Guía de Estudio para la Asignatura de

ECONOMETRÍA I

Prof. Luca Gambetti

Universitat Autonoma de Barcelona

Julio 2009

Introducción

El curso de Econometría I constituye uno de los dos cursos de ECONOMETRÍA del Programa Universitat Empresa y es un curso obligatorio del primer semestre del primer año del Programa (tercero de carrera). El objetivo principal del curso es familiarizar al estudiante con los elementos básicos del trabajo econométrico proporcionándole todos los conocimientos necesarios para poder desarrollar un análisis cuantitativo de los modelos económicos. Por otro lado, es también objetivo del curso familiarizar al estudiante con la utilización de software econométrico, como el programa GRETL, para la investigación empírica. Por la naturaleza del los modelos estudiados, el curso requiere el conocimiento de determinados conceptos estadísticos y matemáticos que, por ello, se revisan en la primera parte del curso.

La presente guía de estudios tiene como objetivo fundamental ser un instrumento útil al alumno en el aprendizaje del contenido temático del curso. La idea es que pueda servir como material complementario a los apuntes y de profundización de algunos de los temas estudiados en clase. La guía contiene apuntes sobre todos los temas teóricos tratados en el curso, además de ejemplos, ejercicios, aplicaciones empíricas con GRETL, exámenes pasados y referencias bibliográficas.

Capítulo 1

Modelo de Regresión Lineal Simple

1.1. El modelo

Sean y y x dos variables económicas de interés. Queremos un modelo econométrico que proporcione respuestas a preguntas como: (i) ¿cuáles son los efectos sobre y de un cambio en x? (ii) ¿cuanto podemos explicar de y a través de x? El modelo econométrico que estudiaremos en este curso es el modelo de regresión lineal. En este capítulo, en particular, nos concentraremos en el modelo "simple". El modelo de regresión lineal simple se describe a través de la siguiente relación *lineal*

$$y = \beta_0 + \beta_1 x + u \tag{1.1}$$

Donde:

1. y −es la variable que queremos explicar y recibe el nombre de *variable dependiente* o *variable explicada.*

2. x − es la variable a través de la cual queremos explicar y y recibe el nombre de *variable independiente*, porque se determina fuera de nuestro modelo, o *variable explicativa.*

3. u − se denomina término de error, es una variable aleatoria y representa factores non observables distintos a x que afectan a y.

4. $\beta_0, \beta_1 -$ son los parámetros del modelo.

Sea ahora $\{(x_i, y_i) : i = 1, ...n\}$ una muestra de la población. Asumiendo que estos datos sean generados por el modelo de regresión (1.1) podemos asumir que por cada i

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{1.2}$$

Al supuesto de linealidad del modelo añadimos los supuestos siguientes:

S1 *La esperanza de u_i es zero: $E(u_i) = 0$ por cada i.*

S2 *La variable independiente es no aleatoria o fija en muestras repetidas.*

S3 *La varianza de u_i es constante: $Var(u_i) = \sigma^2$* para todos i.

S4 *u_i y u_j son independientes por cada i y cada j con $i \neq j$.*

El supuesto **S3** se conoce también como supuesto de *homoscedasticidad*. El supuesto **S4** implica que la covarianza entre u_i y u_j con $i \neq j$ debe ser igual a cero o sea $E(u_i u_j) = 0$. Tomando el valor esperado de y obtenemos

$$\begin{aligned} E(y_i) &= E(\beta_0 + \beta_1 x_i + u_i) \\ &= E(\beta_0) + E(\beta_1 x_i) + E(u_i) \\ &= \beta_0 + \beta_1 x + E(u_i) \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

donde la primera y segunda igualdad derivan de las propiedades del valor esperado y la tercera de **S2**. La media condicional de y es una recta que se conoce como *función de regresión poblacional* y es también una función lineal de x.

El parámetro β_1 es el parámetro que a menudo mas "interesa" a los economistas porque describe la relación existente entre y y x. ¿Cómo interpretamos este parámetro? Considerese un cambio Δy de y en (1.1). Este será $\Delta y = \beta_1 \Delta x + \Delta u$. Ahora supongamos que u se mantenga constante, $\Delta u = 0$, de manera que $\Delta y = \beta_1 \Delta x$. Entonces $\beta_1 = \frac{\Delta y}{\Delta x}$ nos dice cuánto varía y si x varia de Δx y u se mantiene constante. Si

$\Delta x = 1$, o sea un cambio unitario en x, $\beta_1 = \Delta y$ representa la variación de y. Queda claro ahora la implicación del supuesto de linealidad del modelo: el efecto de x sobre y es lineal sobre y porque es simplemente β_1 por el cambio en x. Se considere ahora la función de regresión poblacional. En este caso β_1 nos dice de cuánto varía en promedio y si x varía de Δx. Podemos interpretar β_1 como la variación de y provocada por un cambio unitario de x si todos los demás factores son constantes, o como la variación promedia de y provocada por un cambio unitario de x. La interpretación de β_0 es mas problemática. De hecho β_0 nos dice cual es el valor promedio de y cuando x es cero. Sin embargo para poder interpretar β_0, x_i debe poder tomar valor igual a cero. En caso contrario no podemos interpretar este parámetro.

Obs.	y	x
1	1.8987	8.0000
2	7.0240	9.0000
3	3.1294	10.0000
4	5.3579	11.0000
5	4.3279	12.0000
6	6.1567	13.0000
7	4.1183	14.0000
8	3.5169	15.0000
9	3.7885	16.0000
10	6.7904	17.0000
11	6.1492	18.0000
12	6.4227	19.0000
13	9.0072	20.0000
14	7.7124	21.0000
15	7.8797	22.0000
16	10.1453	23.0000

Tabla 1:

Para comprender mejor el modelo, consideramos ahora su interpretación gráfica. Consideremos la muestra aleatoria de 16 observaciones de y y x en la Tabla 1. Tal muestra se supone generada por el siguiente modelo

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + u_i \\ &= 1 + 0{,}3x_i + u_i \end{aligned}$$

El grafico 1 enseña todas las observaciones (y_i, x_i) de la muestra y la recta de regresión poblacional $y_i = 1 + 0{,}3x_i$. El parámetro $\beta_0 = 1$ representa la intercepta y β_1 la pendiente de la recta de regresión. El modelo descompone cada observación y_i en dos partes. Una parte, la esperanza de y_i, o sea $y_i = 1 + 0{,}3x_i$, que está encima de la recta de regresión poblacional (los triángulos) y otra parte, el error de regresión u_i, que representa la distancia vertical entre los puntos y los triángulos.

1.2. Estimación

Trataremos ahora la importante cuestión de la estimación de los parámetros del modelo de regresión. La idea es que nosotros, como investigadores, no conocemos ninguna característica cualitativa (el signo) o cuantitativa de los parámetros del modelo. La finalidad del procedimiento de estimación es la de utilizar una muestra de datos, que se suponen generados por el modelo de regresión, para obtener valores para los parámetros, o sea unas estimaciones.

1.2.1. Estimador Mínimos Cuadrados Ordinarios (MCO)

En este curso el único estimador que estudiaremos y utilizaremos es el estimador de **Mínimos Cuadrados Ordinarios**. El criterio que nos permite obtener los estimadores de MCO de β_0, β_1 es el de la minimización de la suma de los cuadrados de

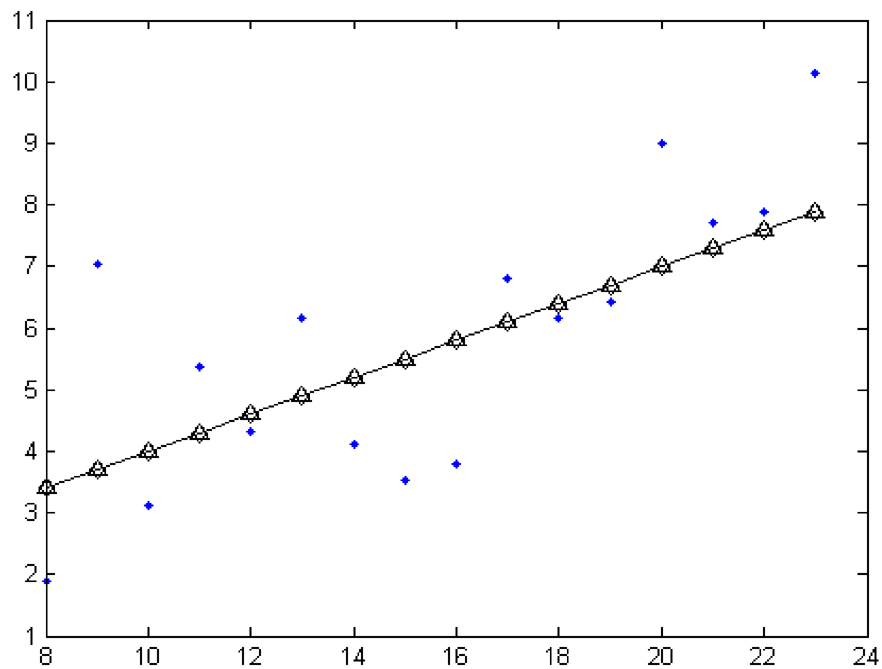

Figura 1:

los residuos de regresión. Para entender el método necesitamos algunas definiciones preliminares. Sean $\hat{\beta}_0, \hat{\beta}_1$ dos estimadores de β_0, β_1. Definimos el **valor ajustado** de y_i como $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, y el **residuo** de regresión como $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$. El residuo de regresión para la observación i es la diferencia entre el valor verdadero y_i y su valor ajustado $\hat{y}_i$ y expresa la parte de y_i que no podemos explicar con nuestra variable explicativa x. El método de mínimos cuadrado se basa en la minimización de la suma de todos los residuos al cuadrado. La idea es que queremos escoger $\hat{\beta}_0, \hat{\beta}_1$ de manera que la parte de variabilidad de y que no podemos explicar, o sea la suma de todos los residuos al cuadrado, $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i)^2$, sea mínima.

Formalmente $\hat{\beta}_0, \hat{\beta}_1$ son los estimadores que solucionan el siguiente problema de minimización:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \equiv Q$$

Para solucionar este problema hay que encontrar las condiciones del primer orden. Estas condiciones son las derivadas parciales de la función objetivo, Q, respecto a los dos estimadores igualadas a cero.

$$\frac{\partial Q}{\hat{\beta}_0} = 0 \Rightarrow \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$
$$\frac{\partial Q}{\hat{\beta}_1} = 0 \Rightarrow \sum_{i=1}^n 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Desde la primera condición obtenemos

$$\begin{aligned} \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\beta}_0 - \sum_{i=1}^n \hat{\beta}_1 x_i = 0 \\ \frac{\sum_{i=1}^n y_i}{n} - \frac{\sum_{i=1}^n \hat{\beta}_0}{n} - \frac{\sum_{i=1}^n \hat{\beta}_1 x_i}{n} = 0 \\ \bar{y} - \frac{n\hat{\beta}_0}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} = 0 \\ \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0 \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \tag{1.3}$$

que es una ecuación para $\hat{\beta}_0$ en términos de $\hat{\beta}_1$. Ahora podemos utilizar la segunda condición de primer orden para encontrar una ecuación para $\hat{\beta}_1$. De la segunda obtenemos

$$
\begin{aligned}
\sum_{i=1}^{n}\left[y_i - (\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i\right] x_i = 0 \\
\sum_{i=1}^{n}(y_i - \bar{y})x_i - \sum_{i=1}^{n}\hat{\beta}_1(x_i - \bar{x})x_i = 0 \\
\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) - \hat{\beta}_1\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0 \\
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}
\tag{1.4}
$$

La última igualdad se obtiene utilizando las propiedades del operador suma. Podemos observar que $\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})$ implica $\sum_{i=1}^{n} \bar{x}(x_i - \bar{x}) = 0$ que es cierto porque $\sum_{i=1}^{n} \bar{x}(x_i - \bar{x}) = \bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}(\bar{x})^2 = n(\bar{x})^2 - n(\bar{x})^2$.

Una segunda manera a través de la cual se puede derivar el mismo estimador se conoce como el método de los momentos. La idea es muy sencilla. El supuesto **S1** nos dice que la media de u es cero ($E(u) = 0$) mientras que el supuesto **S1** y **S2** implican que la covarianza entre x y u es cero ($Cov(x, u) = 0$) o sea $E(xu) = 0$ puesto que $E(u) = 0$. Entonces

$$
\begin{aligned}
E(y - \beta_0 - \beta_1 x) = 0 \\
E(x(y - \beta_0 - \beta_1 x)) = 0
\end{aligned}
$$

El método de los momentos elige los estimadores $\hat{\beta}_0, \hat{\beta}_1$ de manera que solucionen las contrapartidas muestrales de las dos ecuaciones de arriba. O sea

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
n^{-1}\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0
\end{aligned}
$$

Multiplicando las dos ecuaciones por $2n$ obtenemos las mismas ecuaciones obtenidas de la minimización de la suma de los residuos al cuadrado. Resulta claro entonces que las soluciones deberán ser las mismas y los estimadores que encontraremos en este caso coincidirán con (1.3) y (1.4).

Podemos notar que $\hat{\beta}_1$ es simplemente el ratio entre la covarianza muestral y la varianza muestral de x

$$\begin{aligned}\hat{\beta}_1 &= \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}}{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \\ &= \frac{\hat{Cov(x,y)}}{\hat{Var(x)}}\end{aligned}$$

La interpretación de $\hat{\beta}_1, \hat{\beta}_0$ coincide a la de β_1, β_0 de la que hemos hablado anteriormente, ahora en términos de valores ajustados. Repetimos que $\hat{\beta}_1$ representa el cambio en $\hat{y}$ producido por un cambio unitario en x y $\hat{\beta}_0$ representa el valor de $\hat{y}$ cuando $x_i = 0$. Considerese el siguiente ejemplo.

Ejemplo 1.1 Utilizando datos para salario promedio horario (medidos en dólares por hora) y años de educación para 526 individuos, contenidos en el archivo WAGE1 en la base de datos Wooldridge en GRETL hemos estimado con MCO la siguiente regresión:

$$\hat{salario_i} = -0{,}90 + 0{,}54 educ_i$$

El valor estimado de la pendiente significa que un año más de educación hace que el salario aumente en 0,54 dólares por hora. ¿Cuál será el salario horario previsto para un individuo con ocho años de educación? La respuesta es $-0{,}90 + 0{,}54(8) = 3{,}42$ dólares por hora.

Una implicación clave de la estimación del modelo es que la recta de regresión muestral *no coincide* con la recta de regresión poblacional. Si aplicamos MCO a los datos en la Tabla 1 obtenemos

$$\hat{y}_i = 0{,}3458 + 0{,}3544 x_i$$

que es diferente de la regresión poblacional debido a que las estimaciones de los parámetros no coinciden con sus contrapartidas teóricas. El grafico 2 muestra la recta de regresión poblacional (la línea continua) vista antes y la nueva recta de regresión

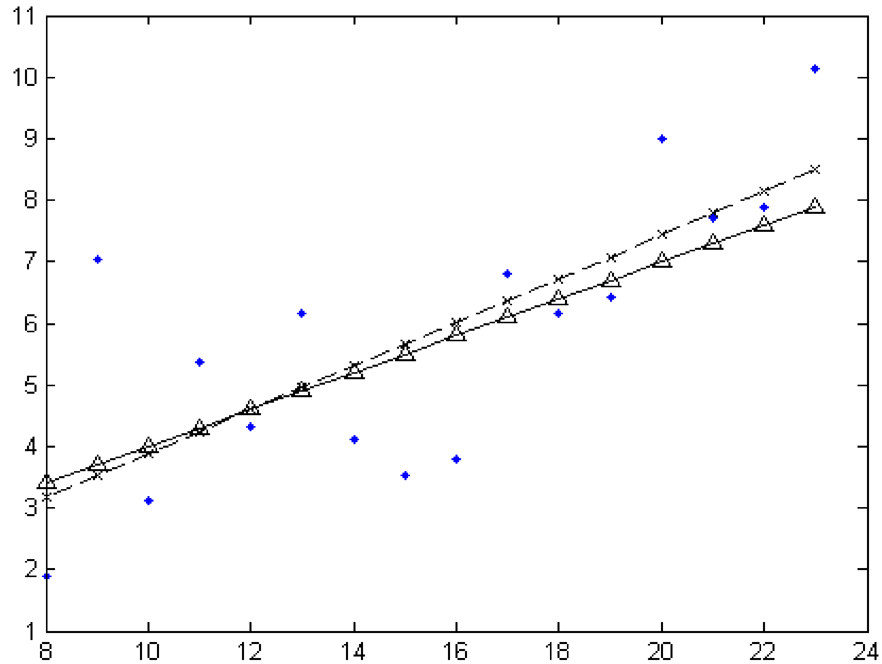

Figura 2.

muestral (la línea discontinua) estimada en este ejemplo. La razón que explica esta diferencia resultará clara más adelante.

1.2.2. Propiedades de los estimadores

Describiremos ahora las propiedades de los estimadores de MCO. Primero estudiaremos las propiedades algébricas y después las propiedades estadísticas de tales estimadores.

Propiedades algebraicas

P1 *La suma de los residuos es cero*, $\sum_{i=1}^{n} \hat{u}_i = 0$.

$$\begin{aligned} \sum_{i=1}^{n} \hat{u}_i &= \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 x_i \\ &= n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1\bar{x} \\ &= n\bar{y} - n(\bar{y} - \hat{\beta}_1\bar{x}) - n\hat{\beta}_1\bar{x} \\ &= 0. \end{aligned}$$

Las primeras tres igualdades se obtienen a través de las propiedades del operador suma y la cuarta substituyendo la formula MCO de $\hat{\beta}_0$.

P2 *La covarianza muestral entre regresores y residuos es cero*, $\sum_{i=1}^{n} \hat{u}_i x_i = 0$.

Ésta es la segunda condición del primer orden que hemos utilizado antes para derivar el estimador de mínimos cuadrados. Por esta razón esta condición siempre se cumple.

P3 *La media muestral de los valores ajustados es igual a la media muestral del los valores originales*: $\bar{\hat{y}} = \bar{y}$.

$$\hat{u}_i = y_i - \hat{y}_i$$

$$\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i \\
\sum_{i=1}^{n} y_i &= \sum_{i=1}^{n} \hat{y}_i \\
\bar{\hat{y}} &= \bar{y}
\end{aligned}$$

donde la tercera igualdad se obtiene a través de la propiedad **P1**.

Propiedades estadísticas

Antes de ver las propiedades estadísticas de los estimadores necesitamos repasar algunos conceptos y resultados estadísticos clave. Sea x una variable aleatoria con función de densidad normal con media μ y varianza σ^2

$$x \sim N(\mu, \sigma^2)$$

Si ahora restamos la media a x y dividimos por su desviación típica obtenemos la variable normal estandardizada

$$y = \frac{x - \mu}{\sigma} \sim N(0, 1)$$

Sean $z_i,\ i = 1, ..., k$ k variables normales estandardizadas independientes. Un resultado muy importante es que

$$w = \sum_{i=1}^{k} z_i^2 \sim \chi_k$$

o sea la suma de las z_i al cuadrado es una variable con distribución chi-cuadrado con k grados de libertad. Si además w es independiente de y entonces

$$\frac{y}{\sqrt{\frac{w}{k}}} \sim t_k$$

o sea una variable con distribución $t-$student con k grados de libertad. Ahora sean w_1 y w_2 dos variables χ^2 independientes con grados de libertad a y b respectivamente. El siguiente ratio

$$\frac{w_1/a}{w_2/b} \sim F_{a,b}$$

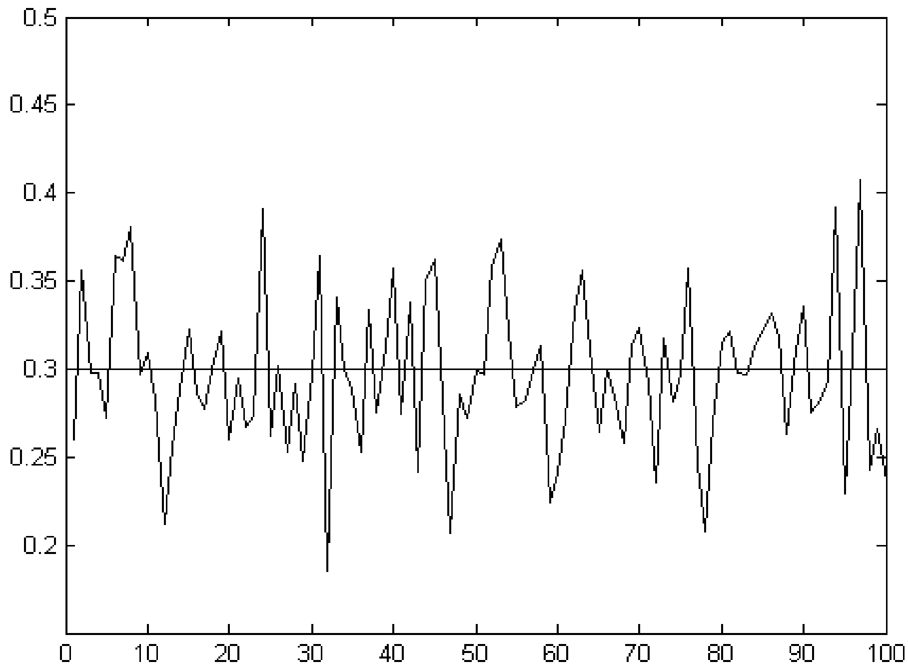

Figura 3

se distribuye con una distribución F con a y b grados de libertad.

Vamos a estudiar ahora las propiedades estadísticas de los estimadores MCO. Bajo los supuestos hechos anteriormente valen las siguientes propiedades.

P4 *Los estimadores $\hat{\beta}_0, \hat{\beta}_1$ son variables aleatorias.*

Veremos el caso de $\hat{\beta}_1$. Empezando con la fórmula del estimador MCO tenemos

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0 + \sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i + \sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}$$

Vamos a analizar los primeros dos términos a la derecha del igual. El primero

$$\begin{aligned}\frac{\sum_{i=1}^{n}(x_i-\bar{x})\beta_0}{\sum_{i=1}^{n}(x_i-\bar{x})^2} &= \frac{\beta_0\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ &= 0\end{aligned}$$

porque $\sum_{i=1}^{n}(x_i-\bar{x}) = n\bar{x}-\sum_{i=1}^{n}x_i = n\bar{x}-n\bar{x}$. El segundo

$$\begin{aligned}\frac{\sum_{i=1}^{n}(x_i-\bar{x})\beta_1 x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} &= \frac{\beta_1 x_i\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ &= \frac{\beta_1 x_i\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ &= \beta_1\end{aligned}$$

donde la segunda igualdad se obtiene de

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}) &= \sum_{i=1}^{n}(x_i-\bar{x})x_i-\sum_{i=1}^{n}(x_i-\bar{x})\bar{x} \\ &= \sum_{i=1}^{n}(x_i-\bar{x})x_i-\bar{x}\sum_{i=1}^{n}(x_i-\bar{x}) \\ &= \sum_{i=1}^{n}(x_i-\bar{x})x_i-\bar{x}(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\bar{x}) \\ &= \sum_{i=1}^{n}(x_i-\bar{x})x_i-\bar{x}(n\bar{x}-n\bar{x}) \\ &= \sum_{i=1}^{n}(x_i-\bar{x})x_i\end{aligned}$$

Resumiendo,

$$\hat{\beta}_1 = \beta_1+\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{1.5}$$

La (1.5) nos dice que el estimador MCO $\hat{\beta}_1$ es igual al parámetro β_1 más otro término que es una combinación lineal de las variables aleatorias u_i. Por esta razón $\hat{\beta}_1$ es también una variable aleatoria. Una directa implicación de este resultado es que para muestras aleatorias distintas la estimación del parámetro obtenida con MCO será distinta.

Ejemplo 1.2 Para entender mejor esta propiedad haremos el siguiente ejercicio. Consideremos el modelo que hemos tratado antes, $y_i = 1 + 0{,}3x_i + u_i$. Generamos 100 muestras aleatorias de 16 observaciones de u_i utilizando una $N(0, 0{,}16)$. Con cada una de estas muestras generamos 16 observaciones de la variables dependiente y. Por cada muestra volvemos a estimar el modelo de regresión anterior. La figura 3 muestra, por cada una de las muestra (eje x), el valor estimado de $\hat{\beta}_1$. Es evidente que cada una de las muestras proporciona una diferente estimación. La razón está en la propiedad que acabamos de ver: $\hat{\beta}_1$ es una variable aleatoria.

P5 *Los estimadores MCO son insesgados:* $E(\hat{\beta}_0) = \beta_0$ y $E(\hat{\beta}_1) = \beta_1$.

Esta propiedad se puede demostrar utilizando el resultado anterior. Empezamos con $\hat{\beta}_1$. Tomando la esperanza de $\hat{\beta}_1$ utilizando la (1.5) se obtiene

$$
\begin{aligned}
E(\hat{\beta}_1) &= E\left[\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] \\
&= \beta_1 + E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] \\
&= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})E(u_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \beta_1
\end{aligned}
$$

La segunda igualdad se obtiene aplicando las propiedades del valor esperando. La tercera se obtiene utilizando **S2** y la última con el supuesto **S1**.

Vamos ahora a demostrar la insesgadez de $\hat{\beta}_0$. Tomando el valor esperado en la fórmula del estimador obtenemos

$$
\begin{aligned}
E(\hat{\beta}_0) &= E(\bar{y} - \hat{\beta}_1\bar{x}) \\
&= E(\beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}) \\
&= E[\beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}] \\
&= \beta_0 + E[(\beta_1 - \hat{\beta}_1)\bar{x}] + E(\bar{u}) \\
&= \beta_0 + E[(\beta_1 - \hat{\beta}_1)\bar{x}] + E(\bar{u})
\end{aligned}
$$

$$
\begin{aligned}
&= \beta_0 + E\left(\frac{\sum_{i=1}^n u_i}{n}\right) \\
&= \beta_0 + \frac{\sum_{i=1}^n E(u_i)}{n} \\
&= \beta_0
\end{aligned}
$$

La segunda igualdad se obtiene substituyendo $\bar{y}$ y la última utilizando la propiedad de insesgadez de $\hat{\beta}_1$ vista antes, la definición de $\bar{u}$ y la propiedad **P1**. Ahora podemos establecer la siguiente propiedad relativa a la varianza de los estimadores MCO

P6 $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ y $Var(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$

Demostraremos sólo la primera parte de la propiedad. La demostración de la segunda parte la dejamos como ejercicio. Considerese la ecuación (1.5). Ésta implica

$$
\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}
$$

Utilizando **P5**, tomando el cuadrado y el valor esperado de los elementos obtenemos la siguiente expresión para la varianza de $\hat{\beta}_1$

$$
E\left[\hat{\beta}_1 - E(\hat{\beta}_1)\right]^2 = E\left[\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2\right].
$$

Sea $k_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Podemos escribir la ecuación de arriba como

$$
E\left[\hat{\beta}_1 - E(\hat{\beta}_1)\right]^2 = E\left(\sum_{i=1}^n k_i u_i\right)^2
$$

Para entender mejor el valor esperado a la derecha del igual supongamos que $n = 2$. En este caso

$$
\begin{aligned}
E(\sum_{i=1}^2 k_i u_i)^2 &= E\left(k_1^2 u_1^2 + k_2^2 u_2^2 + 2k_1 u_1 k_2 u_2\right) \\
&= E(k_1^2 u_1^2) + E(k_2^2 u_2^2) + E(2k_1 u_1 k_2 u_2) \\
&= k_1^2 E(u_1^2) + k_2^2 E(u_2^2) + 2k_1 k_2 E(u_1 u_2) \\
&= k_1^2 \sigma^2 + k_2 \sigma^2 \\
&= \sigma^2 (k_1^2 + k_2)
\end{aligned}
$$

Las primeras tres igualdades se obtienen utilizando las propiedades del valor esperado. La cuarta se obtiene utilizando **S3** u **S4**. En general, por un n cualquiera

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} k_i u_i\right)^2 &= \sigma^2 \sum_{i=1}^{n} k_i^2 \\
&= \sigma^2 \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (1.6)
\end{aligned}
$$

Otra propiedad del estimador MCO es la siguiente, conocida como Teorema de Gauss-Markov:

P7 *El estimador MCO es el estimador insesgado y lineal con varianza mínima.*

Esta propiedad nos dice que no existe un estimador con la propiedad de insesgadez y linealidad con varianza menor que el estimador MCO. Haremos ahora una prueba sencilla de esta propiedad para $\hat{\beta}_1$. Definimos otra vez $k_i = \frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$. Entonces, el estimador MCO es $\hat{\beta}_1 = \sum_{i=1}^{n} k_i y_i$. Definimos otro estimador lineal de β_1, le llamaremos β_i^* como $\beta_i^* = \sum_{i=1}^{n} w_i y_i$. Si substituimos y_i obtenemos $\beta_i^* = \alpha \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i u_i$. Tomando el valor esperado

$$
\begin{aligned}
E(\beta_i^*) &= E\left(\alpha \sum_{i=1}^{n} w_i\right) + E\left(\beta_1 \sum_{i=1}^{n} w_i x_i\right) + E\left(\sum_{i=1}^{n} w_i u_i\right) \\
&= \alpha \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i x_i \\
&= \beta_1.
\end{aligned}
$$

Para que se cumpla la propiedad de insesgadez de este nuevo estimador $\sum_{i=1}^{n} w_i = 0$ y $\sum_{i=1}^{n} w_i x_i = 1$. Vamos ahora a calcular la varianza de este nuevo estimador.

$$
\begin{aligned}
Var(\beta_1^*) &= Var\left(\sum_{i=1}^{n} w_i y_i\right) \\
&= \sum_{i=1}^{n} w_i^2 Var(y_i)
\end{aligned}
$$

$$
\begin{aligned}
&= \sigma^2 \sum_{i=1}^{n} w_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(w_i - \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(w_i^2 - \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)^2 + \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} + \\
&\quad +2\sigma^2 \sum_{i=1}^{n} \left(w_i^2 - \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right) \left(\frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)
\end{aligned}
$$

el último elemento a la derecha del igual es cero. El segundo $\sigma^2 \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$. Entonces la varianza de β_1^* será

$$
Var(\beta_1^*) = \sigma^2 \sum_{i=1}^{n} \left(w_i^2 - \frac{(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)^2 + \sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1.7}
$$

Esto implica que la varianza es mínima cuando $w_i = \frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ o sea β_1^* tiene que ser el estimador MCO $\hat{\beta}_1$.

1.2.3. Estimar la varianza del error

Hasta ahora hemos concentrado nuestra atención en los parámetros β_0, β_1 y en su estimación. Sin embargo, también σ^2 es un parámetro del cual no tenemos ninguna información. Este parámetro es particularmente importante porque, como hemos visto, de éste dependen las varianzas de los estimadores. Como será claro en la siguiente sección, obtener una estimación de la varianza del término de error es básico para poder hacer inferencia.

Antes de estudiar el estimador de la varianza del error, vamos a insistir en la diferencia existente entre los errores u_i y los residuos $\hat{u}_i$ de regresión. Hemos visto que el residuo de regresión es

$$
\begin{aligned}
\hat{u}_i &= y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i \\
&= \beta_0 + \beta_1 x_i + u_i - \hat{\beta}_0 + \hat{\beta}_1 x_i \\
&= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i
\end{aligned}
$$

Es claro que $\hat{u}_i$ u u_i no son iguales simplemente porque parámetros y estimadores no coinciden. La relación que hay entre residuo y error es que sus medias coinciden gracias a la propiedad de insesgadez de los estimadores MCO.

Volvemos ahora a la estimación de σ^2. Como $\sigma^2 = E(u_i^2)$ podríamos utilizar $\frac{\sum_{i=1}^n u_i^2}{n}$. El problema es que los errores no son observados. Lo que observamos son los valores estimados de los errores o sea los residuos de regresión. Entonces, una posibilidad sería utilizar la fórmula anterior con los residuos $\frac{\sum_{i=1}^n \hat{u}_i^2}{n}$. El problema es que este estimador es sesgado. El estimador insesgado que utilizamos en este curso es el siguiente:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-2} \tag{1.8}$$

La insesgadez se obtiene a través de la normalización por $n-2$ en lugar de n. Una vez obtenida una estimación para la varianza del error es muy sencillo estimar la varianza del estimador. Simplemente hay que substituir (1.8) en (1.6).

1.2.4. Bondad del ajuste

A menudo necesitamos una medida que nos proporcione información sobre la capacidad de la variable independiente x de explicar la variable dependiente y, o, en otras palabras, sobre hasta que punto la recta de regresión MCO se *ajusta* a los datos. Empezamos definiendo la Suma Total de los Cuadrados (STC), la Suma Explicada de los Cuadrados (SEC) y la Suma de los Cuadrados de los Residuos (SCR) de la siguiente manera:

$$\begin{aligned} STC &\equiv \sum_{i=1}^n (y_i - \bar{y})^2 \\ SEC &\equiv \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ SCR &\equiv \sum_{i=1}^n \hat{u}_i^2 \end{aligned}$$

STC mide la variabilidad muestral de las y_i ($STC/(n-1)$ es la varianza muestral). SEC mide la varianza muestral de las $\hat{y}_i$ (sabemos de **P3** que $\bar{\hat{y}}_i = \bar{y}$) y SCE mide la variación muestral de $\hat{u}_i$ (sabemos de **P1** que $\bar{\hat{u}} = 0$). Ahora vamos a demostrar que entre las tres cantidades existe la siguiente relación

$$STC = SEC + SCR$$

o sea la suma total de los cuadrados es igual a la suma de los cuadrados de los residuos y la suma explicada. De su definición

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(\hat{u}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}\hat{u}_i^2 + \sum_{i=1}^{n}2\hat{u}_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
&= SCR + \sum_{i=1}^{n}2\hat{u}_i(\hat{y}_i - \bar{y}) + SEC
\end{aligned}$$

Lo único que necesitamos demostrar es que el segundo término de la expresión de arriba es cero.

$$\begin{aligned}
\sum_{i=1}^{n}2\hat{u}_i(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n}2\hat{u}_i(\hat{y}_i - \bar{y}) \\
&= \sum_{i=1}^{n}2\hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}) \\
&= \sum_{i=1}^{n}2\hat{u}_i(\bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1 x_i - \bar{y}) \\
&= \sum_{i=1}^{n}2\hat{u}_i\hat{\beta}_1(x_i - \bar{x}) \\
&= 2\hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i x_i - 2\hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i\bar{x} \\
&= 2\hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i x_i - 2\hat{\beta}_1\bar{x}\sum_{i=1}^{n}\hat{u}_i \\
&= 0
\end{aligned}$$

por las propiedades **P1** y **P2**. Si dividimos todo por STC obtenemos

$$1 = \frac{SEC}{STC} + \frac{SCR}{STC}$$

Definimos el *R-cuadrado* de la regresión, o coeficiente de determinación, de la siguiente manera:

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

y se interpreta como *la fracción de la variación muestral de y_i explicada por x_i*. El R^2 siempre está entre cero y uno.

Ejemplo 1.1 (cont) En el modelo estimado en el ejemplo 1 encontramos que $R^2 = 0{,}16$. Este valor para R^2 significa que el 16 % de la variabilidad de la variable *salario* es explicada por nuestra variable independiente *educ*.

1.2.5. Cambio de escala y formas funcionales

Cambio de escala

En el ejemplo anterior, hemos medido el salario en miles de dólares. Conocer la unidad de medida de las variables es necesario para poder interpretar correctamente los parámetros estimados. Vamos a estudiar ahora cómo cambian las estimaciones de nuestro modelo cuando cambiamos la escala de las variables dependiente y independiente.

Caso 1: multiplicamos y por una constante c

Si multiplicamos la variable dependiente por una constante c y estimamos la nueva regresión, el estimador MCO de la pendiente será

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(cy_i - c\bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= c\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= c\hat{\beta}_1
\end{aligned}$$

mientras que el nuevo estimador de β_0 será

$$\tilde{\beta}_0 = c\bar{y} - \hat{\beta}_1^*\bar{x}$$

$$
\begin{aligned}
&= c\bar{y} - c\hat{\beta}_1\bar{x} \\
&= c\hat{\beta}_0
\end{aligned}
$$

En este caso, con la nueva variable dependiente cy, los dos estimadores serán los estimadores obtenidos con y multiplicados por la constante c.

Caso 2: multiplicación de x por una constante c

Si multiplicamos la variable dependiente por una constante c y estimamos la nueva regresión, el estimador MCO de la pendiente será

$$
\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum_{i=1}^{n}(cx_i - c\bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(cx_i - c\bar{x})^2} \\
&= c\frac{c\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{c^2\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{1}{c}\hat{\beta}_1
\end{aligned}
\tag{1.9}
$$

mientras que el nuevo estimador de β_0 será

$$
\begin{aligned}
\tilde{\beta}_0 &= \bar{y} - \hat{\beta}_1^*\bar{c}x \\
&= \bar{y} - \frac{1}{c}\hat{\beta}_1 c\bar{x} \\
&= \hat{\beta}_0
\end{aligned}
$$

En este caso, con la nueva variable dependiente, la estimación de la pendiente se divide por la constante c, mientras que la estimación de β_0 no varía.

Formas funcionales

Hasta ahora hemos considerados relaciones lineales entre variables, por ejemplo salario y educación. Sin embargo, en muchos trabajos econométricos, encontraremos otros tipos de relaciones. En particular, encontraremos a menudo casos en que una o las dos variables aparecen en forma logarítmica. En este apartado, estudiaremos la interpretación de los parámetros en el caso en que una o las dos variables del modelo de regresión sean especificadas en logaritmos.

Caso 1: log-nivel

Supongamos que nuestro modelo de regresión sea el siguiente

$$\log(y) = \beta_0 + \beta_1 x + u$$

Si $\Delta u = 0$ entonces $\beta_1 = \frac{\Delta \log(y)}{\Delta x}$. Esto significa que $100\beta_1$ representa el cambio en términos porcentuales en y si x aumenta en una unidad.

Ejemplo 3 Consideremos el siguiente modelo

$$\log(salario_i) = \beta_0 + \beta_1 educ_i + u_i$$

Utilizando los datos del Ejemplo 1 obtenemos las siguientes estimaciones

$$\log(s\hat{a}lario_i) = 0{,}58 + 0{,}08 educ_i$$

donde el parámetro $\hat{\beta}_1 = 0{,}08$ nos dice que un año más de educación determina un aumento en el salario horario de un 8 %. Observese la diferencia con el caso anterior donde las dos variables eran especificadas en niveles. En ese caso, β_1 expresaba el cambio en unidades de *salario* (miles de euros) cuando *educ* aumentaba en una unidad (año).

Caso 2: nivel-log

Supongamos que nuestro modelo de regresión ahora sea el siguiente

$$y = \beta_0 + \beta_1 \log(x) + u$$

Si $\Delta u = 0$, entonces $\beta_1 = \frac{\Delta y}{\Delta \log(x)}$. Esto significa que $\beta_1/100$ representa el cambio en de y en términos de su unidad de medida si x aumenta en un 1 %.

Ejemplo 4 Ahora utilizando el archivo CEOSALES1 de la bases de datos Wooldridge en GRETL estimamos la siguiente regresión

$$salary_i = \beta_0 + \beta_1 \log(sales_i) + u_i$$

donde $salary_i$ se refiere al salario de un ejecutivo de la empresa i en miles de dólares y $sales$ las ventas de esta empresa en miliones de dólares. Utilizando el estimador MCO obtenemos

$$sal\hat{a}ry_i = -898{,}93 + 262{,}9 \log(sales_i)$$

donde el parámetro $\hat{\beta}_1 = 262{,}9$ nos dice que si la ventas aumentan en un 1 %, el salario de un ejecutivo aumenta de $262{,}9/100 = 2{,}629$ miles de dólares.

Caso 3: log-log

Supongamos que nuestro modelo de regresión ahora sea el siguiente

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

Si $\Delta u = 0$ entonces $\beta_1 = \frac{\Delta \log(y)}{\Delta \log(x)}$. En este caso β_1 tiene la interpretación de elasticidad y expresa de cuánto varía en términos porcentuales y si x aumenta en un 1 %.

Ejemplo 4 (cont.) Utilizando el mismo archivo del ejemplo anterior CEOSALES1 estimamos la siguiente regresión:

$$\log(salary_i) = \beta_0 + \beta_1 \log(sales_i) + u_i$$

las estimaciones nos dan

$$\log(s\hat{a}lary_i) = 4{,}82 + 0{,}25 \log(sales_i)$$

donde el parámetro $\hat{\beta}_1 = 0{,}25$ ahora nos dice si la ventas aumentan en un 1 % el salario de un ejecutivo aumenta de 0,25 %.

La tabla 2 resume la interpretación en todos los diferentes casos.

Model	Dependiente	Independiente	Interpretación β_1
nivel-nivel	y	x	$\Delta y = \beta_1 \Delta x$
nivel-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\,\%\Delta x$
log-nivel	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1\,\%\Delta x$

Tabla 2

1.3. Inferencia

Vamos a tratar ahora el tema de la inferencia. Básicamente estudiaremos el contraste de una hipótesis estadística, el contraste t, y los intervalos de confianza para un único parámetro.

Para poder desarrollar y comprender el análisis inferencial necesitamos añadir el siguiente supuesto al modelo:

S5 *El error se distribuye normalmente con media cero y varianza* σ^2, $u_i \sim N(0, \sigma^2)$

Este supuesto implica otra propiedad del estimador MCO o sea que el estimador se distribuye normalmente con media el verdadero parámetro y varianza la obtenida anteriormente en la propiedad **P6**.

P8 $\beta_i \sim N(\beta_i, Var(\hat{\beta}_i)), i = 0, 1.$

hemos visto antes, en la derivación de la **P4**, que el estimador MCO es una combinación lineal de las u_i que, bajo el supuesto **S5** son variables normales independientes. Un resultado fundamental nos dice que tal combinación es también una variable normal.

Además de este nuevo supuesto, necesitamos también repasar algunos resultados estadísticos adicionales. Primero, como visto antes, si al estimador restamos su media y dividimos por su desviación típica obtenemos una variable normal estandardizada

$$\frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)} \sim N(0,1) \tag{1.10}$$

donde $se(\hat{\beta}_i) = \sqrt{Var(\hat{\beta}_i)} = \sigma[\sum_{i=1}^n (x_i - \bar{x})^2]^{-1/2}$. Segundo (que se demostrará mas adelante)

$$\frac{\sum_{i=1}^n \hat{u}_i^2}{\sigma^2} = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}$$

o sea el ratio entre el estimador de la varianza multiplicado por $n-2$ y la varianza de término de error es una variable con distribución χ^2 con $n-2$ grados de libertad. Tercero, si substituimos σ con $\hat{\sigma}$ en (1.10) obtenemos una variable con distribución $t-$student

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}[\sum_{i=1}^n (x_i - \bar{x})^2]^{-1/2}} = \frac{\hat{\beta}_i - \beta_i}{\left(\frac{\hat{\sigma}}{\sigma}\right)([\sigma \sum_{i=1}^n (x_i - \bar{x})^2]^{-1/2})}$$

El numerador de este ratio es una variable normal estandardizada y el denominador es la raíz cuadrada de una variable con distribución χ_{n-2} dividido por sus grados de libertad, o sea, la definición de una variable $t-$student con $n-2$ grados de libertad. Entonces, el segundo resultado importante es que

$$\frac{\hat{\beta}_i - \beta_i}{\widehat{se(\hat{\beta}_i)}} \sim t_{n-2} \tag{1.11}$$

donde $\widehat{se(\hat{\beta}_i)} = \sqrt{\widehat{Var(\hat{\beta}_i)}} = \hat{\sigma}[\sum_{i=1}^n (x_i - \bar{x})^2]^{-1/2}$.

1.3.1. Contraste de hipóthesis, el contraste t

El primer instrumento de análisis inferencial que vamos a estudiar es el contraste de hipótesis. El contraste de hipótesis sirve para *contrastar* o averiguar determinadas hipótesis estadísticas sobre los parámetros del modelo. Para poder desarrollar el contraste necesitamos antes aclarar el concepto de *hipótesis nula* y de *hipótesis alternativa.* La hipótesis nula es una hipótesis estadística, una afirmación sobre algunos de los parámetros del modelo, que se asume ser cierta al principio del contraste. En este curso, especificaremos la hipótesis nula (H_0) de la manera siguiente:

$$H_0 : \beta_i = \beta_i^0$$

Una hipótesis que contrastaremos a menudo es que la pendiente de la recta de regresión sea igual a cero

$$H_0 : \beta_1 = 0.$$

El contraste de esta hipótesis es un caso muy interesante porque trata de averiguar si entre la variable independiente y dependiente existe alguna relación. En el ejemplo anterior de salario y educación contrastar la hipótesis $H_0 : \beta_1 = 0$ equivale a contrastar la afirmación que la educación no sea importante para explicar el salario. Si no podemos rechazar tal hipótesis tendríamos que concluir que no hay ninguna relación estadísticamente significativa entre estas dos variables.

La hipótesis alternativa es la hipótesis frente a la cual evaluamos la hipótesis nula, o sea la hipótesis que "aceptamos" si rechazamos la hipótesis nula. Consideraremos dos tipos distintos de hipótesis alternativas correspondientes a contrastes a una o dos colas. Un primer tipo de hipótesis alternativa (contraste a una sola cola) es el siguiente

$$H_0 : \beta_i < \beta_i^0 \quad \text{o} \quad H_0 : \beta_i > \beta_i^0$$

Un segundo tipo de hipótesis, correspondiente a contraste a dos colas, es

$$H_0 : \beta_1 \neq \beta_i^0.$$

Como se ha dicho anteriormente, el contraste de hipótesis trata de averiguar si una determinada hipótesis nula se puede considerar compatible o no con lo datos. Una vez especificada la hipótesis nula y la alternativa, necesitamos una medida de discrepancia entre los datos y la hipótesis nula. O sea, necesitamos una cantidad que permita juzgar si la hipótesis que estamos contrastando es compatible con lo que observamos en los datos. Esta medida toma el nombre de *estadístico de contraste.* El estadístico que se utiliza es (1.11). Como hemos dicho antes, por definición, la hipótesis nula se considera cierta al principio de la prueba de hipótesis. Esto significa que cuando consideramos nuestro estadístico de contraste tenemos que hacerlo *bajo* el supuesto que la hipótesis nula es cierta. En general, bajo la hipótesis nula $H_0 : \beta_i = \beta_i^0$ el estadístico de contraste será

$$\frac{\hat{\beta}_i - \beta_i^0}{\hat{se}(\hat{\beta}_i)} \sim t_{n-2}$$

Por ejemplo, si $H_0 : \beta_1 = 0$, nuestro estadístico de contraste será

$$\frac{\hat{\beta}_i}{\hat{se}(\hat{\beta}_i)} \sim t_{n-2}$$

Es importante entender porqué este estadístico puede proporcionar información que nos permita decidir si la H_0 es compatible o no con los datos. Hemos visto que este estadístico es una variable con distribución $t-$student. Una vez obtenidas las estimaciones de los parámetros del modelo, podemos calcular el valor del estadístico de contraste bajo la hipótesis nula. Esto nos proporcionará un número, digamos $\hat{t}_{n-2}$. Supongamos ahora que este número se encuentre muy alejado del valor promedio de la distribución. Esto implica que si los datos fueran de verdad generados por el modelo de regresión con el valor del parámetro implicado por la hipótesis nula, sería "improbable" observar lo que de hecho observamos en los datos, o, en otras palabras, seria "improbable" obtener los valores obtenidos para las estimaciones de β_i y $se(\hat{\beta}_i)$. Esto tendría que sugerir que la hipótesis nula es incompatible con nuestros datos y por esta razón debeamos rechazarla.

El problema es: ¿cuánto el estadístico de contraste tiene que alejarse de su valor medio para poder rechazar la hipótesis nula? Para esto necesitamos especificar un *nivel de significatividad* α para el contraste. Éste representa la probabilidad de rechazar la hipótesis nula cuando ésta es cierta, o sea, la probabilidad de cometer un error de tipo I. Una vez establecido el nivel de significatividad, podemos encontrar el valor critico de la distribución t correspondiente a ese nivel. Definimos el valor crítico $t^*_{n-2,\frac{\alpha}{2}}$, asociado a α, como aquel valor tal que $p(|t| > t^*_{n-2,\frac{\alpha}{2}}) = \alpha$, o sea aquel valor tal que la probabilidad que la variable en valor absoluto tome valores mayores que el valor crítico es igual al nivel de significatividad α.

Vamos ahora a ver en detalle como desarrollar el contraste de hipótesis. Consideraremos dos contrastes distintos: a una y dos colas.

Contraste a dos colas

Supongamos de querer contrastar la $H_0 : \beta_i = \beta_i^0$ frente a la alternativa $H_1 : \beta_i \neq \beta_i^0$. El contraste de hipótesis se desarrollará a través de los pasos siguientes:

1. Calcular, utilizando las estimaciones obtenidas, el estadístico de contraste bajo la hipótesis nula $\frac{\hat{\beta}_i - \beta_i^0}{\hat{se}(\hat{\beta}_i)} = \hat{t}_{n-2}$.

2. Establecer un nivel de significatividad α.

3. Comparar el estadístico de contraste con el valor crítico para un nivel de significatividad α. Si $|\hat{t}_{n-2}| > t^*_{n-2\frac{\alpha}{2}}$ rechazo la H_0. En caso contrario no rechazo la hipótesis nula. La idea que está detrás de este criterio de rechazo es que si $|\hat{t}_{n-2}| > t^*_{n-2\frac{\alpha}{2}}$ esto significa que $p(|t_{n-2}| > \hat{t}_{n-2}) < \alpha$, o sea el valor $\hat{t}_{n-2}$ estará muy alejado del promedio de la distribución y por tanto concluimos que la hipótesis nula es poco compatible con los datos.

Ejemplo 1 (cont.) En el modelo estimado en el ejemplo 1 encontramos estos valores para las desviaciones típicas de los estimadores: $\hat{se}(\hat{\beta}_0) = 0{,}16$, $\hat{se}(\hat{\beta}_1) = 0{,}05$. Podemos contrastar la hipótesis nula que los años de educación no tienen efecto alguno sobre el salario. Esta hipótesis se puede formular como $H_0 : \beta_1 = 0$. Para desarrollar el contraste tenemos que calcular el valor de estadístico de contraste

$$\begin{aligned} \frac{\hat{\beta}_1}{\hat{se}(\hat{\beta}_1)} &= \frac{0{,}54}{0{,}05} \\ &= 10{,}8 \end{aligned}$$

y comparar el valor obtenido con el valor crítico de la distribución $t-$student con $n-2 = 526-2 = 524$ grados de libertad a un determinado nivel de significatividad, digamos $\alpha = 0{,}05$. El valor crítico es $t^*_{524,0{,}025} = 1{,}9645 < \hat{t}$. Entonces, rechazamos la hipótesis nula al 5 %. Supongamos ahora que la hipótesis estadística que nos interesa contrastar sea que un año más de educación corresponde a un dólar por hora mas. Esta hipótesis se puede especificar como $H_0 : \beta_1 = 1$. Si calculamos el nuevo valor del estadístico de contraste obtenemos

$$\begin{aligned} \left|\frac{\hat{\beta}_1 - 1}{s\hat{e(\hat{\beta}_1)}}\right| &= \left|\frac{-0{,}44}{0{,}05}\right| \\ &= 8{,}8 \end{aligned}$$

Otra vez rechazamos la hipótesis nula al 5 % porque el valor absoluto del estadístico de contraste es mayor que el valor crítico.

Una propiedad muy importante del contraste de hipótesis es que si hemos rechazado una hipótesis a un nivel de significatividad α, rechazaremos la misma hipótesis para todos los niveles de significatividad mayor que α. La razón es que si $\alpha' > \alpha$ entonces $t^*_{n-2,\frac{\alpha'}{2}} < t^*_{n-2,\frac{\alpha}{2}}$ porque $\frac{\alpha}{2}$ representa el área bajo la distribución en el intervalo entre $t^*_{n-2,\frac{\alpha}{2}}$ y infinito.

Contraste a una cola

Supongamos de querer contrastar la $H_0 : \beta_i = \beta_i^0$ frente a la alternativa $H_1 : \beta_i > \beta_i^0$. El contraste de hipótesis se desarrollará a través de los pasos siguientes:

1. Calcular, utilizando las estimaciones obtenidas, el estadístico de contraste bajo la hipótesis nula $\frac{\hat{\beta}_i - \beta_i^0}{s\hat{e(\hat{\beta}_i)}} = \hat{t}_{n-2}$.

2. Establecer un nivel de significatividad α.

3. Comparar el estadístico de contraste con el valor crítico para un nivel de significatividad α. Si $\hat{t}_{n-2} > t^*_{n-2,\alpha}$ rechazo la H_0. En caso contrario no rechazo la hipótesis nula.

Supongamos que queremos contrastar la $H_0 : \beta_i = \beta_i^0$ frente a la alternativa $H_1 : \beta_i < \beta_i^0$. El contraste de hipótesis se desarrollará a través de los pasos siguientes:

1. Calcular, utilizando las estimaciones obtenidas, el estadístico de contraste bajo la hipótesis nula $\frac{\hat{\beta}_i - \beta_i^0}{se(\hat{\beta}_i)} = \hat{t}_{n-2}$.

2. Establecer un nivel de significatividad α.

3. Comparar el estadístico de contraste con el valor crítico para un nivel de significatividad α. Si $\hat{t}_{n-2} < -t^*_{n-2,\alpha}$ rechazo la H_0. En caso contrario no rechazo la hipótesis nula.

Notese que mientras en el contraste a dos colas el valor crítico de referencia es $t^*_{n-2,\frac{\alpha}{2}}$ para un nivel de significatividad α, en el contraste a una cola será $t^*_{n-2,\alpha}$ o $-t^*_{n-2,\alpha}$

Valor-p

A veces distintos niveles de confianza conducen a distintas decisiones sobre la hipótesis nula. El problema es que la elección del nivel de confianza es una decisión subjetiva del investigador ya que no existe un nivel *correcto*. Para esta razón, a menudo, es informativo calcular el valor p para el contraste. El valor p es el nivel de significación mínimo al que se puede rechazar la hipótesis nula y se define (en el contraste a dos colas) como

$$\text{valor-}p = p(|t| > |\hat{t}|)$$

o sea, indica la probabilidad que la variable t tome, en valor absoluto, valores mayores que el valor obtenido para el estadístico de contraste $\hat{t}$. Queda claro entonces porque el valor-p indica el nivel de significatividad mínimo al cual se puede rechazar la H_0. Por cada $\alpha >$ valor-p el valor crítico correspondiente es menor que $\hat{t}$ y por esta razón rechazamos la hipótesis nula.

1.3.2. Intervalos de confianza

El segundo instrumento para la inferencia es el de los intervalos de confianza para un parámetro. En este caso, de manera diferente al contraste de hipótesis, donde tenemos que decidir sobre una hipótesis nula, la idea es de encontrar un intervalo tal que, con una determinada probabilidad, contenga el verdadero valor del parámetro.

Utilizando lo que hemos visto antes sabemos que $1-\alpha$ corresponde a la probabilidad que el estadístico de contraste (una variable $t-$student) tome valores en el intervalo definido por menos y mas el valor crítico a nivel de significatividad α. Formalmente

$$p\left(-t^*_{n-2,\frac{\alpha}{2}} < \frac{\hat{\beta}_i - \beta_i}{\widehat{se(\hat{\beta}_i)}} < t^*_{n-2,\frac{\alpha}{2}}\right) = 1-\alpha \tag{1.12}$$

Vamos a ver ahora como de aquí podemos construir un intervalo de confianza para el parámetro β_i. Desarrollando la probabilidad de arriba obtenemos

$$\begin{aligned} p\left(-t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)} < \hat{\beta}_i - \beta_i < t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}\right) &= 1-\alpha \\ p\left(-\hat{\beta}_i - t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)} < -\beta_i < -\hat{\beta}_i - t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}\right) &= 1-\alpha \\ p\left(\hat{\beta}_i - t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)} < \beta_i < \hat{\beta}_i + t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}\right) &= 1-\alpha \end{aligned}$$

Esto nos dice que con probabilidad $1-\alpha$ el parámetro β_i estará incluido en el intervalo $[\hat{\beta}_i - t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}, \hat{\beta}_i + t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}]$. Esto es un intervalo de confianza para el parámetro β_i.

Ejemplo 1.1 (cont.) Es muy sencillo encontrar un intervalo de confianza al 5 % para β_1 utilizando las estimaciones del Ejemplo 1. Sabemos que $\hat{\beta}_1 = 0{,}54$, $\widehat{se(\hat{\beta}_1)} = 0{,}05$ y $t*_{524,0,025} = 1{,}9645$. Aplicando la fórmula vista arriba obtenemos el siguiente intervalo de confianza $[\hat{\beta}_i - t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}, \hat{\beta}_i + t^*_{n-2,\frac{\alpha}{2}}\widehat{se(\hat{\beta}_i)}] = [0{,}54 - (0{,}05)(1{,}96), 0{,}54 + (0{,}05)(1{,}96)] = [0{,}44, 0{,}64]$. Podemos notar que el intervalo no incluye el valor cero y esto es consistente con el resultado del contraste de la hipótesis $H_0 : \beta_1 = 0$ que hemos visto antes.

Ejercicios

(1) Queremos analizar el efecto del gasto en campaña electoral sobre el porcentaje de votos. Disponemos de una base de datos de 173 observaciones que incluye: $voteA$, el porcentaje de votos al candidato A, $expendA$, el gasto del candidato A en su campaña electoral (en miles de dólares), $expendB$, el gasto del candidato B (solo hay dos candidatos) en su campaña electoral (en miles de dólares). El resultado de la estimación por mínimos cuadrados nos da la siguiente regresión

$$voteA_i = 43{,}1712 + 0{,}0236 expendA_i + \hat{u}_i \quad R^2 = 0{,}15614$$

(a) ¿Cuál es el efecto de un aumento de 1000 dólares en el gasto de A sobre el porcentaje de votos de A?

(b) Contrastar al 5 % de significación la hipótesis nula $H_0 : \beta_1 = 0$ utilizando como valor crítico 1.96 y sabiendo que la desviación típica de $\hat{\beta}_1$ es 0,01 .

(c) La suma de los cuadrados de los residuos (SCR) es 40891.1 Hallar la desviación típica de los residuos $\hat{u}_i$.

(d) ¿Qué parte de la variabilidad de $voteA$ explica nuestro modelo de regresión?

(e) Calcular la varianza muestral de y.

(2) Este es un ejercicio relativo al contraste de la racionalidad en el cálculo de los precios de las viviendas. En el modelo de regresión simple

$$price = \beta_0 + \beta_1 assess + u$$

donde $price$ es el precio de la vivienda y $assess$ es la valoración de la vivienda antes de la venta, la valoración es racional si $\beta_1 = 1$ y $\beta_0 = 0$. La ecuación estimada es

$$price = -14{,}47 + 0,976 assess$$

$$n = 88, SRC = 165644{,}51, R^2 = 0,820 \tag{1.13}$$

donde 16,27 es la desviación típica de $\hat{\beta}_0$ y 0,049 es la desviación típica de $\hat{\beta}_1$.

(a) Contrastar la hipótesis H_0: $\beta_0 = 0$.

(b) Contrastar H_0: $\beta_1 = 1$.

(c) Encontrar un intervalo de confianza al 95 % para β_1.

(d) Encontrar un intervalo de confianza al 95 % para β_0.

(3) Demostrar la propiedad **P6** para $\hat{\beta}_0$.

(4) Demostrar que R^2 equivale al cuadrado del coeficiente de correlación entre x y y.

(5) Demostrar que si un intervalo de confianza al 95 % para β_1 no incluye el cero, esto implica que rechazamos la hipótesis nula $H_0 : \beta_1 = 0$ al 5 % utilizando el contraste t.

(6) Supongamos que $y_i = \beta_0 + \beta_1 x_i + u_i$ y que todos los supuestos vistos en este capítulo sean satisfechos. Utilzando los datos de la tabla siguiente

y	x
6.3000	2.0000
9.5000	4.0000
7.4000	3.0000
2.4000	1.0000
0.9000	0

(a) Estimar los parámetros del modelo con MCO

(b) Estimar la varianza del error.

(c) Encontrar R^2.

(d) Contrastar la hipótesis nula $H_0 : \beta_1 = 0$.

(e) Contrastar la hipótesis nula $H_0 : \beta_0 = 1$.

(7) Estamos interesados en analizar la relación entre dos variables x_i y y_i. Por esto se proponen los dos modelos siguientes

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$x_i = \gamma_0 + \gamma_1 y_i + v_i$$

Si estimamos los parámetros de los dos modelos utilizando MCO ¿es siempre cierto que $\beta_1 = \frac{1}{\gamma_1}$? Si no ¿en que casos?

Aplicaciones con GRETL

(1) Ejercicio 2.10 Wooldridge.

(2) Ejercicio 2.11 Wooldridge.

(3) Ejercicio 2.12 Wooldridge.

Capítulo 2

Modelo de Regresión Lineal con k-Variables

2.1. El modelo

En muchas aplicaciones es natural pensar que una variable económica de interés pueda depender de más de una variable exógena. Por esta razón, veremos ahora cómo generalizar el modelo estudiado en el capitulo anterior. Estudiaremos en esta sección el modelo de regresión lineal con k-variables. A diferencia del modelo simple este modelo admite más de una variable como regresor. El modelo de regresión lineal con k-variables puede expresarse a través de la siguiente ecuación

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i \tag{2.1}$$

donde i, como antes, se refiere a la observación i en nuestra muestra y

1. $y_i -$ es la variable que queremos explicar y recibe el nombre de *variable dependiente* o *variable explicada.*

2. $x_{ij}, \quad j = 1, ..., k -$ son las k variables a través de las cuales queremos explicar y y reciben, como antes, el nombre de *variables independientes* o *explicativas.*

3. $u_i -$ es el término de error, es una variables aleatoria y representa factores no observables distintos a x_j que afectan a y.

4. $\beta_j, \;\; j = 0, ..., k -$ son los parámetros del modelo.

En el caso de k variables es muy útil escribir el modelo en forma matricial. Sea n el número de observaciones en nuestra muestra aleatoria. Definimos

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

un vector $(n \times 1)$ de todas las observaciones de la variable dependiente

$$\mathbf{X} = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_n} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

una matriz $(n \times k)$ con todas las observaciones en las filas y las variables independientes en las columnas

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

un vector $(n \times 1)$ de errores y

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

un vector $((k+1) \times 1)$ donde el primer elemento es el coeficiente del término constante y los demás son los coeficientes de las variables exógenas. Entonces podemos escribir el modelo como

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \tag{2.2}$$

Como en el capitulo anterior, añadimos al supuesto de linealidad del modelo los supuestos siguientes:

S1' *La esperanza de u_i es zero: $E(u_i) = 0$ por cada i.*

S2' *Las variables dependientes son non aleatorias o fijas en muestras repetidas.*

S3' *La varianza de u_i es constante: $Var(u_i) = E(u_i^2) = \sigma^2$ por cada i.*

S4' *u_i y u_j son independientes por cada i y cada j con $i \neq j$.*

S5' *El rango de $\mathbf{X}$ es $k+1$.*

Como el caso del modelo simple un supuesto alternativo a S2 que podemos hacer es el siguiente.

2.2. Estimación

2.2.1. Estimador Mínimos Cuadrados Ordinarios (MCO)

Como en el modelo simple, el único estimador que veremos en el modelo a k-variables es el estimador MCO. Para poder derivar el estimador necesitamos algunos resultados y definiciones preliminares. Sea $\hat{\beta}$ un estimador de β. Definimos el vector de residuos de regresión como $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$ y el vector de valores ajustados $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$. Como antes, el estimador de MCO es el estimador que minimiza la suma de los residuos al cuadrado $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \sum_{i=1}^{n} \hat{u}_i^2$. Antes de planear y solucionar el problema de minimización es útil desarrollar esta suma de la manera siguiente:

$$
\begin{aligned}
\hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\
&= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\
&= \mathbf{y}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - 2\hat{\beta}'\mathbf{X}'y
\end{aligned}
$$

Se puede notar que $2\hat{\beta}'\mathbf{X}'y = \hat{\beta}'\mathbf{X}'y + y'\mathbf{X}\hat{\beta}$ porque $\hat{\beta}'\mathbf{X}'y = y'\mathbf{X}\hat{\beta}$. La razón es que el término a la derecha es el traspuesto del término a la izquierda que es un escalar y esto implica que los dos serán iguales. Entonces podemos formular el problema de minimización de la manera siguiente:

$$\min_{\hat{\beta}} \quad \mathbf{y}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - 2\hat{\beta}'\mathbf{X}'\mathbf{y}$$

Como antes, el problema se soluciona tomando las condiciones del primer orden, o sea las derivadas parciales respecto a $\hat{\beta}$ igualadas a cero. La única diferencia con al caso anterior es que aquí tenemos que utilizar las reglas de derivaciones respecto a un vector y no un escalar. Las condiciones del primer orden son las siguientes

$$\mathbf{X}'\mathbf{X}\hat{\beta} - \mathbf{X}'\mathbf{y} = 0.$$

Solucionando respecto al vector de estimadores obtenemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.3}$$

Se puede demostrar que cada uno de los estimadores contenidos en el vector $\hat{\beta}$ se puede escribir como

$$\hat{\beta}_j = \frac{\sum_{i=1}^n y_i \hat{r}_{ij}}{\sum_{i=1}^n \hat{r}_{ij}^2}$$

donde $\hat{r}_j$ es el residuo de la regresión de x_j sobre las demás variables independientes. Este resultado es muy útil para poder entender la interpretación de cada uno de los elementos en el vector $\hat{\beta}$. Dado que $\hat{r}_j$ es la parte de x_j que no está relacionada con las demás variables independientes, $\hat{\beta}_j$ mide el efecto de x_j sobre y una vez que descontamos los efectos de las demás variables exógenas. Por eso, podemos interpretar $\hat{\beta}_j$ como el efecto parcial de x_j sobre y o sea el efecto de x_j cuando las demás variables se mantienen fijas.

Ejemplo 2.1 En el modelo utilizado en el Ejemplo 1.1 para explicar el salario horario, añadimos otro regresor: los años trabajados en el actual puesto de trabajo. Utilizando el estimador que acabamos de ver obtenemos las siguientes estimaciones

$$\log(\widehat{salario_i}) = 0{,}216 + 0{,}097 educ + 0{,}010 exper.$$

La interpretación de $\hat{\beta}_1 = 0{,}097$ es que ahora un año más de educación produce un incremento del salario de 9.7 %, mientras que un año más de experiencia laboral aumenta el salario en un 1 %.

El ejemplo anterior evidencia una caracteristica muy importante del modelo a k-variables. En general, si añadimos una variable al modelo las estimaciónes de los parámetros que ya eran incluidos cambian. Este resultado queda claro en el ejemplo que acabamos de ver, donde en el caso de un solo regresor $\hat{\beta}_1 = 0{,}8$ mientras que añadiendo *exper* cambia a 0,97. Hay dos excepciones a este resultado. Se consideren dos modelos, el primero con una sola variable dependiente, $y = \beta_0 + \beta_1 x_{i1} + u_i$, y el segundo con dos variables independientes $y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$. Es evidente que si $\hat{\beta}_2 = 0$ entonces el estimador de β_1 coincidirá en los dos modelos. Un segundo caso en que $\hat{\beta}_1$ coincide en los dos modelos es cuando x_1 y x_2 no están correlacionadas. El estimador MCO de β_1 en el segundo modelo es $\hat{\beta}_i = \frac{\sum_{i=1}^n y_i \hat{r}_{i1}}{\sum_{i=1}^n \hat{r}_{i1}^2}$, donde $\hat{r}_{i1}$ es el residuo de la regresión de x_1 sobre x_2. Si los dos regresores no están correlacionados esto significa que la covarianza es cero y que el efecto de x_2 sobre x_1 es cero. Pero esto implica que $\hat{r}_{i1} = x_{i1} - \hat{\beta}_0 = x_{i1} - \bar{x}_1$ y los estimadores de $\hat{\beta}_1$ coinciden en los dos modelos.

Para la varianza del error, el estimador insesgado que utilizaremos en este capitulo es parecido al anterior. La única diferencia es que ahora hay que normalizar por el número de observaciones menos el número total de parámetros. Así que en el modelo de k-variables el estimador de la varianza del error es

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k-1} \tag{2.4}$$

Ejemplo 2.2 Se consideren los datos[1] siguientes

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1} &= \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \\
\mathbf{X}'\mathbf{y} &= \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix} \\
& \mathbf{y\prime y} = 80, \quad n = 90
\end{aligned} \tag{2.5}
$$

1. Calcular $\hat{\beta}$. Aplicando la formula vista antes

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 11 \\ -7 \\ 12 \end{pmatrix}
\end{aligned} \tag{2.6}
$$

2. Calcular $\hat{\sigma}^2$. Aplicando la formula vista antes

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\hat{\mathbf{u}}\prime\hat{\mathbf{u}}}{n-k-1} \\
&= \frac{\mathbf{y\prime y} - \hat{\beta}'\mathbf{X}'\mathbf{y}}{n-k-1}
\end{aligned}
$$

[1] Los datos son tomado desde Johnston and DiNardo (2001).

Sabemos que $\mathbf{y}\prime\mathbf{y} = 80$. Sólo nos falta calcular el segundo elemento

$$\hat{\beta}'\mathbf{X}'\mathbf{y} = \begin{pmatrix} 11 & -7 & 12 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix} = 37$$

entonces $\hat{\sigma}^2 = \frac{80-37}{86} = 0{,}5$.

2.2.2. Propiedades de los estimadores

Describiremos ahora las propiedades de los estimadores de MCO. Primero estudiaremos las propiedades algébricas y después las propiedades estadísticas de tales estimadores.

Propiedades algébricas

Además de **P1** y **P3** vale la siguiente propiedad

P1' $\mathbf{X}'\hat{\mathbf{u}} = 0$

$$\begin{aligned} \mathbf{X}'\hat{\mathbf{u}} &= \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= 0 \end{aligned}$$

Propiedades estadísticas

Antes de ver las propiedades estadísticas de los estimadores, primero necesitamos repasar media y varianza para un vector de variables aleatorias. Sea $\mathbf{x} = [x_1, ..., x_k]'$ un vector de n variables aleatorias. Definimos su valor esperado

$$E(\mathbf{x}) = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_k) \end{pmatrix}$$

y su varianza

$$Var(\mathbf{x}) = E\left[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))'\right]$$

La varianza así definida es una matriz de dimensión $k \times k$ donde en la diagonal principal hay los términos $E\left[(x_i - E(x_i))^2\right]$, o sea las varianzas de las variables contenidas en el vector $\mathbf{x}$, mientras que fuera de la diagonal principal hay los términos $E\left[(x_i - E(x_i))(x_j - E(x_j))\right]$, o sea las covarianzas entre las variables en $\mathbf{x}$. Si $\mathbf{x}$ es un vector de variables aleatoria normales, o sea si $\mathbf{x}$ tiene distribución normal multivariante con valor esperado μ y varianza Σ escribimos

$$\mathbf{x} \sim N(\mu, \Sigma)$$

Vamos a estudiar ahora las propiedades estadísticas de los estimadores MCO. Podemos ahora volver a escribir en forma matricial los supuestos **S2'** y **S3'**. El primero implica que $E(\mathbf{u}) = 0$, el segundo que $Var(\mathbf{u}) = \sigma^2 I$ donde I es una matriz identidad de dinemsión $n \times n$.

P2' *Los estimadores MCO son insesgados,* $E(\hat{\beta}) = \beta$.

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}$$

Tomando el valor esperado obtenemos

$$\begin{aligned}
E(\hat{\beta}) &= E[\beta + E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})] \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}) \\
&= \beta
\end{aligned}$$

donde la última igualdad se obtiene utilizando la propiedad P1'.

P3' $Var(\hat{\beta}) = \sigma^2(\mathbf{X'X})^{-1}$

Utilizando la definición de varianza de un vector de variables aleatorias

$$\begin{aligned} Var(\hat{\beta}) &= E\left[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'\right] \\ &= E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] \end{aligned}$$

Desde la propiedad anterior sabemos que

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

Entonces

$$\begin{aligned} Var(\hat{\beta}) &= E\left[(\mathbf{X'X})^{-1}\mathbf{X'uu'X}(\mathbf{X'X})^{-1}\right] \\ &= (\mathbf{X'X})^{-1}\mathbf{X'}E(\mathbf{uu'})\mathbf{X}(\mathbf{X'X})^{-1} \\ &= \sigma^2(\mathbf{X'X})^{-1} \end{aligned}$$

P4' *El estimador de la varianza del término de error visto antes es insesgado* $E(\hat{\sigma}^2) = \sigma^2$.

P5' *El vector de residuos* $\hat{\mathbf{u}}$ *es una transformación lineal del vector término de error* $\mathbf{u}$.

De su definición tenemos

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\hat{\beta} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} \\ &= (I - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'})\mathbf{y} \\ &= (I - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'})(\mathbf{X}\beta + \mathbf{u}) \\ &= (I - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'})\mathbf{u} \\ &= \mathbf{Mu} \end{aligned}$$

donde $\mathbf{M} = (I - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'})$ es una matriz simétrica y idempotente, una matriz tal que $\mathbf{MM} = \mathbf{M}$.

P6' *El vector de residuos* $\hat{\mathbf{u}}$ *tiene esperanza cero,* $E(\hat{\mathbf{u}}) = 0$ *y varianza* $E(\hat{\mathbf{u}}\hat{\mathbf{u}}') = \sigma^2\mathbf{M}$.

Desde la propiedad de antes $E(\hat{\mathbf{u}}) = \mathbf{M}E(\mathbf{u}) = 0$. Además

$$\begin{aligned} E(\hat{\mathbf{u}}\hat{\mathbf{u}}') &= E(\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}) \\ &= \mathbf{M}E(\mathbf{u}\mathbf{u}')\mathbf{M} \\ &= \sigma^2\mathbf{M} \end{aligned}$$

donde la última igualdad se obtiene gracias al hecho de que $\mathbf{M}$ es idempotente y simétrica.

P7' *Teorema de Gauss-Markov* $\hat{\beta}$ es el estimador lineal insesgado óptimo (ELIO) de β.

Este teorema, como hemos visto antes, nos dice que no hay un estimador lineal y insesgado de β con menor varianza. Vamos a ver ahora la demostración de este teorema. Sea $\beta^* = \mathbf{A}'y$ otro estimador lineal de β donde $\mathbf{A}$ es una matriz $n(k+1)$ de elementos non aleatorios. Substituyendo la difinición de y obtenemos $\beta^* = \mathbf{A}'(\mathbf{X}\beta + \mathbf{u}) = \mathbf{A}'\mathbf{X}\beta + \mathbf{A}'\mathbf{u})$. Tomando el valor esperado $E(\beta^*) = E(\mathbf{A}'\mathbf{X}\beta + \mathbf{A}'\mathbf{u}) = E(\mathbf{A}'\mathbf{X}\beta) + E(\mathbf{A}'\mathbf{u})$. Para que el estimador sea insesgado las iguientes condiciones deben cumplirse: a) $E(\mathbf{A}'\mathbf{X}\beta) = \beta$ $(\mathbf{A}'X = I)$ y b) $E(\mathbf{A}'\mathbf{u}) = 0$. La varianza del estimador será $Var(\beta*) = \mathbf{A}'[Var(\mathbf{u})]\mathbf{A} = \sigma^2\mathbf{A}'\mathbf{A}$. Entonces

$$\begin{aligned} Var(\beta*) - Var(\hat{\beta}) &= \sigma^2[\mathbf{A}'\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}] \\ &= \sigma^2[\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}] \\ &= \sigma^2\mathbf{A}'[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{A} \\ &= \sigma^2\mathbf{A}'\mathbf{M}\mathbf{A} \end{aligned}$$

donde $\mathbf{A}'\mathbf{M}\mathbf{A}$ es una matriz semidefinida positiva. Sea $\mathbf{c}$ cualquier vector de dimensión $(k+1)$. La varianza de la combinación lineal $\mathbf{c}'\hat{\beta}$ y $\mathbf{c}'\beta^*$ es respectivamente $Var(\mathbf{c}'\hat{\beta}) = \mathbf{c}'Var(\hat{\beta})\mathbf{c}$ y $Var(\mathbf{c}'\beta^*) = \mathbf{c}'Var(\beta^*)\mathbf{c}$. Además

$$\mathbf{c}'\left[Var(\hat{\beta}) - Var(\beta^*)\right]\mathbf{c} \geq 0$$

por definición de matríz semidefinida positiva y en particular $Var(\hat{\beta}_j) \leq Var(\beta_j^*)$ por cualquier j.

2.2.3. Bondad del ajuste

Para poder estudiar la bondad del ajuste en el caso de k-variables necesitamos algunos resultados algébricos relativos a nuestro modelo. Primero, podemos observar que

$$\begin{aligned} STC &\equiv \sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2 \\ SEC &\equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2 \\ SCR &\equiv \sum_{i=1}^{n}\hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}} \end{aligned}$$

Segundo, como en el modelo simple $STC = SEC + SCR$. Para poder demostrar esto empezamos por la definición de SCR.

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} \end{aligned} \tag{2.7}$$

donde para obtener la última igualdad hemos substituido la formula MCO de $\hat{\beta}$. Entonces

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= \hat{\mathbf{u}}'\hat{\mathbf{u}} + \hat{\beta}'\mathbf{X}'\mathbf{y} \\ \mathbf{y}'\mathbf{y} - n\bar{y}^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} + \hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 \end{aligned}$$

Lo único que hay que observar ahora es que $\hat{\mathbf{y}}'\hat{\mathbf{y}} = \hat{\beta}'\mathbf{X}'\mathbf{y}$. Esto es fácil de demostrar:

$$\begin{aligned} \hat{\beta}'\mathbf{X}'\mathbf{y} &= \hat{\beta}'\mathbf{X}'(\mathbf{X}\hat{\beta} - \hat{\mathbf{u}}) \\ &= \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - \hat{\beta}'\mathbf{X}'\hat{\mathbf{u}} \\ &= \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} \end{aligned}$$

donde la tercera igualdad se obtiene con la propiedad **P1'** y la última por definición de $\hat{\mathbf{y}}$. Hemos entonces establecido que la suma total de cuadrados es igual a la suma de la suma explicada y residual como en el modelo simple. Vamos ahora a ver la definición de R^2. De manera equivalente al modelo anterior definimos el coeficiente de determinación múltiple como

$$\begin{aligned} R^2 &= \frac{SEC}{STC} \\ &= \frac{\hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} \\ &= \frac{\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} \end{aligned}$$

¿Cómo interpretamos el R^2 en este caso? Como anteriormente R^2 se interpreta como *la fracción de la variación muestral de y_i explicada por todas las x_i conjuntamente.* Existe una relación muy importante entre la varianza de estimador MCO y R^2. De hecho podemos escribir la varianza de cada uno del los $\hat{\beta}_j$ como

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ij} - \bar{x})^2(1 - R_j^2)} \tag{2.8}$$

donde R_j^2 es el R-cuadrado que se obtiene de la regresión de x_j sobre todas las demás variables exógenas.

Ejercicio 2.2 (cont.) Utilizando los datos del ejemplo 2.2 calcular R^2. Sabemos que $R^2 = \frac{\hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}$. El primer elemento de $\mathbf{X}'\mathbf{y}$ es $\sum_{i=1}^{n} y_i = 3$. Entonces el numerador es igual a $37 - 90(3/90)^2 = 36{,}9$. El denominador es $80 - 90(3/90)^2 = 79{,}9$ y $R^2 = 36{,}9/79{,}9 = 0{,}46$.

Se puede demostrar que el R^2 aumenta si añadimos variables exógenas al modelo. La idea es que como la suma total de cuadrados es constante, si añadimos variables independientes con una varianza muestral distinta de cero, la suma de los residuos al cuadrado tiene que disminuir y por esta razón el coeficiente de determinación tiene que aumentar. Esto implica que no podemos utilizar el R^2 como una medida de bondad del ajuste para comparar modelos con un número diferente de regresores. O sea no podemos utilizar el coeficiente de determinación para eligir cuántos y cuáles regresores utilizar. Por esta razón, en el modelo de k-variables definimos el R^2-corregido, que es otra medida de bondad del ajuste del modelo. La idea es que corregimos el R^2 de manera que si añadimos un regresor queremos que éste aumente sólo si la nueva variable exógena es útil para explicar la variable dependiente. Gracias a esta corrección $\bar{R}^2$ nos permite comparar la bondad del ajuste en modelos con un número diferente de regresores. Definimos el coeficiente de determinación corregido como

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

La corrección está en el denominador del segundo elemento a la derecha del igual. Cuando añadimos un regresor k aumenta. Esto hace aumentar $\frac{n-1}{n-k}(1-R^2)$ y disminuir $\bar{R}^2$ contrastando el efecto opuesto debido al aumento de R^2.

A menudo la comparación entre $\bar{R}^2$ de diferentes modelos se toma como criterio de elección entre modelos. Para poder hacer esto, dos condiciones deben cumplirse: el tamaño de la muestra debe ser el mismo y la variables dependiente debe ser la misma. Hay que hacer una advertencia aquí. El hecho de encontrar un $\bar{R}^2$ bajo no significa necesariamente que no podamos confiar en los resultados obtenidos. Ésta es una situación que ocurre a menudo en los trabajos empíricos. Más importante que alcanzar una alta proporción de variabilidad explicada es poder obtener resultados confiables para los contrastes de hipótesis para los parámetros del modelo de interés económico.

2.3. Inferencia, el contraste F

Antes de empezar a estudiar el tema de la inferencia en el modelo de k-variables necesitamos repasar algunos resultados estadísticos preliminares para formas cuadráticas y añadir un supuesto, como hemos en el capitulo anterior, al modelo. El supuesto es que el término de error se distribuye normalmente

S6' *El término de error se distribuye normalmente* $\mathbf{u} \sim N(0, \sigma^2 I)$

Del supuesto anterior deriva otra propiedad del estimador MCO, o sea

P8' $\hat{\beta}$ *tiene distribución normal multivariante,* $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$

Los resultados mencionados antes son los siguientes. Sea $\mathbf{x}$ un vector $k \times 1$ de variables aleatorias.

(i) Si $\mathbf{x} \sim N(0, \Sigma)$ entonces

$$\mathbf{x}'\Sigma^{-1}\mathbf{x} \sim \chi^2_k$$

(ii) Si $\mathbf{x} \sim N(0, I)$ y A es una matriz simétrica y idempotente con rango r

$$\mathbf{x}'A\mathbf{x} \sim \chi^2_r$$

En particular si $\mathbf{x} \sim N(0, \sigma^2 I)$

$$\frac{\mathbf{x}'A\mathbf{x}}{\sigma^2} \sim \chi^2_r$$

Aplicando (i) y (ii) a nuestros estimadores obtenemos dos resultados muy importantes y clave para entender el contraste de hipótesis:

1) Desde la proprieded **P4'** sabemos que bajo las hipótesis del modelo $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. Entonces

$$\hat{\beta} - \beta \sim N(0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

y por una matrix R cualquiera de rango máximo y dimensión $q \times (k+1)$ (con $q \leq k+1$)

$$R(\hat{\beta} - \beta) \sim N(0, R\sigma^2(\mathbf{X}'\mathbf{X})^{-1}R').$$

Por el resultado (i) visto antes

$$(R(\hat{\beta} - \beta))'(R\sigma^2(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R(\hat{\beta} - \beta)) \sim \chi^2_q \qquad (2.9)$$

2) Desde la proprieded **P7'** y desde el supuesto **S5'** sabemos que $\hat{\mathbf{u}} = M\mathbf{u}$ y $\mathbf{u} \sim N(0, \sigma^2 I)$. Entonces desde el resultado (ii)

$$\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma^2} = \frac{\mathbf{u}'M\mathbf{u}}{\sigma^2} \sim \chi^2_{n-k-1}$$

3) Si ahora substituimos en (2.9) el estimador de la varianza del término de error visto antes obtenemos

$$\frac{(R(\hat{\beta}-\beta))'(R(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R(\hat{\beta}-\beta))}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k-1}}$$

Si dividimos el numerador por q y multiplicamos y dividimos por σ^2 obtenemos

$$\frac{(R(\hat{\beta}-\beta))'(R(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R(\hat{\beta}-\beta))/q}{\frac{\sigma^2}{\sigma^2}\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k-1}} \tag{2.10}$$

Podemos notar que esta expresión es el ratio entre dos variables χ^2 cada una dividida por sus grados de libertad. La primera es

$$\frac{(R(\hat{\beta}-\beta))'(R(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R(\hat{\beta}-\beta))/q}{\sigma^2}$$

y la segunda es

$$\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma^2(n-k-1)} = \frac{\hat{\sigma}^2}{\sigma^2},$$

como hemos visto en el capitulo anterior, el ratio entre dos χ^2 divididas por sus grados de libertad es una variable con distribución F. Entonces (2.10) es una variable $F_{q,n-k-1}$ donde q son los grados de libertad del numerador y $n-k-1$ los del denominador.

2.3.1. Contraste de hipótesis para restricciones lineales múltiples: el contraste de Wald

En este apartado explicaremos como hacer un contraste de hipótesis para restricciones lineales múltiples. Como en el modelo simple los ingredientes básicos son una hipótesis nula y una alternativa, un estadístico de contraste y un criterio de rechazo de la hipótesis nula. Empezamos explicando cómo podemos expresar en términos formales una hipótesis nula que implique un conjunto de restricciones lineales sobre los parámetros del modelo. Sea R una matriz de dimensión $q \times (k+1)$ y sea r un vector de dimensión $q \times 1$. Podemos expresar cualquier combinación lineal de los parámetros de la siguiente manera:

$$R\beta = r$$

Supongamos por ejemplo que $k = 3$ y se considere el siguiente sistema de ecuaciones lineales (restricciones lineales de los parámetros)

$$\begin{aligned} \beta_1 + \beta_2 &= 1 \\ \beta_3 &= -1 \end{aligned}$$

Este se puede expresar como

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

o $R\beta = r$ donde

$$\begin{aligned} R &= \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ r &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned}$$

Entonces especificaremos la hipótesis nula (H_0) de la manera siguiente

$$H_0: \quad R\beta = r$$

y la alternativa como

$$H_0: \quad R\beta \neq r$$

Como estadístico de contraste utilizaremos el estadístico (). Bajo la hipótesis nula el estadístico es

$$\frac{(R\hat{\beta} - r)'(R(\mathbf{X}'\mathbf{X})^{-1}R')^{-1}(R\hat{\beta} - r)/q}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k-1}} \sim F_{q,n-k-1} \tag{2.11}$$

o sea una variable con distribución F con q y $n-k-1$ grados de libertad. (2.11) se conoce cono contraste de Wald para restricciones lineales de los parámetros. El criterio de rechazo de la hipótesis nula es como el del capítulo anterior. Utilizando nuestra muestra de datos y las restricciones implicadas por la hipótesis nula calculamos el valor para el estadístico de contraste $\hat{F}$. Si $\hat{F} > F^*_{q,n-k-1,\alpha}$, donde $F^*_{q,n-k-1,\alpha}$ es el valor crítico para la distribución F con q y $n-k-1$ grados de libertad, rechazamos la hipótesis nula al nivel de confianza α. Si por lo contrario $\hat{F} < F^*_{q,n-k-1,\alpha}$ no rechazamos la hipótesis. Otra vez la idea que está detrás del criterio de rechazo es la de establecer si bajo la hipótesis nula es "probable" o no observar lo que de hecho observamos en nuestra muestra de datos. Es muy importante entender el hecho de que cualquier decisión sobre la hipótesis nula es una decisión sobre todas las restricciones lineales contenidas en la hipótesis. O sea, si rechazamos la hipótesis nula rechazamos todas las restricciones conjuntamente.

Ejemplo 2.2 (cont) Utilizando los datos del ejercicio ()

1. Contrastar la siguiente hipótesis nula

$$H_0 : 2\beta_1 + \beta_2 = 3 \tag{2.12}$$

Primero necesitamos encontrar el estadístico de contraste 2.11. Empezamos definiendo R y r. $R = \begin{pmatrix} 0 & 2 & 1 & 0 \end{pmatrix}$, $= r = 3$. Entonces

$$R\hat{\beta} - r = \begin{pmatrix} 0 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 11 \\ -7 \\ 12 \\ 3 \end{pmatrix} - 3 = -5$$

Podemos encontrar ahora $R(\mathbf{X}'\mathbf{X})^{-1}R'$

$$R(\mathbf{X}'\mathbf{X})^{-1}R' \quad = \quad \begin{pmatrix} 0 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -4 & 10 & 0 & -5 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix} = 20$$

Juntando todas las piezas obtenemos

$$\hat{F} = \frac{(-5)\frac{1}{20}(-5)}{0{,}5} = 2{,}5$$

El valor crítico 5 % es $F^*_{1,86} = 3{,}95$, por lo tanto no rechazamos la hipótesis nula.

2. Contrastar la siguiente hipótesis nula

$$H_0 : \begin{cases} 2\beta_1 + \beta_2 = 3 \\ \beta_3 = 6 \end{cases} \tag{2.13}$$

En este caso

$$\begin{aligned} R &= \begin{pmatrix} 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ r &= \begin{pmatrix} 3 \\ 6 \end{pmatrix} \end{aligned}$$

entonces

$$\begin{aligned} R\hat{\beta} - r &= \begin{pmatrix} 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 11 \\ -7 \\ 12 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \begin{pmatrix} -5 \\ 3 \end{pmatrix} \end{aligned}$$

Además

$$\begin{aligned} R(\mathbf{X}'\mathbf{X})^{-1}R' &= \begin{pmatrix} 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 20 & -5 \\ -5 & 4 \end{pmatrix} \end{aligned}$$

Entonces el estadístico de contraste será

$$\begin{pmatrix} -5 & -3 \end{pmatrix} \begin{pmatrix} 20 & -5 \\ -5 & 4 \end{pmatrix}^{-1} \begin{pmatrix} -5 \\ -3 \end{pmatrix} = 7{,}8$$

que es mayor del valor crítico al 5 % o sea 3.1. Por lo tanto rechazamos la hipótesis nula.

2.3.2. Dos forma alternativa para contraste de hipótesis

Hay dos formas alternativas para el contraste de hipótesis. La primera se basa en la suma de los residuos al cuadrado en el modelo restringido. La segunda se basa en el R^2 del modelo restringido. Vamos primero a aclarar el concepto de modelo restringido y después a estudiar estas dos formas alternativas para el contraste de hipótesis.

Se considere el siguiente modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$$

y la siguiente hipótesis

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 1 \end{cases} . \tag{2.14}$$

Nos preguntamos ¿cuál es el modelo bajo esta hipótesis? Simplemente es el modelo donde las restricciones anteriores se cumplen. En este caso el modelo restringido es

$$y_i = \beta_0 + x_{i1} + \beta_3 x_{i3} + u_i$$

¿Cómo podemos estimar los parámetros de este modelo? Podemos escribir la ecuación anterior como

$$y_i - x_{i1} = \beta_0 + \beta_3 x_{i3} + u_i$$
$$y_i^* = \beta_0 + \beta_3 x_{i3} + u_i$$

Considerando $y_i^* = y_i - x_{i1}$ como nueva variable dependiente, podemos utilizar MCO para estimar los parámetros del modelo. En particular obtenemos

$$\hat{\beta}_3^* = \frac{\sum_{i=1}^n (y_i^* - \bar{y}^*)(x_{i3} - \bar{x}_3)}{\sum_{i=1}^n (x_{i3} - \bar{x}_3)^2}, \qquad \hat{\beta}_0^* = \bar{y}^* - \hat{\beta}_3^* \bar{x}_3,$$

Una vez obtenidas las estimaciones podemos calcular la suma de los residuos al cuadrado $SCR_r = \sum_{i=1}^{n} \hat{u}_{ri}^2$ con $\hat{u}_{ri} = \hat{y}_i^* - \hat{\beta}_3^* x_{i3}$ (el índice r se refiere al modelo restringido). Con estas definiciones podemos utilizar el siguiente estadístico de contraste

$$\frac{(SCR_r - SCR)/q}{SCR/(n-k-1)} \sim F_{q,n-k-1} \tag{2.15}$$

o, utilizando sus definiciones,

$$\frac{(\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/q}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} \sim F_{q,n-k-1} \tag{2.16}$$

Otra vez si el estadístico es mayor que el valor crítico rechazamos la H_0. En general podemos contrastar una hipótesis nula cualquiera utilizando el estadístico (2.15). La clave es poder escribir el modelo restringido y calcular la suma de los residuos al cuadrado. Es importante saber que el estadístico (2.15) es equivalente al estadístico (2.11), en el sentido que los números calculados en los dos casos coinciden.

Otra forma alternativa para el contraste de hipótesis se basa en los coeficientes de determinación del modelo restringido y no restringido. Este tipo de estadístico de contraste sólo se puede utilizar en el caso en que las variables dependientes coincidan en los dos modelos. Considerese (2.16). Si dividimos y multiplicamos por la suma total de cuadrados $\mathbf{y'y} - n\bar{y}^2$ obtenemos

$$\frac{(\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/q}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} = \frac{\left(\frac{\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r}{\mathbf{y'y}-n\bar{y}^2} - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y'y}-n\bar{y}^2}\right)/q}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y'y}-n\bar{y}^2}/(n-k-1)}$$

Desde la definición de coeficiente de determinación $\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\mathbf{y'y}-n\bar{y}^2} = 1-R^2$ y $\frac{\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r}{\mathbf{y'y}-n\bar{y}^2} = 1-R_r^2$ donde R_r^2 es el coeficiente de determinación en el modelo restringido. Entonces

$$\begin{aligned} \frac{(\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/q}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} &= \frac{(1-R_r^2-1+R^2)/q}{(1-R^2)/(n-k-1)} \\ &= \frac{(R^2-R_r^2)/q}{(1-R^2)/(n-k-1)} \sim F_{q,n-k-1} \end{aligned} \tag{2.17}$$

(2.17) es la segunda forma alternativa para contraste de hipótesis. La expresión (2.17) pone en evidencia porque para poder utilizar este contraste necesitamos que la variable dependiente sea la misma en los dos modelos. Para llegar a esta expresión necesitamos poder interpretar $\frac{\hat{\mathbf{u}}_r{}'\hat{\mathbf{u}}_r}{\mathbf{y}'\mathbf{y}-n\bar{y}^2}$ como uno menos el coeficiente de determinación en el modelo restringido. Esto es el caso sólo si $\mathbf{y}'\mathbf{y}-n\bar{y}^2$ es la suma de cuadrados totales o sea si y es la variable dependiente en el modelo restringido también. Esto significa que este estadistico no se puede utilizar para contrastar la hipótesis nula (2.14) ya que, bajo esta hipótesis, la variable dependiente es $y_i^* = y_i - x_{i1} \neq y_i$.

Ejemplo 2.1 (cont.). En el modelo utilizado anteriormente hemos obtenido las siguientes estimaciones:

$$\log(\hat{salario_i}) = 0{,}216 + 0{,}097 educ + 0{,}010 exper$$

com $SCR = 111{,}345$ y $R_r^2 = 0{,}24934$. Si excluimos *exper* del modelo y volvemos a estimar obtenemos

$$\log(\hat{salario_i}) = 0{,}583 + 0{,}082 educ$$

y la suma de residuos al cuadrado es $SCR_r = 120{,}769$ y $R^2 = 0{,}18581$.

1. Contrastar al 5 % la hipótesis nula $H_0 : \beta_2 = 0$ utilizando (2.16) y (2.17). Utilizando (2.16) obtenemos

$$\frac{(120{,}769 - 111{,}345)/1}{111{,}345/519} = 43{,}92$$

   y rechazamos la hipótesis nula porque este valor es mayor que el valor crítico al 5 % $F_{1,524}^* = 3{,}86$. Utilizando (2.17) obtenemos

$$\frac{(0{,}24934 - 0{,}18581)/1}{(1 - 0{,}24934)/519} = 43{,}92$$

2. Contrastar la hipótesis nula $H_0 : \beta_2 = \beta_3 = 0$. Contrastar esta hipótesis implica contrastar que todos los coeficientes del modelo no sean significativos. Bajo esta restricción el modelo es un modelo con solo constante y término de error. El R^2 de este modelo restringido es cero. Entonces podemos utilizar (2.17) con $R_r^2 = 0$. Asi obtenemos

$$\frac{R^2/q}{(1-R^2)/(n-k-1)} = \frac{0{,}24934/1}{(1-0{,}24934)/519} = 172{,}39$$

que es mayor que el valor crítico al 5 % y por lo tanto rechazamos la hipótesis nula.

2.3.3. La relación entre contraste t y F

Hemos visto que con el estadístico F podemos contrastar cualquier restricción lineal sobre los parámetros del modelo. En particular este contraste se puede utilizar para contrastar $H_0 : \beta_i = \beta_i^0$. En el capítulo anterior hemos visto que para contrastar esta hipótesis podíamos utilizar el contraste t. Entonces en el modelo de k-variables ¿cuál estadístico tenemos que utilizar? La respuesta es que podemos utilizar cualquiera de los dos porque existe una relación que nos dice que $t^2_{n-k-1} = F_{1,n-k-1}$. Para contrastar la hipótesis nula $H_0 : \beta_i = \beta_i^0$ podemos emplear el estadístico t con $n - k - 1$ grados de libertad o el estadístico $F_{1,n-k-1}$. El resultado (rechazar o no rechazar H_0) será exactamente igual en los dos contraste. Nótese que en el caso de k-variables si utilizamos el contraste t los grados de libertad no serán $n - 2$ como en el capítulo anterior si no $n - k - 1$ o sea el número de observaciones menos el número total de parámetros.

Ejercicios

(1) Volvemos a considerar el modelo del ejercicio **(2)** en el precedente capitulo

$$price = \beta_0 + \beta_1 assess + u$$

donde *price* es el precio de la vivienda y *assess* es la valoración de la vivienda antes de la venta, la valoración es racional si $\beta_1 = 1$ y $\beta_0 = 0$. La ecuación estimada es

$$price = -14{,}47 + 0,976assess$$
$$n = 88, SCR = 165644{,}51, R2 = 0,820 \tag{2.18}$$

donde 16,27 es la desviación típica de $\hat{\beta}_0$ y 0,049 es la desviación típica de $\hat{\beta}_1$.

(a) Para contrastar la hipótesis conjunta de $\beta_0 = 0$ y $\beta_1 = 1$, necesitamos la la suma de los cuadrados de los residuos del modelo restringido (SCR_r). El resultado es $SCR_r = 209448{,}99$. Llevar a cabo el contraste al 5 % de la hipótesis conjunta con un test F (al 5 % $F_{(2,86)} = 3{,}1$). ¿Cuál es el residuo en este modelo restringido?

Ahora extendemos el modelo de la siguiente manera

$$price = \beta_0 + \beta_1 assess + \beta_2 sqrft + \beta_3 bdrms + u$$

donde la variable $sqrft$ es el número de metros cuadrados de la casa y $bdrms$ el número de habitaciones. El R^2 de la estimación de este modelo usando la muestra de 88 casas es 0,829.

(c) Con MCO encontramos $\hat{\beta}_1 = 0{,}95$ y $\hat{\beta}_2 = -0{,}0048$ y $\hat{\beta}_2 = 11{,}83$. Interpretar esto valores.

(d) Contrastar al 5 % $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ (al 5 % $F_{(3,84)} = 2{,}71$) .

(e) Contrastar al 5 % $H_0 : \beta_2 = 0, \beta_3 = 0$ (al 5 % $F_{(2,84)} = 3{,}105$).

(f) ¿Podemos, sabiendo que bajo la $H_0 : \beta_1 = 1, \beta_2 = 0$ el R^2 es 0,7, contrastar esta H_0?.

(2) Considerar el siguiente modelo de regresión

$$Y = X\beta + u$$

donde X es una matriz 80×4, β es un vector 4×1, Y y u son vectores 80×1. Supongamos que

$$(X'X)^{-1} = \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix} \quad (X'Y)^{-1} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

y además $Y'Y = 80$.

(a) Contrastar con test F al 5 % la hipótesis nula

$$H_0 : \begin{cases} \beta_1 + \beta_2 = 1 \\ \beta_3 = 1 \end{cases}$$

sabiendo que el valor crítico al 5 % es $F_{(2,86)} = 3{,}1$.

(b) Contrastar con test F al 5 % la hipótesis nula

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \\ \beta_3 = 0 \end{cases}$$

sabiendo que el valor crítico al 5 % es $F_{(3,86)} = 2{,}71$.

(c) Del ejercicio 2.2 sabemos que $R^2 = 0{,}49$. Contrastar al 5 % la hipótesis nula en (b).

(3) Considerar el siguiente modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

donde se conocen las siguientes cantidades: $\sum_i X_i = 37{,}2$ $\sum_i X_i^2 = 147{,}18$ $\sum_i Y_i = 75{,}50$ $\sum_i Y_i^2 = 597{,}03$ $\sum_t Y_i X_i = 295{,}95$, $N = 10$.

a) Estimar β_0, β_1 utilizando el estimador de mínimos cuadrados.

b) Calcular y interpretar el R^2 de la regresión.

c) Contrastar al 5 % la hipótesis que la pendiente de la recta de regresión sea igual a cero.

Valore críticos:

$t_{0,025,8} = 2{,}28$, $t_{0,025,702} = 1{,}96$, $F_{2,702} = 3{,}01$, $F_{3,702} = 2{,}62$

(4) Considerar el siguiente modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

Supongamos de añadir otro regresor x_{i2} correlacionado con x_{i1}. ¿Cual será el efecto sobre la varianza de $\hat{\beta}_1$?.

(5) Demostrar la propiedad **P4'**.

Aplicaciones con GRETL

(1) Ejercicio 4.12 Wooldridge.

(2) Ejercicio 4.17 Wooldridge.

(3) Ejercicio 4.19 Wooldridge.

Capítulo 3

Extensiones

3.1. Errores de Especificación

El término *error de especificación* se refiere al caso en que cometemos un error en la decisión sobre cuales variables hay que incluir en el modelo. Veremos dos casos: el primero en que en que incluimos una variable irrelevante y el segundo en que omitimos una variable relevante.

3.1.1. Inclusión de una variable irrelevante

Supongamos que nuestro modelo de regresión sea

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i$$

pero añadimos la variable x_2 que no tiene ningún efecto sobre y y especificamos nuestro modelo como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

¿Cuáles son las consecuencias? No muchas. Simplemente obtendremos una estimación de β_2 que tendría que acercarse al cero, ya que su valor medio en muestras repetidas será igual a cero. Esto deriva del hecho que los estimadores siguen insesgados porque no hay ninguna violación de los supuestos del modelo. Sin embargo, hay alguna implicación sobre la precisión de las estimaciones. Desde el primer capítulo sabemos que la varianza de $\hat{\beta}_1$ es

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2(1 - R_1^2)} \tag{3.1}$$

donde R_1^2 es el coeficiente de determinación de de la regresión de x_1 sobre x_2. Si x_1 y x_2 están correlacionadas, $R_1^2 > 0$ y esto implica una varianza para $\hat{\beta}_1$ mayor que en el caso en que el modelo fuese bien especificado. Entonces, aunque los estimadores siguen insesgados, la varianza de los estimadores será mayor si las variables están correlacionadas, o sea tendremos menor precisión de la estimación.

3.1.2. Omisión de una variable relevante

Las consecuencias de excluir variables relevantes son mas graves porque crea sesgo en los estimadores MCO. Supongamos ahora que el modelo de regresión correctamente especificado sea

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

y que cometemos un error: excluimos del modelo x_2 y consideramos el modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i$$

Podemos verlo formalmente. Se considere el estimador MCO de β_1 en el modelo sin x_2

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(y_i - \bar{y}_i)}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)}$$

substituyendo $_i$ con el veradero modelo obtenemos

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(\beta_1 x_{i1} + \beta_2 x_{i2} + u_i)}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} \\
&= \beta_1 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i1}}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} \\
&= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} \qquad (3.2)
\end{aligned}$$

Si tomamos el valor esperado obtenemos

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_i 1 - \bar{x}_1)^2} \tag{3.3}$$

o sea el estimador es sesgado y el sesgo depende de la correlación muestral entre las dos variables independientes.

3.2. Estimación Bajo Restricciones

A veces la teoría económica nos proporciona restricciones sobre los parámetros del modelo. Por ejemplo, podemos pensar en los rendimientos de escala en una función de producción que queremos suponer constantes. En esta sección estudiaremos como estimar el modelo de regresión bajo restricciones lineales sobre los parámetros, o sea estudiaremos el método de Mínimos Cuadrados Restringidos.

Empezamos estableciendo un conjunto de q restricciones entre los parámetros del siguiente modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + u_i$$

donde $q < k$ y que estas restricciones se pueden expresar, como hemos visto ya en el contraste de hipótesis, como $R\beta = r$. El objetivo es de estimar el modelo anterior bajo estas restricciones. Sean $\hat{\mathbf{u}}_r$ los residuos del modelo. Utilizando el método de mínimos cuadrados queremos encontrar el estimador que minimiza $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r = \mathbf{y}'\mathbf{y} + \beta^{*\prime}\mathbf{X}'\mathbf{X}\beta^* - 2\beta^{*\prime}\mathbf{X}'\mathbf{y}$ y que al mismo tiempo satisface $R\beta^* = r$ donde β^* es el estimador de Mínimos Cuadrados Restringido de β. Formalmente escogemos el estimador que soluciona el siguiente problema de minimización

$$\begin{aligned} \min_{\beta^*} \quad & \mathbf{y}'\mathbf{y} + \beta^{*\prime}\mathbf{X}'\mathbf{X}\beta^* - 2\beta^{*\prime}\mathbf{X}'\mathbf{y} \\ \text{s.a.} \quad & R\beta^* = r \end{aligned}$$

Solucionando este problema obtenemos

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

donde $\hat{\beta}$ es el estimador MCO. Su varianza es

$$Var(\beta^*) = \sigma^2 \left[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1} R' (R(\mathbf{X}'\mathbf{X})^{-1} R')^{-1} R(\mathbf{X}'\mathbf{X})^{-1} \right]$$

y se puede demostrar que esta varianza es menor que la varianza del $\hat{\beta}$. Vamos a ver ahora formalmente que el test de Wald visto antes coincide con la forma alternativa basada en la suma de los cuadrados del modelo restringido.

$$\begin{aligned}
\hat{\mathbf{u}}_r &= \mathbf{y} - \mathbf{X}\beta^* \\
&= \mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{X}(\beta^* - \hat{\beta}) \\
&= \hat{\mathbf{u}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} R' (R(\mathbf{X}'\mathbf{X})^{-1} R')^{-1} (R\hat{\beta} - r)
\end{aligned}$$

entonces

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

y

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

que es el numerador del estadístico de contraste (2.9).

Ejercicios

(1) Supongamos que el verdadero modelo de regresión sea

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

y cometemos un error de especificación y estimamos el modelo sin x_2

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i$$

Demostrar que si la correlación entre x_1 y x_2 es positiva entonces el sesgo $E(\hat{\beta}_1) - \beta_1$ es positivo si $\beta_2 > 0$ y negativo si $\beta_2 < 0$.

Apéndices

A.1 GRETL

GRETL (http://gretl.sourceforge.net/win32/) es un programa de econometría que se puede bajar gratis desde internet. Está instalado en las aulas de informática 21-22-23. Podéis bajar el programa e instalarlo en vuestro ordenador privado. Funciona con Windows, Mac, y Linux. Está disponible en varios idiomas incluyendo el español. GRETL ya tiene instalada varias bases de datos. En la misma página podéis bajar e instalar en GRETL otras bases de datos como Wooldridge, de la cual hemos utilizado algunos archivos en esta guía, o Stock y Watson. El programa es relativamente fácil de utilizar. En el programa hay una guía de usuario muy útil que introduce el uso del programa.

En esta mini-guía veremos algunos ejemplos sencillos de aplicaciones de los conceptos teóricos que hemos visto.

Datos

La figura 4 muestra la pantalla principal de GRETL una vez abierto el programa. Para poder cargar una base de datos en GRETL tenemos que, como se ve en figura 5, seleccionar en el menú `File, Open data, Sample file`. La figura 6 muestra la pantalla con todas las bases de datos que hay disponibles en GRTL. Como ejemplo, seleccionamos `Wooldride, wage1` que es el archivo que hemos utilizado en el capí-

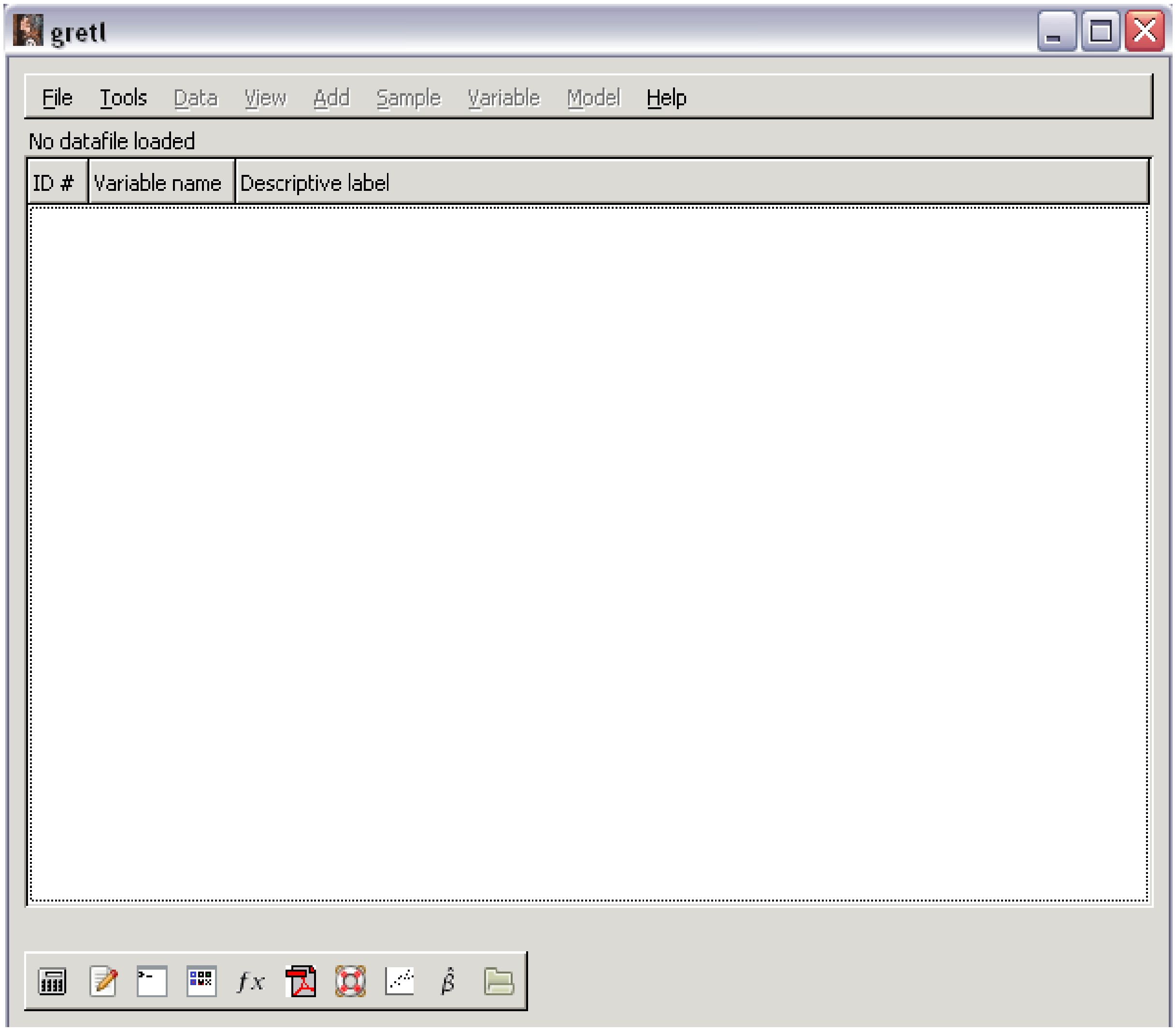

Figura 4

tulo 1. La figura 7 muestra una nueva pantalla con todas las variables contenidas en el fichero y sus descripciones. Podemos observar como ahora (figura 8) se activan opcciones en el menú que antes eran disponibles. Por ejemplo, seleccionando `Datos` tenemos varias opcciones que, entre otras cosas, nos permiten crear nuevas variables o hacer transformaciones de las variables existentes en el fichero. En particular, la primera opción es la de crear una variable definida como el logaritmo de la variable en negrita que hemos seleccionada en la pantalla principal.

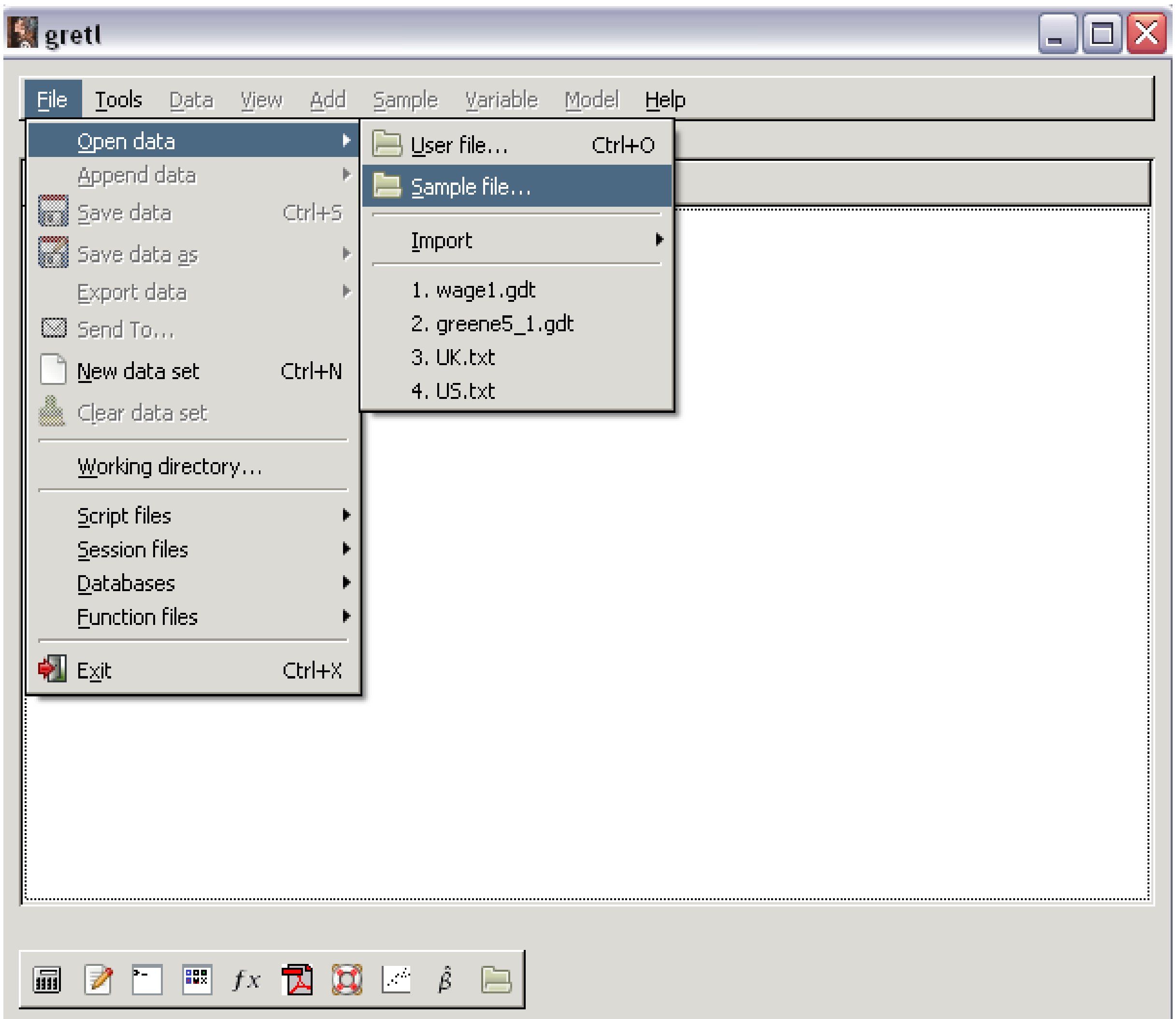
gretl
File
Tools
Data
View
Add
Sample
Variable
Model
Help
Open data
Append data
Save data Ctrl+S
Save data as
Export data
Send To...
New data set Ctrl+N
Clear data set
Working directory...
Script files
Session files
Databases
Function files
Exit Ctrl+X
User file... Ctrl+O
Sample file...
Import
1. wage1.gdt
2. greene5_1.gdt
3. UK.txt
4. US.txt

Figura 5

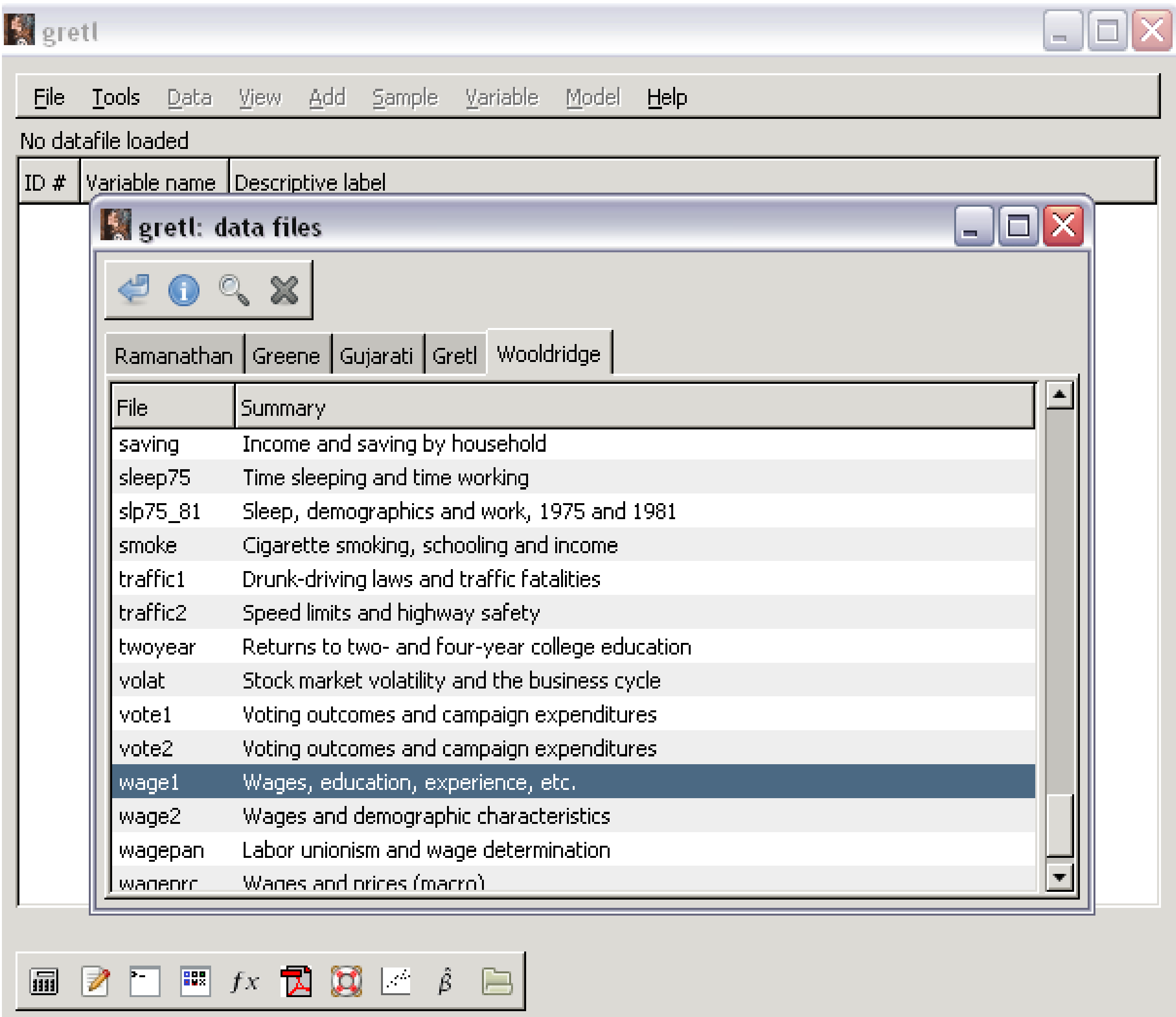
gretl
File Tools Data View Add Sample Variable Model Help
No datafile loaded
ID # Variable name Descriptive label
gretl: data files
Ramanathan Greene Gujarati Gretl Wooldridge
File Summary
saving Income and saving by household
sleep75 Time sleeping and time working
slp75_81 Sleep, demographics and work, 1975 and 1981
smoke Cigarette smoking, schooling and income
traffic1 Drunk-driving laws and traffic fatalities
traffic2 Speed limits and highway safety
twoyear Returns to two- and four-year college education
volat Stock market volatility and the business cycle
vote1 Voting outcomes and campaign expenditures
vote2 Voting outcomes and campaign expenditures
wage1 Wages, education, experience, etc.
wage2 Wages and demographic characteristics
wagepan Labor unionism and wage determination

Figura 6

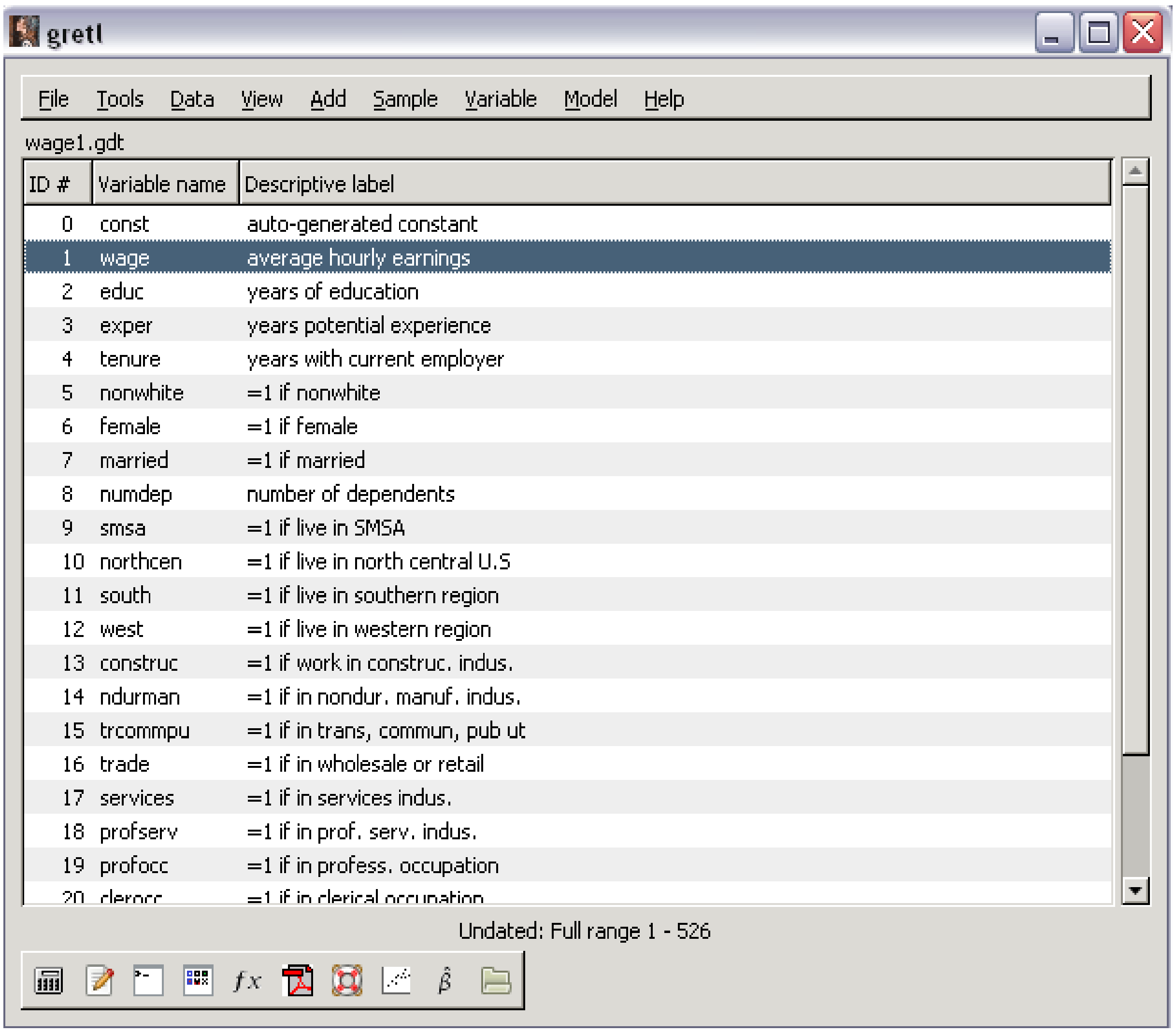
gretl
File Tools Data View Add Sample Variable Model Help
wage1.gdt
ID # Variable name Descriptive label
0 const auto-generated constant
1 wage average hourly earnings
2 educ years of education
3 exper years potential experience
4 tenure years with current employer
5 nonwhite =1 if nonwhite
6 female =1 if female
7 married =1 if married
8 numdep number of dependents
9 smsa =1 if live in SMSA
10 northcen =1 if live in north central U.S
11 south =1 if live in southern region
12 west =1 if live in western region
13 construc =1 if work in construc. indus.
14 ndurman =1 if in nondur. manuf. indus.
15 trcommpu =1 if in trans, commun, pub ut
16 trade =1 if in wholesale or retail
17 services =1 if in services indus.
18 profserv =1 if in prof. serv. indus.
19 profocc =1 if in profess. occupation
Undated: Full range 1 - 526

Figura 7

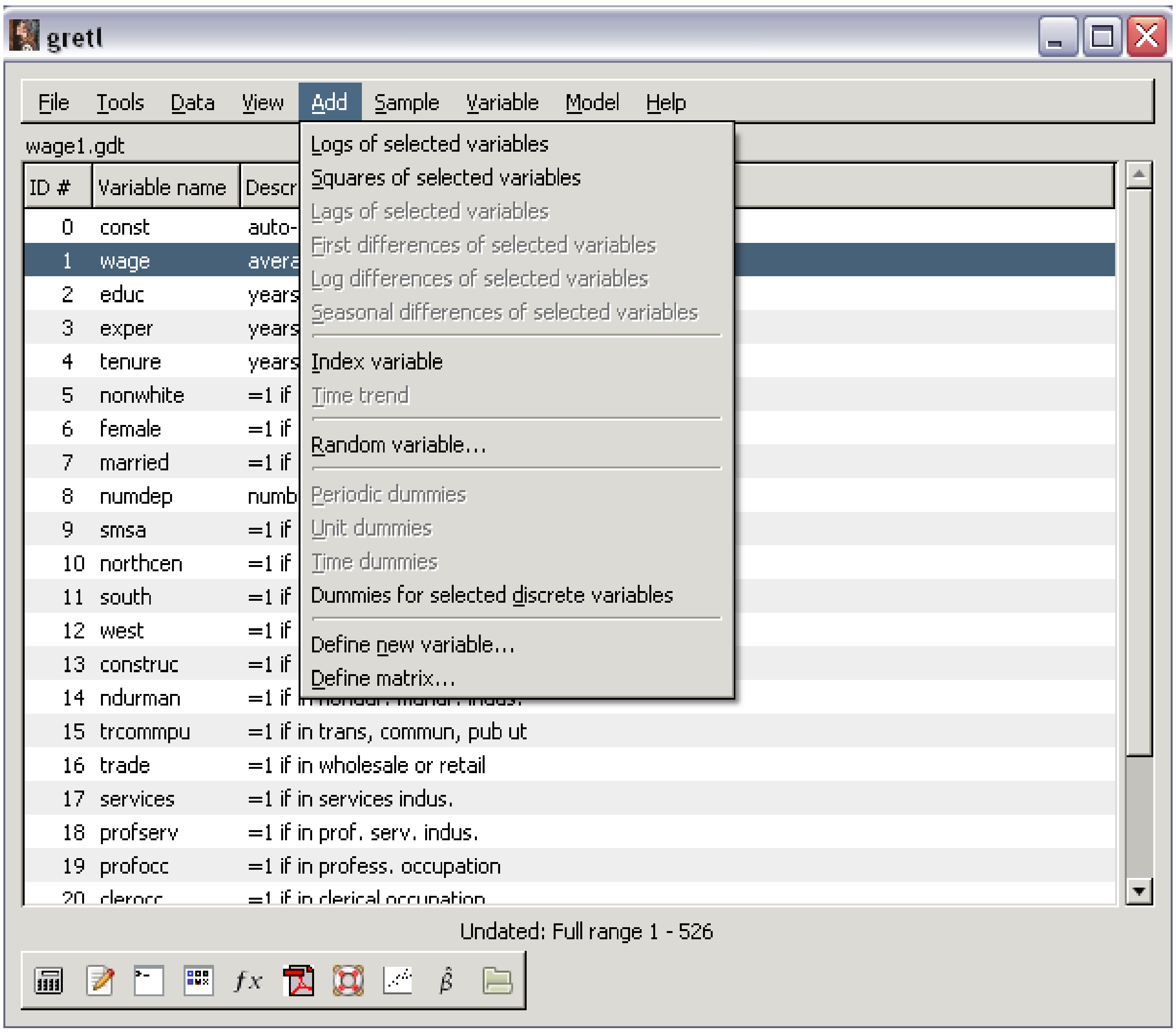
gretl
File
Tools
Data
View
Add
Sample
Variable
Model
Help
wage1.gdt
ID #
Variable name
Descr
0 const auto-
1 wage avera
2 educ years
3 exper years
4 tenure years
5 nonwhite =1 if
6 female =1 if
7 married =1 if
8 numdep numb
9 smsa =1 if
10 northcen =1 if
11 south =1 if
12 west =1 if
13 construc =1 if
14 ndurman =1 if
15 trcommpu =1 if in trans, commun, pub ut
16 trade =1 if in wholesale or retail
17 services =1 if in services indus.
18 profserv =1 if in prof. serv. indus.
19 profocc =1 if in profess. occupation
Logs of selected variables
Squares of selected variables
Lags of selected variables
First differences of selected variables
Log differences of selected variables
Seasonal differences of selected variables
Index variable
Time trend
Random variable...
Periodic dummies
Unit dummies
Time dummies
Dummies for selected discrete variables
Define new variable...
Define matrix...
Undated: Full range 1 - 526

Figura 8

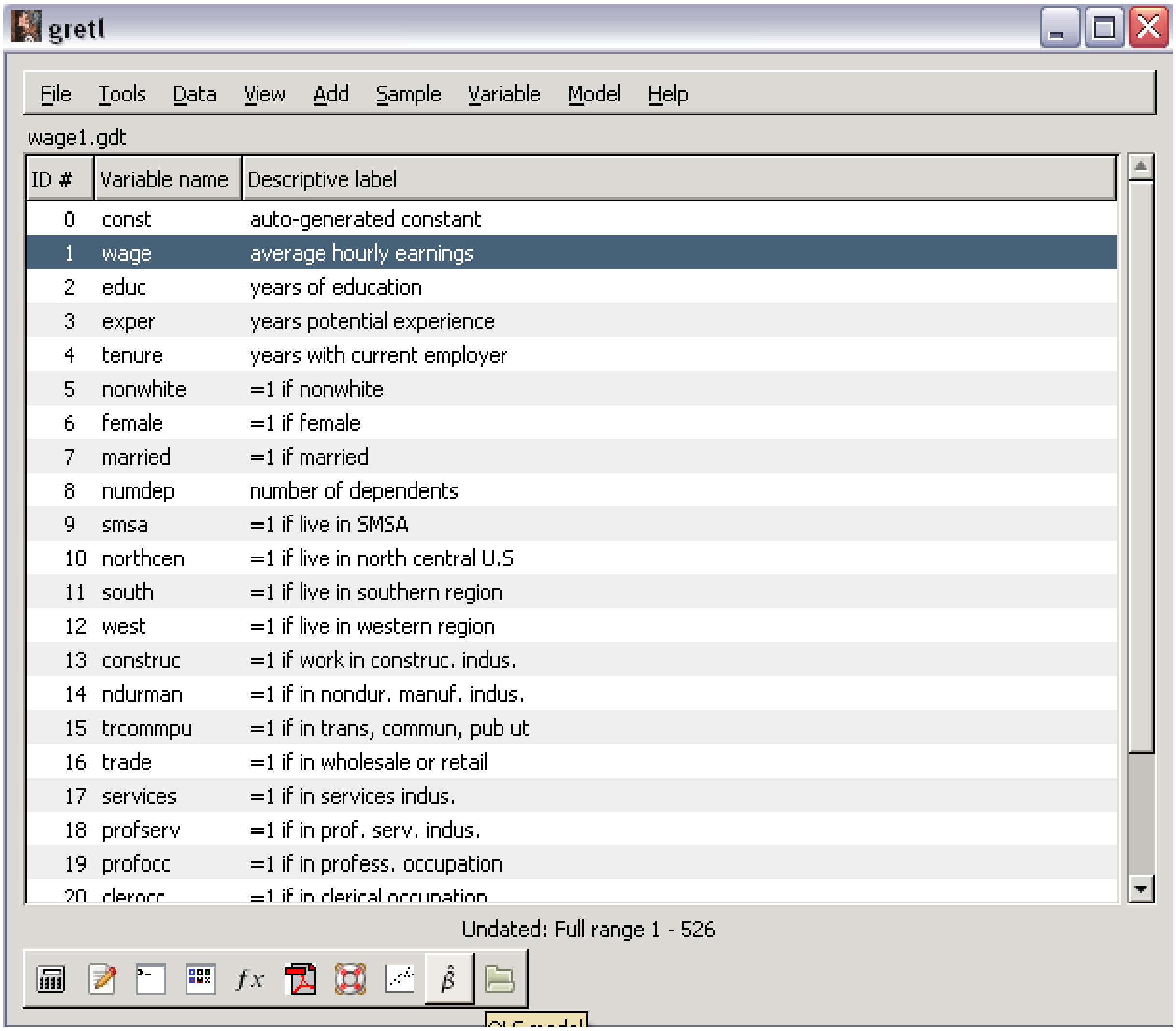

Figura 9

MCO

Vamos a ver ahora como estimar un modelo de regresión con MCO. Si volvemos a la pantalla principal podemos observar que en el menú de la parte baja de la pantalla tenemos un $\hat{\beta}$. Ésta es la instrucción que nos permite estimar un modelo de regresión con MCO. Si seleccionamos este comando se abre otra pantalla (figura 10). En esta nueva pantalla tenemos que seleccionar la variable dependiente (en la parte alta) y las variables independiente (en la parte baja). En este ejemplo seleccionamos `wage` como variable dependiente y `const` (el término constante) y `educ` como variables independientes. Una vez seleccionado `ok` GRETL estima con MCO el modelo y nos proporciona la tabla de resultados que vemos en figura 11. En esta tabla tenemos: 1) las estimaciones en la columna `coefficient`. 2) las desviaciones típicas de los parámetros en la columna `std. error` . 3) El valor de los estadisticos de contraste para la hipótesis nula $H_0 : \beta_i = 0$ en la columna `t-ratio` y 4) los valores p en la última columna. Además GRETL nos proporcina otros resultados como el R^2 o el $\bar{R}^2$.

En la figura 12 podemos ver los resultados de la estimación si repetimos el procedimiento visto antes pero añadiendo otro regresor `exper`. Utilizamos ahora estas nuevas estimaciones para ver como contrastar la hipótesis nula $H_0 : \beta_1 + \beta_2 = 0$. En el menú de la pantalla de los resultados hay una opción que es `tests`. Si seleccionamos esta oppcion vemos (figura 13) que hay varias sub-opcciones entre las cuales hay `linear restrictions`. Seleccionando tal opccion se abre otra página (figure 14) en la que hay que escribir cual hipótesis nula queremos contrastar. Una vez declarada la hipótesis nula seleccionamos `ok` y obtenemos, figura 15, los resultados del contraste. GRETL nos proporciona el valor del estadistico de contraste F (en este caso igual a 123,858) y el valor-p correspondiente.

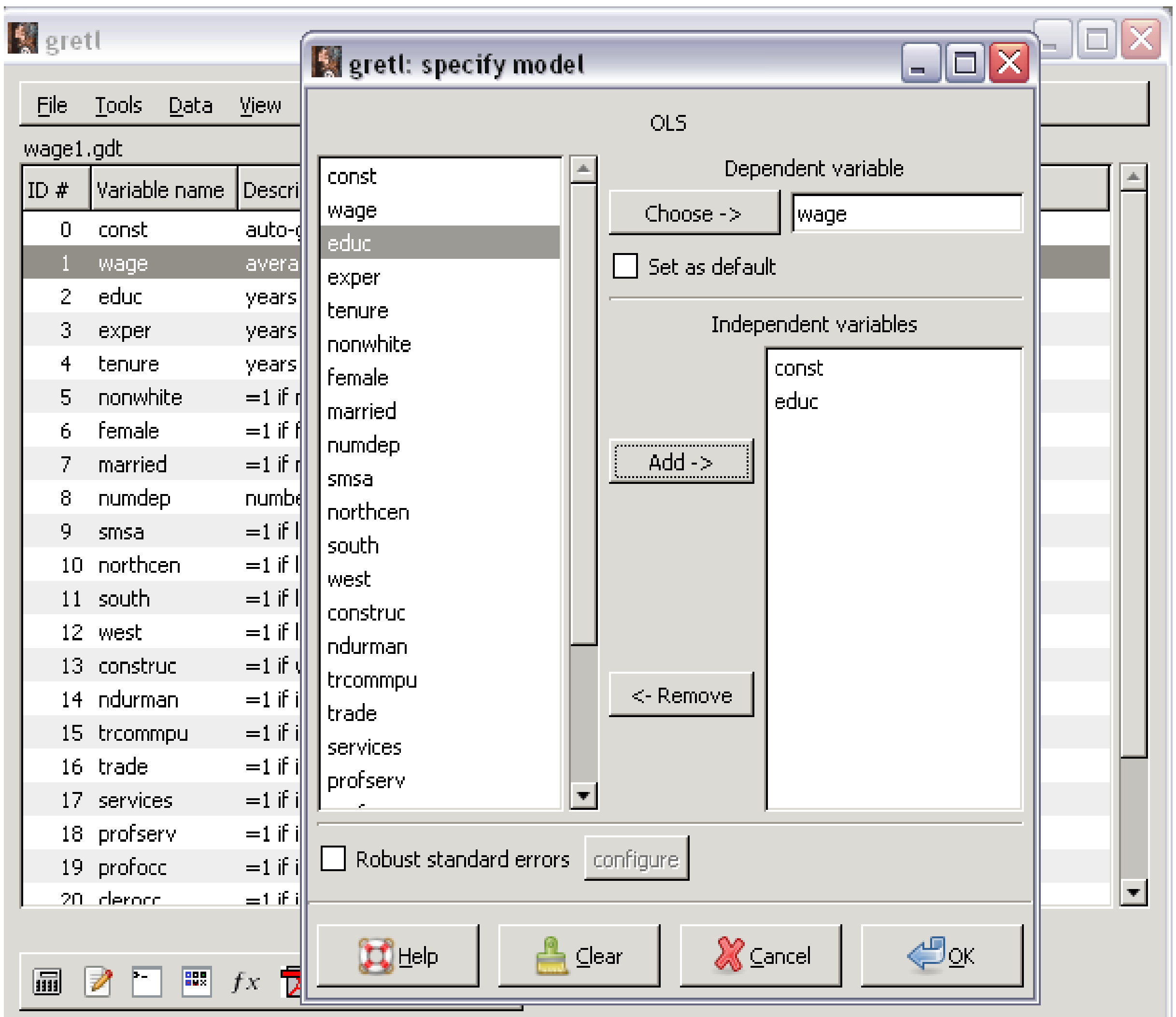
gretl
File Tools Data View
wage1.gdt
ID # Variable name Descri
0 const
1 wage
2 educ
3 exper
4 tenure
5 nonwhite
6 female
7 married
8 numdep
9 smsa
10 northcen
11 south
12 west
13 construc
14 ndurman
15 trcommpu
16 trade
17 services
18 profserv
19 profocc
gretl: specify model
OLS
const
wage
educ
exper
tenure
nonwhite
female
married
numdep
smsa
northcen
south
west
construc
ndurman
trcommpu
trade
services
profserv
Dependent variable
Choose ->
wage
Set as default
Independent variables
Add ->
const
educ
<- Remove
Robust standard errors
configure
Help
Clear
Cancel
OK

Figura 10

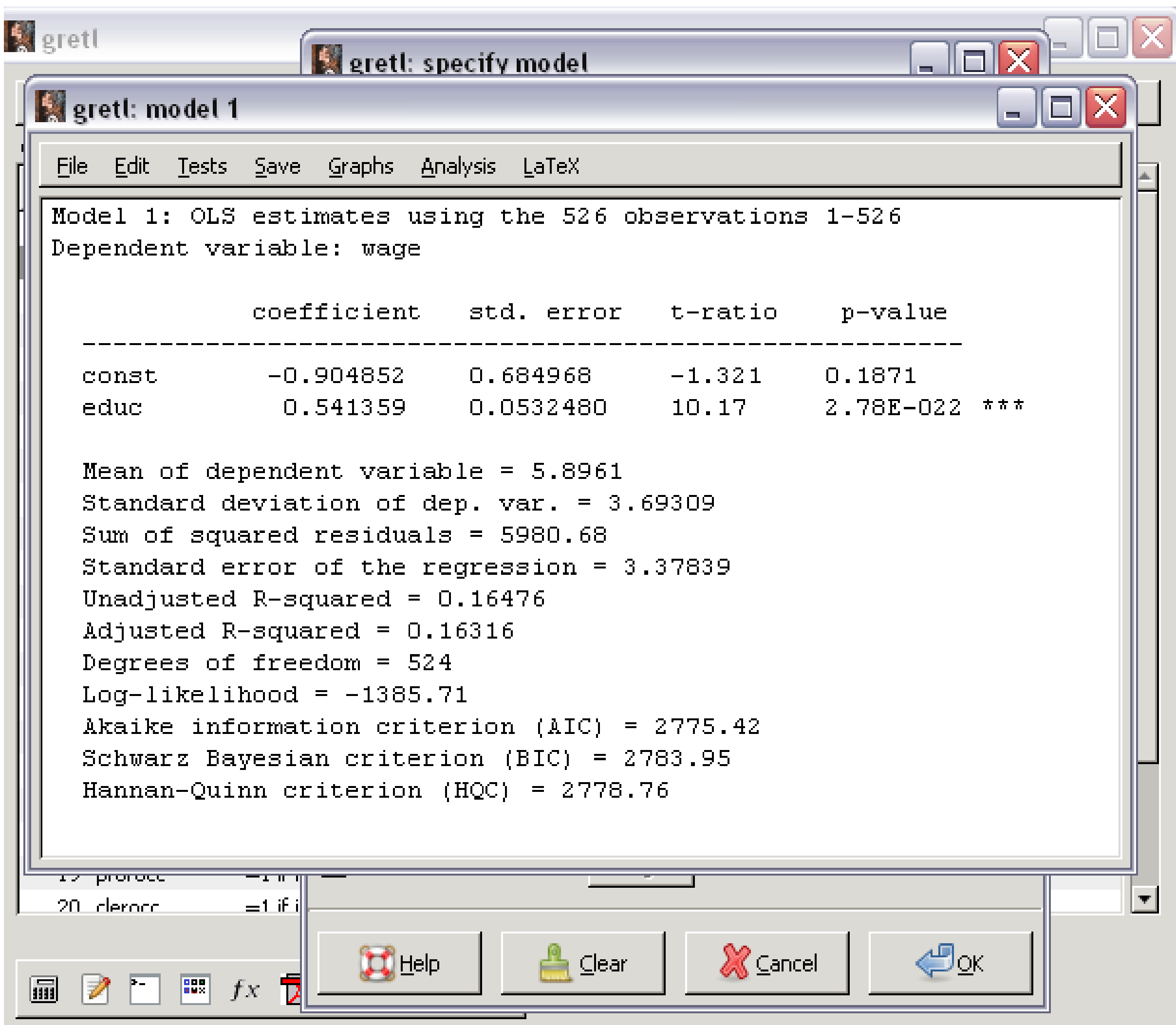
gretl
gretl: specify model
gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Model 1: OLS estimates using the 526 observations 1-526
Dependent variable: wage
coefficient std. error t-ratio p-value
const -0.904852 0.684968 -1.321 0.1871
educ 0.541359 0.0532480 10.17 2.78E-022 ***
Mean of dependent variable = 5.8961
Standard deviation of dep. var. = 3.69309
Sum of squared residuals = 5980.68
Standard error of the regression = 3.37839
Unadjusted R-squared = 0.16476
Adjusted R-squared = 0.16316
Degrees of freedom = 524
Log-likelihood = -1385.71
Akaike information criterion (AIC) = 2775.42
Schwarz Bayesian criterion (BIC) = 2783.95
Hannan-Quinn criterion (HQC) = 2778.76
20 clerocc =1 if i
Help
Clear
Cancel
OK

Figura 11

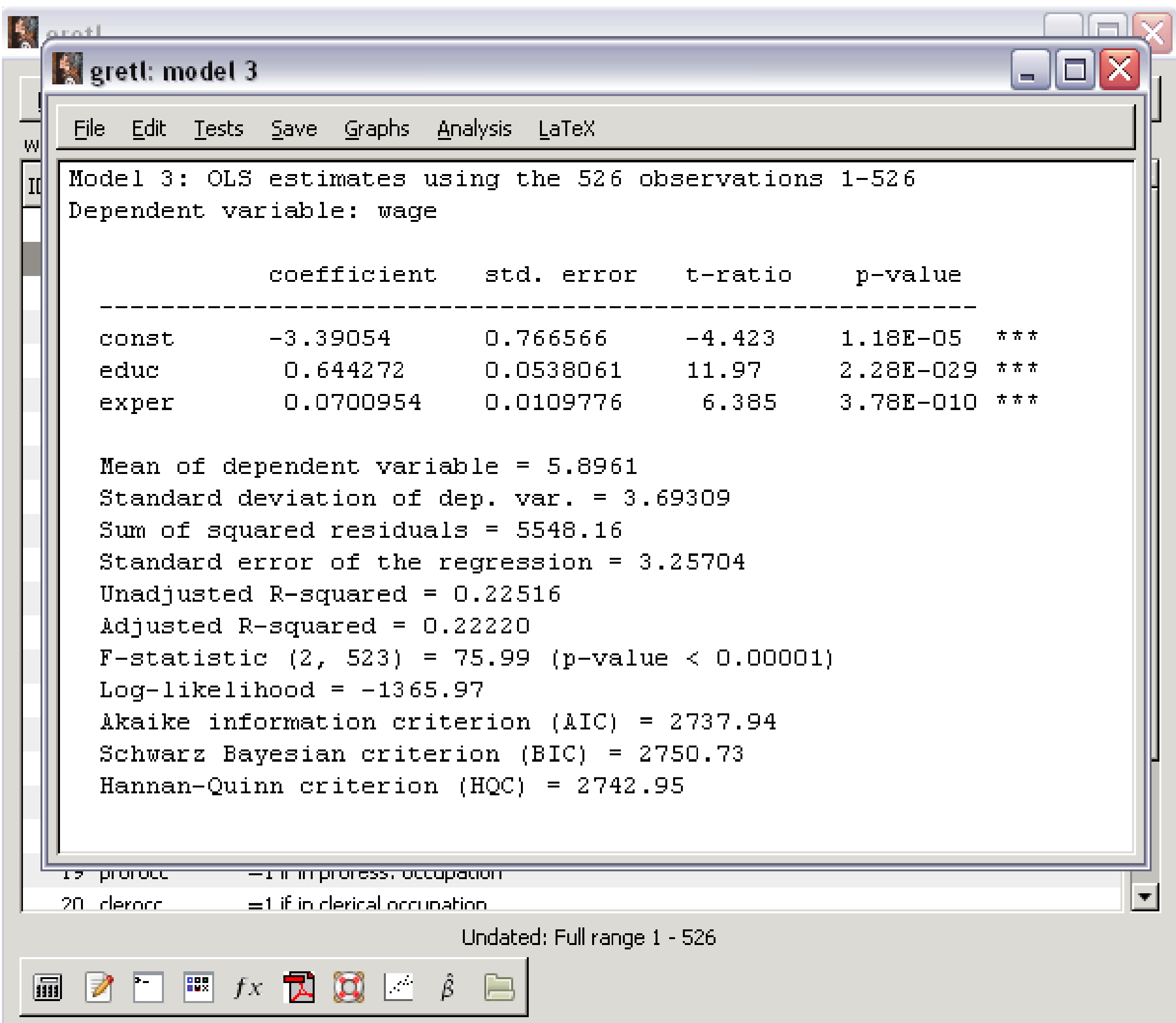
gretl: model 3
File Edit Tests Save Graphs Analysis LaTeX
Model 3: OLS estimates using the 526 observations 1-526
Dependent variable: wage
coefficient std. error t-ratio p-value
const -3.39054 0.766566 -4.423 1.18E-05 ***
educ 0.644272 0.0538061 11.97 2.28E-029 ***
exper 0.0700954 0.0109776 6.385 3.78E-010 ***
Mean of dependent variable = 5.8961
Standard deviation of dep. var. = 3.69309
Sum of squared residuals = 5548.16
Standard error of the regression = 3.25704
Unadjusted R-squared = 0.22516
Adjusted R-squared = 0.22220
F-statistic (2, 523) = 75.99 (p-value < 0.00001)
Log-likelihood = -1365.97
Akaike information criterion (AIC) = 2737.94
Schwarz Bayesian criterion (BIC) = 2750.73
Hannan-Quinn criterion (HQC) = 2742.95
Undated: Full range 1 - 526

Figura 12

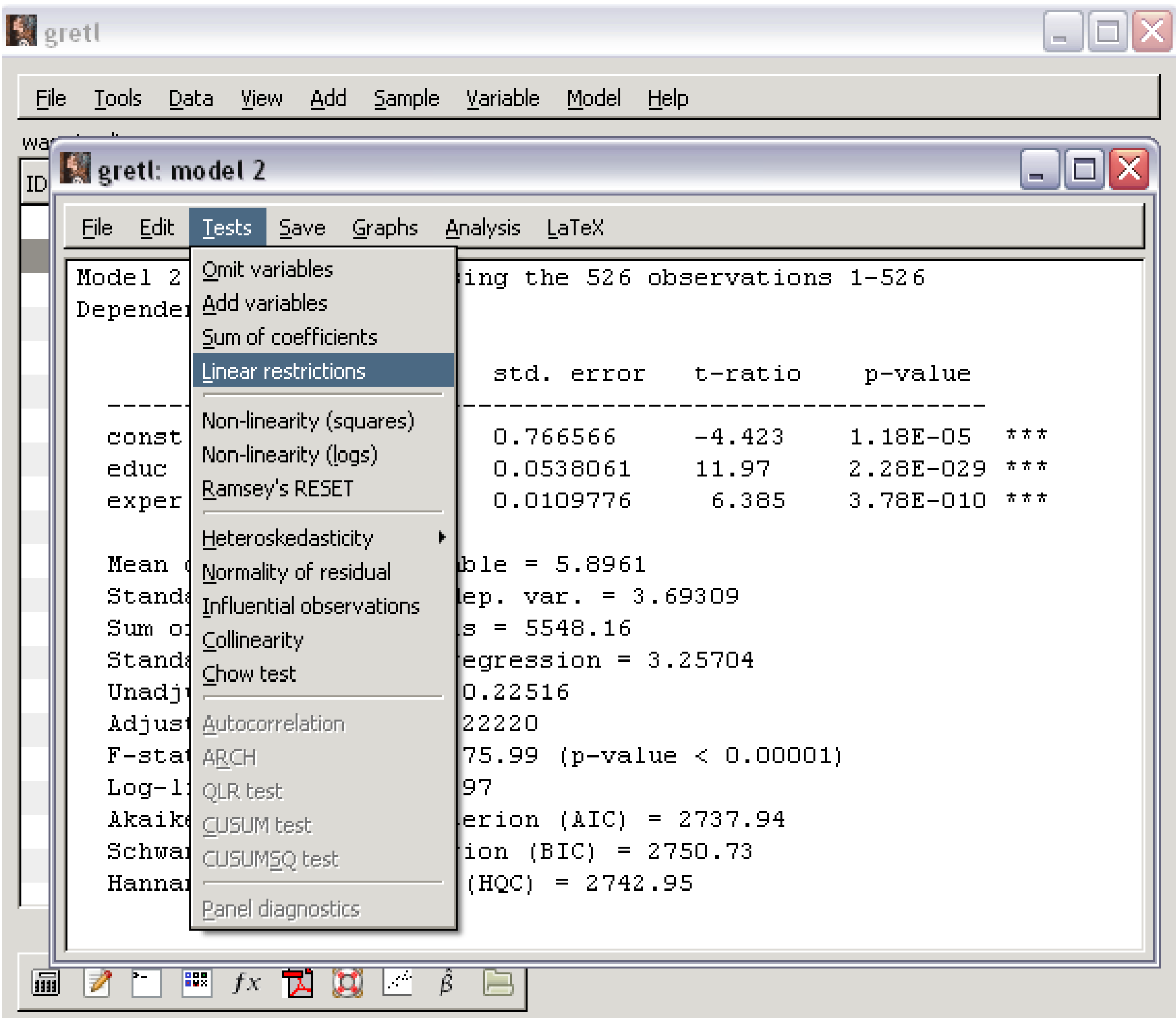
gretl
File Tools Data View Add Sample Variable Model Help
gretl: model 2
File Edit Tests Save Graphs Analysis LaTeX
Omit variables
Add variables
Sum of coefficients
Linear restrictions
Non-linearity (squares)
Non-linearity (logs)
Ramsey's RESET
Heteroskedasticity
Normality of residual
Influential observations
Collinearity
Chow test
Autocorrelation
ARCH
QLR test
CUSUM test
CUSUMSQ test
Panel diagnostics
Model 2 ing the 526 observations 1-526
std. error t-ratio p-value
const 0.766566 -4.423 1.18E-05 ***
educ 0.0538061 11.97 2.28E-029 ***
exper 0.0109776 6.385 3.78E-010 ***
ble = 5.8961
dep. var. = 3.69309
s = 5548.16
egression = 3.25704
0.22516
22220
75.99 (p-value < 0.00001)
97
erion (AIC) = 2737.94
ion (BIC) = 2750.73
(HQC) = 2742.95

Figura 13

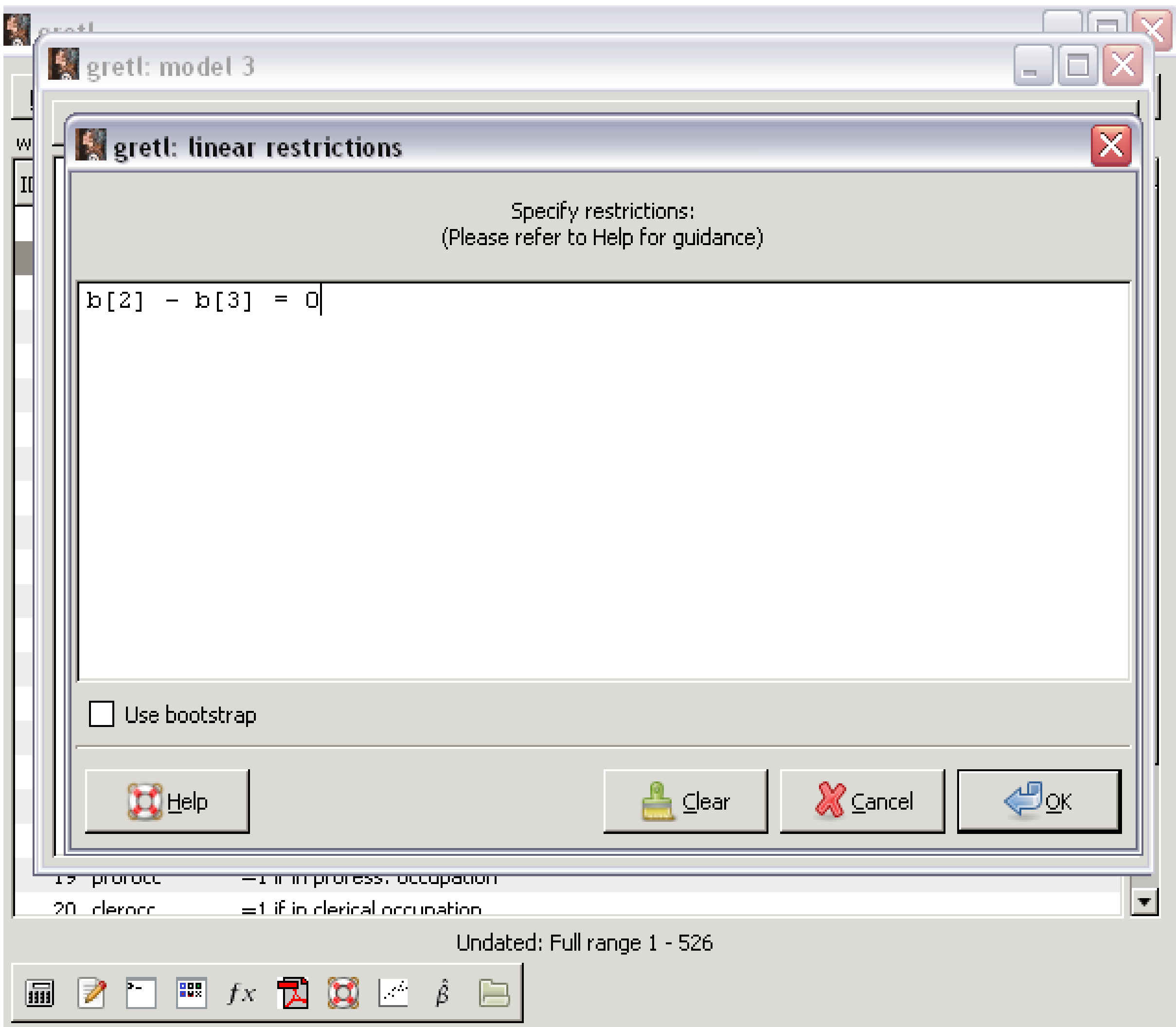
gretl: model 3
gretl: linear restrictions
Specify restrictions:
(Please refer to Help for guidance)
b[2] - b[3] = 0
Use bootstrap
Help
Clear
Cancel
OK
20 clerocc =1 if in clerical occupation
Undated: Full range 1 - 526

Figura 14

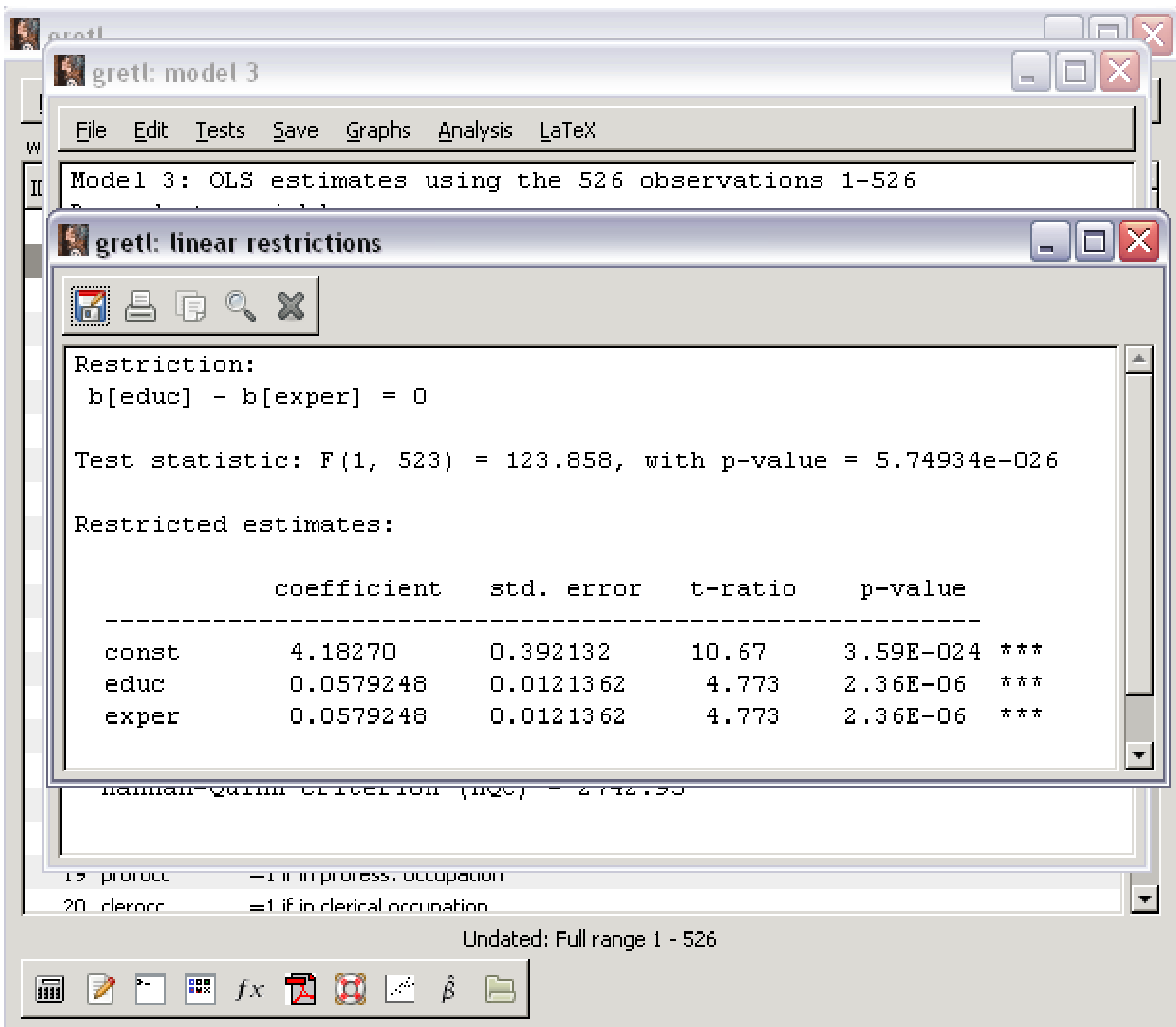
gretl: model 3
File Edit Tests Save Graphs Analysis LaTeX
Model 3: OLS estimates using the 526 observations 1-526
gretl: linear restrictions
Restriction:
b[educ] - b[exper] = 0
Test statistic: F(1, 523) = 123.858, with p-value = 5.74934e-026
Restricted estimates:
coefficient std. error t-ratio p-value
const 4.18270 0.392132 10.67 3.59E-024 ***
educ 0.0579248 0.0121362 4.773 2.36E-06 ***
exper 0.0579248 0.0121362 4.773 2.36E-06 ***
20 clerocc =1 if in clerical occupation
Undated: Full range 1 - 526

Figura 15

A.2 Exámenes

Examen Econometría (PUE) Diciembre 2007

1. El siguiente modelo es una versión simplificada del modelo utilizado por Biddle y Hamermesh (1990) para estudiar el *trade off* entre tiempo dedicado a dormir y a trabajar:

$$sleep = \beta_0 + \beta_1 totwork + \beta_2 educ + \beta_3 age + u$$

donde *sleep* y *totwork* son el tiempo (en minutos) dormido y trabajado por semana y *educ* y *age* son respectivamente educación y edad (en años). Utilizando 706 observaciones se obtienen las siguientes estimaciones

$$\hat{sleep} = 3638{,}25 - 0{,}148 totwork - 11{,}13 educ + 2{,}20 age \quad R^2 = 0{,}113$$

a) Se interpreten desde un punto de vista cuantitativo todos y cada uno de los coeficientes estimados.

b) Si alguien trabaja cinco horas mas por semana, ¿de cuanto cambia *sleep* según el modelo estimado?

c) ¿Que mide R^2? ¿Cual es la interpretación de este valor en este caso concreto?

d) ¿Suponiendo que $\sqrt{\frac{\sum(sleep_i - \bar{sleep})^2}{N-k}} = 445{,}3620$ (k = número de variables), estimar la desviación típica de los errores de regresión utilizando un estimador insesgado.

Sean $112{,}28, 0{,}017, 5{,}88, 1{,}45$ las desviaciones típicas de $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ respectivamente.

e) Contrastar al 5 % la hipótesis nula de que cada uno de los coeficientes sean separadamente iguales a cero.

f) Contrastar al 5 % la hipótesis nula de que un minuto mas trabajado corresponde a un minuto menos dormido.

g) Contrastar al 5 % la hipótesis nula de que los coeficientes $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ sean cero conjuntamente.

Supongamos ahora que sacamos *educ* y *age* de nuestro modelo. La estimación nos da

$$\hat{sleep} = 3638{,}38 - 0{,}151 totwork \quad R^2 = 0{,}103$$

h) ¿El hecho de incluir *age* y *educ* cambia substancialmente la relación entre trabajar y dormir?

i) ¿Podría ser R^2 en este caso mayor del que hemos obtenido en la estimación anterior? ¿Por que?

l) ¿Contrastar al 5 % la hipótesis nula de que *age* y *educ* sean conjuntamente iguales a cero.

2. Considere el siguiente modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

donde se conocen las siguientes cantidades: $\sum_i X_i = 37{,}2$ $\sum_i X_i^2 = 147{,}18$ $\sum_i Y_i = 75{,}50$ $\sum_i Y_i^2 = 597{,}03$ $\sum_t Y_i X_i = 295{,}95$, $N = 10$.

a) Estimar β_0, β_1 utilizando el estimador de mínimos cuadrados.

b) Calcular y interpretar el R^2 de la regresión.

c) Contrastar al 5 % la hipótesis que la pendiente de la recta de regresión sea igual a cero.

Valore críticos:

$t_{0,025,8} = 2{,}28$, $t_{0,025,702} = 1{,}96$, $F_{2,702} = 3{,}01$, $F_{3,702} = 2{,}62$

Examen Econometría (PUE) Septiembre 2008

La siguiente ecuación describe el precio (*price*) de las viviendas medido en miles de dólares en términos de los pies cuadrados de construcción (*sqrft*) y del número de habitaciones (*bdrms*)

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

Se satisfacen todos los supuestos del modelo de regresión lineal que hemos estudiado en clase. Usando los datos se obtiene la siguiente ecuación por MCO:

$$price = -19{,}315 + 0{,}128436 sqrft + 15{,}1982 bdrms + u, \quad n = 88, \quad R^2 = 0{,}631918$$

(i) ¿Cuál es la interpretación de $\hat{\beta}_1 = 0{,}128436$?

(ii) ¿Qué porcentaje de la variación en el precio de una casa se explica por la superficie y el número de habitaciones?

(iii) La desviación típica de $\hat{\beta}_1$ es 0.0138245. Contrastar al 5 % hipótesis nula $H_0 : \beta_1 = 1$.

(iv) Contrastar al 5 % hipótesis nula

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \end{cases}$$

Ahora especificamos el precio de las casas y la superficie en logaritmos. Utilizando los datos se obtiene la siguiente ecuación por MCO:

$$log(price) = -0{,}623398 + 0{,}808254 log(sqrft) + 0{,}0381107 bdrms + u,$$
$$n = 88,\ R^2 = 0{,}561136$$

(v) ¿Cuál es el aumento estimado en el precio de una casa cuando se añade una habitación adicional, manteniendo constantes los pies cuadrados?

(vi) ¿Cuál es ahora la interpretación de $\hat{\beta}_1 = 0{,}808254$?

Añadimos ahora otras dos variables, y se considere el siguiente modelo

$$log(price) \;=\; \beta_0 + \beta_1 log(sqrft) + \beta_2 bdrms + \beta_3 log(assess) + \beta_4 log(lotsize) + u$$

donde *assess* es la tasación de la casa antes de la venta y *lotsize* es el tamaño del solar. Usando los datos se obtiene la siguiente ecuación por MCO:

$$\begin{aligned} log(price) &= 0{,}263745 - 0{,}103239 log(sqrft) + 0{,}0338392 bdrms + ... \\ &\quad 1{,}04306 log(assess) + +0{,}00743824 log(lotsize) + u, \\ &\quad n = 88, \quad R^2 = 0{,}772809 \qquad (3.4) \end{aligned}$$

(vii) Contrastar al 5 % la hipótesis nula de que los coeficientes de $log(assess)$ y $log(lotsize)$ sean conjuntamente iguales a cero.

(viii) En el modelo siguiente

$$log(price) - log(assess) \;=\; \beta_0 + u$$

la suma de los residuos al cuadrados es SCR=1.880. En el modelo (1) la SCR=1.822. Contrastar la hipótesis nula

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}$$

Valores críticos: $t_{0,025,85} = 1{,}98$, $F_{2,85} = 3{,}1$, $F_{2,83} = 3{,}11$, $F_{4,83} = 2{,}48$

Examen Econometría I (PUE) Diciembre 2008

Consideremos el siguiente modelo para comparar el rendimiento de la educación en escuelas de formación profesional y en universidades.

$$log(wage) = \alpha + \beta_1 jc + \beta_2 univ + u \tag{3.5}$$

donde *wage* es el salario por hora, *jc* es el número de años como estudiante en una escuela de formación profesional y *univ* es el número de años en una universidad. Utilizando los datos de Kane y Rouse (1995), obtenemos la siguiente estimación:

$$log(\hat{wage}) = 2{,}091 + 0{,}070 jc + 0{,}069 univ, \quad n = 6763, \quad R^2 = 0{,}109, \quad SCR = 1432{,}93.$$

(i) ¿Cual es la interpretación de $\hat{\beta}_1 = 0{,}070$ y $\hat{\beta}_2 = 0{,}069$?

(ii) ¿Cual es el valor predicho de $log(wage)$ para un individuo con cuatro años de estudios universitarios y sin estudios en escuela de formación profesional?

(iii) ¿Qué porcentaje de la variación en el logaritmo del salario se explica por las variables independientes incluidas en el modelo?

(iv) La desviación típica de $\hat{\beta}_2$ es 0.0024. Contrastar al 5 % la hipótesis nula de que la educación universitaria no es importantes para explicar el logaritmo del salario ($H_0 : \beta_2 = 0$).

(v) Contrastar al 5 % la hipótesis nula de que los coeficientes β_1 y β_2 sean cero conjuntamente.

Ahora añadimos al modelo la variable *exper* que mide la experiencia acumulada en el trabajo

$$log(wage) = \alpha + \beta_1 jc + \beta_2 univ + \beta_3 exper + u. \tag{3.6}$$

Con los datos se obtiene la siguiente estimación:

$$\hat{log(wage)} = 1{,}472 + 0{,}0667jc + 0{,}0769univ + 0{,}0049exper$$
$$n = 6763, \quad SCR = 1250{,}54.$$

(vi) ¿Puede ser R^2 en este caso menor que 0,109?¿Por qué?

(vii) Contrastar al 5 % la hipótesis nula de que la experiencia no es importante para explicar el logaritmo del salario ($H_0 : \beta_3 = 0$).

(viii) La hipótesis nula $H_0 : \beta_1 = \beta_2$ es muy interesante porque bajo esta hipótesis un año adicional en una escuela de formación profesional o en una universidad conducen al mismo aumento porcentual en el salario cuando son constantes todos los demás factores. Esto nos permite comparar el rendimiento de diferentes tipos de educación. Sabiendo que $(R(X'X)^{-1}R')^{-1} = 3846$ contrastar dicha H_0 al 5 %.

(ix) Sea $\theta = \beta_1 - \beta_2$ y sea $totcol = jc + univ$. Mostrar que el modelo (2) se puede escribir como

$$log(wage) = \alpha + \theta jc + \beta_2 totcol + \beta_3 exper + u \tag{3.7}$$

y explicar un procedimiento para contrastar la hipótesis nula $H_0 : \beta_1 = \beta_2$ en este modelo.

Valores críticos: $t_{0,025,6760} = 1{,}96$, $F_{2,6760} = 2{,}99$, $F_{1,6759} = 3{,}84284$.

Bibliografia

1. Greene, W. (1998), Análisis Econométrico, Macmillan Publishing Company, New York.

2. Gujarati, D. (2003), Econometría, Ed. McGraw-Hill 4.a edición.

3. Johnston, J y DiNardo, J. (2001), Métodos de Econometría, Ed. Vicens-Vives 3.a edición.

4. Stock, J.H. y M.W. Watson (2003): Introduction to Econometrics. Pearson Education, International Edition

5. Wooldridge, J.M. (2006), Introducción a la Econometría: un Enfoque Moderno. Paraninfo Thompson Learning, 2ª Ed.