regresando -que puede ser la variable endógena o una transformación de las variables endógenas-, es una función lineal de $k$ regresores correspondientes a las variables explicativas -o a transformaciones de las mismas- y una perturbación aleatoria o error. El modelo también incluye un término independiente. Si designamos por $y$ al regresando, por $x_2$, $x_3$,..., $x_k$ a los regresores y por $u$ al error o perturbación aleatoria, el modelo poblacional de regresión lineal múltiple vendrá dado por la siguiente expresión:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + u \qquad (3\text{-}9)$$

Los parámetros $\beta_1, \beta_2, \beta_3, \cdots, \beta_k$ son fijos y desconocidos.

En el segundo miembro de (3-9) se pueden distinguir dos componentes: un componente sistemático $\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k$ y la perturbación aleatoria *u*. Llamando $\mu_y$ al componente sistemático, podemos escribir:

$$\mu_y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k \tag{3-10}$$

Esta ecuación es conocida como función de regresión poblacional (*FRP*) o *hiperplano poblacional*. Cuando *k*=2, la *FRP* es específicamente una línea recta, cuando *k*=3, la *FRP* es específicamente un plano y, por último, cuando *k*>3, la *FRP* es denominada genéricamente hiperplano, que no es susceptible de ser representado físicamente.

De acuerdo con (3-10), $\mu_y$ es una función lineal en los parámetros $\beta_1, \beta_2, \beta_3, \cdots, \beta_k$. Ahora, supongamos que tenemos una muestra aleatoria de tamaño *n,* $\{(y_i\ , x_{2i}, x_{31}, \cdots, x_{ki}) : \ i \ = \ 1, 2, \cdots, n\}$, extraída de la población estudiada. Si expresamos el modelo poblacional para todas las observaciones de la muestra, se obtiene el siguiente el sistema:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + \beta_k x_{k2} + u_2 \\ \cdots & \qquad \cdots \qquad\qquad \cdots \quad \cdots \\ y_n &= \beta_1 + \beta_2 x_{2n} + \beta_3 x_{3n} + \cdots + \beta_k x_{kn} + u_n \end{aligned} \tag{3-11}$$

El anterior sistema de ecuaciones puede expresarse de una forma más compacta usando la notación matricial. Así, vamos a denominar

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{21} & x_{31} & \ldots & x_{k1} \\ 1 & x_{22} & x_{32} & \ldots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & x_{3n} & \ldots & x_{kn} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{bmatrix}$$

La matriz **X** es la matriz de regresores. Entre los regresores también se incluye el regresor correspondiente al término independiente. Este regresor, que a menudo se denomina regresor *ficticio*, toma el valor 1 para todas las observaciones.

El modelo de regresión lineal múltiple (3-11) expresado en notación matricial es el siguiente:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \ldots & x_{k1} \\ 1 & x_{22} & x_{32} & \ldots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & x_{3n} & \ldots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{3-12}$$

Si se tiene en cuenta las denominaciones dadas a vectores y matrices, el modelo de regresión lineal múltiple puede ser expresado de forma compacta de la siguiente manera:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3-13}$$

donde, de acuerdo con la notación utilizada, **y** es un vector $n\times 1$, **X** es una matriz $n\times k$, $\boldsymbol{\beta}$ es un vector $k\times 1$ y **u** es un vector $n\times 1$.

### 3.1.2 Función de regresión muestral

La idea básica de la regresión consiste en estimar los parámetros poblacionales $\beta_1,\beta_2,\beta_3,\cdots,\beta_k$, a partir de una muestra dada.

La *función de regresión muestral* (*FRM*) es la contrapartida de la función de regresión poblacional (*FRP*). Dado que *FRM* se obtiene de una muestra dada, una nueva muestra generará diferentes estimaciones.

La *FRM*, que es una estimación de la *FRP*, que viene dada por

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i}+\cdots+\hat{\beta}_kx_{ki}\qquad i=1,2,\cdots,n \tag{3-14}$$

nos permite calcular el *valor ajustado* ($\hat{y}_i$) correspondiente a cada $y_i$. En la *FRM,* $\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\cdots,\hat{\beta}_k$ son los estimadores de los parámetros $\beta_1,\beta_2,\beta_3,\cdots,\beta_k$.

Se denomina residuo a la diferencia entre $y_i$ e $\hat{y}_i$. Esto es

$$\hat{u}_i=y_i-\hat{y}_i=y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki} \tag{3-15}$$

En otras palabras, el residuo $\hat{u}_i$ es la diferencia entre un valor muestral y su correspondiente valor ajustado.

El sistema de ecuaciones (3-14) puede expresarse de una forma más compacta utilizando notación matricial. Así, vamos a denotar

$$\hat{\mathbf{y}}=\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\\ \cdots\\ \hat{y}_n\end{bmatrix}\qquad \hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \vdots\\ \hat{\beta}_k\end{bmatrix}\qquad \hat{\mathbf{u}}=\begin{bmatrix}\hat{u}_1\\ \hat{u}_2\\ \cdots\\ \hat{u}_n\end{bmatrix}$$

El modelo ajustado correspondiente, para todas las observaciones de la muestra, será el siguiente:

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}} \tag{3-16}$$

El vector de los residuos es igual a la diferencia entre el vector de valores observados y el vector de valores ajustados, es decir,

$$\hat{\mathbf{u}}=\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}} \tag{3-17}$$

## 3.2 Obtención de estimaciones de mínimos cuadrados, interpretación de los coeficientes, y otras características

### 3.2.1 Obtención de estimadores *MCO*

Denominando *S* a la suma de los cuadrados de los residuos,

$$S=\sum_{i=1}^{n}\hat{u}_i^2=\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right]^2 \tag{3-18}$$

para aplicar el criterio de mínimos cuadrados en el modelo de regresión lineal múltiple, calculamos la primera derivada de $S$ con respecto a cada $\hat{\beta}_j$ en la expresión (3-18):

$$\begin{aligned}
\frac{\partial S}{\partial\hat{\beta}_1}&=2\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right][-1]\\
\frac{\partial S}{\partial\hat{\beta}_2}&=2\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right][-x_{2i}]\\
\frac{\partial S}{\partial\hat{\beta}_3}&=2\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right][-x_{3i}]\\
\cdots\quad&\qquad\cdots\qquad\qquad\cdots\qquad\qquad\cdots\\
\frac{\partial S}{\partial\hat{\beta}_k}&=2\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right][-x_{ki}]
\end{aligned} \tag{3-19}$$

Los estimadores de mínimos cuadráticos se obtienen al igualar a 0 las derivadas anteriores:

$$\begin{aligned}
&\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right]=0\\
&\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right]x_{2i}=0\\
&\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right]x_{3i}=0\\
&\cdots\qquad\cdots\qquad\qquad\cdots\qquad\qquad\cdots\\
&\sum_{i=1}^{n}\left[y_i-\hat{\beta}_1-\hat{\beta}_2x_{2i}-\hat{\beta}_3x_{3i}-\cdots-\hat{\beta}_kx_{ki}\right]x_{ki}=0
\end{aligned} \tag{3-20}$$

o, con notación matricial,

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y} \tag{3-21}$$

Al sistema anterior se le denomina genéricamente sistema de ecuaciones normales del *hiperplano*.

En notación matricial ampliada, el sistema de ecuaciones normales es el siguiente:

$$\begin{bmatrix}
n & \sum_{i=1}^{n}x_{2i} & \cdots & \sum_{i=1}^{n}x_{ki}\\
\sum_{i=1}^{n}x_{2i} & \sum_{i=1}^{n}x_{2i}^2 & \cdots & \sum_{i=1}^{n}x_{2i}x_{ki}\\
\vdots & \vdots & \ddots & \vdots\\
\sum_{i=1}^{n}x_{ki} & \sum_{i=1}^{n}x_{ki}x_{2i} & \cdots & \sum_{i=1}^{n}x_{ki}^2
\end{bmatrix}
\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix}
=\begin{bmatrix}\sum_{i=1}^{n}y_i\\ \sum_{i=1}^{n}x_{2i}y_i\\ \vdots\\ \sum_{i=1}^{n}x_{ki}y_i\end{bmatrix} \tag{3-22}$$

Obsérvese que:

a) a) $\mathbf{X}'\mathbf{X}/n$ es la matriz de momentos muestrales de segundo orden, con respecto al origen, de los regresores, entre los cuales se incluye el regresor ficticio ($x_{1i}$) asociado al término independiente, que toma el valor $x_{1i}$=1 para todo $i$.

b) $\mathbf{X}'\mathbf{y}/n$ es el vector de momentos muestrales de segundo orden, con respecto al origen, entre el regresando y los regresores.

En este sistema hay $k$ ecuaciones y $k$ incógnitas $(\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\cdots,\hat{\beta}_k)$. Este sistema puede resolverse fácilmente utilizando álgebra matricial. Con el fin de resolver unívocamente el sistema (3-21) con respecto a $\hat{\boldsymbol{\beta}}$, es preciso que el rango de la matriz $\mathbf{X}'\mathbf{X}$ sea igual a $k$. Si esto se cumple, ambos miembros de (3-21) pueden ser premultiplicados por $\left[\mathbf{X}'\mathbf{X}\right]^{-1}$:

$$\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y}$$

obteniéndose la expresión del vector de estimadores de mínimos cuadrados, o más exactamente, el vector de estimadores de mínimos cuadrados ordinarios (*MCO*), porque $\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{X}=\mathbf{I}$. Por lo tanto, la solución es la siguiente:

$$\begin{bmatrix}\hat{\beta}_1\\\hat{\beta}_2\\\vdots\\\hat{\beta}_k\end{bmatrix}=\hat{\boldsymbol{\beta}}=\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y} \tag{3-23}$$

Como la matriz de segundas derivadas, $2\mathbf{X}'\mathbf{X}$, es una matriz definida positiva, la conclusión es que $S$ presenta un mínimo en $\hat{\boldsymbol{\beta}}$.

### 3.2.2 Interpretación de los coeficientes

El coeficiente $\hat{\beta}_j$ mide el efecto *parcial* del regresor $x_i$, manteniendo los otros regresores fijos. Vamos a ver el significado de esta expresión.

El modelo estimado para la observación $i$-ésima viene dado por

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i}+\cdots+\hat{\beta}_jx_{ji}+\cdots+\hat{\beta}_kx_{ki} \tag{3-24}$$

Consideremos ahora el modelo estimado para la observación $h$-ésima, en el que los valores de las variables explicativas y, en consecuencia, de $y$ habrán cambiado con respecto a la (3-24):

$$\hat{y}_h=\hat{\beta}_1+\hat{\beta}_2x_{2h}+\hat{\beta}_3x_{3h}+\cdots+\hat{\beta}_jx_{jh}+\cdots+\hat{\beta}_kx_{kh} \tag{3-25}$$

Restando (3-25) de (3-24), tenemos que

$$\Delta\hat{y}=\hat{\beta}_2\Delta x_2+\hat{\beta}_3\Delta x_3+\cdots+\hat{\beta}_j\Delta x_j+\cdots+\hat{\beta}_k\Delta x_k \tag{3-26}$$

donde $\Delta\hat{y}=\hat{y}_i-\hat{y}_h,\Delta x_2=x_{2i}-x_{2h},\Delta x_3=x_{3i}-x_{3h},\cdots\Delta x_k=x_{ki}-x_{kh}$.

La expresión anterior capta la variación de $\hat{y}$ debida a cambios en todos los regresores. Si sólo cambia $x_j$, tendremos que

$$\Delta\hat{y} = \hat{\beta}_j \Delta x_j \tag{3-27}$$

Si $x_k$ se incrementa en una unidad, tenemos

$$\Delta\hat{y} = \hat{\beta}_j \quad \text{for } \Delta x_j = 1 \tag{3-28}$$

En consecuencia, el coeficiente $\hat{\beta}_j$ mide el cambio en $y$ cuando $x_j$ aumenta en 1 unidad, *manteniendo fijos los regresores* $x_2, x_3, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k$. Es muy importante en la interpretación de los coeficientes tener en cuenta esta cláusula *ceteris paribus*.

Esta interpretación no es válida, por supuesto, para el término independiente.

***EJEMPLO 3.1 Cuantificando la influencia de la edad y del salario sobre el absentismo en la empresa Buenosaires***

Buenosaires es una empresa dedicada a la fabricación de ventiladores, habiendo tenido resultados relativamente aceptables en los últimos años. Los directivos consideran, sin embargo, que los resultados habrían sido mejores si el absentismo en la empresa no fuera tan alto. Para este propósito, el modelo que se propone es el siguiente:

$$absent = \beta_1 + \beta_2 age + \beta_3 tenure + \beta_4 wage + u$$

donde la ausencia, *absent,* se mide en días por año, el salario, *wage,* en miles de euros al año; los años trabajados en la empresa, *tenure,* y la edad, *age*, se expresan en años.

Utilizando una muestra de tamaño 48 (fichero *absent*), se ha estimado la siguiente ecuación:

$$\widehat{absent} = \underset{(1.603)}{14.413} - \underset{(0.048)}{0.096}\, age - \underset{(0.067)}{0.078}\, tenure - \underset{(0.007)}{0.036}\, wage$$

$$R^2 = 0.694 \qquad n = 48$$

La interpretación de $\hat{\beta}_2$ es la siguiente: manteniendo fijo el salario y los años trabajados en la empresa, si la edad se incrementa en un año, el absentismo laboral se reducirá en 0.096 días al año. La interpretación de $\hat{\beta}_3$ es como sigue: manteniendo fijo el salario y la edad, si los años trabajados en la empresa se incrementan en un año, el absentismo laboral se reducirá en 0.078 días al año. Finalmente, la interpretación de $\hat{\beta}_4$ es la siguiente: manteniendo fija la edad y los años trabajados en la empresa, si el salario se incrementa en 1000 euros al año, el absentismo laboral se reducirá en 0.036 días por año.

***EJEMPLO 3.2 Demanda de servicios hoteleros***

Para explicar la demanda de servicios hoteleros se formuló el siguiente modelo:

$$\ln(hostel) = \beta_1 + \beta_2 \ln(inc) + \beta_3 hhsize + u \tag{3-29}$$

donde *hostel* es el gasto en servicios hoteleros e *inc* es la renta disponible; ambas variables están expresadas en euros por mes. La variable *hhize* es el número de miembros del hogar.

La ecuación estimada con una muestra de 40 hogares, utilizando el fichero *hostel*, es la siguiente:

$$\widehat{\ln(hostel_i)} = -27.36 + 4.442 \ln(inc_i) - 0.523 hhsize_i$$

$$R^2 = 0.738 \qquad n = 40$$

A la vista de estos resultados, podemos decir que los servicios hoteleros son un bien de lujo, ya que la elasticidad de la demanda/renta para este bien es muy alta (4.44). Esto significa que si la renta se incrementa en un 1%, el gasto en servicios hoteleros se incrementará un 4.44%, manteniendo fijo el tamaño de la familia. Por otro lado, si el tamaño del hogar aumenta en un miembro, entonces el gasto en servicios hoteleros disminuirá en un 52%.

***EJEMPLO 3.3 Una regresión hedónica para coches***

El modelo hedónico de medición de precios se basa en el supuesto de que el valor de un bien depende del valor de sus diferentes características. Así, el precio de un coche dependerá del valor que el comprador asigne a sus atributos: cualitativos (por ejemplo, cambio automático, potencia, diesel, dirección asistida, aire acondicionado), y cuantitativos (por ejemplo, consumo de combustible, peso, etc.).

La base de datos para este ejercicio es el fichero *hedcarsp* (precios hedónicos de los coches en España) y cubre los años 2004 y 2005. Un primer modelo basado sólo en atributos cuantitativos es el siguiente:

$$\ln(price) = \beta_1 + \beta_2 volume + \beta_3 fueleff + u$$

donde *volume* es longitud×anchura×altura en $m^3$ y *fueleff* es la *ratio* litros por 100 km/caballos de vapor expresada en porcentaje.

La ecuación estimada con una muestra de 214 observaciones es la siguiente:

$$\widehat{\ln(price)}_i = \underset{(0.151)}{14.97} + \underset{(0.009)}{0.0956}\, volume_i - \underset{(0.010)}{0.1608}\, fueleff_i$$

$$R^2=0.765 \qquad n=214$$

La interpretación de $\hat{\beta}_2$ y $\hat{\beta}_3$ es la siguiente. Manteniendo *fueleff* fijo, si aumenta *volume* en 1 $m^3$, el precio de los coches se incrementarán en un 9.56%. Manteniendo fijo *volume*, si la ratio litros por 100 km/caballos de vapor aumenta en un punto porcentual, el precio de los automóviles se reducirá en un 16,08%.

***EJEMPLO 3.4. Ventas y publicidad: el caso de Lydia E. Pinkham***

En este caso se va a estimar un modelo con datos de series temporales con objeto de medir el efecto que puedan tener los gastos de publicidad, realizados a lo largo de distintos períodos de tiempo, sobre las ventas del momento actual. Designando por $V_t$ y $P_t$ a las ventas y a los gastos en publicidad, realizados en el momento *t*, el modelo planteado inicialmente para explicar las ventas, en función de los gastos en publicidad presentes y pasados, es el siguiente:

$$V_t = \alpha + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 P_{t-2} + \cdots + u_t \tag{3-30}$$

En la expresión anterior los puntos suspensivos indican que los gastos en publicidad realizados en el pasado siguen ejerciendo influencia de forma indefinida, aunque, se supone, que con un impacto decreciente sobre las ventas del momento actual. Naturalmente, el modelo anterior no es operativo, ya que tiene un número indefinido de coeficientes. Para solucionar el problema se pueden adoptar, en principio, dos enfoques. El primer enfoque consiste en fijar *a priori* el número máximo de períodos durante los cuales la publicidad mantiene sus efectos sobre las ventas. En el segundo enfoque, se postula que los coeficientes se comportan de acuerdo con alguna ley que permite determinar su valor en función de un número reducido de parámetros, posibilitando además una ulterior simplificación.

En el primer enfoque el problema que surge es que en general no existen criterios precisos e información suficiente que permitan la fijación *a priori* del número máximo de períodos. Por esta razón, vamos a ver un caso particular del segundo enfoque que tiene gran interés por la plausibilidad del supuesto y su fácil aplicación. El caso que vamos a considerar consiste en establecer que los coeficientes $\beta_i$ disminuyen de forma geométrica a medida que nos alejamos hacia atrás en el tiempo según el esquema:

$$\beta_i = \beta_1 \lambda^i \quad \forall i \qquad 0 < \lambda < 1 \tag{3-31}$$

A la anterior transformación se le denomina transformación de Koyck, ya que fue este autor el que la introdujo en 1954 para el estudio de la inversión.

Sustituyendo (3-31) en (3-30), se obtiene que

$$V_t = \alpha + \beta_1 P_t + \beta_1 \lambda P_{t-1} + \beta_1 \lambda^2 P_{t-2} + \cdots + u_t \tag{3-32}$$

El modelo anterior sigue teniendo infinitos términos, pero sólo tres parámetros y además se puede simplificar. En efecto, si expresamos la ecuación (3-32) para el período *t*-1 y multiplicamos ambos miembros por $\lambda$ se obtiene que

$$\lambda V_{t-1} = \alpha\lambda + \beta_1 \lambda P_{t-1} + \beta_1 \lambda^2 P_{t-2} + \beta_1 \lambda^3 P_{t-3} + \cdots + \lambda u_{t-1} \tag{3-33}$$

Restando miembro a miembro (3-33) de (3-32), y teniendo en cuenta que los factores $\lambda^i$ tienden a 0 al tender *i* a infinito, se llega al siguiente resultado:

$$V_t = \alpha(1-\lambda) + \beta_1 P_t + \lambda V_{t-1} + u_t - \lambda u_{t-1} \tag{3-34}$$

El modelo ha quedado simplificado de forma que solamente tiene tres regresores, aunque, a cambio, se ha pasado a un término de perturbación compuesto. Antes de ver la aplicación de este modelo se va analizar el significado del coeficiente $\lambda$ y el problema de la duración de los efectos de los gastos en publicidad sobre las ventas. El parámetro $\lambda$ es la tasa a que decaen los efectos de los gastos en publicidad presentes sobre las ventas presentes y futuras. Los efectos acumulados del gasto en publicidad de una unidad monetaria sobre las ventas después de transcurridos $m$ períodos vienen dados por

$$\beta_1(1+\lambda+\lambda^2+\lambda^3+\cdots+\lambda^m) \tag{3-35}$$

Para calcular la suma acumulada de los efectos, dada en (3-35), vamos a tener en cuenta que esta expresión es la suma de los términos de una progresión geométrica[1], con lo que se puede expresar de la siguiente forma:

$$\frac{\beta_1(1-\lambda^m)}{1-\lambda} \tag{3-36}$$

Cuando $m$ tiende a infinito, entonces la suma de los efectos acumulados viene dada por

$$\frac{\beta_1}{1-\lambda} \tag{3-37}$$

Una cuestión interesante es determinar cuántos períodos de tiempo se requieren para que se obtenga el $p$% (por ejemplo, el 50%) del efecto total. Designando por $h$ el número de períodos requeridos para obtener dicho porcentaje, se puede establecer que

$$p=\frac{\text{Efecto en } h \text{ periodos}}{\text{Efecto total}}=\frac{\dfrac{\beta_1(1-\lambda^h)}{1-\lambda}}{\dfrac{\beta_1}{1-\lambda}}=1-\lambda^h \tag{3-38}$$

Fijado $p$ se puede calcular $h$ de acuerdo con (3-38). En efecto, despejando $h$ en esta expresión se obtiene que

$$h=\frac{\ln(1-p)}{\ln\lambda} \tag{3-39}$$

Este modelo lo utilizó Kristian S. Palda en su tesis doctoral publicada en 1964, titulada *The Measurement of Cumulative Advertising Effects*, para analizar los efectos acumulados de los gastos en publicidad, en el caso de la compañía Lydia E. Pinkham. Este caso, así como el estudio de Palda, han sido la base a partir de la cual se ha desarrollado la investigación de los efectos de los gastos en publicidad. Vamos a ver a continuación algunas características de este caso:

1) La Lydia E. Pinkham Medicine Company fabricaba desde 1873 un extracto de hierbas diluido en una solución alcohólica. Este producto se anunciaba inicialmente como un analgésico y también como un remedio para una enorme variedad de enfermedades.

2) En general, en los distintos tipos de productos suele haber competencia entre distintas marcas, como pueda ser el caso paradigmático de la Coca-Cola y la Pepsi-Cola en el campo de las colas. Cuando esto ocurre, para analizar los efectos de los gastos en publicidad hay tener en cuenta el comportamiento de los principales competidores. Lydia E. Pinkham tenía la ventaja de no tener competidores, y de que, en su línea de producto, actuaba en la práctica como monopolista.

3) Otra característica del caso Lydia E. Pinkham era que la mayor parte de los gastos de distribución se asignaban a la publicidad, ya que la compañía no tenía agentes comerciales, siendo muy elevada la relación gastos en publicidad/ventas.

4) El producto pasó a lo largo del tiempo por distintos avatares. Así, en 1914 la Food and Drug Administration (organismo de los Estados Unidos que establece los controles para los productos

---

[1] Designando por $a_p$, $a_u$ y $r$ al primer término, al último término y a la razón respectivamente, la suma de los términos de una progresión geométrica convergente viene dada por

$$\frac{a_p-a_u}{1-r}$$

alimenticios y los medicamentos) le acusó de publicidad engañosa por lo que tuvo que cambiar sus mensajes publicitarios. También la Internal Revenue (oficina de impuestos) le amenazó con aplicarle una tasa sobre bebidas alcohólicas, ya que el contenido alcohólico del producto era del 18%. Por todos estos motivos se produjeron cambios en la presentación y contenido durante el período 1915-1925. En 1925 la Food and Drug Administration prohibió que el producto se anunciara como medicina, pasando a distribuirse como bebida tónica. En el período 1926-1940 se incrementaron considerablemente los gastos en publicidad para después decaer.

La estimación del modelo (3-34) con datos desde 1907 a 1960, recogidos en el fichero *pinkham*, es la siguiente:

$$\widehat{ventas}_t = \underset{(95.7)}{138.7} + \underset{(0.156)}{0.3288}\, gpub_t + \underset{(0.0915)}{0.7593}\, ventas_{t-1}$$

$$R^2 = 0.877 \qquad n = 53$$

La suma de los efectos acumulados de los gastos en publicidad sobre las ventas se obtiene aplicando la fórmula (3-37):

$$\frac{\hat{\beta}_1}{1-\hat{\lambda}} = \frac{0.3288}{1-0.7593} = 1.3660$$

De acuerdo con dicho resultado, por cada unidad monetaria adicional gastada en publicidad se obtiene un efecto acumulado total en las ventas de 1.366 unidades monetarias. Dado que es importante no solo determinar el efecto total, sino también como se distribuyen estos efectos a lo largo del tiempo, vamos a contestar ahora a la siguiente pregunta: ¿Cuantos períodos de tiempo se requieren para alcanzar la mitad de los efectos totales? Aplicando la fórmula (3-39) para el caso de $p$=0,5, se obtiene el siguiente resultado:

$$\hat{h}(0.5) = \frac{\ln(1-0.5)}{\ln(0.7593)} = 2.5172$$

### 3.2.3 Implicaciones algebraicas de la estimación

Las implicaciones algebraicas de la estimación se derivan exclusivamente de la aplicación del método de *MCO* al modelo de regresión lineal múltiple:

1. *La suma de los residuos de MCO es igual a 0*:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \tag{3-40}$$

De la definición de residuos

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \cdots - \hat{\beta}_k x_{ki} \qquad i = 1, 2, \cdots, n \tag{3-41}$$

Si sumamos para las $n$ observaciones, obtenemos:

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} - \cdots - \hat{\beta}_k \sum_{i=1}^{n} x_{ki} \tag{3-42}$$

Por otro lado, la primera ecuación del sistema de ecuaciones normales (3-20) es igual a

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} - \cdots - \hat{\beta}_k \sum_{i=1}^{n} x_{ki} = 0 \tag{3-43}$$

Si comparamos (3-42) y (3-43), llegamos a la conclusión de que (3-40) se cumple.

Tenga en cuenta que, si (3-40) se cumple, esto implica que

$$\sum_{i=1}^{n} y = \sum_{i=1}^{n} \hat{y}_i \tag{3-44}$$

y, dividiendo (3-40) y (3-44) por *n,* obtenemos

$$\overline{\hat{u}} = 0 \qquad \overline{y} = \overline{\hat{y}} \tag{3-45}$$

2. *El hiperplano MCO pasa siempre a través del punto de medias muestrales* $\left(\overline{y}, \overline{x}_2, \cdots, \overline{x}_k\right)$.

En efecto, dividiendo la ecuación (3-43) por $n$ se tiene que:

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{x}_2 + \cdots + \hat{\beta}_k \overline{x}_k \tag{3-46}$$

*3. El producto cruzado muestral entre cada uno de los regresores y los residuos MCO es cero*.

Es decir,

$$\sum_{i=1}^{n} x_{ji} \hat{u}_i = 0 \qquad\qquad j = 2, 3, \cdots, k \tag{3-47}$$

Utilizando las últimas $k$ ecuaciones normales (3-20) y teniendo en cuenta que, por definición $\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i} - \cdots - \hat{\beta}_k x_{ki}$, podemos ver que

$$\begin{aligned} \sum_{i=1}^{n} \hat{u}_i x_{2i} &= 0 \\ \sum_{i=1}^{n} \hat{u}_i x_{3i} &= 0 \\ \cdots \qquad &\cdots \\ \sum_{i=1}^{n} \hat{u}_i x_{ki} &= 0 \end{aligned} \tag{3-48}$$

4. *El producto cruzado muestral entre los valores ajustados (* $\hat{y}$ *) y los residuos MCO es cero.*

Es decir,

$$\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0 \tag{3-49}$$

Teniendo en cuenta (3-40) y (3-48), obtenemos

$$\begin{aligned} \sum_{i=1}^{n} \hat{y}_i \hat{u}_i &= \sum_{i=1}^{n} (\hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}) \hat{u}_i = \hat{\beta}_1 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} \hat{u}_i + \cdots \hat{\beta}_k \sum_{i=1}^{n} x_{ki} \hat{u}_i \\ &= \hat{\beta}_1 \times 0 + \hat{\beta}_2 \times 0 + \cdots \hat{\beta}_k \times 0 = 0 \end{aligned} \tag{3-50}$$

## 3.3 Supuestos y propiedades estadísticas de los estimadores de *MCO*

Antes de estudiar las propiedades estadísticas de los estimadores de *MCO* en el modelo de regresión lineal múltiple*,* necesitamos formular un conjunto de supuestos estadísticos. Específicamente, al conjunto de supuestos que vamos a formular se les denomina *supuestos del modelo lineal clásico* (*MLC*). Es importante destacar que los supuestos estadísticos del *MLC* son muy simples, y que los estimadores de *MCO* tienen, bajo estos supuestos, muy buenas propiedades.

### 3.3.1 Supuestos estadísticos del *MLC* en la regresión lineal múltiple

***a) Supuesto sobre la forma funcional***

*1) La relación entre el regresando, los regresores y el error es lineal en los parámetros:*

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{3-51}$$

o, alternativamente, para todas las observaciones,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3-52}$$

***b) Supuestos sobre los regresores***

*2) Los valores de $x_2, x_3, \cdots, x_k$ son fijos en repetidas muestras, o la matriz* **X** *es fija en repetidas muestras*:

Este es un supuesto fuerte en el caso de las ciencias sociales, donde, en general, no es posible experimentar. Una hipótesis alternativa puede formularse así:

*2*) Los regresores $x_2, x_3, \cdots, x_k$ se distribuyen independientemente de la perturbación aleatoria.* Formulado de otra manera, **X** se distribuye de forma independiente del vector de perturbaciones aleatorias, lo que implica que $E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$

Como hicimos en el capítulo 2, vamos a adoptar también el supuesto 2).

*3) La matriz de regresores,* **X***, no contiene errores de medición.*

*4) La matriz de regresores*, **X**, *tiene rango igual a k*:

$$\rho(\mathbf{X}) = k \tag{3-53}$$

Recordemos que la matriz de regresores contiene *k* columnas, correspondientes a los *k* regresores del modelo, y *n* filas, correspondientes al número de observaciones. El supuesto 4 tiene dos implicaciones:

1. El número de observaciones, *n*, debe ser igual o mayor que el número de regresores, *k*. Intuitivamente, esto tiene sentido: para estimar *k* parámetros, necesitamos al menos *k* observaciones
2. Cada regresor debe ser linealmente independiente, lo que implica que no existen relaciones lineales exactas entre los regresores. Si un regresor es una combinación lineal exacta de otros regresores, entonces se dice que hay *multicolinealidad perfecta*, y el modelo no puede estimarse.

Si existe una relación lineal aproximada - es decir, no una relación exacta -, entonces se pueden estimar los parámetros, aunque su fiabilidad se verá afectada. En este caso, se dice que existe *multicolinealidad no perfecta*.

***c) Supuesto sobre los parámetros***

*5) Los parámetros $\beta_1, \beta_2, \beta_3, \cdots, \beta_k$ son constantes, o* $\boldsymbol{\beta}$ *es un vector constante*

***d) Supuestos sobre la perturbación aleatoria***

*6) La media de las perturbaciones es cero,*

$$E(u_i)=0, \quad i=1,2,3,\ldots,n \quad \text{o} \quad E(\mathbf{u})=\mathbf{0} \tag{3-54}$$

*7) Las perturbaciones tienen una varianza constante (supuesto de homoscedasticidad)*:

$$var(u_i)=\sigma^2 \quad i=1,2,\cdots n \tag{3-55}$$

*8) Las perturbaciones con diferentes subíndices no están correlacionadas entre sí (supuesto de no autocorrelación)*:

$$E(u_i u_j)=0 \quad i\neq j \tag{3-56}$$

La formulación de los supuestos de homoscedasticidad y no autocorrelación permite especificar la matriz de covarianzas del vector de perturbaciones:

$$\begin{aligned}
&E\left[[\mathbf{u}-E(\mathbf{u})][\mathbf{u}-E(\mathbf{u})]'\right]=E\left[[\mathbf{u}-\mathbf{0}][\mathbf{u}-\mathbf{0}]'\right]=E\left[[\mathbf{u}][\mathbf{u}]'\right]\\
&=E\left[\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}\begin{bmatrix}u_1 & u_2 & \cdots & u_n\end{bmatrix}\right]=E\begin{bmatrix}u_1^2 & u_1u_2 & \cdots & u_1u_n\\ u_2u_1 & u_2^2 & \cdots & u_2u_n\\ \vdots & \vdots & \ddots & \vdots\\ u_nu_1 & u_nu_2 & \cdots & u_n^2\end{bmatrix}\\
&=\begin{bmatrix}E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n)\\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n)\\ \vdots & \vdots & \ddots & \vdots\\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2)\end{bmatrix}=\begin{bmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2\end{bmatrix}
\end{aligned} \tag{3-57}$$

Para obtener la última igualdad se ha tenido en cuenta que la varianza de cada elemento es constante e igual a $\sigma^2$, de acuerdo con (3-55), y que la covarianza entre cada par de elementos es 0, de acuerdo con (3-56).

El resultado anterior puede expresarse de forma compacta del siguiente modo:

$$E(\mathbf{uu}')=\sigma^2\mathbf{I} \tag{3-58}$$

A la matriz dada en (3-58) se le denomina matriz e*scalar*, puesto que es un escalar ($\sigma^2$, en este caso) multiplicado por la matriz identidad.

*9) La perturbación u tiene una distribución normal*

Teniendo en cuenta los supuestos 6 a 9, tenemos que

$$u_i \sim NID(0,\sigma^2) \quad i=1,2,\cdots,n \qquad \text{o} \qquad \mathbf{u}\sim N(\mathbf{0},\sigma^2\mathbf{I}) \tag{3-59}$$

donde el *NID* significa que la perturbación está *normal e independiente distribuida*.

### 3.3.2 Propiedades estadísticas del estimador de *MCO*

Bajo los supuestos del *MLC*, el estimador de *MCO* poseen buenas propiedades. En las demostraciones de este apartado implícitamente se tendrán en cuenta siempre los supuestos 3, 4 y 5.

***Linealidad e insesgadez del estimador de MCO***

Ahora, vamos a demostrar que el estimador de *MCO* es linealmente insesgado. En primer lugar expresaremos $\hat{\boldsymbol{\beta}}$ como una función del vector **u**, utilizando el supuesto 1, de acuerdo con (3-52):

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\left[\mathbf{X}\boldsymbol{\beta}+\mathbf{u}\right] = \boldsymbol{\beta}+\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{u} \qquad (3\text{-}60)$$

El estimador de *MCO* puede expresarse del siguiente modo con el fin de ver de forma más clara la propiedad de linealidad:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}+\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{u} = \boldsymbol{\beta}+\mathbf{A}\mathbf{u} \qquad (3\text{-}61)$$

donde $\mathbf{A} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'$ es fija bajo el supuesto 2. Así pues, $\hat{\boldsymbol{\beta}}$ es una función lineal de **u** y, consecuentemente, es un estimador *lineal*.

Tomando las esperanzas en (3-60) y aplicando el supuesto 6, se obtiene

$$E\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}+\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'E\left[\mathbf{u}\right] = \boldsymbol{\beta} \qquad (3\text{-}62)$$

Por lo tanto, $\hat{\boldsymbol{\beta}}$ es un estimador *insesgado*.

***Varianza del estimador de MCO***

Para calcular la matriz de covarianzas de $\hat{\boldsymbol{\beta}}$ son necesarios los supuestos 7 y 8, además de los 6 primeros:

$$\begin{aligned}\mathrm{var}(\hat{\boldsymbol{\beta}}) &= E\left[\hat{\boldsymbol{\beta}}-E(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}}-E(\hat{\boldsymbol{\beta}})\right]' = E\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]\left[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right]' \\ &= E\left[\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\right] = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1} \\ &= \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'E(\sigma^2\mathbf{I})\mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1} = \sigma^2\left[\mathbf{X}'\mathbf{X}\right]^{-1}\end{aligned} \qquad (3\text{-}63)$$

En el tercer paso de la demostración anterior se ha tenido en cuenta que, de acuerdo con (3-60), $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{u}$. El supuesto 2 se ha tenido en cuenta en el cuarto paso. Finalmente, los supuestos 7 y 8 se han utilizado en el último paso.

Por lo tanto, $\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2\left[\mathbf{X}'\mathbf{X}\right]^{-1}$ es la matriz de covarianzas del vector $\hat{\boldsymbol{\beta}}$. En esta matriz de covarianzas, la varianza de cada elemento $\hat{\beta}_j$ aparece en la diagonal principal, mientras que las covarianzas entre cada par de elementos se encuentran fuera de la diagonal principal. Específicamente, la varianza de $\hat{\beta}_j$ (para $j$=2,3,…,$k$) es igual a $\sigma^2$ multiplicada por el elemento correspondiente de la diagonal principal de $\left[\mathbf{X}'\mathbf{X}\right]^{-1}$. Después de operar, la varianza de $\hat{\beta}_j$ puede expresarse como

$$\mathrm{var}(\hat{\beta}_j) = \frac{\sigma^2}{nS_j^2(1-R_j^2)} \qquad (3\text{-}64)$$

donde $R_j^2$ es el $R$ cuadrado de la regresión de cada $x_j$ sobre el resto de regresores, $n$ es el tamaño de la muestra y $S_j^2$ es la varianza muestral del regresor $x_j$.

La fórmula (3-64) es válida para todos los coeficientes de pendiente, pero no para el término independiente.

A la raíz cuadrada de (3-64) se le denomina *desviación estándar* (*de*) de $\hat{\beta}_j$:

$$de(\hat{\beta}_j) = \frac{\sigma}{\sqrt{nS_j^2(1-R_j^2)}} \tag{3-65}$$

***Los estimadores de MCO son ELIO***

Bajo los supuestos 1 a 8 del *MLC*, que son denominados supuestos de Gauss-Markov, los estimadores de *MCO* son *estimadores lineales insesgados y óptimos (ELIO)*.

El teorema de Gauss Markov establece que los estimadores *MCO* son estimadores óptimos dentro la clase de los estimadores lineales insesgados. En este contexto *óptimo*, significa que es un estimador con la varianza más pequeña para un determinado tamaño de muestra. Vamos ahora a comparar la varianza de un elemento de $\hat{\boldsymbol{\beta}}$ ($\hat{\beta}_j$), con cualquier otro estimador $\tilde{\beta}_j$ que sea lineal (tal que $\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$) e insesgado (de forma que los pesos, $w_j$, deben cumplir algunas restricciones). La propiedad de que $\hat{\beta}_j$ es un estimador *ELIO* tiene las siguientes implicaciones al comparar su varianza con la varianza de $\tilde{\beta}_j$:

1) La varianza de un coeficiente $\tilde{\beta}_j$ o es mayor que, o igual a, la varianza de $\hat{\beta}_j$ obtenido por *MCO*:

$$\text{var}(\tilde{\beta}_j) \geq \text{var}(\hat{\beta}_j) \qquad \text{j} = 1, 2, \ldots, k \tag{3-66}$$

2) La varianza de cualquier combinación lineal de $\tilde{\beta}_j$ es mayor que, o igual a, la varianza de la correspondiente combinación lineal de $\hat{\beta}_j$.

En el apéndice 3.1 puede verse la demostración del teorema de Gauss-Markov.

***Estimador de la varianza de la perturbación***

Teniendo en cuenta el sistema de ecuaciones normales (3-20), si conocemos *n–k* residuos, podemos obtener los otros k residuos utilizando las restricciones que impone ese sistema a los residuos.

Por ejemplo, la primera ecuación normal nos permite obtener el valor de $\hat{u}_n$ en función de los residuos restantes:

$$\hat{u}_n = -\hat{u}_1 - \hat{u}_2 - \cdots - \hat{u}_{n-1}$$

Por lo tanto, sólo hay *n-k* grados de libertad en los residuos de *MCO*, a diferencia de los *n* grados de libertad en las perturbaciones. Recuerde que los grados de libertad se definen como la diferencia entre el número de observaciones y el número de parámetros estimados.

El estimador insesgado de $\sigma^2$ se ajusta teniendo en cuenta los grados de libertad:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k} \tag{3-67}$$

Bajo los supuestos 1 a 8, se obtiene que

$$E(\hat{\sigma}^2) = \sigma^2 \tag{3-68}$$

Véase el apéndice 3.2 para la demostración.

A la raíz cuadrada de (3-67), $\hat{\sigma}$ se le denomina *error estándar de la regresión* y es un estimador de. $\sigma$

***Estimadores de la varianzas de*** $\hat{\boldsymbol{\beta}}$ **y** ***del coeficiente de pendiente*** $\hat{\beta}_j$

El estimador de la matriz de covarianzas de $\hat{\boldsymbol{\beta}}$ viene dado por

$$\widehat{Var(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2 \left[\mathbf{X}'\mathbf{X}\right]^{-1} = \begin{bmatrix} \widehat{\mathrm{var}(\hat{\beta}_1)} & \widehat{Cov(\hat{\beta}_1,\hat{\beta}_2)} & \cdots & \widehat{Cov(\hat{\beta}_1,\hat{\beta}_j)} & \cdots & \widehat{Cov(\hat{\beta}_1,\hat{\beta}_k)} \\ \widehat{Cov(\hat{\beta}_2,\hat{\beta}_1)} & \widehat{\mathrm{var}(\hat{\beta}_2)} & \cdots & \widehat{Cov(\hat{\beta}_2,\hat{\beta}_j)} & \cdots & \widehat{Cov(\hat{\beta}_2,\hat{\beta}_k)} \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ \widehat{Cov(\hat{\beta}_j,\hat{\beta}_1)} & \widehat{Cov(\hat{\beta}_j,\hat{\beta}_2)} & \cdots & \widehat{\mathrm{var}(\hat{\beta}_j)} & \cdots & \widehat{Cov(\hat{\beta}_j,\hat{\beta}_k)} \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ \widehat{Cov(\hat{\beta}_k,\hat{\beta}_1)} & \widehat{Cov(\hat{\beta}_k,\hat{\beta}_2)} & \cdots & \widehat{Cov(\hat{\beta}_k,\hat{\beta}_j)} & \cdots & \widehat{\mathrm{var}(\hat{\beta}_k)} \end{bmatrix} \tag{3-69}$$

La varianza del coeficiente de la pendiente $\hat{\beta}_j$, dada en (3-64), es una función del parámetro desconocido $\sigma^2$. Cuando $\sigma^2$ se sustituye por su estimador $\hat{\sigma}^2$, se obtiene un estimador de la varianza de $\hat{\beta}_j$ :

$$\widehat{\mathrm{var}(\hat{\beta}_j)} = \frac{\hat{\sigma}^2}{nS_j^2(1-R_j^2)} \tag{3-70}$$

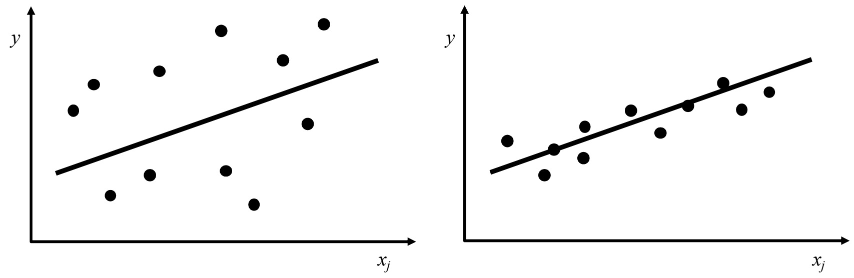
De acuerdo con la expresión anterior, el estimador de la varianza de $\hat{\beta}_j$ viene afectado por los siguientes factores:

a) Cuanto mayor es $\hat{\sigma}^2$, mayor es la varianza del estimador. Esto no es sorprendente en absoluto: cuanto más "ruido" exista en la ecuación, y, en consecuencia, más grande será $\hat{\sigma}^2$, con lo que será más difícil estimar con precisión el efecto parcial de cualquier regresor sobre *y*. (Véase figura 3.1).

b) A medida que se incrementa el tamaño de la muestra, la varianza del estimador se reduce.

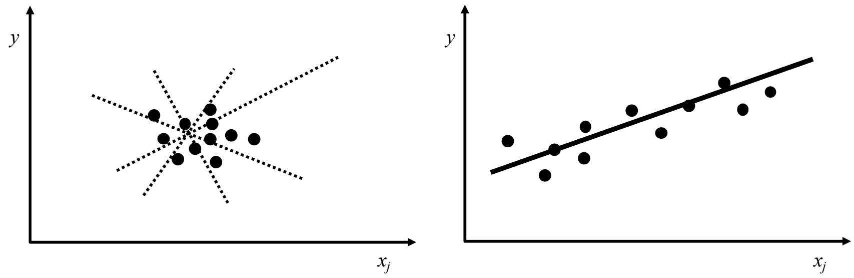
c) Cuanto más pequeña sea la varianza muestral de un regresor, mayor es la variación del coeficiente correspondiente. Manteniendo los demás factores igual, para estimar $\beta_j$ es preferible que la variación muestral de $x_j$ sea lo más grande posible, tal como se ilustra en la figura 3.2. Como se puede ver hay muchas líneas hipotéticas que podrían ajustarse a los datos cuando la varianza muestral de $x_j$, ($S_j^2$), es pequeña como puede verse en la parte a) de la figura. En cualquier caso, no está permitido por el supuesto 4 que $S_j^2$=0.

d) Cuanto mayor sea $R_j^2$, (es decir, cuanto mayor sea la correlación del regresor $j$-ésimo con el resto de los regresores), mayor será la varianza de $\hat{\beta}_j$.




a) $\hat{\sigma}^2$ grande                b) $\hat{\sigma}^2$ pequeña

**FIGURA 3.1. Influencia de $\hat{\sigma}^2$ sobre el estimador de la varianza.**




a) $S_j^2$ pequeño                b) $S_j^2$ grande

**FIGURA 3.2. Influencia de $S_j^2$ sobre el estimador de la varianza.**

A la raíz cuadrada de (3-70) se le denomina *error estándar* (*ee*) de $\hat{\beta}_j$:

$$ee(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{nS_j^2(1-R_j^2)}} \tag{3-71}$$

***Otras propiedades de los estimadores MCO***

Bajo los supuestos 1 a 6 del *MLC*, el estimador de *MCO,* $\hat{\boldsymbol{\beta}}$, es consistente, como puede verse en el apéndice 3.3, *asintótica y normalmente distribuido,* y también *asintóticamente eficiente* dentro de la clase de los estimadores consistentes y asintóticamente normales.

Bajo los supuestos 1 a 9 del *MLC*, el estimador *MCO* es también el estimador de *máxima verosimilitud* (*MV*), como se prueba en el apéndice 3.4, y es el *estimador de mínima varianza insesgado* (*EMVI*). Esto último significa que el estimador de *MCO* tiene la menor varianza entre todos los estimadores insesgados, sean lineales o no.

## 3.4 Más sobre formas funcionales

En este apartado vamos a examinar dos temas sobre formas funcionales: el uso de los logaritmos en modelos econométricos y las funciones polinomiales.

### 3.4.1 Utilización de logaritmos en los modelos econométricos

Algunas variables se utilizan a menudo en forma logarítmica. Así es en el caso de las variables monetarias que, en general, son positivas o de otras variables con valores elevados. La utilización de modelos con transformaciones logarítmicas tiene además sus ventajas. Una de ellas es que los coeficientes tienen interpretaciones atractivas (elasticidades o semi-elasticidades). Otra es la invariancia de los coeficientes de pendiente cuando hay cambios de escala en las variables. Tomar logaritmos puede ser conveniente debido a que reduce el rango de las variables, lo que hace que las estimaciones sean menos sensibles a los valores extremos de las variables. Los supuestos del *MLC* se satisfacen más a menudo en modelos que aplican logaritmos a la variable endógena, que en los modelos que no aplican ninguna transformación. Así, sucede que la distribución condicional de $y$ es frecuentemente heteroscedástica, mientras que $\ln(y)$ puede ser homoscedástica.

Una limitación de la transformación logarítmica es que no se puede utilizar cuando la variable original toma valores cero o negativos. Por otro lado, algunas variables que se miden en años y en otras variables que son una proporción o un porcentaje, se utiliza la variable original sin ninguna transformación.

### 3.4.2 Funciones polinomiales

Las funciones polinomiales se han utilizado ampliamente en la investigación econométrica. Cuando en el modelo solo hay regresores correspondientes a una función polinomial decimos que es un *modelo polinomial.* La forma general del *modelo polinomial de grado k* puede expresarse como

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \cdots + \beta_k x^k + u \qquad (3\text{-}72)$$

#### *Funciones cuadráticas*

Un caso interesante de funciones polinomiales es la *función cuadrática*, que es una *función polinomial de segundo grado*. Cuando hay sólo regresores correspondientes a la función cuadrática, tenemos un *modelo cuadrático*:

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + u \qquad (3\text{-}73)$$

Las funciones cuadráticas se utilizan muy a menudo en economía aplicada para captar la disminución o el aumento de los efectos marginales. Es importante observar que, en tal caso, $\beta_2$ no mide el cambio en $y$ con respecto a $x$, porque no tiene sentido mantener $x^2$ fijo, mientras cambia $x$. El efecto marginal de $x$ sobre $y$, que depende linealmente del valor de $x$, es el siguiente:

$$em = \frac{dy}{dx} = \beta_2 + 2\beta_3 x \qquad (3\text{-}74)$$

En una aplicación específica, este efecto marginal se evaluará para valores específicos de $x$. Si $\beta_2$ y $\beta_3$ tienen signos opuestos el punto de cambio está situado en

$$x^* = -\frac{\beta_2}{2\beta_3} \qquad (3\text{-}75)$$

Si $\beta_2>0$ y $\beta_3<0$, el efecto marginal de $x$ sobre y es positivo al principio, pero será negativo cuando $x$ sea mayor que $x^*$. Si $\beta_2<0$ y $\beta_3>0$, el efecto marginal de $x$ sobre $y$ es negativo al principio, pero será positivo para $x$ mayor que $x^*$.

***Ejemplo 3.5 Salarios y años de antigüedad en la empresa***

Utilizando los datos de *ceosal2* para estudiar el tipo de relación entre el salario (*salary)* de los consejeros delegados (CEO) en Estados Unidos y los años de permanencia en la empresa como CEO de la compañía (*ceoten*), se estimó el siguiente modelo:

$$\widehat{\ln(salary)} = \underset{(0.086)}{6.246} + \underset{(0.0001)}{0.0006}\, profits + \underset{(0.0156)}{0.0440}\, ceoten - \underset{(0.00052)}{0.0012}\, ceoten^2$$

$$R^2=0.1976 \quad n=177$$

donde los beneficios de las compañías (*profits*) están expresados en millones de dólares y el salario es la remuneración anual expresada en miles de dólares.

El efecto marginal de *ceoten* sobre *salary* expresado en porcentaje es el siguiente:

$$\widehat{em_{salario/ceoten}}\% = 4.40 - 2\times 0.12 ceoten$$

Así, para un consejero delegado con 10 años en su compañía, si está un año más en la empresa, su salario se incrementará en un 2%. Igualando a cero la expresión anterior y despejando *ceoten*, nos encontramos con que el efecto máximo de permanencia como consejero delegado sobre el salario se alcanza a los 18 años. Es decir, hasta los 18 años como CEO el efecto marginal del salario con respecto a los años de permanencia en la compañía es positivo. Por el contrario, desde los 18 años en adelante, este efecto marginal es negativo.

## *Funciones cúbicas*

Otro caso interesante es la *función cúbica* o *función polinomial de tercer grado*. Si en el modelo hay sólo regresores correspondientes a la función cúbica, tenemos un *modelo cúbico*:

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + u \tag{3-76}$$

Los modelos cúbicos se utilizan muy a menudo en economía aplicada para captar variaciones en los efectos marginales, particularmente en las funciones de costes. El efecto marginal (*em*) de *x* sobre *y*, que depende, según una forma cuadrática, del valor de *x*, será el siguiente:

$$em = \frac{dy}{dx} = \beta_2 + 2\beta_3 x + 3\beta_4 x^2 \tag{3-77}$$

El mínimo de *em* se producirá cuando

$$\frac{dem}{dx} = 2\beta_3 + 6\beta_4 x = 0 \tag{3-78}$$

Por lo tanto,

$$em_{\min} = \frac{-\beta_3}{3\beta_4} \tag{3-79}$$

En un modelo cúbico de una función de costes, debe cumplirse la restricción $\beta_3^2 < 3\beta_4\beta_2$ para garantizar que $em_{mi}$ sea positivo. Otras restricciones que la función de costes debe cumplir son las siguientes: $\beta_1$, $\beta_2$, and $\beta_4$>0; y $\beta_3$<0.

***Ejemplo 3.6 Efecto marginal en una función de costes***

Utilizando los datos de 11 empresas de plantas de celulosa (fichero *costfunc*) para estudiar la función de costes, se estimó el siguiente modelo:

$$\widehat{cost} = \underset{(1.602)}{29.16} + \underset{(0.2167)}{2.316}\, output - \underset{(0.0081)}{0.0914}\, output^2 + \underset{(0.000086)}{0.0013}\, output^3$$

$$R^2=0.9984 \quad n=11$$

donde *output* es la producción de pasta de papel en miles de toneladas y *cost* es el coste total en millones de euros.

El *coste marginal* es el siguiente:

$$\widehat{marcost} = 2.316 - 2\times 0.0914 output + 3\times 0.0013 output^2$$

Por lo tanto, en una empresa con una producción de 30 mil toneladas de pasta de papel, si la empresa aumenta la producción de celulosa en mil de toneladas, el coste se incrementará en 0.754 millones de euros. Calculando el mínimo de la expresión anterior y resolviendo para el *output*, nos encontramos con que el coste marginal mínimo es igual a una producción de 23222 toneladas de pasta de papel.

## 3.5 Bondad del ajuste y selección de regresores

Una vez que se han aplicado los mínimos cuadrados, es conveniente tener alguna medida de la bondad del ajuste del modelo a los datos. En el caso de que se hayan estimado varios modelos alternativos, las medidas de la bondad del ajuste podrían ser utilizadas para seleccionar el modelo más apropiado.

En la literatura econométrica existen numerosas medidas de bondad del ajuste. La más popular es el coeficiente de determinación, que se designa por $R^2$ o $R$-cuadrado, y el coeficiente de determinación ajustado, que se designa por $\bar{R}^2$ o $R$-cuadrado ajustado. Dado que estas medidas tienen algunas limitaciones, nos referiremos también al criterio de información de Akaike (*AIC*) y al criterio de Schwarz (*SC*).

### 3.5.1 Coeficiente de determinación

Como vimos en el capítulo 2, el coeficiente de determinación se basa en la siguiente descomposición:

$$SCT = SCT + SCR \qquad (3\text{-}80)$$

donde *SCT* es la *suma de cuadrados totales*, *SCE* es la *suma de cuadrados* explicados y *SCR* es la *suma de cuadrados residual*.

Basándose en esta ecuación, el coeficiente de determinación se define como:

$$R^2 = \frac{SCE}{SCT} \qquad (3\text{-}81)$$

Alternativamente, y de una forma equivalente, el coeficiente de determinación se puede definir como

$$R^2 = 1 - \frac{SCR}{SCT} \qquad (3\text{-}82)$$

Los valores extremos del coeficiente de determinación son: 0, cuando la varianza explicada es cero, y 1, cuando la varianza residual es cero, es decir, cuando el ajuste es perfecto. Por lo tanto,

$$0 \le R^2 \le 1 \qquad (3\text{-}83)$$

Un $R^2$ pequeño implica que la varianza de la perturbación ($\sigma^2$) es grande en relación a la variación de $y$, lo que significa que $\beta_j$ no puede ser estimada con precisión. Pero hay que recordar, que una varianza de la perturbación grande puede compensarse con un tamaño muestral elevado, de forma que si $n$ es suficientemente grande, podemos ser capaces de estimar los coeficientes con precisión a pesar de que no se hayan controlado muchos de los factores no observados.

Para interpretar el coeficiente de determinación adecuadamente, se deben tener en cuenta las siguientes cautelas:

a) Cuando se añaden nuevas variables explicativas, el coeficiente de determinación aumenta su valor o, al menos, mantiene el mismo valor. Esto sucede a pesar de que la variable o variables añadidas no tengan relación con la variable endógena. Así pues, siempre se verifica que

$$R_j^2 \geq R_{j-1}^2 \tag{3-84}$$

donde $R_{j-1}^2$ es el *R* cuadrado en un modelo con *j*-1 regresores, y $R_j^2$ es el *R* cuadrado en un modelo con un regresor adicional. Es decir, si se añaden variables a un modelo determinado, $R^2$ nunca disminuirá, incluso si estas variables no tienen una influencia significativa.

b) Si el modelo no tiene término independiente, el coeficiente de determinación no tiene una interpretación clara, porque la descomposición dada (3-80) no se cumple. Además, las dos formas de cálculo mencionadas -(3-81) y (3-82)- por lo general conducen a resultados diferentes, que en algunos casos pueden quedar fuera del intervalo [0,1].

c) El coeficiente de determinación no se puede utilizar para comparar modelos en los que la forma funcional de la variable endógena es diferente. Por ejemplo, $R^2$ no se puede aplicar para comparar dos modelos en los que el regresando es la variable original en uno, *y*, y ln(*y*) en el otro.

### 3.5.2 R cuadrado ajustado

Para superar una de las limitaciones del $R^2$, este coeficiente se puede "ajustar" de manera que tenga en cuenta el número de variables incluidas en un modelo dado. Para ver cómo el $R^2$ usual podría ajustarse, es útil expresarlo como

$$R^2 = 1 - \frac{SCR / n}{SCT / n} \tag{3-85}$$

donde, en el segundo término del segundo miembro, aparece la varianza residual dividida por la varianza del regresando.

El $R^2$, tal como está definido en (3-85), es una medida *muestral*. Ahora bien, si deseamos una medida poblacional ($R_{POB}^2$), ésta se podría definir como

$$R_{POB}^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2} \tag{3-86}$$

Sin embargo, hay que tener en cuenta que se dispone de una mejor estimación de las varianzas, $\sigma_u^2$ y $\sigma_y^2$, que las utilizadas en (3-85). En su lugar, vamos a utilizar estimaciones insesgadas de estas varianzas:

$$\bar{R}^2 = 1 - \frac{SCR / (n-k)}{SCT / (n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{3-87}$$

Esta medida se denomina *R cuadrado ajustado*, o $\bar{R}^2$. El principal atractivo del $\bar{R}^2$ es que impone una penalización al añadir otros regresores a un modelo. Si se añade un regresor al modelo la *SCR* decrece o, en el peor de los casos queda, igual. Por otra parte, los *grados de libertad* de la regresión ($n-k$) siempre disminuyen. Por ello, el $\bar{R}^2$ puede crecer o decrecer cuando se añade un nuevo regresor al modelo. Es decir:

$$\bar{R}_j^2 \geq \bar{R}_{j-1}^2 \quad \text{o} \quad \bar{R}_j^2 \leq \bar{R}_{j-1}^2 \tag{3-88}$$

Un resultado algebraico interesante es el hecho de que si añadimos un nuevo regresor a un modelo, el $\bar{R}^2$ se incrementa si, y solo si, el estadístico $t$ del nuevo regresor es mayor que uno en valor absoluto. Así, vemos inmediatamente que $\bar{R}^2$ podría ser utilizado para decidir si un determinado regresor adicional debe ser incluido en el modelo. El $\bar{R}^2$ tiene una cota superior que es igual a 1, pero estrictamente no tiene una cota inferior, ya que puede tomar un valor negativo, aunque muy cerca de 0.

Las observaciones b) y c) hechas para el $R$ cuadrado siguen siendo válidas para el $R$ cuadrado ajustado.

### 3.5.3 Criterio de información de Akaike (*AIC*) y criterio de Schwarz (*SC*)

Estos dos criterios -criterio de información de Akaike (*AIC*) y criterio de Schwarz (*SC*)- tienen una estructura muy similar. Por esta razón, se examinarán conjuntamente.

El estadístico *AIC*, propuesto por Akaike (1974) y basado en la teoría de la información, tiene la siguiente expresión:

$$AIC = -\frac{2l}{n} + \frac{2k}{n} \tag{3-89}$$

donde $l$ es el logaritmo de la función de verosimilitud (suponiendo que las perturbaciones tengan una distribución normal) evaluada para los valores estimados de los coeficientes.

El estadístico *SC* propuesto por Schwarz (1978), tiene la siguiente expresión:

$$SC = -\frac{2l}{n} + \frac{k\ln(n)}{n} \tag{3-90}$$

Los estadísticos *AIC* y *SC*, a diferencia de los coeficientes de determinación ($R^2$ y $\bar{R}^2$), indican mejores ajustes cuanto más bajos sean sus valores. Es importante destacar que los estadísticos *AIC* y *SC* no tienen cotas, a diferencia del $R^2$.

a) Los estadísticos *AIC* y *SC* penalizan la introducción de nuevos regresores. En el caso de *AIC*, como puede verse en el segundo término del segundo miembro de (3-89), el número de regresores $k$ aparece en el numerador. Por lo tanto, el crecimiento de $k$ incrementará el valor del *AIC* y por lo tanto empeorará la bondad del ajuste, si no se ve compensado por un crecimiento suficiente de $l$. En el caso del *SC*, como puede verse en el segundo término del segundo miembro de (3-90), el numerador es $k\ln(n)$. Para $n>7$, ocurre lo siguiente: $k\ln(n)>2k$. Por lo tanto, el *SC* impone una penalización mayor a la introducción de regresores que el *AIC* cuando el tamaño de la muestra es mayor de 7.

b) Los estadísticos *AIC* y *SC* se puede aplicar a modelos estadísticos sin término independiente.

c) Los estadísticos *AIC* y *SC* no son medidas relativas como lo son los coeficientes de determinación. Por lo tanto, su magnitud, en sí misma, no ofrece ninguna información.

d) Los estadísticos *AIC* y *SC* se puede aplicar para comparar modelos en los que las variables endógenas tienen diferentes formas funcionales. En particular, vamos a comparar dos modelos en los que los regresandos son *y* y ln(*y*). Cuando el regresando es *y*, se aplica la fórmula (3-89) en el caso del *AIC*, o (3-90) en el caso del *SC*. Cuando el regresando es ln(*y*), y además queremos comparar con otro modelo en el que el regresando es *y*, hay que corregir esos estadísticos de la siguiente manera:

$$AIC_C = AIC + 2\overline{\ln(Y)} \tag{3-91}$$

$$SC_C = SC + 2\overline{\ln(Y)} \tag{3-92}$$

donde $AIC_C$ y $SC_C$ son los estadísticos corregidos, y *AIC* y *SC* son los estadísticos que suministra cualquier paquete econométrico como, por ejemplo, el E-views.

***Ejemplo 3.7 Selección del mejor modelo***

Para analizar los determinantes del gasto en productos lácteos, se han considerados los siguientes modelos alternativos:

1) $dairy = \beta_1 + \beta_2 inc + u$
2) $dairy = \beta_1 + \beta_2 \ln(inc) + u$
3) $dairy = \beta_1 + \beta_2 inc + \beta_3 punder5 + u$
4) $dairy = \beta_2 inc + \beta_3 punder5 + u$
5) $dairy = \beta_1 + \beta_2 inc + \beta_3 hhsize + u$
6) $\ln(dairy) = \beta_1 + \beta_2 inc + u$
7) $\ln(dairy) = \beta_1 + \beta_2 inc + \beta_3 punder5 + u$
8) $\ln(dairy) = \beta_2 inc + \beta_3 punder5 + u$

donde *inc* es la renta disponible de los hogares, *hhsize* es el número de miembros del hogar y *punder5* es la proporción de niños menores de cinco años en el hogar.

Utilizando una muestra de 40 hogares (fichero *demand*), y teniendo en cuenta que $\overline{\ln(dairy)} =$ 2.3719, los estadísticos de bondad del ajuste obtenidos para los 8 modelos se muestran en el cuadro 3.1. En particular, el estadístico *AIC* corregido para el modelo 6) se ha calculado como sigue:

$$AIC_C = AIC + 2\overline{\ln(Y)} = 0.2794 + 2\times 2.3719 = 5.0232$$

*Conclusiones*

a) El *R*-cuadrado puede ser utilizado para comparar los siguientes pares de modelos: 1) con 2), y 3) con 5).
b) El *R*-cuadrado ajustado sólo se puede utilizar para comparar los modelos 1) con 2), 3) y 5); y 6) con 7.
c) El mejor de los ocho modelos es el modelo de 7) de acuerdo con los criterios *AIC* y *SC*.

CUADRO 3.1. **Medidas de bondad de ajuste de ocho modelos.**

| ***Número de modelo*** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| ***Regresando*** | *dairy* | *dairy* | *dairy* | *dairy* | *dairy* | ln(*dairy*) | ln(*dairy*) | ln(*dairy*) |
| ***Regresores*** | *intercept* *inc* | *intercept* ln(*inc)* | *intercept* *inc* *punder5* | *inc* *punder5* | *intercept* *Inc* *househsize* | *intercept* *inc* | *intercept* *inc* *punder5* | *inc* *punder5* |
| R-cuadrado | 0.4584 | 0.4567 | 0.5599 | 0.5531 | 0.4598 | 0.4978 | 0.5986 | -0.6813 |
| R-cuadrado ajustado | 0.4441 | 0.4424 | 0.5361 | 0.5413 | 0.4306 | 0.4846 | 0.5769 | -0.7255 |
| Criterio de información de Akaike | 5.2374 | 5.2404 | 5.0798 | 5.0452 | 5.2847 | 0.2794 | 0.1052 | 1.4877 |
| Criterio de Schwarz | 5.3219 | 5.3249 | 5.2065 | 5.1296 | 5.4113 | 0.3638 | 0.2319 | 1.5721 |
| Criterio de información de Akaike corregido | | | | | | 5.0232 | 4.8490 | 6.2314 |
| Criterio de Schwarz corregido | | | | | | 5.1076 | 4.9756 | 6.3159 |

## Ejercicios

**Ejercicio 3.1** Considere el modelo de regresión lineal $\mathbf{y} = \mathbf{X\beta} + \mathbf{u}$, donde $\mathbf{X}$ es una matriz 50×5.

Conteste de forma razonada a las siguientes cuestiones:

*a*) ¿Cuáles son las dimensiones de los vectores $\mathbf{y}$, $\boldsymbol{\beta}$, $\mathbf{u}$ ?

*b*) ¿Cuántas ecuaciones hay en el sistema de ecuaciones normales $\mathbf{X'X\hat{\beta}} = \mathbf{X'y}$ ?

*c*) ¿Qué condición debe cumplirse para poder obtener $\hat{\boldsymbol{\beta}}$ ?

**Ejercicio 3.2** Dado el modelo

$$y_i=\beta_1+\beta_2 x_{2i}+\beta_3 x_{3i}+u_i$$

y los siguientes datos:

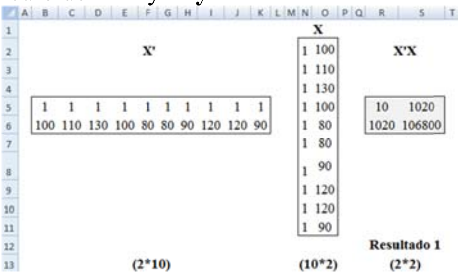
| $y$ | $x_2$ | $x_3$ |
|---|---|---|
| 10 | 1 | 0 |
| 25 | 3 | -1 |
| 32 | 4 | 0 |
| 43 | 5 | 1 |
| 58 | 7 | -1 |
| 62 | 8 | 0 |
| 67 | 10 | -1 |
| 71 | 10 | 2 |

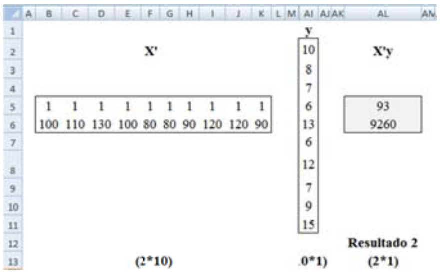
*a)* Estime $\beta_1$, $\beta_2$ y $\beta_3$ por *MCO*.
*b)* Calcule la suma de los cuadrados de los residuos.
*c)* Obtenga la varianza residual.
*d)* Obtenga la varianza explicada por la regresión.
*e)* Obtenga la varianza de la variable endógena.
*f)* Calcule el coeficiente de determinación.
*g)* Obtenga una estimación insesgada de $\sigma^2$.
*h)* Estime la varianza de $\hat{\beta}_2$.

Para responder a estas preguntas puede utilizar Excel. Véase el recuadro 3.1.

**Recuadro 3.1**

a) Cálculo de **X'X** y **X'y**




Explicación para **X'X**

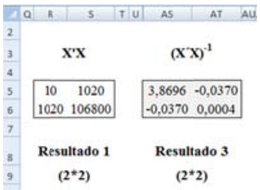
a) Introduzca las matrices **X'** y **X**: B5:K6 y N2:O11 en Excel

b) El producto **X'X** se calcula seleccionando previamente las celdas donde desea colocar la matriz resultante (R5:S6).

c) Una vez seleccionadas las celdas para la matriz resultante, y mientras aún está resaltada, escriba la fórmula siguiente: = MMULT (B5:K6;N2:O11)

d) Cuando la fórmula se haya introducido, pulse la *tecla Ctrl* y la *tecla Shift* simultáneamente, entonces, teniendo presionadas estas dos teclas, pulse la *tecla Enter* también.

2) Cálculo of $(\mathbf{X'X})^{-1}$



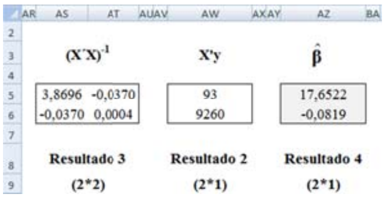

a) Introduzca la matriz **X'X** en Excel: R5:S6

b) Encontramos la inversa de la matriz X'X, seleccionando previamente las celdas donde deseamos colocar la matriz resultante (AS5:AT6)

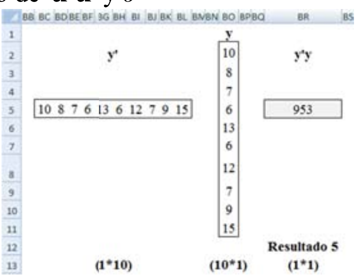
c) Una vez seleccionadas las celdas para la matriz resultante, y mientras aún está resaltada, escriba la fórmula siguiente: = MINVERSA (AO5:AP6).

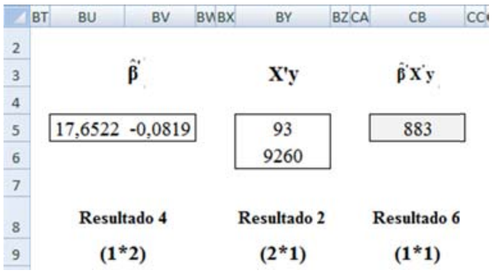
d) Cuando la fórmula se haya introducido, pulse la *tecla Ctrl* y la *tecla Shift* simultáneamente, entonces, teniendo presionadas estas dos teclas, pulse la *tecla Enter* también.

3) Cálculo de vector $\hat{\boldsymbol{\beta}}$




4) Cálculo de $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ y $\sigma^2$




$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{y}}'\hat{\mathbf{y}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = R.5 - R.6 = 953 - 883 = 70$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-2} = \frac{70}{8} = 8.6993$$

5) Cálculo de la matriz de covarianzas de $\hat{\boldsymbol{\beta}}$

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2\left[\mathbf{X}'\mathbf{X}\right]^{-1} = 8.6993\begin{pmatrix} 3.8696 & -0.0370 \\ -0.0370 & 0.0004 \end{pmatrix} = \begin{pmatrix} 33.6624 & -0.3215 \\ -0.3215 & 0.0032 \end{pmatrix}$$

**Ejercicio 3.3** El siguiente modelo ha sido estimado para explicar las *ventas* anuales de empresas fabricantes de productos de limpieza doméstica en función de un índice de precios relativo (*ipr*) y de gastos de publicidad (*gpub*):

$$ventas = \beta_1 + \beta_2 ipr + \beta_3 gpu + u$$

donde las *ventas* están expresadas en millones de euros, *ip* es un índice de precios relativos (precios de la empresa/precios de la empresa 1 de la muestra) y *gpub* son los gastos anuales realizados en publicidad y campañas de promoción y difusión, expresados también en millones de euros.

Para ello se dispone de los siguientes datos sobre diez empresas fabricantes de productos de limpieza doméstica:

| *firm* | *ventas* | *rpi* | *gpub* |
|---|---|---|---|
| 1 | 10 | 100 | 300 |
| 2 | 8 | 110 | 400 |
| 3 | 7 | 130 | 600 |
| 4 | 6 | 100 | 100 |
| 5 | 13 | 80 | 300 |
| 6 | 6 | 80 | 100 |
| 7 | 12 | 90 | 600 |
| 8 | 7 | 120 | 200 |
| 9 | 9 | 120 | 400 |
| 10 | 15 | 90 | 700 |

Utilizando una hoja excel:

*a*) Estime los parámetros del modelo propuesto.
*b*) Estime la matriz de covarianzas.
*c*) Calcule el coeficiente de determinación.

*Nota*: En el recuadro 3.1 se estima el modelo $ventas = \beta_1 + \beta_2 rpi + u$ utilizando Excel. Allí también pueden verse las instrucciones para hacerlo.

**Ejercicio 3.4** Un investigador, que está elaborando un modelo econométrico con el que desea explicar el comportamiento de la renta, formula la siguiente especificación:

$$renta=\alpha+\beta cons+\gamma ahorro+u \qquad [1]$$

donde *renta* es la renta disponible de las familias, *cons* es el consumo total y *ahorro* es el ahorro total de las familias.

El investigador no tuvo en cuenta que las tres magnitudes anteriores están ligadas por la identidad

$$renta=cons+ahorro \qquad [2]$$

La equivalencia entre los modelos [1] y [2] exige que, además de desaparecer el término de perturbación, los parámetros del modelo [1] tomen los siguientes valores:

$$\alpha=0, \beta=1, \gamma=1$$

Si se emplean los datos de un país para ajustar la ecuación [1] por MCO, ¿Se puede esperar, *en general*, que las estimaciones obtenidas tomen los valores

$$\hat{\alpha}=0, \hat{\beta}=1, \hat{\gamma}=0?$$

Justifíquese la respuesta, utilizando notación matemática.

**Ejercicio 3.5** Un investigador plantea el siguiente modelo econométrico para explicar los ingresos totales por turismo en un país determinado (*ingtotur*):

$$ingtotur = \beta_1 + \beta_2 ingmetur + \beta_3 numtur + u$$

donde *ingmetur* es el gasto medio por turista y *numtur* es el número total de turistas.

*a*) Es obvio que *turtot*, *turmean* y *numtur* están ligados también por la relación *ingtotur*=*turmean*×*numtur*; ¿afectará este hecho de alguna forma a las estimaciones de los parámetros del modelo propuesto?

*b*) ¿Existe otra forma funcional del modelo que implique restricciones más fuertes sobre los parámetros? Si la hubiera, indíquela.

*c*) ¿Le parece razonable utilizar el modelo indicado para explicar el comportamiento de los ingresos por turismo?

**Ejercicio 3.6** Supongamos que se tiene que estimar el modelo

$$\ln(y) = \beta_1 + \beta_2 \ln(x_2) + \beta_3 \ln(x_3) + \beta_4 \ln(x_4) + u$$

utilizando las siguientes observaciones:

| $x_2$ | $x_3$ | $x_4$ |
|---|---|---|
| 3 | 12 | 4 |
| 2 | 10 | 5 |
| 4 | 4 | 1 |
| 3 | 9 | 3 |
| 2 | 6 | 3 |
| 5 | 5 | 1 |

¿Qué problemas puede plantear la estimación de este modelo?

**Ejercicio 3.7** Conteste a las siguientes preguntas:

*a*) Explique que miden los coeficientes de determinación ($R^2$) y de determinación corregido ($\bar{R}^2$). ¿Para qué se pueden utilizar? Razone la respuesta.

*b*) Dados los modelos

$$\ln(y)=\beta_1+\beta_2\ln(x)+u \quad (1)$$
$$\ln(y)=\beta_1+\beta_2\ln(x)+\beta_3\ln(z)+u \quad (2)$$
$$\ln(y)=\beta_1+\beta_2\ln(z)+u \quad (3)$$
$$y=\beta_1+\beta_2 z+u \quad (4)$$

indique qué medida de bondad del ajuste es adecuada para comparar los siguientes pares de modelos: (1)-(2); (1)-(3); y (1)-(4). Razone su respuesta.

**Ejercicio 3.8** Se estima por *MCO* el siguiente modelo:

$$\ln(y) = \beta_1 + \beta_2 \ln(x) + \beta_3 \ln(z) + u$$

*a)* ¿Pueden ser todos los residuos mínimo cuadráticos positivos? Razone la respuesta.

*b*) Bajo la hipótesis básica de no autocorrelación de las perturbaciones, ¿son independientes los residuos minimocuadráticos? Razone la respuesta.

*c)* Suponiendo que las perturbaciones no tengan distribución normal, ¿los estimadores minimocuadráticos son insesgados? Razone la respuesta.

**Ejercicio 3.9** Considere el modelo de regresión

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$$

donde $\mathbf{y}$ y $\mathbf{u}$ son vectores 8×1, $\mathbf{X}$ es una matriz 8×3 y $\boldsymbol{\beta}$ es un vector 3×1 de parámetros desconocidos. Además se dispone de la siguiente información:

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 3\end{bmatrix} \qquad\qquad \hat{\mathbf{u}}'\hat{\mathbf{u}}=22$$

Responda a las siguientes preguntas, justificando la respuesta:

*a)* Indique el tamaño de la muestra, el número de regresores, el número de parámetros y los grados de libertad de la suma de los cuadrados de los residuos.

*b)* Deduzca la matriz de covarianzas del vector $\hat{\boldsymbol{\beta}}$, explicitando los supuestos utilizados. Estime las varianzas de los estimadores de los parámetros del modelo.

*c)* ¿Contiene el modelo de regresión término constante? ¿Qué implicaciones tiene la contestación a esta pregunta en el significado del $R^2$ en este modelo?

**Ejercicio 3.10** Argumente la veracidad o falsedad de las siguientes afirmaciones:

*a)* En un modelo de regresión lineal, la suma de los residuos es cero.

*b)* El coeficiente de determinación ($R^2$) es siempre una buena medida de la calidad del modelo.

*c)* El estimador por mínimos cuadrados es un estimador sesgado.

**Ejercicio 3.11** El siguiente modelo se formula para explicar el tiempo empleado en dormir:

$$sleep=\beta_1+\beta_2\ totalwrk+\beta_3 leisure+u$$

donde el tiempo dedicado a dormir (*sleep)*, al trabajo -remunerado y no remunerado- (*totalwrk* ), y al ocio (*leisure*) (tiempo no dedicado a dormir o trabajar) están medidos en minutos por día.

La ecuación estimada con una muestra de 200 observaciones, utilizando el fichero *timuse03*, es la siguiente:

$$\widehat{sleep}=1440-1\times\ total\_work-1\times leisure$$

$$R^2\text{=1.000} \qquad n\text{=1000}$$

*a)* ¿Cuál es su opinión acerca de estos resultados?

*b)* ¿Cuál es el significado del término independiente estimado?

**Ejercicio 3.12** Utilizando una submuestra de la Encuesta de Estructura Salarial para España en 2006 (archivo *wage06sp*) se estimó el siguiente modelo para explicar el salario (*wage*):

$$\widehat{\ln(wage)}=1.565+0.0448educ+0.0177tenure+0.0065age$$

$$R^2\text{=0.337} \qquad n\text{=800}$$

donde educación (*educ*), permanencia en la empresa (*tenure)* y edad (*age*) están medidos en años y el salario en euros por hora.

*a)* ¿Cuál es la interpretación de los coeficientes *educ*, *tenure* y *age*?

*b)* ¿Cuántos años tiene que aumentar la edad para que tenga un efecto similar al incremento de 1 año en la educación, manteniendo fijos en cada caso, los otros dos regresores?

*c)* Sabiendo que $\overline{educ}$=10.2, $\overline{tenure}$=7.2 y $\overline{age}$=42.0, calcule las elasticidades de los salarios con respecto a la educación, permanencia en la empresa y edad, manteniendo fijos los otros regresores. ¿Considera usted que estas elasticidades son altas o bajas?

**Ejercicio 3.13** La siguiente ecuación describe el precio de la vivienda en términos del número de dormitorios de la casa (*bedrooms*), del número de baños completos (*bathrms*) y del tamaño de la parcela en pies cuadrados (*lotsize*):

$$price = \beta_1 + \beta_2 bedrooms + \beta_3 bathrms + \beta_4 lotsize + u$$

donde el precio (*price*) de la vivienda se mide en dólares.

Utilizando los datos de la ciudad de Windsor contenidos en el fichero *housecan*, se estima el siguiente modelo:

$$\widehat{price} = -2418 + 5827 bedrooms + 19750 bathrms + 5.411 lotsize$$

$$R^2\text{=0.486} \quad n\text{=546}$$

*a)* ¿Cuál es el aumento estimado en el precio de una casa con un dormitorio y un baño adicionales, manteniendo *lotsize* constante?

*b)* ¿Qué porcentaje de variación en el precio se explica por el número de dormitorios, el número de baños completos y el tamaño de la vivienda en conjunto?

*c)* Determine el precio de venta predicho para una casa de la muestra con *bedrooms*=3, *bathrms*=2 y *lotsize*=3880.

*d)* El precio de venta real de la casa en *c*) fue de 66000$. Encuentre el valor del residuo para esta casa. A la vista de este resultado, ¿el comprador pagó de más o de menos por la casa?

**Ejercicio 3.14** Para examinar los efectos de los rendimientos de las empresas sobre los salarios de sus consejeros delegados se formuló el siguiente modelo:

$$\ln(salary) = \beta_1 + \beta_2 roa + \beta_3 \ln(sales) + \beta_4 profits + \beta_5 tenure + u$$

donde *roa*, es la ratio beneficios/activos expresada en porcentaje y *tenure* es el número de años en la empresa como consejero delegado (=0 si es menor de 6 meses). El salario (*salary*) está expresado en miles de dólares, mientras que las ventas (*sales*) y los beneficios (*profits*) están en millones de dólares.

Se ha utilizado el fichero *ceoforbes* para la estimación del modelo. Este archivo contiene datos sobre 447 ejecutivos de las 500 empresas más grandes de EE.UU. (52 de las 500 empresas fueron excluidas por falta de datos sobre una o más variables. Apple Computer también fue excluido porque Steve Jobs, consejero delegado de Apple en 1999, no recibió ninguna compensación durante ese período.) Los datos de las empresas provienen de la revista Fortune y se refieren a 1999, los datos de los consejeros delegados provienen de la revista Forbes y se refieren también a 1999. Los resultados obtenidos fueron los siguientes:

$$\widehat{\ln(salary)} = 4.641 + 0.0054 roa + 0.2893 \ln(sales) + 0.0000564 profits + 0.0122 tenure$$

$$R^2\text{=0.232} \quad n\text{=447}$$

*a)* Interprete el coeficiente del regresor *roa*.

*b)* Interprete el coeficiente del regresor ln(*sales*). ¿Cuál es su opinión sobre la magnitud de la elasticidad del *salary/sales*?
*c)* Interprete el coeficiente del regresor *profits*.
*d)* ¿Cuál es la elasticidad de *salary/profits* para el punto de las medias muestrales ($\overline{salary}$=2028 y $\overline{profits}$=700).

**Ejercicio 3.15** (Continuación del ejercicio 2.21) Utilizando una base de datos de 1983 empresas encuestada en el año 2006 (fichero *rdspain*), se estimó la siguiente ecuación:

$$\widehat{rdintens} = -1.8168 + 0.1482\ln(sales) + 0.0110 exponsal$$

$$R^2 = 0.048 \quad n=1983$$

donde *rdintens* es el gasto en investigación y desarrollo (I+D) expresado como porcentaje de las ventas, las ventas (*sales*) se miden en millones de euros, y *exponsal* son las exportaciones tomadas como porcentaje de las ventas.

*a)* Interprete el coeficiente de ln(*sales*). En particular, si las ventas aumentan en un 100%, ¿cuál es el porcentaje de cambio estimado de *rdintens*? ¿Es éste un efecto económico grande?
*b)* Interprete el coeficiente de *exponsal*. ¿Es grande este coeficiente desde un punto de vista económico?
*c)* ¿Qué porcentaje de la variación en *rdintens* se explica por las ventas y por las exportaciones tomadas como porcentaje de las ventas?
*d)* ¿Cuál es la elasticidad *rdintens*/*sales* para la media muestral ($\overline{rdintens}$= 0.732 y $\overline{sales}$=63544960 ). Comente el resultado.
*e)* ¿Cuál es la elasticidad *rdintens*/*exponsal* para la media muestral ($\overline{rdintens}$= 0.732 y $\overline{exponsal}$ 17.657)? Comente el resultado.

**Ejercicio 3.16** La siguiente regresión hedónica se formuló para explicar los precios de los coches (véase ejemplo 3.3):

$$\ln(price) = \beta_1 + \beta_2 cid + \beta_3 hpweight + \beta_4 fueleff + u$$

donde *cid*, es el desplazamiento en pulgadas cúbicas, *hpweight* es la *ratio* potencia/ peso en kg, expresada en porcentaje y *fueleff* es la ratio litros por 100 km/caballos de vapor expresada en porcentaje.

*a)* ¿Cuáles son los signos probables de $\beta_2$, $\beta_3$ y $\beta_4$? Explíquelo.
*b)* Estime el modelo utilizando el fichero *hedcarsp* y exprese los resultados en forma de ecuación.
*c)* Interprete el coeficiente del regresor *cid*.
*d)* Interprete el coeficiente del regresor *hpweight*.
*e)* Expanda el modelo, introduciendo un regresor relativo al tamaño de coche, como el volumen o el peso. ¿Qué pasa si se introducen los dos en la regresión? ¿Cuál cree que es la relación entre el peso y el volumen?

**Ejercicio 3.17** El concepto de trabajo cubre un amplio espectro de actividades posibles en la economía productiva. Una parte importante del trabajo es no remunerado, no pasa por el mercado y, por lo tanto, no tiene precio. El trabajo no remunerado más importante es el trabajo realizado en el hogar (*houswork*) llevado a cabo principalmente por mujeres. Con el fin de analizar los factores que influyen en el trabajo del hogar, se ha formulado el siguiente modelo:

$$houswork = \beta_1 + \beta_2 educ + \beta_3 hhinc + \beta_4 age + \beta_5 paidwork + u$$

donde *educ* son los años de educación alcanzados, *hhinc* son los ingresos de los hogares en euros por mes, *age* es la edad de la persona entrevistada y *paidwork* es el trabajo remunerado. Las variables *houswork* y *paidwork* están medidas en minutos por día.

Utilice los datos contenidos en el fichero *timuse03* para estimar el modelo. Este archivo contiene 1000 observaciones correspondientes a una submuestra aleatoria extraída de la encuesta de *Empleo del Tiempo* en España que se llevó a cabo en el período 2002-2003.

*a)* ¿Qué signos esperaría para $\beta_2$, $\beta_3$, $\beta_4$ y $\beta_5$? Explíquelo.
*b)* Exprese los resultados en forma de ecuación
*c)* ¿Cree usted que hay factores relevantes omitidos en la ecuación anterior? Explíquelo.
*d)* Interprete los coeficientes de los regresores *educ*, *hhinc*, *age* y *paidwork*.

**Ejercicio 3.18** (Continuación del ejercicio 2.20) Para explicar la satisfacción general de las personas (*stsfglo*) se formula el siguiente modelo:

$$stsfglo = \beta_1 + \beta_2 gnipc + \beta_3 lifexpec + u$$

donde *gnipc* es la renta nacional bruta per cápita expresada en dólares (USA) PPA (paridad del poder adquisitivo) a precios de 2008 y *lifexpec* es la esperanza de vida al nacer, es decir, el número de años que un recién nacido puede esperar vivir. Cuando una magnitud se expresa en dólares estadounidenses PPA, eso significa que se ha convertido a dólares internacionales usando tasas PPA. (Un dólar internacional tiene el mismo poder adquisitivo que un dólar de los EE.UU. en los Estados Unidos.)

Utilice el fichero *HDR2010* para la estimación del modelo.

*a)* ¿Qué signos esperaría para $\beta_2$ y $\beta_3$? Explíquelo.
*b)* ¿Cuál sería la de satisfacción global media de un país cuyos habitantes tienen una esperanza de vida al nacer de 80 años y que tienen una renta nacional bruta per cápita de 30.000 dólares PPA expresados en dólares de 2008 de Estados Unidos?
*c)* Interprete los coeficientes de *gnipc* y *lifexpe*.
*d)* Teniendo en cuenta un país cuya esperanza de vida al nacer es igual a 50 años, ¿Cual debería ser la renta nacional bruta per cápita para obtener una satisfacción global igual a 5?

**Ejercicio 3.19** (Continuación del ejercicio 2.24) Debido a los problemas surgidos en el modelo keynesiano, Brown introdujo en la función de consumo, además de la renta, el consumo retardado para reflejar la persistencia de hábitos del consumidor:

$$conspc = \beta_1 + \beta_2 incpc + \beta_3 conspc(-1) + u$$

Como en este modelo se incluye el consumo retardado, hay que distinguir entre propensión marginal al consumo a corto plazo y a largo plazo. La propensión marginal a corto plazo se calcula de igual forma que en la función de consumo de Keynes. Para calcular la propensión marginal a largo plazo se debe considerar una situación de equilibrio, en la que no hay variaciones en las variables. Designando por $conspc^e$ y $incpc^e$ al consumo y a la renta de equilibrio y prescindiendo de la perturbación aleatoria, el modelo anterior en situación de equilibrio viene dado por

$$conspc^e = \beta_1 + \beta_2 incpc^e + \beta_3 conspc^e$$

La función de consumo de Brown se estimó con datos de la economía española para el periodo 1954-2010 (fichero *consumsp*), obteniéndose los siguientes resultados:

$$\widehat{conspc_t} = -7.156 + 0.3965 incpc_t + 0.5771 conspc_{t-1}$$

$$R^2=0.997 \qquad n=56$$

*a)* Interprete el coeficiente de *incpc*. En su interpretación, ¿incluiría la clausula “mantenido fijo el otro regresor? Justifique la respuesta.

*b)* Calcule la elasticidad a corto plazo para las medias muestrales ($\overline{conspc}$ =8084, $\overline{incpc}$ =8896).

*c)* Calcule la elasticidad a largo plazo para las medias muestrales.

*d)* Comente la diferencia entre los valores obtenidos para los dos tipos de elasticidad.

**Ejercicio 3.20** Para explicar la influencia de los incentivos y los gastos de publicidad en las ventas, se han formulado los modelos alternativos siguientes:

$$sales = \beta_1 + \beta_2 advert + \beta_3 incent + u \tag{1}$$

$$\ln(sales) = \beta_1 + \beta_2 \ln(advert) + \beta_3 \ln(incent) + u \tag{2}$$

$$\ln(sales) = \beta_1 + \beta_2 advert + \beta_3 incent + u \tag{3}$$

$$sales = \beta_2 advert + \beta_3 incent + u \tag{4}$$

$$\ln(sales) = \beta_1 + \beta_2 \ln(incent) + u \tag{5}$$

$$sales = \beta_1 + \beta_2 incent + u \tag{6}$$

*a)* Utilizando una muestra de 18 áreas de venta (fichero *advincen*), estime los modelos anteriores:

*b)* Dentro de cada uno de los siguientes grupos seleccione el mejor modelo, indicando cuáles han sido los criterios que se han utilizado. Justifique su respuesta.

- *b1)* (1) y (6)
- *b2)* (2) y (3)
- *b3)* (1) y (4)
- *b4)* (2), (3) y (5)
- *b5)* (1), (4) y (6)
- *b6)* (1), (2), (3), (4), (5) y (6)

## Apéndices

### Apéndice 3.1 Demostración del Teorema de Gauss-Markov

Para demostrar este teorema, se utilizan los supuestos 1 a 8 del *MLC*.

Consideremos otro estimador $\tilde{\boldsymbol{\beta}}$ que es una función de $\mathbf{y}$ (recuerde que $\hat{\boldsymbol{\beta}}$ es también una función de $\mathbf{y}$), dado por

$$\tilde{\boldsymbol{\beta}} = \left[ [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}' + \mathbf{A} \right] \mathbf{y} \tag{3-93}$$

donde $\mathbf{A}$ es una matriz arbitraria, $k \times n$, que es función de $\mathbf{X}$ y/o otras variables no estocásticas, pero no es función de $\mathbf{y}$. Para que $\tilde{\boldsymbol{\beta}}$ sea insesgado, se han de cumplir ciertas condiciones.

Teniendo en cuenta (3-52), tenemos que

$$\tilde{\boldsymbol{\beta}} = \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' + \mathbf{A} \right] \left[ \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \right] = \boldsymbol{\beta} + \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' + \mathbf{A} \right] \mathbf{u} \qquad (3\text{-}94)$$

Tomando las esperanzas en ambos miembros de (3-94), tenemos que

$$E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' + \mathbf{A} \right] E(\mathbf{u}) = \boldsymbol{\beta} + \mathbf{A}\mathbf{X}\boldsymbol{\beta} \qquad (3\text{-}95)$$

Para que $\tilde{\boldsymbol{\beta}}$ sea insesgado, es decir, $E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, se debe cumplir lo siguiente:

$$\mathbf{A}\mathbf{X} = \mathbf{I} \qquad (3\text{-}96)$$

Consecuentemente,

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' + \mathbf{A} \right] \mathbf{u} \qquad (3\text{-}97)$$

Teniendo en cuenta los supuestos 7 y 8, y (3-96), la $Var(\tilde{\boldsymbol{\beta}})$ es igual a

$$\begin{aligned} Var(\tilde{\boldsymbol{\beta}}) &= E((\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})') = E\left[ \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' + \mathbf{A} \right] \mathbf{u}\mathbf{u}' \left[ \mathbf{X} \left[ \mathbf{X}'\mathbf{X} \right]^{-1} + \mathbf{A}' \right] \right] \\ &= E\left[ \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' \right] \mathbf{u}\mathbf{u}' \left[ \mathbf{X} \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \right] + \mathbf{A}\mathbf{A}' \right] = \sigma^2 \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} + \mathbf{A}\mathbf{A}' \right] \end{aligned} \qquad (3\text{-}98)$$

La diferencia entre ambas varianzas es la siguiente:

$$Var(\tilde{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[ \left[ \mathbf{X}'\mathbf{X} \right]^{-1} + \mathbf{A}\mathbf{A}' - \left[ \mathbf{X}'\mathbf{X} \right]^{-1} \right] = \sigma^2 \mathbf{A}\mathbf{A}' \qquad (3\text{-}99)$$

El producto de una matriz por su transpuesta es siempre una matriz semidefinida positiva. Por lo tanto,

$$Var(\tilde{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{A}\mathbf{A}' \geq 0 \qquad (3\text{-}100)$$

La diferencia entre la varianza de un estimador $\tilde{\boldsymbol{\beta}}$-arbitrario, pero lineal e insesgado - y la varianza del estimador $\hat{\boldsymbol{\beta}}$ es una matriz semidefinida positiva. En consecuencia, $\hat{\boldsymbol{\beta}}$ es un Estimador Lineal Insesgado Óptimo, es decir, es un estimador *ELIO*.

## Apéndice 3.2 Demostración: $\hat{\sigma}^2$ es un estimador insesgado de la varianza de la perturbación

Con el fin de ver cuál es el estimador más conveniente de $\sigma^2$, se van a analizar primero las propiedades de la suma de los cuadrados de los residuos. Este es precisamente el numerador de la varianza residual.

Teniendo en cuenta (3-17) y (3-23), vamos a expresar el vector de residuos como una función del regresando

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}\left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}'\mathbf{y} = \left[ \mathbf{I} - \mathbf{X}\left[ \mathbf{X}'\mathbf{X} \right]^{-1} \mathbf{X}' \right] \mathbf{y} = \mathbf{M}\mathbf{y} \qquad (3\text{-}101)$$

donde **M** es la matriz idempotente.

Alternativamente, el vector de residuos se puede expresar como una función del vector de las perturbaciones:

$$
\begin{aligned}
\hat{\mathbf{u}} &= \left[\mathbf{I} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\right]\mathbf{y} = \left[\mathbf{I} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\right]\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\right] \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{u} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\boldsymbol{\beta}\mathbf{u} \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} + \left[\mathbf{I} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\right]\mathbf{u} = \left[\mathbf{I} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\right]\mathbf{u} \\
&= \mathbf{M}\mathbf{u}
\end{aligned}
\qquad (3\text{-}102)
$$

Teniendo en cuenta (3-102), la suma de los cuadrados de los residuos (*SRC*) se puede expresar en la forma siguiente:

$$
\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u} \qquad (3\text{-}103)
$$

Ahora bien, teniendo en cuenta que estamos buscando un estimador insesgado de $\sigma^2$, vamos a calcular la esperanza de la expresión anterior:

$$
\begin{aligned}
E\left[\hat{\mathbf{u}}'\hat{\mathbf{u}}\right] &= E\left[\mathbf{u}'\mathbf{M}\mathbf{u}\right] = trE\left[\mathbf{u}'\mathbf{M}\mathbf{u}\right] = E\left[tr\mathbf{u}'\mathbf{M}\mathbf{u}\right] \\
&= E\left[tr\mathbf{M}\mathbf{u}\mathbf{u}'\right] = tr\mathbf{M}E\left[\mathbf{u}\mathbf{u}'\right] = tr\mathbf{M}\sigma^2\mathbf{I} \\
&= \sigma^2 tr\mathbf{M} = \sigma^2(n-k)
\end{aligned}
\qquad (3\text{-}104)
$$

En la deducción de (3-104), hemos utilizado la propiedad de la traza de que $tr(\mathbf{AB}) = tr(\mathbf{BA})$ ). Teniendo en cuenta esta propiedad de la traza, se obtiene el valor de $tr\mathbf{M}$:

$$
\begin{aligned}
tr\mathbf{M} &= tr\left[\mathbf{I}_{n\times n} - \mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\right] = tr\mathbf{I}_{n\times n} - tr\mathbf{X}\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}' \\
&= tr\mathbf{I}_{n\times n} - tr\mathbf{I}_{k\times k} = n-k
\end{aligned}
$$

De acuerdo con (3-104), se cumple que

$$
\sigma^2 = \frac{E\left[\hat{\mathbf{u}}'\hat{\mathbf{u}}\right]}{n-k} \qquad (3\text{-}105)
$$

A la vista de (3-105), un estimador de la varianza insesgado vendrá dado por:

$$
\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} \qquad (3\text{-}106)
$$

puesto que, de acuerdo con (3-104),

$$
E(\hat{\sigma}^2) = E\left[\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\right] = \frac{E(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{n-k} = \frac{\sigma^2(n-k)}{n-k} = \sigma^2 \qquad (3\text{-}107)
$$

El denominador de (3-106) son los grados de libertad que corresponden a la *SCR* que aparece en el numerador. Este resultado se justifica por el hecho de que las ecuaciones normales del hiperplano imponen *k* restricciones sobre los residuos. Por lo tanto, el número de grados de libertad de la *SCR* es igual al número de observaciones (*n*) menos el número de restricciones *k*.

**Apéndice 3.3 La consistencia del estimador de *MCO***

En el apéndice 2.8 hemos probado en el modelo de regresión simple la consistencia del estimador *MCO*. Ahora vamos a probar la consistencia del vector $\hat{\boldsymbol{\beta}}$ obtenido por *MCO*.

En primer lugar, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$, dado en (3-23) puede expresarse así

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{1}{n}\mathbf{X'X}\right)^{-1}\left(\frac{1}{n}\mathbf{X'u}\right) \tag{3-108}$$

Ahora, tomamos límites al último factor de (3-108) y llamamos $\mathbf{Q}$ al resultado:

$$\lim_{n\to\infty}\frac{1}{n}\mathbf{X'X} = \mathbf{Q} \tag{3-109}$$

Si $\mathbf{X}$ es fija en muestras repetidas, de acuerdo con el supuesto 2, entonces (3-109) implica que $\mathbf{Q}$=(l/$n$)$\mathbf{X'X}$. De acuerdo con el supuesto 3, y debido a que la matriz inversa es una función continua de la matriz original, existe $\mathbf{Q}^{-1}$. Por lo tanto, podemos escribir que

$$\text{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{Q}^{-1}\text{plim}\left[\frac{1}{n}\mathbf{X'u}\right]$$

El último término de (3-108) se puede expresar como

$$\begin{aligned}\frac{1}{n}\mathbf{X'u} &= \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ x_{21} & x_{22} & \cdots & x_{2i} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{ji} & \cdots & x_{jn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{ki} & \cdots & x_{kn} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_n \end{bmatrix} \\ &= \frac{1}{n}\begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_i & \cdots & \mathbf{x}_n\end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_n \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i u_i = \overline{\mathbf{x}_i u_i}\end{aligned} \tag{3-110}$$

donde $\mathbf{x}_i$ es el vector de la columna correspondiente a la observación *i-ésima*.

Ahora, vamos a calcular la esperanza y la varianza de (3-110),

$$E\left[\overline{\mathbf{x}_i u_i}\right] = \frac{1}{n}\sum_{i=1}^{n}E\left[\mathbf{x}_i u_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i E\left[u_i\right] = \frac{1}{n}\mathbf{X'}E\left[\mathbf{u}\right] = \mathbf{0} \tag{3-111}$$

$$var\left[\overline{\mathbf{x}_i u_i}\right] = E\left[\overline{\mathbf{x}_i u_i}\,\overline{(\mathbf{x}_i u_i)}'\right] = \frac{1}{n}\mathbf{X'}E\left[\mathbf{uu}'\right]\mathbf{X} = \frac{1}{n}\frac{\sigma^2}{n}\frac{\mathbf{X'X}}{n} = \frac{\sigma^2}{n^2}\mathbf{Q} \tag{3-112}$$

ya que $E\left[\mathbf{uu}'\right] = \sigma^2\mathbf{I}$ de acuerdo con los supuestos 7 y 8.

Tomando límites en (3-112), se sigue entonces que

$$\lim_{n\to\infty}\text{var}\left[\overline{\mathbf{x}_i\mathbf{u}_i}\right] = \lim_{n\to\infty}\frac{\sigma^2}{n^2}\mathbf{Q} = 0(\mathbf{Q}) = \mathbf{0} \tag{3-113}$$

Dado que la esperanza de $\overline{\mathbf{x}_i\mathbf{u}_i}$ es idéntica a cero y su varianza converge a cero, $\overline{\mathbf{x}_i\mathbf{u}_i}$ converge en media cuadrática a cero. La convergencia en media cuadrática implica la convergencia en probabilidad, por lo que plim($\overline{\mathbf{x}_i\mathbf{u}_i}$)=0. Por lo tanto,

$$\text{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{Q}^{-1}\text{plim}(\mathbf{x}_i\mathbf{u}_i) = \boldsymbol{\beta} + \mathbf{Q}^{-1}\text{plim}\left[\frac{1}{n}\mathbf{X}'\mathbf{u}\right] = \boldsymbol{\beta} + \mathbf{Q}^{-1} \times 0 = \boldsymbol{\beta} \quad (3\text{-}114)$$

En consecuencia, $\hat{\boldsymbol{\beta}}$ es un estimador consistente.

**Apéndice 3.4 Estimador de máxima verosimilitud**

El método de máxima verosimilitud es un método ampliamente utilizado en econometría. Este método propone como estimadores de los parámetros aquellos valores para los que la probabilidad de obtener las observaciones dadas es máxima. En la estimación de mínimos cuadrados no se adoptó *a priori* ningún supuesto; por el contrario, la estimación por máxima verosimilitud requiere establecer *a priori* supuestos estadísticos sobre los diversos elementos del modelo. Así, en la estimación por máxima verosimilitud vamos a adoptar todos los supuestos del modelo lineal clásico (*MLC*).

Por lo tanto, en la estimación por máxima verosimilitud de $\boldsymbol{\beta}$ y $\sigma^2$ en el modelo (3-52), se toman como estimadores a aquellos valores que maximizan la probabilidad de obtener las observaciones de una muestra dada.

Vamos a ver el procedimiento para obtener los estimadores de máxima verosimilitud $\boldsymbol{\beta}$ y $\sigma^2$ . De acuerdo con los supuestos del *MLC*

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad (3\text{-}115)$$

La esperanza y la varianza de la distribución de $\mathbf{y}$ están dadas por

$$E(\mathbf{y}) = E\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\right] = \mathbf{X}\boldsymbol{\beta} + E(\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} \quad (3\text{-}116)$$

$$\text{var}(\mathbf{y}) = E\left[\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\right] = E\left[\mathbf{u}\mathbf{u}'\right] = \sigma^2\mathbf{I} \quad (3\text{-}117)$$

Por lo tanto,

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad (3\text{-}118)$$

La función de densidad de $\mathbf{y}$ (o función de verosimilitud), considerando $\mathbf{X}$ e $\mathbf{y}$ fijas y $\boldsymbol{\beta}$ y $\sigma^2$ variables, será de acuerdo con (3-118) igual a

$$L = f(y \mid \boldsymbol{\beta}, \sigma^2) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\left(1/2\sigma^2\right)\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right) \quad (3\text{-}119)$$

Dado que el máximo para $L$ se alcanza en el mismo punto que ln($L$), por ser la función logaritmo monótona, podemos, a efectos de maximización, trabajar con ln($L$) en lugar de con $L$. Entonces,

$$\ln(L) = -\frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - \frac{1}{2\sigma^2} - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (3\text{-}120)$$

Para maximizar ln($L$), hay que derivar con respecto a $\boldsymbol{\beta}$ y $\sigma^2$:

$$\frac{\delta \ln(L)}{\delta \boldsymbol{\beta}} = -\frac{1}{2\sigma^2}(-2\mathbf{X'y} + 2\mathbf{X'X\beta}) \tag{3-121}$$

$$\frac{\delta \ln(L)}{\delta \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y - X\beta})'(\mathbf{y - X\beta})}{2\sigma^4} \tag{3-122}$$

Igualando (3-121) a cero, vemos que el estimador de máxima verosimilitud de $\boldsymbol{\beta}$, denotado por $\tilde{\boldsymbol{\beta}}$, satisface que

$$\mathbf{X'X}\tilde{\beta} = \mathbf{X'y} \tag{3-123}$$

Como se supone que $\mathbf{X'X}$ es invertible, tenemos que

$$\tilde{\boldsymbol{\beta}} = \left[\mathbf{X'X}\right]^{-1} = \mathbf{X'y} \tag{3-124}$$

En consecuencia, el estimador de máxima verosimilitud de $\boldsymbol{\beta}$, bajo los supuestos del *MLC*, coincide con el estimador *MCO*, es decir,

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} \tag{3-125}$$

Por lo tanto,

$$(\mathbf{y - X\tilde{\beta}})'(\mathbf{y - X\tilde{\beta}}) = (\mathbf{y - X\hat{\beta}})'(\mathbf{y - X\hat{\beta}}) = \mathbf{\hat{u}'\hat{u}} \tag{3-126}$$

Igualando (3-122) a cero y sustituyendo $\boldsymbol{\beta}$ por $\tilde{\boldsymbol{\beta}}$, obtenemos:

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{\mathbf{\hat{u}'\hat{u}}}{2\tilde{\sigma}^4} = 0 \tag{3-127}$$

donde hemos designado por $\tilde{\sigma}^2$ al estimador de máxima verosimilitud de la varianza de las perturbaciones aleatorias. De (3-127) se deduce que

$$\tilde{\sigma}^2 = \frac{\mathbf{\hat{u}'\hat{u}}}{n} \tag{3-128}$$

Como puede verse el estimador de máxima verosimilitud no es igual al estimador insesgado que se ha obtenido en (3-106). De hecho, si tomamos esperanzas en (3-128),

$$E\left[\tilde{\sigma}^2\right] = \frac{1}{n}E\left[\mathbf{\hat{u}'\hat{u}}\right] = \frac{n-k}{n}\sigma^2 \tag{3-129}$$

Es decir, el estimador de máxima verosimilitud, $\tilde{\sigma}^2$, es un estimador sesgado, aunque su sesgo tiende a cero cuando $n$ tiende a infinito, ya que

$$\lim_{n\to\infty}\frac{n-k}{n} = 1 \tag{3-130}$$