

## PROGRAMA DE ECONOMETRÍA

NOTA: Los libros propuestos como bibliografía básica son:

J. Jonston – J. Dinardo “Métodos de Econometría” Ed. Vicens Vvives. 2001

Alfonso Novales: “Econometría” Ed. McGraw-Hill (2ª Edición. 1993)

Para las clases prácticas se utilizará : Pena et al. “Cien ejercicios de Econometría” Ed. Pirámide 1999

### Tema 1.-El modelo lineal general

- 1.1 Introducción al modelo lineal general. (*JyD pg. 80*)
- 1.2 Especificaciones. (*JyD 3.1 pg 81 e hipótesis 3.4.1*)
- 1.3 Estimadores mínimo cuadrados ordinarios (MCO) (*JyD 3.1.1*)
- 1.4 Propiedades de los estimadores MCO:
  - 1.4.1 Propiedades de los residuos (*JyD 3.1.1 pg.83*)
  - 1.4.2 Descomposición de la Suma de Cuadrados (*JyD 3.1.2*)
  - 1.4.3 Ecuación en forma de desviaciones (*JyD 3.1.3*)
  - 1.4.4 Medidas de la bondad del ajuste: Coeficiente de correlación múltiple (Coeficiente de determinación) y otros criterios (*JyD 3.1.3 pg 85 y 86*)
  - 1.4.5 Tratamiento general de los coeficientes de Correlación Parcial y de Regresión múltiple (*JyD 3.2.3*)
- 1.5 Contraste de normalidad. (*Novales 3.5*)
- 1.6 Estimador de Máxima Verosimilitud (MV)
  - 1.6.1 Estimadores de MV (*JyD 5.1 pg.164*)
  - 1.6.2 Propiedades de los estimadores de MV (*JyD 5.1.1*)
  - 1.6.3 Estimación MV del Modelo Lineal (*JyD 5.2*)
- 1.7 Errores de especificación (*JyD pg.125 y 4.1 pg.126 a 128*)

### Tema 2.-Inferencia en el modelo lineal

- 2.1 Hipótesis (*JyD 3.4.1*)
  - 2.1.1 Media y varianza de **b** (*JyD 3.4.2*)
  - 2.1.2 La estimación de  $\sigma^2$  (*JyD 3.4.3*)
  - 2.1.4 El teorema de Gauss-Markov (*JyD 3.4.4*)
- 2.2 Contraste de hipótesis
  - 2.2.1 Tratamiento general (*JyD 3.4.5 pg. 104-106*)
  - 2.2.2 Contraste acerca de un coeficiente del modelo (*JyD pg 107*)
  - 2.2.3 Contraste de un subconjunto paramétrico. (*JyD pg 109*)
  - 2.2.4 Contraste de significación global del modelo (*JyD pg 108 y 109*)
  - 2.2.5 Intervalos y Regiones de confianza. (*Novales 4.6*)
  - 2.2.6 Contraste de cambio estructural: Test de Chow y de Hansen (*JyD 4.3*)
- 2.3 Estimación recursiva y contrastes basados en estimación recursiva:
  - Contrastes CUSUM y CUSUMSQ (*JyD 4.3.3, 4.3.4 y 4.3.5*)
- 2.4 Contraste más general de cambio estructural (*JyD 4.5*)
- 2.5 Contraste general de Errores de especificación: El contraste RESET de Ramsey (*JyD 4.3.6*)
- 2.6 Estimación bajo restricciones
  - 2.6.1 Regresiones Restringidas y No Restringidas (*JyD 3.4.6*)

- 2.6.2 Ajuste de la Regresión Restringida (*JyD 3.4.7*)
- 2.7 Predicción en el modelo lineal (*JyD 3.5*)

### **Tema 3 Contraste de la Razón de verosimilitud, de Wald y de Multiplicadores de Lagrange**

- 3.1 Contraste de Razón de Verosimilitudes (RV) (*5.3.1*)
- 3.2 El contraste de Wald (W) (*JyD 5.3.2*)
- 3.3 Contraste de Multiplicadores de Lagrange (ML) (*JyD 5.3.3*)
- 3.4 Visión de conjunto de los tres tests (*José M<sup>a</sup> Otero "Econometría. Series temporales y predicción" Ed.AC1993 pg.364*)

### **Tema 4.-Modelo lineal con perturbaciones no esféricas**

- 4.1 Perturbaciones no esféricas (*JyD 5.4 pg. 175*)
- 4.2 Propiedades del estimador MCO en presencia de perturbaciones no esféricas (*JyD 6.1*)
- 4.3 El estimador de mínimos cuadrados generalizado (MCG) (*JyD 5.4.1*)
- 4.4 El estimador MV (*JyD 5.4*)

### **Tema 5.-Heteroscedasticidad**

- 5.1 La heteroscedasticidad y sus posibles causas (*JyD pg.189*)
- 5.2 Propiedades de los estimadores MCO en presencia de heteroscedasticidad. (*JyD 6.1*)
- 5.3 Contraste de heteroscedasticidad (*JyD 6.2*)
  - 5.3.1 Contraste de White (*JyD 6.2.1*)
  - 5.3.2 Contraste de Breusch-Pagan/ Godfrey (*JyD 6.2.2*)
  - 5.3.3 Contraste de Goldfeld-Quandt (*JyD 6.2.3*)
  - 5.3.4 Extensiones del contraste de Goldfeld-Quandt (*JyD 6.2.4*)
  - 5.3.5 Contraste con residuos recursivos (*Johnston "Métodos de Econometría" pg. 472*)
- 5.4 Estimación bajo heteroscedasticidad (*JyD 6.3*)
- 5.5 Transformación Box-Cox (*Novales 6.6*)
- 5.6 Heteroscedasticidad condicional autorregresiva (ARCH) (*JyD 6.9*)

### **Tema 6.-Autocorrelación**

- 6.1 Perturbaciones autocorrelacionadas (*JyD 6.4*)
- 6.2 Naturaleza y causa de autocorrelación (*JyD 6.4.1 y 6.4.2*)
- 6.3 Consecuencias de la autocorrelación (*JyD 6.5*)
- 6.4 Contrastes de autocorrelación (*JyD 6.6*)
- 6.5 Estimación del modelo con perturbaciones autocorrelacionadas (*JyD 6.7*)
- 6.6 Predicción con perturbaciones autocorrelacionadas (*JyD 6.8*)

### **Tema 7.- Modelos dinámicos**

- 7.1 Introducción (*Novales 9.1*)
- 7.2 Justificación teórica de los modelos econométricos dinámicos (*Novales 9.2*)
- 7.3 Modelos de retardos infinitos (*Novales 9.3*)
- 7.4 Estimación con retardos en la variable endógena (*Novales 9.4*)
- 7.5 Contraste de exogeneidad de Wu-Hausman (*Novales 9.4.d*)
- 7.6 El estimador de variables instrumentales (V.I.) (*Novales 9.4.c*)
- 7.7 Eficiencia relativa de los estimadores V.I. (*Novales 9.5*)
- 7.8 Estimación de los modelos con expectativas racionales (*Novales 9.8*)

## **Tema 8.- Multicolinealidad y modelos no lineales**

- 8.1 Multicolinealidad: Concepto y consecuencias (*Novales 10.1*)
- 8.2 Detección de la multicolinealidad (*Novales 10.4*)
- 8.3 Remedios contra la multicolinealidad (*Novales 10.5*)
- 8.4 Observaciones influyentes (*Novales 10.9 pg.366*)
- 8.5 Especificación de modelos no lineales (*Novales 11.1*)
- 8.6 Una aproximación lineal al modelo no lineal (*Novales 11.1*)
- 8.7 Mínimos cuadrados no lineales (*Novales 11.2*)
- 8.8 El estimador de Máxima Verosimilitud (*Novales 11.3*)

## **Tema 9.- Datos de panel**

- 9.1 Descripción del problema (*Novales 15.1*)
- 9.2 El modelo de efectos aleatorios (*Novales 15.2*)
- 9.3 Estimación (*Novales 15.3 y 15.4*)
- 9.4 Contraste de especificación (*Novales 15.5*)
- 9.5 Modelos dinámicos (*Novales 15.6*)
- 9.6 Identificación de efectos individuales en el estimador intragrupos (*Novales 15.7*)

## **Tema 10.-Variables dependientes cualitativas y limitadas**

- 10.1 Tipos de modelos de elección discreta (*JyD 13.1*)
- 10.2 El modelo de probabilidad lineal (*JyD 13.2*)
- 10.3 Formulación de un modelo de probabilidad (*JyD 13.4*)
- 10.4 Modelo Probit (*JyD 13.5*)
- 10.5 Modelo Logit (*JyD 13.6*)
- 10.6 Error de especificación en modelos de variables dependiente binaria (*JyD 13.7*)
- 10.7 Extensiones del modelo básico: Datos agrupados (*JyD 13.8*)
- 10.8 Probit ordenado (*JyD 13.9*)
- 10.9 Modelos Tobit (*JyD 13.10*)

## **Tema 11.- Modelos de ecuaciones simultáneas**

- 11.1 Especificación (*Novales 17.1*)
- 11.2 Forma estructural y forma reducida (*Novales 17.2*)
- 11.3 El problema de la identificación (*Novales 17.3*)
- 11.4 Estimación (*Novales 18.1*)
  - 11.4.1 Estimación por mínimos cuadrados indirectos (MCI) (*Novales 18.2*)
  - 11.4.2 Estimación por variables instrumentales (V.I.) (*Novales 18.3*)
  - 11.4.3 Estimación por mínimos cuadrados bietápicas (MC2E) (*Novales 18.4*)
  - 11.4.4 Estimador de máxima verosimilitud con información limitada (MVIL) (*Novales 18.5*)
  - 11.4.5 Estimación por mínimos cuadrados trietápicas (MC3E) (*Novales 18.6*)
  - 11.4.6 Estimación por máxima verosimilitud con información completa (MVIC) (*Novales 18.7*)
- 11.5 Sistemas recursivos (*Novales 18.8*)
- 11.6 Comparación entre distintos estimadores (*Novales 18.9*)



## **Tema 1.-El modelo lineal general**

- 1.1 Introducción al modelo lineal general. (*JyD pg. 80*)
- 1.2 Especificaciones. (*JyD 3.1 pg 81 e hipótesis 3.4.1*)
- 1.3 Estimadores mínimo cuadrados ordinarios (MCO) (*JyD 3.1.1*)
- 1.4 Propiedades de los estimadores MCO:
  - 1.4.1 Propiedades de los residuos (*JyD 3.1.1 pg.83*)
  - 1.4.2 Descomposición de la Suma de Cuadrados (*JyD 3.1.2*)
  - 1.4.3 Ecuación en forma de desviaciones (*JyD 3.1.3*)
  - 1.4.4 Medidas de la bondad del ajuste: Coeficiente de correlación múltiple (Coeficiente de determinación) y otros criterios (*JyD 3.1.3 pg 85 y 86*)
  - 1.4.5 Tratamiento general de los coeficientes de Correlación Parcial y de Regresión múltiple (*JyD 3.2.3*)
- 1.5 Contraste de normalidad. (*Novales 3.5*)
- 1.6 Estimador de Máxima Verosimilitud (MV)
  - 1.6.1 Estimadores de MV (*JyD 5.1 pg.164*)
  - 1.6.2 Propiedades de los estimadores de MV (*JyD 5.1.1*)
  - 1.6.3 Estimación MV del Modelo Lineal (*JyD 5.2*)
- 1.7 Errores de especificación (*JyD pg.125 y 4.1 pg.126 a 128*)

# La Ecuación Lineal de $k$ Variables

En los capítulos 1 y 2 se han desarrollado las herramientas y procedimientos estadísticos básicos para el análisis de relaciones de dos variables. Sin embargo, es evidente que el marco de dos variables es demasiado restrictivo para aplicarlo al análisis real de fenómenos económicos. Tanto el sentido común como la teoría económica indican la necesidad de especificar y analizar relaciones *multivariantes*. Los modelos económicos postulan generalmente la existencia de varias relaciones conjuntas y simultáneas, cada una de ellas con más de dos variables. Así pues, el objetivo último de la econometría es el análisis de *sistemas de ecuaciones simultáneas*. De momento, sin embargo, restringiremos el análisis a una *única ecuación*, aunque con la posibilidad de incluir  $k$  variables, donde  $k$ , en general, es un número mayor que dos.

Especifiquemos dicha relación como

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + u_t \quad t = 1, \dots, n \quad (3.1)$$

La ecuación identifica  $k - 1$  variables explicativas (regresores):  $X_2, X_3, \dots, X_k$ , que influyen sobre la variable dependiente (regresado). Para simplificar, denominaremos a todas las variables explicativas  $X_{jt}$ , donde el primer subíndice indica la variable en cuestión y el segundo indica la observación en particular de dicha variable. Como en los ejemplos del Capítulo 2, es posible que las  $X$  sean transformaciones de otras variables, aunque la relación sea lineal en los coeficientes  $\beta$ . Suponemos, como en la ecuación (1.21), que las perturbaciones son ruido blanco. Así pues, el modelo posee  $k + 1$  parámetros, las  $\beta$  y la varianza de la perturbación  $\sigma^2$ . La relación multivariante ofrece un espectro de cuestiones de inferencia mucho más rico y amplio que el que aparecía en el modelo de dos variables. La forma más sencilla de abordarlas es mediante la notación matricial, que elimina un gran número de signos de sumatorios, subíndices, etc. En el Apéndice A desarrollamos las nociones básicas del álgebra matricial; las secciones del Apéndice siguen el desarrollo de éste y posteriores capítulos de modo que, en cuanto nos ha sido posible, las distintas operaciones matriciales siguen la secuencia del texto principal.

## 3.1

### FORMULACIÓN MATRICIAL DEL MODELO DE $k$ VARIABLES

Indicaremos las matrices y vectores con negrita (negrita mayúscula para matrices y minúscula para vectores). En general, y siempre que no se indique lo contrario, los vectores serán vectores columna. Por ejemplo,

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2n} \end{bmatrix}$$

son vectores  $n \times 1$ , denominados también  $n$ -vectores, que incluyen las observaciones muestrales de  $Y$  y  $X_2$ . Utilizando terminología vectorial, las  $n$  observaciones muestrales de la ecuación (3.1) se pueden escribir como

$$\begin{bmatrix} \vdots \\ \mathbf{y} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_k \mathbf{x}_k + \mathbf{u} \\ \vdots \end{bmatrix} \quad (3.2)$$

El vector  $\mathbf{y}$  se expresa como una combinación lineal de los vectores  $\mathbf{x}$  y el vector de perturbación  $\mathbf{u}$ . El vector  $\mathbf{x}_1$  es una columna de unos que permite la existencia de un término intersección en la ecuación. Agrupando todos los vectores  $\mathbf{x}$  en una matriz  $\mathbf{X}$  y los coeficientes  $\beta$  en un vector  $\beta$  conseguimos una representación más sencilla si cabe:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \quad (3.3)$$

donde<sup>1</sup>

$$\mathbf{X} = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

#### 3.1.1 El Álgebra de Mínimos Cuadráticos

Reemplazando el vector desconocido  $\beta$  de la ecuación (3.3) por una estimación de  $\mathbf{b}$ , se define un vector de residuos  $\mathbf{e}$ ,

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

El principio de mínimo cuadrados consiste en elegir  $\mathbf{b}$  para minimizar la suma de cuadrados de los residuos,  $\mathbf{e}'\mathbf{e}$ , esto es,

<sup>1</sup> El orden de los subíndices en la matriz  $\mathbf{X}$  no sigue la notación convencional, puesto que nosotros utilizaremos el primer subíndice para señalar la variable y el segundo para indicar la observación. De este modo,  $X_{25}$ , se refiere a la quinta observación de la variable  $X_2$  y se encuentra en la intersección de la quinta fila y la segunda columna de  $\mathbf{X}$ .

$$\text{SCR} = e'e$$

$$= (y - Xb)'(y - Xb)$$

$$= y'y - b'X'y - y'Xb + b'X'Xb$$

$$= y'y - 2b'X'y + b'X'Xb$$

donde este desarrollo utiliza el hecho de que el transpuesto de un escalar es el mismo escalar, por lo que  $y'Xb = (y'Xb)' = b'X'y$ . Como se muestra en el Apéndice A, las condiciones de primer orden son

$$\frac{\partial(\text{SCR})}{\partial b} = -2X'y + 2X'Xb = 0$$

dando las **ecuaciones normales** representadas en

$$(X'X)b = X'y \quad (3.4)$$

Dichas ecuaciones muestran como el vector mínimo cuadrático  $b$  se relaciona con los datos.

**EJEMPLO 3.1. ECUACIONES NORMALES PARA EL CASO DE DOS VARIABLES.** Para ilustrar la ecuación matricial, concretaremos la ecuación (3.4) al caso de dos variables y confirmaremos las ecuaciones normales derivadas en el Capítulo 1. El proceso corresponde a  $k = 2$  y la ecuación se formula como  $Y = \beta_1 + \beta_2 X + u$ . La matriz  $X$  es

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

Por lo tanto,

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

dando

$$\begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$\begin{aligned} nb_1 + b_2 \sum X &= \sum Y \\ b_1 \sum X + b_2 \sum X^2 &= \sum XY \end{aligned}$$

como en la ecuación (1.28).

**EJEMPLO 3.2. ECUACIÓN DE TRES VARIABLES.** De modo similar, puede demostrarse que las ecuaciones normales que ajustan por mínimos cuadrados una ecuación de tres variables son

$$\begin{aligned} nb_1 + b_2 \sum X_2 + b_3 \sum X_3 &= \sum Y \\ b_1 \sum X_2 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3 &= \sum X_2 Y \\ b_1 \sum X_3 + b_2 \sum X_2 X_3 + b_3 \sum X_3^2 &= \sum X_3 Y \end{aligned}$$

Si en la ecuación (3.4) se reemplaza  $y$  por  $Xb + e$ , se obtiene

$$\begin{aligned} (X'X)b &= X'(Xb + e) = (X'X)b + X'e \\ X'e &= 0 \end{aligned} \quad (3.5)$$

Por lo tanto,

que es otro resultado mínimo cuadrático fundamental. El primer elemento de la ecuación (3.5) da  $\sum e_i = 0$ , y en consecuencia,

$$\bar{e} = \bar{Y} - b_1 - b_2 \bar{X}_2 - \dots - b_k \bar{X}_k = 0$$

Los residuos tienen media cero y el plano de regresión pasa a través del punto medio del espacio de  $k$  dimensiones. Los restantes elementos de la ecuación (3.5) son de la forma

$$\sum_i X_{it} e_i = 0 \quad i = 2, \dots, k$$

Como vimos en la nota al pie de página número 16 del Capítulo 1, tal condición significa que todos los regresores tienen correlación muestral cero con los residuos. Este hecho implica, a su vez, que  $y (= Xb)$ , es decir, que el vector de los valores de  $Y$  sobre la regresión, no se halla correlacionado con  $e$  puesto que

$$\hat{y}'e = (Xb)'e = b'X'e = 0$$

### 3.1.2 Descomposición de la Suma de los Cuadrados

Las covarianzas nulas entre regresores y residuos ocultan la descomposición de la suma de los cuadrados. Descompongamos el vector  $y$  entre la parte explicada y la no explicada por la regresión,

$$y = \hat{y} + e = Xb + e$$

de donde

$$y'y = (\hat{y} + e)'(\hat{y} + e) = \hat{y}'\hat{y} + e'e = b'X'Xb + e'e$$

En cualquier caso,  $y'y = \sum_{i=1}^n Y_i^2$  es la suma de los cuadrados de los valores de  $Y$ . El interés suele centrarse en analizar la variación de  $Y$ , medida por la suma de cuadrados de las desviaciones respecto de la media muestral, esto es,

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i Y_i^2 - n\bar{Y}^2$$

Sustrayendo  $n\bar{Y}^2$  en cada lado de la igualdad, obtenemos una nueva descomposición,

$$(y' - n\bar{Y}^2) = (b'X'Xb - n\bar{Y}^2) + e'e \quad (3.6)$$

$$\text{SCT} = \text{SCE} + \text{SCR}$$

donde SCT indica la suma total de cuadrados de  $Y$ , mientras que SCE y SCR indican, respectivamente, la suma de cuadrados explicada y residual (no explicada).

### 3.1.3 Ecuación en Forma de Desviaciones

Un enfoque alternativo consiste en expresar todos los datos como desviaciones de las medias muestrales. La ecuación mínimo cuadrática es

$$Y_t = b_1 + b_2X_{2t} + b_3X_{3t} + \dots + b_kX_{kt} + e_t \quad t = 1, \dots, n$$

Calculando el promedio de las observaciones muestrales, se obtiene

$$\bar{Y} = b_1 + b_2\bar{X}_2 + b_3\bar{X}_3 + \dots + b_k\bar{X}_k$$

que no incluye término  $e$  porque  $\bar{e}$  es cero. Restando la segunda ecuación de la primera obtenemos

$$y_t = b_2x_{2t} + b_3x_{3t} + \dots + b_kx_{kt} + e_t \quad t = 1, \dots, n$$

donde, como en el Capítulo 1, las minúsculas indican las desviaciones de las medias muestrales. Cuando se opera con el formato en desviaciones, el término de intersección  $b_1$  desaparece, aunque puede recuperarse luego haciendo

$$b_1 = \bar{Y} - b_2\bar{X}_2 - \dots - b_k\bar{X}_k$$

Los coeficientes de las pendientes mínimo cuadráticas  $b_2, \dots, b_k$ , son idénticos en ambas formas de la ecuación de regresión, e igual sucede con los residuos.

Reuniendo las  $n$  observaciones, podremos formular la ecuación en desviaciones mediante una matriz de transformación,

$$A = I_n - 1 \left( \frac{1}{n} \right) i i' \quad (3.7)$$

donde  $i$  es un vector columna con  $n$  unos.

Como se muestra en el Apéndice A, se trata de una matriz simétrica e idempotente que, al multiplicarse por un vector de  $n$  observaciones, transforma dicho vector en desviaciones. Por lo tanto,  $Ae = e$  y  $Ai = 0$ . Las ecuaciones mínimo cuadráticas se escriben como

$$y = Xb + e = [i \ X_2] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + e$$

donde  $X_2$  es la matriz  $n \times (k-1)$  de observaciones de los regresores y  $b_2$  es el vector  $k-1$  que incluye los coeficientes  $b_2, b_3, \dots, b_k$ . Multiplicando por  $A$ , se obtiene

$$Ay = \begin{bmatrix} 0 & AX_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + Ae = (AX_2)b_2 + e$$

$$y_* = X_*b_2 + e \quad (3.8)$$

donde  $y_* = Ay$  y  $X_* = AX_2$  proporcionan los datos en forma de desviaciones. Como  $X'e = 0$ , resulta que  $X'_*e = 0$ . Premultiplicando la ecuación (3.8) por  $X'_*$  se obtiene

$$X'_*y_* = (X'_*X_*)b_2$$

que son las ya conocidas ecuaciones normales (como la ecuación (3.4)), excepto que los datos vienen expresados ahora en forma de desviaciones y que el vector  $b_2$  incluye los  $k-1$  coeficientes de las pendientes y excluye el término de intersección. Mediante la ecuación (3.8), expresamos la descomposición de la suma de los cuadrados como

$$y'^1y_* = b'^1X'_*X_*b_2 + e'e \quad (3.9)$$

$$\text{SCT} = \text{SCE} + \text{SCR}$$

El coeficiente de correlación múltiple,  $R$ , se define como la raíz cuadrada positiva de

$$R^2 = \frac{\text{SCE}}{\text{TSS}} = 1 - \frac{\text{SCR}}{\text{TSS}} \quad (3.10)$$

$R^2$  mide la proporción de variación total de  $Y$  explicada por la combinación lineal de los regresores. Casi todos los programas informáticos proporcionan también de manera rutinaria el cálculo de un coeficiente  $\bar{R}^2$  corregido, denominado  $\bar{R}^2$ . Dicho estadístico tiene en cuenta el número de regresores utilizados en la ecuación. En ocasiones es útil comparar el resultado de ajustar varias especificaciones que difieren entre sí debido a la adición o eliminación de variables explicativas. Si se añade una nueva variable al conjunto de regresores, el coeficiente  $R^2$  sin corregir jamás disminuye. La SCE permanece constante cuando la que se añade es una variable totalmente irrelevante. Sin embargo, al añadir variables de bajo nivel explicativo, puede darse el caso de que el coeficiente corregido disminuya. Este coeficiente se define como

$$\bar{R}^2 = 1 - \frac{\text{RSS}/(n-k)}{\text{TSS}/(n-1)} \quad (3.11)$$

Como veremos posteriormente, el numerador y denominador del lado derecho de la ecuación (3.11) son, respectivamente, estimadores insesgados de la varianza de perturbación y de la varianza de  $Y$ . La relación entre los coeficientes corregido y no corregido es

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{n-1}{n-k} (1 - R^2) \\ &= \frac{1-k}{n-k} + \frac{n-1}{n-k} R^2 \end{aligned} \quad (3.12)$$



Existen otros dos criterios que son muy utilizados para comparar el ajuste de varias especificaciones de acuerdo con el distinto número de regresores utilizados. Se trata del **criterio de Schwarz**<sup>2</sup>,

$$CS = \ln \frac{e'e}{n} + \frac{k}{n} \ln n$$

y el **criterio de información de Akaike (CIAK o AIC)**<sup>3</sup>,

$$CIAK = \ln \frac{e'e}{n} + \frac{2k}{n}$$

Habitualmente se buscan especificaciones capaces de reducir la suma de cuadrados de los residuos; sin embargo, todos los criterios llevan implícita una penalización que aumenta con el número de regresores.

**EJEMPLO 3.3.** Un breve ejemplo numérico servirá para ilustrar las fórmulas anteriores. Se utilizan números muy sencillos con objeto de no ocultar la naturaleza de las operaciones con aritmética complicada. Más adelante especificaremos aplicaciones informáticas más realistas. Los datos muestrales son

$$y = \begin{bmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{bmatrix} \quad y \quad X = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 5 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{bmatrix}$$

donde se inserta una columna de unos en la primera columna de X. A partir de estos datos calculamos rápidamente

$$X'X = \begin{bmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{bmatrix} \quad y \quad X'y = \begin{bmatrix} 20 \\ 76 \\ 109 \end{bmatrix}$$

Las ecuaciones normales (3.4) son

$$\begin{bmatrix} 5 & 15 & 25 \\ 15 & 55 & 81 \\ 25 & 81 & 129 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 20 \\ 76 \\ 109 \end{bmatrix}$$

Mediante el procedimiento de resolución de ecuaciones basado en el método de eliminación de Gauss, restamos tres veces la primera fila de la segunda y cinco veces la primera fila de la tercera. Realizadas las operaciones, obtenemos el siguiente sistema

<sup>2</sup> Schwarz, G., "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 1978, 461-464.

<sup>3</sup> Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle", en B. Petrov y F. Csake, eds., *2nd International Symposium on Information Theory*, Budapest, Akademiai Kiado, 1973.

$$\begin{bmatrix} 5 & 15 & 25 \\ 0 & 10 & 6 \\ 0 & 6 & 4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 20 \\ 16 \\ 9 \end{bmatrix}$$

A continuación, restamos seis-décimos de la segunda fila de la tercera, obteniendo

$$\begin{bmatrix} 5 & 15 & 25 \\ 0 & 10 & 6 \\ 0 & 6 & 0,4 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 20 \\ 16 \\ -0,6 \end{bmatrix}$$

La tercera ecuación da como resultado  $0,4b_3 = -0,6$ , es decir

$$b_3 = -1,5$$

Sustituyendo  $b_3$  en la segunda ecuación, descubrimos que

$$10b_2 + 6b_3 = 16$$

$$b_2 = 2,5$$

Finalmente, la primera ecuación

$$5b_1 + 15b_2 + 25b_3 = 20$$

da<sup>4</sup>

$$b_1 = 4$$

La ecuación de regresión será pues

$$\hat{Y} = 4 + 2,5X_2 - 1,5X_3$$

De modo alternativo, si los datos se transforman al formato en desviaciones resulta

$$y_* = Ay = \begin{bmatrix} -1 \\ -3 \\ 4 \\ -1 \\ 1 \end{bmatrix} \quad y \quad X_* = AX_2 = \begin{bmatrix} 0 & 0 \\ -2 & -1 \\ 2 & 1 \\ -1 & -1 \\ 1 & 1 \end{bmatrix}$$

Las ecuaciones normales son entonces

$$\begin{bmatrix} 10 & 6 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 16 \\ 9 \end{bmatrix}$$

Se trata de la segunda y tercera ecuación obtenidas mediante la eliminación de Gauss antes descrita<sup>5</sup>. Por lo tanto, las soluciones para  $b_2$  y  $b_3$  coinciden con las obtenidas anteriormente. Del mismo modo,  $b_1$  será también igual porque la ecuación final de la sustitución anterior es

$$b_1 = Y - b_2X_2 - b_3X_3$$

Partiendo del vector  $y_*$  puede comprobarse que SCT es igual a 28. SCE se calcula a partir de

<sup>4</sup> Los datos muestrales conducen a una única solución para el vector b. La condición requerida para que aparezca una única solución se examina más adelante en este mismo capítulo.

<sup>5</sup> Véase Problema 3.1.

cidad del precio es  $\beta_2$ , mientras que  $\beta_3$  indica la tasa de cambio de la cantidad demandada por unidad de tiempo. Tomando logaritmos, obtenemos

$$\ln Q = \beta_1 + \beta_2 \ln P + \beta_3 T$$

La elasticidad del precio se estima ajustando directamente esta regresión múltiple, o bien excluyendo la tendencia lineal tanto en  $\ln Q$  como en  $\ln P$  y estimando la pendiente de regresión del primer residuo sobre el segundo. Destacaremos, sin embargo, que en el caso de existir tendencias separadas, ninguno de los anteriores coeficientes de la variable tiempo es un estimador del parámetro de cambio  $\beta_3$ <sup>11</sup>.

### 3.2.3 Tratamiento General de los Coeficientes de Correlación Parcial y de Regresión Múltiple

En las condiciones que mostramos en la sección siguiente, las ecuaciones normales se resolverán para  $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ . En este caso, expresaremos los residuos de la regresión MC como

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y} \quad (3.17)$$

donde

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$$

$\mathbf{M}$  es una matriz simétrica e idempotente que posee, además, las propiedades de que  $\mathbf{M}\mathbf{X} = \mathbf{0}$  y  $\mathbf{M}\mathbf{e} = \mathbf{e}$ . Formulamos ahora la regresión general en forma particionada como

$$\mathbf{y} = \begin{bmatrix} x_2 & X_* \end{bmatrix} \begin{bmatrix} b_2 \\ \mathbf{b}_{(2)} \end{bmatrix} + \mathbf{e}$$

$x_2$  es el vector de  $n \times 1$  observaciones de  $X_2$ , con coeficiente  $b_2$ , mientras que  $X_*$  es la matriz  $n \times (k-1)$  de las variables restantes (incluyendo la columna de unos) con vector coeficiente  $\mathbf{b}_{(2)}$ <sup>12</sup>. Las ecuaciones normales son

$$\begin{bmatrix} x_2'x_2 & x_2'X_* \\ X_*'x_2 & X_*'X_* \end{bmatrix} \begin{bmatrix} b_2 \\ \mathbf{b}_{(2)} \end{bmatrix} = \begin{bmatrix} x_2'y \\ X_*'y \end{bmatrix}$$

Resolvemos este sistema de ecuaciones para  $b_2$  e interpretamos los resultados en términos de la pendiente de una regresión simple. La solución es<sup>13</sup>

$$b_2 = (x_2'M_*x_2)^{-1}(x_2'M_*y) \quad (3.18)$$

$$\mathbf{M}_* = \mathbf{I} - X_*(X_*'X_*)^{-1}X_*'$$

donde

<sup>11</sup>Véase Problema 3.4.

<sup>12</sup>Nótese que este es un uso diferente del símbolo asterisco que en una sección anterior fue utilizado para indicar que los datos están en desviaciones.

<sup>13</sup>Véase Apéndice 3.2.

$\mathbf{M}_e$  es una matriz simétrica e idempotente con las propiedades de  $\mathbf{M}_*X_* = \mathbf{0}$  y  $\mathbf{M}_*e = e$ . Siguiendo la ecuación (3.17), sabemos que

$\mathbf{M}_*y$  es el vector de residuos cuando regresamos  $y$  sobre  $X_*$

y  $\mathbf{M}_*x_2$  es el vector de residuos cuando regresamos  $x_2$  sobre  $X_*$ .

Al realizar la regresión del primer vector sobre el segundo se obtiene un coeficiente de la pendiente que, haciendo uso de la simetría e idempotencia de  $\mathbf{M}_*$ , nos da el coeficiente  $b_2$  definido anteriormente en la ecuación (3.18). Este resultado general ha sido ilustrado antes para el caso de tres variables.

Idéntico resultado se obtiene a continuación de un modo más sencillo y elegante. Escribimos la regresión particionada como sigue

$$\mathbf{y} = x_2b_2 + X_*\mathbf{b}_{(2)} + \mathbf{e}$$

Premultiplicando por  $\mathbf{M}_*$ , obtenemos

$$\mathbf{M}_*y = (\mathbf{M}_*x_2)b_2 + \mathbf{e}$$

Finalmente, premultiplicando por  $x_2'$ , tenemos

$$x_2'\mathbf{M}_*y = (x_2'\mathbf{M}_*x_2)b_2$$

que equivale a la ecuación (3.18).

El coeficiente de correlación parcial entre  $Y$  y  $X_2$ , condicional en  $X_3, \dots, X_k$ , se define como

$$r_{12.3\dots k} = \text{coeficiente de correlación entre } (\mathbf{M}_*y) \text{ y } (\mathbf{M}_*x_2) \\ = \frac{x_2'\mathbf{M}_*y}{\sqrt{x_2'\mathbf{M}_*x_2} \sqrt{y'\mathbf{M}_*y}} \quad (3.19)$$

Comparando la ecuación (3.19) con la (3.18), tenemos

$$b_2 = r_{12.3\dots k} \frac{\sqrt{y'\mathbf{M}_*y}}{\sqrt{x_2'\mathbf{M}_*x_2}} \quad (3.20) \\ = r_{12.3\dots k} \frac{s_{1.3\dots k}}{s_{2.3\dots k}}$$

donde  $s_{1.3\dots k}$  es la desviación estándar de los residuos de la regresión de  $Y$  sobre una constante y  $X_3, \dots, X_k$ ; mientras que  $s_{2.3\dots k}$  es la desviación estándar de los residuos de la regresión de  $X_2$  sobre las mismas variables. La ecuación (3.20) es la versión multivariante de lo que la ecuación (1.30) era para el modelo de dos variables. Los restantes coeficientes de correlación parcial y coeficientes de regresión múltiple se obtendrán sustituyendo  $x_2$  por  $x_i$  ( $i = 3, \dots, k$ ) en las ecuaciones (3.19) y (3.20) y realizando los cambios correspondientes en  $\mathbf{M}_*$ .

## **Tema 1.5. Contraste de normalidad (Novales 3.5)**

### 3.5. CONTRASTE DE NORMALIDAD

Puesto que algunas de las propiedades del estimador MCO que hemos visto dependen del supuesto de Normalidad de término de error del modelo, tiene interés disponer de un contraste de tal hipótesis. El contraste más habitual, propuesto por Bera y Jarque (1981), se basa en los coeficientes de asimetría

---

<sup>(3)</sup> El significado de los estadísticos denotados por DW y Q será analizado en capítulos posteriores.

TABLA 3.3

Año	Stock de capital	Empleo	PIB (pesetas constantes) de 1980
64: 1	11.302,5	11.703,0	7.527,4
65: 1	12.578,4	11.990,9	8.004,2
66: 1	14.024,7	12.123,4	8.568,9
67: 1	15.471,3	12.198,2	8.939,1
68: 1	17.036,7	12.256,5	9.544,7
69: 1	18.814,3	12.333,6	10.398,0
70: 1	20.536,9	12.330,7	10.822,3
71: 1	22.061,6	12.427,1	11.318,0
72: 1	23.888,2	12.650,4	12.227,1
73: 1	26.010,6	12.875,4	13.166,9
74: 1	28.214,8	13.041,7	13.866,5
75: 1	30.094,6	12.823,0	13.940,9
76: 1	31.819,3	12.587,4	14.397,2
77: 1	33.396,0	12.581,8	14.829,2
78: 1	34.774,5	12.373,4	15.044,0
79: 1	35.909,7	12.173,9	15.023,1
80: 1	36.991,9	11.811,5	15.209,1
81: 1	37.892,5	11.471,8	15.171,3
82: 1	38.755,6	11.358,3	15.355,9
83: 1	39.488,6	11.279,3	15.633,2
84: 1	39.966,1	10.954,9	15.914,5
85: 1	40.607,3	10.854,9	16.282,8
86: 1	41.504,4	11.111,2	16.816,4
87: 1	42.851,5	11.452,2	17.748,7
88: 1	44.656,5	11.780,6	18.676,6

y curtosis que introdujimos en el Capítulo 2, aplicados a los residuos del modelo. Estos autores probaron que la distribución del estadístico

$$BJ = T \cdot \left( \frac{(\text{asimetría})^2}{6} + \frac{(\text{curtosis} - 3)^2}{24} \right)$$

es, para muestras grandes, una chi-cuadrado con 2 grados de libertad.

La distribución Normal, por ser simétrica, tiene coeficiente de asimetría igual a cero; además, su coeficiente de curtosis es igual a 3, por lo que el valor del estadístico BJ sería cero. Si el término de error del modelo es Normal, entonces, en la medida en que los residuos pueden interpretarse como una *estimación del término de error*, tendrán un valor pequeño del estadístico BJ.

Si el valor del estadístico BJ calculado para los residuos de un modelo estimado exceden del valor de las tablas de una distribución  $\chi^2_2$ , debe rechazarse la hipótesis nula de Normalidad, manteniéndose si el valor del estadístico BJ es inferior al de las tablas.

# Máxima Verosimilitud (MV), Mínimos Cuadrados Generalizados (MCG) y Estimadores de Variable Instrumental (VI)

ca encontrar un valor específico, por ejemplo  $\hat{\theta}$ , que maximice la probabilidad de obtener los valores muestrales ya observados. Se dice entonces que  $\hat{\theta}$  es el estimador de máxima verosimilitud (EMV) del vector de parámetros desconocidos.

Lo más sencillo, en casi todas las aplicaciones, es maximizar el logaritmo de la función de verosimilitud. Definiremos el logaritmo de la verosimilitud como

$$l = \ln L$$

Entonces

$$\frac{\partial l}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}$$

y el  $\hat{\theta}$  que maximice  $l$ , significa que también maximizará a  $L$ . La derivada de  $l$  respecto a  $\theta$  se conoce como el **gradiente (score)**,  $s(\theta; y)$ . El EMV,  $\hat{\theta}$ , se obtiene igualando dicho gradiente a cero, esto es, hallando el valor de  $\theta$  que resuelva la ecuación

$$s(\theta; y) = \frac{\partial l}{\partial \theta} = 0 \quad (5.2)$$

La frecuente utilización de estimadores de máxima verosimilitud se debe principalmente a la variedad de propiedades deseables que poseen y que resumimos en la siguiente sección.

## 5.1.1 Propiedades de los Estimadores de Máxima Verosimilitud

Las propiedades principales de los EMV son las *asintóticas* o de *muestras grandes*. Se deducen bajo condiciones bastante generales.

### 1. Consistencia

$$\text{plim } (\hat{\theta}) = \theta$$

### 2. Normalidad asintótica

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta))$$

Esta expresión indica que la distribución asintótica de  $\hat{\theta}$  es normal, con media  $\theta$  y una varianza obtenida a partir de la inversa de  $I(\theta)$ .  $I(\theta)$  es la **matriz de información** definida de dos maneras equivalentes

$$I(\theta) = E \left[ \left( \frac{\partial l}{\partial \theta} \right) \left( \frac{\partial l}{\partial \theta} \right)' \right] = -E \left[ \frac{\partial^2 l}{\partial \theta \partial \theta'} \right] \quad (5.3)$$

En la práctica, la segunda fórmula es más utilizada puesto que resulta mucho más sencilla de obtener. Siendo  $\theta$  un vector de  $k$  elementos,  $\partial/\partial\theta$  indica un vector columna de  $k$  derivadas parciales, esto es,

En el Capítulo 2 introducíamos el principio de la máxima verosimilitud. Ha llegado el momento de tratarlo con mayor profundidad.

## 5.1 ESTIMADORES DE MÁXIMA VEROSIMILITUD

En los últimos años ha tenido lugar un rápido desarrollo de nuevos contrastes económicos basados en los enfoques de Wald y de los multiplicadores de Lagrange. Del mismo modo, ha resurgido el interés por el enfoque de la máxima verosimilitud.

Supongamos que  $y' = [y_1 \ y_2 \ \dots \ y_n]$  es un vector de  $n$  valores muestrales que depende de un vector de  $k$  parámetros desconocidos,  $\theta' = [\theta_1 \ \theta_2 \ \dots \ \theta_k]$ . Formulemos la densidad conjunta como  $f(y; \theta)$ , que indica su dependencia de  $\theta$ . Dicha densidad es susceptible de doble interpretación. Para un  $\theta$  dado, indica la probabilidad de un conjunto de resultados muestrales. Por otro lado, puede también interpretarse como una función de  $\theta$  condicionada por un conjunto de resultados muestrales. Esta última interpretación es la denominada función de *verosimilitud*. La definición formal es

$$\text{Función de verosimilitud} = L(\theta; y) = f(y; \theta) \quad (5.1)$$

El orden de los símbolos de la función suele invertirse para destacar el nuevo punto de interés de la función. Maximizar la función de verosimilitud respecto a  $\theta$ , signifi-

$$\frac{\partial l}{\partial \theta} = \begin{bmatrix} \frac{\partial l}{\partial \theta_1} \\ \frac{\partial l}{\partial \theta_2} \\ \vdots \\ \frac{\partial l}{\partial \theta_k} \end{bmatrix}$$

Cada elemento de este vector de scores (o gradiente) es, en sí mismo, una función de  $\theta$ , por lo que es necesario diferenciarlos parcialmente respecto a cada uno de los elementos de  $\theta$ . Por ejemplo,

$$\frac{\partial [\partial l / \partial \theta]}{\partial \theta_1} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1^2} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \end{bmatrix}$$

donde las derivadas de segundo orden aparecen formuladas como un vector fila. Siguiendo el desarrollo, obtenemos la matriz de derivadas de segundo orden, que es cuadrada y simétrica, y que recibe el nombre de matriz hessiana,

$$\frac{\partial^2 l}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_k \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_k^2} \end{bmatrix}$$

Nótese que no se trata de la misma matriz que  $(\partial l / \partial \theta)(\partial l / \partial \theta)'$  (matriz también cuadrada y simétrica de  $k \times k$ ), en la que su elemento  $i, j$ -ésimo es el producto  $(\partial l / \partial \theta_i)(\partial l / \partial \theta_j)$ .

3. **Eficiencia asintótica.** Siendo  $\hat{\theta}$  el estimador de máxima verosimilitud de un parámetro único  $\theta$ , la propiedad anterior significa que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$$

para una constante finita  $\sigma^2$ . Si denominamos con  $\hat{\theta}$  cualquier otro estimador consistente y asintóticamente normal de  $\theta$ , entonces  $\sqrt{n}\hat{\theta}$  posee una distribución normal límite cuya varianza es mayor o igual que  $\sigma^2$ . El EMV tiene la mínima varianza entre toda la clase de estimadores consistentes y asintóticamente norma-

les. El término **varianza asintótica** se refiere a la varianza de una distribución límite. Por lo tanto, la varianza asintótica de  $\sqrt{n}\hat{\theta}$  es  $\sigma^2$ . Sin embargo, también se utiliza este mismo término para describir la varianza de una aproximación asintótica a la distribución muestral finita desconocida.

Así pues, tenemos la afirmación equivalente de que la varianza asintótica de  $\hat{\theta}$  es  $\sigma^2/n$ . Si  $\theta$  es un vector de parámetros y  $\hat{\theta}$  es el EMV, entonces

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$$

para una matriz positiva definida  $V$ . Si  $\hat{V}$  indica la matriz de varianzas de otro estimador cualquiera, consistente y asintóticamente normal,  $\hat{V} - V$  es una matriz positiva semidefinida.

4. **Invarianza.** Si  $\hat{\theta}$  es el EMV de  $\theta$  y  $g(\theta)$  es una función continua de  $\theta$ ,  $g(\hat{\theta})$  es el EMV de  $g(\theta)$ .

5. **El gradiente tiene media cero y varianza  $I(\theta)$ .** Para demostrar que la media es igual a cero, deberemos observar que la integración de la función de densidad conjunta sobre todos los valores posibles de  $y$  es igual a uno, esto es,

$$\int \cdots \int f(y_1, y_2, \dots, y_n; \theta) dy_1 \cdots dy_n = \int \cdots \int L dy = 1$$

En efecto, diferenciando ambos lados respecto a  $\theta$ , obtenemos

$$\int \cdots \int \frac{\partial L}{\partial \theta} dy = 0$$

Pero

$$E(s) = \int \cdots \int \frac{\partial l}{\partial \theta} L dy$$

$$\int \cdots \int \frac{\partial L}{\partial \theta} dy$$

$$= 0$$

Por lo tanto, la varianza de  $s$  es

$$\text{var}(s) = E(ss') = E \left[ \left( \frac{\partial l}{\partial \theta} \right) \left( \frac{\partial l}{\partial \theta} \right)' \right] = I(\theta)$$

## 5.2 ESTIMACIÓN MV DEL MODELO LINEAL

Esta sección se refiere a la estimación de máxima verosimilitud del modelo lineal (que incluye la mayoría de las aplicaciones econométricas). La ecuación es

$$y = X\beta + u \text{ con } u \sim N(0, \sigma^2 I)$$

La función de densidad normal multivariante de  $u$  es

$$f(u) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(u'u)}$$

La función de densidad multivariante de  $y$ , condicionada por  $X$ , es

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

donde  $|(\partial u / \partial y)|$  es el valor absoluto del determinante formado a partir de la matriz  $n \times n$  de derivadas parciales de los elementos de  $u$  respecto a los elementos de  $y$ <sup>1</sup>. En este caso, la matriz es la matriz identidad. Por lo tanto, la función del logaritmo de la verosimilitud es

$$\begin{aligned} l = \ln f(y | X) &= \ln f(u) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} u'u \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \end{aligned} \quad (5.4)$$

El vector de parámetros desconocidos,  $\theta$ , tiene  $k + 1$  elementos,

$$\theta' = [\beta', \sigma^2]$$

Después de tomar derivadas parciales obtenemos

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= -\frac{1}{\sigma^2} (-X'y + X'X\beta) \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta) \end{aligned}$$

Igualando a cero las derivadas parciales, los EMV son

$$\hat{\beta} = (X'X)^{-1} X'y \quad (5.5)$$

y

$$\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/n \quad (5.6)$$

<sup>1</sup> Véase Apéndice 5.1 sobre el cambio de variables en funciones de densidad.

El EMV  $\hat{\beta}$  es el estimador MCO  $b$  y  $\sigma^2$  es  $e'e/n$ , donde  $e = y - Xb$  es el vector de residuos MCO<sup>2</sup>. La teoría de los mínimos cuadrados establece que  $E(e'e/(n-k)) = \sigma^2$ . Por lo tanto,  $E(\hat{\sigma}^2) = \sigma^2(n-k)/n$ , y de este modo resulta que  $\hat{\sigma}^2$  es un estimador sesgado de  $\sigma^2$ , aunque  $\hat{\beta}$  sea un estimador insesgado de  $\beta$ . Las derivadas de segundo orden son

$$\frac{\partial^2 l}{\partial \beta \partial \beta'} = -\frac{X'X}{\sigma^2} \text{ con } -E\left(\frac{\partial^2 l}{\partial \beta \partial \beta'}\right) = \frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \beta \partial \sigma^2} = -\frac{X'u}{\sigma^4} \text{ con } -E\left(\frac{\partial^2 l}{\partial \beta \partial \sigma^2}\right) = 0$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{u'u}{\sigma^6} \text{ con } -E\left(\frac{\partial^2 l}{\partial (\sigma^2)^2}\right) = \frac{n}{2\sigma^4}$$

ya que  $E(u'u) = n\sigma^2$ .

La matriz de información es

$$I(\theta) = I\left(\begin{matrix} \beta \\ \sigma^2 \end{matrix}\right) = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

y su inversa

$$I^{-1}\left(\begin{matrix} \beta \\ \sigma^2 \end{matrix}\right) = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}$$

Los términos que están fuera de la diagonal de esta matriz son iguales a cero e indican que  $\beta$  y  $\sigma^2$  se distribuyen independientemente.

Sustituyendo los valores EMV de las ecuaciones (5.5) y (5.6) en la función del logaritmo de la verosimilitud, ecuación (5.4), y deshaciendo luego la transformación logarítmica obtenemos el máximo de la función de verosimilitud:

$$\begin{aligned} L(\hat{\beta}, \hat{\sigma}^2) &= (2\pi e)^{-n/2} (\hat{\sigma}^2)^{-n/2} \\ &= \left(\frac{2\pi e}{n}\right)^{-n/2} (e'e)^{-n/2} \\ &= \text{constante} \cdot (e'e)^{-n/2} \end{aligned} \quad (5.7)$$

donde la constante no depende de ninguno de los parámetros del modelo.

<sup>2</sup> No debe confundirse  $e$ , o sus elementos, con la constante matemática  $e = 2.71828$ .



# Contrastes de Errores de Especificación de la Ecuación Lineal de $k$ Variables

## 4.1 ERROR DE ESPECIFICACIÓN

La especificación del modelo lineal se centra en el vector de términos de perturbación  $u$  y en la matriz  $X$ . Recordemos los supuestos del Capítulo 3:

$$y = X\beta + u \quad (4.1)$$

$$u_i \text{ son iid } (0, \sigma^2) \quad i = 1, \dots, n \quad (4.2a)$$

$$0, \quad u_i \text{ son iid } N(0, \sigma^2) \quad i = 1, \dots, n \quad (4.2b)$$

$$E(X_i u_s) = 0 \text{ para todo } i = 1, \dots, k \text{ y } s = 1, \dots, n \quad (4.3)$$

$$X \text{ es una matriz no estocástica de rango pleno igual a } k \quad (4.4)$$

El supuesto (4.2a) postula que se trata de perturbaciones de **ruido blanco** y el (4.2b) que son perturbaciones **gaussianas de ruido blanco**. Bajo el supuesto de regresor fijo, la ecuación (4.3) se obtiene de manera trivial a partir del supuesto de media cero para el término de perturbación.

¿Qué podría funcionar mal? Indicaremos algunas de las posibilidades de partida de dichos supuestos. Se trata ahora, por el momento, de una visión preliminar y reservamos para los capítulos siguientes el desarrollo de otros temas importantes.

### 4.1.1 Posibles Problemas con $u$

1. El supuesto (4.2a) se sostiene, pero el (4.2b), no. Como indicamos anteriormente en el Capítulo 2, no se trata de que este hecho destruya las propiedades ELIO de MCO, sin embargo, los procedimientos de inferencia sólo serán válidos asintóticamente.
2.  $E(uu') = \text{diag}[\sigma_1^2 \dots \sigma_n^2]$ . La matriz de varianzas-covarianzas para  $u$  es diagonal, con distintas varianzas en la diagonal principal y ceros en el resto de la matriz. Por lo tanto, se viola el supuesto de **homoscedasticidad**. Se trata de la forma más sencilla de **heteroscedasticidad**, muy frecuente en aplicaciones con datos de corte transversal, aunque ésta y otras formas más complicadas se pueden encontrar también en aplicaciones con datos de series temporales. En el Capítulo 6 analizaremos tanto los procedimientos para la detección de la heteroscedasticidad como el desarrollo de los procedimientos de inferencia adecuados.
3.  $E(u_i u_{i-s}) \neq 0, (s \neq 0)$ . En este caso suponemos que las perturbaciones se correlacionan dos a dos. En las aplicaciones de series temporales se dan fuertes correlaciones entre perturbaciones adyacentes y, tal vez, correlaciones menores entre perturbaciones más alejadas entre sí. De modo similar, y cuando trabajamos con datos de corte transversal, es posible que ciertas unidades compartan perturbaciones comunes. En el Capítulo 6 discutiremos la verificación del supuesto de **perturbaciones autocorrelacionadas** y los procedimientos de inferencia relevantes.

La técnica de los mínimos cuadrados explicada en el Capítulo 3 es el caballo de batalla de los *económetras* y se utiliza de modo rutinario en el análisis de una gran variedad de conjuntos de datos. Por derivar del conjunto de supuestos más sencillos de la ecuación, suele denominarse método de **Mínimos Cuadrados Ordinarios** (MCO). En base a los supuestos del Capítulo 3, los estimadores mínimo cuadrados poseen las propiedades deseables allí mencionadas y, por ello, pueden emplearse también en un atractivo espectro de procedimientos exactos de inferencia. Sin embargo, nos enfrentamos a una pregunta crucial. *¿Cómo saber si los supuestos que ocultan los MCO son válidos para un conjunto determinado de datos?* ¿Cómo conocer las propiedades del término de perturbación no observable? ¿Cómo saber qué variables incluir en la matriz  $X$  y en qué forma funcional hacerlo? Cuando alguno de los supuestos subyacentes carece de validez, ¿qué sucede con los estimadores MCO? ¿Siguen siendo útiles o resultan confusos? ¿Existen estimadores y procedimientos de inferencia alternativos que resulten más apropiados bajo supuestos alternativos? En este capítulo y en los siguientes responderemos a estas preguntas.

El **error de especificación** aparece cuando alguno de los supuestos está equivocado. Ciertos errores de especificación tienen implicaciones menores; otros, sin embargo, las tienen muy graves. Resulta tremendamente importante estar alertado de posibles errores de especificación y verificar su presencia. En capítulos posteriores veremos cómo muchas veces es necesario utilizar y desarrollar especificaciones y procedimientos de inferencia más complejos que subyacen en la técnica de los MCO.

### 4.1.2 Posibles Problemas con $X$

1. Exclusión de variables relevantes. La teoría económica enseña que el ingreso y los precios afectan conjuntamente a la demanda, por lo tanto, si aislamos el ingreso de la ecuación de la demanda no esperamos obtener un buen estimador para la elasticidad del precio. Sin embargo, y en situaciones más complicadas, no suele ser tan evidente averiguar cuáles son las variables a incorporar en una relación, lo que puede llegar a convertirse en un importante problema de especificación.
2. Inclusión de variables irrelevantes. Este es el caso contrario al presentado en el Problema 1. Ahora, la hipótesis incluye variables que no deberían estar presentes en la ecuación. Este hecho tiene ciertas consecuencias sobre los procedimientos de inferencia aunque, en general, suelen ser menos graves que aquellas relacionadas con la exclusión de variables relevantes.
3. Forma funcional incorrecta. Puede darse el caso de que las variables sean las correctas pero la forma funcional que las relaciona sea incorrecta. A veces, el contexto de modelo lineal es suficiente para manejar el problema. Por ejemplo, una relación  $Y = f(X_2, X_3)$  puede especificarse como

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

o, tal vez, como

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_2 X_2^2 + \gamma_3 X_3^2 + \delta(X_2 X_3) + u$$

La segunda ecuación permite tanto una respuesta cuadrática a los regresores como un efecto de interacción. El efecto de interacción se basa en una nueva variable, el producto de los dos regresores. Por lo tanto, el efecto esperado de un cambio unitario en  $X_2$  será  $\beta_2 + 2\gamma_2 X_2 + \delta X_3$ , dependiendo pues de  $\beta_2$  y de los niveles de  $X_2$  y  $X_3$ . Del mismo modo, el efecto esperado de un cambio unitario en  $X_3$  dependerá tanto del nivel de  $X_2$ , como del de  $X_3$ . Cuando el error de especificación consiste en utilizar la primera ecuación en lugar de la segunda, aquél se corrige fácilmente añadiendo los términos  $X_2^2$ ,  $X_3^2$ , y  $(X_2 X_3)$ . En otros casos, será necesaria una especificación intrínsecamente no lineal, como en algunos de los ejemplos del Capítulo 2.

4. La matriz  $X$  no es de rango pleno. Esto impide la estimación de un único vector  $b$ . A menudo, los regresores se acercan a la dependencia lineal, en cuyo caso podemos calcular el vector de estimadores MCO, aunque es probable que sus elementos tengan errores estándar muy elevados. Se trata del problema de **colinealidad**.

5. Correlaciones distintas de cero entre los regresores y la perturbación. En este caso se viola el supuesto (4.3). Esta situación puede aparecer en distintas formas. Como explicamos en el Capítulo 2, puede presentarse cuando el regresor es un valor retardado de  $Y$ . Dicho valor carecerá de correlación tanto con las perturbaciones actuales como futuras, aunque sí estará correlacionado con perturbaciones pasadas.

Los estimadores MCO serán sesgados para muestras finitas, aunque consistentes y asintóticamente distribuidos siguiendo una ley normal siempre que pueda aplicarse el teorema de Mann-Wald descrito en el Capítulo 2. Otra situación distinta es la que se produce cuando un regresor se correlaciona con la perturbación actual. En dicho caso, los estimadores MCO serán sesgados e inconsistentes. Tal condición tiene lugar cuando el regresor (o regresores) presentan errores de medida o cuando la ecuación considerada forma parte de un sistema de ecuaciones simultáneas. Discutiremos estas situaciones en capítulos posteriores.

6. Variables no estacionarias. Mencionamos brevemente en el Capítulo 2 que la mayoría de los procedimientos de inferencia tradicionales suponen que las variables son estacionarias. Cuando no se da este caso nos enfrentamos a procedimientos de inferencia no estándar y nos introducimos en el reino de las variables integradas, la cointegración, los modelos de corrección del error, etc., que discutiremos más adelante.

### 4.1.3 Posibles Problemas con $\beta$

La ecuación (4.1) asume de forma implícita que  $\beta$  es un vector constante, tanto en el conjunto de observaciones actuales como en otras observaciones muestrales posibles. Existen situaciones en las que es frecuente toparse con coeficientes con cambios estructurales repentinos o, por el contrario, con evoluciones lentas debidas a cambios en el entorno social o ambiental. Sería poco razonable suponer que la elasticidad de la demanda de manzanas fuera la misma en el Jardín del Edén que la que pudiera darse en Los Angeles a finales del siglo XX. Sin embargo, dicha circunstancia no excluiría el desarrollo de una función de demanda que tuviera aplicaciones prácticas útiles en la situación actual.

## 4.2. EVALUACIÓN DEL MODELO Y PRUEBAS DE DIAGNÓSTICO

Durante mucho tiempo, las prácticas econométricas habituales se resumieron en (i) formular un modelo basado en teoría o en anteriores descubrimientos econométricos, (ii) estimar los parámetros del modelo mediante los datos muestrales relevantes disponibles, y (iii) examinar los estimadores resultantes y estadísticos asociados con el fin de juzgar la validez del modelo especificado. Dicho examen solía centrarse en el ajuste global, en la concordancia con los signos de unos coeficientes previamente supuestos, en la significación estadística de los coeficientes y en la comprobación de la autocorrelación de las perturbaciones. Si el modelo cumplía dichos criterios “satisfactoriamente”, la nueva ecuación pasaba a engrosar la literatura de la materia y podría utilizarse para realizar predicciones con datos externos a la escala

## **Tema 2.-Inferencia en el modelo lineal**

### **2.1 Hipótesis (*JyD 3.4.1*)**

2.1.1 Media y varianza de **b** (*JyD 3.4.2*)

2.1.2 La estimación de  $\sigma^2$  (*JyD 3.4.3*)

2.1.4 El teorema de Gauss-Markov (*JyD 3.4.4*)

### **2.2 Contraste de hipótesis**

2.2.1 Tratamiento general (*JyD 3.4.5 pg. 104-106*)

2.2.2 Contraste acerca de un coeficiente del modelo (*JyD pg 107*)

2.2.3 Contraste de un subconjunto paramétrico .(*JyD pg 109*)

2.2.4 Contraste de significación global del modelo (*JyD pg 108 y109*)

2.2.5 Intervalos y Regiones de confianza. (*Novales 4.6*)

2.2.6 Contraste de cambio estructural: Test de Chow y de Hansen (*JyD 4.3*)

### **2.3 Estimación recursiva y contrastes basados en estimación recursiva: Contrastes CUSUM y CUSUMSQ (*JyD 4.3.3, 4.3.4 y 4.3.5*)**

### **2.4 Contraste más general de cambio estructural (*JyD 4.5*)**

### **2.5 Contraste general de Errores de especificación: El contraste RESET de Ramsey (*JyD 4.3.6*)**

### **2.6 Estimación bajo restricciones**

2.6.1 Regresiones Restringidas y No Restringidas (*JyD 3.4.6*)

2.6.2 Ajuste de la Regresión Restringida (*JyD 3.4.7*)

### **2.7 Predicción en el modelo lineal (*JyD 3.5*)**

Cuando los vectores columna son linealmente dependientes,  $\hat{y}$  seguirá expresándose únicamente como la perpendicular trazada desde  $y$  hasta el espacio columna; sin embargo, habrá uno o más vectores  $c$  que satisfagan  $Xc = 0$ . Entonces

$$y = Xb + Xc = X(b + c)$$

que significa que  $\hat{y}$  no puede expresarse como una combinación lineal única de las  $x_i$  y que, por lo tanto, las ecuaciones normales no se resolverán para un único  $b$ .

Resumiendo, los mínimos cuadrados necesitan que  $X$  tenga rango  $k$  con el fin de que  $(X'X)$  sea no singular y las ecuaciones normales se resuelvan para un único  $b$ .

### 3.4 INFERENCIA EN LA ECUACIÓN DE $k$ VARIABLES

A continuación, necesitamos establecer las propiedades estadísticas de los estimadores mínimo cuadráticos y derivar los procedimientos de inferencia adecuados. Dependerán de qué supuestos se realicen al especificar la relación.

#### 3.4.1 Hipótesis

1.  $X$  es no estocástica y tiene un rango columna completo igual a  $k$ .

Las inferencias son condicionales a los valores muestrales de las variables  $X$ : por lo tanto, para muestras repetidas, los elementos de la matriz  $X$  se tratan como fijos. Como se ilustró en la Sección 3.3, para obtener una única determinación del vector  $b$  es necesario que las columnas de  $X$  sean linealmente independientes.

2. Las perturbaciones tiene las siguientes propiedades:

$$E(u) = 0 \quad (3.21)$$

$$\text{var}(u) = E(uu') = \sigma^2 I \quad (3.22)$$

Cuando se aplica a un vector o matriz el operador esperanza matemática  $E$ , éste se aplica a todos y cada uno de los elementos de ese vector o matriz. Por lo tanto, partiendo de la ecuación (3.21), obtenemos

$$E(u) = E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

y de la ecuación (3.22) resulta

$$E(uu') = E \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ u_2 & u_1 & \dots & u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n & u_n & \dots & u_n \end{pmatrix} = \begin{pmatrix} E(u_1^2) & E(u_1u_2) & \dots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \dots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \dots & E(u_n^2) \end{pmatrix}$$

$$= \begin{pmatrix} \text{var}(u_1) & \text{cov}(u_1, u_2) & \dots & \text{cov}(u_1, u_n) \\ \text{cov}(u_2, u_1) & \text{var}(u_2) & \dots & \text{cov}(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(u_n, u_1) & \text{cov}(u_n, u_2) & \dots & \text{var}(u_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 I$$

Esta matriz es la matriz de varianzas-covarianzas del término de perturbación. Las varianzas aparecen en la diagonal principal y las covarianzas en las posiciones que están fuera de dicha diagonal. La matriz la simbolizaremos mediante la abreviatura  $\text{var}(u)$ . A veces se indica también como  $V(u)$  o  $\text{Cov}(u)$ .

La matriz de varianzas incorpora dos importantes supuestos. El primero es que la varianza de la perturbación es constante en todos los puntos de la muestra. Dicha condición se denomina **homoscedasticidad** y la condición opuesta, aquella que supone que las varianzas de la perturbación son distintas en todos los puntos, se denomina **heteroscedasticidad**. El segundo supuesto es que las perturbaciones *no están correlacionadas a pares*. Por ejemplo, en un estudio de corte transversal sobre el comportamiento de los gastos en vivienda, implicaría covarianzas nulas entre las perturbaciones de distintas viviendas. Con datos de series temporales, implica covarianzas nulas entre las perturbaciones que tengan lugar en distintos periodos de tiempo. Cuando falla dicha condición, se dice que las perturbaciones están **autocorrelacionadas** o **serialmente correlacionadas**.

### 3.4.2 Media y Varianza de $b$

Con fines teóricos, resulta más sencillo reformular las ecuaciones normales como

$$b = (X'X)^{-1} X'y$$

Sustituyendo y por su valor se obtiene

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

y a partir de aquí

$$b - \beta = (X'X)^{-1} X'u \quad (3.23)$$

Al tomar esperanzas, el operador esperanza matemática se trasladará a la derecha de los términos no estocásticos, como  $X$ , pero se aplicará en cambio a cualquier variable estocástica. Por lo tanto,

$$E(b - \beta) = (X'X)^{-1} X'E(u) = 0$$

dando

$$E(b) = \beta \quad (3.24)$$

Así pues, *bajo los supuestos establecidos para este modelo*, los coeficientes MC son estimadores insesgados de los parámetros  $\beta$ . La matriz de varianzas-covarianzas de los estimadores MC se establece como sigue. Partiendo de los primeros principios, como hicimos en el desarrollo de la ecuación (3.22), resulta

$$\text{var}(b) = E[(b - \beta)(b - \beta)']$$

Sustituyendo la ecuación (3.23),

$$\begin{aligned} E[(b - \beta)(b - \beta)'] &= E[(X'X)^{-1} X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1} X'E[uu']X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Por lo tanto,

$$\text{var}(b) = \sigma^2(X'X)^{-1} \quad (3.25)$$

Esta expresión es una matriz  $k \times k$  con las varianzas muestrales de  $b_i$  en la diagonal principal y las covarianzas situadas en las posiciones fuera de la diagonal.

**EJEMPLO 3.5. ERRORES ESTÁNDAR EN LA ECUACIÓN DE DOS VARIABLES.** Como se ha mostrado anteriormente en el Ejemplo 3.1,  $X'X$  es (en este caso)

$$X'X = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{D} \begin{bmatrix} \sum X^2 & -\sum X \\ -\sum X & n \end{bmatrix}$$

Por lo tanto,

donde  $D$  es el determinante de  $(X'X)$ ,

$$D = n \sum X^2 - (\sum X)^2 = n \sum x^2$$

Entonces, indicando la intersección MC y la pendiente por  $a$  y  $b$ , respectivamente, tenemos

$$\text{var}(b) = \frac{\sigma^2}{\sum x^2}$$

que confirma la ecuación (1.40). De modo similar,

$$\begin{aligned} \text{var}(a) &= \frac{\sigma^2 \sum X^2}{n \sum x^2} \\ &= \frac{\sigma^2 (\sum X^2 + n \bar{X}^2)}{n \sum x^2} \\ &= \sigma^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum x^2} \right) \end{aligned}$$

que confirma la ecuación (1.42). Las raíces cuadradas de esas varianzas suelen denominarse **errores estándar**. Son las desviaciones estándar de las distribuciones marginales de  $a$  y  $b$ <sup>15</sup>. Finalmente, vemos que

$$\text{cov}(a, b) = -\sigma^2 \frac{\bar{X}}{\sum x^2}$$

que confirma la ecuación (1.43).

**EJEMPLO 3.6. UNA ECUACIÓN DE TRES VARIABLES.** Casi todas las aplicaciones económicas centran más su interés en los coeficientes de las *pendientes* de la regresión que en los términos de intersección. Por eso trabajaremos con los datos en forma de desviaciones. Sirven todavía ecuaciones como la (3.25), convirtiéndose más en un problema bidimensional que tridimensional. Así pues,

$$\text{var}(b) = \begin{bmatrix} \text{var}(b_2) & \text{cov}(b_2, b_3) \\ \text{cov}(b_2, b_3) & \text{var}(b_3) \end{bmatrix} = \sigma^2 \begin{bmatrix} \sum x_2^2 & \sum x_2 x_3 \\ \sum x_2 x_3 & \sum x_3^2 \end{bmatrix}^{-1}$$

Con algunas operaciones algebraicas puede demostrarse que

$$\text{var}(b_2) = \frac{\sigma^2}{\sum x_2^2(1 - r_{23}^2)} \quad \text{y} \quad \text{var}(b_3) = \frac{\sigma^2}{\sum x_3^2(1 - r_{23}^2)}$$

donde  $r_{23}$  es la correlación entre  $X_2$  y  $X_3$ . Si las variables explicativas no están correlacionadas, las varianzas muestrales se reducirán a las de las regresiones simples de  $Y$  sobre  $X_2$  y de  $Y$  sobre  $X_3$ . Sin embargo, cuando la correlación entre las variables explicativas se incrementan también los errores estándar, más allá de aquellos que corresponderían al caso ortogonal. Cuanto más se asemejen las  $X$ , menos preciso será el intento de estimar sus efectos relativos. Dicha situación recibe el nombre de *multicolinealidad* o *colinealidad*. Con una colinealidad perfecta o exacta los errores estándar son infinitos. Colinealidad exacta significa que las columnas de  $X$  son linealmente dependientes y que, por lo tanto, no puede estimarse el vector MC.

### 3.4.3 La Estimación de $\sigma^2$

La matriz de varianzas-covarianzas de la ecuación (3.25) incluye la varianza de la perturbación  $\sigma^2$ , que es desconocida. Resulta razonable utilizar entonces un estimador basado en la suma de cuadrados de los residuos de la regresión ajustada. Partiendo de la ecuación (3.17), tenemos  $\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}(X\boldsymbol{\beta} + \mathbf{u}) = \mathbf{M}\mathbf{u}$ , ya que  $\mathbf{M}\mathbf{X} = \mathbf{0}$ .

Por lo tanto,

$$E(\mathbf{e}'\mathbf{e}) = E(\mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u}) = E(\mathbf{u}'\mathbf{u})$$

<sup>15</sup> Dichas fórmulas no son operativas porque  $\sigma^2$  es desconocida. Cuando se reemplaza por el estimador,  $\hat{\sigma}^2$  (derivado en la Sección 3.4.3), tenemos los errores estándar *estimados*. Por lo tanto, el término "error estándar" se utiliza tanto referido a errores estándar verdaderos como estimados, dependiendo del contexto.

Aprovechando el hecho de que la traza de un escalar es el mismo escalar, resulta

$$\begin{aligned} E(\mathbf{u}'\mathbf{M}\mathbf{u}) &= E[\text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})] \\ &= E[\text{tr}(\mathbf{u}\mathbf{u}'\mathbf{M})] \\ &= \sigma^2 \text{tr}(\mathbf{M}) \\ &= \sigma^2 \text{tr}[\mathbf{I} - \sigma^2 \text{tr}[(X'X)^{-1}X']] \\ &= \sigma^2 \text{tr}[\mathbf{I} - \sigma^2 \text{tr}[(X'X)^{-1}(X'X)]] \\ &= \sigma^2(n - k) \end{aligned}$$

Por lo tanto,

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n - k} \quad (3.26)$$

define un estimador insesgado de  $\sigma^2$ . Su raíz cuadrada es la desviación estándar de los valores  $Y$  sobre el plano de regresión. Este resultado suele denominarse **error estándar del estimador** o **error estándar de la regresión** (EER).

### 3.4.4 El Teorema de Gauss-Markov

Es el teorema fundamental de los mínimos cuadrados. Establece que, de acuerdo con los supuestos establecidos anteriormente, no existe otro estimador lineal insesgado de los coeficientes  $\boldsymbol{\beta}$  que tenga varianzas muestrales menores que los del estimador mínimo cuadrático de la ecuación (3.25). Demostraremos un resultado más general con cualquier combinación lineal de los coeficientes  $\boldsymbol{\beta}$ . Sea  $\mathbf{c}$  un vector arbitrario de  $k$  elementos de constantes conocidas. Definamos una cantidad escalar  $\mu$  como

$$\mu = \mathbf{c}'\boldsymbol{\beta}$$

Eligiendo  $\mathbf{c}' = [0 \quad 1 \quad 0 \dots 0]$ , entonces  $\mu = \beta_2$ . Podemos, por lo tanto, elegir entre los  $\boldsymbol{\beta}$  el elemento que deseemos. Si escogemos

$$\mathbf{c}' = [1 \quad X_{2,n+1} \quad X_{3,n+1} \dots X_{k,n+1}]$$

entonces

$$\mu = E(Y_{n+1})$$

que es el valor esperado de la variable dependiente  $Y$  en el periodo  $n + 1$ , condicionado por los valores  $X$  en ese periodo en concreto. En general,  $\mu$  representa cualquier combinación lineal de los elementos de  $\boldsymbol{\beta}$ . A continuación, consideramos la clase de los estimadores lineales insesgados de  $\mu$ . Para ello definimos un escalar  $m$  que hará las veces de estimador lineal de  $\mu$ , esto es,

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

donde  $\mathbf{a}$  es algún vector columna de  $n$  elementos. La linealidad queda asegurada mediante dicha definición. Para asegurar la inesgadez tenemos

$$\begin{aligned} E(\mathbf{m}) &= \mathbf{a}'\mathbf{X}\boldsymbol{\beta} + \mathbf{a}'E(\mathbf{u}) \\ &= \mathbf{a}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{c}'\boldsymbol{\beta} \end{aligned} \quad \mathbf{a}'\mathbf{X} = \mathbf{c}' \quad (3.27)$$

que se cumple sólo si

El problema estriba en encontrar el  $n$ -vector  $\mathbf{a}$  que, sujeto a las  $k$  condiciones de la ecuación (3.27), minimice la varianza de  $\mathbf{m}$ . La varianza de  $\mathbf{m}$  es

$$\text{var}(\mathbf{m}) = E(\mathbf{a}'\mathbf{u}\mathbf{u}'\mathbf{a}) = \sigma^2\mathbf{a}'\mathbf{a}$$

donde, gracias al hecho de que  $\mathbf{a}'\mathbf{u}$  es un escalar, podrá formularse su cuadrado como el producto de dicho escalar por su traspuesta. El problema estriba ahora en descubrir un  $\mathbf{a}$  que minimice  $\mathbf{a}'\mathbf{a}$ , condicionado a  $\mathbf{X}'\mathbf{a} = \mathbf{c}$ . La solución es<sup>16</sup>

$$\begin{aligned} \mathbf{a} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \\ \mathbf{m} &= \mathbf{a}'\mathbf{y} \\ &= \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{c}'\mathbf{b} \end{aligned}$$

que implica

Dicho resultado significa lo siguiente:

1. Cada uno de los coeficientes MC,  $b_i$ , es el mejor estimador lineal insesgado del correspondiente parámetro poblacional  $\beta_i$ .
2. El mejor estimador lineal insesgado (ELIO) de cualquier combinación lineal de  $\boldsymbol{\beta}$  es esa misma combinación lineal de las  $b_i$ .
3. El ELIO de  $E(Y_j)$  es

$$\hat{Y}_j = b_1 + b_2X_{2s} + b_3X_{3s} + \dots + b_kX_{ks}$$

que es el valor hallado insertando un vector relevante de valores  $X$  en la ecuación de regresión.

### 3.4.5 Comprobación de Hipótesis Lineales de $\boldsymbol{\beta}$

Hemos establecido las propiedades del estimador MC de  $\boldsymbol{\beta}$ . Queda por ilustrar cómo utilizar dicho estimador para verificar diversas hipótesis sobre  $\boldsymbol{\beta}$ . Consideremos los siguientes ejemplos de hipótesis típicas.

- (i)  $H_0: \beta_i = 0$ , es decir, la hipótesis de que el regresor  $X_i$  no influye en  $Y$ . Las verificaciones de este estilo son muy comunes y reciben el nombre de contrastes de

*significación*.

<sup>16</sup>Véase Apéndice I.1.

- (ii)  $H_0: \beta_i = \beta_{i0}$ . En este caso,  $\beta_{i0}$  es un valor especificado. Por ejemplo, en el caso de que  $\beta_i$  indicara una elasticidad precio, se podría verificar que la elasticidad es  $-1$ .
- (iii)  $H_0: \beta_2 + \beta_3 = 1$ . Si las  $\beta$  indicaran las elasticidades del trabajo y capital en una función de producción, esta formulación sirve para establecer la hipótesis de rendimientos constantes de escala.

- (iv)  $H_0: \beta_3 = \beta_4$ , o  $\beta_3 - \beta_4 = 0$ . Esta hipótesis supone que  $X_3$  y  $X_4$  poseen idéntico coeficiente.

- (v)  $H_0:$

$$\begin{bmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Se trata de la hipótesis de que todo el conjunto de regresores no ejerce efecto alguno sobre  $Y$ . Verificar la significación conjunta de la relación. La hipótesis no incluye el término de intersección ya que, al centrarse el interés en la variación de  $Y$  respecto a su media, el nivel de las series suele ser irrelevante.

- (vi)  $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ . En este caso, el vector  $\boldsymbol{\beta}$  se divide en dos subvectores,  $\boldsymbol{\beta}_1$  y  $\boldsymbol{\beta}_2$  que incluyen, respectivamente,  $k_1$  y  $k_2 (= k - k_1)$  elementos. Así se establece la hipótesis de que un subconjunto específico de regresores no influye significativamente en la determinación de  $Y$ .

Los seis ejemplos se ajustan al formato lineal general

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad (3.28)$$

donde  $\mathbf{R}$  es una matriz  $q \times k$  de constantes conocidas, con  $q < k$ , y  $\mathbf{r}$  es un vector  $q$  de constantes conocidas. Cada hipótesis nula determina los elementos relevantes de  $\mathbf{R}$  y  $\mathbf{r}$ . Para los ejemplos anteriores tendremos

- (i)  $\mathbf{R} = [0 \dots 0 \ 1 \ 0 \dots 0]$  con 1 situado en la posición  $i$ -ésima.  
 $\mathbf{r} = 0$   $q = 1$
- (ii)  $\mathbf{R} = [0 \dots 0 \ 1 \ 0 \dots 0]$  con 1 situado en la posición  $i$ -ésima.  
 $\mathbf{r} = \beta_{i0}$   $q = 1$
- (iii)  $\mathbf{R} = [0 \ 1 \ 1 \ 0 \dots 0]$   
 $\mathbf{r} = 1$   $q = 1$
- (iv)  $\mathbf{R} = [0 \ 0 \ 1 \ -1 \ 0 \dots 0]$   
 $\mathbf{r} = 0$   $q = 1$
- (v)  $\mathbf{R} = [0 \ \mathbf{I}_{k-1}]$  donde  $\mathbf{0}$  es un vector de  $k - 1$  ceros.  
 $\mathbf{r} = \mathbf{0}$   $q = k - 1$
- (vi)  $\mathbf{R} = [\mathbf{0}_{k_2 \times k_1} \ \mathbf{I}_{k_2}]$  donde  $\mathbf{0}$  es un vector de  $k - 1$  ceros.  
 $\mathbf{r} = \mathbf{0}$   $q = k_2$

El modo más eficiente de proceder consiste en proporcionar un procedimiento de verificación para la hipótesis lineal general

$$H_0: \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$$

El contraste general deberá ser tal que sirva para cualquier tipo de aplicación. Conociendo el estimador MC, un paso obvio es calcular el vector  $(\mathbf{R}\mathbf{b} - \mathbf{r})$ . Este vector

mide la discrepancia entre la esperanza y la observación. Si, en algún sentido, se trata de un vector "grande", entonces se pone en duda la validez de la hipótesis nula; si, por el contrario, es un vector "pequeño", entonces no se contradice la hipótesis nula. Como en cualquier procedimiento convencional de verificación, la diferencia entre grande y pequeño viene determinada por la distribución muestral relevante bajo la hipótesis nula que, en este caso, es la distribución de  $\mathbf{Rb}$  cuando  $\mathbf{R}\beta = \mathbf{r}$ . A partir de la ecuación (3.24), obtenemos

$$E(\mathbf{Rb}) = \mathbf{R}\beta \quad (3.29)$$

De este modo, de la ecuación (3.25)

$$\begin{aligned} \text{var}(\mathbf{Rb}) &= E[\mathbf{R}(\mathbf{b} - \beta)(\mathbf{b} - \beta)' \mathbf{R}'] \\ &= \mathbf{R} \text{var}(\mathbf{b}) \mathbf{R}' \\ &= \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \end{aligned} \quad (3.30)$$

Conocemos así la media y la varianza del vector  $\mathbf{Rb}$ . Necesitamos un supuesto adicional que determine la forma de la distribución muestral. Siendo  $\mathbf{b}$  función del vector  $\mathbf{u}$ , la distribución muestral de  $\mathbf{Rb}$  vendrá determinada por la distribución de  $\mathbf{u}$ . Las ecuaciones (3.21) y (3.22) representan los supuestos sobre  $\mathbf{u}$  realizados hasta este momento. Supondremos ahora, además, que las  $u_i$  se hallan distribuidas normalmente y combinaremos los tres supuestos en una única afirmación

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (3.31)$$

Puesto que las combinaciones lineales de variables normales se hallan también distribuidas normalmente, se sigue directamente que

$$\mathbf{b} \sim N[\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}] \quad (3.32)$$

$$\mathbf{Rb} \sim N[\mathbf{R}\beta, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'] \quad (3.33)$$

$$\mathbf{R}(\mathbf{b} - \beta) \sim N[\mathbf{0}, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'] \quad (3.34)$$

En caso de que la hipótesis nula  $\mathbf{R}\beta = \mathbf{r}$  sea cierta

$$(\mathbf{Rb} - \mathbf{r}) \sim N[\mathbf{0}, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'] \quad (3.35)$$

La ecuación nos da la distribución muestral de  $\mathbf{Rb}$ ; como veremos en el Apéndice B, derivaremos una variable  $\chi^2$

$$(\mathbf{Rb} - \mathbf{r})' [\sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{r}) \sim \chi^2(q) \quad (3.36)$$

El único problema con el que choca la aplicación práctica de la ecuación (3.36) es la presencia de un  $\sigma^2$  desconocido. Sin embargo, en el Apéndice B se demuestra que

$$\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2(n - k) \quad (3.37)$$

y que dicho estadístico se distribuye independientemente de  $\mathbf{b}$ . Por lo tanto, las ecua-

ciones (3.36) y (3.37) deben combinarse para dar lugar a un estadístico calculable que tiene una distribución  $F$  cuando la hipótesis nula es cierta y que viene dado por

$$\frac{(\mathbf{Rb} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{r})/q}{\mathbf{e}'\mathbf{e}/(n - k)} \sim F(q, n - k) \quad (3.38)$$

El procedimiento de verificación rechazará la hipótesis  $\mathbf{R}\beta = \mathbf{r}$  siempre que el valor calculado de  $F$  exceda de un valor crítico preseleccionado. A continuación veremos si el procedimiento de verificación es viable para las aplicaciones específicas indicadas anteriormente.

En ciertos casos resulta útil reformular la ecuación (3.38) como sigue

$$(\mathbf{Rb} - \mathbf{r})' [\mathbf{s}^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{r})/q \sim F(q, n - k) \quad (3.39)$$

con  $\mathbf{s}^2$  tal como se definió en la ecuación (3.26). Por lo tanto,  $\mathbf{s}^2 (\mathbf{X}'\mathbf{X})^{-1}$  es la matriz estimada de varianzas-covarianzas de  $\mathbf{b}$ . Por su parte,  $c_{ij}$  indicará el elemento  $i, j$ -ésimo de  $(\mathbf{X}'\mathbf{X})^{-1}$ , entonces

$$\mathbf{s}^2 c_{ii} = \text{var}(b_i) \quad \text{y} \quad \mathbf{s}^2 c_{ij} = \text{cov}(b_i, b_j) \quad i, j = 1, 2, \dots, k$$

En todas las aplicaciones, las formas específicas de  $\mathbf{R}$  y  $\mathbf{r}$  se sustituyen en la ecuación (3.38) o (3.39).

(i)  $H_0: \beta_i = 0$ .  $\mathbf{Rb}$  se convierte en  $b_i$  y  $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'$  en  $c_{ii}$ , el elemento  $i$ -ésimo de la diagonal de  $(\mathbf{X}'\mathbf{X})^{-1}$ . La ecuación (3.39) se convierte entonces en

$$F = \frac{b_i^2}{\mathbf{s}^2 c_{ii}} = \frac{b_i^2}{\text{var}(b_i)} \sim F(1, n - k)$$

o, tomando la raíz cuadrada,

$$t = \frac{b_i}{\mathbf{s} \sqrt{c_{ii}}} = \frac{b_i}{\mathbf{s} \cdot \text{e.e.}(b_i)} \sim t(n - k)$$

Por lo tanto, la hipótesis nula que sostiene que  $X_i$  carece de asociación con  $Y$  se verifica dividiendo el  $i$ -ésimo coeficiente estimado por su error estándar estimado, y relacionando dicho ratio con la distribución  $t$ .

(ii)  $H_0: \beta_i = \beta_{i0}$ . Verificamos esta hipótesis de forma similar a la anterior

$$t = \frac{b_i - \beta_{i0}}{\text{e.e.}(b_i)} \sim t(n - k)$$

En vez de contrastar hipótesis específicas acerca de  $\beta_i$ , podemos también calcular un intervalo de confianza del 95% para  $\beta_i$ . Éste se obtiene mediante

$$b_i \pm t_{0.025} \cdot \text{e.e.}(b_i)$$



(iii)  $H_0: \beta_2 + \beta_3 = 1$ .  $Rb$  es la suma de dos coeficientes estimados,  $b_2 + b_3$ . Pre-multiplicando  $(X'X)^{-1}$  por  $R$ , se obtiene un vector fila cuyos elementos son la suma de los elementos correspondientes en la segunda y tercera fila de  $(X'X)^{-1}$ . Este último resultado combinado con  $R'$  da la suma de los elementos segundo y tercero del vector fila, es decir,  $c_{22} + 2c_{23} + c_{33}$ , con  $c_{23} = c_{32}$ . Por lo tanto,

$$\begin{aligned} R_3^2 (X'X)^{-1} R' &= s^2 (c_{22} + 2c_{23} + c_{33}) \\ &= \text{var}(b_2) + 2\text{cov}(b_2, b_3) + \text{var}(b_3) \\ &= \text{var}(b_2 + b_3) \end{aligned}$$

El contraste estadístico es

$$t = \frac{(b_2 + b_3 - 1)}{\sqrt{\text{var}(b_2 + b_3)}} \sim t(n - k)$$

De modo alternativo, podemos calcular un intervalo de confianza del 95% para la suma  $(\beta_2 + \beta_3)$  como

$$(b_2 + b_3) \pm t_{0.025} \sqrt{\text{var}(b_2 + b_3)}$$

(iv)  $H_0: \beta_3 = \beta_4$ . De forma similar a (iii), el contraste estadístico será en este caso

$$t = \frac{b_3 - b_4}{\sqrt{\text{var}(b_3 - b_4)}} \sim t(n - k)$$

(v)  $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$ . Los primeros cuatro ejemplos implicaban una única hipótesis, razón por la cual hemos elegido pruebas equivalentes  $F$  y  $t$ . El presente ejemplo contempla una hipótesis que incluye los  $k - 1$  regresores. Ahora,  $R(X'X)^{-1}R'$  acabará siendo igual a la submatriz cuadrada de orden  $k - 1$  situada en la esquina inferior de  $(X'X)^{-1}$ . Para evaluar dicha submatriz, particionaremos la matriz  $X$  como  $\begin{bmatrix} i' \\ X_2 \end{bmatrix}$ , donde  $X_2$  es la matriz de observaciones de los  $k - 1$  regresores. Entonces

$$X'X = \begin{bmatrix} i' & i' \\ X_2 & X_2 \end{bmatrix} \begin{bmatrix} i \\ X_2 \end{bmatrix} = \begin{bmatrix} n & i'X_2 \\ X_2' i & X_2' X_2 \end{bmatrix}$$

De la fórmula del Apéndice A que proporciona la inversa de una matriz particionada, podemos expresar la submatriz requerida como

$$[X_2'X - X_2' i i' X_2]^{-1} = [X_2' A X_2]^{-1} = [X_2' X_*]^{-1}$$

donde  $A$  es la matriz, definida anteriormente en la ecuación (3.7), que transforma las observaciones en desviaciones, y  $X_* = AX_2 Rb = b_2$ , es el vector de los estimadores MC de los coeficientes de los  $k - 1$  regresores. Sin tener en cuenta el divisor  $q$ , el numerador de la ecuación (3.38) es

$b_2' X_*' X_* b_2$ , que, tal y como se demostró en la ecuación (3.9), es la SCE de la regresión. Por lo tanto, el estadístico  $F$  que verificará la significación conjunta del total de regresores es

$$F = \frac{\text{SCE}/(k - 1)}{\text{SCR}/(n - k)} \sim F(k - 1, n - k) \quad (3.40)$$

Utilizando la ecuación (3.10), expresaremos también el estadístico como

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)} \sim F(k - 1, n - k) \quad (3.41)$$

Lo que cuestiona el contraste es, básicamente, si el cuadrado medio originado por la regresión es significativamente mayor que el cuadrado medio del residuo.

(vi)  $H_0: \beta_2 = 0$ . La hipótesis afirma que un *subconjunto* de coeficientes de regresión es un vector cero, a diferencia del ejemplo anterior, donde *todos* los coeficientes de regresión debían ser cero. Particionemos la ecuación de regresión como sigue

$$y = [X_1 \quad X_2] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + e = X_1 b_1 + X_2 b_2 + e$$

donde  $X_1$  tiene  $k_1$  columnas, incluyendo una columna de unos,  $X_2$  tiene  $k_2$  ( $= k - k_1$ ) columnas, y  $b_1$  y  $b_2$  son los correspondientes subvectores de los coeficientes de regresión. La partición de la matriz  $X$  nos da

$$X'X = \begin{bmatrix} X_1' X_1 & X_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{bmatrix}$$

$R(X'X)^{-1}R'$  se convierte en la matriz cuadrada de orden  $k_2$  situada en la esquina inferior derecha de  $(X'X)^{-1}$ . De modo similar a lo ocurrido en el Ejemplo (v), esta matriz resultará ser  $(X_2' M_1 X_2)^{-1}$  donde  $M_1 = I - X_1(X_1' X_1)^{-1}X_1'$ . Se trata de una matriz simétrica e idempotente de las del tipo que definíamos en la ecuación (3.17). Posee asimismo la propiedad de que  $M_1 X_1 = 0$  y  $M_1 e = e$ . Además,  $M_1 y$  nos da el vector de residuos cuando se hace la regresión de  $y$  sobre  $X_1$ . El numerador de la ecuación (3.38) es

$$b_2'(X_2' M_1 X_2) b_2 / k_2$$

Con el fin de entender lo que mide dicho numerador, premultiplicaremos la ecuación de regresión particionada por  $M_1$  obteniendo

$$M_1 y = M_1 X_2 b_2 + e$$

Elevando ambos lados al cuadrado

$$y' M_1 y = b_2'(X_2' M_1 X_2) b_2 + e'e$$



**Tema 2.2.5. Intervalos y Regiones de confianza. (Novales 4.6)**

tico  $F$  escrito como función de las sumas residuales de estos dos modelos como se muestra en [4.12]. Este resultado acerca de la equivalencia entre las dos formas de llevar a cabo el contraste de hipótesis, que facilita enormemente el trabajo empírico, puede generalizarse al contraste de cualquier hipótesis del tipo  $H_0: \mathbf{R}\beta = \mathbf{r}$ , como demostraremos en la Sección 4.8.

#### 4.6. INTERVALOS Y REGIONES DE CONFIANZA

Proporcionar un estimador puntual de un coeficiente desconocido no es una información de excesiva utilidad a no ser que se acompañe de un intervalo de confianza para dicha estimación. Para un nivel de confianza dado es importante conocer el rango de valores admisibles del coeficiente que se estima y no tan sólo qué valor consideramos como más probable. Distinguiremos el caso en que se estima un coeficiente del caso en que se estiman varios.

##### 4.6.a. Intervalo de confianza para un solo coeficiente

Hemos visto en las secciones previas que el estadístico

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}}$$

sigue una distribución  $t_{T-k}$ , por lo que si se quiere un intervalo de confianza del  $(1 - \alpha)\%$  para el parámetro  $\beta_i$  (es decir, de significación del  $\alpha\%$ ), basta obtener de las tablas de esta distribución el valor  $\lambda_\alpha$  correspondiente, y entonces:

$$1 - \alpha = P \left[ -\lambda_\alpha \leq \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \leq \lambda_\alpha \right] = P[\hat{\beta}_i - \lambda_\alpha \hat{\sigma}_u \sqrt{a_{ii}} \leq \beta_i \leq \hat{\beta}_i + \lambda_\alpha \hat{\sigma}_u \sqrt{a_{ii}}]$$

y puesto que  $\hat{\beta}_i$ ,  $\alpha$ ,  $\lambda_\alpha$ ,  $\hat{\sigma}_u$  y  $a_{ii}$  son todos valores numéricos conocidos, esta expresión nos da el intervalo de confianza que buscamos. Una vez que se ha construido el intervalo, se puede contrastar una hipótesis  $H_0: \beta_i = \beta_i^0$  al nivel de significación  $\alpha$  sin más que ver si  $\beta_i^0$  está dentro del intervalo (en cuyo caso no rechazamos la hipótesis nula), o fuera de él (en cuyo caso rechazamos la hipótesis nula).

Por ejemplo, en el modelo estimado en la Sección 4.3.b, los intervalos del 95 por 100 de confianza para cada coeficiente son: (7,84, 14,16), (-10,54, -3,46), (9,17, 14,83), (0,17, 5,83), por lo que todos ellos son significativamente diferentes de cero.

#### 4.6.b. Regiones de confianza para varios coeficientes

Cuando se busca un rango de valores para varios coeficientes simultáneamente, no se tiene un intervalo, sino una región de confianza. El razonamiento es análogo al del caso anterior, pero en vez de utilizar la distribución  $t_{T-k}$  como hicimos para un solo coeficiente, utilizaremos el estadístico

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T-k}} \quad [4.13]$$

que se distribuye como una variable  $F_{q, T-k}$ .

Como se trata de obtener regiones de confianza acerca de coeficientes y no de restricciones lineales más generales, la matriz  $\mathbf{R}$  es la que selecciona dichos coeficientes de entre todo el vector  $\boldsymbol{\beta}$ , de modo que el estadístico [4.13] se convierte en:

$$\frac{(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)' [(\mathbf{X}'\mathbf{X})_2^{-1}]^{-1} (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)/s}{\hat{\sigma}_u^2}$$

con distribución  $F_{s, T-k}$ , donde  $\boldsymbol{\beta}_2$  denota el subvector  $s \times 1$ ,  $s \leq k$  cuya región de confianza se quiere construir y  $(\mathbf{X}'\mathbf{X})_2^{-1}$  denota la submatriz  $s \times s$  correspondiente, extraída de la matriz  $(\mathbf{X}'\mathbf{X})^{-1}$ . Si el subvector  $\boldsymbol{\beta}_2$  se coloca al final del modelo, entonces ya vimos en la Sección 4.4.d que  $(\mathbf{X}'\mathbf{X})_2^{-1} = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}$ .

En definitiva, la región de confianza se obtendría mediante:

$$1 - \alpha = P[(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)' (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2) (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \leq \lambda_\alpha s \hat{\sigma}_u^2] \quad [4.14]$$

donde  $\lambda_\alpha$  es el valor numérico que aparece en las tablas de la distribución  $F_{s, T-k}$  al nivel de significación del  $\alpha\%$ .

En la obtención práctica de tales regiones es muy útil recordar que  $\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2$  no es sino la submatriz de  $\mathbf{X}'\mathbf{X}$  correspondiente a los coeficientes incluidos en  $\boldsymbol{\beta}_2$ .

**Ejemplo 4.3.** Así, si disponemos de un modelo como el estimado en la Sección 4.3.b, la región de confianza conjunta de los coeficientes  $\beta_2$  y  $\beta_3$  resulta ser:

$$-0,2043(-7 - \beta_2)^2 + 2(0,4731)(-7 - \beta_2)(12 - \beta_3) - 0,0430(12 - \beta_3)^2 \leq 2\lambda_\alpha \sigma_u^2$$

es decir:

$$-95,6835 - 0,2043\beta_2^2 - 0,0430\beta_3^2 - 14,2146\beta_2 + 7,6554\beta_3 + 0,9462\beta_2\beta_3 \leq \lambda_\alpha$$

El miembro izquierdo de la desigualdad que aparece en [4.14] es un polinomio en el que aparecen tan sólo los coeficientes cuya región de

confianza se construye. La dificultad estriba en que el grado del polinomio es igual al número de tales coeficientes, por lo que la utilidad de estas regiones de confianza se restringe al caso de dos coeficientes. Entonces la región de confianza es una elipse, centrada en el estimador MCO. Al igual que ocurre con los intervalos de confianza para un solo coeficiente, las dimensiones de la elipse dependen directamente de las varianzas con que se han estimado los coeficientes: cuanto mayor sea la precisión de estas estimaciones, menores serán las varianzas, y más pequeña será la elipse de confianza del  $(1 - \alpha)\%$ . Por último, la inclinación de la elipse depende de la covarianza entre las estimaciones de los dos coeficientes para los que se ha construido: la elipse asciende de izquierda a derecha si tal covarianza es positiva, descendiendo en caso contrario.

Al igual que en el caso de un solo coeficiente, se puede contrastar la hipótesis conjunta  $H_0: (\beta_1, \beta_2) = (\beta_1^0, \beta_2^0)$  sin más que examinar si el punto  $(\beta_1^0, \beta_2^0)$  está dentro o fuera de la región de confianza construida. Se mantiene la hipótesis nula en el primer caso, rechazándola si se da la segunda situación.

#### 4.7. ESTIMACION BAJO RESTRICCIONES

Un aspecto básico de la inferencia estadística que se lleva a cabo en Economía es que el investigador sólo contrasta hipótesis en cuya validez está dispuesto a creer a priori, de modo que si su contraste no las rechaza, entonces pasa a imponerlas en la representación estructural que está considerando.

Supongamos por un momento que se contrastara la hipótesis de que el dinero no tiene ningún efecto sobre el nivel de precios, hipótesis en la que el investigador no cree, y el resultado de dicho contraste es que la hipótesis nula no se rechaza. El investigador se encontraría así en una situación difícil de explicar. Generalmente, cuando los resultados empíricos están en total desacuerdo con las sugerencias de la Teoría Económica, es más conveniente interpretarlos como una «casualidad estadística» que como un nuevo resultado que revolucionará la Ciencia Económica, por estar en contradicción directa con los resultados hasta entonces admitidos.

En definitiva, las características de *potencia* de un contraste hacen que sea siempre posible mantener una hipótesis nula incluso si es falsa; por ello es que sólo tiene sentido contrastar hipótesis nulas *aceptables* desde el punto de vista conceptual.

De esta discusión se deduce que si la hipótesis nula no se rechaza, entonces sería muy interesante disponer de un procedimiento para estimar de nuevo el modelo, pero esta vez imponiendo ese conjunto de hipótesis que hemos contrastado y no rechazado. La idea de eficiencia está ligada a la utilización óptima de toda la información disponible. Si se cree que los coeficientes del modelo satisfacen ciertas restricciones, entonces se ganaría eficiencia introduciendo dichas restricciones en el proceso de estimación.

El objetivo de esta sección es describir el modo en que este proceso puede llevarse a cabo. Vamos a encontrar el estimador que minimice la suma de cuadrados de los residuos, de igual modo que hicimos en la Sección 3.2, pero

temporal o al rango empírico de la muestra. En caso de que el modelo se clasificara de "insatisfactorio", el investigador seguía intentando hallar aquella reformulación que cumpliera los requisitos necesarios. Dichos procesos de investigación apenas se reflejaban en la literatura puesto que la **minería de datos** estaba mal vista y, por otro lado, porque resultaba prácticamente imposible determinar valores  $P$  y coeficientes de confianza correctos para los estadísticos finales.<sup>1</sup>

En los últimos años la tendencia está cambiando. La nueva corriente de opinión fue iniciada por Denis Sargan, de la London School of Economics, que escribía en 1975:

A pesar de los problemas asociados con la "minería de datos", considero que la especificación sugerida debe comprobarse de todas las formas posibles y que sólo deberán utilizarse aquellas especificaciones que sobrevivan a este proceso de prueba y que correspondan a un modelo económico razonable.<sup>2</sup>

Dicho enfoque ha sido muy desarrollado últimamente, especialmente por David Hendry y sus colegas.<sup>3</sup> El resultado es una auténtica batería de pruebas de diagnóstico que no pueden utilizarse ni de forma automática ni rutinaria, ya que requieren una buena dosis de juicio, intuición económica o sentido común. Algunos de los contrastes resaltan un error o errores de especificación en particular. Otros indican que determinada especificación no funciona bien sin señalar explícitamente un problema preciso. Finalmente, puede ocurrir que sobrevivan a este proceso de prueba u que algunas especificaciones superen un cierto tipo de pruebas estadísticas pero no otras.

#### 4.3 PRUEBAS ACERCA DE LA CONSTANCIA DE LOS PARÁMETROS

Uno de los criterios de adecuación más importantes para una ecuación estimada es que debería resultar relevante para *datos externos a los datos muestrales utilizados en la estimación*. Dicho criterio se conoce como **constancia de los parámetros**, es decir, el vector  $\beta$  debería resultar aplicable tanto fuera como dentro de la muestra. Examinaremos la constancia de parámetros de distintos modos. Uno de los más efectivos será la prueba de la capacidad predictiva del modelo o de precisión de las predicciones.

<sup>1</sup> Véase, por ejemplo, Michael C. Lowell, "Data Mining", *The Review of Economics and Statistics*, LXV, 1983, 1-12.

<sup>2</sup> J.D. Sargan. Discussion on Misspecification en *Modelling the Economy*, ed. G.A. Renton, Heinemann, 1975, mencionado en Adrian R. Pagan. "Model Evaluation by Variable Addition", Capítulo 5, *Econometrics and Quantitative Economics*, eds. David F. Hendry y Kenneth Wallis, Blackwell, 1984, p. 131.

<sup>3</sup> Véase para más detalles cualquier manual reciente de PC-GIVE (David F. Hendry et al, Institute of Economics and Statistics, University of Oxford, UK).

### 4.3.1 El Contraste de Predicción de Chow<sup>4</sup>

Cuando el vector de parámetros es constante, las predicciones externas a la muestra tendrán probabilidades específicas de situarse dentro de los límites calculados a partir de los datos muestrales. Por lo tanto, errores de predicción “grandes” pondrán en duda la hipótesis de constancia, mientras que con los “pequeños” se dará el caso contrario. El procedimiento sugerido no estima la totalidad de observaciones muestrales, sino que divide el conjunto de datos en  $n_1$  observaciones a estimar y en  $n_2 = n - n_1$  observaciones que se utilizarán para el contraste. Cuando se trabaja con series temporales, las primeras  $n_1$  observaciones suelen ser las utilizadas para la estimación, mientras que las últimas  $n_2$  se dedicarán a la comprobación. En aplicaciones de corte transversal, la base de datos se dividirá en dos submuestras de acuerdo con los valores de una variable de *tamaño* como, por ejemplo, el ingreso del hogar o los beneficios de una empresa, ganancias, empleo, etc. No existen normas rápidas ni estrictas que determinen los valores relativos de  $n_1$  y  $n_2$ . Lo normal es reservar para la comprobación el 5, 10 o 15% del total de observaciones disponibles.

El contraste de precisión de las predicciones, conocido como test de Chow en honor a la influencia del artículo publicado por Chow en 1960, consiste en lo siguiente:

1. Estimar el vector de parámetros por MCO con  $n_1$  observaciones, obteniendo

$$\mathbf{b}_1 = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y}_1 \quad (4.5)$$

donde  $\mathbf{X}_i$ ,  $\mathbf{y}_i$  ( $i = 1, 2$ ) indican la partición de los datos en  $n_1$  y  $n_2$  observaciones.

2. Utilizar  $\mathbf{b}_1$  para obtener una predicción del vector  $\mathbf{y}_2$  de la forma

$$\hat{\mathbf{y}}_2 = \mathbf{X}_2 \mathbf{b}_1 \quad (4.6)$$

3. Obtener el vector de errores de predicción y analizar su distribución muestral bajo la hipótesis nula de constancia en el vector de parámetros.

El vector de errores de predicción es

$$\mathbf{d} = \mathbf{y}_2 - \hat{\mathbf{y}}_2 = \mathbf{y}_2 - \mathbf{X}_2 \mathbf{b}_1 \quad (4.7)$$

Si la ecuación  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ , con  $E(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{I}$ , es válida para ambos conjuntos de datos, podremos formular el vector de errores de predicción como

$$\mathbf{d} = \mathbf{y}_2 - \mathbf{X}_2 \mathbf{b}_1 = \mathbf{u}_2 - \mathbf{X}_2(\mathbf{b}_1 - \boldsymbol{\beta})$$

Por lo tanto,  $E(\mathbf{d}) = \mathbf{0}$ , y puede demostrarse<sup>5</sup> que la matriz de varianzas-covarianzas para  $\mathbf{d}$  es

$$\text{var}(\mathbf{d}) = E(\mathbf{d}\mathbf{d}')$$

$$\begin{aligned} &= \sigma^2 \mathbf{I}_{n_2} + \mathbf{X}_2' \cdot \text{var}(\mathbf{b}_1) \cdot \mathbf{X}_2' \\ &= \sigma^2 [\mathbf{I}_{n_2} + \mathbf{X}_2' (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_2'] \end{aligned} \quad (4.8)$$

Suponiendo perturbaciones gaussianas,

$$\mathbf{d} \sim N[\mathbf{0}, \text{var}(\mathbf{d})]$$

y, por lo tanto

$$\mathbf{d}' [\text{var}(\mathbf{d})]^{-1} \mathbf{d} \sim \chi^2(n_2)$$

Además,

$$\mathbf{e}_1' \mathbf{e}_1 / \sigma^2 \sim \chi^2(n_1 - k)$$

donde  $\mathbf{e}_1' \mathbf{e}_1$  es la suma de cuadrados de los residuos de la regresión estimada. Ambos estadísticos  $\chi^2$  se distribuyen independientemente. Por lo tanto, bajo la hipótesis de constancia de parámetros, se tiene que

$$F = \frac{\mathbf{d}' [\mathbf{I}_{n_2} + \mathbf{X}_2' (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_2']^{-1} \mathbf{d} / n_2}{\mathbf{e}_1' \mathbf{e}_1 / (n_1 - k)} \sim F(n_2, n_1 - k) \quad (4.9)$$

Los valores elevados de este estadístico  $F$  rechazarán la hipótesis de que el mismo vector  $\boldsymbol{\beta}$  sirve tanto dentro como fuera de los datos utilizados para la estimación. La derivación de  $\text{var}(\mathbf{d})$  supone que  $\sigma^2$  es idéntica en todos los subconjuntos de datos. Por lo tanto, el contraste  $F$  en la ecuación (4.9) estará condicionado por dicho supuesto. Cuando las perturbaciones sean heteroscedásticas, el estadístico  $F$  puede sobredimensionar el verdadero nivel de significación.

Pagan y Nicholls, siguiendo un artículo anterior de Salkever<sup>6</sup>, ilustran un modo alternativo de desarrollar este contraste de precisión de la predicción. El proceso de cálculo resulta más sencillo. Supongamos que, en el periodo a predecir, permitimos la posible existencia de un vector de coeficientes distinto al de la muestra utilizada para la estimación. El modelo completo se formularía como

$$\mathbf{y}_1 = \mathbf{X}_1 \boldsymbol{\beta} + \mathbf{u}_1$$

$$\mathbf{y}_2 = \mathbf{X}_2 \boldsymbol{\alpha} + \mathbf{u}_2 = \mathbf{X}_2 \boldsymbol{\beta} + \mathbf{X}_2(\boldsymbol{\alpha} - \boldsymbol{\beta}) + \mathbf{u}_2 = \mathbf{X}_2 \boldsymbol{\beta} + \boldsymbol{\gamma} + \mathbf{u}_2 \quad (4.10)$$

donde  $\boldsymbol{\gamma} = \mathbf{X}_2(\boldsymbol{\alpha} - \boldsymbol{\beta})$ . Si,  $\boldsymbol{\gamma} = \mathbf{0}$ , entonces  $\boldsymbol{\alpha} = \boldsymbol{\beta}$ , y el vector de coeficientes es constante en toda el recorrido de los datos utilizados ya sea en los relativos a la estimación como en el periodo de la predicción. En forma resumida, formulamos el modelo como

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{I}_{n_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \quad (4.11)$$

<sup>5</sup> Véase Apéndice 4.1.

<sup>6</sup> Adrian Pagan y D. Nicholls, “Estimating Prediction Errors and Their Standard Deviations Using Constructed Variables”, *Journal of Econometrics*, 24, 1984, 293–310; y D. Salkever, “The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals”, *Journal of Econometrics*, 4, 1976, 393–397.

<sup>4</sup> G.C. Chow, “Tests of Equality between Sets of Coefficients in Two Linear Regressions”, *Econometrica*, 52, 1960, 211–22.



<sup>7</sup> Bruce E. Hansen, "Testing for Parameter Instability in Linear Models", *Journal of Policy Modeling*, **14**, 1992, 517-533.

El ajuste por el método de los MCO ofrece las conocidas condiciones

$$\sum_{i=1}^n x_{it} e_i = 0 \quad i = 1, \dots, k$$

y

$$\sum_{i=1}^n (e_i^2 - \hat{\sigma}^2) = 0 \quad (4.18)$$

Esta última condición permite definir un estimador de la varianza de la perturbación como  $\hat{\sigma}^2 = \sum_{i=1}^n e_i^2/n$ , que se diferencia ligeramente del estimador insesgado  $s^2 = \sum_{i=1}^n e_i^2/(n-k)$ . Definimos ahora

$$f_{it} = \begin{cases} x_{it}e_i & i = 1, \dots, k \\ e_i^2 - \hat{\sigma}^2 & i = k+1 \end{cases}$$

Combinando las ecuaciones (4.17) y (4.18), obtenemos

$$\sum_{i=1}^n f_{it} = 0 \quad i = 1, \dots, k+1 \quad (4.19)$$

El contraste estadístico de Hansen se basa en los *sumatorios acumulados* de  $f_{it}$ , esto es

$$S_{it} = \frac{1}{n} \sum_{j=1}^i f_{ij} \quad (4.20)$$

Hansen desarrolla contrastes para comprobar de forma individual la estabilidad de cada parámetro, así como la estabilidad conjunta de todos los parámetros. Los estadísticos de prueba de parámetros individualmente considerados son

$$L_i = \frac{1}{nV_i} \sum_{i=1}^n S_{it}^2 \quad i = 1, \dots, k+1 \quad (4.21)$$

donde

$$V_i = \sum_{i=1}^n f_{it}^2 \quad (4.22)$$

Para la comprobación de la estabilidad conjunta, tenemos

$$f_t = [f_{1t} \quad \dots \quad f_{k+1,t}]'$$

y

$$s_t = [S_{1t} \quad \dots \quad S_{k+1,t}]'$$

El estadístico de verificación de la estabilidad conjunta será

$$L_c = \frac{1}{n} \sum_{i=1}^n s_t' V^{-1} s_t \quad (4.23)$$

donde

$$V = \sum_{i=1}^n f_i f_i' \quad (4.24)$$

Bajo la hipótesis nula, la suma acumulada tenderá a distribuirse alrededor de cero. Por lo tanto, los valores "grandes" de los estadísticos de prueba sugieren rechazar la hipótesis nula. La teoría de la distribución no es estándar y sólo disponemos de los valores críticos asintóticos. Los encontramos tabulados en la Tabla 7 del Apéndice D. Hay una línea que se refiere al número de parámetros utilizados que toma valores del 1 al 20. La primera línea de la tabla nos da, en consecuencia, los valores críticos del contraste para un coeficiente individualmente considerado. El valor crítico al 5% es 0,407, lo que da lugar a la regla de que en caso de un valor del estadístico de prueba que sea mayor que 0,5 podemos suponer que se trata de un parámetro inestable. De manera parecida, para cinco parámetros incluyendo la varianza, el valor crítico al 5% es aproximadamente 1,5. Algunos programas informáticos económicos incorporan ya el contraste de Hansen.

### 4.3.3 Contrastes Basados en Estimación Recursiva

El modelo de la ecuación (4.1) puede formularse como

$$y_t = x_t' \beta + u_t \quad t = 1, \dots, n \quad (4.25)$$

donde  $y_t$  es la observación que ocupa el lugar  $t$  de la variable dependiente, y  $x_t' = [1 \quad x_{2t} \quad \dots \quad x_{kt}]$  es el  $t$ -ésimo vector fila de regresores en donde hemos utilizado de nuevo las letras de los subíndices para indicar los niveles de las variables. La matriz completa del conjunto de regresores es

$$X = \begin{bmatrix} \dots & x_1' & \dots \\ \dots & x_2' & \dots \\ \vdots & \vdots & \vdots \\ \dots & x_n' & \dots \end{bmatrix}$$

La base de la estimación recursiva es muy sencilla. Ajustamos el modelo a las primeras  $k$  observaciones. El ajuste es perfecto, ya que sólo tenemos  $k$  coeficientes de regresión que estimar. A continuación utilizamos los primeros  $k+1$  datos y calculamos de nuevo el vector de coeficientes. Seguiremos así, añadiendo cada vez un nuevo dato, hasta obtener el último vector de coeficientes, basado en la totalidad  $n$  de datos. El proceso genera una *secuencia* de vectores,  $b_k, b_{k+1}, \dots, b_n$ , donde los subíndices indican el número de datos utilizados en la estimación. En general,

$$b_t = (X_t' X_t)^{-1} X_t' y_t \quad (4.26)$$

donde  $X_t$  es la matriz  $t \times k$  de regresores para los primeros  $t$  datos de la muestra, y  $y_t$  es el vector  $t$  de las primeras  $t$  observaciones de la variable dependiente. Los errores

estándar de los distintos coeficientes se calculan en cada paso del proceso recursivo, exceptuando el primero, ya que la SCR es igual a cero cuando  $t = k$ . Algunos programas informáticos empiezan los cálculos recursivos en cualquier  $m > k$ , generando la secuencia  $b_m, b_{m+1}, \dots, b_n$ . Los gráficos muestran la evolución de todos los coeficientes con el valor de  $m$ , y menos, dos veces el respectivo error estándar. La inspección de los gráficos sugerirá, o no, la constancia de los parámetros. Es posible que a medida que añadimos más datos los gráficos muestren un movimiento vertical que podrá alcanzar un nivel superior a los límites de confianza previamente estimados. Dicho fenómeno suele ser consecuencia del resultado del propio modelo ensayado y sugiere la existencia de un cambio estructural que induce a sospechar la inconstancia de los parámetros. La estimación recursiva, al proporcionar un orden único de los datos, resulta un procedimiento atractivo con las series temporales. El procedimiento, asimismo, es fácilmente aplicable a datos de corte transversal que, en caso necesario, podrán ordenarse según una variable de "tamaño" adecuado.

#### 4.3.4 Errores de Predicción de un Paso en Adelante

Cuando utilizamos todos los datos disponibles e incluimos un periodo  $t - 1$ , la predicción un paso en adelante de  $y_t$  será  $\mathbf{x}'_t \mathbf{b}_{t-1}$ . El error de predicción un paso adelante será pues,

$$v_t = y_t - \mathbf{x}'_t \mathbf{b}_{t-1} \quad (4.27)$$

Si partimos de la ecuación (4.8), la varianza del error de predicción un paso adelante será

$$\text{var}(v_t) = \sigma^2 [1 + \mathbf{x}'_t (\mathbf{X}'_{t-1} \mathbf{X}_{t-1})^{-1} \mathbf{x}_t] \quad (4.28)$$

El valor desconocido de  $\sigma^2$  en la ecuación (4.28) se sustituye por la varianza residual estimada de las primeras observaciones ( $t - 1$ ), dado que  $t - 1 > k$ . La raíz cuadrada nos da la estimación del error estándar de la regresión (E.E.R.). Trazaremos una línea en el valor de dos veces este error estándar recursivo y otra en el valor de menos dos veces dicho error estándar recursivo. Estas líneas se trazan en un mismo gráfico alrededor de la línea del valor cero y alrededor de los errores de predicción actuales (llamados también residuos recursivos). Los residuos situados fuera de las bandas señaladas por los errores estándar suelen sugerir la inconstancia de los parámetros. En cada punto, la probabilidad del error observado, bajo la hipótesis nula, se calcula a partir de la correspondiente distribución  $t$ , tal como ya se hizo en la ecuación (3.48).

#### 4.3.5. Contrastes CUSUM y CUSUMSQ

Los **residuos recursivos reescalados** se definen como

$$w_t = \frac{v_t}{\sqrt{[1 + \mathbf{x}'_t (\mathbf{X}'_{t-1} \mathbf{X}_{t-1})^{-1} \mathbf{x}_t]}} \quad t = k + 1, \dots, n \quad (4.29)$$

Bajo las hipótesis de las ecuaciones (4.1) y (4.2b) resulta ahora que

$$w_t \sim N(0, \sigma^2)$$

Se demuestra, asimismo, que los residuos recursivos reescalados se hallan incorrelacionados dos a dos<sup>8</sup>. Por lo tanto,

$$\mathbf{w} = N(\mathbf{0}, \sigma^2 \mathbf{I}_{n-k}) \quad (4.30)$$

Brown et al., basándose en estos residuos recursivos reescalados, sugieren dos contrastes de constancia de los parámetros. El primer estadístico de prueba es la cantidad **CUSUM**,

$$W_t = \sum_{j=k+1}^t w_j / \hat{\sigma} \quad t = k + 1, \dots, n \quad (4.31)$$

donde  $\sigma^2 = \text{SCR}_n / (n - k)$ , siendo  $\text{SCR}_n$  la suma de cuadrados de los residuos calculada a partir de la regresión de la totalidad de la muestra.  $W_t$  es el sumatorio *acumulado* calculado respecto a  $t$ . Cuando los parámetros son constantes,  $E(W_t) = 0$ , pero cuando no lo sean  $W_t$  tenderá a ser distinto a dicho valor. La significación del hecho de diferir de la línea que representa el valor igual a cero para la suma acumulada se obtiene calculando un par de líneas rectas que pasan por los puntos

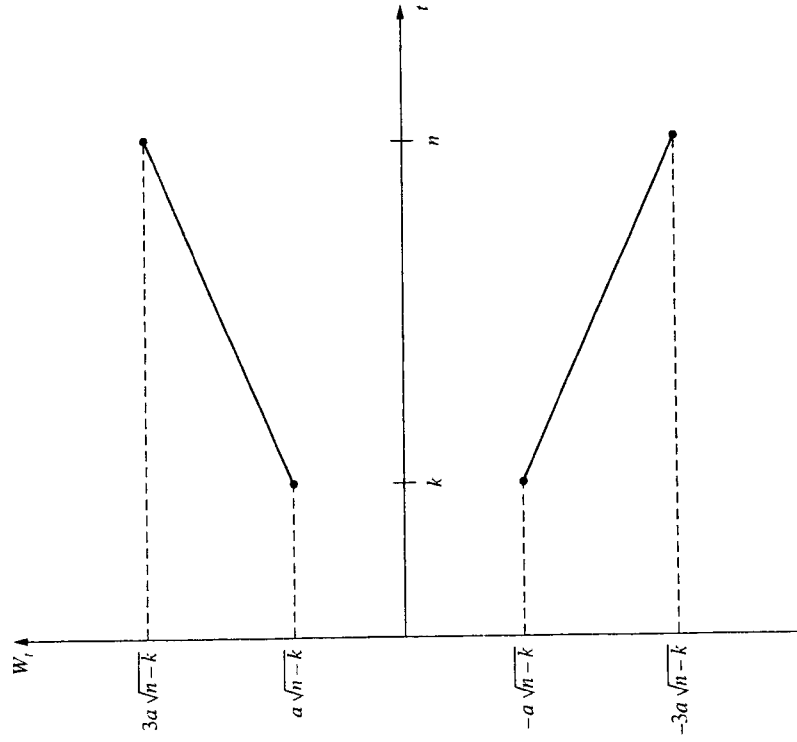
$$(k, \pm a \sqrt{n - k}) \quad \text{y} \quad (n, \pm 3a \sqrt{n - k})$$

donde  $a$  es un parámetro que depende del nivel de significación escogido para el contraste. La correspondencia para ciertos niveles de significación convencionales es

$\alpha = 0.01$	$a = 1.143$
$\alpha = 0.05$	$a = 0.948$
$\alpha = 0.10$	$a = 0.850$

La Figura 4.1 muestra las líneas para decidir la significación de la prueba.

<sup>8</sup> R.L. Brown, J. Durbin y J.M. Evans, "Techniques for Testing the Constancy of Regression Relationships over Time", *Journal of the Royal Statistical Society, Series B*, **35**, 1975, 149-192.



**FIGURA 4.1.**  
Gráfico CUSUM

El segundo estadístico de contraste se basa en los sumatorios acumulados de los cuadrados de los residuos,

$$S_t = \frac{\sum_{j=k+1}^t w_j^2}{\sum_{j=k+1}^n w_j^2} \quad t = k+1, \dots, n \quad (4.32)$$

Bajo la hipótesis nula, el cuadrado de las  $w$  son variables independientes distribuidas como una  $\chi^2(1)$ . Por lo tanto, el numerador tiene un valor esperado igual a  $t-k$  y el denominador, un valor esperado igual a  $n-k$ . Dado el valor esperado *aproximado* del estadístico de prueba bajo la hipótesis nula, la línea de valor medio es

$$E(S_t) = \frac{t-k}{n-k}$$

que va desde cero, cuando  $t = k$ , hasta la unidad cuando  $t = n$ . La significación de las discrepancias de la línea de valor esperado se calculan trazando un par de líneas paralelas a la línea  $E(S_t)$  a una distancia, por encima y por debajo, igual a  $c_0$ .

En el Apéndice D se tabulan los valores de  $c_0$  que están tomados del artículo de Brown, Durbin y Evans para distintos tamaños muestrales y niveles de significación.

Hansen (*op. cit.*) sugiere que el contraste CUSUM equivale a su contraste  $L_1$  (de estabilidad del término de intersección) y que el contraste CUSUMSQ equivale a su contraste  $L_{k+1}$  (de estabilidad de la varianza).

#### 4.3.6. Un Contraste Más General de Errores de Especificación: el Contraste RESET de Ramsey

Ramsey argumenta que los diversos errores de especificación mostrados en la Sección 4.1 (variables omitidas, formas funcionales incorrectas, correlación entre  $X$  y  $u$ ), originan un vector  $u$  distinto de cero<sup>9</sup>. Por lo tanto, las hipótesis nula y alternativa serán

$$H_0: u \sim N(0, \sigma^2 I)$$

$$H_1: u \sim N(\mu, \sigma^2 I) \quad \mu \neq 0$$

El contraste de la hipótesis nula  $H_0$  se basa en una regresión aumentada

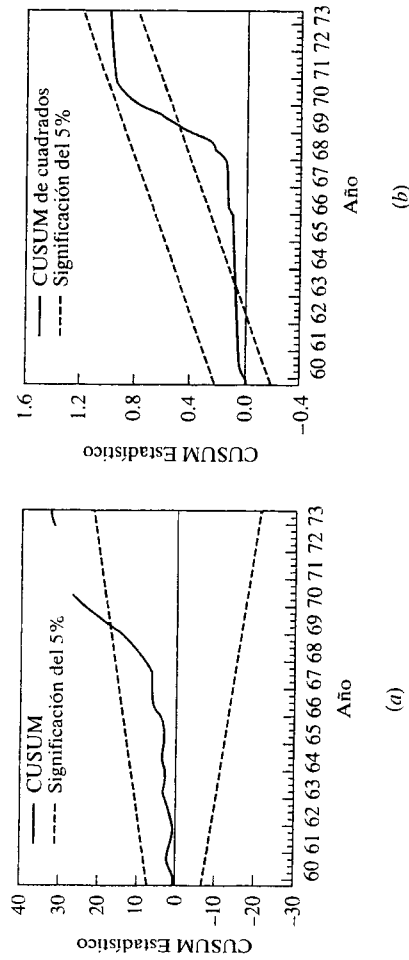
$$y = X\beta + Z\alpha + u$$

Así pues, el contraste de error de especificación es equivalente a probar que  $\alpha = 0$ . Ramsey sugiere que  $Z$  debería incluir las potencias de los *valores predichos de la variable dependiente*. Las potencias de segundo, tercer y cuarto grado nos dan

$$Z = [\hat{y}^2 \quad \hat{y}^3 \quad \hat{y}^4]$$

donde  $\hat{y} = X\beta$  y  $y^2 = [\hat{y}_1^2 \hat{y}_2^2 \dots \hat{y}_n^2]'$ , etc. Se excluye la primera potencia,  $\hat{y}$ , por ser una combinación lineal exacta de las columnas de  $X$ . En caso de incluirse, la matriz de regresión  $[X \ Z]$  no sería una matriz de rango pleno.

<sup>9</sup> J.B. Ramsey, "Tests for Specification Error in Classical Linear Least Squares Analysis", *Journal of the Royal Statistical Society, Series B*, 31, 1969, 350-371. Véase también J.B. Ramsey y P. Schmidt, "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests", *Journal of American Statistical Association*, 71, 1976, 389-390.



**FIGURA 4.6**  
Contrastes CUSUM de la ecuación de gasolina de la Sección 4.4.

Los contrastes CUSUM de la Fig. 4.6 confirman el mensaje de los gráficos anteriores. Por último, el contraste RESET de Ramsey, utilizando únicamente  $\hat{y}^2$ , nos da  $F = 47.2$ , que es un indicador muy fuerte del error de especificación cometido.

El proceso es similar al examen clínico de un paciente realizado por un habilitado médico. Algunos signos vitales se situarán en un rango aceptable de valores, aunque otros proporcionen lecturas contrarias. El médico deberá establecer el estado de salud global del paciente y si es necesaria alguna terapia. Y, lo que es más importante ¿podrá el médico sugerir las terapias adecuadas para los problemas graves? En este caso, parece evidente que nos encontramos ante un paciente enfermo. En capítulos posteriores examinaremos las terapias disponibles.

## 4.5 CONTRASTES DE CAMBIO ESTRUCTURAL

El contraste de predicción de Chow nos conduce, de forma natural, a contrastes más generales de cambio estructural. Un cambio estructural o una ruptura estructural se da cuando los parámetros de una relación difieren entre conjuntos de datos. Naturalmente, existen diversos subconjuntos de datos con la posibilidad de diversas rupturas estructurales. De momento, consideraremos solamente dos subconjuntos de  $n_1$  y  $n_2$  observaciones que dan una muestra total de  $n = n_1 + n_2$  observaciones. Supongamos, por ejemplo, que deseamos investigar si el consumo agregado en un país difiere en tiempos de paz y en tiempos de guerra, y que poseemos las observaciones de las variables relevantes de  $n_1$  años de paz y  $n_2$  años de guerra. Podríamos realizar un contraste de Chow utilizando la función estimada de tiempos de paz para predecir el

consumo en tiempos de guerra. Sin embargo, siempre que  $n_2 > k$ , se utilizará de forma alternativa la función estimada de tiempos de guerra para predecir el consumo en tiempos de paz. La elección no es evidente y ambos procedimientos proporcionarían respuestas distintas. Siempre que los subconjuntos sean lo suficientemente grandes, será mejor estimar ambas funciones y verificar los parámetros comunes.

#### 4.5.1. Contraste de un Cambio Estructural

El contraste de cambio estructural puede llevarse a cabo de tres modos distintos, aunque equivalentes. Supongamos que  $y_i$ ,  $X_i$  ( $i = 1, 2$ ) indican una partición adecuada de los datos. Formularemos el modelo no restringido como

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + u \quad u \sim N(0, \sigma^2 I) \quad (4.34)$$

donde  $\beta_1$  y  $\beta_2$  son los  $k$  vectores de coeficientes para los tiempos de paz y guerra, respectivamente. La hipótesis nula de inexistencia de cambio estructural es

$$H_0: \beta_1 = \beta_2 \quad (4.35)$$

El primer enfoque se basa en la aplicación directa del contraste de restricciones lineales definido en las ecuaciones (3.28) y (3.38). La hipótesis nula define  $R = [I_k - I_k]$  y  $r = 0$ . Sustituyendo la ecuación (3.38), obtenemos  $Rb - r = b_1 - b_2$ , donde  $b_1$  y  $b_2$  son los estimadores MCO de los vectores de coeficientes de la ecuación (4.34). Ajustando la ecuación (4.34) obtenemos la SCR del modelo no restringido,  $e'e$ . Los coeficientes MCO se formulan como

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'y_1 \\ X_2'y_2 \end{bmatrix} + \begin{bmatrix} (X_1'y_1)^{-1}X_1'y_1 \\ (X_2'y_2)^{-1}X_2'y_2 \end{bmatrix}$$

Por lo tanto, el modelo no restringido se estimará disponiendo de todos los datos, como en la ecuación (4.34), y realizando de una vez la estimación MCO, o bien ajustando por separado la ecuación a los datos de tiempos de paz y luego a los de tiempos de guerra. En este último caso, deberán sumarse ambas SCR para obtener la SCR del modelo no restringido, es decir,  $e'e = e_1'e_1 + e_2'e_2$ . La sustitución en la ecuación (3.38) permite contrastar las restricciones lineales.

En el Capítulo 3 hemos visto que la verificación de restricciones lineales puede formularse también en términos de una SCR no restringida y una SCR restringida. En este caso, la hipótesis nula nos da el siguiente modelo restringido

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + u \quad (4.36)$$

Denominaremos a la suma de cuadrados de los residuos, SCR, del modelo que ajus-

ta la ecuación (4.36) como  $e'e^*$ . La verificación de la hipótesis nula vendrá dada por

$$F = \frac{(e'e^* - e'e)/k}{e'e/(n-2k)} \sim F(k, n-2k)$$

Para la tercera posibilidad, consideramos una puesta en escena alternativa del modelo no restringido,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ X_2 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 - \beta_1 \end{bmatrix} + u \quad (4.37)$$

En este caso, la verificación de  $H_0$  es un simple contraste de significación para los últimos  $k$  regresores. La elección del procedimiento más adecuado dependerá mucho del programa informático utilizado.

#### 4.5.2 Contrastes acerca de la Pendiente

Las investigaciones económicas suelen interesarse por los coeficientes que representan la pendiente de la ecuación sin imponer restricciones al término de intersección. Con objeto de explicar dichos contrastes particionaremos las matrices  $X$  como sigue

$$X_1 = [i_1 \quad X_1'] \quad X_2 = [i_2 \quad X_2']$$

donde  $i_1$ ,  $i_2$  son  $n_1$  y  $n_2$  vectores de unos y las  $X_i'$  son matrices de las  $k-1$  variables regresores. La partición de los vectores  $\beta$  es

$$\beta_1 = [\alpha_1 \quad \beta_1'] \quad \beta_2 = [\alpha_2 \quad \beta_2']$$

$\beta_1^* = \beta_2^*$

Ahora la hipótesis nula es

El modelo no restringido es

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1' & 0 \\ 0 & i_2 & 0 & X_2' \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1' \\ \beta_2' \end{bmatrix} + u \quad (4.38)$$

El modelo restringido es

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1' \\ 0 & i_2 & X_2' \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta^* \end{bmatrix} + u \quad (4.39)$$

La verificación de la hipótesis nula puede basarse en las SCR de ambas regresiones. Ciertos programas informáticos que realizan la regresión proporcionan un término

de intersección de forma automática (insertando una columna de unos en los regresores). Dicho paso debe eludirse en el ajuste de las ecuaciones (4.38) y (4.39) con el fin de evitar la generación de regresores linealmente dependientes.

Una formulación alternativa del modelo no restringido es

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1^* & 0 \\ i_2 & i_2 & X_2^* & X_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta_1^* \\ \beta_2^* - \beta_1^* \end{bmatrix} + u \quad (4.40)$$

El contraste de significación conjunta de los últimos  $k - 1$  regresores es una prueba de la hipótesis nula planteada anteriormente. Idéntica advertencia sobre el término de intersección debe aplicarse a la estimación de la ecuación (4.40).

### 4.5.3 Contraste acerca de las Intersecciones

Podría parecer que la verificación de  $H_0: \alpha_1 = \alpha_2$  viene dada por la comprobación del segundo regresor de la ecuación (4.40). Una verificación de este tipo tiene, sin embargo, poco sentido. La estimación de la ecuación (4.40) no impone restricción alguna a las pendientes y, por lo tanto, lo que cuestiona la hipótesis de intersección es si existen dos superficies de regresión *distintas* que cortan al eje y en el mismo punto. El contraste de intersecciones distintas adquiere más sentido cuando el supuesto de pendientes de regresión *comunes* se convierte en una condición. La pregunta ahora es si el plano de regresión está situado más arriba o más abajo en la dirección que marca la variable dependiente. El modelo no restringido para dicho contraste es el especificado en la ecuación (4.39), que aparecía como modelo restringido en el contraste de las pendientes. El modelo restringido es

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & X_1^* \\ i_2 & X_2^* \end{bmatrix} \begin{bmatrix} \alpha \\ \beta^* \end{bmatrix} + u \quad (4.41)$$

La comparación de las SCR de las ecuaciones (4.39) y (4.41) proporciona una prueba de la igualdad de las intersecciones cuando las pendientes son iguales.

Una forma alternativa de mostrar el modelo no restringido es

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1^* \\ i_2 & i_2 & X_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta^* \end{bmatrix} + u \quad (4.42)$$

El contraste de significación del segundo regresor verifica la hipótesis condicional de que los términos de intersección son iguales.

### 4.5.4 Resumen

Los tres modelos guardan entre sí un orden jerárquico:

I: $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & X_1^* \\ i_2 & X_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta^* \end{bmatrix} + u$	Parámetros comunes
II: $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1^* \\ 0 & i_2 & X_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta^* \end{bmatrix} + u$	Términos de intersección distintos, Vector de pendientes común
III: $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1^* & 0 \\ 0 & i_2 & 0 & X_2^* \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{bmatrix} + u$	Términos de intersección distintos, Vectores de pendientes distintas

La aplicación de los MCO a cada modelo conducirá a una suma de cuadrados de los residuos, SCR, con grados de libertad asociados  $n - k$ ,  $n - k - 1$  y  $n - 2k$ , respectivamente. Los estadísticos de verificación para distintas hipótesis serán los siguientes:

$$H_0: \alpha_1 = \alpha_2 \quad \text{Contraste de términos de intersección distintos}$$

$$F = \frac{SCR_1 - SCR_2}{SCR_2 / (n - k - 1)} \sim F(1, n - k - 1)$$

$$H_0: \beta_1^* = \beta_2^* \quad \text{Contraste de vectores de pendientes distintos}$$

$$F = \frac{SCR_2 - SCR_3 / (k - 1)}{SCR_3 / (n - 2k)} \sim F(k - 1, n - 2k)$$

$$H_0: \beta_1 = \beta_2 \quad \text{Contraste de parámetros distintos (intersecciones distintas y pendientes distintas)}$$

$$F = \frac{SCR_1 - SCR_3 / k}{SCR_3 / (n - 2k)} \sim F(k, n - 2k)$$

Los grados de libertad de los numeradores es igual al número de restricciones impuestas por el cambio de modelo no restringido a modelo restringido. Dicho número equivale también a la diferencia en los grados de libertad de las sumas de cuadrados de los residuos del numerador.

El término de la izquierda de la ecuación es la SCR cuando únicamente se realiza la regresión de  $y$  sobre  $X_1$ . El último término,  $e'e$ , es la SCR cuando se realiza la regresión de  $y$  sobre  $[X_1 \ X_2]$ . Por lo tanto, el término medio mide el *incremento* sufrido por la SCE (o, de forma equivalente, la *disminución* de la SCR) cuando añadimos  $X_2$  al conjunto de regresores. Para contrastar la hipótesis planteada es preciso realizar dos regresiones distintas. Calcularemos primero la regresión de  $y$  sobre  $X_1$ , indicando la SCR como  $e'e_*$ . Después realizaremos la regresión para todas las  $X_s$ , obteniendo con ello la SCR que, como es habitual, vendrá indicada por  $e'e$ . Partiendo de la ecuación (3.38), el contraste estadístico es

$$F = \frac{(e'e_* - e'e)/k_2}{e'e/(n - k)} \sim F(k_2, n - k) \quad (3.42)$$

### 3.4.6 Regresiones Restringidas y No Restringidas

Es evidente que  $(v)$  es un caso especial de  $(vi)$  y, en consecuencia,  $(v)$  puede interpretarse también como el resultado de dos regresiones distintas. En la ecuación (3.9) comentábamos que la SCE puede expresarse como  $SCE = y'y_* - e'e$ , donde  $y_* = Ay$ , con  $A$  definido en la ecuación (3.7). Se demuestra que  $y'y_*$  es la SCR cuando se realiza la regresión de  $y_*$  sobre  $x_1$  ( $=f$ )<sup>17</sup>. Sustituyendo SCE en la ecuación (3.40), resulta que el estadístico de prueba resultante tiene idéntica forma que el de la ecuación (3.42).

En ambos casos,  $(v)$  y  $(vi)$ , denominamos a la primera regresión como **regresión restringida**, mientras que la segunda es una **regresión no restringida**. Del mismo modo,  $e'e_*$  es la SCR *restringida* y  $e'e$  es la SCR *no restringida*. En la regresión restringida, las restricciones incluidas en  $H_0$  se imponen en la ecuación *estimada*. Así pues, la regresión restringida de  $(v)$  omite los  $X_2, X_3, \dots, X_k$  de la regresión o, de forma equivalente, el conjunto de coeficientes  $b_2, b_3, \dots, b_k$  son todos iguales a cero. En  $(vi)$  la regresión restringida utiliza únicamente las variables de  $X_1$ . La regresión no restringida utiliza todas las variables de la matriz  $X$ . Siguiendo idéntico razonamiento, el Ejemplo (i) es también un caso especial de  $(vi)$ . En la regresión restringida se utilizan todas las variables excepto  $X_i$ . Por lo tanto, el contraste de significación de  $\beta_i$  cuestiona si existe una disminución significativa de la SCR (o un incremento de la SCE) cuando añadimos  $X_i$  al conjunto de regresores.

A los estudiantes suele resultarles difícil determinar el valor correcto de  $q$  en esos contrastes. Podemos calcularlo de distintas formas:

1. El número de filas de la matriz  $R$ .
2. El número de elementos en la hipótesis nula.

3. La diferencia entre el número de coeficientes  $\beta$  estimados en las ecuaciones restringida y no restringida.
4. La diferencia entre los grados de libertad asociados a  $e'e_*$  y  $e'e$ .

En los seis ejemplos anteriores, hemos desarrollado los estadísticos de prueba implicando en ellos los coeficientes  $b_i$  de la regresión no restringida. En los Ejemplos (i), (v) y (vi) hemos visto, sin embargo, que también es posible expresar los estadísticos de prueba en términos de la diferencia entre las SCR de las regresiones restringida y no restringida. En los tres casos, la regresión restringida se obtenía fácilmente excluyendo de la regresión las variables relevantes. Es ahora cuando surge la pregunta de si, en el resto de ejemplos, es posible realizar una interpretación similar en términos de la diferencia entre dos sumas de cuadrados de residuos. Para responder a la pregunta, necesitaremos examinar cómo se ajusta la regresión restringida en dichos casos.

### 3.4.7 Ajuste de la Regresión Restringida

El ajuste puede realizarse de dos modos. El primero consiste en desarrollar cada caso específico desde el principio; el segundo, en obtener una fórmula general que se ajuste a cada caso específico. Utilizaremos el Ejemplo (iii), con la regresión en forma de desviaciones, como ejemplo del primer método,

$$y = b_2x_2 + b_3x_3 + e$$

Impondremos la restricción de que  $b_2 + b_3 = 1$ . Sustituyendo la restricción en la regresión obtenemos una nueva ecuación:

$$y = b_2x_2 + (1 - b_2)x_3 + e_*$$

o

$$(y - x_3) = b_2(x_2 - x_3) + e_*$$

Se crean, por lo tanto, dos nuevas variables,  $(y - x_3)$  y  $(x_2 - x_3)$ . La regresión simple de la primera variable sobre la segunda (sin término de intersección) proporciona el estimador restringido de  $b_2$ . La SCR de esta regresión es la SCR restringida,  $e'_*e_*$ .

El método general requiere un vector  $b_*$  que minimice la SCR y que se halle sujeto a las restricciones  $Rb_* = r$ . Para obtenerlo establecemos la función siguiente

$$\phi = (y - Xb_*)'(y - Xb_*) - 2\lambda'(Rb_* - r)$$

donde  $\lambda$  es un vector de  $q$  multiplicadores de Lagrange. Las condiciones de primer orden son

$$\frac{\partial \phi}{\partial b_*} = -2X'y + 2(X'X)b_* - 2R'\lambda = 0$$

$$\frac{\partial \phi}{\partial \lambda} = -2(Rb_* - r) = 0$$

<sup>17</sup>Véase Problema 3.5.



La solución de  $\mathbf{b}_*$  es<sup>18</sup>

$$\mathbf{b}_* = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\mathbf{b}) \quad (3.43)$$

donde  $\mathbf{b}$  es el estimador MC habitual y no restringido  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . Los residuos de la regresión restringida son

$$\begin{aligned} \mathbf{e}_* &= \mathbf{y} - \mathbf{X}\mathbf{b}_* \\ &= \mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{X}(\mathbf{b}_* - \mathbf{b}) \\ &= \mathbf{e} - \mathbf{X}(\mathbf{b}_* - \mathbf{b}) \end{aligned}$$

Transponiendo y multiplicando, obtenemos

$$\mathbf{e}'_*\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\mathbf{b}_* - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}_* - \mathbf{b})$$

Utilizando el proceso de sustitución de la ecuación (3.43) para  $(\mathbf{b}_* - \mathbf{b})$  y simplificando, tenemos

$$\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e} = (\mathbf{r} - \mathbf{R}\mathbf{b})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\mathbf{b})$$

donde, dejando a un lado  $q$ , la expresión del lado derecho es la misma que la que aparece en el numerador del estadístico  $F$  de la ecuación (3.38). Por lo tanto, una expresión alternativa al estadístico de prueba para  $H_0: \mathbf{R}\mathbf{b} = \mathbf{r}$ , es

$$F = \frac{(\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e})/q}{\mathbf{e}'\mathbf{e}/(n-k)} \sim F(q, n-k) \quad (3.44)$$

Los seis ejemplos seguirán pasos idénticos.

Resumiendo, las pruebas tipo  $\mathbf{R}\mathbf{b} = \mathbf{r}$  podrán implementarse ajustando la regresión no restringida y sustituyendo el vector  $\mathbf{b}$  resultante en la ecuación (3.38). De modo alternativo, podremos ajustar también la regresión restringida y establecer pruebas estadísticas basadas en la diferencia  $(\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e})$  entre las SCR restringidas y no restringidas, como en el caso de la ecuación (3.44). Los siguientes ejemplos numéricos ilustran cada uno de estos procedimientos.

**EJEMPLO 3.7.** Seguiremos utilizando los datos del Ejemplo 3.3.

(i)  $H_0: \beta_3 = 0$ . El estadístico de prueba adecuado es  $t = b_3/s \sqrt{c_{33}}$ , donde  $c_{33}$  es el elemento inferior derecho de  $(\mathbf{X}'\mathbf{X})^{-1}$ . Partiendo de los resultados obtenidos anteriormente,  $b_3 = -1.5$  y  $s = \sqrt{\text{SCR}/(n-k)} = \sqrt{1.5/2} = \sqrt{0.75}$ . El término  $c_{33}$  se obtiene calculando directamente el determinante de la matriz  $3 \times 3$   $(\mathbf{X}'\mathbf{X})$  y dividiendo el adjunto relevante por ese determinante. Evaluar directamente el determinante resulta tedioso. Añadir múltiplos de las filas (columnas) a una fila (columna) de una matriz no altera el determinante y por ello resulta más sencillo buscar el determinante de la matriz obtenido anteriormente en la eliminación de Gauss del Ejemplo 3.3,

$$|\mathbf{X}'\mathbf{X}| = \begin{vmatrix} 5 & 15 & 25 \\ 0 & 10 & 6 \\ 0 & 0 & 0.4 \end{vmatrix} = 20$$

donde el determinante es, simplemente, el producto de los elementos de la diagonal principal. El adjunto relevante es

$$\begin{vmatrix} 5 & 15 \\ 15 & 55 \end{vmatrix} = 50$$

Así pues,  $c_{33} = 50/20 = 2.5$  y

$$t = \frac{-1.5}{\sqrt{0.75} \sqrt{2.5}} = -\sqrt{1.2} = 1.095$$

que está muy cerca de cualquier valor crítico convencional.

Cuando la hipótesis no incluye el término de intersección, suele ser más sencillo trabajar con los datos en forma de desviaciones porque se reduce la dimensionalidad del problema. En este caso,  $c_{33}$  será el elemento inferior derecho del inverso de  $(\mathbf{X}'_*\mathbf{X}_*)$ . Volviendo al Ejemplo 3.3, tenemos

$$[(\mathbf{X}'_*\mathbf{X}_*)]^{-1} = \begin{bmatrix} 10 & 6 \\ 6 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}$$

que es el resultado obtenido anteriormente.

Alternativamente, podemos analizar el problema en términos de la reducción de la SCR cuando añadimos  $X_3$  al conjunto de regresores. Según la Tabla 3.2,  $\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e} = 0.9$ . Por lo tanto,

$$F = \frac{\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e}}{\mathbf{e}'\mathbf{e}/(n-k)} = \frac{0.9}{1.5/2}$$

y  $t = \sqrt{F} = \sqrt{1.2}$ , igual que antes.

(ii)  $H_0: \beta = -1$ . El estadístico de prueba relevante en este caso es

$$t = \frac{-1.5 - (-1)}{s \sqrt{c_{33}}} = \frac{-0.5}{\sqrt{0.75} \sqrt{2.5}} = -0.365$$

que resulta no significativo. Podemos calcular también un intervalo de confianza del 95% para  $\beta_3$ . En la distribución  $t$ ,  $t_{0.025}(2) = 4.303$ . El intervalo es

$$b_3 \pm t_{0.025} \text{s.e.}(b_3)$$

es decir,  $-1.5 \pm 4.303 \sqrt{0.75} \sqrt{2.5}$

o sea,  $-7.39$  a  $4.39$

La amplitud del intervalo confirma que las hipótesis (i) y (ii) no se rechazan. Naturalmente, se trata de un ejemplo simplista en extremo, y el pequeño tamaño de la muestra no permite realizar inferencias de ningún tipo.

<sup>18</sup>Véase Apéndice 3.4.

(iii)  $H_0: \beta_2 + \beta_3 = 0$ . Partiendo de  $(X'X)^{-1}$  del caso (i)

$$\text{var}(b_2 + b_3) = 0,75[1 + 2,5 - 2(1,5)] = 0,375$$

$$\text{dando } t = \frac{1}{\sqrt{0,375}} = 1,63$$

que resulta no significativo.

(iv)  $H_0: \beta_2 = \beta_3 = 0$ . Deben tomarse precauciones y observar que existen diferencias entre ésta y las hipótesis previas. Sustituyendo en la ecuación (3.40), obtenemos

$$F = \frac{26,5/2}{1,5/2} = 17,67$$

Sin embargo,  $F_{0,05}(2,2) = 19,00$ , por lo tanto, y a pesar del impresionante valor de  $R^2$ , la hipótesis no resulta rechazada a un nivel del 5%.

### 3.5 PREDICCIÓN

Supongamos que hemos ajustado una ecuación de regresión y que consideramos algunos valores específicos del vector de regresión

$$c' = [1 \quad X_{2f} \quad \dots \quad X_{kf}]$$

Estos valores de las  $X$  pueden ser valores hipotéticos en el caso de aquel investigador que estudia los posibles efectos de escenarios distintos. También puede darse el caso de que se trate de valores observados nuevos. En cualquiera de estas dos situaciones, deseamos predecir el valor de  $Y$  condicional a  $c$ . Cualquier predicción de este tipo se basará en el supuesto de que el modelo ajustado permanece inalterado en el periodo de predicción. Es posible verificar dicho supuesto de estabilidad cuando se observa también un nuevo valor  $Y_f$ . Obtendremos una atractiva predicción por punto insertando los valores dados de  $X$  en la ecuación de regresión

$$\hat{Y}_f = b_1 + b_2 X_{2f} + \dots + b_k X_{kf} = c'b \quad (3.45)$$

En la discusión acerca el teorema de Gauss-Markov demostrábamos que  $c'b$  es el mejor estimador lineal insesgado de  $c'\beta$ . En el contexto actual  $c'\beta = E(Y_f)$ . Por lo tanto,  $\hat{Y}_f$  es un predictor óptimo de  $E(Y_f)$ . Además, la ecuación (3.30) demostraba que  $\text{var}(Rb) = R\text{var}(b)R'$ . Reemplazando  $R$  por  $c'$ , obtenemos

$$\text{var}(c'b) = c' \text{var}(b) c$$

Si suponemos que el término de perturbación es normal, tenemos que

$$\frac{c'b - c'\beta}{\sqrt{\text{var}(c'b)}} \sim N(0,1)$$

Cuando el valor desconocido de  $\sigma^2$  en  $\text{var}(b)$  se sustituye por  $s^2$ , tiene lugar la transformación habitual hacia la distribución  $t$ , resultando

$$\frac{\hat{Y}_f - E(Y_f)}{s\sqrt{c'(X'X)^{-1}c}} \sim t(n-k)$$

y el intervalo de confianza del 95% para  $E(Y_f)$  es

$$\hat{Y}_f \pm t_{0,025} s \sqrt{c'(X'X)^{-1}c} \quad (3.47)$$

**EJEMPLO 3.8.** Prosigamos con los datos del Ejemplo 3.3. Deseamos predecir  $E(Y)$  si  $X_2 = 10$  y  $X_3 = 10$ . La predicción por punto es

$$\hat{Y}_f = 4 + 2,5(10) - 1,5(10) = 14$$

$$\text{Invirtiendo } (X'X) \text{ se obtiene } = \begin{bmatrix} 26,7 & 4,5 & -8,0 \\ 4,5 & 1,0 & -1,5 \\ -8,0 & -1,5 & 2,5 \end{bmatrix}$$

$$\text{Por lo tanto, } c'(X'X)^{-1}c = \begin{bmatrix} 1 & 10 & 10 \end{bmatrix} \begin{bmatrix} 26,7 & 4,5 & -8,0 \\ 4,5 & 1,0 & -1,5 \\ -8,0 & -1,5 & 2,5 \end{bmatrix} = 6,7$$

y  $s^2 = 0,75$ , como antes. Por lo tanto, el intervalo de confianza del 95% para  $E(Y_f)$  es

$$14 \pm 4,303 \sqrt{0,75} \sqrt{6,7}$$

0                      4,35                      hasta                      23,65

Si no se dispone de PC y no deseamos invertir matrices  $3 \times 3$ , podemos reformular el Ejemplo 3.8 en términos de desviaciones, lo que disminuye la dimensionalidad en una unidad<sup>19</sup>. A veces lo que deseamos es obtener un intervalo de confianza para  $Y_f$  en lugar de buscarlo para  $E(Y_f)$ . La diferencia estriba en la perturbación  $u_f$  que esperamos aparezca en el periodo de predicción. Su valor es impredecible y, por lo tanto, la predicción por punto es la misma que la obtenida utilizando la ecuación (3.45). Se trata aún del mejor predictor lineal insesgado, pero la incertidumbre acerca de la predicción aumenta. Tenemos  $\hat{Y}_f = c'b$ , como antes, y ahora  $Y_f = c'\beta + u_f$ . El error de predicción es

$$e_f = Y_f - \hat{Y}_f = u_f - c'(b - \beta)$$

Elevando ambos lados al cuadrado y tomando esperanzas obtendremos la varianza del error de predicción

<sup>19</sup> Véase Problema 3.6.

$$\text{var}(e_f) = \sigma^2 + c' \text{var}(b) c$$

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

de donde derivamos un estadístico  $t$

$$\frac{\hat{Y}_f - Y_f}{s\sqrt{1 + c'(X'X)^{-1}c}} \sim t(n-k) \quad (3.4)$$

Comparando con la ecuación (3.46) se demuestra que la única diferencia estriba en el aumento de una unidad en el término que aparece debajo del símbolo de la raíz cuadrada. Por lo tanto, el intervalo de confianza revisado de los datos del Ejemplo 3.8 es

$$14 \pm 4.303 \sqrt{0.75} \sqrt{7.7}$$

o

$$3.66 \quad \text{hasta} \quad 24.34$$

## APÉNDICE

### APÉNDICE 3.1

**Para comprobar  $r_{12,3} = (r_{12} - r_{13}r_{23}) / \sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}$**

Recordemos que en el Capítulo 1 afirmábamos que, para una regresión de dos variables,  $y = bx + e$ , en forma de desviaciones,  $b = r_{y/x}/s_x$  y  $\sum e^2 = \sum y^2(1 - r^2) = ns_y^2(1 - r^2)$ , donde  $s$  indica la desviación muestral estándar de la variable del subíndice. Por lo tanto

$$\sum e_{1,3} = ns_1^2(1 - r_{13}^2) \quad \text{y} \quad \sum e_{2,3}^2 = ns_2^2(1 - r_{23}^2)$$

$$\text{También} \quad e_{1,3} = y - r_{13} \frac{s_1}{s_3} x_3 \quad \text{y} \quad e_{2,3} = x_2 - r_{23} \frac{s_2}{s_3} x_3$$

Simplificando se obtiene

$$\sum e_{1,3} e_{2,3} = ns_1 s_2 (r_{12} - r_{13} r_{23})$$

$$\text{y, por lo tanto} \quad r_{12,3} = \frac{\sum e_{1,3} e_{2,3}}{\sqrt{\sum e_{1,3}^2} \sqrt{\sum e_{2,3}^2}} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

### APÉNDICE 3.2

**Obtención de uno solo de los coeficientes de regresión en una regresión múltiple**  
Las ecuaciones normales son

$$\begin{bmatrix} x_1' x_1 & x_1' x_2 & x_1' x_3 \\ x_2' x_1 & x_2' x_2 & x_2' x_3 \\ x_3' x_1 & x_3' x_2 & x_3' x_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} x_1' y \\ x_2' y \\ x_3' y \end{bmatrix}$$

donde  $x_2$  es el  $n$ -vector de observaciones en  $X_2$  y  $X_*$  es la matriz  $n \times (k-1)$  de observaciones de todas las variables del lado derecho, incluyendo una columna de



### **Tema 3 Contraste de la Razón de verosimilitud, de Wald y de Multiplicadores de Lagrange**

3.1 Contraste de Razón de Verosimilitudes (RV)(5.3.1)

3.2 El contraste de Wald (W) (JyD 5.3.2)

3.3 Contraste de Multiplicadores de Lagrange (ML) (JyD 5.3.3)

3.4 Visión de conjunto de los tres tests ( *José M<sup>a</sup> Otero "Econometría. Series temporales y predicción" Ed.AC1993 pg.364*)

### 5.3 CONTRASTES DE LA RAZÓN DE VEROSIMILITUDES, DE WALD Y DE MULTIPLICADORES DE LAGRANGE

Ilustraremos todos estos contrastes en un contexto de hipótesis lineales sobre  $\beta$ . La hipótesis lineal general es

$$H_0: R\beta = r \quad (5.8)$$

donde  $R$  es una matriz  $q \times k$  ( $q < k$ ) de constantes conocidas y  $r$ , un vector conocido  $q \times 1$ . Los contrastes más importantes son los de RV, W y ML.

#### 5.3.1 Contraste de Razón de Verosimilitudes (RV)

Los EMV de las ecuaciones (5.5) y (5.6) maximizan la función de verosimilitud sin imponer restricción alguna. El valor resultante de  $L(\hat{\beta}, \hat{\sigma}^2)$  de la ecuación (5.7) es la máxima verosimilitud *no restringida* y se expresa como función de la *suma de cuadrados residual no restringida*,  $e'e$ . El modelo es susceptible asimismo de estimarse de forma *restringida*, maximizando la verosimilitud sujeta a las restricciones,  $R\beta = r$ . Los estimadores resultantes los simbolizaremos mediante  $\tilde{\beta}$  y  $\tilde{\sigma}^2$ . En este caso, la máxima verosimilitud se consigue sustituyendo dichos valores en la función de verosimilitud para obtener  $L(\tilde{\beta}, \tilde{\sigma}^2)$ . El máximo restringido no puede exceder del valor del máximo no restringido aunque, si las restricciones son válidas, esperaremos que el máximo restringido esté muy "cercano" al máximo no restringido. Definimos la razón de verosimilitudes como

$$\lambda = \frac{L(\tilde{\beta}, \tilde{\sigma}^2)}{L(\hat{\beta}, \hat{\sigma}^2)}$$

de modo que, intuitivamente, esperaremos rechazar la hipótesis nula cuando  $\lambda$  es "pequeño". En algunos casos, es posible realizar contrastes para muestras exactas y finitas para verificar la "pequeñez" de  $\lambda$ . Sin embargo, el contraste general para *muestras grandes* es de la forma:

$$RV = -2 \ln \lambda = 2[\ln L(\hat{\beta}, \hat{\sigma}^2) - \ln L(\tilde{\beta}, \tilde{\sigma}^2)] \xrightarrow{d} \chi^2(q) \quad (5.9)$$

Los EMV restringidos se obtienen maximizando

$$l^* = l - \mu' (R\beta - r) \quad (5.10)$$

donde  $\mu$  es un vector  $q \times 1$  de multiplicadores de Lagrange y  $l$  es el logaritmo de la verosimilitud especificada en la ecuación (5.4). Resulta fácil demostrar que  $\tilde{\beta}$  es, simplemente, el vector  $b_*$  restringido obtenido anteriormente en el análisis estándar de mínimos cuadrados [véase ecuación (3.43)]. Dicho vector satisface la limitación  $b_* = r$ . Si formulamos los correspondientes residuos como

$$e_* = y - Xb_*$$

el EMV restringido de  $\sigma^2$  es  $\tilde{\sigma}^2 = e_*' e_* / n$ , y, por lo tanto

$$L(\tilde{\beta}, \tilde{\sigma}^2) = \text{constante} \cdot (e_*' e_*)^{-n/2} \quad (5.11)$$

Sustituyendo las ecuaciones (5.7) y (5.11) en la ecuación (5.9), obtenemos el estadístico de prueba RV como  $RV = n(\ln e_*' e_* - \ln e'e)$ . Con el fin de referenciarlos a él posteriormente, apuntamos algunos formatos alternativos en la definición del estadístico RV:

$$\begin{aligned} LR &= n(\ln e_*' e_* - \ln e'e) \\ &= n \ln \left( 1 + \frac{e_*' e_* - e'e}{e'e} \right) \\ &= n \ln \left( \frac{1}{1 - (e_*' e_* - e'e)/e'e} \right) \end{aligned} \quad (5.12)$$

Como puede observarse, el cálculo del estadístico RV requiere ajustar tanto el modelo restringido como el no restringido.

#### 5.3.2 El Contraste de Wald (W)

En el procedimiento de Wald se calcula únicamente el vector  $\hat{\beta}$  no restringido. En este caso, el vector  $(R\hat{\beta} - r)$  indicará hasta qué punto los estimadores no restringidos MV ajustan la hipótesis nula. Cuando el mencionado vector se "acerque" a cero, la hipótesis nula tenderá a cumplirse; por otro lado, los valores "grandes" tenderán a contradecirla. Como  $\hat{\beta}$  se distribuye normal y asintóticamente, con vector media  $\beta$  y matriz de varianzas-covarianzas  $I^{-1}(\beta)$ , tendremos que con  $H_0$ ,  $(R\hat{\beta} - r)$  se distribuye asintóticamente según una normal multivariante con vector de medias igual a cero y matriz de varianzas-covarianzas dada por  $RI^{-1}(\beta)R'$ , donde  $I^{-1}(\beta) = \sigma^2(X'X)^{-1}$ . Como anteriormente comentamos, la matriz de información para el modelo de regresión lineal es una matriz diagonal por bloques y, por lo tanto, bastará por el momento que nos concentremos en la submatriz relacionada con  $\beta$ . Por consiguiente<sup>3</sup>

$$(R\hat{\beta} - r)' [RI^{-1}(\beta)R']^{-1} (R\hat{\beta} - r) \xrightarrow{d} \chi^2(q) \quad (5.13)$$

donde  $q$  es el número de restricciones incluidas en  $R$ . La distribución asintótica sigue existiendo cuando la  $\sigma^2$  desconocida en  $I^{-1}(\beta)$  se reemplaza por un estimador consistente  $\sigma^2 = e'e/n$ . El resultado es el estadístico de Wald,

<sup>3</sup> Como vimos en el Capítulo 3, y cuando las perturbaciones están distribuidas normalmente, dicho estadístico tiene una distribución muestral finita y exacta ( $q$ ).

Para evaluar el vector de gradientes en el valor del estimador restringido  $\hat{\theta}$ , sustituiremos  $u$  por  $e_* = y - X\hat{\beta}$  y  $\sigma^2 = e_*'e_*/n$ . El vector  $\hat{\beta}$  satisface la condición  $R\hat{\beta} = r$ . Por lo tanto,

$$s(\hat{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2} X'e_* \\ 0 \end{bmatrix}$$

Anteriormente mostramos el inverso de la matriz de información. Si la evaluamos en  $\hat{\theta}$ , obtenemos

$$I^{-1}(\hat{\theta}) = \begin{bmatrix} \hat{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

Sustituyendo en la ecuación (5.16), resulta

$$\begin{aligned} \text{ML} &= \begin{bmatrix} \frac{1}{\hat{\sigma}^2} e_*'X & 0 \end{bmatrix} \begin{bmatrix} \hat{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}^2} X'e_* \\ 0 \end{bmatrix} \\ &= \frac{e_*'X(X'X)^{-1}X'e_*}{\hat{\sigma}^2} \\ &= \frac{n e_*'X(X'X)^{-1}X'e_*}{e_*'e_*} \end{aligned} \quad (5.17)$$

Recopilaremos las fórmulas del Capítulo 3 referentes a las sumas de cuadrados explicada y total de la regresión múltiple. Así pues, el estadístico ML para la ecuación (5.17) es

$$\text{ML} = nR^2$$

donde  $R^2$  es el cuadrado del coeficiente de correlación múltiple de la regresión de  $e_*$  sobre  $X$ . Cuando la media de  $e_*$  es distinta de cero,  $R^2$  equivale al  $R^2$  no centrado. Sin embargo, frecuentemente ocurre que las restricciones se refieren únicamente a los coeficientes representativos de las pendientes,  $\beta_2, \beta_3, \dots, \beta_k$ . En este caso,  $e_*$  tiene media cero y el  $R^2$  es el estadístico centrado que aparece en los programas informáticos convencionales que calculan la regresión<sup>4</sup>. Por lo tanto, podemos llevar a cabo el contraste ML en dos pasos. Calcularemos primero el estimador restringido  $\hat{\beta}$  y obtendremos el vector residual  $e_*$ . A continuación realizamos la regresión de

<sup>4</sup> Para una explicación del  $R^2$  centrado y no centrado, véase Apéndice 5.2.

$$W = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R'(R\hat{\beta} - r)]}{\hat{\sigma}^2} \sim \chi^2(q) \quad (5.14)$$

En la ecuación (3.44), demostrábamos que podemos también formular el numerador de la ecuación (5.14) como  $(e_*'e_* - e'e)$ . Así pues, la fórmula alternativa del estadístico de Wald para verificar (5.8), será

$$W = \frac{n(e_*'e_* - e'e)}{e'e} \sim \chi^2(q) \quad (5.15)$$

### 5.3.3 Contraste de Multiplicadores de Lagrange (ML)

El test ML, conocido también como **contraste del gradiente (score)**, se basa en el vector de gradientes (scores),

$$s(\theta) = \frac{\partial \ln L}{\partial \theta} = \frac{\partial l}{\partial \theta}$$

Calculemos en primer lugar el estimador no restringido,  $\hat{\theta}$ , resolviendo  $s(\hat{\theta}) = \theta$ , donde  $s(\hat{\theta})$  indica el vector gradiente evaluado en  $\hat{\theta}$ . En general, cuando evaluamos el vector gradiente en  $\hat{\theta}$ , el estimador restringido será distinto de cero. Sin embargo, en el caso de que las restricciones sean válidas, el máximo restringido,  $l(\hat{\theta})$ , debería ser similar al máximo no restringido,  $l(\hat{\theta})$  y, por lo tanto, el gradiente anterior debería ser también cercano a cero. Anteriormente afirmábamos que el vector de gradientes tiene media cero y matriz varianzas-covarianzas proporcionada por la matriz de información  $I(\theta)$ . La forma cuadrática,  $s'(\theta)I^{-1}(\theta)s(\theta)$ , tiene una distribución  $\chi^2$ . Evaluando la fórmula cuadrática cuando  $\theta = \hat{\theta}$ , obtenemos la verificación de la hipótesis nula. Bajo la hipótesis nula, el resultado básico es que

$$\text{ML} = s'(\hat{\theta})I^{-1}(\hat{\theta})s(\hat{\theta}) \sim \chi^2(q) \quad (5.16)$$

Destacaremos que todos los elementos de la ecuación (5.16) se evalúan en  $\hat{\theta}$ . A diferencia de lo que sucedía con el contraste de Wald, ahora únicamente necesitamos calcular el estimador restringido. La popularidad de los contrastes ML se debe al hecho de que, en la mayoría de los casos, resulta mucho más fácil calcular el estimador restringido que el no restringido.

Partiendo de los desarrollos iniciados en las ecuaciones (5.5) y (5.6), el vector de gradientes es

$$s(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} X'u \\ -\frac{n}{2\sigma^2} + \frac{u'u}{2\sigma^4} \end{bmatrix}$$

$e_*$  sobre  $X$  y compararemos el resultado de hacer  $nR^2$  con el de la distribución  $\chi^2(q)$ . El procedimiento en dos pasos suele aplicarse en aquellos casos en los que maximizar la verosimilitud equivale a minimizar una suma de cuadrados. Demostramos<sup>5</sup> así que la ecuación (5.17) puede reformularse como

$$ML = \frac{n(e_*' - e'e)}{e_*'e_*} \quad (5.18)$$

Ilustraremos ahora la famosa desigualdad de los tres contrastes estadísticos para el caso de un modelo lineal, esto es,  $W \geq RV \geq ML$ . Aplicando los primeros dos términos de la expansión logarítmica  $\ln(1+z) = z - 1/2 z^2 + \dots$  en la segunda expresión de  $RV$  de la ecuación (5.12), obtenemos

$$\begin{aligned} LR &= \frac{n(e_*' - e'e)}{e'e} - \frac{n}{2} \left( \frac{e_*'e - e'e}{e'e} \right)^2 \\ LR &= -n \ln \left( 1 - \frac{e_*'e - e'e}{e'e} \right) \\ &= \frac{n(e_*'e - e'e)}{e'e} + \frac{n}{2} \left( \frac{e_*'e - e'e}{e'e} \right)^2 \end{aligned}$$

que implica  $RV \leq W$ . De modo similar, utilizando la tercera expresión de la ecuación (5.12), obtenemos

$$LR = -n \ln \left( 1 - \frac{e_*'e - e'e}{e'e} \right)$$

$$= \frac{n(e_*'e - e'e)}{e'e} + \frac{n}{2} \left( \frac{e_*'e - e'e}{e'e} \right)^2$$

por lo tanto,  $RV \geq ML$  y, finalmente,  $W \geq RV \geq ML$ . Los contrastes son asintóticamente equivalentes aunque, en muestras finitas en general, se obtendrán resultados numéricos distintos.

**EJEMPLO 5.1.** Volvamos a los datos del Ejemplo 3.3 para verificar  $H_0: \beta_3 = 0$  para esos contrastes asintóticos. Partiendo de la Tabla 3.2, vemos que la regresión no restringida de  $Y$  sobre  $X_2$  y  $X_3$  da como resultado  $e'e = 1.5$ ; y la regresión restringida (al excluir  $X_3$ ), da como resultado  $e_*'e = 2.4$ .

Sustituyendo, en este orden, las ecuaciones (5.15), (5.12) y (5.18), obtenemos

$$W = 5(2.4 - 1.5)/1.5 = 3.00$$

$$RV = 5 \ln(2.4/1.5) = 2.35$$

$$ML = 5(2.4 - 1.5)/2.4 = 1.875$$

El 5% de  $\chi^2(1)$  es 3.841 y, por lo tanto, ninguno de los contrastes rechazará la hipótesis nula.

<sup>5</sup> Véase Apéndice 5.3.

De todos modos, pocas conclusiones deben esperarse cuando se utilizan contrastes asintóticos en muestras pequeñas como la empleada en este ejemplo.

Cuando se utiliza la serie residual de la regresión restringida de  $Y$  sobre  $X$  como la variable dependiente de una regresión de  $X_2$  y  $X_3$ ,  $R^2$  equivaldrá a 0.375 y, según el valor del anterior estadístico LM,  $nR^2 = 5(0.375) = 1.875$ .

## 5.4 ESTIMACIÓN MV DEL MODELO LINEAL CON PERTURBACIONES NO ESFÉRICAS

Consideraremos ahora el modelo

$$y = X\beta + u \quad \text{con } u \sim N(0, \sigma^2\Omega) \quad (5.19)$$

donde  $\Omega$  es una matriz finita positiva de orden  $n$ . Este modelo se conoce como el caso de perturbaciones no esféricas, opuesto a  $\text{var}(u) = \sigma^2 I$ , que es el caso de perturbaciones esféricas. Supondremos que los elementos de  $\Omega$  son conocidos. Por ejemplo, cuando la varianza de la perturbación de todos los puntos de la muestra es proporcional al cuadrado de uno de los regresores, supongamos  $X_2$ , tendremos

$$\text{var}(u_i) = \sigma^2 X_{2i}^2 \quad i = 1, 2, \dots, n$$

donde  $\sigma^2$  es un factor de escala. La matriz de varianzas-covarianzas de las perturbaciones es

$$\begin{aligned} \text{var}(u) = \sigma^2 \begin{bmatrix} X_{21}^2 & 0 & \dots & 0 \\ 0 & X_{22}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{2n}^2 \end{bmatrix} &= \sigma^2 \text{diag} \{X_{21}^2, X_{22}^2, \dots, X_{2n}^2\} \end{aligned}$$

Partiendo de la ecuación (5.19), vemos que la función de densidad normal multivariante para  $u$ , es

$$f(u) = (2\pi)^{-n/2} |\sigma^2\Omega|^{-1/2} \exp \left[ -\frac{1}{2} u' (\sigma^2\Omega)^{-1} u \right]$$

Dado que  $|\sigma^2\Omega| = \sigma^{2n} |\Omega|$ , podemos reformular la función de densidad como

$$f(u) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp \left[ (-1/2\sigma^2) u' \Omega^{-1} u \right]$$

Entonces, el logaritmo de verosimilitud es

$$l = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta) \quad (5.20)$$

Diferenciando respecto a  $\beta$  y  $\sigma^2$ , obtenemos



#### 11.1.4. Visión de conjunto de los tres tests

Los tres tests difieren en la clase de información que utilizan. En efecto, *RV* utiliza información referida tanto a la estimación restringida como a la estimación sin restringir, *W* utiliza sólo información relativa a la estimación sin restricciones y *ML* sólo usa las estimaciones restringidas. En algunos casos, las dificultades de cálculo que conlleva la estimación con restricciones pueden llegar a decantar la elección en el test de Wald.

Los tres procedimientos son asintóticamente equivalente pero, en general, sus propiedades con muestras pequeñas difieren. En el caso de que los tres tests se apliquen para verificar un conjunto de *restricciones lineales* referidas a los coeficientes de un modelo de regresión lineal para el que se conoce la matriz de varianzas y covarianzas del término de perturbación, se obtiene el siguiente resultado:

$$RV = W = ML \quad (11.15)$$

En tal caso, la distribución de los tres tests es una  $\chi^2(q)$ , idéntica a todos los tamaños muestrales. Este resultado se mantiene para cualquier verificación de restricciones lineales en que la función de log-verosimilitud sea cuadrática.

Si en el modelo lineal la matriz de varianzas y covarianzas es desconocida, como ocurre en la práctica, hay que estimarla, por lo que la función de log-verosimilitud deja de ser cuadrática. En este caso se produce sistemáticamente el siguiente resultado

$$W \geq RV \geq ML \quad (11.16)$$

Los tres tests suelen utilizarse para la estimación de sistemas no lineales porque sus resultados pueden diferir con muestras pequeñas.

### 11.2. TESTS DE RESIDUOS

Se contemplan a continuación dos tests que vienen gozando de bastante popularidad en el análisis de los residuos. El primero sirve para verificar la ausencia de autocorrelación (Test de Breusch-Godfrey). El segundo se aplica para detectar de determinado tipo de heteroscedasticidad (ARCH). Ambos se basan en el criterio del multiplicador de Lagrange. Siguen otros dos tests de residuos muy usados para verificar heteroscedasticidad general y ciertos tipos de especificación errónea.

#### 11.2.1. Test de Breusch-Godfrey de errores no autocorrelacionados

El test de Durbin, expuesto en §5.3.3, si bien sirve para tratar con regresores que sean valores retardados de la variable dependiente, está limitado al caso de autoco-

rrrelaci  
to en  
las ap  
ge. Se

en do

siend  
de  $r$   
es la

y los  
a toc

en d  
(11.  
mul  
 $\chi^2(r)$   
a ur  
cha  
moc

11.2

Tra  
sos  
tar  
inte  
cor  
el p  
(19

† E  
t

