Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid

Econometría I

Tema 1 Especificación y Estimación del Modelo Lineal General

Marcos Bujosa

Material de apoyo para el curso *Econometría I*

 marcos.bujosa@ccee.ucm.es
Actualizado el: 9 de octubre de 2007
Version 2.01

Índice

Este es un material de apoyo a las clases. En ningún caso sustituye a los libros de texto que figuran en el programa de la asignatura; textos que el alumno debe estudiar para afrontar el examen final con ciertas garantías de éxito.

Referencias recomendadas para la asignatura: Novales (1993), Wooldridge (2006), Verbeek (2004)

Otra referencia seguida en la elaboración de este material es el capítulo 1 de Hayashi (2000), que se puede descargar desde:
http://www.pupress.princeton.edu/chapters/s6946.pdf

Especificación y Estimación del Modelo Lineal General

- Capítulos 1, 2 y 3 y secciones 4.1, 4.2, 6.2 y 6.3 de Wooldridge (2006)
- Apéndices E1, E2 y E3 de Wooldridge (2006)

1. Introducción

- Léase el Capítulo 1 de Wooldridge (2006)

Otra referencia seguida en la elaboración de este material es el capítulo 1 de Hayashi (2000), que se puede descargar desde: `http://www.pupress.princeton.edu/chapters/s6946.pdf`

1.1. El punto de vista estadístico: Regresión como descomposición ortogonal

Descomposición ortogonal y causalidad 1

$$Y = \mathrm{E}(Y \,|\, \mathcal{D}) + U$$

donde el conjunto de información es $\mathcal{D}: \ (\mathbf{X} = \mathbf{x})$; por tanto

$$Y = \mathrm{E}(Y \,|\, \mathbf{X}) + U$$

donde $\mathrm{E}(Y \,|\, \mathbf{x})$ es una función arbitraria

lectura estadística: de izquierda a derecha.

Siempre es cierta. No implica causalidad ni conclusiones teóricas

lectura teórica: de derecha a izquierda.

Interpretación puede ser falsa (regresiones espurias)

De Spanos (1999, Capítulo 7, en particular la Sección 7.5.3)

Sea Y una variable aleatoria con segundo momento finito, es decir, $\mathrm{E}\left(|Y|^2\right) < \infty$, y un conjunto de información $\mathcal{D}$; entonces siempre podemos encontrar una descomposición de Y como la siguiente:

$$Y = \mathrm{E}(Y \,|\, \mathcal{D}) + U \tag{1.1}$$

donde

$\mathrm{E}(Y \,|\, \mathcal{D})$**:** es el componente sistemático [1]

U**:** es el componente NO-sistemático

La existencia de dicha descomposición[2] está garantizada siempre que $\mathrm{E}\left(|Y|^2\right) < \infty$.

Ambos componentes de Y satisfacen las siguientes propiedades

1. $\mathrm{E}(U \,|\, \mathcal{D}) = 0$
2. $\mathrm{E}\left(U^2 \,|\, \mathcal{D}\right) = \mathrm{Var}(Y \,|\, \mathcal{D}) < \infty$
3. $\mathrm{E}\left(U \cdot \left[\mathrm{E}(Y \,|\, \mathcal{D})\right]\right) = 0$ por tanto **ambos componentes son ortogonales**.

Supondremos que disponemos de una sucesión de variables aleatorias Y_n (para $n = 1, \ldots, N$) y de una matriz de variables aleatorias $\underset{[N \times k]}{\mathbf{X}}$; y que nuestro conjunto de información $\mathcal{D}$ es

$$\mathcal{D}: \ (\mathbf{X} = \mathbf{x})$$

es decir, *el conjunto de variables aleatorias* $\mathbf{X}$ (en total $N \times k$ variables) *ha tomado conjuntamente la matriz de valores* $\mathbf{x}$.

Siendo así, la descomposición ortogonal para cada Y_n queda como sigue:

$$Y_n = \mathrm{E}(Y_n \,|\, \mathbf{X}) + U_n$$

[1] vea la Sección ??, en la página~??, del Tema 2 del curso de Introducción a la Econometría de LECO

[2] Si interpretamos las variables aleatorias con varianza finita como elementos de un espacio vectorial, entonces $\mathrm{E}(Y \,|\, \mathcal{D})$ representa una proyección ortogonal, y la descomposición (1.1) es análoga al teorema de proyección ortogonal (Luenberger, 1968), con $\mathrm{E}(Y \,|\, \mathcal{D})$ como el mejor predictor en el sentido de la propiedad ECSV4 en la página~?? del Tema 2 del curso de Introducción a la Econometría de LECO.

Nótese que esta es una descomposición *puramente estadística*. Únicamente nos dice que si disponemos de cierta información acerca de las variables $\mathbf{X}$, podemos descomponer la variable Y_n en dos partes. Pero no hay una *teoría económica* detrás; por tanto no dice si hay relaciones de causalidad entre las variables. Podría ocurrir que:

1. bien las variables $\mathbf{X}$ generaran parcialmente a Y (y por tanto, al conocer $\mathcal{D}: \ (\mathbf{X} = \mathbf{x})$ sabemos qué parte de Y es debida a $\mathbf{X}$ y qué parte no)
2. o bien que Y causa (o genera) las variables $\mathbf{X}$ (y por tanto, al observar $\mathcal{D}: \ (\mathbf{X} = \mathbf{x})$ sabemos qué cabe esperar que ha ocurrido con la variable causante Y; como cuando vemos llover por la ventana, y entonces sabemos que hay nubes en el cielo
3. o bien, que hay alguna otra causa común (y quizá desconocida) que genera conjuntamente tanto a Y como a $\mathbf{X}$ (y observar lo que ha ocurrido con $\mathbf{X}$ (la información $\mathcal{D}$) nos indica que cabe esperar que ha ocurrido con Y (puesto que tienen un causante común).

La descomposición ortogonal

$$Y_n = \mathrm{E}(Y_n \,|\, \mathbf{X}) + U_n$$

se lee de **izquierda a derecha** (es decir, *"puedo descomponer Y_n en las dos partes descritas a la derecha"*), y no hay una *teoría* detrás.

1.2. El punto de vista del Análisis Económico: Regresión como modelo explicativo

Como economistas deseamos que la descomposición estadística de más arriba sea reflejo de las relaciones teóricas entre $\mathbf{X}$ y Y. En este sentido queremos leer la relación de **derecha a izquierda**, es decir Y (por ejemplo el consumo) está generado por una función de las variables $\mathbf{X}$ (por ejemplo una función de la renta) junto a otras causas distintas de la renta (U).

Esta visión sugiere algunos de los nombres dados tanto para Y como para $\mathbf{X}$. No obstante (y a pesar de los nombres), no debemos nunca perder de vista que la descomposición ortogonal es una relación estadística que siempre[3] podemos encontrar; pero que en general no permite sacar conclusiones teóricas de ella (regresiones espurias). Sólo en aquellos casos en que las variables situadas a derecha e izquierda provienen de un modelo teórico bien establecido, que nos sugiere qué variables son causantes (y por ello las situamos a derecha) y cuáles son causadas (izquierda) quizá podamos sacar conclusiones. La palabra "quizá", se debe a que con frecuencia los datos disponibles no miden aquellos conceptos empleados en los modelos teóricos (consumo permanente, preferencias, nivel de precios, utilidades, aversión al riesgo, etc.), o bien a que los modelos no están correctamente especificados (temas que se verán en otros cursos de econometría).

⇑ **Modelo de regresión** 2

$$Y_n = h\big(\mathbf{X}\big) + U_n = h\big(1,\ \boldsymbol{X}_{\blacktriangledown 2}, \ldots,\ \boldsymbol{X}_{\blacktriangledown k}\big) + U_n$$

donde :

- Y_n: Vble. endógena, objetivo, explicada (o regresando)
- $\mathbf{X} = \big[\mathbf{1},\ \boldsymbol{X}_{\blacktriangledown 1}, \ldots, \boldsymbol{X}_{\blacktriangledown k}\big]$ Vbles. exógenas, de control, explicativas (o regresores)
- U_n: factor desconocido o perturbación

Suponemos que la variable aleatoria Y en el momento n, es decir, Y_n es función del vector $\boldsymbol{X}_{n\triangleright}$ y de U_n.

Llamamos a Y *vble. endógena* (porque consideramos que se determina su valor o características a través del modelo), *vble. objetivo* (porque es una magnitud que deseamos controlar, por ejemplo la inflación si somos la autoridad monetaria) o simplemente ***regresando***.

La matriz $\mathbf{X} = \big[\mathbf{1},\ \boldsymbol{X}_{\blacktriangledown 1}, \ldots, \boldsymbol{X}_{\blacktriangledown k}\big]$: esta constituida por k columnas de variables que llamamos *exógenas* (porque consideramos que vienen dadas de manera externa al modelo), o *vbles. de control* (porque tenemos capacidad de alterar su valor para, a través del modelo, *controlar* Y; por ejemplo fijar la oferta monetaria o los tipos de interés en el ejemplo anterior), o simplemente ***regresores***.

U_n es el efecto conjunto de otras variables o circunstancias que influyen en la observación de Y_n, y que decidimos no contemplar en el modelo por alguna razón (dificultad o imposibilidad de observarlas) o sencillamente que desconocemos. También puede ser sencillamente un error cometido al medir Y_n. Llamamos a U_n ***perturbación***.

[3]siempre y cuando $\mathrm{E}\left(|Y_n|^2\right) < \infty$

⇑ **Tipos de datos** 3

- Datos temporales (series de tiempo)
- Sección cruzada
- Datos de panel

2. Modelo Clásico de Regresión Lineal

⇑ **Modelo Clásico de Regresión Lineal** 4

Modelo especial en el que la descomposición ortogonal

$$Y_n = \mathrm{E}(Y_n \mid \mathbf{X}) + U_n$$

es tal que $\mathrm{E}(Y_n \mid \mathbf{x})$ es una función lineal de $\boldsymbol{x}_{n\triangleright}$
$\mathrm{Var}(Y_n \mid \mathbf{x})$ es una constante (**homocedasticidad**)

¿QUÉ DEBO SUPONER PARA QUE ESTO SE CUMPLA?
(¡al menos como lectura estadística!)

En el análisis de regresión estamos interesados en estimar los dos primeros momentos de Y_n condicionados a $\mathbf{X} = \mathbf{x}$, es decir, $\mathrm{E}(Y_n \mid \mathbf{x})$ y $\mathrm{Var}(Y_n \mid \mathbf{x})$.

El modelo *Modelo Clásico de Regresión Lineal* es un caso particular en el que $\mathrm{E}(Y_n \mid \mathbf{x})$ es función lineal de $\boldsymbol{x}_{n\triangleright}$ (los regresores con subíndice n, es decir, del instante n, o de la empresa n, o del país n, o del individuo n, ...) y $\mathrm{Var}(Y_n \mid \mathbf{x})$ es una función constante (por tanto $Y_n \mid \mathbf{x}$ es *homocedástica*).

A continuación, vamos a describir los tres supuestos de un modelo econométrico que garantizan la existencia de una descomposición ortogonal como la del modelo clásico de regresión lineal. El cuarto supuesto, que garantiza que la estimación de la relación lineal es única, lo veremos en la sección siguiente.

2.1. Tres primeros supuestos en el Modelo Clásico de Regresión Lineal

- Capítulos 2 y 3 de Wooldridge (2006)
- Sección 6.2 de Wooldridge (2006)
- Apéndice E1 de Wooldridge (2006)

⇑ **Supuesto 1: linealidad** 5

$h(\cdot)$ es lineal: $Y_n = h\big(\boldsymbol{X}_{n\triangleright}\big) + U_n = a_1 + a_2 X_{t2} + a_3 X_{t3} + \cdots + a_k X_{tk} + U_n$
por lo tanto

$$\begin{aligned} Y_1 &= \quad a_1 + a_2 X_{12} + a_3 X_{13} + \cdots + a_k X_{1k} + U_1 \\ Y_2 &= \quad a_1 + a_2 X_{22} + a_3 X_{23} + \cdots + a_k X_{2k} + U_2 \\ \cdots & \qquad\qquad\qquad \cdots \\ Y_N &= \quad a_1 + a_2 X_{N2} + a_3 X_{N3} + \cdots + a_k X_{Nk} + U_N \end{aligned}$$

o

$$Y_n = \boldsymbol{X}_{n\triangleright}\boldsymbol{\beta} + U_n$$

donde $\boldsymbol{\beta} = (a_1, \ldots, a_k)'$, y $\boldsymbol{X}_{n\triangleright} = \begin{bmatrix} 1 & X_{n2} & X_{n3} & \cdots & X_{nk} \end{bmatrix}$ es decir

$$\underset{[N\times 1]}{\boldsymbol{Y}} = \underset{[N\times k]}{\mathbf{X}} \; \underset{[k\times 1]}{\boldsymbol{\beta}} + \underset{[N\times 1]}{\boldsymbol{U}}$$

donde

$$\boldsymbol{Y} = \big[Y_1, \ldots, Y_N\big]', \; \mathbf{X} = \big[\mathbf{1}, \; \boldsymbol{X}_{\blacktriangledown 2}, \ldots, \; \boldsymbol{X}_{\blacktriangledown k}\big], \; \boldsymbol{U} = \big[U_1, \ldots, U_N\big]'$$

es decir,

$$\mathbf{X} = \begin{bmatrix} 1 & X_{12} & X_{13} & \ldots & X_{1k} \\ 1 & X_{22} & X_{23} & \ldots & X_{2k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{N2} & X_{N3} & \ldots & X_{Nk} \end{bmatrix};$$

$$\text{o bien } \mathbf{X} = \begin{bmatrix} \boldsymbol{X}_{1\triangleright} \\ \boldsymbol{X}_{2\triangleright} \\ \vdots \\ \boldsymbol{X}_{N\triangleright} \end{bmatrix} = \left[\mathbf{1}, \boldsymbol{X}_{\blacktriangledown 2}, \ldots, \ \boldsymbol{X}_{\blacktriangledown k}\right]; \qquad \text{donde} \qquad \boldsymbol{X}_{\blacktriangledown j} = \begin{bmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{Nj} \end{bmatrix}$$

por tanto

$$\begin{aligned} \boldsymbol{Y} =& \left[\mathbf{1}, \boldsymbol{X}_{\blacktriangledown 2}, \ldots, \ \boldsymbol{X}_{\blacktriangledown k}\right] \boldsymbol{\beta} + \boldsymbol{U} \\ =& a_1 + a_2 \boldsymbol{X}_{\blacktriangledown 2} + a_3 \boldsymbol{X}_{\blacktriangledown 3} + \cdots + a_k \boldsymbol{X}_{\blacktriangledown k} + \boldsymbol{U} \end{aligned}$$

es decir

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = a_1 \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + a_2 \cdot \begin{bmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{N2} \end{bmatrix} + a_3 \cdot \begin{bmatrix} X_{13} \\ X_{23} \\ \vdots \\ X_{N3} \end{bmatrix} + \cdots + a_k \cdot \begin{bmatrix} X_{1k} \\ X_{2k} \\ \vdots \\ X_{Nk} \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}$$

⇑ **Supuesto 1: linealidad** 6

Modelo	Interpretación	
$Y_n = \beta X_n + U_n$	$\beta = \frac{dY_n}{dX_n}$	Cambio esperado en nivel de Y_n cuando X_n aumenta una unidad
$\ln(Y_n) = \beta \ln(X_n) + U_n$	$\beta = \frac{X_n}{Y_n}\frac{dY_n}{dX_n}$	Cambio porcentual (en tanto por uno) esperado en Y_n cuando X_n aumenta un uno por ciento (en tanto por uno, ie, 0.01)
$\ln(Y_n) = \beta X_n + U_n$	$\beta = \frac{1}{Y_n}\frac{dY_n}{dX_n}$	Cambio porcentual (en tanto por uno) esperado en Y_n cuando X_n aumenta una unidad
$Y_n = \beta \ln(X_n) + U_n$	$\beta = X_n\frac{dY_n}{dX_n}$	Cambio esperado en el nivel de Y_n cuando X_n aumenta un uno por ciento (en tanto por uno)

Más tipos de modelos lineales en Ramanathan (1998, Capítulo 6, pp. 232 y siguientes) y en el material preparado por J. Alberto Mauricio http://www.ucm.es/info/ecocuan/jam/ectr1/Ectr1-JAM-Guion.pdf

Ejemplo 1. [**función de consumo:**]

$$CON_n = \beta_1 + \beta_2 RD_n + U_n$$

donde CON_n y RD_n son el consumo y la renta disponible del individuo n-ésimo respectivamente, y U_n son otros factores que afectan al consumo del individuo n-ésimo distintos a su renta disponible (activos financieros, estado de ánimo, etc.).

Aquí la variable exógena Y es el consumo (CON), y los regresores son X_1 =1 (una constante) y X_2 la renta disponible (RD).

Ejemplo 2. [**ecuación de salarios:**] Supongamos el siguiente modelo no-lineal en los parámetros

$$SALAR_n = e^{\beta_1 + \beta_2 EDUC_n + \beta_3 ANTIG_n + \beta_4 EXPER_n + U_n};$$

donde $SALAR_n$ es el salario del individuo n-ésimo, $EDUC_n$ son sus años de educación, $ANTIG_n$ sus años de antigüedad en la empresa, y $EXPER_n$ sus años de experiencia en el sector de la empresa.

Al tomar logaritmos tenemos un nuevo modelo para $\ln(SALAR_n)$ que es lineal en los parámetros:

$$\ln(SALAR_n) = \beta_1 + \beta_2 EDUC_n + \beta_3 ANTIG_n + \beta_4 EXPER_n + U_n$$

En este caso la interpretación de un valor como $\beta_2 = 0.03$ es que *un año adicional en la formación educativa implica un incremento esperado del salario del 3 %.*

Ejemplo 3. [**función de producción Cobb-Douglas:**] Pensemos en la clásica función de producción

$$Q_n = cK_n{}^{\beta_2} L_n{}^{\beta_3}$$

donde Q_n es la producción el el momento n, K_n es el capital empleado en el instante n; L_n el trabajo empleado en n. Supongamos, además, que hay un efecto aleatorio adicional ν_n debido a otras causas o factores

$$Q_n = cK_n{}^{\beta_2} L_n{}^{\beta_3}\nu_n;$$

tomando logaritmos tenemos

$$\ln Q_n = \beta_1 + \beta_2 \ln K_n + \beta_3 \ln L_n + U_n,$$

donde $\beta_1 = \ln c$, y $U_n = \ln \nu_n$ (es decir, $\nu_n = e^{U_n}$.)

En este caso, un valor como $\beta_2 = 0.05$ es interpretado como que un incremento de capital del 1 % (0.01) aumenta la producción en un 5 % (0.05).

Nota 1. Definimos la esperanza de una matriz $\mathbf{X}$ como la matriz de las esperanzas de sus elementos, es decir

$$\mathrm{E}(\mathbf{X}) \equiv \mathrm{E}\left(\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NN} \end{bmatrix}\right) \equiv \begin{bmatrix} \mathrm{E}(x_{11}) & \mathrm{E}(x_{12}) & \cdots & \mathrm{E}(x_{1N}) \\ \mathrm{E}(x_{21}) & \mathrm{E}(x_{22}) & \cdots & \mathrm{E}(x_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(x_{N1}) & \mathrm{E}(x_{N2}) & \cdots & \mathrm{E}(x_{NN}) \end{bmatrix}$$

⇑ **Supuesto 2: Esperanza condicional de $\boldsymbol{U}$ – Estricta exogeneidad** 7

$$\boxed{\mathrm{E}(\boldsymbol{U} \mid \mathbf{x}) = \underset{[N\times 1]}{\mathbf{0}}}$$

es decir

$$\mathrm{E}(\boldsymbol{U} \mid \mathbf{x}) = \begin{bmatrix} \mathrm{E}(U_1 \mid \mathbf{x}) \\ \mathrm{E}(U_2 \mid \mathbf{x}) \\ \vdots \\ \mathrm{E}(U_N \mid \mathbf{x}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathrm{E}(U_n \mid \mathbf{x}) \equiv \mathrm{E}(U_n \mid \boldsymbol{x}_{\blacktriangledown 2},\ \ldots,\ \boldsymbol{x}_{\blacktriangledown k}) \equiv \mathrm{E}(U_n \mid \boldsymbol{x}_{1\triangleright};\ \ldots;\ \boldsymbol{x}_{N\triangleright})$$

para $n = 1, \ldots, N$.

$$\mathrm{E}(U_n \mid \mathbf{x}) \equiv \mathrm{E}(U_n \mid \boldsymbol{x}_{\blacktriangledown 2},\ \ldots,\ \boldsymbol{x}_{\blacktriangledown k}) \equiv \mathrm{E}\left(U_n \mid \begin{bmatrix} \boldsymbol{x}_{1\triangleright} \\ \vdots \\ \boldsymbol{x}_{N\triangleright} \end{bmatrix}\right)$$

para $n = 1, \ldots, N$.

Ejemplo 4. [**función de consumo:** (continuación del Ejemplo 1 en la página anterior)]

Estricta exogeneidad implica que para el individuo n-ésimo

$$\mathrm{E}(U_n \mid \mathbf{1},\ \boldsymbol{rd}) = \mathrm{E}(U_n \mid (rd_2,\ rd_3,\ \cdots,\ rd_k)) = 0,$$

es decir, la esperanza de la perturbación n-ésima, condicionada a todas y cada una de las rentas disponibles, es cero.

Ejemplo 5. [**ecuación de salarios:** (continuación del Ejemplo 2 en la página anterior)]

Estricta exogeneidad implica que para el individuo n-ésimo

$$\mathrm{E}(U_n \mid \mathbf{1},\ \boldsymbol{educ},\ \boldsymbol{antig},\ \boldsymbol{exper}) = 0,$$

es decir, la esperanza de la perturbación del individuo n-ésimo, condicionada —no solo a los años de educación, antigüedad y experiencia de dicho individuo sino a los años de *educación, antigüedad y experiencia de todos los trabajadores*— es cero.

Supuesto 2: Esperanza condicional de $\boldsymbol{U}$ – Estricta exogeneidad 8

$$\boxed{\mathrm{E}(\boldsymbol{U} \mid \mathbf{x}) = \underset{[N\times 1]}{\mathbf{0}}} \Rightarrow \begin{cases} \mathrm{E}(U_n\,\mathbf{X}) = \mathbf{0} \quad \text{ortogonalidad } \; U_n \perp \mathbf{X} \\ \mathrm{E}(U_n) = 0 \\ \text{por tanto } \; \mathrm{Cov}(U_n, \mathbf{X}) = 0 \end{cases}$$

(ortogonalidad entre lo que conozco $\mathbf{X}$ y lo que desconozco U_n)

Comentario. En el caso de regresión con datos temporales, la exogeneidad estricta implica que los regresores son ortogonales a las perturbaciones *pasadas, presentes y futuras*. Esta es una restricción muy fuerte, que no se cumple en general con datos temporales (se discutirá en el segundo trimestre [Econometría II]).

A continuación aparecen las demostraciones de la transparencia anterior $\boxed{\text{T8}}$:

Proposición 2.1. *Si* $\mathrm{E}(U_n \mid \mathbf{x}) = 0$, *entonces* $\mathrm{E}(U_n\,\mathbf{X}) = \underset{[N\times k]}{\mathbf{0}}$

Demostración.

$$\begin{aligned}
\mathrm{E}(U_n\,\mathbf{X}) &= \int \cdots \int u_n\,\mathbf{x}\, f(u_n, \mathbf{x})\, du_n\; dx_{kN} \cdots dx_{11} \\
&= \int \cdots \int u_n\,\mathbf{x}\, f(u_n \mid \mathbf{x})\, f(\mathbf{x})\, du_n\; dx_{kN} \cdots dx_{11} \\
&= \int u_n \left[\int \cdots \int \mathbf{x}\, f(\mathbf{x})\, dx_{kN} \cdots dx_{11} \right] f(u_n \mid \mathbf{x})\, du_n \\
&= \int u_n\, [\mathrm{E}(\mathbf{X})]\, f(u_n \mid \mathbf{x})\, du_n \\
&= [\mathrm{E}(\mathbf{X})] \int u_n f(u_n \mid \mathbf{x})\, du_n \\
&= \mathrm{E}(\mathbf{X}) \cdot \mathrm{E}(U_n \mid \mathbf{x}) \\
&= \mathrm{E}(\mathbf{X}) \cdot 0 \; = \underset{[N\times k]}{\mathbf{0}} \qquad \text{por hipótesis}
\end{aligned}$$

□

Una importante implicación de $\mathrm{E}(U_n \mid \mathbf{x}) = 0$, es que entonces $\mathrm{E}(U_n) = 0$ ya que

$$\begin{aligned}
\mathrm{E}(U_n) &= \mathrm{E}(\mathrm{E}(U_n \mid \mathbf{x})) && \text{por el T}^{\text{a}} \text{ de las esperanzas iteradas.} \\
&= \mathrm{E}(0) = 0 && \text{por ser } \mathrm{E}(U_n \mid \mathbf{x}) \text{ las realizaciones de } \mathrm{E}(U_n \mid \mathbf{x})
\end{aligned}$$

Y de los dos resultados anteriores se deriva que

$$\mathrm{Cov}(U_n, \mathbf{X}) = \mathrm{E}(U_n\,\mathbf{X}) - \mathrm{E}(U_n) \cdot \mathrm{E}(X) = \underset{[N\times k]}{\mathbf{0}} - 0 \cdot \mathrm{E}(X) = \underset{[N\times k]}{\mathbf{0}}$$

EJERCICIO 6. **[Relación *si y sólo si* entre la función de regresión lineal y los supuestos 1 y 2]** Demuestre que los supuestos 1 y 2 implican la primera condición del *Modelo Clásico de Regresión Lineal*, esto es, que la función de regresión de Y_n sobre los regresores es lineal

$$\mathrm{E}(Y_n \mid \mathbf{x}) = \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta}\,.$$

Recíprocamente, demuestre que si dicha condición se verifica para todo $n = 1, \ldots, N$, entonces necesariamente se satisfacen los supuestos 1 y 2.

Solución:

$$\begin{aligned}
\mathrm{E}(Y_n \mid \mathbf{x}) &= \mathrm{E}(\boldsymbol{X}_{n\triangleright}\,\boldsymbol{\beta} + U_n \mid \mathbf{x}) && \text{por el Supuesto 1}\\
&= \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta} + \mathrm{E}(U_n \mid \mathbf{x}) && \text{puesto que } \boldsymbol{X}_{n\triangleright} = \boldsymbol{x}_{n\triangleright}\\
&= \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta} && \text{por el Supuesto 2.}
\end{aligned}$$

Recíprocamente, suponga que $\mathrm{E}(Y_n \mid \mathbf{x}) = \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta}$ para todo $n = 1, \ldots, N$. Definamos $U_n = Y_n - \mathrm{E}(Y_n \mid \mathbf{x})$. Entonces, por construcción el Supuesto 1 se satisface ya que $U_n = Y_n - \boldsymbol{X}_{n\triangleright}\,\boldsymbol{\beta}$. Por otra parte

$$\begin{aligned}
\mathrm{E}(U_n \mid \mathbf{x}) &= \mathrm{E}(Y_n \mid \mathbf{x}) - \mathrm{E}(\mathrm{E}(Y_n \mid \mathbf{x}) \mid \mathbf{x}) && \text{por la definición que aquí damos a } U_n\\
&= 0;
\end{aligned}$$

pues $\mathrm{E}(\mathrm{E}(Y_n \mid \mathbf{x}) \mid \mathbf{x}) = \mathrm{E}(Y_n \mid \mathbf{x})$, ya que:

$$\begin{aligned}
\mathrm{E}(\mathrm{E}(Y_n \mid \mathbf{x}) \mid \mathbf{x}) &= \int \left[\int y_t f(U_n \mid \mathbf{x})\ du_n\right] f(U_n \mid \mathbf{x})\ du_n\\
&= \int \left[\int (U_n + \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta}) f(U_n \mid \mathbf{x})\ du_n\right] f(U_n \mid \mathbf{x})\ du_n\\
&= \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta} + \int \left[\int U_n f(U_n \mid \mathbf{x})\ du_n\right] f(U_n \mid \mathbf{x})\ du_n\\
&= \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta} + \mathrm{E}(\mathrm{E}(U_n \mid \mathbf{x}) \mid \mathbf{x})\\
&= \boldsymbol{x}_{n\triangleright}\,\boldsymbol{\beta} + \mathrm{E}(U_n \mid \mathbf{x}) \;=\; \mathrm{E}(\boldsymbol{X}_{n\triangleright}\,\boldsymbol{\beta} + U_n \mid \mathbf{x}) = \mathrm{E}(Y_n \mid \mathbf{x})
\end{aligned}$$

Ejercicio 6

⇑ **Supuesto 3: Perturbaciones esféricas** 9

- **homocedasticidad**

$$\mathrm{E}\left({U_n}^2 \mid \mathbf{x}\right) = \sigma^2 \qquad \text{para } n = 1, 2, \ldots, N$$

- **no autocorrelación**

$$\mathrm{E}(U_i U_j \mid \mathbf{x}) = 0 \quad \text{si } i \neq j \qquad \text{para } i, j = 1, 2, \ldots, N$$

Definición 1. Definimos la matriz de varianzas y covarianzas de un vector columna $\boldsymbol{Y}$ como

$$\mathrm{Var}(\boldsymbol{Y}) \equiv \mathrm{E}\left(\left(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\right)\left(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\right)'\right) \tag{2.1}$$

EJERCICIO 7. Demuestre que $\mathrm{Var}(\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\,\mathrm{E}(\boldsymbol{Y}')$.

Nota 2. Por tanto la matriz de varianzas y covarianzas de un vector columna $\boldsymbol{Y}$ es de la forma

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{Y}) &\equiv \mathrm{Var}\left(\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}\right) \equiv \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\,\mathrm{E}(\boldsymbol{Y}')\\
&= \begin{bmatrix} \mathrm{E}({Y_1}^2) & \mathrm{E}(Y_1Y_2) & \cdots & \mathrm{E}(Y_1Y_N) \\ & \mathrm{E}({Y_2}^2) & \cdots & \mathrm{E}(Y_2Y_N) \\ & & \ddots & \vdots \\ & & & \mathrm{E}({Y_N}^2) \end{bmatrix} - \begin{bmatrix} [\mathrm{E}(Y_1)]^2 & \mathrm{E}(Y_1)\mathrm{E}(Y_2) & \cdots & \mathrm{E}(Y_1)\mathrm{E}(Y_N) \\ & [\mathrm{E}(Y_2)]^2 & \cdots & \mathrm{E}(Y_2)\mathrm{E}(Y_N) \\ & & \ddots & \vdots \\ & & & [\mathrm{E}(Y_N)]^2 \end{bmatrix}\\
&= \begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1Y_2} & \cdots & \sigma_{Y_1Y_N} \\ & \sigma^2_{Y_2} & \cdots & \sigma_{Y_2Y_N} \\ & & \ddots & \vdots \\ & & & \sigma^2_{Y_N} \end{bmatrix}
\end{aligned}$$

Aplicando la definición de varianza al vector de perturbaciones, y teniendo en cuenta los dos supuestos anteriores, tenemos que la matriz de varianzas y covarianzas de las perturbaciones es

$$\begin{aligned}
\operatorname{Var}(\boldsymbol{U} \mid \mathbf{x}) =& \operatorname{E}\left(\boldsymbol{U}\boldsymbol{U}' \mid \mathbf{x}\right) - \operatorname{E}(\boldsymbol{U} \mid \mathbf{x})\operatorname{E}\left(\boldsymbol{U}' \mid \mathbf{x}\right)\\
=&\operatorname{E}\left(\begin{bmatrix}U_1\\ \vdots \\ U_N\end{bmatrix}\begin{bmatrix}U_1 & \cdots & U_N\end{bmatrix} \middle| \mathbf{x}\right) - \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}\begin{bmatrix}0 & \cdots & 0\end{bmatrix} && \text{por el Supuesto 2}\\
=&\begin{bmatrix}\operatorname{E}\left(U_1{}^2 \mid \mathbf{x}\right) & \operatorname{E}(U_1U_2 \mid \mathbf{x}) & \dots & \operatorname{E}(U_1U_N \mid \mathbf{x})\\ \operatorname{E}(U_2U_1 \mid \mathbf{x}) & \operatorname{E}\left(U_2{}^2 \mid \mathbf{x}\right) & \dots & \operatorname{E}(U_2U_N \mid \mathbf{x})\\ \vdots & \vdots & \ddots & \vdots\\ \operatorname{E}(U_NU_1 \mid \mathbf{x}) & \operatorname{E}(U_NU_2 \mid \mathbf{x}) & \dots & \operatorname{E}\left(U_N{}^2 \mid \mathbf{x}\right)\end{bmatrix} - \begin{bmatrix}0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix}\\
=&\begin{pmatrix}\sigma^2 & 0 & \dots & 0\\ 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \sigma^2\end{pmatrix} && \text{por el Supuesto 3}
\end{aligned}$$

⇑ **Supuestos 2 y 3: Implicación conjunta** **10**

$$\begin{aligned}
\operatorname{Var}(\boldsymbol{U} \mid \mathbf{x}) =& \begin{pmatrix}\operatorname{Var}(U_1 \mid \mathbf{x}) & \operatorname{Cov}(U_1,U_2 \mid \mathbf{x}) & \dots & \operatorname{Cov}(U_1,U_N \mid \mathbf{x})\\ \operatorname{Cov}(U_2,U_1 \mid \mathbf{x}) & \operatorname{Var}(U_2 \mid \mathbf{x}) & \dots & \operatorname{Cov}(U_2,U_N \mid \mathbf{x})\\ \vdots & \vdots & \ddots & \vdots\\ \operatorname{Cov}(U_N,U_1 \mid \mathbf{x}) & \operatorname{Cov}(U_N,U_2 \mid \mathbf{x}) & \dots & \operatorname{Var}(U_N \mid \mathbf{x})\end{pmatrix}\\
=&\begin{pmatrix}\sigma^2 & 0 & \dots & 0\\ 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \sigma^2\end{pmatrix} = \sigma^2 \underset{[N\times N]}{\mathbf{I}}
\end{aligned}$$

El supuesto de que la matriz de varianzas y covarianzas de la perturbaciones (condicionada a $\mathbf{x}$) es σ^2 veces la matriz identidad (estructura denominada *perturbaciones esféricas*)

$$\boldsymbol{\Sigma} = \begin{bmatrix}\sigma^2 & 0 & 0 & \dots & 0\\ 0 & \sigma^2 & 0 & \dots & 0\\ 0 & 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & \sigma^2\end{bmatrix}$$

es una restricción muy fuerte, ya que implica:

1. que la dispersión (la varianza) del efecto de término perturbación asociada a cada observación (o a cada instante, o a cada individuo, etc) es idéntica a la de las demás (no sabemos exactamente a que se debe la perturbación que afecta a cada Y_n pero la dispersión (incertidumbre) de ese efecto es idéntica para todos).

   Dicho de otra forma: las perturbaciones U_n son *hocedásticas*, ya que
   $$\operatorname{Var}(U_n \mid \mathbf{x}) = \sigma^2 \quad \text{para todo } n = 1:N.$$

2. que la covarianza entre las perturbaciones de observaciones distintas (o de instantes ,o individuos diferentes) es cero. Dicho de otra forma: las perturbaciones no tienen *correlación serial*, ya que
   $$\operatorname{Cov}(U_i, U_j \mid \mathbf{x}) = 0 \qquad \text{para } i \neq j.$$

   Esto añadido al supuesto de distribución conjunta Normal (ver Supuesto 5 más adelante T31) significará que las perturbaciones son *independientes* para las distintas observaciones.

Ejemplo 8. [**ecuación de salarios:** (continuación del Ejemplo 2 en la página˜6)]

Estricta exogeneidad y perturbaciones esféricas implican conjuntamente que: aunque el factor desconocido U_n de cada el individuo n-ésimo *es desconocido*; la incertidumbre (la varianza) de dicho factor

—condicionada a los años de educación, antigüedad y experiencia de *todos* los individuos— es la misma en cada caso (¡Supuesto curioso! ¿no?).

Hay cierto factor que influye en los salarios de Pepito y Juanito; no sé qué es, pero la incertidumbre que tengo sobre él es la misma (la dispersión del efecto que tiene el factor desconocido es la misma) para ambos casos.

Nota 3 (Relación entre la función cedástica contante y los supuestos 1 y 3). Nótese que con los supuestos 1 y 3 también se cumple la segunda condición del modelo clásico de regresión lineal ya que

$$\operatorname{Var}(Y_n \mid \boldsymbol{x}) = \operatorname{Var}(\beta_1 + \beta_2 X_n + U_n \mid \boldsymbol{x}) = \operatorname{Var}(U_n \mid \boldsymbol{x}) = \sigma^2$$

2.2. Variación de los supuestos 2 y 3 en algunos casos especiales:

2.2.1. Supuestos del Modelo con Muestras Aleatorias

Si $(\boldsymbol{Y}, \mathbf{X})$ es una muestra aleatoria simple, i.e.. $\{Y_n, \boldsymbol{X}_{n\triangleright}\}$ es i.i.d. para $n = 1, \dots, N$; entonces,

$$\begin{aligned} \mathrm{E}(U_n \mid \mathbf{x}) &= \mathrm{E}(U_n \mid \boldsymbol{x}_{n\triangleright}) \\ \mathrm{E}\left(U_n{}^2 \mid \mathbf{x}\right) &= \mathrm{E}\left(U_n{}^2 \mid \boldsymbol{x}_{n\triangleright}\right) \\ \text{y también } \mathrm{E}(U_iU_j \mid \mathbf{x}) &= \mathrm{E}(U_i \mid \boldsymbol{x}_{i\triangleright})\,\mathrm{E}(U_j \mid \boldsymbol{x}_{j\triangleright}) \qquad \text{para } i \neq j \end{aligned}$$

Con lo que los los supuestos 2 $\boxed{\text{T7}}$ y 3 $\boxed{\text{T9}}$ quedan reducidos a

supuesto 2': $\mathrm{E}(U_n \mid \boldsymbol{x}_{n\triangleright}) = 0$

supuesto 3': $\mathrm{E}\left(U_n{}^2 \mid \boldsymbol{x}_{n\triangleright}\right) = \sigma^2 > 0$

para todo $n = 1, \dots, N$

(Nótese que los regresores están referidos exclusivamente a la observación n-ésima)

En general *este supuesto no es adecuado para modelos con datos de series temporales* ya que las muestras no son i.i.d. (no son muestras aleatorias simples puesto que suele haber correlación entre los datos).

Ejemplo 9. [**ecuación de salarios:** (continuación del Ejemplo 2 en la página~6)]

Con muestras aleatorias, estricta exogeneidad implica que para el individuo n-ésimo

$$\mathrm{E}(U_n \mid \mathbf{1},\ \boldsymbol{educ},\ \boldsymbol{antig},\ \boldsymbol{exper}) = \mathrm{E}(U_n \mid 1,\ educ_n,\ antig_n,\ exper_n) = 0,$$

es decir, la esperanza de la perturbación del individuo n-ésimo, condicionada —exclusivamente a los años de educación, antigüedad y experiencia de dicho individuo— es cero, independientemente de lo que ocurra con el resto de trabajadores. Por supuesto, también ocurre con la varianza condicionada:

$$\operatorname{Var}(U_n \mid \mathbf{1},\ \boldsymbol{educ},\ \boldsymbol{antig},\ \boldsymbol{exper}) = \operatorname{Var}(U_n \mid 1,\ educ_n,\ antig_n,\ exper_n) = \sigma^2\,\mathbf{I},$$

EJERCICIO 10. Demuestre que

$$\mathrm{E}(U_iU_j \mid \mathbf{x}) = \mathrm{E}(U_i \mid \boldsymbol{x}_{i\triangleright})\,\mathrm{E}(U_j \mid \boldsymbol{x}_{j\triangleright}) \qquad \text{para } i \neq j$$

para el caso de **muestras aleatorias simples** ($m.a.s.$)

Pista.

$$\mathrm{E}(U_iU_j \mid \mathbf{x}) = \mathrm{E}(\mathrm{E}(U_i \mid \mathbf{X}\, U_j) \cdot U_j \mid \mathbf{x})$$

debido a que $\{U_i, \boldsymbol{X}_{i\triangleright}\}$ es independiente de $\{U_j, \boldsymbol{X}_{1\triangleright}, \dots, \boldsymbol{X}_{i-1\triangleright},\ \boldsymbol{X}_{i+1\triangleright}, \dots, \boldsymbol{X}_{N\triangleright}\}$ para $i \neq j$, junto con el teorema de las esperanzas iteradas.

2.2.2. Supuestos del Modelo con Regresores No Estocásticos

Si los regresores son no estocásticos, es decir son la matriz determinista $\mathbf{x}$, entonces no es necesario distinguir entre funciones de densidad condicionales, $f\left(u_n \mid \mathbf{x}\right)$, e incondicionales, $f\left(u_n\right)$; por tanto los supuestos 2 $\boxed{\text{T7}}$ y 3 $\boxed{\text{T9}}$ quedan reducidos a

supuesto 2": $\mathrm{E}(U_n) = 0$

supuesto 3": $\mathrm{E}\big({U_n}^2\big) = \sigma^2 > 0 \quad$ y $\quad \mathrm{E}\big(U_iU_j\big) = 0 \quad$ para $i \neq j$

para todo $n, i, j = 1, \ldots, N$

(Estos son los supuestos empleados en la mayoría de libros de texto, como por ejemplo en Novales (1993))

Este caso no puede suponerse con modelos autorregresivos o de ecuaciones simultáneas.

La interpretación geométrica de estos supuestos aparece en la Sección A en la página˜46 del Apéndice.

Queda un cuarto supuesto acerca del rango de la matriz de regresores y un quinto supuesto acerca de la distribución conjunta de $\boldsymbol{U}$ que enunciaremos más adelante (véase Supuesto 4 $\boxed{\text{T13}}$ y Supuesto 5 $\boxed{\text{T31}}$)

3. Estimación MCO (Mínimos Cuadrados Ordinarios)

- Capítulos 2 y 3 de Wooldridge (2006)
- Apéndice E1 de Wooldridge (2006)

⇑ **Término de error** **11**

Las perturbaciones U_n no son observables
Pero las podemos estimar para un hipotético valor $\widetilde{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ y una muestra concreta $\{y_n,\ \boldsymbol{x}_{n\triangleright}\}_{n=1}^N$ de $\{Y_n,\ \boldsymbol{X}_{n\triangleright}\}_{n=1}^N$.

$$\widetilde{e}_n \;=\; y_n - \boldsymbol{x}_{n\triangleright}\widetilde{\boldsymbol{\beta}} \;=\; y_n - \widetilde{y_n}$$

Consideremos la Suma de los Residuos al Cuadrado para todo n

$$SRC(\widetilde{\boldsymbol{\beta}}) \equiv \sum_{n=1}^{N}\left(y_n - \boldsymbol{x}_{n\triangleright}\widetilde{\boldsymbol{\beta}}\right)^2 = (\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) = \widetilde{\boldsymbol{e}}'\,\widetilde{\boldsymbol{e}}$$

⇑ **Mínimos cuadrados ordinarios: Ecuaciones normales** **12**

El Supuesto 2 del modelo implica que $\quad U_n \perp \mathbf{X} \quad$ (ortogonalidad).
La $SRC(\widetilde{\boldsymbol{\beta}})$ es mínima para valores $\widehat{\boldsymbol{\beta}}$ tales que los errores

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}}$$

son ortogonales a los regresores de la muestra $\mathbf{x}$

$$\widehat{\boldsymbol{e}} \perp \mathbf{x} \quad \Rightarrow \quad \mathbf{x}'\,\widehat{\boldsymbol{e}} = \mathbf{0}\,.$$

Así

$$\mathbf{x}'\,\widehat{\boldsymbol{e}} = \mathbf{0}; \quad \Rightarrow \quad \mathbf{x}'\left(\boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}}\right) = \mathbf{0}; \quad \Rightarrow \quad \mathbf{x}'\,\boldsymbol{y} - \mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}} = \mathbf{0}$$

es decir

$$\boxed{\mathbf{x}'\,\boldsymbol{y} = \mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}}} \tag{3.1}$$

$\boxed{\text{Estimación MCO es la solución } \widehat{\boldsymbol{\beta}} \text{ a dichas ecuaciones}}$

Proposición 3.1. *La suma de residuos al cuadrado* $SRC(\widetilde{\boldsymbol{\beta}})$ *es mínima para* $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$.

Demostración. Sea $\widetilde{\boldsymbol{\beta}}$ una estimación de $\boldsymbol{\beta}$, entonces

$$\begin{aligned}\widetilde{\boldsymbol{e}}'\,\widetilde{\boldsymbol{e}} = (\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) =& (\boldsymbol{y}-\mathbf{x}\widehat{\boldsymbol{\beta}}+\mathbf{x}\widehat{\boldsymbol{\beta}}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widehat{\boldsymbol{\beta}}+\mathbf{x}\widehat{\boldsymbol{\beta}}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) \qquad \text{sumando y restando } \mathbf{x}\widehat{\boldsymbol{\beta}}\\ =& \Big(\widehat{\boldsymbol{e}}+\mathbf{x}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}})\Big)'\Big(\widehat{\boldsymbol{e}}+\mathbf{x}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}})\Big)\\ =& \widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}+(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}})'\,\mathbf{x}'\,\mathbf{x}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}) \qquad \text{ya que } \mathbf{x}'\,\widetilde{\boldsymbol{e}}=\mathbf{0}\,.\end{aligned}$$

Y puesto que $(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}})'\,\mathbf{x}'\,\mathbf{x}(\widehat{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}})$ es una suma de cuadrados (por tanto semi-definido positivo), se deduce que

$$SRC(\text{cualquier } \widetilde{\boldsymbol{\beta}}) \;=\; \widetilde{\boldsymbol{e}}'\,\widetilde{\boldsymbol{e}} \;\geq\; \widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} \;=\; SRC(\widehat{\boldsymbol{\beta}}).$$

□

Para una interpretación geométrica, véase también la Sección A.1 en la página˜47 del apéndice.

La demostración anterior es, para mi gusto, más elegante que la que aparece en la mayoría de los manuales (búsqueda del mínimo de la suma residual igualando a cero las primeras derivadas). No obstante, en la Sección B en la página˜48 del apéndice se muestra la derivación tradicional de las ecuaciones normales.

Para que la solución al sistema de ecuaciones normales (3.1) sea única es necesario que se cumpla un cuarto supuesto.

3.1. Cuarto supuesto del Modelo Clásico de Regresión Lineal

⇑ **Supuesto 4: Independencia lineal de los regresores** **13**

El rango de $\underset{[N\times k]}{\mathbf{X}}$ es k con probabilidad 1.

- número de observaciones $\geq k$
- Vectores columna $\mathbf{1}$, $\boldsymbol{X}_{\blacktriangledown 2}$, $\dots$, $\boldsymbol{X}_{\blacktriangledown k}$ linealmente indep.

Este supuesto implica que $\mathbf{x}'\,\mathbf{x}$ es de rango completo, es decir, que existe la matriz $(\mathbf{x}'\,\mathbf{x})^{-1}$.

Se dice que existe *multicolinealidad perfecta* cuando el Supuesto 4 NO se satisface; es decir, cuando hay dependencia lineal entre los regresores, o lo que es lo mismo: *hay multicolinealidad perfecta cuando alguno de los coeficientes de correlación lineal entre dos regresores es uno en valor absoluto.*

El Supuesto 4 garantiza la unicidad de las soluciones. Si no se cumple no es posible encontrar *"la estimación"* MCO de los parámetros (pues hay infinitas soluciones posibles).

Ejemplo 11. [**ecuación de salarios:** (continuación del Ejemplo 2 en la página˜6)]

¿Que pasa si todos los individuos de la muestra nunca han cambiado de empresa?

Entonces *años de experiencia* y *años de antigüedad* coinciden. Por tanto no es posible discriminar el efecto por separado de ambas variables; sólo podemos calcular su efecto conjunto.

$$\ln(SALAR_n) = \beta_1 + \beta_2 EDUC_n + (\beta_3+\beta_4)EXPER_n + U_n$$

Volveremos sobre esto en la Sección 3 sobre Multicolinealidad en la página˜8 del Tema 3

3.2. Algunas expresiones que serán empleadas frecuentemente

Las expresiones que aparecen a continuación serán empleadas repetidamente durante el curso.

Denotamos a la media aritmética de los elementos del vector $\boldsymbol{y}$ de orden N como:

$$\overline{\boldsymbol{y}} = (\sum y_n)/N.$$

Nota 4. Sean $\boldsymbol{x}$ e $\boldsymbol{y}$ vectores de orden N, entonces

$$\sum_n (x_n - \overline{x})(y_n - \overline{y}) = \sum_n y_n(x_n - \overline{x}) \quad \text{para } n = 1, \dots, N.$$

Demostración.

$$\begin{aligned}\sum_n (x_n - \overline{x})(y_n - \overline{y}) &= \sum_n y_n(x_n - \overline{x}) - \overline{y}\sum_n (x_n - \overline{x}) \\ &= \sum_n y_n(x_n - \overline{x}) - \overline{y}\cdot 0 = \sum_n y_n(x_n - \overline{x}) \qquad \text{para } n = 1, \dots, N.\end{aligned}$$

□

Nota 5. Sean $\boldsymbol{x}$ e $\boldsymbol{y}$ vectores de orden N, entonces

$$\sum_n (x_n - \overline{x})(y_n - \overline{y}) \;=\; \sum_n y_n x_n - N\overline{y}\,\overline{x} \;=\; \boldsymbol{y}'\,\boldsymbol{x} - N\overline{y}\,\overline{x}.$$

EJERCICIO 12. Compruebe la igualdad de la nota anterior.

Así pues, del ejercicio anterior, $N s_{\boldsymbol{x}\,\boldsymbol{y}} = \sum_n (x_n - \overline{x})(y_n - \overline{y}) = \boldsymbol{y}'\,\boldsymbol{x} - N\overline{y}\,\overline{x}$, es decir

$$s_{\boldsymbol{x}\,\boldsymbol{y}} = \frac{\sum_n (x_n - \overline{x})(y_n - \overline{y})}{N} = \frac{\boldsymbol{y}'\,\boldsymbol{x}}{N} - \overline{y}\,\overline{x}; \tag{3.2}$$

donde $s_{\boldsymbol{x}\,\boldsymbol{y}}$ es la covarianza muestral entre los elementos de $\boldsymbol{x}$ e $\boldsymbol{y}$; por tanto, la expresión de más arriba es el análogo muestral de $\operatorname{Cov}(X, Y) = \operatorname{E}([X - \operatorname{E}(X)][Y - \operatorname{E}(Y)]) = \operatorname{E}(XY) - \operatorname{E}(X)\operatorname{E}(Y)$.

Nota 6. Sea $\boldsymbol{z}$ un vector de orden N, entonces $\sum_n (z_n - \overline{\boldsymbol{z}})^2 \;=\; \sum_n z_n^2 - N\overline{\boldsymbol{z}}^2 \;=\; \boldsymbol{z}'\,\boldsymbol{z} - N\overline{\boldsymbol{z}}^2$

Demostración. De la **Nota** 4 sabemos que $\sum_n (z_n - \overline{\boldsymbol{z}})(y_n - \overline{y}) = \sum_n y_n(z_n - \overline{\boldsymbol{z}})$, por tanto, si $\boldsymbol{y} = \boldsymbol{z}$

$$\begin{aligned}\sum_n (z_n - \overline{\boldsymbol{z}})^2 &= \sum_n z_n(z_n - \overline{\boldsymbol{z}}) \\ &= \sum_n z_n^2 - \overline{\boldsymbol{z}}\sum_n z_n \;=\; \sum_n z_n^2 - N\overline{\boldsymbol{z}}^2 \;=\; \boldsymbol{z}'\,\boldsymbol{z} - N\overline{\boldsymbol{z}}^2 \qquad \text{para } n = 1, \dots, N;\end{aligned}$$

□

Es decir,

$$s_{\boldsymbol{z}}^2 = \frac{\sum_n (z_n - \overline{\boldsymbol{z}})^2}{N} = \frac{\boldsymbol{z}'\,\boldsymbol{z}}{N} - \overline{\boldsymbol{z}}^2; \tag{3.3}$$

donde $s_{\boldsymbol{z}}^2$ es la varianza muestral de los elementos de $\boldsymbol{z}$; por tanto, la expresión anterior es el análogo muestral de $\operatorname{Var}(Z) = \operatorname{E}\big([Z - \operatorname{E}(Z)]^2\big) = \operatorname{E}\big(Z^2\big) - [\operatorname{E}(Z)]^2$.

3.3. Algunos casos particulares

3.3.1. Modelo con sólo una constante

⇑ **Modelo 1: No vbles explicativas** 14

"Si no sé nada ($\mathcal{D}: \emptyset$)" ; $\boldsymbol{Y} = h(\mathbf{1}) + \boldsymbol{U}$ donde $g(\cdot)$ es lineal; por lo tanto

$$Y_n = a \cdot 1 + U_n$$

$\operatorname{E}(Y_n \,|\text{conjunto de información vacío}) = \operatorname{E}(Y_n) = a$
Veamos que nos da la estimación MCO

$$\mathbf{x}'\,\boldsymbol{y} = \mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}}$$

es decir

$$\mathbf{1}'\,\boldsymbol{y} = \mathbf{1}'\,\mathbf{1}\,\widehat{a}$$

y calculando los productos escalares,

$$\sum y_n = N\,\widehat{a}; \qquad \Rightarrow \qquad \widehat{a} = \frac{\sum y_n}{N} = \overline{y} \tag{3.4}$$

Nótese como la estimación MCO consiste en sustituir el momentos teórico $\mathrm{E}(Y_n)$ por su análogo muestral (la media aritmética).

En este caso los residuos del modelo son las deviaciones de los datos respecto a su media, ya que

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \overline{y}. \tag{3.5}$$

3.3.2. Modelo Lineal Simple

⇑ **Modelo 2: Modelo Lineal Simple** **15**

"Si $(\mathcal{D} : \boldsymbol{X}_{\blacktriangledown} = \boldsymbol{x}_{\blacktriangledown})$*"* ; $\boldsymbol{Y} = h(\mathbf{1}, \boldsymbol{X}_{\blacktriangledown}) + \boldsymbol{U}$ donde $g(\cdot)$ es lineal; por lo tanto

$$Y_n = a + bX_n + U_n;$$

entonces

$$\begin{aligned}\mathrm{E}(Y_n \mid x_n) =& \mathrm{E}(a + bX_n + U_n \mid x_n)\\ =& a + bx_n + \mathrm{E}(U_n \mid x_n) \qquad = a + bx_n.\end{aligned}$$

Por lo tanto, es función lineal y

$$\mathrm{E}(Y_n \mid x_n) = \underbrace{\mathrm{E}(Y) - \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}\mathrm{E}(X)}_{a} + \underbrace{\frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}}_{b} \cdot x_n; \tag{3.6}$$

para todo $x_n \in \mathbb{R}_X$,

Véanse las ecuaciones (**??**) y (**??**) Sección **??** (**??**) del Tema 2 del curso de Introducción a la Econometría de LECO, página **??**.

⇑ **Modelo 2: Modelo Lineal Simple** **16**

Sea $Y_n = a + bX_n + U_n$; entonces

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}; \quad \mathbf{x} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}; \quad \widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix}$$

y loas ecuaciones normales son

$$\mathbf{x}'\boldsymbol{y} = \mathbf{x}'\mathbf{x}\widehat{\boldsymbol{\beta}}$$

es decir

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix}$$

⇑ **Modelo 2: Modelo Lineal Simple** **17**

$$\begin{aligned} \sum y_n &= \widehat{a}\, N + \widehat{b} \sum x_n \\ \sum x_n y_n &= \widehat{a} \sum x_n + \widehat{b} \sum x_n^2 \end{aligned}; \tag{3.7}$$

dividiendo por N la primera igualdad, despejando $\widehat{a}$ y sustituyendo en la segunda, y empleando (3.2) y (3.3)

$$\begin{aligned} \overline{y} &= \widehat{a} + \widehat{b}\,\overline{x} \\ s_{\boldsymbol{x}\,\boldsymbol{y}} &= \widehat{b} s_{\boldsymbol{x}}^2 \end{aligned} \tag{3.8}$$

es decir

$$\widehat{b} = \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s_{\boldsymbol{x}}^2} \tag{3.9}$$

y

$$\widehat{a} = \overline{y} - \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s_{\boldsymbol{x}}^2}\,\overline{x} = \overline{y} - \widehat{b}\,\overline{x} \tag{3.10}$$

Supuesto 4 (independencia lineal de regresores) ⇒ solución única.

Nótese como las estimaciones MCO consisten en sustituir los momentos teóricos de la Ecuación (3.6) por sus análogos muestrales.

Ejercicio 13. Empleando el sistema de ecuaciones (3.7), obtenga el segundo sistema (3.8) de la transparencia anterior.

Ejercicio 14. ¿Cómo afectaría al problema de estimación que la variable $\boldsymbol{x}$ fuera un vector de constantes $\boldsymbol{c}$?

Ejemplo 15. [**precio de las viviendas:**]

n	Precio	Superficie
1	199.9	1065
2	228.0	1254
3	235.0	1300
4	285.0	1577
5	239.0	1600
6	293.0	1750
7	285.0	1800
8	365.0	1870
9	295.0	1935
10	290.0	1948
11	385.0	2254
12	505.0	2600
13	425.0	2800
14	415.0	3000

Cuadro 1: Superficie (en pies al cuadrado) y precio de venta de los pisos (en miles de dólares) (Ramanathan, 1998, pp. 78)

Planteamos el modelo $Y_n = a + bX_n + U_n$, donde Y_n es el precio del piso n-ésimo, X_n es su superficie, y U_n son otros factores que influyen en el precio del piso, pero "ortogonales" al la superficie del mismo (situación, estado de mantenimiento, servicios, etc.) Deseamos saber cual es el efecto *marginal* del incremento de la superficie de un piso en su precio. Por lo tanto necesitamos estimar el valor del parámetro b.

Puesto que

$$\sum_n x_n = 26\,753 \qquad \sum_n x_n^2 = 55\,462\,515$$

$$\sum_n y_n = 4\,444.9 \qquad \sum_n x_n y_n = 9\,095\,985.5$$

De 3.7 en la página anterior tenemos el sistema de ecuaciones lineales

$$\begin{array}{rcccc} 4\,444.9 & = & \widehat{a} \cdot 14 & + & \widehat{b} \cdot 26\,753 \\ 9\,095\,985.5 & = & \widehat{a} \cdot 26\,753 & + & \widehat{b} \cdot 55\,462\,515 \end{array}$$

cuya solución nos da la estimación por mínimos cuadrados de a y b:

$$\widehat{a} = 52.3509 \qquad \widehat{b} = 0.13875;$$

que también podemos calcular a partir de (3.9) y (3.10) en la página anterior

$$\widehat{a} = \overline{y} - \overline{x}\frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s_{\boldsymbol{x}}^2} = 52.3509 \qquad \widehat{b} = \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s_{\boldsymbol{x}}^2} = 0.13875$$

Estimaciones MCO utilizando las 14 observaciones 1–14
Variable dependiente: price

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	52,3509	37,2855	1,4041	0,1857
sqft	0,138750	0,0187329	7,4068	0,0000

Media de la var. dependiente	317,493
D.T. de la variable dependiente	88,4982
Suma de cuadrados de los residuos	18273,6
Desviación típica de los residuos ($\hat{\sigma}$)	39,0230
R^2	0,820522
$\bar{R}^2$ corregido	0,805565
Grados de libertad	12
Criterio de información de Akaike	144,168
Criterio de información Bayesiano de Schwarz	145,447

Salida del programa "libre" Gretl (Gnu Regression, Econometrics and Time-series Library)

$$\widehat{\text{price}} = \underset{(1,404)}{52,3509} + \underset{(7,407)}{0,138750}\,\text{sqft}$$

$$N = 14 \quad \bar{R}^2 = 0,8056 \quad F(1,12) = 54,861 \quad \hat{\sigma} = 39,023$$

(entre paréntesis, los estadísticos t)

Por lo tanto, el precio de venta *esperado* de un piso con una superficie de 1800 pies cuadrados, $\mathrm{E}(\,Y\,|\,1800)$, será de

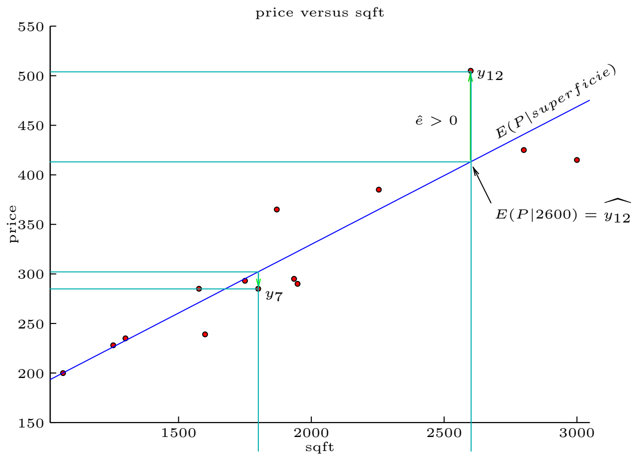
$$\widehat{y_7} = 52.3509 + 0.139 \cdot 1800 = 302101.5$$

sin embargo $y_7 = 285$. Esta discrepancia (el error $\widehat{e_7}$ puede deberse a que dicho piso esta en una mala situación, dispone de pocos servicios, etc.)

n	Precio	Superficie	Precio estimado $\mathrm{E}(\,P\,|\,\text{superficie})$	Error $\widehat{e}$
1	199.9	1065	200.1200	-0.22000
2	228.0	1254	226.3438	1.65619
3	235.0	1300	232.7263	2.27368
4	285.0	1577	271.1602	13.83984
5	239.0	1600	274.3514	-35.35142
6	293.0	1750	295.1640	-2.16397
7	285.0	1800	302.1015	-17.10148
8	365.0	1870	311.8140	53.18600
9	295.0	1935	320.8328	-25.83278
10	290.0	1948	322.6365	-32.63653
11	385.0	2254	365.0941	19.90587
12	505.0	2600	413.1017	91.89826
13	425.0	2800	440.8518	-15.85180
14	415.0	3000	468.6019	-53.60187

Cuadro 2: Superficie (en pies al cuadrado), precio de venta (en miles de dólares), precio estimado, y errores estimados.

⇑ **Estimación MCO: Interpretación gráfica** **18**



"regresión a ojo" GNU Gretl (este ejemplo) "data list"

Continuación del ejemplo "precio de las viviendas" en la página 34

3.3.3. Modelo con tres regresores

Ejercicio 16. Repita los pasos dados en la transparencia $\boxed{\text{T16}}$ y llegue hasta el sistema de ecuaciones equivalente a (3.7 en la página˜15) para los siguientes modelos:
(a) $Y_n = aX_{1n} + bX_{2n} + cX_{3n} + U_n$
(b) $Y_n = a + bX_{2n} + cX_{3n} + U_n$

Ejercicio 17. Obtenga la siguiente solución del segundo sistema de ecuaciones del ejercicio anterior.

$$\widehat{a} = \overline{y} - \widehat{b} \cdot \overline{\boldsymbol{x_2}} - \widehat{c} \cdot \overline{\boldsymbol{x_3}} \tag{3.11}$$

$$\widehat{b} = \frac{s_{\boldsymbol{x_2}\,\boldsymbol{y}} \cdot s^2_{\boldsymbol{x_3}} - s_{\boldsymbol{x_3}\,\boldsymbol{y}} \cdot s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}}{s^2_{\boldsymbol{x_2}} \cdot s^2_{\boldsymbol{x_3}} - \left(s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}\right)^2} \tag{3.12}$$

$$\widehat{c} = \frac{s_{\boldsymbol{x_3}\,\boldsymbol{y}} \cdot s^2_{\boldsymbol{x_2}} - s_{\boldsymbol{x_2}\,\boldsymbol{y}} \cdot s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}}{s^2_{\boldsymbol{x_2}} \cdot s^2_{\boldsymbol{x_3}} - \left(s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}\right)^2} \tag{3.13}$$

Nótese que si la covarianza entre $\boldsymbol{x_2}$ y $\boldsymbol{x_3}$ es cero, la estimación de $\widehat{b}$ del modelo $Y_n = a + bX_{2n} + cX_{3n} + U_n$ coincide exactamente con la estimación de $\widehat{b}$ en el modelo restringido $Y_n = a + bX_{2n} + U_n$ en el que se ha quitado el regresor X_{3n}.

Ejercicio 18. Si la covarianza entre $\boldsymbol{x_2}$ y $\boldsymbol{x_3}$ es cero, ¿Con la estimación de qué modelo restringido coincide la estimación de $\widehat{c}$?

Nota 7. *Si los regresores de una regresión múltiple tienen correlación muestral cero entre si (por tanto son ortogonales), entonces las estimaciones de las pendientes de la regresión múltiple son las mismas que las estimaciones de las pendientes de las regresiones simples.*

Multicolinealidad perfecta: Ejercicio 19. ¿Cómo afectaría al problema de estimación que los regresores $\boldsymbol{x_2}$ y $\boldsymbol{x_3}$ tuvieran un coeficiente de correlación muestral con valor absoluto igual a uno?

Relación entre los modelos de tres regresores y los de dos. Considere los siguientes modelos de regresión simple

1. $\boldsymbol{Y} = a_{yx_2} + b_{yx_2}\, \boldsymbol{X_2} + \boldsymbol{U}$: Regresión de $\boldsymbol{Y}$ sobre $\boldsymbol{X_2}$
2. $\boldsymbol{Y} = a_{yx_3} + b_{yx_3}\, \boldsymbol{X_3} + \boldsymbol{U}^*$: Regresión de $\boldsymbol{Y}$ sobre $\boldsymbol{X_3}$
3. $\boldsymbol{X_2} = a_{x_2x_3} + b_{x_2x_3}\, \boldsymbol{X_3} + \boldsymbol{U}^{**}$: Regresión de $\boldsymbol{X_2}$ sobre $\boldsymbol{X_3}$

(Nótese como los subíndices de los coeficientes describen cada regresión)

¿Que relación tienen las estimaciones MCO de estos tres modelos con las estimaciones MCO del modelo

$$\boldsymbol{Y} = a + b\,\boldsymbol{X_2} + c\,\boldsymbol{X_3} + \boldsymbol{U} : \qquad \text{Regresión de } \boldsymbol{Y} \text{ sobre } \boldsymbol{X_2} \text{ y } \boldsymbol{X_3}$$

descritas en las ecuaciones (3.12) y (3.12)?

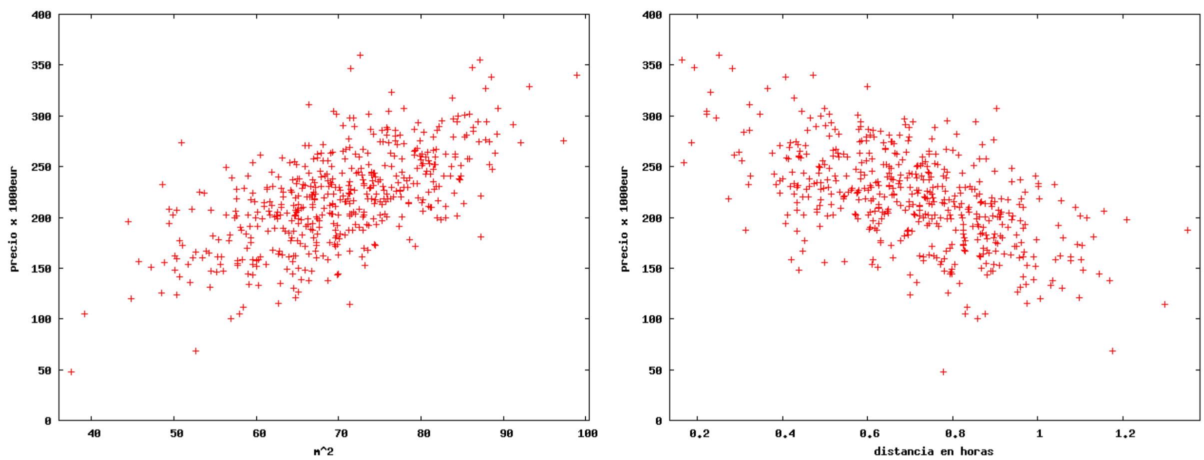
Si multiplicamos y dividimos (3.12) y (3.12) por $s^2_{\boldsymbol{x_2}} \cdot s^2_{\boldsymbol{x_3}}$ obtenemos las siguientes expresiones en términos de los coeficientes MCO de las tres regresiones anteriores:

$$\widehat{b} = \frac{\widehat{b_{yx_2}} - \widehat{b_{yx_3}}\,\widehat{b_{x2x_3}}}{1 - r^2_{\boldsymbol{x_2x_3}}} \tag{3.14}$$

$$\widehat{c} = \frac{\widehat{b_{yx_3}} - \widehat{b_{yx_2}}\,\widehat{b_{x2x_3}}}{1 - r^2_{\boldsymbol{x_2x_3}}} \tag{3.15}$$

donde $r_{\boldsymbol{x_2x_3}}$ es la correlación muestral entre ambos regresores.

Modelo simulado $P_n = 100 + 3S_n - 130D_n + U_n$



Modelo simulado $P_n = 100 + 3S_n - 130D_n + U_n$

Modelo 1 $P_n = \beta_1 + \beta_2 S_n + U_n$

Modelo 1: estimaciones MCO utilizando las 500 observaciones 1–500
Variable dependiente: precio

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	8,86429	11,7399	0,7551	0,4506
superfic	2,99968	0,166441	18,0225	0,0000

Media de la var. dependiente	218,374
D.T. de la variable dependiente	47,0678
Suma de cuadrados de los residuos	669080,
Desviación típica de los residuos ($\hat{\sigma}$)	36,6542
R^2	0,394756
$\bar{R}^2$ corregido	0,393541
Grados de libertad	498
Criterio de información de Akaike	5022,46
Criterio de información Bayesiano de Schwarz	5030,89

Modelo simulado $P_n = 100 + 3S_n - 130D_n + U_n$

Modelo 2 $P_n = \beta_1 + \beta_2 D_n + U_n$

Modelo 2: estimaciones MCO utilizando las 500 observaciones 1–500
Variable dependiente: precio

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	310,482	6,32078	49,1208	0,0000
distanci	−130,54	8,61143	−15,1599	0,0000

Media de la var. dependiente	218,374
D.T. de la variable dependiente	47,0678
Suma de cuadrados de los residuos	756399,
Desviación típica de los residuos ($\hat{\sigma}$)	38,9727
R^2	0,315768
$\bar{R}^2$ corregido	0,314394
Grados de libertad	498
Criterio de información de Akaike	5083,80
Criterio de información Bayesiano de Schwarz	5092,23

Modelo simulado: $P_n = 100 + 3S_n - 130D_n + U_n$

Modelo 3 $P_n = \beta_1 + \beta_2 S_n + \beta_3 D_n + U_n$

Modelo 3: estimaciones MCO utilizando las 500 observaciones 1–500
Variable dependiente: precio

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	98,9950	8,70328	11,3744	0,0000
superfic	3,06214	0,111940	27,3553	0,0000
distanci	−133,93	5,44707	−24,5876	0,0000

Media de la var. dependiente	218,374
D.T. de la variable dependiente	47,0678
Suma de cuadrados de los residuos	301877,
Desviación típica de los residuos ($\hat{\sigma}$)	24,6454
R^2	0,726925
$\bar{R}^2$ corregido	0,725826
$F(2, 497)$	661,506
Criterio de información de Akaike	4626,52
Criterio de información Bayesiano de Schwarz	4639,17

EJERCICIO 20. ¿Coinciden los valores estimados para los parámetros β_2 y β_3 en el modelo $P_n = \beta_1 + \beta_2 S_n - \beta_3 D_n + U_n$ con los valores obtenidos para las pendientes en los modelos restringidos anteriores? ¿Qué podemos afirmar entonces sobre la covarianza muestral de los regresores *distancia* y *superficie*?

3.3.4. Modelo Lineal General

⇑ Modelo Lineal General 19

En general tenemos más de una variable exógena por lo que "$(\mathcal{D} : \mathbf{X} = \mathbf{x})$";

$$Y_n = \boldsymbol{X}_{n\triangleright}\boldsymbol{\beta} + U_n = \begin{bmatrix}1, & X_{n2}, & \dots, X_{nk}\end{bmatrix} \underset{[k \times 1]}{\boldsymbol{\beta}} + U_n;$$

entonces

$$\begin{aligned} \mathrm{E}(Y_n \mid \boldsymbol{x}_{n\triangleright}) = & \ \mathrm{E}\Big(\begin{bmatrix}1, & X_{n2}, & \dots, X_{nk}\end{bmatrix}\boldsymbol{\beta} + U_n \Bigm| \boldsymbol{x}_{n\triangleright}\Big) = \\ = & \ \mathrm{E}\Big(\begin{bmatrix}1, & x_{t2}, & \dots, x_{tk}\end{bmatrix}\boldsymbol{\beta} + U_n \Bigm| \boldsymbol{x}_{n\triangleright}\Big) = \\ = & \ \mathrm{E}(a_1 + a_2 x_{n2} + \dots + a_k x_{nk} + U_n \mid \boldsymbol{x}_{n\triangleright}) = \\ = & \ a_1 + a_2 x_{n2} + \dots + a_k x_{nk} + \mathrm{E}(U_n \mid \boldsymbol{x}_{n\triangleright}) \\ = & \ a_1 + a_2 x_{n2} + \dots + a_k x_{nk} = \boldsymbol{x}_{n\triangleright}\boldsymbol{\beta}; \end{aligned}$$

donde $\boldsymbol{x}_{n\triangleright} = (1, \ x_{n2}, \ \dots, x_{nk})$.
Necesitamos conocer el valor de los elementos de $\boldsymbol{\beta}$,

$$(a_1, \ a_2, \ \cdots, \ a_k).$$

que dependen de las varianzas y covarianzas de $\begin{bmatrix}Y_n, & \boldsymbol{X}_{n\triangleright}\end{bmatrix}$.
(Véase la Sección C.1 del apéndice)

La expresión general de las ecuaciones normales es

$$\mathbf{x}'\boldsymbol{y} = \mathbf{x}'\mathbf{x}\widehat{\boldsymbol{\beta}}.$$

El Supuesto 4 garantiza (con probabilidad 1) que la matriz $\mathbf{x}'\mathbf{x}$ es invertible. Por tanto la estimación MCO del vector $\boldsymbol{\beta}$ se puede expresar como

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y}.$$

(Véase la Sección D para una interpretación de esta expresión.)

4. Propiedades algebraicas de la estimación MCO

4.1. Propiedades básicas

- Capítulos 2 y 3 de Wooldridge (2006)
- Apéndice E1 de Wooldridge (2006)

⇑ Mínimos cuadrados ordinarios: Propiedades algebraicas 20

El vector de residuos evaluado en $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$ es

$$\underset{[N \times 1]}{\widehat{\mathbf{e}}} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}}$$

Reordenando las ecuaciones normales $\quad \mathbf{x}'\boldsymbol{y} = \mathbf{x}'\mathbf{x}\widehat{\boldsymbol{\beta}} \quad$ tenemos

$$\mathbf{x}'(\boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}}) = 0; \quad \Rightarrow \boxed{\mathbf{x}'\,\widehat{\boldsymbol{e}} = \mathbf{0}} \quad \Rightarrow \boxed{\widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{e}} = 0} \qquad (4.1)$$

La propiedad

$$\mathbf{x}'\,\widehat{\boldsymbol{e}} = \mathbf{0}$$

es el análogo muestral de las condiciones de ortogonalidad derivadas del Supuesto 2 $\boxed{\text{T8}}$ (recuérdese que dos vectores de números $\boldsymbol{a}$ y $\boldsymbol{b}$ son ortogonales si $\boldsymbol{a}'\boldsymbol{b} = \sum a_i b_i = 0$.)

Esta propiedad indica que el término de error estimado, $\widehat{\boldsymbol{e}}$, es ortogonal a todos y cada uno de los regresores.

Del mismo modo que hemos definido $\widehat{\boldsymbol{e}}$ como $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}}$, definimos los valores ajustados $\widehat{\boldsymbol{y}}$ como

$$\widehat{\boldsymbol{y}} = \mathbf{x}\widehat{\boldsymbol{\beta}};$$

entonces $\widehat{\boldsymbol{y}}' = \widehat{\boldsymbol{\beta}}'\mathbf{x}'$, y por tanto

$$\widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{\beta}}'\mathbf{x}'\,\widehat{\boldsymbol{e}} = \widehat{\boldsymbol{\beta}}'\,\mathbf{0} = 0.$$

Practica 21. Con algún programa econométrico estime un modelo del tipo

$$Y_n = \beta_1 + \beta_2 X_{n2} + \beta_3 X_{n3} + U_n.$$

Obtenga los residuos $\widehat{\boldsymbol{e}}$ y los valores ajustados $\widehat{\boldsymbol{y}}$. Compruebe que

$$\begin{aligned}\boldsymbol{x_1}'\widehat{\boldsymbol{e}} =&0\\ \boldsymbol{x_2}'\widehat{\boldsymbol{e}} =&0\\ \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} =&0\end{aligned}$$

Calcule los valores medios de $\widehat{\boldsymbol{e}}$, $\widehat{\boldsymbol{y}}$ e $\boldsymbol{y}$. Explique los resultados. Añadir script de Gretl

⇑ **Mínimos cuadrados ordinarios: Más propiedades algebraicas** **21**

$$\boxed{\boldsymbol{y}'\boldsymbol{y} \;=\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} \;+\; \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}} \quad (\text{T}^{\text{a}} \text{ Pitágoras } \boxed{\text{T46}}\,) \tag{4.2}$$

Ya que

$$\begin{aligned}\boldsymbol{y}'\boldsymbol{y} =& (\widehat{\mathbf{y}} + \widehat{\mathbf{e}})'(\widehat{\mathbf{y}} + \widehat{\mathbf{e}}) && \text{puesto que } \widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\mathbf{y}}\\ =&\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + 2\widehat{\mathbf{y}}'\widehat{\mathbf{e}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} && \text{desarrollando el producto}\\ =&\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \;\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} && \text{ya que de (4.1) } \widehat{\mathbf{y}}'\widehat{\mathbf{e}} = 0\end{aligned}$$

⇑ **Sumas de cuadrados** **22**

$$SRC \equiv \sum_{n=1}^{N} \widehat{e_n}^2 = \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}$$

$$STC \equiv \sum_{n=1}^{N} (y_n - \overline{y})^2 = \boldsymbol{y}'\boldsymbol{y} - N\overline{y}^2$$

$$SEC \equiv \sum_{n=1}^{N} (\widehat{y_n} - \overline{y})^2 = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + N\overline{y}^2 - 2N\overline{y}\overline{\widehat{y}}$$

Por tanto, $STC = Ns^2_{\boldsymbol{y}}$ donde $s^2_{\boldsymbol{y}}$ es la varianza muestral de $\boldsymbol{y}$; por el contrario, las sumas SRC y SEC no son necesariamente N veces las varianzas de $\widehat{\boldsymbol{e}}$ y $\widehat{\boldsymbol{y}}$ (aunque veremos que así ocurre si el modelo tiene término cte.).

Ejercicio 22. Verifique las igualdades de la transparencia anterior.

Caso especial (Modelos con término constante). *Cuando hay término constante en el modelo (el primer regresor es un vector de unos — tal y como hemos presentado el modelo aquí) se verifica que*

$$\mathbf{1}'\widehat{\boldsymbol{e}} = 0; \quad \Rightarrow \quad \sum_{n=1}^{N} \widehat{e_n} = 0 \quad \Rightarrow \quad \boxed{\overline{\widehat{e}} = 0}.$$

Y puesto que para cada n, se verifica que $y_n = \widehat{y_n} + \widehat{e_n}$, entonces sumando para $n = 1, \ldots, N$

$$\sum_{n=1}^{N} y_n = \sum_{n=1}^{N} \widehat{y_n} + 0 \quad \text{o bien} \quad \mathbf{1}'\boldsymbol{y} = \mathbf{1}'\widehat{\boldsymbol{y}} \quad \Rightarrow \quad \boxed{\overline{y} = \overline{\widehat{y}}}$$

Además, de (4.2)

$$\sum y_n^2 = \sum \widehat{y_n}^2 + \sum \widehat{e}^2;$$

restando a derecha e izquierda $N\overline{y}^2$ (que es igual a $N\overline{\widehat{y}}^2$),

$$\sum y_n^2 - N\overline{y}^2 = \sum \widehat{y_n}^2 - N\overline{\widehat{y}}^2 + \sum \widehat{e}^2;$$

*y empleando el resultado de la **Nota** 6 en la página˜14*

$$\sum_{n=1}^{N} (y_n - \overline{y})^2 = \sum_{n=1}^{N} (\widehat{y_n} - \overline{\widehat{y}})^2 + \sum_{n=1}^{N} \widehat{e_n}^2 \quad \text{o bien} \quad (\boldsymbol{y} - \overline{y})'(\boldsymbol{y} - \overline{y}) = (\widehat{\boldsymbol{y}} - \overline{\widehat{y}})'(\widehat{\boldsymbol{y}} - \overline{\widehat{y}}) + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}.$$

Dividiendo por N tenemos

$$\boxed{s^2_{\boldsymbol{y}} = s^2_{\widehat{\boldsymbol{y}}} + s^2_{\widehat{\boldsymbol{e}}}}$$

ya que $\overline{\widehat{\boldsymbol{e}}} = 0$; *y donde* $s^2_{\boldsymbol{z}}$ *es la varianza muestral de* $\boldsymbol{z}$.

Ejercicio 23. Demuestre que $\widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{y}}'\,\boldsymbol{y}$; es decir, $\sum \widehat{y_n}^2 = \sum \widehat{y_n} y_n$.

Caso especial (Modelos con término constante). *La suma explicada de cuadrados, SEC, se puede expresar como:*

$$\begin{aligned} SEC =& \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} + N\overline{y}^2 - 2N\overline{y}\overline{\widehat{y}} \\ =& \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} - N\overline{y}^2 && \text{ya que } \overline{y} = \overline{\widehat{y}} \text{ por haber término cte.} \\ =& N s^2_{\widehat{\boldsymbol{y}}} && \text{por la \textbf{Nota} 6} \end{aligned}$$

otras expresiones son:

$$\begin{aligned} =& \widehat{\boldsymbol{\beta}}'\,\mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}} - N\overline{y}^2 && \text{sustituyendo } \widehat{\boldsymbol{y}} \text{ por } \mathbf{x}\widehat{\boldsymbol{\beta}} \\ =& \widehat{\boldsymbol{y}}'\,\boldsymbol{y} - N\overline{\widehat{y}}\overline{y} && \text{por Ejercicio 23 y por } \overline{y} = \overline{\widehat{y}} \\ =& N s_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}} && \text{por la \textbf{Nota} 4} \end{aligned}$$

Además, en este caso en particular, la suma total de cuadrados, STC, se puede descomponer en la suma:

$$STC = SEC + SRC$$

ya que

$$\begin{aligned} \boldsymbol{y}'\,\boldsymbol{y} =& \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} && \text{de (4.2) (página 22)} \\ \boldsymbol{y}'\,\boldsymbol{y} - N\overline{y}^2 =& \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} - N\overline{y}^2 + \widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} && \text{restando a ambos lados } N\overline{y}^2 \\ STC =& \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} - N\overline{y}^2 + SRC && \text{por definición de } STC \text{ y } SRC \\ STC =& SEC + SRC && \text{por haber término constante } \overline{y} = \overline{\widehat{y}} \end{aligned}$$

Esta relación sugiere el nombre de "suma explicada de cuadrados", ya que descomponemos la variabilidad de la variable que queremos estudiar ($\boldsymbol{y}$) en dos partes: SRC es la variabilidad de los residuos (aquello que el modelo no "explica") y SEC es la variabilidad de $\widehat{\boldsymbol{y}}$, que es la estimación de la esperanza condicionada a los datos (aquello que "explica" el modelo).

En esta discusión se debe tener presente que el término "explicación" es engañoso. En el ejemplo del precio de las viviendas y su superficie, es sensato suponer que los precios dependen de las características de las viviendas, y en particular, que parte de las variaciones de los precios se deben a la variación en la superficie de las viviendas; por ello, el nombre de "suma explicada de cuadrados" toma todo su sentido.

Ahora bien, suponga que estima el modelo:

$$S_n = \beta_1 + \beta_2 P_n + U_n.$$

En este modelo, la superficie es función del precio de la vivienda, y por ser un modelo lineal con término constante, la relación algebraica $STC = SEC + SRC$ se cumple. Pero no tiene sentido suponer que las características de la vivienda se deben al precio; de lo contrario podríamos suponer que si el piso experimenta un alza en su precio, entonces, en consecuencia su superficie aumentará. Esto es absurdo, y podemos concluir que la relación $STC = SEC + SRC$ es puramente algebraica, y que su interpretación sólo es posible cuando el modelo estimado "tiene sentido" desde el punto de vista de la Teoría Económica.

La única interpretación posible a las estimaciones es de carácter puramente estadístico (y no de Teoría Económica): si un piso tiene un precio muy elevado, cabe "esperar" que el piso sea grande. (Este es un buen momento para que lea de nuevo la Introducción a este Tema 1 en la página˜3).

4.2. Más propiedades algebraicas.

4.2.1. Proyecciones

Si se cumple el cuarto supuesto, entonces $\mathbf{x}'\mathbf{x}$ es de rango completo y existe la matriz $(\mathbf{x}'\mathbf{x})^{-1}$. Sólo entonces, es posible "despejar" $\widehat{\boldsymbol{\beta}}$ en las ecuaciones normales (3.1) para obtener la expresión:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y}\,.$$

Llamamos estimación MCO de $\boldsymbol{y}$ a

$$\widehat{\boldsymbol{y}} = \mathbf{x}\widehat{\boldsymbol{\beta}}$$

que es igual a

$$\widehat{\boldsymbol{y}} = \mathbf{x}\widehat{\boldsymbol{\beta}} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y}\,.$$

Por otra parte,

$$\begin{aligned}\widehat{\boldsymbol{e}} &= \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{y} - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y} \\ &= (\mathbf{I} - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}')\,\boldsymbol{y}\end{aligned}$$

Si llamamos $\mathbf{p} \equiv \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$ y $\mathbf{m} \equiv \mathbf{I} - \mathbf{p}$, entonces

$$\widehat{\boldsymbol{y}} = \mathbf{p}\boldsymbol{y} \equiv \boldsymbol{y}_{\subset\mathbf{x}}; \qquad \widehat{\boldsymbol{e}} = \mathbf{m}\boldsymbol{y} \equiv \boldsymbol{y}_{\perp\mathbf{x}}\,.$$

donde $\boldsymbol{y}_{\subset\mathbf{x}}$ es la parte de $\boldsymbol{y}$ que se puede expresar como función lineal de las $\mathbf{x}$; e $\boldsymbol{y}_{\perp\mathbf{x}}$ es la parte de $\boldsymbol{y}$ que no se puede expresar como función lineal de las $\mathbf{x}$, es decir, la parte de $\boldsymbol{y}$ ortogonal a las $\mathbf{x}$.

Además sabemos que $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$, por tanto

$$\boldsymbol{y} = \mathbf{p}\boldsymbol{y} + \mathbf{m}\boldsymbol{y} = \boldsymbol{y}_{\subset\mathbf{x}} + \boldsymbol{y}_{\perp\mathbf{x}}\,.$$

(véase la figura de la Transparencia T46); y $\mathbf{p} + \mathbf{m} = \mathbf{I}$.

Nota 8. La inversa de una matriz simétrica es simétrica, así pues, $(\mathbf{x}'\mathbf{x})^{-1}$ es una matriz simétrica, y por tanto $\left[(\mathbf{x}'\mathbf{x})^{-1}\right]' = (\mathbf{x}'\mathbf{x})^{-1}$. La traspuesta de un producto de matrices $\mathbf{a}$ y $\mathbf{b}$ es $[\mathbf{ab}]' = \mathbf{b}'\mathbf{a}'$.

Ejercicio 24. Cual será la expresión de la traspuesta del producto de tres matrices $(\mathbf{abc})'$?

Ejercicio 25. Demuestre que $\mathbf{p}'\mathbf{m} = \mathbf{p}'(\mathbf{I} - \mathbf{p}) = \mathbf{0}$.

Se puede verificar (empleando el resultado del ejercicio anterior) que $\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} = 0$, pues

$$\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{e}} = (\mathbf{p}\boldsymbol{y})'\mathbf{m}\boldsymbol{y} = \boldsymbol{y}'\mathbf{p}'\mathbf{m}\boldsymbol{y} = \boldsymbol{y}'\mathbf{0}\boldsymbol{y} = 0;$$

resultado que ya vimos en la Ecuación 4.1 en la página˜21. Por tanto, podemos concluir que:

> *La estimación MCO separa el vector $\boldsymbol{y}$ en dos componentes, $\widehat{\boldsymbol{y}}$ y $\widehat{\boldsymbol{e}}$, ortogonales entre si (perpendiculares). La primera componente $\widehat{\boldsymbol{y}}$ es una combinación lineal de los regresores (la parte de $\boldsymbol{y}$ que se puede describir mediante un modelo lineal con las variables explicativas). La segunda componente es la parte de $\boldsymbol{y}$ ortogonal a los regresores (lo que no se puede describir linealmente con los regresores, ni siquiera de manera aproximada).*

Ejercicio 26. Demuestre que $\mathbf{m}' = \mathbf{m}$ y que $\mathbf{m}'\mathbf{m} = \mathbf{m}$,

De los ejercicios y resultados anteriores, se deduce que

$$\begin{aligned}\boldsymbol{y}'\boldsymbol{y} &= (\mathbf{p}\boldsymbol{y} + \mathbf{m}\boldsymbol{y})'(\mathbf{p}\boldsymbol{y} + \mathbf{m}\boldsymbol{y}) \\ &= \boldsymbol{y}'\mathbf{p}'\mathbf{p}\boldsymbol{y} + \boldsymbol{y}'\mathbf{m}'\mathbf{m}\boldsymbol{y} && \text{pues } \mathbf{p}'\mathbf{m} = \mathbf{p}\mathbf{m}' = \mathbf{0} \\ &= \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} && \text{(expresión que ya obtuvimos en (4.2); T. de Pitágoras)}\end{aligned}$$

(véase la figura de la Transparencia T46).

La estimación MCO de $\boldsymbol{y}$, es decir el vector $\widehat{\boldsymbol{y}} = \mathbf{p}\boldsymbol{y}$, se obtiene proyectando $\boldsymbol{y}$ sobre el conjunto de todas las combinaciones lineales de los regresores (todos los posibles modelos lineales generados con los regresores $\mathbf{x}$), para seleccionar aquel cuya suma de residuos al cuadrado $\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}$ es menor. (compare la figura de la Transparencia T46 con la figura inmediatamente anterior).

De manera análoga, los residuos $\widehat{\boldsymbol{e}} = \mathbf{m}\boldsymbol{y}$ son la proyección del vector $\boldsymbol{y}$ sobre el espacio ortogonal al anterior (al de los modelos lineales obtenidos como combinaciones lineales de los regresores $\mathbf{x}$). Es decir, $\widehat{\boldsymbol{e}}$ es la parte de $\boldsymbol{y}$ que no es expresable en función de un modelo lineal de $\mathbf{x}$ (o lo que es lo mismo, no es explicable como combinación lineal de los regresores).

Por tanto, la matriz $\mathbf{p}$ es una aplicación lineal que "proyecta" el vector $\boldsymbol{y}$ sobre las $\mathbf{x}$ (sobre el espacio vectorial expandido por las columnas —los regresores— de la matriz $\mathbf{x}$); y la matriz $\mathbf{m}$ es una aplicación lineal que "proyecta" el vector $\boldsymbol{y}$ sobre el espacio ortogonal a las $\mathbf{x}$ (sobre el espacio vectorial ortogonal al expandido por las columnas de la matriz $\mathbf{x}$);

Proyectores ortogonales

Definición 2. Decimos que una matriz $\mathbf{q}$ es simétrica si se verifica que $\mathbf{q}' = \mathbf{q}$.

Definición 3. Decimos que una matriz $\mathbf{q}$ es idempotente si se verifica que $\mathbf{qq} = \mathbf{q}$.

Definición 4. Sea $\mathbf{q}$ una matriz idempotente ($\mathbf{qq} = \mathbf{q}$). Si además la matriz es simétrica ($\mathbf{q} = \mathbf{q}'$), entonces se dice que la matriz $\mathbf{q}$ es un *proyector ortogonal*.

Ejercicio 27. Verifique que $\mathbf{p}$ y $\mathbf{m}$ son proyectores ortogonales.

4.2.2. Regresión particionada

Wooldridge (páginas 85 y ejercicio 3.17 de la página 119 2006). Pero mejor en:

- Johnston y Dinardo (páginas 88 a 95 y 116 a 118 2001)
- Novales (paginas 85 a 86 1993)
- Peña (paginas 390 a 392 2002)

En la parte de contrastación de hipótesis será necesario, en ocasiones, tener expresiones explícitas de sub-vectores de $\widehat{\boldsymbol{\beta}}$

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta_1}} \\ \cdots \\ \widehat{\boldsymbol{\beta_2}} \end{bmatrix}$$

Para ello vamos a reescribir el modelo lineal de la forma $\boldsymbol{Y} = \mathbf{X_1}\,\boldsymbol{\beta_1} + \mathbf{X_2}\,\boldsymbol{\beta_2} + \boldsymbol{U}$ y también las ecuaciones normales 3.1 en la página~12 del siguiente modo

$$\left[\begin{pmatrix} \mathbf{x_1}' \\ \mathbf{x_2}' \end{pmatrix} \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} \end{bmatrix}\right] \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix} = \begin{bmatrix} \mathbf{x_1}'\boldsymbol{y} \\ \mathbf{x_2}'\boldsymbol{y} \end{bmatrix}$$

o mejor aún

$$\begin{aligned} \mathbf{x_1}'\mathbf{x_1}\widehat{\boldsymbol{\beta_1}} + \mathbf{x_1}'\mathbf{x_2}\widehat{\boldsymbol{\beta_2}} &= \mathbf{x_1}'\boldsymbol{y} \\ \mathbf{x_2}'\mathbf{x_1}\widehat{\boldsymbol{\beta_1}} + \mathbf{x_2}'\mathbf{x_2}\widehat{\boldsymbol{\beta_2}} &= \mathbf{x_2}'\boldsymbol{y} \end{aligned} \tag{4.3}$$

donde $\mathbf{x} = \left[\, \mathbf{x_1} \vdots \mathbf{x_2} \,\right]$, es decir, hemos dividido la matriz de regresores en dos conjuntos de columnas, cada uno asociado a los parámetros de los vectores $\widehat{\boldsymbol{\beta_1}}$ y $\widehat{\boldsymbol{\beta_2}}$.

Si pre-multiplicamos la primera de las ecuaciones por $\mathbf{x_2}'\mathbf{x_1}(\mathbf{x_1}'\mathbf{x_1})^{-1}$ y la restamos de la segunda, tenemos

$$\left(\mathbf{x_2}'\mathbf{x_2} - \mathbf{x_2}'\mathbf{x_1}(\mathbf{x_1}'\mathbf{x_1})^{-1}\mathbf{x_1}'\mathbf{x_2}\right)\widehat{\boldsymbol{\beta_2}} = \mathbf{x_2}'\boldsymbol{y} - \mathbf{x_2}'\mathbf{x_1}(\mathbf{x_1}'\mathbf{x_1})^{-1}\mathbf{x_1}'\boldsymbol{y} \tag{4.4}$$

Vamos ha definir los proyectores

$$\mathbf{p_1} = \mathbf{x_1}(\mathbf{x_1}'\mathbf{x_1})^{-1}\mathbf{x_1}' \quad \text{y} \quad \mathbf{m_1} = \mathbf{I} - \mathbf{p_1}$$

El primero de ellos es una aplicación lineal que "proyecta" cualquier vector $\boldsymbol{z}$ sobre el primer conjunto de regresores $\mathbf{x_1}$, y el segundo lo "proyecta" sobre el espacio ortogonal al primero. Por tanto $\mathbf{p_1}\boldsymbol{z}$ realiza la regresión MCO del vector $\boldsymbol{z}$ sobre los regresores $\mathbf{x_1}$ y $\mathbf{m_1}\boldsymbol{z}$ son los residuos (los errores) de dicha regresión.

Sustituyendo $\mathbf{p_1}$ y $\mathbf{m_1}$ en (4.4) tenemos

$$\widehat{\boldsymbol{\beta_2}} = (\mathbf{x_2}'\mathbf{m_1}\mathbf{x_2})^{-1}\mathbf{x_2}'\mathbf{m_1}\boldsymbol{y} \tag{4.5}$$

y sustituyendo esta expresión en las ecuaciones normales (4.3)

$$\widehat{\boldsymbol{\beta_1}} = (\mathbf{x_1}'\mathbf{x_1})^{-1}\mathbf{x_1}'(\boldsymbol{y} - \mathbf{x_2}\widehat{\boldsymbol{\beta_2}}) \tag{4.6}$$

Es sencillo verificar que, de nuevo, $\mathbf{m_1}' = \mathbf{m_1}$. y que $\mathbf{m_1}'\,\mathbf{m_1} = \mathbf{m_1}$. Por lo que (4.5) se puede escribir como

$$\widehat{\boldsymbol{\beta_2}} = (\mathbf{x_2}'\,\mathbf{m_1}'\,\mathbf{m_1}\,\mathbf{x_2})^{-1}\,\mathbf{x_2}'\,\mathbf{m_1}'\,\mathbf{m_1}\,\boldsymbol{y}$$

En esta expresión, $\mathbf{m_1}\boldsymbol{y}$ son los residuos obtenidos al realizar la regresión de $\boldsymbol{y}$ sobre el subconjunto de regresores $\mathbf{x_1}$ (la parte de $\boldsymbol{y}$ ortogonal a $\mathbf{x_1}$). Y $\mathbf{m_1}\mathbf{x_2}$ es una matriz cuyas columnas son los residuos obtenidos realizando la regresión de cada una de las columnas de $\mathbf{x_2}$ sobre $\mathbf{x_1}$ (la parte de $\mathbf{x_2}$ ortogonal a $\mathbf{x_1}$).

Nótese que si llamamos $\boldsymbol{y}_{\perp\mathbf{x_1}} = \mathbf{m_1}\boldsymbol{y}$ a los residuos de la primera regresión, y $\mathbf{x_2}_{\perp\mathbf{x_1}} = \mathbf{m_1}\mathbf{x_2}$ a la matriz de residuos de las regresiones de las columnas de $\mathbf{x_2}$, entonces (4.5) se puede escribir como

$$\widehat{\boldsymbol{\beta_2}} = (\mathbf{x_2}_{\perp\mathbf{x_1}}{}'\,\mathbf{x_2}_{\perp\mathbf{x_1}})^{-1}\mathbf{x_2}_{\perp\mathbf{x_1}}{}'\,\boldsymbol{y}_{\perp\mathbf{x_1}}$$

Este resultado nos indica que podemos estimar $\widehat{\boldsymbol{\beta_2}}$ mediante regresiones auxiliares:

1. Realizamos la regresión de $\boldsymbol{y}$ sobre el primer conjunto de regresores $\mathbf{x_1}$ y obtenemos el vector de residuos $\boldsymbol{y}_{\perp\mathbf{x_1}}$
2. Realizamos las regresiones de cada una de las columnas de $\mathbf{x_2}$ sobre las variables $\mathbf{x_1}$, almacenando los residuos de cada regresión en las columnas de $\mathbf{x_2}_{\perp\mathbf{x_1}}$.
3. por último, $\widehat{\boldsymbol{\beta_2}}$ se obtiene de la regresión de $\boldsymbol{y}_{\perp\mathbf{x_1}}$ sobre $\mathbf{x_2}_{\perp\mathbf{x_1}}$, es decir, $\widehat{\boldsymbol{\beta_2}} = (\mathbf{x_2}_{\perp\mathbf{x_1}}{}'\,\mathbf{x_2}_{\perp\mathbf{x_1}})^{-1}\mathbf{x_2}_{\perp\mathbf{x_1}}{}'\,\boldsymbol{y}_{\perp\mathbf{x_1}}$
4. las estimaciones de $\widehat{\boldsymbol{\beta_1}}$ se pueden recuperar de (4.6)

Nótese que si $\boldsymbol{\beta_2} = \beta_2$; es decir, si el sub-vector se reduce a un escalar (un único parámetro), entonces la expresión (4.5) se reduce a

$$\widehat{\boldsymbol{\beta_2}} = \widehat{\beta_2} = (\underset{[1\times N]}{\boldsymbol{x_2}'}\;\underset{[N\times N]}{\mathbf{m_1}}\;\underset{[N\times 1]}{\boldsymbol{x_2}})^{-1}\boldsymbol{x_2}'\,\mathbf{m_1}\boldsymbol{y} = \frac{\boldsymbol{x_2}'\,\mathbf{m_1}\boldsymbol{y}}{\underset{[1\times N]}{\boldsymbol{x_2}'}\;\underset{[N\times N]}{\mathbf{m_1}}\;\underset{[N\times 1]}{\boldsymbol{x_2}}} \tag{4.7}$$

Regresión ortogonal particionada. Suponga que ambos grupos de regresores $\left[\mathbf{x_1} \vdots \mathbf{x_2}\right]$, son ortogonales entre si ($\mathbf{x_1}'\,\mathbf{x_2} = \mathbf{0}$), es decir, están incorrelados. En este caso, las ecuaciones 4.3 en la página anterior se reducen a

$$\begin{aligned} \mathbf{x_1}'\,\mathbf{x_1}\,\widehat{\boldsymbol{\beta_1}} &= \mathbf{x_1}'\,\boldsymbol{y} \\ \mathbf{x_2}'\,\mathbf{x_2}\,\widehat{\boldsymbol{\beta_2}} &= \mathbf{x_2}'\,\boldsymbol{y} \end{aligned};$$

y por lo tanto los vectores de coeficientes $\widehat{\boldsymbol{\beta_1}}$ y $\widehat{\boldsymbol{\beta_2}}$ se pueden estimar por separado mediante las regresiones de Y sobre $\mathbf{X_1}$, y de Y sobre $\mathbf{X_2}$. Esta es una generalización de la Nota 7 en la página˜18.

4.2.3. Regresión en desviaciones respecto a la media

Wooldridge (páginas 63, 64, 90 2006). Pero mejor:

- Novales (paginas 86 a 91 1993)
- Johnston y Dinardo (páginas 84 a 88 2001)
- Gujarati (Sección 6.1 2003, hay versión castellana de este manual)

Un caso particular de la regresión particionada es que el primer grupo de regresores se limite a la columna de unos. Es decir $\mathbf{x} = \left[\mathbf{1} \vdots \mathbf{x_2}\right]$, donde $\mathbf{x_1} = \mathbf{1}$. En este caso

$$\mathbf{p_1} = \mathbf{x_1}(\mathbf{x_1}'\,\mathbf{x_1})^{-1}\mathbf{x_1}' = \mathbf{1}(\mathbf{1}'\,\mathbf{1})^{-1}\mathbf{1}' = \frac{\mathbf{1}\,\mathbf{1}'}{N} = \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \cdots & \cdots & \ddots & \vdots \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}$$

por lo que

$$\mathbf{m_1}\boldsymbol{y} = (\mathbf{I} - \mathbf{p_1})\,\boldsymbol{y} = \begin{bmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \\ \vdots \\ y_N - \overline{y} \end{bmatrix} = \ddot{\boldsymbol{y}} \equiv \boldsymbol{y}_{\perp\mathbf{1}}$$

es decir, $\ddot{\boldsymbol{y}} = \mathbf{m_1}\boldsymbol{y}$ son las desviaciones de los elementos del vector columna $\boldsymbol{y}$ respecto de su media muestral $\overline{y}$ (son los residuos $\boldsymbol{y}_{\perp\mathbf{x_1}} \equiv \boldsymbol{y}_{\perp\mathbf{1}}$ de la primera regresión en el paso 1; aquí $\mathbf{x_1} = \mathbf{1}$. Véase la Ecuación 3.5 en la página~15). De manera similar, $\mathbf{m_1}\mathbf{x_2}$ da como resultado una matriz $\mathbf{x_2}_{\perp\mathbf{1}} \equiv \ddot{\mathbf{x}}_\mathbf{2}$ en la que aparecen las desviaciones de los datos de cada una de las columnas de $\mathbf{X_2}$ respecto de sus respectivas medias (son los residuos de las regresiones auxiliares del paso 2).

Ahora es inmediato estimar $\widehat{\boldsymbol{\beta_2}}$ como (paso 3)

$$\widehat{\boldsymbol{\beta_2}} = (\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\mathbf{x}}_\mathbf{2})^{-1}\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\boldsymbol{y}} \tag{4.8}$$

es decir, en un modelo con término constante, la estimación de todos los parámetros excepto el de la constante. se pueden obtener mediante la regresión de las variables del modelo en desviaciones respecto a su media. Por último (paso 4)

$$\widehat{\beta_1} = (\mathbf{1}'\,\mathbf{1})^{-1}\mathbf{1}'(\boldsymbol{y} - \mathbf{x_2}\widehat{\boldsymbol{\beta_2}}) = \frac{\mathbf{1}'(\boldsymbol{y} - \mathbf{x_2}\widehat{\boldsymbol{\beta_2}})}{N} = \overline{y} - \widehat{\beta_2}\overline{x_2} - \widehat{\beta_3}\overline{x_3} - \cdots - \widehat{\beta_k}\overline{x_k} \tag{4.9}$$

En definitiva, si en el modelo $Y_n = \beta_1 + \beta_2 X_{2n} + \cdots + \beta_k X_{kn}$ deseamos estimar por MCO sólo β_2, β_3, ..., β_k. Basta restar la media muestral a cada una de las variables del modelo, y realizar la regresión en un nuevo modelo sin término constante y con las nuevas variables transformadas. $\ddot{Y}_n = \beta_2\ddot{X}_{2n} + \cdots + \beta_k\ddot{X}_{kn}$.

Practica 28. Verifique con un programa econométrico la afirmación anterior.

Nótese además, que la expresión (4.8) se puede reescribir como:

$$\widehat{\boldsymbol{\beta_2}} = \left(\frac{1}{N}\,\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\mathbf{x}}_\mathbf{2}\right)^{-1}\left(\frac{1}{N}\,\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\boldsymbol{y}}\right);$$

donde $\frac{1}{N}\,\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\mathbf{x}}_\mathbf{2}$ es la matriz de covarianzas muestrales de los regresores, y $\frac{1}{N}\,\ddot{\mathbf{x}}_\mathbf{2}'\,\ddot{\boldsymbol{y}}$ es el vector de covarianzas muestrales entre los regresores y el regresando (que es la contrapartida muestral de la Ecuación C.1 en la página~49).

4.2.4. Añadiendo regresores

Suponga que ha estimado por MCO el siguiente modelo

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{U}.$$

Posteriormente decide incluir como regresor adicional la variable Z; entonces el nuevo modelo ampliado será:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}^* + c\,\boldsymbol{Z} + \boldsymbol{U}^*.$$

Podemos aplicar los resultados de la regresión particionada para obtener el coeficiente, c, asociado al nuevo regresor $\boldsymbol{Z}$ del siguiente modo (de 4.5 en la página~25):

$$c = (\boldsymbol{z}'\,\mathbf{m}\,\boldsymbol{z})^{-1}\boldsymbol{z}'\,\mathbf{m}\boldsymbol{y} = (\boldsymbol{z}_{\perp\mathbf{x}}'\,\boldsymbol{z}_{\perp\mathbf{x}})^{-1}\boldsymbol{z}_{\perp\mathbf{x}}'\,\boldsymbol{y}_{\perp\mathbf{x}}; \tag{4.10}$$

donde $\boldsymbol{y}_{\perp\mathbf{x}}$ son los residuos de la regresión MCO de $\boldsymbol{y}$ sobre $\mathbf{x}$ (la parte de $\boldsymbol{y}$ que no se puede expresar como función lineal de las $\mathbf{x}$, es decir, la parte de $\boldsymbol{y}$ ortogonal a las $\mathbf{x}$), y $\boldsymbol{z}_{\perp\mathbf{x}}$ son los residuos de la regresión MCO de $\boldsymbol{z}$ sobre $\mathbf{x}$ (la parte de $\boldsymbol{z}$ ortogonal a las $\mathbf{x}$), es decir $\boldsymbol{z}_{\perp\mathbf{x}} = \mathbf{m}\boldsymbol{z}$, e $\boldsymbol{y}_{\perp\mathbf{x}} = \mathbf{m}\boldsymbol{y}$; donde $\mathbf{m} = \left[\mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\right]$.

Practica 29. Verifique con un programa econométrico la afirmación anterior. Los pasos a seguir son

1. Calcule los residuos MCO con el modelo reducido.
2. Calcule los coeficientes estimados en el modelo ampliado. Fíjese en el valor obtenido para el coeficiente c asociado al nuevo regresor[4].
3. Calcule los residuos en la regresión de la nueva variable explicativa $\boldsymbol{z}$ sobre los antiguos regresores $\mathbf{x}$.
4. Calcule por MCO la regresión de los residuos del punto 3 sobre los residuos del punto 1; y compare el valor estimado con el obtenido en el punto 2.

[4]Nótese que el resto de coeficientes puede diferir respecto de los obtenidos en la nueva regresión. Esto será así siempre que el nuevo regresor tenga correlación con los del modelo inicial.

Suma de residuos: Cuando se añaden regresores a un modelo, la suma de residuos al cuadrado nunca crece; de hecho suele disminuir. Esto se cumple incluso si la variable añadida no tiene ningún sentido dentro del modelo (ninguna relación teórica). Veámoslo:

Del modelo inicial obtendremos los residuos

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}};$$

por otra parte, los residuos con el modelo ampliado son

$$\widehat{\boldsymbol{e}^*} = \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}^*} - \boldsymbol{z}\,\widehat{c}.$$

(nótese que si $\mathbf{x}'\,\boldsymbol{z} \neq \mathbf{0}$ entonces $\widehat{\beta} \neq \widehat{\beta^*}$; y que si $\widehat{c} \neq 0$ entonces $\widehat{\boldsymbol{e}} \neq \widehat{\boldsymbol{e}^*}$.)

De (4.6) sabemos que

$$\widehat{\boldsymbol{\beta}^*} = (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'(\boldsymbol{y} - \boldsymbol{z}\,\widehat{c}) = \widehat{\beta} - (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{z}\,\widehat{c}.$$

Sustituyendo $\widehat{\boldsymbol{\beta}^*}$ en $\widehat{\boldsymbol{e}^*}$ obtenemos

$$\begin{aligned}
\widehat{\boldsymbol{e}^*} &= \boldsymbol{y} - \mathbf{x}\widehat{\boldsymbol{\beta}} + \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{z}\,\widehat{c} - \boldsymbol{z}\,\widehat{c} \\
&= \widehat{\boldsymbol{e}} - \mathbf{m}\boldsymbol{z}\,\widehat{c} \\
&= \widehat{\boldsymbol{e}} - \boldsymbol{z}_{\perp\mathbf{x}}\widehat{c} && \text{de (4.10)}
\end{aligned}$$

Así pues,

$$\widehat{\boldsymbol{e}^*}'\,\widehat{\boldsymbol{e}^*} = \widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} + \widehat{c}^2\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big) - 2\widehat{c}\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\widehat{\boldsymbol{e}}$$

Teniendo en cuenta que de (4.10) $\widehat{c} = (\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}})^{-1}\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{y}_{\perp\mathbf{x}}$ y que $\widehat{\boldsymbol{e}} = \mathbf{m}\boldsymbol{y} = \boldsymbol{y}_{\perp\mathbf{x}}$ tenemos

$$\widehat{c}^2\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big) = \widehat{c}\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big)\widehat{c} = \widehat{c}\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big)\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big)^{-1}\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{y}_{\perp\mathbf{x}} = \widehat{c}\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{y}_{\perp\mathbf{x}} = \widehat{c}\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\widehat{\boldsymbol{e}}\,.$$

Por lo que finalmente

$$\underbrace{\widehat{\boldsymbol{e}^*}'\,\widehat{\boldsymbol{e}^*}}_{SRC^*} = \underbrace{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}_{SRC} - \underbrace{\widehat{c}^2\big(\boldsymbol{z}_{\perp\mathbf{x}}{}'\,\boldsymbol{z}_{\perp\mathbf{x}}\big)}_{\geq 0} \tag{4.11}$$

por lo que la suma de residuos al cuadrado del modelo ampliado SRC^* nunca será mayor que la del modelo reducido SRC.

4.2.5. Correlaciones parciales

Suponga que tiene tres variables; por ejemplo, la renta $\boldsymbol{r}$, la edad $\boldsymbol{e}$ y el número de años de estudio o formación $\boldsymbol{f}$ de una serie de individuos.

$$R_n = \beta_1 + \beta_2 F_n + \beta_3 E_n + U_n$$

Querríamos saber el grado de relación lineal entre dos de ellas, una vez "descontado" la relación lineal que la tercera tiene con ellas. En nuestro ejemplo nos podría interesar conocer el grado de relación lineal de la renta con la formación, una vez "descontado el efecto lineal" que la edad tiene con ambas (nótese que tanto para formarse como para generar rentas es necesario el transcurso del tiempo, por lo que generalmente hay una relación directa entre la edad y las otras dos variables).

La solución es "tomar" la parte de ambas variables, "renta" y "educación", ortogonal a la tercera, la"edad"; y observar la correlación de dichas partes (que ya no mantienen relación lineal ninguna con la variable "edad").

El modo de hacerlo es sencillo una vez visto lo anterior:

1. Se toman los residuos de la regresión de la variable renta $\boldsymbol{r}$ sobre la variable edad $\boldsymbol{e}$ y la constante (modelo lineal simple); es decir, se obtiene $\boldsymbol{r}_{\perp\mathbf{e}}$.
2. Se toman los residuos de la regresión de la variable formación $\boldsymbol{f}$ sobre la variable edad $\boldsymbol{e}$ y la constante (modelo lineal simple); es decir, se obtiene $\boldsymbol{f}_{\perp\mathbf{e}}$.
3. Por último se calcula el coeficiente de correlación simple de ambos residuos $r_{\boldsymbol{r}_{\perp\mathbf{e}}\boldsymbol{f}_{\perp\mathbf{e}}}$.

Dicho coeficiente es la correlación parcial de la variable renta $\boldsymbol{r}$ con la variable formación $\boldsymbol{f}$, una vez "descontado" el efecto de la edad $\boldsymbol{e}$ sobre ambas variables. Nótese que ambos residuos tiene media cero por ser residuos de un modelo con término constante.

Suponga por tanto que dividimos la matriz de regresores $\mathbf{x}$ en dos columnas; por ejemplo la primera variable no cte. $\boldsymbol{x_2}$ y el resto de $k-1$ regresores (incluyendo el término cte.) $\mathbf{w}$.

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{x_2} & \vdots & \mathbf{w} \end{bmatrix}$$

entonces el *coeficiente de correlación parcial de* $\boldsymbol{y}$ *con* $\boldsymbol{x_2}$ *una vez descontado el efecto de las demás variables (incluida la constante)* $\mathbf{w}$ es

$$r_{(\boldsymbol{y},\boldsymbol{x_2})_{\perp\mathbf{z}}} = \frac{\boldsymbol{y}'\,\mathbf{m_w}\,\boldsymbol{x_2}}{\sqrt{\boldsymbol{y}'\,\mathbf{m_w}\,\boldsymbol{y}}\sqrt{\boldsymbol{x_2}'\,\mathbf{m_w}\,\boldsymbol{x_2}}} = \frac{s_{\boldsymbol{y}_{\perp\mathbf{w}}\boldsymbol{x_2}_{\perp\mathbf{w}}}}{\sqrt{s^2_{\boldsymbol{y}_{\perp\mathbf{w}}}}\sqrt{s^2_{\boldsymbol{x_2}_{\perp\mathbf{w}}}}},$$

donde $\mathbf{m_w} = \mathbf{I} - \mathbf{w}(\mathbf{w}'\,\mathbf{w})^{-1}\mathbf{w}'$.

Ejercicio 30. Resuelva el ejercicio propuesto nº 2 del profesor José Alberto Mauricio.
http://www.ucm.es/info/ecocuan/ectr1/index.html#Material.

Ejercicio 31. Resuelva el ejercicio propuesto nº 3 del profesor José Alberto Mauricio.
http://www.ucm.es/info/ecocuan/ectr1/index.html#Material.

4.3. Medidas de ajuste

Las medidas de ajuste sirven para

- Cuantificar la reducción de incertidumbre que proporciona el modelo estimado.
- Comparar la bondad de modelos alternativos para la misma muestra

⇑ **Medidas de ajuste: Coeficiente de determinación R^2** 23

$$R^2 \equiv 1 - \frac{SRC}{STC}; \qquad R^2 \leq 1 \qquad \text{(no acotado inferiormente)}$$

Cuando hay término constante

$$R^2 = \frac{SEC}{STC}; \qquad 0 \leq R^2 \leq 1 \qquad \text{(acotado)}$$

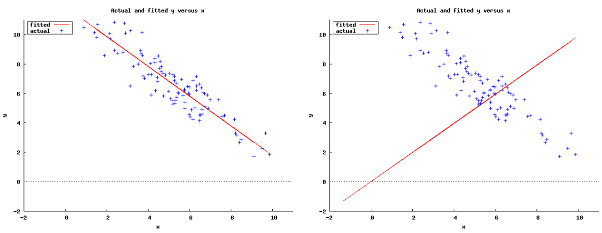
Coeficiente de Determinación o R^2 es una medida de ajuste frecuente. Cuando el modelo contiene un regresor constante, muestra el poder explicativo de los regresores no constantes. Se define como

$$R^2 \equiv 1 - \frac{SRC}{STC};$$

y puesto que SRC y STC son siempre mayores o iguales a cero, $R^2 \leq 1$.

Cuando el modelo no tiene cte. SRC puede ser mayor que STC, por lo que R^2 no está acotado inferiormente.

GNU Gretl: ejemplo simulado



Caso especial (Modelos con término constante). *Si el modelo tiene término constante, el coeficiente R^2 mide el porcentaje de variación de y "explicado" por los regresores no constantes del modelo; ya que*

$$R^2 = 1 - \frac{SRC}{STC} = \frac{STC - SRC}{STC} = \frac{SEC}{STC}$$

y por tanto $0 \leq R^2 \leq 1$.

Nótese además que

$$R^2 = \frac{SEC}{STC} = \frac{SEC^2}{STC \times SEC} = \frac{\left(N s_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}}\right)^2}{N s^2_{\boldsymbol{y}} \times N s^2_{\widehat{\boldsymbol{y}}}} = \frac{N^2}{N^2}\left(\frac{s_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}}}{\sqrt{s^2_{\boldsymbol{y}} \times s^2_{\widehat{\boldsymbol{y}}}}}\right)^2 = \left(r_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}}\right)^2, \tag{4.12}$$

donde $r_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}} = \frac{s_{\widehat{\boldsymbol{y}}\,\boldsymbol{y}}}{s_{\widehat{\boldsymbol{y}}} \times s_{\boldsymbol{y}}}$ *es el coeficiente de correlación lineal simple entre* $\widehat{\boldsymbol{y}}$ *y* $\boldsymbol{y}$.

Ejercicio 32. Calcule el coeficiente de determinación R^2 para el el ejemplo del precio de las viviendas

Ejercicio 33. Calcule el coeficiente de determinación para el **Modelo 1**: $Y_n = a + U_n$
Pista. piense cuanto vale SEC en este caso.

Ejercicio 34. Verifique que, para el caso del Modelo Lineal Simple $Y_n = a + bX_n + U_n$, el coeficiente de determinación R^2 es el cuadrado del coeficiente de correlación simple entre el regresando $\boldsymbol{y}$ y el regresor $\boldsymbol{x}$; es decir, que en este caso $R^2 = r^2_{\boldsymbol{y}\,\boldsymbol{x}}$. (Nótese que este resultado es diferente de (4.12)).

El coeficiente de determinación R^2 tiene algunos problemas al medir la bondad del ajuste.

- añadir nuevas variables al modelo (cuales quiera que sean) nunca hace crecer SRC pero esta suma si pude disminuir (véase la Sección 4.2.4)
- Por tanto el R^2 del modelo ampliado nunca puede ser menor que el del modelo inicial.
- Para evitar este efecto se emplea el coeficiente de determinación corregido (o ajustado) $\bar{R}^2$

El coeficiente de determinación *corregido* $\bar{R}^2$ de define como

$$\bar{R}^2 \equiv 1 - \frac{\frac{SRC}{N-k}}{\frac{STC}{N-1}}; = 1 - \frac{\mathfrak{s}^2_{\widehat{e}}}{\mathfrak{s}^2_{y}}$$

es decir, uno menos la fracción de la cuasivarianza de los errores con la cuasivarianza muestral del regresando. Por ello también es siempre menor o igual a uno.

1. compara estimadores insesgados de la varianza residual y de la varianza de la variable dependiente
2. penaliza modelos con un elevado numero de parámetros, al corregir por el número de grados de libertad $N - k$.

⇑ **Otras medidas de ajuste** **24**

R^2 corregido (mejor cuanto más elevado)

$$\bar{R}^2 \equiv 1 - \frac{\frac{SRC}{N-k}}{\frac{STC}{N-1}} = 1 - \frac{N-1}{N-k}(1 - R^2) \qquad \leq 1$$

Criterios de información de Akaike y de Schwartz (mejor cuanto más bajos)

$$AIC = N\ln(2\pi) + N\ln\left(\frac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N}\right) + N + 2(k+1)$$
$$SBC = N\ln(2\pi) + N\ln\left(\frac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N}\right) + N + (k+1)\ln(N)$$

Volver al recuadro del ejemplo del precio de las viviendas (página 17).

Otras medidas de la bondad del ajuste son los criterios de información de Akaike y de Schwartz (mejor cuanto más bajos)

Akaike prima la capacidad predictiva del modelo (pero tiende a sobreparametrizar)

Schwartz prima la correcta especificación

El programa Gretl (Gnu Regression, Econometrics and Time-series Library) realiza un cálculo especial de R^2 cuando el modelo no tiene término cte. En este caso el R-cuadrado es calculado como el cuadrado de la correlación entre los valores observado y ajustado de la variable dependiente (Véase Ramanathan, 1998, Sección 4.2).

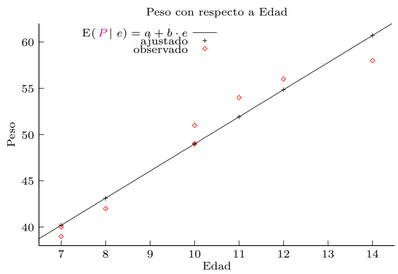
Los coeficientes de determinación nos dan información sobre el grado de ajuste del modelo, pero ¡ojo! nos pueden conducir a engaños. No es recomendable darles demasiada importancia, hay otras cuestiones sobre el modelo de mayor relevancia a la hora de valorarlo. . .

Ejemplo 35. [**peso de niños según su edad:**]

n	Peso Kg	Edad
1	39	7
2	40	7
3	42	8
4	49	10
5	51	10
6	54	11
7	56	12
8	58	14

Cuadro 3: Peso (en kilogramos) y edad (en años)

(Modelo 1 $P_n = \beta_1 + \beta_2 EDAD_n + U_n$)

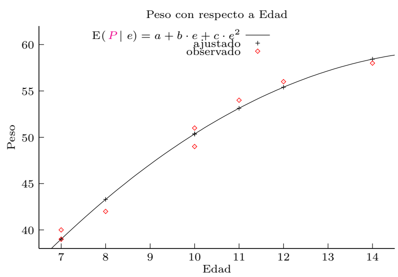


$$\widehat{\text{Peso\_Kg}} = \underset{(6{,}999)}{19,6910} + \underset{(10{,}564)}{2,93003}\,\text{Edad}$$

$$T = 8 \quad \bar{R}^2 = 0,9405 \quad F(1,6) = 111,6 \quad \hat{\sigma} = 1,8161$$

(entre paréntesis, los estadísticos t)

(Modelo 2 $P_n = \beta_1 + \beta_2 EDAD_n + \beta_3 {EDAD_n}^2 + U_n$)

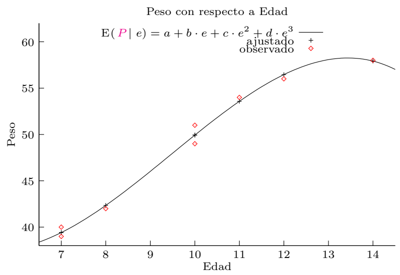


$$\widehat{\text{Peso\_Kg}} = \underset{(-0{,}664)}{-5,11497} + \underset{(5{,}159)}{8,06835}\,\text{Edad} - \underset{(-3{,}305)}{0,252102}\,\text{Edad2}$$

$$T = 8 \quad \bar{R}^2 = 0,9776 \quad F(2,5) = 153,57 \quad \hat{\sigma} = 1,1148$$

(entre paréntesis, los estadísticos t)

(Modelo 3 $P_n = \beta_1 + \beta_2 EDAD_n + \beta_3 EDAD_n{}^2 + \beta_4 EDAD_n{}^3 + U_n$)



$$\widehat{\text{Peso\_Kg}} = \underset{(1{,}904)}{81,7714} - \underset{(-1{,}419)}{18,5964}\,\text{Edad} + \underset{(1{,}845)}{2,37778}\,\text{Edad2} - \underset{(-2{,}043)}{0,0836541}\,\text{Edad3}$$

$$T = 8 \quad \bar{R}^2 = 0,9863 \quad F(3,4) = 168,75 \quad \hat{\sigma} = 0,87188$$

(entre paréntesis, los estadísticos t)

5. Propiedades estadísticas de los estimadores MCO

- Capítulos 2 y 3 de Wooldridge (2006)
- Apéndice E2 de Wooldridge (2006)

⇑ **Estimador MCO $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$** **25**

Los coeficientes *estimados* verifican

$$\mathbf{x}'\,\boldsymbol{y} = \mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}}$$

por Supuesto 4 $\boxed{\text{T13}}$ de independencia lineal podemos despejar $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{y}$$

que es una *estimación*.
El *estimador* de los coeficientes es $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{Y}$ o bien

$$\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} \equiv \widehat{\boldsymbol{\beta}}\,\Big|\,\mathbf{x} = (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{Y} = \mathbf{a}\,\boldsymbol{Y} = \boldsymbol{\beta} + \mathbf{a}\,\boldsymbol{U}$$

donde $\boldsymbol{Y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{U}$ suponiendo conocidas las realizaciones de los regresores.

Nota 9. Nótese las dimensiones de la matriz:

$$\underset{[k\times N]}{\mathbf{a}} \equiv (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}' = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kN} \end{bmatrix};$$

por lo tanto, $\widehat{\boldsymbol{\beta}}$ son k combinaciones lineales de los N datos del vector $\boldsymbol{y}$, donde los coeficientes específicos de cada combinación son los elementos de cada una de las filas de la matriz $(\mathbf{x}'\,\mathbf{x})^{-1}\,\mathbf{x}'$.

Del mismo modo, cada uno de los elementos del vector aleatorio $\widehat{\boldsymbol{\beta}}$ es una combinación lineal de las N variables aleatorias Y_n.

Nótese además que

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} \equiv \widehat{\boldsymbol{\beta}}\,\Big|\,\mathbf{x} &= \mathbf{a}\,\boldsymbol{Y} \\ &= \mathbf{a}\left[\mathbf{x}\boldsymbol{\beta} + \boldsymbol{U}\right] \\ &= \boldsymbol{\beta} + \mathbf{a}\boldsymbol{U} \\ &= \boldsymbol{\beta} + (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{U} \end{aligned}$$

es decir:

> $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ *es igual al verdadero valor de los parámetros más una combinación lineal (o suma ponderada) de las perturbaciones determinada por los coeficientes a_{ij} de la matriz* $\mathbf{a}$.

5.1. Esperanza de los estimadores MCO $\widehat{\boldsymbol{\beta}}_{\mid \mathbf{x}}$

⇑ Esperanza del estimador MCO $\widehat{\boldsymbol{\beta}}_{\mid \mathbf{x}}$ **26**

Denotemos $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ por $\underset{[k\times N]}{\mathbf{A}}$

$$\begin{aligned}\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\;\middle|\;\mathbf{x}\right) =& \mathrm{E}(\boldsymbol{\beta}+\mathbf{A}\boldsymbol{U}\mid \mathbf{x})\\ =& \mathrm{E}(\boldsymbol{\beta}+\mathbf{a}\boldsymbol{U}\mid \mathbf{x})\\ =& \boldsymbol{\beta}+\mathbf{a}\cdot\mathrm{E}(\boldsymbol{U}\mid \mathbf{x})\\ =& \boldsymbol{\beta}\end{aligned}$$

por lo tanto es un estimador insesgado.

Si los regresores son NO estocásticos, la demostración es más sencilla aún

$$\begin{aligned}\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\right) =& \mathrm{E}(\boldsymbol{\beta}+\mathbf{a}\boldsymbol{U})\\ =& \boldsymbol{\beta}+\mathbf{a}\cdot\mathrm{E}(\boldsymbol{U})\\ =& \boldsymbol{\beta}\end{aligned}$$

Modelo 2. [Modelo Lineal Simple (caso particular $\boxed{\mathbf{T16}}$).]
De 3.7 en la página˜15 resulta

$$\widehat{b} = \frac{\sum_n (x_n-\overline{x})(y_n-\overline{y})}{\sum_n (x_n-\overline{x})^2} = \frac{\sum_n y_n(x_n-\overline{x})}{\sum_n (x_n-\overline{x})^2}.$$

es decir,

$$\widehat{b} = \sum_n m_n y_n, \tag{5.1}$$

donde

$$m_n = \frac{x_n-\overline{x}}{\sum_n (x_n-\overline{x})^2}.$$

Por tanto, $\widehat{b}$ es una combinación lineal de los datos y_n (donde m_n son los coeficientes de dicha combinación); y entonces $\widehat{a}$ también es combinación lineal de los datos y_n (véase 3.10 en la página˜15).

Por 5.1 sabemos que $\widehat{b}_{\mid \mathbf{x}} = \sum m_n Y_n$, donde

$$m_n = \frac{x_n-\overline{x}}{\sum (x_n-\overline{x})^2}.$$

Se puede verificar que

1. $\sum m_n = 0$
2. $\sum m_n^2 = \frac{1}{\sum x_n^2} = \frac{1}{\sum (x_n-\overline{x})^2} = \frac{1}{N s_{\boldsymbol{x}}^2}$
3. $\sum m_n(x_n-\overline{x}) = \sum m_n x_n = 1.$

Entonces,

$$\begin{aligned}\widehat{b}_{\mid \mathbf{x}} =& \sum m_n(a+bx_n+U_n)\\ =& a\sum m_n + b\sum m_n x_n + \sum m_n U_n \;=\; b+\sum m_n U_n\end{aligned}$$

y

$$\mathrm{E}\left(\widehat{b}\;\middle|\;\mathbf{x}\right) = b+\sum m_n \mathrm{E}(U_n\mid \mathbf{x}) = b.$$

(Novales, 1997; Gujarati, 2003, pag. 488–491 y pag. 100 respectivamente).

Por otra parte, de 3.10 en la página˜15 sabemos que

$$\widehat{a} \quad = \quad \overline{y}-\widehat{b}\,\overline{x} \quad = \quad \frac{1}{N}\sum y_n - \widehat{b}\frac{1}{N}\sum x_n.$$

Por lo tanto el estimador condicionado es

$$\widehat{a}_{\mid \mathbf{x}} = \frac{1}{N}\sum Y_n - \left(\widehat{b}_{\mid \mathbf{x}}\right)\frac{1}{N}\sum x_n$$

cuya esperanza es

$$\begin{aligned}
\mathrm{E}(\widehat{a}\mid \mathbf{x}) &= \frac{1}{N}\sum \mathrm{E}(Y_n\mid \mathbf{x}) - \mathrm{E}\left(\widehat{b}\,\middle|\,\mathbf{x}\right)\frac{1}{N}\sum x_n\\
&= \frac{1}{N}\sum \mathrm{E}(Y_n\mid \mathbf{x}) - b\frac{1}{N}\sum x_n\\
&= \frac{1}{N}\sum \mathrm{E}(a+bx_n+U_n\mid \mathbf{x}) - b\frac{1}{N}\sum x_n\\
&= \frac{1}{N}\sum a + b\frac{1}{N}\sum x_n + \frac{1}{N}\sum \mathrm{E}(U_n\mid \mathbf{x}) - b\frac{1}{N}\sum x_n\\
&= a.
\end{aligned}$$

Ejercicio 36. Verifique que el estimador MCO del parámetro a del **Modelo 1** (constante como único regresor) es insesgado.

5.2. Varianza de los estimadores MCO $\widehat{\boldsymbol{\beta}}_{|\mathbf{x}}$

⇑ **Varianza del estimador MCO $\widehat{\boldsymbol{\beta}}_{|\mathbf{x}}$** 27

Aplicando la def. de la Ecuación (1) tenemos:

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) =&\mathrm{E}\left(\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\,\middle|\,\mathbf{x}\right)\\
=&\mathrm{E}\left((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{U}\boldsymbol{U}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\,\middle|\,\mathbf{x}\right)\\
=&(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\,\mathrm{E}\left(\boldsymbol{U}\boldsymbol{U}'\,\middle|\,\mathbf{x}\right)\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\\
=&\sigma^2(\mathbf{x}'\mathbf{x})^{-1}
\end{aligned}$$

Modelo 2. [Modelo Lineal Simple] Sabemos de (3.7) en la página~15 que $\mathbf{x}'\mathbf{x} = \begin{pmatrix} N & \sum x_n \\ \sum x_n & \sum x_n^2\end{pmatrix}$ cuyo determinante es

$$\det \mathbf{x}'\mathbf{x} \equiv |\mathbf{x}'\mathbf{x}| = N\sum x_n^2 - \left(\sum x_n\right)^2 = N\sum(x_n-\overline{x})^2;$$

Por tanto la matriz de varianzas y covarianzas del estimador es:

$$\sigma^2(\mathbf{x}'\mathbf{x})^{-1} = \frac{\sigma^2}{N\sum(x_n-\overline{x})^2}\cdot\begin{pmatrix}\sum x_n^2 & -\sum x_n\\ -\sum x_n & N\end{pmatrix}.$$

Nótese que

$$\sum(x_n-\overline{x})^2 = N\cdot s_{\boldsymbol{x}}^2.$$

Así pues, podemos deducir que

$$\mathrm{Var}(\widehat{a}\mid \mathbf{x}) = \frac{\sigma^2\sum x_n^2}{N\sum(x_n-\overline{x})^2} = \frac{\sigma^2\overline{\boldsymbol{x}^2}}{N\cdot s_{\boldsymbol{x}}^2};\quad \text{y}\quad \mathrm{Var}\left(\widehat{b}\,\middle|\,\mathbf{x}\right) = \frac{\sigma^2}{\sum(x_n-\overline{x})^2} = \frac{\sigma^2}{N\cdot s_{\boldsymbol{x}}^2}. \tag{5.2}$$

Además, ambos estimadores tienen una covarianza igual a

$$\mathrm{Cov}\left(\widehat{a},\widehat{b}\,\middle|\,\mathbf{x}\right) = \frac{-\sigma^2\sum x_n}{N\sum(x_n-\overline{x})^2} = \frac{-\sigma^2\cdot\overline{x}}{N\cdot s_{\boldsymbol{x}}^2} \tag{5.3}$$

Ejemplo 37. [**continuación de "precio de las viviendas":**]

Podemos calcular la inversa de $\mathbf{x}'\mathbf{x}$:

$$(\mathbf{x}'\mathbf{x})^{-1} = \begin{bmatrix} 9.1293e-01 & -4.4036e-04\\ -4.4036e-04 & 2.3044e-07\end{bmatrix};$$

así pues, las desviaciones típicas de $\widehat{a}_{|\mathbf{x}}$ y $\widehat{b}_{|\mathbf{x}}$ son (véase 5.2 en la página anterior)

$$\begin{aligned}\operatorname{Dt}(\widehat{a}\mid\mathbf{x}) =& \sqrt{\sigma^2\cdot(9.1293e-01)} = \sqrt{\frac{\sigma^2\overline{\boldsymbol{x}^2}}{N\cdot s_{\boldsymbol{x}}^2}}\\ \operatorname{Dt}\left(\widehat{b}\,\middle|\,\mathbf{x}\right) =& \sqrt{\sigma^2\cdot(2.3044e-07)} = \sqrt{\frac{\sigma^2}{N\cdot s_{\boldsymbol{x}}^2}}.\end{aligned}$$

Pero no conocemos $\sigma^2_{U_n}$.

Continuación del ejemplo “precio de las viviendas” en la página 41

Practica 38. Observe los resultados de las estimaciones del ejemplo del “precio de las viviendas”. ¿Qué estimación cree que es más fiable, la de la pendiente o la de la constante? Con los datos del ejemplo del “precio de las viviendas”, repita la regresión pero con las siguientes modificaciones:

1. con todos los datos excepto los de la última vivienda
2. con todos los datos excepto los de las últimas dos viviendas
3. con todos los datos excepto los de la primera y la última viviendas

¿Confirman los resultados de estas regresiones su respuesta a la primera pregunta?
ejemplo del “precio de las viviendas’ en GNU Gretl

Nota 10. Sea $\underset{[m\times N]}{\mathbf{a}}$, entonces, aplicando la definición de la Nota 2

$$\begin{aligned}\operatorname{Var}(\mathbf{a}\boldsymbol{Y}) =& \operatorname{E}\left(\mathbf{a}\boldsymbol{Y}\boldsymbol{Y}'\mathbf{a}'\right) - \operatorname{E}(\mathbf{a}\boldsymbol{Y})\operatorname{E}\left(\boldsymbol{Y}'\mathbf{a}'\right)\\ =& \,\mathbf{a}\left[\operatorname{E}\left(\boldsymbol{Y}\boldsymbol{Y}'\right) - \operatorname{E}(\boldsymbol{Y})\operatorname{E}\left(\boldsymbol{Y}'\right)\right]\mathbf{a}' && \text{sacando factores comunes}\\ =& \,\mathbf{a}\operatorname{Var}(\boldsymbol{Y})\,\mathbf{a}'\end{aligned}$$

Nota 11. Sean $\underset{[n\times N]}{\mathbf{q}}$ y $\underset{[m\times N]}{\mathbf{r}}$ matrices, y $\boldsymbol{v}$ y $\boldsymbol{w}$ vectores de orden n y m respectivamente. Entonces

$$\operatorname{E}(\mathbf{q}\boldsymbol{U}+\boldsymbol{v}) = \operatorname{E}(\mathbf{q}\boldsymbol{U}) + \operatorname{E}(\boldsymbol{v}) = \mathbf{q}\operatorname{E}(\boldsymbol{U}) + \boldsymbol{v},$$

y

$$\operatorname{Var}(\mathbf{q}\boldsymbol{U}+\boldsymbol{v}) = \operatorname{Var}(\mathbf{q}\boldsymbol{U}) = \mathbf{q}\operatorname{Var}(\boldsymbol{U})\,\mathbf{q}',$$

además

$$\operatorname{Cov}(\mathbf{q}\boldsymbol{U}+\boldsymbol{v},\mathbf{r}\boldsymbol{U}+\boldsymbol{w}) = \operatorname{Cov}(\mathbf{q}\boldsymbol{U},\mathbf{r}\boldsymbol{U}) = \mathbf{q}\operatorname{Cov}(\boldsymbol{U},\boldsymbol{U})\,\mathbf{r}' = \mathbf{q}\operatorname{Var}(\boldsymbol{U})\,\mathbf{r}'$$

Nota 12. Sean $\underset{[n\times N]}{\mathbf{Q}} = f(\mathbf{X})$ y $\underset{[m\times N]}{\mathbf{R}} = g(\mathbf{X})$ matrices, y $\boldsymbol{v}$ y $\boldsymbol{w}$ vectores de orden n y m respectivamente; sea además $\mathbf{X}=\mathbf{x}$, por lo que $\mathbf{q}=f(\mathbf{x})$ y $\mathbf{r}=g(\mathbf{x})$. Entonces

$$\operatorname{E}(\mathbf{Q}\boldsymbol{U}+\boldsymbol{v}\mid\mathbf{x}) = \operatorname{E}(\mathbf{q}\boldsymbol{U}\mid\mathbf{x}) + \operatorname{E}(\boldsymbol{v}\mid\mathbf{x}) = \mathbf{q}\operatorname{E}(\boldsymbol{U}\mid\mathbf{x}) + \boldsymbol{v},$$

y

$$\operatorname{Var}(\mathbf{Q}\boldsymbol{U}+\boldsymbol{v}\mid\mathbf{x}) = \operatorname{Var}(\mathbf{q}\boldsymbol{U}\mid\mathbf{x}) = \mathbf{q}\operatorname{Var}(\boldsymbol{U}\mid\mathbf{x})\,\mathbf{q}';$$

además

$$\operatorname{Cov}(\mathbf{Q}\boldsymbol{U}+\boldsymbol{v},\mathbf{R}\boldsymbol{U}+\boldsymbol{w}\mid\mathbf{x}) = \operatorname{Cov}(\mathbf{q}\boldsymbol{U},\mathbf{r}\boldsymbol{U}\mid\mathbf{x}) = \mathbf{q}\operatorname{Var}(\boldsymbol{U}\mid\mathbf{x})\,\mathbf{r}'$$

Ejercicio 39. Denotemos $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ por $\underset{[k\times N]}{\mathbf{A}}$. Sabiendo que $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}\boldsymbol{U}$, calcule de nuevo la expresión de $\operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right)$ empleando las propiedades de la esperanza y la varianza de vectores de las notas anteriores.

⇑ **Eficiencia del estimador MCO $\widehat{\boldsymbol{\beta}}\,\big|\,\mathbf{x}$: Tª de Gauss-Markov** **28**

Con los supuestos 1 a 4,

$\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ **eficiente** entre estimadores *lineales e insesgados*

es decir, para cualquier estimador lineal insesgado $\widetilde{\boldsymbol{\beta}}_{|\,\mathbf{x}}$

$$\operatorname{Var}\left(\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) \geq \operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right)$$

en sentido matricial[a]

Entonces se dice ELIO (BLUE en inglés).

[a] La matriz $\left[\operatorname{Var}\left(\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) - \operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right)\right]$ es definida positiva

De hecho el Tª arriba mencionado implica que

$$\operatorname{Var}\left(\widetilde{\beta_j}\,\middle|\,\mathbf{x}\right) \geq \operatorname{Var}\left(\widehat{\beta_j}\,\middle|\,\mathbf{x}\right) \quad \text{para} \quad j = 1, \ldots, k.$$

es decir, la relación es cierta para cada uno de los estimadores de cada uno de los parámetros individuales.

Teorema 5.1 (Gauss-Markov). *Sea $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ el estimador MCO de $\boldsymbol{\beta}$, y sea $\widetilde{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ otro estimador* lineal *e* insesgado *de $\boldsymbol{\beta}$; entonces bajo los supuestos 1 a 4, para cualquier $\underset{[k\times 1]}{\boldsymbol{v}}$ se verifica que* $\operatorname{Var}\left(\boldsymbol{v}'\,\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) \geq \operatorname{Var}\left(\boldsymbol{v}'\,\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right)$

Demostración. Puesto que $\widetilde{\boldsymbol{\beta}}_{|\,\mathbf{x}} = \mathbf{f}\boldsymbol{Y}$ es un estimador insesgado, $\operatorname{E}\left(\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) = \mathbf{f}\cdot\operatorname{E}(\boldsymbol{Y}|\,\mathbf{x}) = \mathbf{f}\,\mathbf{x}\boldsymbol{\beta} = \boldsymbol{\beta}$. Por tanto la insesgadez implica necesariamente que $\mathbf{fx} = \mathbf{I}$. Sea $\mathbf{g} = \mathbf{a}+\mathbf{f}$, donde $\mathbf{a} = (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'$; entonces $\mathbf{gx} = \mathbf{0}$ (y por tanto $\mathbf{g}\,\mathbf{a}' = \underset{[k\times k]}{\mathbf{0}}$ y, trasponiendo, $\mathbf{a}\,\mathbf{g}' = \underset{[k\times k]}{\mathbf{0}'}$). Puesto que $\operatorname{Var}(\boldsymbol{Y}|\,\mathbf{x}) = \operatorname{Var}(\boldsymbol{U}\,|\,\mathbf{x}) = \sigma^2\,\mathbf{I}$ se deduce que:

$$\operatorname{Var}\left(\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) = \mathbf{f}\operatorname{Var}(\boldsymbol{Y}|\,\mathbf{x})\,\mathbf{f}' = \sigma^2\left[\mathbf{a}+\mathbf{g}\right]\left[\mathbf{a}'+\mathbf{g}'\right] = \sigma^2\left[\mathbf{a}\,\mathbf{a}'+\mathbf{a}\,\mathbf{g}'+\mathbf{g}\,\mathbf{a}'+\mathbf{g}\,\mathbf{g}'\right] = \underbrace{\sigma^2(\mathbf{x}'\,\mathbf{x})^{-1}}_{\operatorname{Var}(\widehat{\beta_j}\,|\,\mathbf{x})}+\sigma^2\,\mathbf{g}\,\mathbf{g}',$$

donde $\mathbf{g}\,\mathbf{g}'$ es semi-definida positiva.

Por tanto, para cualquier vector $\boldsymbol{v}$ de orden k

$$\begin{aligned}\operatorname{Var}\left(\boldsymbol{v}'\,\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) &= \boldsymbol{v}'\operatorname{Var}\left(\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right)\boldsymbol{v}' \\ &= \operatorname{Var}\left(\boldsymbol{v}'\,\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) + \sigma^2\boldsymbol{v}'\,\mathbf{g}\,\mathbf{g}'\,\boldsymbol{v};\end{aligned}$$

que implica

$$\operatorname{Var}\left(\boldsymbol{v}'\,\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) \geq \operatorname{Var}\left(\boldsymbol{v}'\,\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right).$$

□

Ejercicio 40. En particular ¿qué implica el Teorema de Gauss-Markov para el caso particular de un vector $\boldsymbol{v} = \begin{bmatrix}0 & \ldots & 0 & 1 & 0 & \ldots & 0\end{bmatrix}$; es decir, con un 1 en la posición j-ésima y ceros en el resto?

5.3. Momentos de los valores ajustados $\widehat{\boldsymbol{y}}_{|\,\mathbf{x}}$ y de los errores $\widehat{\boldsymbol{e}}_{|\,\mathbf{x}}$

Recuerde las definiciones que aparecen al final de la Subsección 4.2.1 en la página˜25; y resuelva el siguiente ejercicio:

Ejercicio 41. Denotemos $\mathbf{x}\,(\mathbf{x}'\,\mathbf{x})^{-1}\,\mathbf{x}'$ por $\mathbf{p}$.

Nótese que

$$\mathbf{p} \equiv \mathbf{x}\,(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}' = \mathbf{xa}\,.$$

Verifique que $\mathbf{px} = \mathbf{x}$. Demuestre además que $\mathbf{p}' = \mathbf{p}$ y que $\mathbf{pp} = \mathbf{p}$; es decir, que $\mathbf{p}$ es simétrica e idempotente.

⇑ Primeros momentos de los valores ajustados por MCO **29**

Denotemos $\mathbf{x}\,(\mathbf{x}'\,\mathbf{x})^{-1}\,\mathbf{x}'$ por $\mathbf{p}$, entonces

$$\begin{aligned}\widehat{\boldsymbol{y}}_{|\,\mathbf{x}} &= \mathbf{x}\,\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} = \mathbf{x}\left[\boldsymbol{\beta} + (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{U}\right] \\ &= \mathbf{x}\boldsymbol{\beta} + \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{U} = \mathbf{x}\boldsymbol{\beta} + \mathbf{p}\boldsymbol{U} \qquad \boxed{\text{T47}}\end{aligned}$$

así pues:

$$\mathrm{E}(\,\widehat{\boldsymbol{y}}\mid\mathbf{x}) = \mathbf{x}\boldsymbol{\beta} \qquad \text{por el Supuesto 2 } \boxed{\text{T7}}$$

$$\begin{aligned}\mathrm{Var}(\,\widehat{\boldsymbol{y}}\mid\mathbf{x}) &= \mathbf{p}\,\mathrm{Var}(\,\boldsymbol{U}\mid\mathbf{x})\,\mathbf{p}' \\ &= \sigma^2\,\mathbf{p}\,\mathbf{p}' = \sigma^2\,\mathbf{p} \qquad \text{por el Supuesto 3 } \boxed{\text{T9}}\end{aligned}$$

Donde hemos empleado los resultados de la Nota 11 en la página˜35.

Nótese que la matriz de varianzas y covarianzas es (en general) una matriz “llena” (al contrario que la matriz identidad) por tanto los valores ajustados son autocorrelados y heterocedásticos.

Ejercicio 42. Denotemos $\mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'$ por $\mathbf{m}$.
Nótese que

$$\mathbf{m} \equiv \mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}' = \mathbf{I} - \mathbf{p} = \mathbf{I} - \mathbf{x}\mathbf{a}\,.$$

Verifique que $\mathbf{m}\mathbf{x} = \mathbf{0}$, y que $\mathbf{a}\mathbf{m} = \mathbf{0}$. Demuestre además que $\mathbf{m} = \mathbf{m}'$ y que $\mathbf{m}\mathbf{m} = \mathbf{m}$; es decir, que $\mathbf{m}$ es simétrica e idempotente.

⇑ Primeros momentos de los errores MCO **30**

Denotemos $\mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'$ por $\mathbf{m}$, entonces

$$\begin{aligned}\widehat{\boldsymbol{e}}_{|\,\mathbf{x}} = \boldsymbol{Y}_{|\,\mathbf{x}} - \widehat{\boldsymbol{y}}_{|\,\mathbf{x}} &= \left[\mathbf{x}\boldsymbol{\beta} + \boldsymbol{U}\right] - \mathbf{x}\left[\boldsymbol{\beta} + (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{U}\right] \\ &= \left[\mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\right]\boldsymbol{U} = \mathbf{m}\boldsymbol{U} \qquad \boxed{\text{T47}}\end{aligned}$$

por tanto,

$$\mathrm{E}(\,\widehat{\boldsymbol{e}}\mid\mathbf{x}) = \mathbf{0} \qquad \text{por el Supuesto 2 } \boxed{\text{T7}}$$

y

$$\begin{aligned}\mathrm{Var}(\,\widehat{\boldsymbol{e}}\mid\mathbf{x}) &= \mathbf{m}\,\mathrm{Var}(\,\boldsymbol{U}\mid\mathbf{x})\,\mathbf{m}' \\ &= \sigma^2\,\mathbf{m}\,\mathbf{m}' = \sigma^2\,\mathbf{m} \qquad \text{por Supuesto 3 } \boxed{\text{T9}}\end{aligned}$$

Nótese que la matriz de varianzas y covarianzas es (en general) una matriz “llena” (al contrario que la matriz identidad) por tanto los valores ajustados son autocorrelados y heterocedásticos.

Ejercicio 43. Demuestre que el estimador de la suma residual es $\widehat{SRC}_{|\,\mathbf{x}} = \boldsymbol{U}'\,\mathbf{m}\,\boldsymbol{U}$.

6. Distribución de los estimadores MCO bajo la hipótesis de Normalidad

- Secciones 4.1 y 4.2 de Wooldridge (2006)
- Apéndice E3 de Wooldridge (2006)

Nota 13. Distribución conjunta normal implica

1. distribución queda completamente determinada por el vector de esperanzas y la matriz de varianzas y covarianzas (lo que ya hemos calculado).
2. Correlación cero implica independencia
3. Cualquier transformación lineal también es conjuntamente normal

6.1. Quinto supuesto del Modelo Clásico de Regresión Lineal

Supuesto 5: Distribución Normal de las perturbaciones **31**

Para conocer la distribución completa necesitamos un supuesto más sobre la distribución conjunta de $\boldsymbol{U}$:

$$\boldsymbol{U}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\mathbf{0}\,,\,\sigma^2\,\mathbf{I}\right) \quad \Rightarrow \quad \boldsymbol{Y}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\mathbf{x}\boldsymbol{\beta}\,,\,\sigma^2\,\mathbf{I}\right)$$

donde $\mathbf{I}$ es la matriz identidad.
Puesto que

$$\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} = \boldsymbol{\beta} + (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{U} = \boldsymbol{\beta} + \mathbf{A}\boldsymbol{U}$$

es función lineal de $\boldsymbol{U}$, entonces $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ tiene distribución normal multivariante.

$$\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\boldsymbol{\beta}\,,\,\sigma^2(\mathbf{x}'\,\mathbf{x})^{-1}\right)$$

$\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\boldsymbol{\beta}\,,\,\sigma^2(\mathbf{x}'\,\mathbf{x})^{-1}\right)$ es decir (*y si el modelo tiene término constante*)

$$\begin{pmatrix} \widehat{\beta_1} \\ \widehat{\beta_2} \\ \vdots \\ \widehat{\beta_k} \end{pmatrix}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \sigma^2 \begin{pmatrix} \mathbf{1}'\,\mathbf{1} & \mathbf{1}'\,\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \mathbf{1}'\,\boldsymbol{x}_{\blacktriangledown k} \\ \boldsymbol{x}_{\blacktriangledown 2}{}'\,\mathbf{1} & \boldsymbol{x}_{\blacktriangledown 2}{}'\,\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \boldsymbol{x}_{\blacktriangledown 2}{}'\,\boldsymbol{x}_{\blacktriangledown k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_{\blacktriangledown k}{}'\,\mathbf{1} & \boldsymbol{x}_{\blacktriangledown k}{}'\,\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \boldsymbol{x}_{\blacktriangledown k}{}'\,\boldsymbol{x}_{\blacktriangledown k} \end{pmatrix}^{-1} \right)$$

Distribución del estimador MCO $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ **32**

Así pues,

$$\widehat{\beta_j}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\beta_j\,,\,\sigma^2\left[(\mathbf{x}'\,\mathbf{x})^{-1}\right]_{jj}\right)$$

donde $\left[(\mathbf{x}'\,\mathbf{x})^{-1}\right]_{jj}$ es el elemento (j,j) de la matriz $(\mathbf{x}'\,\mathbf{x})^{-1}$.

y

$$\frac{\widehat{\beta_j}_{|\,\mathbf{x}} - \beta_j}{\mathrm{Dt}\left(\widehat{\beta_j}\;\middle|\;\mathbf{x}\right)} \sim \mathrm{N}\left(0\,,\,1\right)$$

(a partir de ahora también denotaremos los estadísticos condicionados, i.e., $\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}$ o $\widehat{\boldsymbol{e}}_{|\,\mathbf{x}}$ sencillamente como $\widehat{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{e}}$)

Modelo 2. [Modelo Lineal Simple.] De la transparencia anterior y de 5.2 en la página~34 podemos afirmar que bajo todos los supuestos del MLS

$$\widehat{a}_{|\,\mathbf{x}} \sim \mathrm{N}\left(a\,,\,\frac{\sigma^2\overline{\boldsymbol{x}^2}}{N\cdot s_{\boldsymbol{x}}^2}\right) \qquad \text{y} \qquad \widehat{b}_{|\,\mathbf{x}} \sim \mathrm{N}\left(b\,,\,\frac{\sigma^2}{N\cdot s_{\boldsymbol{x}}^2}\right). \tag{6.1}$$

Distribución de los estimadores de valores ajustados y residuos **33**

Ambos estimadores son transformaciones lineales de $\boldsymbol{U} \sim \mathrm{N}$; y vistos sus primeros momentos $\boxed{\text{T29}}$ y $\boxed{\text{T30}}$:

$$\widehat{\boldsymbol{y}}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\mathbf{x}\boldsymbol{\beta}\;,\;\sigma^2\,\mathbf{p}\right) \qquad \text{pues } \widehat{\boldsymbol{y}}_{|\,\mathbf{x}} = \mathbf{x}\boldsymbol{\beta} + \mathbf{p}\boldsymbol{U}$$

$$\widehat{\boldsymbol{e}}_{|\,\mathbf{x}} \sim \mathrm{N}\left(\mathbf{0}\;,\;\sigma^2\,\mathbf{m}\right) \qquad \text{pues } \widehat{\boldsymbol{e}}_{|\,\mathbf{x}} = \mathbf{m}\boldsymbol{U}$$

donde $\quad \mathbf{p} = \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'; \qquad$ y $\qquad \mathbf{m} = \mathbf{I} - \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'$

6.2. Estimación de la varianza residual y la matriz de covarianzas

Nota 14. Llamamos "traza" a la suma de los elementos de la diagonal de una matriz.

El operador traza es un operador lineal con la siguiente propiedad: Sean **a** y **b** dos matrices cuadradas, entonces

$$\text{traza}\,(\mathbf{ab}) = \text{traza}\,(\mathbf{ba})$$

Proposición 6.1. $\text{traza}\,(\mathbf{m}) = N - k$;

Demostración.

$$\begin{aligned}
\text{traza}\,(\mathbf{m}) &= \text{traza}\left(\underset{[N\times N]}{\mathbf{I}} - \underset{[N\times N]}{\mathbf{p}}\right) && \text{puesto que } \mathbf{m} \equiv \mathbf{I} - \mathbf{p}\\
&= \text{traza}\,(\mathbf{I}) - \text{traza}\,(\mathbf{p}) && \text{puesto que traza es lineal}\\
&= N - \text{traza}\,(\mathbf{p})
\end{aligned}$$

y

$$\begin{aligned}
\text{traza}\,(\mathbf{p}) &= \text{traza}\left(\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\right) && \text{puesto que } \mathbf{p} \equiv \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' = \mathbf{xa}\\
&= \text{traza}\left((\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}\right) && \text{puesto que } \text{traza}\,(\mathbf{xa}) = \text{traza}\,(\mathbf{ax})\\
&= \text{traza}\left(\underset{[k\times k]}{\mathbf{I}}\right) = k
\end{aligned}$$

Por tanto $\text{traza}\,(\mathbf{m}) = N - k$. □

Proposición 6.2. $\text{E}\left(\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} \mid \mathbf{x}\right) = (N-k)\sigma^2$

Demostración. En $\boxed{\text{T30}}$ vimos que $\widehat{\boldsymbol{e}}_{|\mathbf{x}} = \mathbf{m}\boldsymbol{U}$; por tanto

$$\begin{aligned}
\text{E}\left(\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}} \,\middle|\, \mathbf{x}\right) &= \text{E}\left(\boldsymbol{U}'\,\mathbf{m}'\,\mathbf{m}\,\boldsymbol{U} \mid \mathbf{x}\right) = \text{E}\left(\boldsymbol{U}'\,\mathbf{m}\,\boldsymbol{U} \mid \mathbf{x}\right) && \text{por ser } \mathbf{m} \text{ idempotente}\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} m_{ij}\text{E}\left(U_iU_j \mid \mathbf{x}\right) && \text{pues el operador esperanza es lineal}\\
&= \sum_{i=1}^{N} m_{ii}\sigma^2 && \text{por el supuesto 3 } \boxed{\text{T9}}\\
&= \sigma^2\,\text{traza}\,(\mathbf{m}) = \sigma^2(N-k) && \text{por la } \textbf{Nota } 14 \text{ (Pág. 39) y } \textbf{Proposición } 6.1
\end{aligned}$$

□

Por tanto, $\widehat{\mathfrak{s}^2_{\widehat{e}}} \equiv \frac{\widehat{e}'\,\widehat{e}}{N-k}$ es un estimador insesgado de σ^2. Consecuentemente emplearemos como estimador de la matriz de varianzas y covarianzas la expresión (6.2) de más abajo.

⇑ **Estimación de la varianza residual** **34**

El parámetro σ^2 es desconocido $\boxed{\text{T9}}$

La cuasivarianza de $\widehat{\boldsymbol{e}}$

$$\widehat{\mathfrak{s}^2_{\widehat{e}}} \equiv \frac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N-k}$$

es un estimador insesgado de σ^2 puesto que

$$\text{E}\left(\widehat{\mathfrak{s}^2_{\widehat{e}}} \,\middle|\, \mathbf{x}\right) = \text{E}\left(\frac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N-k} \,\middle|\, \mathbf{x}\right) = \frac{\sigma^2(N-k)}{N-k} = \sigma^2$$

Estimador de la matriz de varianzas y covarianzas de $\widehat{\boldsymbol{\beta}}_{|\mathbf{x}}$

$$\widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}_{|\mathbf{x}}\right) = \widehat{\mathfrak{s}^2_{\widehat{e}}} \cdot (\mathbf{x}'\mathbf{x})^{-1} \tag{6.2}$$

Proposición 6.3. *Si una matriz cuadrada* **q** *es idempotente entonces* $\text{rango}\,(\mathbf{q}) = \text{traza}\,(\mathbf{q})$.

Demostración. (Demostración en Rao, 2002, pp. 28) □

Proposición 6.4. *Sea el vector* $\boldsymbol{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$, *y sea* $\mathbf{q}$ *una matriz simétrica e idempotente, entonces* $\boldsymbol{Z}'\mathbf{q}\boldsymbol{Z} \sim \chi^2_{(\mathrm{rango}(\mathbf{q}))}$.

Demostración. (Demostración en Mittelhammer, 1996, pp. 329) □

⇑ **Distribución cuando la varianza de $\boldsymbol{U}$ es desconocida** **35**

$$\frac{\widehat{\beta_j} - \beta_j}{\sqrt{\sigma^2\big((\mathbf{x}'\mathbf{x})^{-1}\big)_{jj}}} \sim \mathrm{N}(0, 1)$$

sustituyendo σ^2 por su estimador, $\widehat{s^2_{\hat{e}}}$, tenemos el estadístico $\mathcal{T}$ del parámetro j-ésimo:

$$\frac{\widehat{\beta_j} - \beta_j}{\sqrt{\widehat{s^2_{\hat{e}}}\big((\mathbf{x}'\mathbf{x})^{-1}\big)_{jj}}} = \frac{\widehat{\beta_j} - \beta_j}{\sqrt{\Big[\widehat{\mathrm{Var}}\big(\widehat{\boldsymbol{\beta}}\big)\Big]_{jj}}} \equiv \mathcal{T}^j \sim t_{N-k} \tag{6.3}$$

Proposición 6.5. $\frac{N-k}{\sigma^2}\widehat{s^2_{\hat{e}}} = \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2} \sim \chi^2_{(N-k)}$

Demostración.

$$\begin{aligned}
\frac{N-k}{\sigma^2}\widehat{s^2_{\hat{e}}} = \frac{N-k}{\sigma^2}\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{N-k} &= \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\sigma^2} = \frac{1}{\sigma}\hat{\mathbf{e}}'\hat{\mathbf{e}}\frac{1}{\sigma} \\
&= \frac{1}{\sigma}\boldsymbol{U}'\mathbf{m}'\mathbf{m}\boldsymbol{U}\frac{1}{\sigma} && \text{ya que } \hat{\mathbf{e}} = \mathbf{m}\boldsymbol{U} \\
&= \frac{1}{\sigma}\boldsymbol{U}'\mathbf{m}\boldsymbol{U}\frac{1}{\sigma} \sim \chi^2_{(N-k)}
\end{aligned}$$

puesto que $\mathbf{m}$ es idempotente, $\boldsymbol{U}_{|\mathbf{x}} \sim \mathrm{N}\big(\mathbf{0}, \sigma^2\mathbf{I}\big)$, por las **proposiciones** 6.3 y 6.4 y la **Proposición** 6.1 en la página anterior. □

Ejercicio 44. Teniendo en cuenta que si una v.a. $X \sim \chi^2_{N-k}$ entonces $\mathrm{E}(X) = N-k$ y $\mathrm{Var}(X) = 2(N-k)$, y puesto que $\widehat{s^2_{\hat{e}}}$ es una variable aleatoria χ^2_{N-k} multiplicada por $\frac{\sigma^2}{N-k}$; calcule la esperanza y la varianza de $\widehat{s^2_{\hat{e}}}$

Proposición 6.6. *Las variables aleatorias* $\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)_{|\mathbf{x}}$ *y* $\hat{\mathbf{e}}_{|\mathbf{x}}$ *son independientes.*

Demostración. Puesto que $\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)_{|\mathbf{x}} = \mathbf{a}\boldsymbol{U}$ y $\hat{\mathbf{e}}_{|\mathbf{x}} = \mathbf{m}\boldsymbol{U}$, ambas variables son transformaciones lineales de $\boldsymbol{U}$, y por tanto ambas tienen distribución conjunta normal condicionada a $\mathbf{x}$ (**Nota** 13 en la página˜37)

Basta, por tanto, demostrar que ambas variables tienen covarianza nula

$$\begin{aligned}
\mathrm{Cov}(\mathbf{a}\boldsymbol{U}, \mathbf{m}\boldsymbol{U} \mid \mathbf{x}) &= \mathbf{a}\,\mathrm{Var}(\boldsymbol{U} \mid \mathbf{x})\,\mathbf{m}' && \text{por el supuesto 2 y la **Nota** 12 (Página 35)} \\
&= \mathbf{a}\,\sigma^2\,\mathbf{I}\,\mathbf{m}' && \text{por el supuesto 3} \\
&= \sigma^2\,\mathbf{a}\,\mathbf{m} = \sigma^2\,\mathbf{0} = \mathbf{0}
\end{aligned}$$

□

Nota 15. Si dos variables aleatorias X e Y son independientes, entonces transformaciones de ellas, $h(X)$ y $g(Y)$, también son independientes.

Proposición 6.7. *El estadístico* $\mathcal{T}^j$ *de distribuye como una* t *con* $N-k$ *grados de libertad, es decir,* $\mathcal{T}^j \sim t_{N-n}$

Demostración.

$$\frac{\widehat{\beta_j} - \beta_j}{\sqrt{\widehat{s^2_{\hat{e}}}\big((\mathbf{x}'\mathbf{x})^{-1}\big)_{jj}}} = \frac{\widehat{\beta_j} - \beta_j}{\sqrt{\sigma^2\big((\mathbf{x}'\mathbf{x})^{-1}\big)_{jj}}} \cdot \sqrt{\frac{\sigma^2}{\widehat{s^2_{\hat{e}}}}} = \frac{Z}{\sqrt{\frac{\widehat{s^2_{\hat{e}}}}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}/\sigma^2}{N-k}}}$$

donde la parte de numerador es función de $\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)_{|\,\mathbf{x}}$ y la del denominador es función de $\widehat{\boldsymbol{e}}_{|\,\mathbf{x}}$. Así pues, por la **Nota** 15 en la página anterior y la **Proposición** 6.6 en la página anterior el numerador y el denominador son independientes.

Además, en numerador tiene distribución $\mathrm{N}\left(0\,,\,1\right)$. Por tanto tenemos una $\mathrm{N}\left(0\,,\,1\right)$ dividida por la raíz cuadrada de un χ^2 dividida por sus grados de libertad; este cociente tiene distribución t de Student con $N-k$ grados de libertad. □

Ejemplo 45. [**continuación de "precio de las viviendas":**]

La inversa de $\mathbf{x'}\,\mathbf{x}$ es:

$$(\mathbf{x'}\,\mathbf{x})^{-1} = \begin{bmatrix} 9.1293e-01 & -4.4036e-04 \\ -4.4036e-04 & 2.3044e-07 \end{bmatrix};$$

así pues, las desviaciones típicas de $\widehat{a}$ y $\widehat{b}$ son (véase 5.2 en la página 34)

$$\mathrm{Dt}\left(\widehat{a}\right) = \sqrt{\sigma^2\cdot(9.1293e-01)} = \sqrt{\frac{\sigma^2\sum x_n^2}{N\sum(x_n-\overline{x})^2}}$$

$$\mathrm{Dt}\left(\widehat{b}\right) = \sqrt{\sigma^2\cdot(2.3044e-07)} = \sqrt{\frac{\sigma^2}{\sum(x_n-\overline{x})^2}}.$$

No conocemos $\sigma^2_{U_n}$; pero podemos sustituirla por la la cuasi-varianza:

$$\widehat{\mathrm{Dt}}\left(\widehat{a}\right) = \sqrt{(1522.8)\cdot(9.1293e-01)} = \sqrt{\frac{(1522.8)\sum x_n^2}{N\sum(x_n-\overline{x})^2}} = 37.285;$$

$$\widehat{\mathrm{Dt}}\left(\widehat{b}\right) = \sqrt{(1522.8)\cdot(2.3044e-07)} = \sqrt{\frac{(1522.8)}{\sum(x_n-\overline{x})^2}} = 0.01873$$

puesto que $\widehat{\mathfrak{s}^2_{\widehat{e}}} = \frac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N-n} = \frac{18273.6}{14-2} = 1522.8$.

Véase los resultados de estimación en el ejemplo del precio de las viviendas (página 17).

Por otra parte, $\widehat{\mathrm{Cov}}\left(\widehat{a},\widehat{b}\right) = (1522.8)*(-4.4036e-04) = \frac{-\widehat{\mathfrak{s}^2_{\widehat{e}}}\sum x_n}{N\sum(x_n-\overline{x})^2} = -0.671$ (véase 5.3 en la página 34).

6.3. Cota mínima de Cramér-Rao

⇑ Matriz de Información **36**

Función de verosimilitud

$$\ell(\boldsymbol{\theta};\boldsymbol{y},\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{y}-\mathbf{x}\boldsymbol{\beta}\right)'\left(\boldsymbol{y}-\mathbf{x}\boldsymbol{\beta}\right)\right] = f\left(\boldsymbol{y},\mathbf{x};\boldsymbol{\theta}\right);$$

donde $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta}\\ \sigma^2\end{bmatrix}$

Matriz de Información para $\boldsymbol{\theta}$

$$I(\boldsymbol{\theta}) = -\mathrm{E}\left(\frac{\partial^2\ln\ell(\boldsymbol{\theta};\boldsymbol{Y},\mathbf{X})}{\partial\,\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\,\middle|\,\mathbf{x}\right)$$

⇑ **Cota mínima de Cramér-Rao** 37

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\mathbf{x}'\mathbf{x}}{\sigma^2} & \mathbf{0} \\ \mathbf{0}' & \frac{N}{2\sigma^4} \end{bmatrix}$$

Cota mínima es la inversa de la Matriz de Información

$$I(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \sigma^2(\mathbf{x}'\mathbf{x})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{N} \end{bmatrix}$$

Matriz de varianzas y covarianzas de los estimadores MCO

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}, \widehat{s^2_e}} = \begin{bmatrix} \sigma^2(\mathbf{x}'\mathbf{x})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{N-k} \end{bmatrix}$$

$$I(\boldsymbol{\theta}) = -\operatorname{E}\left(\left[\begin{array}{cc} -\frac{\mathbf{x}'\mathbf{x}}{\sigma^2} & -\frac{\left[\mathbf{x}'\boldsymbol{Y} - \mathbf{x}'\mathbf{x}\boldsymbol{\beta}\right]}{\sigma^4} \\ -\frac{\left[\mathbf{x}'\boldsymbol{Y} - \mathbf{x}'\mathbf{x}\boldsymbol{\beta}\right]'}{\sigma^4} & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\left[\boldsymbol{Y} - \mathbf{x}\boldsymbol{\beta}\right]'\left[\boldsymbol{Y} - \mathbf{x}\boldsymbol{\beta}\right] \end{array} \right] \,\middle|\, \mathbf{x} \right)$$
$$= \begin{bmatrix} \frac{\mathbf{x}'\mathbf{x}}{\sigma^2} & \mathbf{0} \\ \mathbf{0}' & \frac{N}{2\sigma^4} \end{bmatrix}$$

1. La matriz de varianzas y covarianzas $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_{|\,\mathbf{x}}}$ alcanza la cota mínima de Cramér-Rao. Es decir es el estimador insesgado de mínima varianza (resultado más fuerte que Tª de Gauss-Markov)
2. La varianza del estimador $\widehat{s^2_e}$ no alcanza la cota mínima de Cramér-Rao. No obstante, no existe ningún estimador *insesgado* de σ^2 con varianza menor a $\frac{2\sigma^4}{N}$.

Ejercicio 46. Revise el ejercicio numérico nº1 del profesor José Alberto Mauricio
http://www.ucm.es/info/ecocuan/jam/ectr1/index.html#Material.

Ejercicio 47. Resuelva el ejercicio propuesto nº 1 del profesor José Alberto Mauricio.
http://www.ucm.es/info/ecocuan/jam/ectr1/index.html#Material.

Para los ejercicios prácticos con ordenador le puede ser útil

- El programa **gratuito** GRETL. (`http://gretl.sourceforge.net/gretl_espanol.html`)
  Tiene documentación en castellano
  - Guía del usuario
  - Guía de instrucciones

  También puede obtener los datos del libro de texto (Wooldridge, 2006) desde `http://gretl.sourceforge.net/gretl_data.html`
- la guia de Eviews del profesor José Alberto Mauricio (material extenso)
  (http://www.ucm.es/info/ecocuan/jam/ectr1/Ectr1-JAM-IntroEViews.pdf).

Ejercicio 48. Anscombe
GNU Gretl: ejemplo Anscombe

Ejercicio 49. Replique con el ordenador la práctica con ordenador propuesta por el profesor Miguel Jerez
http://www.ucm.es/info/ecocuan/mjm/ectr1mj/.

GNU Gretl MLG peso bbs

7. Estimación por máxima verosimilitud

función de verosimilitud vs función de densidad **38**

Los supuestos 1, 2, 3 y 5, implican que

$$\mathbf{Y} \mid \mathbf{x} \sim \mathrm{N}\left(\mathbf{x}\boldsymbol{\beta},\ \sigma^2 \underset{[N\times N]}{\mathbf{I}}\right)$$

por tanto, la función de densidad de Y dado $\mathbf{x}$ es

$$f(\mathbf{y} \mid \mathbf{x}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\,(\boldsymbol{y}-\mathbf{x}\boldsymbol{\beta})'\,(\boldsymbol{y}-\mathbf{x}\boldsymbol{\beta})\right]$$

donde los parámetros $\left(\boldsymbol{\beta},\ \sigma^2\right)$ son desconocidos.

Estimación por Máxima Verosimilitud **39**

Sustituyendo el vector desconocido $\left(\boldsymbol{\beta},\ \sigma^2\right)$ por un hipotético $\left(\widetilde{\boldsymbol{\beta}},\ \widetilde{\sigma}^2\right)$ y tomando logs[a] obtenemos *función de verosimilitud logarítmica*

$$\ln \ell(\widetilde{\boldsymbol{\beta}},\ \widetilde{\sigma}^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\widetilde{\sigma}^2) - \frac{1}{2\widetilde{\sigma}^2}(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})$$

Maximizando

$$\max_{\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2} \quad \ln \ell(\widetilde{\boldsymbol{\beta}},\ \widetilde{\sigma}^2)$$

obtenemos estimaciones *máximo verosímiles* de $\left(\boldsymbol{\beta},\ \sigma^2\right)$.

[a] transformación monótona

Estimación por Máxima Verosimilitud: derivación **40**

Cond. primer orden en maximización:

$$\frac{\partial \ln \ell(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)}{\partial \widetilde{\boldsymbol{\beta}}'} = 0 \quad \Longrightarrow \quad -\tfrac{1}{2\widetilde{\sigma}^2}\tfrac{\partial}{\partial \widetilde{\boldsymbol{\beta}}'}(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) = 0$$

$$\widetilde{\boldsymbol{\beta}}'_{MV} = (\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}'\,\boldsymbol{y}$$

$$\frac{\partial \ln \ell(\widetilde{\boldsymbol{\beta}},\widetilde{\sigma}^2)}{\partial \widetilde{\sigma}^2} = 0 \quad \Longrightarrow \quad -\tfrac{n}{2\widetilde{\sigma}^2} + \tfrac{1}{2\widetilde{\sigma}^4}(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) = 0$$

$$\widetilde{\sigma}^2_{MV} = \tfrac{\widehat{\boldsymbol{e}}'\,\widehat{\boldsymbol{e}}}{N} = s^2_{\widehat{e}} = \tfrac{N-k}{N}\mathfrak{s}^2_{\widehat{e}}$$

Por tanto:
la estimación de $\boldsymbol{\beta}$ coincide con el MCO
la estimación de σ^2 es sesgada

EJERCICIO 50.
(a) Calcule la esperanza de $\widetilde{\sigma}^2_{MV}$. ¿Es un estimador insesgado de σ^2?
(b) Calcule la varianza de $\widetilde{\sigma}^2_{MV}$
(c) Compare su resultado con la cota mínima de Cramér-Rao. Pero ¿es aplicable esta cota a este estimador?

8. Ejercicios

EJERCICIO 51. Demuestre que en el modelo de regresión simple $Y_n = a+bX_n+U_n$ el supuesto $\mathrm{E}(U_n \mid \mathbf{x}) = 0$ implica $\mathrm{E}(Y_n \mid \mathbf{x}) = a + bX_n$; donde los regresores son no-estocásticos, y Ues la perturbación aleatoria del modelo.

Ejercicio 52. (Consta de 5 apartados)
Sean los siguientes datos:

Empresa	y_i	x_i	$x_i y_i$	x_i^2
A	1	1	1	1
B	3	2	6	4
C	4	4	16	16
D	6	4	24	16
E	8	5	40	25
F	9	7	63	49
G	11	8	88	64
H	14	9	126	81
sumas	56	40	364	256

Cuadro 4:

donde y son beneficios, y x son gastos en formación de personal de una empresa.

Además se sabe que las varianzas y covarianzas muestrales son tales que:

$$N \cdot s_{\boldsymbol{y}}^2 = \sum (y_i - \overline{y})^2 = 132,$$

$$N \cdot s_{\boldsymbol{x}}^2 = \sum (x_i - \overline{x})^2 = 56,$$

$$N \cdot s_{\boldsymbol{x}\,\boldsymbol{y}} = \sum (x_i - \overline{x})(y_i - \overline{y}) = 84,$$

donde N es el tamaño muestral.

Suponga que se plantea el siguiente modelo

$$Y_i = a + bx_i + U_i,$$

donde U_i son otros factores que afectan a los beneficios distintos de sus gastos en formación (el término de error). Se sabe que la distribución conjunta de dichos factores es:

$$\boldsymbol{U} \sim N(\boldsymbol{0},\ \sigma^2\,\mathbf{I}),$$

donde $\mathbf{I}$ es una matriz identidad de orden 8, y σ^2 es la varianza de U_i, cuyo valor es desconocido.

(a) Estime por MCO los parámetros a y b del modelo.

(b) ¿Cuál es el beneficio esperado para una empresa que incurriera en unos gastos de formación de personal de 3?

(c) Calcule los residuos de la empresa E y F. ¿Que indica en este caso el signo de los residuos? La comparación de los residuos para estas empresas ¿contradice el hecho de que F tiene mayores beneficios que E? Justifique su respuesta.
(*Los siguientes apartados sólo tras haber estudiado el tema siguiente*)

(d) Estime por MCO un intervalo de confianza del 95 % para el parámetro b del modelo, sabiendo que la suma de los residuos al cuadrado es 6.

(e) Contraste la hipótesis de que "la pendiente del modelo es uno" frente a que "es menor que uno" con un nivel de significación del 10 %. ¿Cuál es el p-valor de la estimación de "dicha pendiente"?

9. Bibliografía

Gujarati, D. N. (2003). *Basic Econometrics*. McGraw-Hill, cuarta ed. ISBN 0-07-112342-3. International edition. 26, 33

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, New Jersey. ISBN 0-691-01018-8. 2, 3

Johnston, J. y Dinardo, J. (2001). *Métodos de Econometría*. Vicens Vives, Barcelona, España, primera ed. ISBN 84-316-6116-x. 25, 26

Luenberger, D. G. (1968). *Optimization by vector space methods*. Series in decision and control. John Wiley & Sons, Inc., New York. 3

Mittelhammer, R. C. (1996). *Mathematical Statistics for Economics and Business*. Springer-Verlag, New York, primera ed. ISBN 0-387-94587-3. 40

Novales, A. (1993). *Econometría*. McGraw-Hill, segunda ed. 2, 12, 25, 26

Novales, A. (1997). *Estadística y Econometría*. McGraw-Hill, Madrid, primera ed. ISBN 84-481-0798-5. 33

Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza Editorial, Madrid. ISBN 84-206-8695-6. 25

Ramanathan, R. (1998). *Introductory Econometrics with Applications*. Harcourt College Publisher, Orlando. 6, 16, 30

Rao, C. R. (2002). *Linear Statistical Inference and Its Applications*. Wiley series in probability and statistics. John Wiley & Sons, Inc., New York, segunda ed. ISBN 0-471-21875-8. 39

Spanos, A. (1999). *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data*. Cambridge University Press, Cambridge, UK. ISBN 0-521-42408-9. 3

Verbeek, M. (2004). *A Guide to Modern Econometrics*. John Wiley & Sons, Inc., segunda ed. 2

Wooldridge, J. M. (2006). *Introducción a la econometría. Un enfoque moderno*. Thomson Learning, Inc., segunda ed. 2, 3, 5, 12, 21, 25, 26, 32, 37, 42

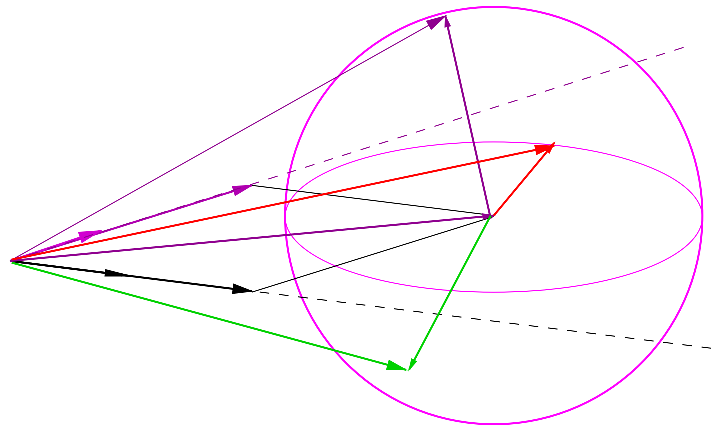
10. Trasparencias

Lista de Trasparencias

A. Geometría del modelo clásico de regresión lineal

Geometría del Modelo lineal **41**

$\mathbf{X} = \begin{bmatrix}\mathbf{1}, & \boldsymbol{X}_{\blacktriangledown 2}\end{bmatrix}$; $\quad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}$; $\quad \boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{U}$



Visión en 3D interactiva

Supuesto 2': Regresores no estocásticos **42**

Suponemos que realmente disponemos de una única realización de $\mathbf{X}$ que denotamos por $\mathbf{x}$.
Es decir, condicionamos a que

$$\mathbf{X} = \mathbf{x}$$

Bajo este supuesto, se mantiene que

$$\mathrm{E}(x_{ij}U_n) = 0 \qquad \text{para} \qquad n, i = 1, \ldots, N; \qquad \text{y} \qquad j = 1, \ldots, k.$$

Esto significa que, como en el caso general, los regresores son ortogonales a los términos de perturbación de *todas las observaciones*

$$\mathrm{E}(x_{ij}U_n) = x_{ij}\mathrm{E}(U_n) = 0 \quad \text{para todo} \quad i, n = 1, \ldots, N; \quad \text{y} \quad j = 1, \ldots, k.$$

por lo que

$$\mathrm{E}(\boldsymbol{x}_{i\triangleright}U_n) = \boldsymbol{x}_{i\triangleright} \cdot \mathrm{E}(U_n) = \boldsymbol{x}_{i\triangleright} \cdot 0 = \underset{[1\times k]}{\mathbf{0}} \quad \text{para todo} \quad i, n = 1, \ldots, N.$$

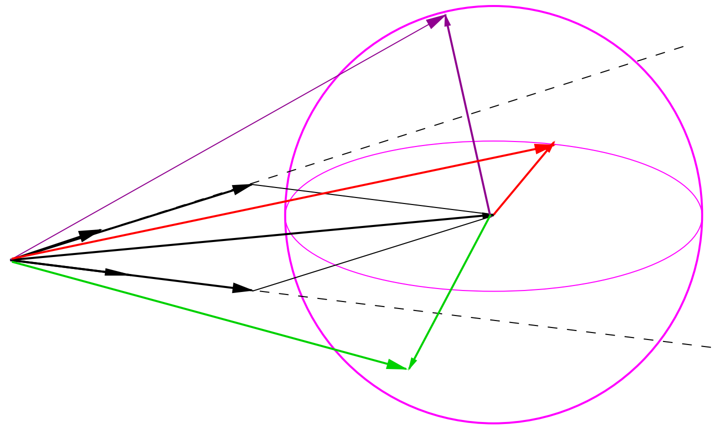
Y la correlación entre los regresores y las perturbaciones es cero, ya que

$$\begin{aligned}\mathrm{Cov}(U_n, x_{ij}) =& \mathrm{E}(x_{ij}U_n) - \mathrm{E}(x_{ij})\,\mathrm{E}(U_n) \\ =& x_{ij}\mathrm{E}(U_n) - x_{ij}\mathrm{E}(U_n) = 0\end{aligned}$$

es decir, regresores no estocásticos en un caso particular del caso general: Supuesto 2 $\boxed{\text{T7}}$ (véase también la Sección 2.2.2 en la página˜12, Página 12)

Geometría del Modelo lineal: regresores no estocásticos **43**

$\mathbf{x} = \left[\mathbf{1},\ \boldsymbol{x}_{\blacktriangledown 2}\right]; \qquad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}; \qquad \boldsymbol{Y} = \mathbf{x}\,\boldsymbol{\beta} + \boldsymbol{U}$



Visión en 3D interactiva

A.1. Geometría del estimador MCO

Estimación de la esperanza condicional: MCO **44**

Tenemos realizaciones de $\boldsymbol{Y}$ y $\mathbf{X}$; es decir, disponemos de

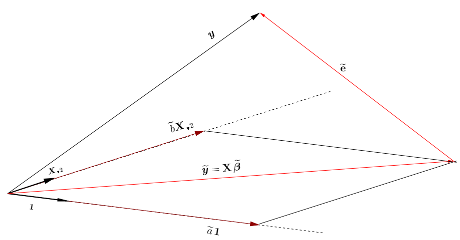
$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$$

y buscamos $\widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \widetilde{a} \\ \widetilde{b} \end{pmatrix}$ tales que

$$\boldsymbol{y} = \mathbf{x}\,\widetilde{\boldsymbol{\beta}} + \widetilde{\mathbf{e}}$$

y $\widetilde{\mathbf{e}}$ sea pequeño.

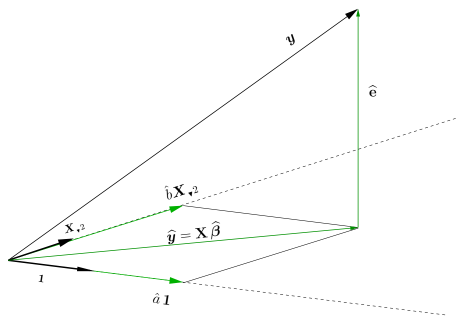
Estimación modelo lineal: geometría MCO **45**



$\mathbf{x} = \left[\mathbf{1},\ \boldsymbol{x}_{\blacktriangledown 2}\right]; \quad \widetilde{\boldsymbol{\beta}} = \begin{bmatrix} \widetilde{a} \\ \widetilde{b} \end{bmatrix}; \quad \boldsymbol{y} = \widetilde{\boldsymbol{y}} + \widetilde{\boldsymbol{e}}; \quad \widetilde{\boldsymbol{y}} = \mathbf{x}\widetilde{\boldsymbol{\beta}};$

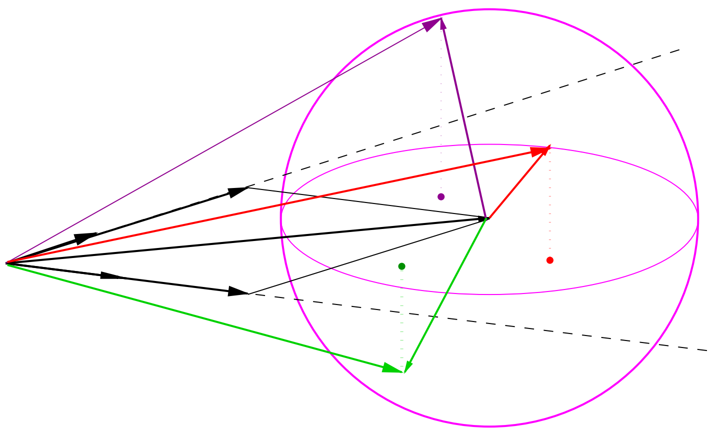
$\widetilde{\boldsymbol{e}} = \boldsymbol{y} - \widetilde{\boldsymbol{y}}$

⇑ Modelo lineal estimado: geometría MCO 46



$\mathbf{x} = \left[\mathbf{1},\ \boldsymbol{x}_{\blacktriangledown 2}\right];\quad \widehat{\boldsymbol{\beta}} = \left[\begin{array}{c}\widehat{a}\\ \widehat{b}\end{array}\right];\quad \boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}};\quad \widehat{\boldsymbol{y}} = \mathbf{x}\widehat{\boldsymbol{\beta}};$
$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$

⇑ Geometría del estimador 47



Visión en 3D interactiva

B. Derivación tradicional de las Ecuaciones Normales

⇑ Mínimos cuadrados ordinarios: Ecuaciones normales (Tradicional) 48

$$SRC(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{y}'\,\boldsymbol{y} - 2\widetilde{\boldsymbol{\beta}}'\,\mathbf{x}'\,\boldsymbol{y} + \widetilde{\boldsymbol{\beta}}'\,\mathbf{x}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}$$

Buscamos un vector $\widehat{\boldsymbol{\beta}}$ que minimice SRC

$$\text{mín}_{\widehat{\boldsymbol{\beta}}}\quad SRC(\widehat{\boldsymbol{\beta}})$$

$$\frac{\partial SRC(\widehat{\boldsymbol{\beta}})}{\partial\,\widehat{\boldsymbol{\beta}}} = 0;\qquad -2\,\mathbf{x}'\,\boldsymbol{y} + 2\,\mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}} = 0$$

con lo que obtenemos las ecuaciones normales

$$\mathbf{x}'\,\boldsymbol{y} = \mathbf{x}'\,\mathbf{x}\,\widehat{\boldsymbol{\beta}} \tag{B.1}$$

Estimación MCO es la solución a dichas ecuaciones

$$\begin{aligned}
SRC(\widetilde{\boldsymbol{\beta}}) =&(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}})\\
=&\left(\boldsymbol{y}'-\widetilde{\boldsymbol{\beta}}'\mathbf{x}'\right)(\mathbf{y}-\mathbf{x}\widetilde{\boldsymbol{\beta}}) && \text{puesto que } (\mathbf{x}\widetilde{\boldsymbol{\beta}})'=\widetilde{\boldsymbol{\beta}}'\mathbf{x}'\\
=&\boldsymbol{y}'\,\boldsymbol{y}-\widetilde{\boldsymbol{\beta}}'\mathbf{x}'\,\boldsymbol{y}-\boldsymbol{y}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}+\widetilde{\boldsymbol{\beta}}'\mathbf{x}'\,\mathbf{x}\widetilde{\boldsymbol{\beta}}\\
=&\boldsymbol{y}'\,\boldsymbol{y}-2\boldsymbol{y}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}+\widetilde{\boldsymbol{\beta}}'\,\mathbf{x}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}
\end{aligned}$$

ya que el escalar $\widetilde{\boldsymbol{\beta}}'\mathbf{x}'\,\boldsymbol{y}$ es igual a su traspuesta $\boldsymbol{y}'\,\mathbf{x}\widetilde{\boldsymbol{\beta}}$ (por ser escalar)

Renombremos algunos términos... por una parte definimos $\boldsymbol{a}\equiv\boldsymbol{y}'\,\mathbf{x}$ y por otra $\mathbf{c}\equiv\mathbf{x}'\,\mathbf{x}$, entonces

$$SRC(\widetilde{\boldsymbol{\beta}})=\boldsymbol{y}'\,\boldsymbol{y}-2\,\boldsymbol{a}\,\widetilde{\beta}+\widetilde{\boldsymbol{\beta}}'\,\mathbf{c}\,\widetilde{\boldsymbol{\beta}}\,.$$

Puesto que $\boldsymbol{y}'\,\boldsymbol{y}$ no depende de $\widetilde{\boldsymbol{\beta}}$ la diferencial de $SRC(\widetilde{\boldsymbol{\beta}})$ respecto de $\widetilde{\boldsymbol{\beta}}$ es

$$\begin{aligned}
\frac{\partial SRC(\widetilde{\boldsymbol{\beta}})}{\partial\,\widetilde{\boldsymbol{\beta}}} &=-\,2\,\boldsymbol{a}+2\,\mathbf{c}\,\widetilde{\boldsymbol{\beta}} && \text{por las propiedades de derivación matricial}\\
&=-\,2\,\mathbf{x}'\,\boldsymbol{y}+2\,\mathbf{x}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}} && \text{sustituyendo } \boldsymbol{a} \text{ y } \mathbf{c};
\end{aligned}$$

que igualando a cero nos da

$$-2\,\mathbf{x}'\,\boldsymbol{y}+2\,\mathbf{x}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}=0 \quad\Rightarrow\quad \mathbf{x}'\,\mathbf{x}\,\widetilde{\boldsymbol{\beta}}=\mathbf{x}'\,\boldsymbol{y}$$

Las condiciones de segundo orden son:

$$\frac{\partial SRC(\widetilde{\boldsymbol{\beta}})}{\partial\,\widetilde{\boldsymbol{\beta}}\,\partial\widetilde{\boldsymbol{\beta}}'}=2\,\mathbf{x}'\,\mathbf{x} \quad \text{que es una matriz definida positiva.}$$

C. Caso General

Sean Y_n, y $\boldsymbol{X^*}_{n\triangleright}\equiv\left[X_{n2},\;\ldots,X_{nk}\right]$ con matriz de varianzas y covarianzas

$$\operatorname{Var}\left(\left[Y_n,\quad \boldsymbol{X^*}_{n\triangleright}\right]\right)=\begin{bmatrix}\sigma^2_{Y_n} & \boldsymbol{\sigma}_{Y_n\boldsymbol{X^*}_{n\triangleright}}\\ \boldsymbol{\sigma}_{\boldsymbol{X^*}'_{n\triangleright}Y_n} & \boldsymbol{\Sigma}_{\boldsymbol{X^*}_{n\triangleright}}\end{bmatrix}$$

entonces siempre podemos encontrar unos parámetros β_1 y $\boldsymbol{\beta^*}=\left[\beta_2,\;\ldots,\beta_k\right]$, tales que

$$Y_n=\beta_1+\boldsymbol{X^*}_{n\triangleright}\,\boldsymbol{\beta^*}+U_n$$

donde $\operatorname{E}(U_n)=0$, y $\operatorname{Var}(U_n)=\boldsymbol{\sigma}_{Y_n\boldsymbol{X^*}_{n\triangleright}}\,\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X^*}_{n\triangleright}}\,\boldsymbol{\sigma}_{\boldsymbol{X^*}'_{n\triangleright}Y_n}$

Dichos parámetros resultan ser

$$\boldsymbol{\beta^*}=\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X^*}_{n\triangleright}}\,\boldsymbol{\sigma}_{\boldsymbol{X^*}'_{n\triangleright}Y_n}; \tag{C.1}$$

(es decir, las covarianzas pre-multiplicadas por la inversa de matriz de varianzas de los regresores) y

$$\beta_1=\operatorname{E}(Y_n)-\boldsymbol{\beta^*}'\operatorname{E}(\boldsymbol{X^*}_{n\triangleright})\,. \tag{C.2}$$

Estos parámetros son la solución a las ecuaciones normales

$$\operatorname{E}\left(\mathbf{X}'\,\boldsymbol{Y}\right)=\operatorname{E}\left(\mathbf{X}'\,\mathbf{X}\right)\begin{pmatrix}\beta_1\\ \boldsymbol{\beta^*}\end{pmatrix}$$

donde la primera columna de $\mathbf{X}$ está exclusivamente compuesta por unos.

Nótese como los parámetros a y b de la Ecuación (3.6) en la página˜15 son un caso particular, donde $a=\beta_1$ y $b=\beta_2$.

Llamamos a $\beta_1+\boldsymbol{X^*}_{n\triangleright}\,\boldsymbol{\beta^*}$ el *mejor predictor lineal* de Y_n dado $\boldsymbol{X^*}'_{n\triangleright}$; puesto que se puede demostrar que β_1 y $\boldsymbol{\beta^*}$ son los valores de b_1 y $\boldsymbol{b^*}$ que minimizan

$$\operatorname{E}\left(\left[Y_n-b_1-\boldsymbol{X^*}'_{n\triangleright}\,\boldsymbol{b^*}\right]^2\right)$$

En este contexto, llamamos a $U_n=Y_n-\left[\beta_1+\boldsymbol{X^*}_{n\triangleright}\,\boldsymbol{\beta^*}\right]$ el *error de predicción.*

Podemos estimar por MCO los parámetros desconocidos, β_1 y $\boldsymbol{\beta^*}$, sustituyendo, en las expresiones anteriores, los momentos poblacionales por sus equivalentes muestrales (véase la Subsección D en la página siguiente). Pero, puesto que aquí no estamos imponiendo las restricciones del *Modelo Clásico de Regresión Lineal*, no podemos, siquiera, conocer la esperanza del estimador. Para ello es necesario especificar algo más sobre la relación entre $\boldsymbol{X^*}_{n\triangleright}$ e Y_n.

C.1. Modelo Clásico de Regresión Lineal General

El modelo lineal general es más restrictivo precisamente es este sentido; puesto que supone que la esperanza condicional $\mathrm{E}(Y_n \mid \boldsymbol{X^*}_{n\triangleright})$ sea función lineal de $\boldsymbol{X^*}_{n\triangleright}$.

Bajo esta hipótesis clásica el predictor lineal de más arriba se convierte en el *mejor predictor* posible en el sentido de que

$$\mathrm{E}\left(\left[Y_n - \mathrm{E}(Y_n \mid \boldsymbol{X^*}_{n\triangleright})\right]^2\right) \leq \mathrm{E}\left(\left[Y_n - g(\boldsymbol{X^*}_{n\triangleright})\right]^2\right)$$

para cualquier función $g(\cdot)$.

C.1.1. Ecuaciones normales en el Modelo Lineal General

Las matrices y vectores de las ecuaciones normales $\mathbf{x}'\boldsymbol{y} = \mathbf{x}'\mathbf{x}\widehat{\boldsymbol{\beta}}$ en el caso general (k regresores) quedan del siguiente modo

$$\underset{[k\times k]}{\mathbf{x}'\mathbf{x}} = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \mathbf{1}'\boldsymbol{x}_{\blacktriangledown k} \\ \boldsymbol{x}_{\blacktriangledown 2}'\mathbf{1} & \boldsymbol{x}_{\blacktriangledown 2}'\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \boldsymbol{x}_{\blacktriangledown 2}'\boldsymbol{x}_{\blacktriangledown k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_{\blacktriangledown k}'\mathbf{1} & \boldsymbol{x}_{\blacktriangledown k}'\boldsymbol{x}_{\blacktriangledown 2} & \cdots & \boldsymbol{x}_{\blacktriangledown k}'\boldsymbol{x}_{\blacktriangledown k} \end{pmatrix}$$

donde cada elemento de la matriz $\mathbf{x}'\mathbf{x}$ es de la forma

$$\boldsymbol{x}_{\blacktriangledown i}'\boldsymbol{x}_{\blacktriangledown j} = \begin{pmatrix} x_{1i} & x_{2i} & \cdots & x_{Ni} \end{pmatrix} \begin{pmatrix} x_{1j} \\ x_{2j} \\ \cdots \\ x_{Nj} \end{pmatrix} = \sum_{n=1}^{N} x_{ni}x_{nj}$$

Además, $\mathbf{1}'\mathbf{1} = N$ y $\mathbf{1}'\boldsymbol{x}_{\blacktriangledown i} = \sum_{n=1}^{N} x_{ni}$. Por otra parte, el vector $\mathbf{x}'\boldsymbol{y}$ es de la forma

$$\underset{[N\times 1]}{\mathbf{x}'\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{x}_{\blacktriangledown 1}'\boldsymbol{y} \\ \boldsymbol{x}_{\blacktriangledown 2}'\boldsymbol{y} \\ \vdots \\ \boldsymbol{x}_{\blacktriangledown k}'\boldsymbol{y} \end{pmatrix} \quad \text{donde cada elemento es} \quad \boldsymbol{x}_{\blacktriangledown i}'\boldsymbol{y} = \begin{pmatrix} x_{1i} & x_{2i} & \cdots & x_{Ni} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{pmatrix} = \sum_{n=1}^{N} x_{ni}y_n$$

D. Una expresión alternativa de las estimaciones MCO

Si suponemos que la matriz $(\mathbf{x}'\mathbf{x})$ es invertible, entonces se puede despejar $\widehat{\boldsymbol{\beta}}$ en las ecuaciones normales para obtener

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y};$$

y puesto que

$$(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\boldsymbol{y} = (\mathbf{x}'\mathbf{x}/n)^{-1}\,\mathbf{x}'\boldsymbol{y}/n$$

las estimaciones MCO se pueden escribir como

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{S}^2_{\mathbf{x}}\right)^{-1}\boldsymbol{S}_{\mathbf{x}\boldsymbol{y}} \tag{D.1}$$

donde

$$\mathbf{S}^2_{\mathbf{x}} = \frac{\mathbf{x}'\mathbf{x}}{n}; \quad \text{y} \quad \boldsymbol{S}_{\mathbf{x}\boldsymbol{y}} = \frac{\mathbf{x}'\boldsymbol{y}}{n};$$

Compare (D.1) con (C.1) y resuelva el ejercicio de más abajo.

Por ejemplo, para $k = 2$

$$\mathbf{S}^2_{\mathbf{x}} = \begin{bmatrix} 1 & \overline{x} \\ \overline{x} & \frac{1}{N}\sum x_n^2 \end{bmatrix}; \quad \text{y} \quad \boldsymbol{S}_{\mathbf{x}\boldsymbol{y}} = \begin{bmatrix} \overline{y} \\ \frac{1}{N}\sum y_n x_n \end{bmatrix}$$

Ejercicio 53.

(a) Verifique las dos igualdades anteriores

(b) Empleando la expresión (D.1) obtenga las expresiones de las ecuaciones (3.9) y (3.10) de la página 15.

Soluciones a los Ejercicios

Ejercicio 7.

$$\begin{aligned}
\operatorname{Var}(\boldsymbol{Y}) =& \mathrm{E}\Big(\big(\boldsymbol{Y}-\mathrm{E}(\boldsymbol{Y})\big)\big(\boldsymbol{Y}-\mathrm{E}(\boldsymbol{Y})\big)'\Big) \\
=& \mathrm{E}\big(\boldsymbol{Y}\boldsymbol{Y}' - \boldsymbol{Y}\mathrm{E}(\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\boldsymbol{Y}' + \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}')\big) && \text{desarollando el producto} \\
=& \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') - \mathrm{E}\big(\boldsymbol{Y}\mathrm{E}(\boldsymbol{Y}')\big) - \mathrm{E}\big(\mathrm{E}(\boldsymbol{Y})\boldsymbol{Y}'\big) + \mathrm{E}\big(\mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}')\big) \\
=& \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}') + \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}') && \text{pues } \mathrm{E}(\boldsymbol{Y}') \text{ es constante} \\
=& \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') - \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y}')
\end{aligned}$$

Ejercicio 7

Ejercicio 10. Puesto que $\{U_i, \boldsymbol{X}_{i\triangleright}\}$ es independiente de $\{U_j, \boldsymbol{X}_{1\triangleright}, \dots, \boldsymbol{X}_{i-1\triangleright}, \boldsymbol{X}_{i+1\triangleright}, \dots, \boldsymbol{X}_{N\triangleright}\}$; tenemos que $\mathrm{E}(U_i \mid \mathbf{x}\, u_j) = \mathrm{E}(U_i \mid \boldsymbol{x}_{i\triangleright})$. Así

$$\begin{aligned}
\mathrm{E}(U_iU_j \mid \mathbf{x}) &= \mathrm{E}\big(\mathrm{E}(U_iU_j \mid \mathbf{X}\, U_j)\cdot U_j \mid \mathbf{x}\big) && \text{por Teorema esperanzas iteradas} \\
&= \mathrm{E}\big(\mathrm{E}(U_i \mid \mathbf{X}\, U_j)\cdot U_j \mid \mathbf{x}\big) && \text{por linealidad de la esperanza condicional} \\
&= \mathrm{E}\big(\mathrm{E}(U_i \mid \boldsymbol{x}_{i\triangleright})\cdot U_j \mid \mathbf{x}\big) && \text{por ser m.a.s.} \\
&= \mathrm{E}(U_i \mid \boldsymbol{x}_{i\triangleright})\,\mathrm{E}(U_j \mid \boldsymbol{x}_{j\triangleright}) && \text{por ser m.a.s.}
\end{aligned}$$

Ejercicio 10

Ejercicio 12. Por la **Nota** 4 en la página˜14 sabemos que $\sum_n (x_n-\overline{x})(y_n-\overline{y}) = \sum_n y_n(x_n-\overline{x})$; por tanto, operando

$$\begin{aligned}
\sum_n (x_n-\overline{x})(y_n-\overline{y}) &= \sum_n y_n(x_n-\overline{x}) \\
&= \sum_n y_nx_n - \overline{x}\sum_n y_n \\
&= \sum_n y_nx_n - N\overline{y}\,\overline{x} \\
&= \boldsymbol{y}'\boldsymbol{x} - N\overline{y}\,\overline{x}.
\end{aligned}$$

Ejercicio 12

Ejercicio 13. Por una parte, dividiendo la primera ecuación de (3.7) por N obtenemos directamente $\boxed{\overline{y} = \widehat{a} + \widehat{b}\,\overline{x}}$; por lo que $\widehat{a} = \overline{y} - \widehat{b}\,\overline{x}$.

Por otra parte, dividiendo la segunda por N tenemos

$$\frac{\sum x_ny_n}{N} = \widehat{a}\,\overline{x} + \widehat{b}\,\frac{\sum x_n^2}{N}$$

o lo que es lo mismo, tenemos

$$\begin{aligned}
\frac{\boldsymbol{x}'\boldsymbol{y}}{N} &= \widehat{a}\,\overline{x} + \widehat{b}\,\frac{\boldsymbol{x}'\boldsymbol{x}}{N} && \text{expresando los sumatorios como productos escalares} \\
&= \big(\overline{y} - \widehat{b}\,\overline{x}\big)\overline{x} + \widehat{b}\,\frac{\boldsymbol{x}'\boldsymbol{x}}{N} && \text{sustituyendo } \widehat{a} \\
&= \overline{x}\cdot\overline{y} - \widehat{b}\,\overline{x}^2 + \widehat{b}\,\frac{\boldsymbol{x}'\boldsymbol{x}}{N} && \text{operando en el paréntesis} \\
&= \overline{x}\cdot\overline{y} + \widehat{b}\Big(\frac{\boldsymbol{x}'\boldsymbol{x}}{N} - \overline{x}^2\Big) && \text{sacando } \widehat{b} \text{ factor común}
\end{aligned}$$

es decir

$$\frac{\boldsymbol{x}'\boldsymbol{y}}{N} - \overline{x}\cdot\overline{y} = \widehat{b}\Big(\frac{\boldsymbol{x}'\boldsymbol{x}}{N} - \overline{x}^2\Big)$$

por lo que empleando (3.2) y (3.3) tenemos la segunda ecuación $\boxed{s_{\boldsymbol{x}\,\boldsymbol{y}} \;=\; \widehat{b}\;s_{\boldsymbol{x}}^2}$

Ejercicio 13

Ejercicio 14. Entonces el Supuesto 4 no se cumpliría, pues $\boldsymbol{x}$ sería conbinación lineal del vector de unos ya que $\boldsymbol{x} = c\,\mathbf{1}$.

En tal situación el sistema de ecuaciones normales (3.7) se reduciría a:

$$\begin{array}{rcrcl} \sum y_n & = & \widehat{a}\ N & + & \widehat{b}\ \sum x_n \\ c \cdot \sum y_n & = & c \cdot \widehat{a}\ N & + & c \cdot \widehat{b}\ \sum x_n \end{array}$$

donde la segunda ecuación es c veces la primera, por lo que realmente tenemos una sóla ecuación con dos incognitas.

Además, la varianza de un vector constante es cero, por lo que $s^2_{\boldsymbol{x}} = 0$ y tambien $s_{\boldsymbol{x}\,\boldsymbol{y}} = \frac{\boldsymbol{y}'\,\boldsymbol{x}}{N} - \overline{y}\,\overline{x} = c\frac{\boldsymbol{y}'\,\boldsymbol{1}}{N} - c\overline{y} = 0$; por lo que la estimación de $\widehat{b}$ está indeterminada, ya que

$$\widehat{b} = \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s^2_{\boldsymbol{x}}} = \frac{0}{0}.$$

Ejercicio 14

Ejercicio 16(a)

$$\begin{aligned} \sum x_{1n} y_n &= \widehat{a} \sum x_{1n}^2 + \widehat{b} \sum x_{1n} x_{2n} + \widehat{c} \sum x_{1n} x_{3n} \\ \sum x_{2n} y_n &= \widehat{a} \sum x_{2n} x_{1n} + \widehat{b} \sum x_{2n}^2 + \widehat{c} \sum x_{2n} x_{3n} \\ \sum x_{3n} y_n &= \widehat{a} \sum x_{3n} x_{1n} + \widehat{b} \sum x_{3n} x_{2n} + \widehat{c} \sum x_{3n}^2 \end{aligned}$$

□

Ejercicio 16(b)

$$\begin{aligned} \sum y_n &= \widehat{a} N + \widehat{b} \sum x_{2n} + \widehat{c} \sum x_{3n} \\ \sum x_{2n} y_n &= \widehat{a} \sum x_{2n} + \widehat{b} \sum x_{2n}^2 + \widehat{c} \sum x_{2n} x_{3n} \\ \sum x_{3n} y_n &= \widehat{a} \sum x_{3n} + \widehat{b} \sum x_{3n} x_{2n} + \widehat{c} \sum x_{3n}^2 \end{aligned}$$

□

Ejercicio 17. Dividiendo la primera ecuación del sistema anterior por N obtenemos

$$\overline{y} = \widehat{a} + \widehat{b} \cdot \overline{\boldsymbol{x_2}} + \widehat{c} \cdot \overline{\boldsymbol{x_3}}$$

esta ecuación indica que el plano de regresión para por el punto de los valores medios de las variables del sistema.

Despejando $\widehat{a}$ tenemos

$$\widehat{a} = \overline{y} - \widehat{b} \cdot \overline{\boldsymbol{x_2}} - \widehat{c} \cdot \overline{\boldsymbol{x_3}}$$

que se puede sustituir en las otras dos ecuaciones del sistema:

$$\begin{aligned} \sum x_{2n} y_n &= \left(\overline{y} - \widehat{b}\overline{\boldsymbol{x_2}} - \widehat{c}\overline{\boldsymbol{x_3}}\right) \sum x_{2n} + \widehat{b} \sum x_{2n}^2 + \widehat{c} \sum x_{2n} x_{3n} \\ \sum x_{3n} y_n &= \left(\overline{y} - \widehat{b}\overline{\boldsymbol{x_2}} - \widehat{c}\overline{\boldsymbol{x_3}}\right) \sum x_{3n} + \widehat{b} \sum x_{3n} x_{2n} + \widehat{c} \sum x_{3n}^2; \end{aligned}$$

operando

$$\begin{aligned} \sum x_{2n} y_n &= \overline{y} \sum x_{2n} - \widehat{b}\overline{\boldsymbol{x_2}} \sum x_{2n} - \widehat{c}\overline{\boldsymbol{x_3}} \sum x_{2n} + \widehat{b} \sum x_{2n}^2 + \widehat{c} \sum x_{2n} x_{3n} \\ \sum x_{3n} y_n &= \overline{y} \sum x_{3n} - \widehat{b}\overline{\boldsymbol{x_2}} \sum x_{3n} - \widehat{c}\overline{\boldsymbol{x_3}} \sum x_{3n} + \widehat{b} \sum x_{3n} x_{2n} + \widehat{c} \sum x_{3n}^2; \end{aligned}$$

puesto que $\sum x_{2n} = N\overline{\boldsymbol{x_2}}$ y $\sum x_{3n} = N\overline{\boldsymbol{x_3}}$, sustituyendo

$$\begin{aligned} \sum x_{2n} y_n &= N\overline{y} \cdot \overline{\boldsymbol{x_2}} - N\widehat{b} \cdot \overline{\boldsymbol{x_2}} \cdot \overline{\boldsymbol{x_2}} - N\widehat{c} \cdot \overline{\boldsymbol{x_3}} \cdot \overline{\boldsymbol{x_2}} + \widehat{b} \sum x_{2n}^2 + \widehat{c} \sum x_{2n} x_{3n} \\ \sum x_{3n} y_n &= N\overline{y} \cdot \overline{\boldsymbol{x_3}} - N\widehat{b} \cdot \overline{\boldsymbol{x_2}} \cdot \overline{\boldsymbol{x_3}} - N\widehat{c} \cdot \overline{\boldsymbol{x_3}} \cdot \overline{\boldsymbol{x_3}} + \widehat{b} \sum x_{3n} x_{2n} + \widehat{c} \sum x_{3n}^2; \end{aligned}$$

sustituyendo los sumatorios que restan por productos escalares:

$$\begin{aligned} \boldsymbol{x_2}'\,\boldsymbol{y} &= N\overline{y} \cdot \overline{\boldsymbol{x_2}} - N\widehat{b} \cdot \overline{\boldsymbol{x_2}} \cdot \overline{\boldsymbol{x_2}} - N\widehat{c} \cdot \overline{\boldsymbol{x_3}} \cdot \overline{\boldsymbol{x_2}} + \widehat{b} \cdot \boldsymbol{x_2}'\,\boldsymbol{x_2} + \widehat{c} \cdot \boldsymbol{x_2}'\,\boldsymbol{x_3} \\ \boldsymbol{x_3}'\,\boldsymbol{y} &= N\overline{y} \cdot \overline{\boldsymbol{x_3}} - N\widehat{b} \cdot \overline{\boldsymbol{x_2}} \cdot \overline{\boldsymbol{x_3}} - N\widehat{c} \cdot \overline{\boldsymbol{x_3}} \cdot \overline{\boldsymbol{x_3}} + \widehat{b} \cdot \boldsymbol{x_3}'\,\boldsymbol{x_2} + \widehat{c} \cdot \boldsymbol{x_3}'\,\boldsymbol{x_3}; \end{aligned}$$

reordenando términos:

$$\boldsymbol{x_2}'\,\boldsymbol{y} - N\overline{y}\cdot\overline{\boldsymbol{x_2}} = \widehat{b}\cdot\left(\boldsymbol{x_2}'\,\boldsymbol{x_2} - N\overline{\boldsymbol{x_2}}^2\right) + \widehat{c}\cdot\left(\boldsymbol{x_2}'\,\boldsymbol{x_3} - N\overline{\boldsymbol{x_3}}\cdot\overline{\boldsymbol{x_2}}\right)$$
$$\boldsymbol{x_3}'\,\boldsymbol{y} - N\overline{y}\cdot\overline{\boldsymbol{x_3}} = \widehat{b}\cdot\left(\boldsymbol{x_3}'\,\boldsymbol{x_2} - N\overline{\boldsymbol{x_2}}\cdot\overline{\boldsymbol{x_3}}\right) + \widehat{c}\cdot\left(\boldsymbol{x_3}'\,\boldsymbol{x_3} - N\overline{\boldsymbol{x_3}}^2\right);$$

y teniendo en cuenta las notas 4 a 6 en la página˜14

$$N s_{\boldsymbol{x_2}\,\boldsymbol{y}} = \widehat{b}\cdot N s^2_{\boldsymbol{x_2}} + \widehat{c}\cdot N s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}$$
$$N s_{\boldsymbol{x_3}\,\boldsymbol{y}} = \widehat{b}\cdot N s_{\boldsymbol{x_2}\,\boldsymbol{x_3}} + \widehat{c}\cdot N s^2_{\boldsymbol{x_3}};$$

o bien;

$$\sum_n y_n(x_{2n} - \overline{\boldsymbol{x_2}}) = \widehat{b}\cdot\sum_n x_{2n}(x_{2n} - \overline{\boldsymbol{x_2}}) + \widehat{c}\cdot\sum_n x_{3n}(x_{2n} - \overline{\boldsymbol{x_2}})$$
$$\sum_n y_n(x_{3n} - \overline{\boldsymbol{x_3}}) = \widehat{b}\cdot\sum_n x_{3n}(x_{2n} - \overline{\boldsymbol{x_2}}) + \widehat{c}\cdot\sum_n x_{3n}(x_{3n} - \overline{\boldsymbol{x_3}});$$

por tanto, resolviendo el sistema, obtenemos los dos últimos resultados

$$\widehat{b} = \frac{s_{\boldsymbol{x_2}\,\boldsymbol{y}}\cdot s^2_{\boldsymbol{x_3}} - s_{\boldsymbol{x_3}\,\boldsymbol{y}}\cdot s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}}{s^2_{\boldsymbol{x_2}}\cdot s^2_{\boldsymbol{x_3}} - \left(s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}\right)^2}$$

$$\widehat{c} = \frac{s_{\boldsymbol{x_3}\,\boldsymbol{y}}\cdot s^2_{\boldsymbol{x_2}} - s_{\boldsymbol{x_2}\,\boldsymbol{y}}\cdot s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}}{s^2_{\boldsymbol{x_2}}\cdot s^2_{\boldsymbol{x_3}} - \left(s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}\right)^2}$$

Ejercicio 17

Ejercicio 18. Con la estimación de la pendiente en el modelo $Y_n = a + cX_{3n} + U_n$

Ejercicio 18

Ejercicio 19. Un coeficiente de correlación muestral con valor absoluto igual a uno significa que hay una dependencia lineal entre los regresores, por lo que el Supuesto 4 deja de cumplirse; y por tanto el sistema de ecuaciones normales tiene infinitas soluciones.

En tal caso las expresiones (3.11), (3.12) y (3.13) dejan de estar definidas ya que, en este caso

$$|\rho_{\boldsymbol{x_2}\boldsymbol{x_3}}| = \left|\frac{s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}}{\sqrt{s^2_{\boldsymbol{x_2}} s^2_{\boldsymbol{x_3}}}}\right| = 1,$$

lo que implica que $|s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}| = \left|\sqrt{s^2_{\boldsymbol{x_2}} s^2_{\boldsymbol{x_3}}}\right|$ y por tanto $s^2_{\boldsymbol{x_2}}\cdot s^2_{\boldsymbol{x_3}} = \left(s_{\boldsymbol{x_2}\,\boldsymbol{x_3}}\right)^2$; y los denominadores de de las expresiones (3.12) y (3.13) son cero.

Ejercicio 19

Ejercicio 20. Los valores estimados en los modelos restringidos y sin restringir difieren. Por lo tanto, podemos afirmar que la covarianza muestral entre los regresores S_n y D_n en esta simulación es distinta de cero.

Ejercicio 20

Ejercicio 22. La primera es inmediata. La segundad también lo es por la **Nota** 6 en la página˜14. La tercera en un poco más complicada (pero no mucho):
Pista. Transforme el producto escalar en un sumatorio. Opere dentro del sumatorio y tenga en cuenta que las medias muestrales son constantes que se pueden sacar fuera de los sumatorios como un “factor común”.

Ejercicio 22

Ejercicio 23.
Pista.

- $\widehat{\boldsymbol{y}} = \boldsymbol{y} - \widehat{\boldsymbol{e}}$
- $\widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{e}} = 0$

Ejercicio 23

Ejercicio 24.

$$(\mathbf{abc})' = ((\mathbf{ab})\,\mathbf{c})' = \mathbf{c}'\,(\mathbf{ab})' = \mathbf{c}'\,\mathbf{b}'\,\mathbf{a}'\,.$$

Ejercicio 24

Ejercicio 25. Por la Nota 8 en la página~24 sabemos que

$$\mathbf{p}' = \left[(\mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1})\,\mathbf{x}'\right]^T = \mathbf{x}\left(\mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\right)^T = \mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}' = \mathbf{p}$$

y entonces

$$\begin{aligned}\mathbf{p}'\,\mathbf{m} &= \mathbf{p}(\mathbf{I}-\mathbf{p}) = \mathbf{p}-\mathbf{p}\mathbf{p}\\ &= \mathbf{p}-\mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\underbrace{\mathbf{x}'\,\mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}}_{\mathbf{I}}\mathbf{x}'\\ &= \mathbf{p}-\mathbf{x}(\mathbf{x}'\,\mathbf{x})^{-1}\mathbf{x}' = \mathbf{p}-\mathbf{p} = \mathbf{0}\end{aligned}$$

Ejercicio 25

Ejercicio 26.

1. $$\mathbf{m}' = \left[\mathbf{I}-\mathbf{p}\right]' = \mathbf{I}'-\mathbf{p}' = \mathbf{I}-\mathbf{p} = \mathbf{m}$$
2. $$\mathbf{m}\mathbf{m}'\,\mathbf{m}\mathbf{m} = \mathbf{m}\mathbf{m} = \left[\mathbf{I}-\mathbf{p}\right]\left[\mathbf{I}-\mathbf{p}\right] = \mathbf{I}-\mathbf{p}-\mathbf{p}+\mathbf{p}\mathbf{p} = \mathbf{I}-\mathbf{p}-\mathbf{p}+\mathbf{p} = \mathbf{I}-\mathbf{p} = \mathbf{m}$$

Ejercicio 26

Ejercicio 32.
Pista. Calcule el coeficiente de correlación lineal simple entre $\widehat{\boldsymbol{y}}$ y $\boldsymbol{y}$ y elévelo al cuadrado.

Solución numérica en el recuadro del ejemplo del precio de las viviendas (página 17).

Ejercicio 32

Ejercicio 33. Por una parte, $SEC = \sum\left(\widehat{y_n}-\overline{y}\right)^2$; pero en este modelo los valores ajustados son constantes iguales a la media muestral de $\boldsymbol{y}$, es decir $\widehat{y_n} = \overline{y}$. Por tanto $SEC = 0$.

Por otra parte, este modelo tiene término cte. y, entonces, $R^2 = \frac{SEC}{STC} = 0$.

Es decir, un modelo que consiste únicamente en un constante, no tiene ninguna capacidad de "explicar" las variaciones de la variable dependiente.

Otra forma de verlo es la siguiente. En este modelo sabemos que $\widehat{y_n} = \overline{y}$. Así que

$$\begin{aligned}SEC = \sum\left(\widehat{y_n}-\overline{y}\right)^2 &= \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} + N\overline{y}^2 - 2N\overline{y}\overline{\widehat{y}} && \text{por } \boxed{\text{T22}}\\ &= \widehat{\boldsymbol{y}}'\,\widehat{\boldsymbol{y}} - N\overline{\widehat{y}}^2 && \text{pues en este caso } \overline{\widehat{y}} = \overline{y}\\ &= \overline{y}\cdot\mathbf{1}'\,\widehat{\boldsymbol{y}} - N\overline{\widehat{y}}^2 && \text{pues } \widehat{\boldsymbol{y}} \text{ es un vector de constantes } \overline{y}\\ &= N\overline{\widehat{y}}^2 - N\overline{\widehat{y}}^2 && \text{pues en este caso } \overline{\widehat{y}} = \overline{y}\end{aligned}$$

Ejercicio 33

Ejercicio 34. En este caso

$$\widehat{b} = \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s_{\boldsymbol{x}}^2} \quad \text{y} \quad \widehat{a} = \overline{y}-\overline{x}\widehat{b},$$

por tanto

$$\widehat{y_n} = \widehat{a}+\widehat{b}x_n = \overline{y}+\widehat{b}(x_n-\overline{x}); \quad \Rightarrow \quad \widehat{y_n}-\overline{y} = \widehat{b}(x_n-\overline{x}).$$

Entonces

$$SEC = \sum(\widehat{y_n}-\overline{y})^2 = \widehat{b}^2\sum(x_n-\overline{x})^2$$

y consiguientemente (por tener un término constante el Modelo Lineal General)

$$R^2 = \frac{SEC}{STC} = \frac{\widehat{b}^2\sum(x_n-\overline{x})^2}{\sum(y_n-\overline{y})^2} = \frac{\left(s_{\boldsymbol{x}\,\boldsymbol{y}}\right)^2}{\left(s_{\boldsymbol{x}}^2\right)^2}\cdot\frac{Ns_{\boldsymbol{x}}^2}{Ns_{\boldsymbol{y}}^2} = \frac{\left(s_{\boldsymbol{x}\,\boldsymbol{y}}\right)^2}{s_{\boldsymbol{x}}^2 s_{\boldsymbol{y}}^2} = r_{\boldsymbol{y}\,\boldsymbol{x}}^2$$

Ejercicio 34

Ejercicio 39.

$$\begin{aligned}\operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\;\middle|\;\mathbf{x}\right) &= \operatorname{Var}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\;\middle|\;\mathbf{x}\right) && \text{ya que } \boldsymbol{\beta} \text{ es cte.}\\ &= \operatorname{Var}(\mathbf{A}\boldsymbol{U}\mid\mathbf{x}) && \text{ya que } \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}+\mathbf{A}\boldsymbol{U}\\ &= \operatorname{E}\left(\mathbf{A}\boldsymbol{U}\,\boldsymbol{U}'\mathbf{A}'\;\middle|\;\mathbf{x}\right) && \text{pues } \operatorname{E}(\mathbf{A}\boldsymbol{U}\mid\mathbf{x}) = \mathbf{0}\\ &= \mathbf{a}\operatorname{Var}(\boldsymbol{U}\mid\mathbf{x})\,\mathbf{a}' && \text{pues } \mathbf{a} \text{ cte. si } \mathbf{X} = \mathbf{x}\\ &= \mathbf{a}\,\sigma^2\,\mathbf{I}\,\mathbf{a}' && \text{por los supuestos 2 y 3 } \boxed{\text{T10}}\\ &= \sigma^2\,\mathbf{a}\,\mathbf{a}' \;=\; \sigma^2(\mathbf{x}'\,\mathbf{x})^{-1}\end{aligned}$$

puesto que $\mathbf{aa'} = (\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}\,\mathbf{x}(\mathbf{x'}\,\mathbf{x})^{-1} = (\mathbf{x'}\,\mathbf{x})^{-1}$.

Si los regresores son NO estocásticos: denotemos $(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}$ por $\underset{[k\times N]}{\mathbf{a}}$

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) =& \mathrm{Var}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) && \text{ya que } \boldsymbol{\beta} \text{ es cte.}\\
=& \mathrm{Var}(\mathbf{a}\boldsymbol{U}) && \text{ya que } \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{a}\boldsymbol{U}\\
=& \mathbf{a}\,\mathrm{Var}(\boldsymbol{U})\,\mathbf{a'} && \text{ya que } \mathbf{a} \text{ es una matriz cte.}\\
=& \mathbf{a}\,\sigma^2\,\mathbf{I}\,\mathbf{a'} && \text{por los supuestos 2 y 3}\\
=& \sigma^2\,\mathbf{a}\,\mathbf{a'}\\
=& \sigma^2(\mathbf{x'}\,\mathbf{x})^{-1}
\end{aligned}$$

Ejercicio 39

Ejercicio 40. En este caso seleccionamos la componente j-ésima del vector $\widehat{\boldsymbol{\beta}}$, por tanto

$$\mathrm{Var}\left(\boldsymbol{v}'\,\widetilde{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right) = \mathrm{Var}\left(\widetilde{\beta_j}\,\middle|\,\mathbf{x}\right) \geq \mathrm{Var}\left(\widehat{\beta_j}\,\middle|\,\mathbf{x}\right) = \mathrm{Var}\left(\boldsymbol{v}'\,\widehat{\boldsymbol{\beta}}\,\middle|\,\mathbf{x}\right).$$

Es decir, el teorema de Gauss-Markov implica que la varianza del estimador de cada parámetro j-ésimo $\mathrm{Var}\left(\widetilde{\beta_j}\,\middle|\,\mathbf{x}\right)$ es mayor o igual que la del estimador MCO $\mathrm{Var}\left(\widehat{\beta_j}\,\middle|\,\mathbf{x}\right)$.

Ejercicio 40

Ejercicio 41.

1. $\mathbf{px} = \mathbf{x}\,\underbrace{(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}\,\mathbf{x}}_{\mathbf{I}} = \mathbf{x}$

2. $$\begin{aligned}
\mathbf{p'} &= \left[\mathbf{x}(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}\right]'\\
&= \left[(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}\right]'\mathbf{x'} && \text{pues } \left[\mathbf{x}\,\mathbf{a}\right]' = \mathbf{a'}\,\mathbf{x'}\\
&= \mathbf{x}\left[(\mathbf{x'}\,\mathbf{x})^{-1}\right]'\mathbf{x'} && \text{idéntica regla de trasposición sobre el corchete}\\
&= \mathbf{x}(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'} = \mathbf{p} && \text{pues } (\mathbf{x'}\,\mathbf{x})^{-1} \text{ es simétrica}
\end{aligned}$$

3. $$\begin{aligned}
\mathbf{pp} &= \mathbf{pxa}\\
&= \mathbf{xa} = \mathbf{p} && \text{pues } \mathbf{px} = \mathbf{x}
\end{aligned}$$

Ejercicio 41

Ejercicio 42.

1. $$\mathbf{mx} = \left[\mathbf{I} - \mathbf{p}\right]\mathbf{x} = \mathbf{x} - \mathbf{px} = \mathbf{x} - \mathbf{x} = \mathbf{0}$$

2. $$\mathbf{am} = \mathbf{a}\left[\mathbf{I} - \mathbf{p}\right] = \mathbf{a} - (\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'}\,\mathbf{x}\;(\mathbf{x'}\,\mathbf{x})^{-1}\mathbf{x'} = \mathbf{a} - \mathbf{a} = \mathbf{0}$$

3. $$\mathbf{m'} = \left[\mathbf{I} - \mathbf{p}\right]' = \mathbf{I'} - \mathbf{p'} = \mathbf{I} - \mathbf{p} = \mathbf{m}$$

4. $$\mathbf{mm} = \left[\mathbf{I} - \mathbf{p}\right]\left[\mathbf{I} - \mathbf{p}\right] = \mathbf{I} - \mathbf{p} - \mathbf{p} + \mathbf{pp} = \mathbf{I} - \mathbf{p} - \mathbf{p} + \mathbf{p} = \mathbf{I} - \mathbf{p} = \mathbf{m}$$

Ejercicio 42

Ejercicio 43.

$$\widehat{SRC}_{|\,\mathbf{x}} = \widehat{\boldsymbol{e}}_{|\,\mathbf{x}}{}'\,\widehat{\boldsymbol{e}}_{|\,\mathbf{x}} = \boldsymbol{U}'\mathbf{m'}\,\mathbf{m}\boldsymbol{U} = \boldsymbol{U}\,\mathbf{m}\boldsymbol{U}$$

por ser $\mathbf{m}$ simétrica e idempotente.

Ejercicio 43

Ejercicio 44.

$$\mathrm{E}\left(\widehat{\mathfrak{s}^2_{\boldsymbol{e}}}\middle|\mathbf{x}\right) = \frac{\sigma^2}{N-k}\cdot(N-k) = \sigma^2 \qquad \text{y} \qquad \mathrm{Var}\left(\widehat{\mathfrak{s}^2_{\boldsymbol{e}}}\middle|\mathbf{x}\right) = \left(\frac{\sigma^2}{N-k}\right)^2\cdot 2(N-K) = 2\frac{\sigma^4}{(N-k)}.$$

Ejercicio 44

Ejercicio 50(a)

$$\mathrm{E}\big(\widetilde{\sigma}^2_{MV}\,\big|\mathbf{x}\big) = \mathrm{E}\Big(\widehat{s^2_{\hat{e}}}\Big|\mathbf{x}\Big) = \left(\frac{\sigma^2}{N}\right)\cdot(N-K) = \frac{(N-k)\sigma^2}{(N)}.$$

□

Ejercicio 50(b)

$$\mathrm{Var}\big(\widetilde{\sigma}^2_{MV}\,\big|\mathbf{x}\big) = \mathrm{Var}\Big(\widehat{s^2_{\hat{e}}}\Big|\mathbf{x}\Big) = \left(\frac{\sigma^2}{N}\right)^2\cdot 2(N-K) = \frac{2\sigma^4}{(N)}.$$

□

Ejercicio 50(c) La varianza coincide con la cota mínima, pero esto no quiere decir nada; esta cota sólo es aplicable a estimadores insesgados, y éste estimador es sesgado.

□

Ejercicio 51. Ya que

$$\begin{aligned}
\mathrm{E}(\,Y_n \mid \mathbf{x}) =&\mathrm{E}(\,a+bX_n+U_n \mid \mathbf{x}) && \\
=&a+bX_n+\mathrm{E}(\,U_n \mid \mathbf{x}) && \text{pues } a,\ b, \text{ y } X_n \text{ son ctes}\\
=&a+bX_n && \text{por el supuesto: } \mathrm{E}(\,U_n \mid \mathbf{x})=0
\end{aligned}$$

Ejercicio 51

Ejercicio 52(a)

1. Por una parte:

$$\widehat{b} = \frac{s_{\boldsymbol{x}\,\boldsymbol{y}}}{s^2_{\boldsymbol{x}}} = \frac{84}{56} = 1.5$$

por otra, las medias muestrales son

$$\overline{x} = \frac{\sum x_i}{8} = \frac{40}{8} = 5; \qquad \overline{y} = \frac{\sum y_i}{8} = \frac{56}{8} = 7;$$

por lo que

$$\widehat{a} = \overline{y} - \widehat{b}\cdot\overline{x} = 7 - 1.5\cdot 5 = -0.5.$$

□

Ejercicio 52(b) Según el modelo estimado, una empresa que incurra en unos gastos de 3 debería tener unos beneficios de

$$\widehat{y} = \widehat{a} + \widehat{b}x = -0.5 + 1.5\cdot 3 = 4$$

□

Ejercicio 52(c) Los residuos de la empresa E serán:

$$y_E - \widehat{y_E} = y_E - \big(\widehat{a} + \widehat{b}x_E\big) = 8 - \big(-0.5 + 1.5\cdot 5\big) = 8 - 7 = 1$$

y los de la empresa F:

$$y_F - \widehat{y_F} = y_F - \big(\widehat{a} + \widehat{b}x_F\big) = 9 - \big(-0.5 + 1.5\cdot 7\big) = 9 - 10 = -1.$$

Puesto que

$$\widehat{y} = \mathrm{E}(\,Y \mid \boldsymbol{x}_{f\triangleright})\,,$$

un signo positivo para el residuo de cierta empresa significa que ésta ha logrado unos beneficios mayores que los esperados (dado su nivel de gasto en formación de personal, x). Por el contrario, un residuo negativo significa que la empresa ha obtenido unos beneficios menores de los esperados por el modelo (dado su gasto en formación).

La comparación entre empresas con distinta inversión en formación no es apropiada para valorar los datos sobre beneficios (sólo lo es entre empresas con mismo nivel de gasto en formación). La empresa F tiene mayores beneficios que los de E, pero, dado su nivel de gasto en formación (7), estos beneficios deberían haber sido aún mayores (el valor esperado es 10).

□

Ejercicio 52(d) El estimador MCO se distribuye Normal con esperanza igual al verdadero valor de los parámetros estimados, y varianza desconocida.

- Buscamos los valores A y B tales que
$$P\left(A \leq \frac{\widehat{b}-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} \leq B\right) = (1-\alpha)$$
Donde $\frac{\widehat{b}-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}}$ se distribuye como una t de Student con $N-2$ grados de libertad; por tanto A y B son los valores que aparecen en las tablas, y que determinan un intervalo centrado en cero con una probabilidad asociada del 95 %; es decir, $A = -2.447$, y $B = 2.447$, y $\widehat{\mathfrak{s}^2_{\widehat{e}}} = 6/(N-2) = 1$. Así pues, la estimación del intervalo de confianza de parámetro desconocido b es
$$\mathrm{IC}^b_{0.95}(\boldsymbol{w}) = \left[1.5 \pm 2.447 \cdot \sqrt{1/56}\right]$$

□

Ejercicio 52(e) Las hipótesis son:
$$H_0 : \ b = 1$$
$$H_1 : \ b < 1$$
La región critica de una sola cola es
$$RC = \left\{ \mathbf{x} \;\middle|\; \frac{\widehat{b}-1}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} < k \right\},$$
donde k es el valor de la tablas para una t de Student de seis grados de libertad, ya que el estadístico de la parte izquierda de la desigualdad tiene dicha distribución. Para $\alpha = 0.1$, tenemos que $k = t_{6,\ 0.1} = -1.44$. Sustituyendo tenemos que
$$\frac{1.5-1}{\sqrt{1/56}} = 3.74 > k = t_{6,0.1} = -1.44$$
por lo que no rechazamos H_0.

El p-valor es la probabilidad de
$$P\left(\widehat{b} \leq 1.5 \;\middle|\; H_0\right) = P\left(\frac{\widehat{b}-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} \leq \frac{1.5-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} \;\middle|\; H_0\right)$$
$$= P\left(W \leq \frac{1.5-1}{\sqrt{1/56}} = 3.74\right) \simeq 0.999$$
donde W se distribuye como una t de Student con seis grados de libertad.

□