Modelo de regresión lineal simple

1 Introducción

Con frecuencia, nos encontramos en economía con modelos en los que el comportamiento de una variable, *Y,* se puede explicar a través de una variable *X;* lo que representamos mediante

$$Y = f(X) \tag{1}$$

Si consideramos que la relación *f,* que liga *Y* con *X*, es lineal, entonces (1) se puede escribir así:

$$Y_t = \beta_1 + \beta_2 X_t \tag{2}$$

Como quiera que las relaciones del tipo anterior raramente son exactas, sino que más bien son aproximaciones en las que se han omitido muchas variables de importancia secundaria, debemos incluir un término de perturbación aleatoria, u_t, que refleja todos los factores – distintos de *X* -que influyen sobre la variable endógena, pero que ninguno de ellos es relevante individualmente. Con ello, la relación quedaría de la siguiente forma:

Modelo de regresión simple

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{3}$$

La expresión anterior refleja una relación lineal, y en ella sólo figura una única variable explicativa, recibiendo el nombre de relación lineal simple. El calificativo de simple se debe a que solamente hay una variable explicativa. Supongamos ahora que disponemos de T observaciones de la variable Y ($Y_1, Y_2, \ldots, Y_T$) y de las correspondientes observaciones de X ($X_1, X_2, \ldots, X_T$). Si hacemos extensiva (3) a la relación entre observaciones, tendremos el siguiente conjunto de T ecuaciones:

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_1 + u_1 \\ Y_2 &= \beta_1 + \beta_2 X_2 + u_2 \\ \cdots &\qquad \cdots \\ Y_T &= \beta_1 + \beta_2 X_T + u_T \end{aligned} \tag{4}$$

El sistema de ecuaciones (4) se puede escribir abreviadamente de la forma siguiente:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T \tag{5}$$

El objetivo principal de la regresión es la determinación o estimación de β_1 y β_2 a partir de la información contenida en las observaciones de que disponemos. Esta estimación se puede llevar a cabo mediante diversos procedimientos. A continuación se analizan en detalle algunos de los métodos posibles.

Interesa, en primer lugar, realizar una aproximación intuitiva a diferentes criterios de ajuste. Para ello se utiliza la representación gráfica de las observaciones (X_t, Y_t), con t = 1, 2,..., T. Si la relación lineal de dependencia entre Y y X fuera exacta, las observaciones se situarían a lo largo de una recta (véase la figura 1). En ese caso, las estimaciones más adecuadas de β_1 y β_2 – de hecho, los verdaderos valores – serían, respectivamente, la ordenada en el origen y la pendiente de dicha recta.

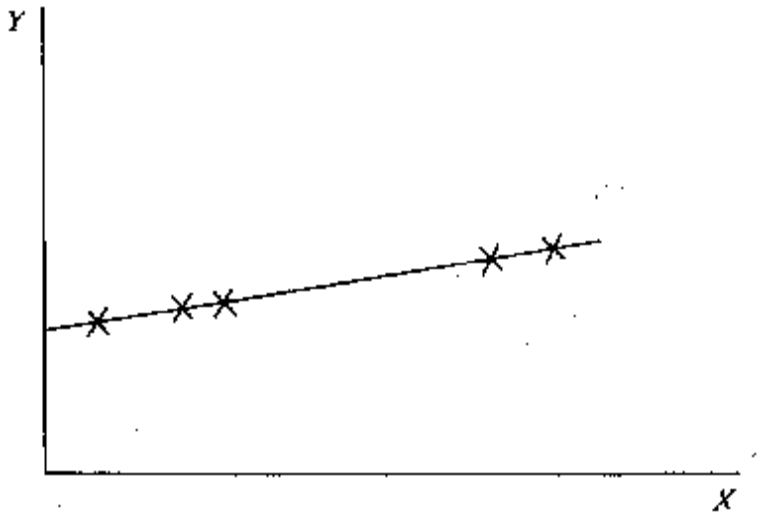

Figura 1

Pero si la dependencia entre Y y X es estocástica, entonces, en general, las observaciones no se alinearán a lo largo de una recta, sino que formarán una nube de puntos, como aparece en la figura 2. En ese caso, podemos contemplar las estimaciones de β_1 y β_2 como la ordenada en el origen y la pendiente de una recta *próxima* a los puntos. Así, si designamos mediante $\hat{\beta}_1$ y $\hat{\beta}_2$ las estimaciones de β_1 y β_2, respectivamente, la ordenada de la recta para el valor X_t vendrá dada por

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t \tag{6}$$

El problema que tenemos planteado es, pues, hallar unos estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ tales que la recta que pasa por los puntos (X_t, $\hat{Y}_t$) se ajuste lo mejor posible a los puntos (X_t, Y_t). Se denomina error o residuo a la diferencia entre el valor observado de la variable endógena y el valor ajustado, es decir,

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t \tag{7}$$

Teniendo en cuenta el concepto de residuo se analizan a continuación diversos criterios de ajuste.

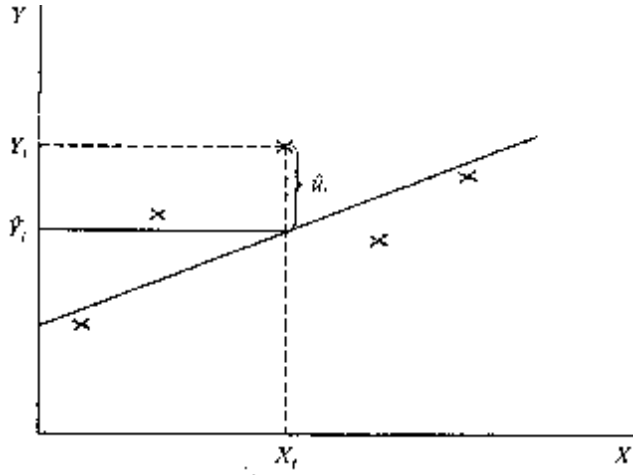

Figura 2

Un primer criterio consistiría en tomar como estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ aquellos valores que hagan la suma de todos los residuos tan próxima a cero como sea posible. Con este criterio la expresión a minimizar sería la siguiente:

$$\left| \sum_{t=1}^{T} \hat{u}_t \right| \tag{8}$$

El problema fundamental de este método de estimación radica en que los residuos de distinto signo pueden compensarse. Tal situación puede observarse gráficamente en la figura 3, en la que se representan tres observaciones alineadas, (X_1, Y_1), (X_2, Y_2) y (X_3, Y_3), tales que

$$\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{Y_3 - Y_1}{X_3 - X_1}.$$

Si se ajusta una recta que pase por los tres puntos, cada uno de los residuos tomará el valor cero, de forma que

$$\left| \sum_{t=1}^{T} \hat{u}_t = 0 \right|$$

Dicho ajuste se podría considerar óptimo. Pero también es posible que $\left| \sum_{t=1}^{3} \hat{u}_t = 0 \right|$ haciendo girar en cualquier sentido la recta si dejamos fijo (X_2, Y_2), como muestra la figura 2, debido a que $\hat{u}_3 = -\hat{u}_1$. Este sencillo ejemplo nos muestra que este criterio no es apropiado para la estimación de β_1 y β_2, debido a

que, para cualquier conjunto de observaciones, existen infinitas rectas que lo satisfacen. Otra forma de evitar la compensación de residuos positivos con negativos consiste en tomar los valores absolutos de los residuos. En este caso se minimizaría la siguiente expresión:

$$\sum_{t=1}^{T} |\hat{u}_t| \tag{9}$$

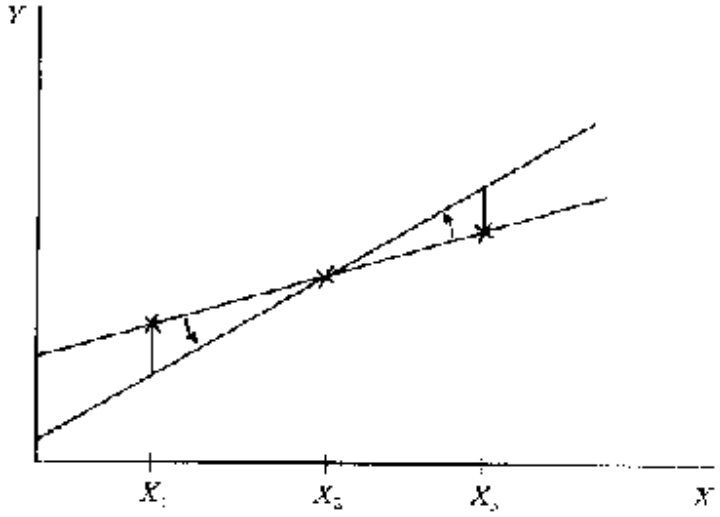

Figura 3

Desgraciadamente, aunque los estimadores así obtenidos tienen algunas propiedades interesantes, su cálculo es complicado, requiriendo la resolución de un problema de programación lineal o la aplicación de un procedimiento de cálculo iterativo. Un tercer método consiste en minimizar la suma de los cuadrados de los residuos, es decir,

$$S = \sum_{t=1}^{T} \hat{u}_t^2 \tag{10}$$

Los estimadores obtenidos con arreglo al criterio expresado en (10) se denominan mínimo-cuadráticos, y gozan de ciertas propiedades estadísticas deseables, que se estudian posteriormente. Por otra parte, frente al primero de los criterios examinados, al tomar los cuadrados de los residuos se evita la compensación de éstos, mientras que, a diferencia del segundo de los criterios, los estimadores mínimo-cuadráticos son sencillos de obtener. Es importante señalar que, desde el momento en que tomamos los cuadrados de los residuos, estamos penalizando más que proporcionalmente a los residuos grandes frente a los pequeños (si un residuo es el doble que otro, su cuadrado será cuatro veces mayor), lo que caracteriza también a la estimación mínimo-cuadrática frente a otros posibles métodos.

2. Obtención de los estimadores mínimo-cuadráticos

A continuación se expone el proceso para la obtención por mínimos cuadrados de los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$. El objetivo es minimizar la suma de los cuadrados de los residuos (*S*). Para ello, en primer lugar expresamos *S* en función de los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$:

$$S = \sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t)^2 \tag{11}$$

Para minimizar *S*, derivamos parcialmente respecto a $\hat{\beta}_1$ y $\hat{\beta}_2$:

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2\sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t)$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2\sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) X_t \tag{12}$$

Los estimadores mínimo-cuadráticos se obtienen igualando las anteriores derivadas a cero:

$$-2\sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) = 0$$

$$-2\sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) X_t = 0 \tag{13}$$

Operando, se tiene que

$$\sum_{t=1}^{T} Y_t = \hat{\beta}_1 T + \hat{\beta}_2 \sum_{t=1}^{T} X_t$$

$$\sum_{t=1}^{T} Y_t X_t = \hat{\beta}_1 \sum_{t=1}^{T} X_t + \hat{\beta}_2 \sum_{t=1}^{T} X_t^2 \tag{14}$$

Las ecuaciones (14) se denominan *ecuaciones normales de la recta de regresión*. Resolviendo este sistema, según puede verse en el recuadro adjunto, a partir de (21) se obtiene de forma inmediata el estimador de $\hat{\beta}_2$:

$$\hat{\beta}_2 = \frac{\sum_{t=1}^{T} (Y_t - \overline{Y})(X_t - \overline{X})}{\sum_{t=1}^{T} (X_t - \overline{X})^2} \tag{15}$$

Resolución del sistema de ecuaciones (14)

Dividiendo la primera ecuación normal en (14) por T se obtiene:

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X} \tag{16}$$

donde

$$\overline{Y} = \frac{\sum_{t=1}^{T} Y_t}{T} \qquad \overline{X} = \frac{\sum_{t=1}^{T} X_t}{T}$$

De acuerdo con la anterior expresión se obtiene

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} \tag{17}$$

Sustituyendo $\hat{\beta}_1$ en la segunda ecuación normal (14) se tienen que

$$\begin{aligned} \sum_{t=1}^{T} Y_t X_t &= (\overline{Y} - \hat{\beta}_2 \overline{X}) \sum_{t=1}^{T} X_t + \hat{\beta}_2 \sum_{t=1}^{T} X_t^2 \\ \sum_{t=1}^{T} Y_t X_t - \overline{Y} \sum_{t=1}^{T} X_t &= \hat{\beta}_2 \left[\sum_{t=1}^{T} X_t^2 - \overline{X} \sum_{t=1}^{T} X_t \right] \end{aligned} \tag{18}$$

Por otra parte,

$$\begin{aligned} \sum_{t=1}^{T} (Y_t - \overline{Y})(X_t - \overline{X}) &= \sum_{t=1}^{T} (Y_t X_t - \overline{X} Y_t - \overline{Y} X_t + \overline{Y}\overline{X}) \\ &= \sum_{t=1}^{T} Y_t X_t - \overline{Y} \sum_{t=1}^{T} X_t - \overline{X} \sum_{t=1}^{T} Y_t + T\overline{Y}\overline{X} = \\ &= \sum_{t=1}^{T} Y_t X_t - \overline{Y} \sum_{t=1}^{T} X_t - \overline{X} T \overline{Y} + T\overline{Y}\overline{X} \\ &= \sum_{t=1}^{T} Y_t X_t - \overline{Y} \sum_{t=1}^{T} X_t \end{aligned} \tag{19}$$

$$\begin{aligned} \sum_{t=1}^{T} (X_t - \overline{X})^2 &= \sum_{t=1}^{T} (X_t^2 - 2\overline{X} X_t + \overline{X}^2) \\ &= \sum_{t=1}^{T} X_t^2 - 2\overline{X} \sum_{t=1}^{T} X_t + T\overline{X}^2 = \sum_{t=1}^{T} X_t^2 - 2\overline{X} \sum_{t=1}^{T} X_t + \overline{X} \sum_{t=1}^{T} X_t \\ &= \sum_{t=1}^{T} X_t^2 - \overline{X} \sum_{t=1}^{T} X_t \end{aligned} \tag{20}$$

Teniendo en cuenta (19) y (20), entonces (18) se puede expresar así:

$$\sum_{t=1}^{T} (Y_t - \overline{Y})(X_t - \overline{X}) = \hat{\beta}_2 \left[\sum_{t=1}^{T} (X_t - \overline{X})^2 \right] \tag{21}$$

A su vez $\hat{\beta}_1$ se obtiene a través de la relación (17). Es decir,

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} \tag{22}$$

Dividiendo numerador y denominador (15) por *T* se tiene que

$$\hat{\beta}_2 = \frac{\dfrac{\sum_{t=1}^{T}(Y_t - \overline{Y})(X_t - \overline{X})}{T}}{\dfrac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T}} = \frac{\text{cov}(X,Y)}{\text{var}(X)} \tag{23}$$

De acuerdo con (23), la estimación de $\hat{\beta}_2$ se obtiene dividiendo la covarianza muestral de *X* e *Y* por la varianza muestral de *X*. Dado que la varianza de *X* no puede ser negativa, el signo de $\hat{\beta}_2$ será el mismo que el de la covarianza muestral de *X* e *Y*.

3. Propiedades descriptivas en la regresión lineal simple

Las propiedades que se exponen a continuación son propiedades derivadas exclusivamente de la aplicación del método de estimación por mínimos cuadrados al modelo de regresión lineal simple, en el que se incluye como primer regresor el término independiente.

1. *La suma de los residuos mínimo-cuadráticos es igual a cero*:

$$\sum_{t=1}^{T} \hat{u}_t = 0 \tag{24}$$

Demostración.

Por definición de residuo

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t \qquad t = 1, 2, \cdots, T \tag{25}$$

Si sumamos para las *T* observaciones, se obtiene:

$$\sum_{t=1}^{T} \hat{u}_t = \sum_{t=1}^{T} Y_t - \sum_{t=1}^{T} \hat{Y}_t = \sum_{t=1}^{T} Y_t - \hat{\beta}_1 T - \hat{\beta}_2 \sum_{t=1}^{T} X_t \tag{26}$$

Por otra parte, la primera ecuación del sistema de ecuaciones normales (14) es igual a

$$\sum_{t=1}^{T} Y_t = T\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^{T} X_t \tag{27}$$

Al comparar (26) y (27), se concluye que necesariamente debe cumplirse (24). Obsérvese que, al cumplirse (24), se cumplirá también que

$$\sum_{t=1}^{T} Y_t = \sum_{t=1}^{T} \hat{Y}_t \tag{28}$$

y, al dividir por *T,* tenemos

$$\overline{Y} = \overline{\hat{Y}} \tag{29}$$

2. La recta de regresión pasa necesariamente por el punto ($\overline{Y}, \overline{X}$).

Demostración.

En efecto, dividiendo por T la ecuación (27) se obtiene:

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X} \tag{30}$$

3. La suma de los productos cruzados entre la variable explicativa y los residuos es igual a 0, es decir,

$$\sum_{t=1}^{T} \hat{u}_t X_t = 0 \tag{31}$$

Demostración.

En efecto,

$$\sum_{t=1}^{T} \hat{u}_t X_t = \sum_{t=1}^{T} (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) X_t = 0$$

Para llegar a (31) se ha tenido en cuenta la segunda ecuación normal de (13).

4. La suma de los productos cruzados entre los valores ajustados y los residuos es igual a 0, es decir,

$$\sum_{t=1}^{T} \hat{u}_t \hat{Y}_t = 0 \tag{32}$$

Demostración.

En efecto, si se tiene en cuenta (17) resulta que

$$\sum_{t=1}^{T}\hat{u}_t\hat{Y}_t=\sum_{t=1}^{T}\hat{u}_t(\hat{\beta}_1+\hat{\beta}_2X_t)=\hat{\beta}_1\sum_{t=1}^{T}\hat{u}_t+\hat{\beta}_2\sum_{t=1}^{T}\hat{u}_tX_t=0$$

Para llegar a (32) se ha tenido en cuenta las propiedades descriptivas 1 y 3.

4 Medidas de la bondad del ajuste. Coeficiente de determinación

Una vez que se ha realizado el ajuste por mínimos cuadrados, conviene disponer de algún indicador que permita medir el grado de ajuste entre el modelo y los datos. En el caso de que se haya estimado varios modelos alternativos podría utilizarse medidas de este tipo, a las que se denomina medidas de la bondad del ajuste, para seleccionar el modelo más adecuado.

Existen en la literatura econométrica numerosas medidas de la bondad del ajuste. La más conocida es el coeficiente de determinación, al que se designa por R^2 o R cuadrado. Como se verá en otro momento, esta medida tienen algunas limitaciones, aunque es válida para comparar modelos de regresión lineal simple.

El coeficiente de determinación se basa en la descomposición de la varianza de la variable endógena, a la que denominaremos *varianza total.* Vamos a ver a continuación como se obtiene esta descomposición.

De acuerdo con (7)

$$Y_t=\hat{Y}_t+\hat{u}_t \tag{33}$$

Restando a ambos miembros $\overline{Y}$, se tiene que

$$Y_t-\overline{Y}=\hat{Y}_t-\overline{Y}+\hat{u}_t \tag{34}$$

Si elevamos al cuadrado ambos miembros se obtiene que

$$\left(Y_t-\overline{Y}\right)^2=\left[(\hat{Y}_t-\overline{Y})+\hat{u}_t\right]^2 \tag{35}$$

es decir,

$$\left(Y_t-\overline{Y}\right)^2=\left(\hat{Y}_t-\overline{Y}\right)^2+\hat{u}_t^2-2\hat{u}_t\left(\hat{Y}_t-\overline{Y}\right) \tag{36}$$

Sumando ambos miembros de la expresión anterior de 1 a T, se tiene

$$\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2=\sum_{t=1}^{T}\left(\hat{Y}_t-\overline{Y}\right)^2+\sum_{t=1}^{T}\hat{u}_t^2-2\sum_{t=1}^{T}\hat{u}_t\left(\hat{Y}_t-\overline{Y}\right) \tag{37}$$

Ahora bien, puede verse que el tercer término del segundo miembro de (37) es

$$2\sum_{t=1}^{T}\hat{u}_t\left(\hat{Y}_t-\overline{Y}\right)=2\sum_{t=1}^{T}\hat{u}_t\hat{Y}_t-2\overline{Y}\sum_{t=1}^{T}\hat{u}_t=0 \tag{38}$$

de acuerdo con (31) y (24). Por lo tanto, (37) queda reducida a

$$\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2=\sum_{t=1}^{T}\left(\hat{Y}_t-\overline{Y}\right)^2+\sum_{t=1}^{T}\hat{u}_t^2 \tag{39}$$

Debe recalcarse que para se cumpla que (38) es igual a 0 es necesario utilizar la relación (24), que a su vez está asociada a la primera ecuación normal de la recta, es decir, a la ecuación correspondiente al término independiente. Si en modelo no hay término independiente, entonces en general no se cumplirá la descomposición obtenida en (39).

Si en la expresión (39) dividimos ambos miembros por *T*, se obtiene que

$$\frac{\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2}{T}=\frac{\sum_{t=1}^{T}\left(\hat{Y}_t-\overline{Y}\right)^2}{T}+\frac{\sum_{t=1}^{T}\hat{u}_t^2}{T} \tag{40}$$

Por lo tanto, la varianza total de la variable endógena se descompone en dos partes: varianza *explicada* por la regresión o varianza de los valores ajustados[1] y varianza residual. Es decir,

Varianza total = varianza "explicada" + varianza residual

A partir de la descomposición anterior, el coeficiente de determinación se define como la proporción de la varianza total explicada por la regresión. Su expresión es la siguiente:

$$R^2=\frac{\sum_{t=1}^{T}(\hat{Y}_t-\overline{Y})^2}{\sum_{t=1}^{T}(Y_t-\overline{Y})^2} \tag{41}$$

Alternativamente, y de forma equivalente, de acuerdo con (39) el coeficiente de determinación se puede definir como 1 menos la proporción no explicada por la regresión, es decir, como

[1] El primer término del segundo miembro de (40) es la varianza de $\hat{Y}_t$, ya que, de acuerdo con (29), se verifica que $\overline{Y}=\overline{\hat{Y}}$

$$R^2 = 1 - \frac{\sum_{t=1}^{T} \hat{u}_t^2}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} \tag{42}$$

Los valores extremos del coeficiente de determinación son: 0, cuando la varianza explicada es nula, y 1,cuando la varianza residual es nula, es decir, cuando el ajuste es perfecto

5 Hipótesis estadísticas del modelo

I Hipótesis sobre la forma funcional

Los elementos del modelo tienen la siguiente relación entre sí:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{43}$$

La relación entre el regresando, los regresores y la perturbación aleatoria es lineal. El regresando y los regresores pueden ser cualquier función de la variable endógena o de las variables predeterminadas, respectivamente, siempre que entre regresando y regresores se mantenga una relación lineal, es decir, el modelo sea lineal en los parámetros. El carácter aditivo de la perturbación aleatoria garantiza su relación lineal con el resto de los elementos.

II Hipótesis sobre la perturbación aleatoria

La perturbación aleatoria u_t es una variable aleatoria no observable con las siguientes propiedades:

a) La esperanza matemática de la perturbación aleatoria u_t es cero.

$$E(u_t) = 0 \qquad t = 1,\ 2, \ldots, T \tag{44}$$

Se adopta aquí el supuesto de que los efectos individuales de las variables incluidas en el término de perturbación tienden a compensarse por término medio. En cualquier caso, aun suponiendo que los efectos individuales no se compensasen exactamente y, por tanto, su valor esperado fuese distinto de cero, dicho valor podría ser acumulado en el término constante del modelo de regresión, con lo cual se podría mantener esta hipótesis sin ningún problema. Por esta razón, si el modelo tiene término constante, es imposible deslindar *a posteriori* la parte estrictamente correspondiente al coeficiente independiente del modelo, de la parte proveniente de la media de la perturbación aleatoria del modelo. Así, pues, ésta seria una hipótesis no contrastable empíricamente.

b) Las perturbaciones aleatorias son homoscedásticas

$$E(u_t^2) = \sigma^2 \qquad t = 1,\ 2, \ldots, T \tag{45}$$

Esta hipótesis indica que todas las perturbaciones aleatorias tienen la misma varianza. Es decir, la varianza de las perturbaciones aleatorias del modelo es constante y, por tanto, independiente del tiempo o de los valores de las variables predeterminadas. Dicha hipótesis es contrastable empíricamente mediante diversos contrastes estadísticos basados en los residuos mínimo-cuadráticos. Asimismo, hay que señalar que, en determinadas situaciones, esta hipótesis resulta poco plausible, sobre todo cuando se trabaja con datos de corte transversal, es decir, con observaciones sobre diferentes unidades muestrales referidas a un mismo momento del tiempo. Si no se cumple esta hipótesis, se dice que las perturbaciones son heteroscedásticas.

c) Las perturbaciones aleatorias con distintos subíndices son independientes entre sí.

$$E(u_t u_s) = 0 \qquad t \neq s \tag{46}$$

Es decir, las perturbaciones correspondientes a distintos momentos del tiempo o a distintas unidades muestrales no están correlacionadas entre si. Este supuesto, al igual que el anterior, es contrastable *a posteriori.* La transgresión del mismo se produce con bastante frecuencia en los modelos en los que se utilizan datos de series temporales, es decir, observaciones realizadas a intervalos regulares de tiempo.

d) La perturbación aleatoria tiene una distribución normal multivariante

Dado que la perturbación aleatoria recoge un conjunto amplio de variables, omitidas del modelo de regresión, que son independientes entre si y también del conjunto de regresores, por el teorema central del limite se puede suponer que el vector de perturbaciones aleatorias tiene una distribución normal multivariante.

Las cuatro hipótesis formuladas sobre las perturbaciones aleatorias se pueden expresar de forma conjunta como

$$u_t \sim NID(0, \sigma^2) \tag{47}$$

donde NID indica que son normales e independientes.

III Hipótesis sobre el regresor *X*

a) Las observaciones de X son fijas en repetidas muestras

De acuerdo con esta hipótesis, los distintos regresores del modelo toman los mismos valores para diversas muestras del regresando. Éste es un supuesto fuerte en el caso de las ciencias sociales, en el que es poco viable experimentar. Los datos se obtienen por observación, y no por experimentación. Para que dicho supuesto se cumpliera, los regresores deberían ser susceptibles de ser controlados por parte del investigador. Es importante señalar que los resultados que se

obtienen utilizando este supuesto se mantendrían prácticamente idénticos si supusiéramos que los regresores son estocásticos, siempre que introdujéramos el supuesto adicional de independencia entre los regresores y la perturbación aleatoria. Este supuesto alternativo se puede formular así:

a) La variable X se distribuye independientemente de la perturbación aleatoria*

En desarrollos posteriores se adoptará el supuesto de que se cumple la hipótesis a).

b) El regresor X no contiene errores de observación o de medida

Ésta es una hipótesis que raramente se cumple en la práctica, ya que los instrumentos de medición en economía son escasamente fiables (piénsese en la multitud de errores que es posible cometer en una recogida de información, mediante encuesta, sobre los presupuestos familiares). Aunque es difícil encontrar instrumentos para contrastar esta hipótesis, la naturaleza del problema y, sobre todo, la procedencia de los datos utilizados pueden ofrecer evidencia favorable o desfavorable a la hipótesis enunciada.

IV Hipótesis sobre los parámetros

β_1 y β_2 son constantes

Si no se adopta esta hipótesis el modelo de regresión sería muy complicado de manejar. En todo caso, puede ser aceptable postular que los parámetros del modelo se mantienen estables en el tiempo (si no se trata de períodos muy extensos) o en el espacio (si está relativamente acotado).

6 Propiedades probabilísticas del modelo

Aleatoriedad del modelo

Dado que u_t es aleatoria, también la variable endógena Y_t será una variable aleatoria por ser una función lineal de la perturbación aleatoria, como se deduce del modelo de regresión lineal (43).

Cuando realizamos una estimación por mínimos cuadrados con datos reales, estamos suponiendo que existe un mecanismo de generación de datos - el modelo de regresión - que ha determinado los valores observados de la variable endógena. Así, cuando realizamos una estimación en un modelo de regresión lineal simple, tal como el modelo (43), estamos suponiendo que los valores observados por el investigador de la variable endógena ($Y_1,Y_2,...,Y_T$) han sido generados por dicha relación que contiene unos parámetros (β_1 y β_2) desconocidos para el investigador, una variable explicativa (X) con valores conocidos y una perturbación aleatoria u cuyos valores son desconocidos. El investigador supone que los valores de la perturbación aleatoria han sido

generados por una distribución normal con media 0 y varianza σ^2, también desconocida.

Así pues, el investigador no observa directamente el proceso de generación de datos, sino los resultados finales de este proceso, es decir, los valores observados de Y: $Y_1,Y_2,...,Y_T$. Precisamente, aplicando el método de mínimos cuadrados (a estos datos y a los datos de la variable explicativa), lo que se persigue es realizar estimaciones de los parámetros del modelo (β_1, β_2 y σ^2), que son desconocidos para el investigador.

Con objeto de comprender mejor el proceso que acabamos de describir, es conveniente invertir los papeles, generando el propio investigador los valores que toma la variable endógena. Cuando se generan los datos de forma artificial, se dice que se está realizando un experimento de Montecarlo. Este nombre proviene del famoso casino de la Costa Azul debido a que en estos experimentos se realizan extracciones de números aleatorios, lo que en definitiva es análogo al resultado del lanzamiento de una bola en la ruleta de un casino.

En la realización de un experimento de Montecarlo se parte del supuesto de que es conocido tanto el mecanismo de generación de datos, como los valores de los parámetros[2].

Aleatoriedad de los estimadores

Los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ son también variables aleatorias puesto que son función de las variables aleatorias Y_t. En efecto,

$$\hat{\beta}_2 = \frac{\sum_{t=1}^{T}(X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} = \frac{\sum_{t=1}^{T}(X_t - \bar{X})Y_t - \sum_{t=1}^{T}(X_t - \bar{X})\bar{Y}}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

$$= \frac{\sum_{t=1}^{T}(X_t - \bar{X})Y_t}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \qquad (48)$$

En el desarrollo anterior se ha tenido en cuenta que

[2] En *Econometría Aplicada* (páginas 60 a 66) puede verse como se generan números aleatorios uniformes y normales mediante rutinas informáticas. Por otra parte, en las páginas 149 a 153 (caso 3.11) se realiza un experimento de Montecarlo con una hipotética función de consumo.

$$\sum_{t=1}^{T}(X_t-\bar{X})\bar{Y}=\bar{Y}\sum_{t=1}^{T}(X_t-\bar{X})=\bar{Y}\left[\sum_{t=1}^{T}X_t-T\bar{X}\right]=\bar{Y}\left[T\bar{X}-T\bar{X}\right]=0$$

Denominando

$$\frac{(X_t-\bar{X})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=c_t$$

entonces el estimador $\hat{\beta}_2$ se puede expresar de la siguiente forma:

$$\hat{\beta}_2=\sum_{t=1}^{T}c_tY_t \tag{49}$$

Si se adopta el supuesto III a), que implica que la variable X_t es no aleatoria, entonces de la expresión anterior se deduce que $\hat{\beta}_2$ es una combinación lineal de la variable Y_t.

Los coeficientes c_t tienen las siguientes propiedades:

$$\sum_{t=1}^{T}c_t=0 \tag{50}$$

$$\sum_{t=1}^{T}c_tX_t=1 \tag{51}$$

En efecto,

$$\sum_{t=1}^{T}c_t=\frac{\sum_{t=1}^{T}(X_t-\bar{X})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=\frac{\sum_{t=1}^{T}X_t-T\bar{X}}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=\frac{T\bar{X}-T\bar{X}}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=0$$

$$\sum_{t=1}^{T}c_tX_t=\frac{\sum_{t=1}^{T}(X_t-\bar{X})X_t}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=\frac{\sum_{t=1}^{T}(X_t-\bar{X})(X_t-\bar{X})}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=\frac{\sum_{t=1}^{T}(X_t-\bar{X})^2}{\sum_{t=1}^{T}(X_t-\bar{X})^2}=1$$

Vamos a expresar a ahora el estimador $\hat{\beta}_2$ en función de las perturbaciones aleatorias. Teniendo en cuenta (49) y (43) resulta que

$$\hat{\beta}_2=\sum_{t=1}^{T}c_t(\beta_1+\beta_2X_t+u_t)$$

$$= \beta_1 \sum_{t=1}^{T} c_t + \beta_2 \sum_{t=1}^{T} c_t X_t + \sum_{t=1}^{T} c_t u_t = \beta_2 + \sum_{t=1}^{T} c_t u_t \qquad (52)$$

Para llegar al resultado final se ha tenido en cuenta (50) y (51). Análogamente,

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} = \beta_1 + \beta_2 \overline{X} + \overline{u} - \hat{\beta}_2 \overline{X} = \beta_1 + \overline{u} - \overline{X}(\hat{\beta}_2 - \beta_2)$$

$$= \beta_1 + \frac{1}{T} \sum_{t=1}^{T} u_t - \overline{X} \sum_{t=1}^{T} c_t u_t \qquad (53)$$

EJEMPLO 1 Estimación de la función de consumo con series simuladas

Con el mismo modelo que el caso 3.11 de *Econometría Aplicada*, en un experimento de Montecarlo al que denominaremos Exp. 1, hemos generado 10 series de consumo (*CONS)* a partir de la relación:

$$CONS_t = 2 + 0,85 \times RENDIS_t + u_t \qquad (54)$$

donde *RENDIS* es la renta disponible y la perturbación *u* se distribuye con media 0 y desviación típica 1. (La única variación con respecto al caso 3.11 es que en dicho caso la desviación típica de la perturbación es 1,2.)

Aplicando mínimos cuadrados utilizando cada una de las muestras generadas de consumo y de la muestra dada de la variable RENDIS (Véase cuadro 3.11 de *Econometría Aplicada*), se han estimado (véase cuadro 1) los parámetros β_1 y β_2 del modelo:

$$CONS_t = \beta_1 + \beta_2 \times RENDIS_t + u_t \qquad (55)$$

CUADRO 1 Resultados Exp. 1
Desviación típica de las perturbaciones: Constante ($\sigma = 1$)
Desviación típica de la muestra de RENDIS: Constante ($S_{RENDIS} = 2.905$)

Núm. muestra			**Desviaciones típicas teóricas**			**Desviaciones típicas estimadas**		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\sigma_{\hat{\beta}_1}$	$\sigma_{\hat{\beta}_2}$	$\hat{\sigma}$	$\hat{\sigma}_{\hat{\beta}_1}$	$\hat{\sigma}_{\hat{\beta}_2}$	R^2
1	2.993	0.803	2.3509	0.1019	0.6285	1.4774	0.0684	0.945
2	-0.408	0.977	2.3509	0.1019	0.7238	1.7014	0.0788	0.951
3	0.759	0.941	2.3509	0.1019	1.1150	2.6210	0.1214	0.883
4	4.077	0.766	2.3509	0.1019	1.0440	2.4541	0.1136	0.853
5	1.062	0.887	2.3509	0.1019	0.6496	1.5271	0.0707	0.951
6	2.197	0.832	2.3509	0.1019	1.3934	3.2755	0.1517	0.790
7	-1.359	0.973	2.3509	0.1019	1.1117	2.6134	0.1210	0.890
8	1.594	0.853	2.3509	0.1019	1.1466	2.6954	0.1248	0.854
9	2.917	0.807	2.3509	0.1019	0.9516	2.2369	0.1036	0.884
10	4.194	0.741	2.3509	0.1019	0.9270	2.1790	0.1009	0.871
Media	**1.803**	**0.858**						

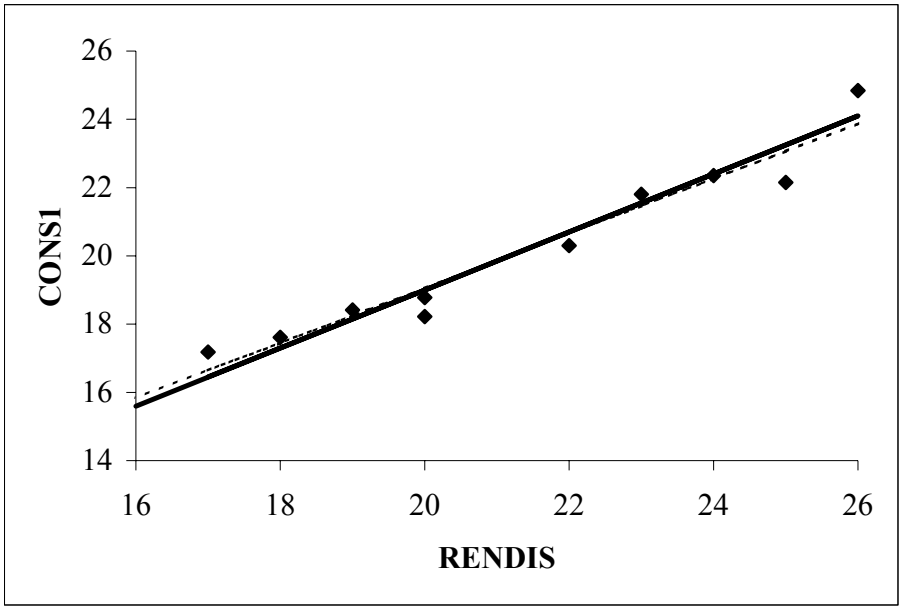

Figura 4. Recta de regresión teórica (trazo continuo) y estimada en la muestra 1 del Exp. 1 (trazo discontinuo)

En la figura 4, además de la nube de puntos, se han representado la recta de regresión teórica (en trazo continuo) y la recta de regresión estimada con los datos de la muestra 1. Como puede verse, la recta ajustada está muy próxima a la recta teórica.

Insesgadez de los coeficientes

Una propiedad deseable en un estimador es que sea insesgado, es decir, que su media teórica coincida con el parámetro que trata de estimar. Veamos

concretamente, y de forma analítica, si se verifica esta propiedad en los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$. Tomando esperanza matemática en (52) y (53), y teniendo en cuenta la hipótesis 2a), se obtiene que

$$E(\hat{\beta}_2) = E\left[\beta_2 + \sum_{t=1}^{T} c_t u_t\right] = E(\beta_2) + \sum_{t=1}^{T} c_t E(u_t) = \beta_2 \qquad (56)$$

$$E(\hat{\beta}_1) = E\left[\beta_1 + \frac{1}{T}\sum_{t=1}^{T} u_t - \overline{X}\sum_{t=1}^{T} c_t u_t\right] = E(\beta_1) + \frac{1}{T}\sum_{t=1}^{T} E(u_t) - \overline{X}\sum_{t=1}^{T} c_t E(u_t) = \beta_1 \qquad (57)$$

Por lo tanto, $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores insesgados de los parámetros β_1 y β_2 respectivamente.

Cuando se está trabajando con series reales no se conocen los valores de los parámetros; por ello, no se puede calcular la diferencia entre estimación y parámetro correspondiente a una muestra en concreto. Sin embargo, si el estimador es insesgado sabemos que si estimáramos el modelo con un gran número de muestras, entonces la media de las estimaciones obtenidas estaría muy próxima a los parámetros que se trata de estimar.

Si un estimador no cumple esta propiedad, se dice que es un estimador sesgado. La diferencia entre el valor esperado del estimador y el estimador se denomina sesgo.

EJEMPLO 1 (continuación) Estimación de la función de consumo con series simuladas

Como en un experimento de Montecarlo se conocen los parámetros, se pueden calcular los sesgos que se han cometido en la estimación. Así, los sesgos cometidos en la estimación con la muestra 1, según muestra el cuadro 1, son los siguientes:

Sesgo ordenada: $\beta_1 - \hat{\beta}_1 = 2,000 - 2,993 = -0,993$

Sesgo pendiente: $\beta_2 - \hat{\beta}_2 = 0,850 - 0,803 = 0,047$

Los resultados anteriores están determinados en parte por el azar, es decir, por la extracción concreta de las perturbaciones aleatorias. Ahora bien, si hacemos varias extracciones y obtenemos la media de todas las estimaciones obtenidas, entonces los sesgos serán en general menores que en una muestra en concreto. Así, la media de las 10 estimaciones realizadas, según puede verse en el cuadro 1 son las siguientes:

$$\overline{\hat{\beta}}_1 = \frac{\sum_{j=1}^{10} \hat{\beta}_{1j}}{10} = 1,803$$

$$\bar{\hat{\beta}}_2 = \frac{\sum_{j=1}^{10} \hat{\beta}_{2j}}{10} = 0,858$$

Los sesgos que se obtienen para estos valores medios son los siguientes:

Sesgo ordenada: $\beta_1 - \bar{\hat{\beta}}_1 = 2,000 - 1,803 = 0,197$

Sesgo pendiente: $\beta_2 - \bar{\hat{\beta}}_2 = 0,850 - 0,858 = -0,008$

Examinando los resultados del cuadro 2, en relación a estos sesgos medios, puede observarse que únicamente en la muestra 8 se obtiene un sesgo menor para la pendiente (0,003), mientras que en la muestra 6 el sesgo de la estimación de la ordenada es igual en valor absoluto al correspondiente sesgo medio.

Precisión de los coeficientes

Otra propiedad deseable de un estimador es que sea preciso, es decir, que la función de densidad se encuentre lo más concentrada posible en torno al valor medio. Una medida de esta precisión la suministra la varianza (o la desviación típica) del estimador.

La varianza del estimador $\hat{\beta}_2$ es la siguiente

$$E(\hat{\beta}_2 - \beta_2)^2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum_{t=1}^{T} (X_t - \bar{X})^2} \tag{58}$$

La demostración de (58) puede verse en el recuadro adjunto.

Denominando S_x^2 a la varianza muestral de X, es decir,

$$S_x^2 = \frac{\sum_{t=1}^{T} (X_t - \bar{X})^2}{T} \tag{59}$$

la varianza de $\hat{\beta}_2$ se puede expresar del siguiente modo

$$E(\hat{\beta}_2 - \beta_2)^2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{TS_x^2} \tag{60}$$

Demostración de (58)

De acuerdo con (52) se tiene que

$$\hat{\beta}_2 - \beta_2 = \sum_{t=1}^{T} c_t u_t$$

Elevando al cuadrado ambos miembros de la expresión anterior, y aplicando el operador esperanza se obtiene

$$E(\hat{\beta}_2 - \beta_2)^2 = E\left[\sum_{t=1}^{T} c_t u_t\right]^2$$

$$= E\left[\sum_{t=1}^{T} c_t^2 u_t^2 + \sum\sum_{t\neq t'} c_t c_{t'} u_t u_{t'}\right] = \sum_{t=1}^{T} c_t^2 E(u_t^2) + \sum\sum_{t\neq t'} c_t c_{t'} E(u_t u_{t'})$$

Teniendo en cuenta las hipótesis II b) y II c) se obtiene

$$= \sigma^2 \sum_{t=1}^{T} c_t^2 = \sigma^2 \frac{\sum_{t=1}^{T} (X_t - \bar{X})^2}{\left[\sum_{t=1}^{T} (X_t - \bar{X})^2\right]^2} = \frac{\sigma^2}{\sum_{t=1}^{T} (X_t - \bar{X})^2}$$

De forma análoga se obtiene la varianza del estimador $\hat{\beta}_1$:

$$E(\hat{\beta}_1 - \beta_1)^2 = \sigma^2_{\hat{\beta}_1} = \sigma^2 \left[\frac{1}{T} + \frac{\bar{X}^2}{\sum_{t=1}^{T} (X_t - \bar{X})^2}\right] = \frac{\sigma^2}{T}\left[1 + \frac{\bar{X}^2}{S_x^2}\right] \quad (61)$$

Por otra parte, puede demostrarse que los estimadores mínimo cuadráticos, son estimadores óptimos, es decir, son los que tiene menor varianza dentro de la clase de estimadores lineales e insesgados. Por ello, suele decirse de los estimadores mínimo-cuadráticos que son ELIO (Estimadores Lineales Insesgados y Óptimos).

De acuerdo con (60) y (61) las desviaciones típicas de los estimadores vendrán dadas por

$$\sigma_{\hat{\beta}_2} = \frac{\sigma}{\sqrt{T} S_x} \quad (62)$$

$$\sigma_{\hat{\beta}_1} = \sigma\sqrt{\frac{1}{T}\left[1+\frac{\overline{X}^2}{S_x^2}\right]} \tag{63}$$

Como puede verse en (62), la desviación típica de $\hat{\beta}_2$ es directamente proporcional a la desviación típica de las perturbaciones e inversamente proporcional a la raíz cuadrada del tamaño de la muestra y a la desviación típica muestral de la variable explicativa. En la expresión (63), depende la desviación típica de $\hat{\beta}_1$ depende de esos mismos factores y, además, de la media de la variable explicativa.

Al ser desconocida la varianza de las perturbaciones (σ^2), las varianzas de los estimadores de los coeficientes de regresión son también desconocidas. Por ello es necesario estimarla. El estimador insesgado de la varianza de las perturbaciones en el modelo de regresión lineal simple viene dado por

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T}\hat{u}_t^2}{T-2} \tag{64}$$

A la desviación típica estimada de la perturbación ($\hat{\sigma}$) se le suele conocer también con la denominación de error típico de regresión.

Si en las varianzas teóricas de los estimadores (expresiones (60) y (61)) se sustituye la varianza de las perturbaciones por el estimador (64), se obtienen las varianzas *estimadas* de los estimadores:

$$\hat{\sigma}^2_{\hat{\beta}_2} = \frac{\hat{\sigma}^2}{TS_x^2} \tag{65}$$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{T}\left[1+\frac{\overline{X}^2}{S_x^2}\right] \tag{66}$$

Análogamente, las desviaciones típicas *estimadas* de los estimadores vendrán dadas por

$$\hat{\sigma}_{\hat{\beta}_2} = \frac{\hat{\sigma}}{\sqrt{T}S_x} \tag{67}$$

$$\hat{\sigma}_{\hat{\beta}_1} = \hat{\sigma}\sqrt{\frac{1}{T}\left[1+\frac{\overline{X}^2}{S_x^2}\right]} \tag{68}$$

EJEMPLO (continuación) Estimación de la función de consumo con series simuladas

En el cuadro 1 se recogen también los resultados obtenidos en las 10 muestras del Exp. 1 para $\hat{\sigma}$, $\hat{\sigma}_{\hat{\beta}_1}$ y $\hat{\sigma}_{\hat{\beta}_2}$.

Con objeto de ver la influencia que tienen σ y S_{RENDIS} en las desviaciones de los estimadores, hemos realizado los experimentos 2 y 3.

En el experimento 2 se utiliza el mismo modelo que en el experimento 1 pero la varianza de la perturbaciones utilizada ha sido distinta en cada una de las muestras. En concreto, en las 5 muestras generadas, según puede verse en el cuadro 2, se han asignado a σ de forma sucesiva los valores 1, 2 3, 4 y 5. Como puede observarse, excepto en el caso de que σ=1, los estimadores obtenidos están muy alejados de los valores de los parámetros. Se comprueba también que según va creciendo σ, lo va haciendo también su estimador, aunque, como es previsible, de una forma menos uniforme que el parámetro.

En la figura 5 se ha representado la recta de regresión verdadera y la correspondiente a estimación con la muestra 5, donde σ=5. Como puede comprobarse están muy alejadas entre sí ambas rectas.

En el experimento 3 se utiliza el mismo modelo que en el experimento 1 y también la misma la varianza de la perturbaciones. Sin embargo hemos utilizado 5 muestras distintas de la variable *X*. Las 5 muestras se caracterizan por tener la misma media (21,4), pero una desviación típica muestral de X variable, con valores que oscilan, según puede verse en el cuadro 3, entre 2,905 de la muestra 1 (igual que en el experimento 1) y 0,290 de la muestra 5. Como puede observarse, las desviaciones típicas de los estimadores crecen de forma drástica a medida que disminuye la desviación típica muestral de la variable explicativa.

CUADRO 2 Resultados Exp. 2

Desviación típica de las perturbaciones: Variable

Desviación típica de la muestra de RENDIS: Constante (S_{RENDIS} = 2.905)

Núm muestra	$\hat{\beta}_1$	$\hat{\beta}_2$	σ	**Desviaciones típicas teóricas**		$\hat{\sigma}$	**Desviaciones típicas estimadas**		R^2
				$\sigma_{\hat{\beta}_1}$	$\sigma_{\hat{\beta}_2}$		$\hat{\sigma}_{\hat{\beta}_1}$	$\hat{\sigma}_{\hat{\beta}_2}$	
1	-0.037	0.931	1	2.3509	0.1019	1.4352	3.3738	0.1562	0.816
2	6.749	0.660	2	4.7018	0.2177	1.3755	3.2335	0.1497	0.708
3	-4.152	1.200	3	7.0527	0.3266	2.3284	5.4735	0.2534	0.737
4	-19.640	1.828	4	9.4036	0.4354	4.2124	9.9022	0.4585	0.665
5	8.654	0.674	5	11.7545	0.5443	6.3668	14.9667	0.6930	0.106
Media	**-1.685**	**1.059**							

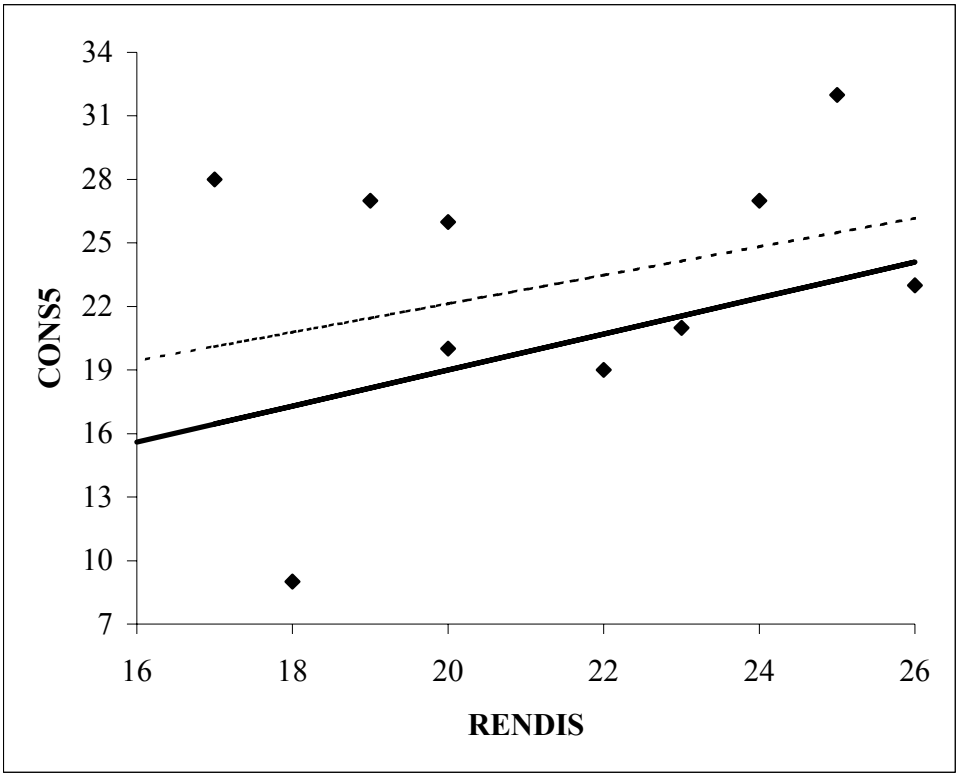

Figura 5. Recta de regresión teórica (trazo continuo) y estimada en la muestra 5 del Exp. 2 (trazo discontinuo)

En la figura 6 se ha representado la recta de regresión verdadera y la correspondiente a estimación con la muestra 5, donde S_{RENDIS} =0,290. Como puede comprobarse en este caso también están muy alejadas entre sí ambas rectas.

CUADRO 3 Renta disponible (RENDIS) utilizada en el Exp. 3

t	RENDIS1	RENDIS2	RENDIS3	RENDIS4	RENDIS5
1	17.00	19.20	20.52	20.85	20.96
2	18.00	19.70	20.72	20.97	21.06
3	19.00	20.20	20.92	21.10	21.16
4	20.00	20.70	21.12	21.22	21.26
5	20.00	20.70	21.12	21.22	21.26
6	22.00	21.70	21.52	21.47	21.46
7	23.00	22.20	21.72	21.60	21.56
8	24.00	22.70	21.92	21.72	21.66
9	25.00	23.20	22.12	21.85	21.76
10	26.00	23.70	22.32	21.97	21.86
Media	**21.400**	**21.400**	**21.400**	**21.400**	**21.400**
Desviación típica	**2.905**	**1.452**	**0.581**	**0.363**	**0.290**

CUADRO 4 Resultados Exp. 3
Desviación típica de las perturbaciones: Constante (σ = 1)
Desviación típica de la muestra de RENDIS: Variable

Núm muestra	$\hat{\beta}_1$	$\hat{\beta}_2$	S_{RENDIS}	**Desviaciones típicas teóricas**		$\hat{\sigma}$	**Desviaciones típicas teóricas**		R^2
				$\sigma_{\hat{\beta}_1}$	$\sigma_{\hat{\beta}_2}$		$\hat{\sigma}_{\hat{\beta}_1}$	$\hat{\sigma}_{\hat{\beta}_2}$	
1	2.213	0.867	2.905	2.3509	0.1019	0.8966	2.1077	0.0996	0.9080
2	4.077	0.768	1.452	4.6714	0.2178	1.2788	5.9714	0.2784	0.4874
3	-1.499	1.014	0.581	11.6519	0.5443	0.8597	10.0170	0.4679	0.3700
4	-25.180	2.141	0.363	18.6453	0.8712	0.9092	16.9454	0.7917	0.4780
5	-19.886	1.849	0.290	23.3376	1.0904	1.1782	27.4482	1.2825	0.2060
Media	**-8.055**	**1.328**							

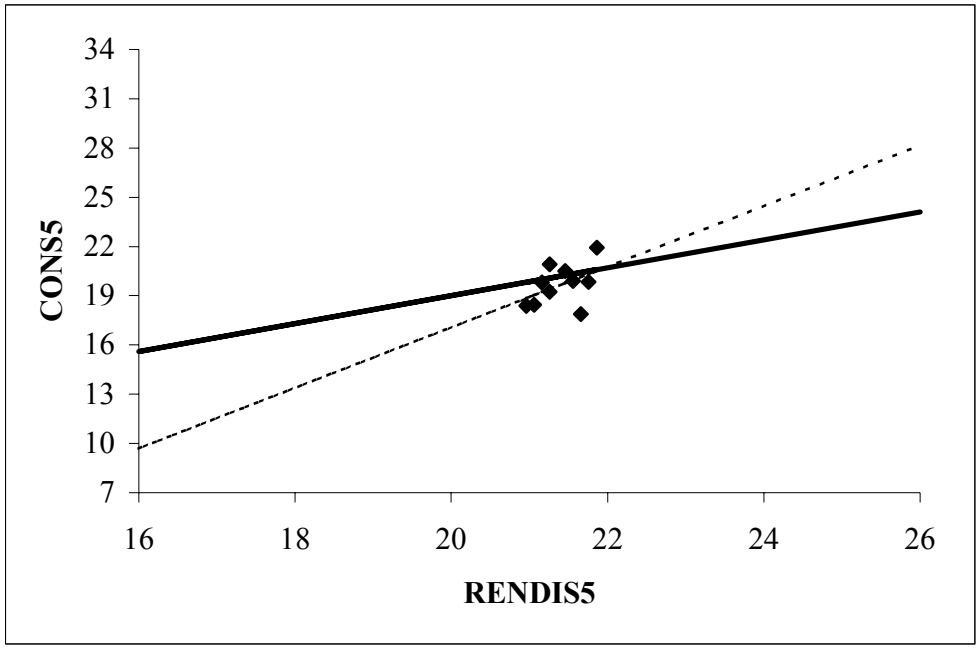

Figura 6. Recta de regresión teórica (trazo continuo) y estimada en la muestra 5 del Exp. 3 (trazo discontinuo)

7 Principios generales del Contraste de hipótesis

El contraste de hipótesis permite realizar inferencias acerca de parámetros poblacionales utilizando datos provenientes de una muestra. Para realizar contrastes de hipótesis en estadística, en general, hay que realizar los siguientes pasos:

1) Establecer una hipótesis nula y una hipótesis alternativa relativas a los parámetros de la población.

2) Construir un estadístico para contrastar las hipótesis formuladas.

3) Definir una regla de decisión para determinar si la hipótesis nula debe ser, o no, rechazada en función del valor que tome el estadístico construido.

Formulación de la hipótesis nula y de la hipótesis alternativa

En la regresión lineal simple vamos a realizar contrates individuales sobre los coeficientes del modelo de regresión. La formulación de la hipótesis nula se realiza mediante una igualdad, que reviste la siguiente forma:

$$H_0 : \beta_i = \beta_i^* \tag{69}$$

donde β_i^* es un valor prefijado por el investigador.

Para formular la hipótesis alternativa se utilizan, según los casos, los operadores "desigualdad", "mayor que" o "menor que". Por tanto, las tres alternativas de hipótesis alternativas que consideraremos son las siguientes:

$$a) H_0 : \beta_i \neq \beta_i^* \qquad b) H_0 : \beta_i > \beta_i^* \qquad c) H_0 : \beta_i < \beta_i^* \tag{70}$$

El caso a) dará lugar a un contraste de 2 colas, mientras que en los casos b) y c) el contraste correspondiente será de una sola cola.

Construcción del estadístico de contraste

Para realizar el contraste se trata de buscar un estadístico que tenga una distribución conocida. La distribución del estadístico dependerá en buena medida de los supuestos que se establezcan en el modelo.

De acuerdo con la hipótesis del modelo de regresión lineal simple II d), la perturbación u_t sigue una distribución normal. Dado que $\hat{\beta}_1$ y $\hat{\beta}_2$ se obtienen como combinación lineal de u_t, seguirán a su vez una distribución normal, es decir,

$$\hat{\beta}_2 \rightarrow N\left[\beta_2, \frac{\sigma}{\sqrt{T}S_x}\right] \tag{71}$$

$$\hat{\beta}_1 \to N\left[\beta_1, \sigma\sqrt{\frac{1}{T}\left[1+\frac{\bar{X}^2}{S_x^2}\right]}\right] \tag{72}$$

o alternativamente, si tipificamos, tendremos que

$$\frac{\hat{\beta}_2 - \beta_2}{\frac{\sigma}{\sqrt{T}S_x}} \to N(0,1) \tag{73}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma\sqrt{\frac{1}{T}\left[1+\frac{\bar{X}^2}{S_x^2}\right]}} \to N(0,1) \tag{74}$$

Supongamos que deseamos realizar un contraste sobre el coeficiente β_2. En concreto, supongamos que deseamos contrastar la siguiente hipótesis nula ((69) frente a la hipótesis alternativa a) de (70)).

Si es cierta la H_0, se verificará que

$$\frac{\hat{\beta}_2 - \beta_2^*}{\frac{\sigma}{\sqrt{T}S_x}} \to N(0,1) \tag{75}$$

El problema que se nos plantea es que no se puede calcular el estadístico anterior porque σ no se conoce cuando trabajamos con datos reales. Cuando se sustituye σ por su estimador $\hat{\sigma}$, entonces el estadístico anterior se distribuye como una *t* con *T-k* grados de libertad, es decir,

$$\frac{\hat{\beta}_2 - \beta_2^*}{\frac{\hat{\sigma}}{\sqrt{T}S_x}} \to t_{T-k} \tag{76}$$

La dispersión de una *t* de Student es mayor que en una *N*(0,1), aunque la dispersión va disminuyendo a medida que aumentan los grados de libertad, verificándose que:

$$t_n \xrightarrow[n\to\infty]{} N(0,1) \tag{77}$$

Así pues, cuando el número de grados de libertad de una *t* de Student tiende hacia infinito converge hacia una distribución *N*(0,1). En el contexto del contraste de hipótesis, si crece el tamaño de la muestra, también lo harán los grados de libertad. Esto implica que para tamaños grandes (por ejemplo, para muestras con un tamaño superior a 60) se puede utilizar, de forma prácticamente

equivalente, la distribución normal para contrastar hipótesis, aún cuando no se conozca la varianza poblacional.

Conviene recordar que una t con n grados de libertad tiene la siguiente relación con una F de 1 grado de libertad en el numerador y n grados de libertad en el denominador:

$$t_n = \pm\sqrt{F_{1,n}} \tag{78}$$

Una variable F toma siempre valores positivos, mientras que una variable t, que tiene una función de densidad simétrica, puede tomar valores positivos y negativos. Obsérvese que a cada valor de una F le corresponden dos valores (uno positivo y otro negativo) en una t. La distribución del estadístico utilizado en el contraste incorpora la H_0, es decir, se construye bajo el cumplimiento de la hipótesis nula.

Regla de decisión para el contraste[3]

[3] Véase página 157 y siguientes de *Econometría Aplicada*