NOTAS SOBRE ECONOMETRIA CON EVIEWS: ESTIMACION

F. MARMOL
Departamento de Estadística y Econometría
Universidad Carlos III de Madrid

Curso 2001-2002

El modelo de interés es el **modelo de regresión lineal múltiple clásico:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_j X_{ji} + \cdots + \beta_K X_{Ki} + u_i,$$

$i = 1, 2, ..., n.$

Como ilustración, estudiaremos el fichero de datos "**estres.txt**" (véase la página web de la asignatura).

El fichero contiene 15 observaciones para analizar la relación existente entre el grado de estrés de los trabajadores (Y), medido a partir de su tensión nerviosa, y el tamaño de la empresa en la que trabajan ($X1$), el número de años que llevan en el puesto de trabajo actual ($X2$), el salario anual percibido ($X3$) y la edad del trabajador ($X4$).

A la hora de especificar el modelo, es siempre aconsejable **en primer lugar realizar una primera inspección gráfica y numérica de las series,** pues esto nos va a permitir analizar el grado y la forma de la relación existente entre las variables.

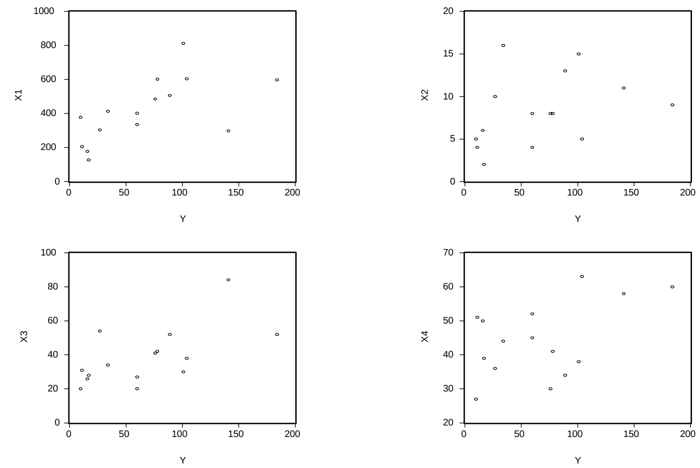

Figura 1:

Por ello, lo primero que deberíamos realizar es abrir las series como un grupo:

Seleccionamos las variables (en orden) mediante la tecla CTRL y el botón izquierdo del ratón.

Hacemos doble clic en cualquier parte del área sombreada en la selección y pulsamos **Open Group** en el cuadro de diálogo resultante. Esta operación nos abre la ventana del objeto grupo deseado. Los gráficos y el análisis descriptivo individual y conjunto de las series se obtienen a partir de las representaciones (**Views**) que aparecen en la ventana del objeto grupo.

Así, por ejemplo, mediante **View/Multiple Graphs/Scatter/First variable against all** obtendríamos la representación de los pares de valores $j = 1, 2, 3, 4$, para el fichero "estres.txt" de la Figura 1.

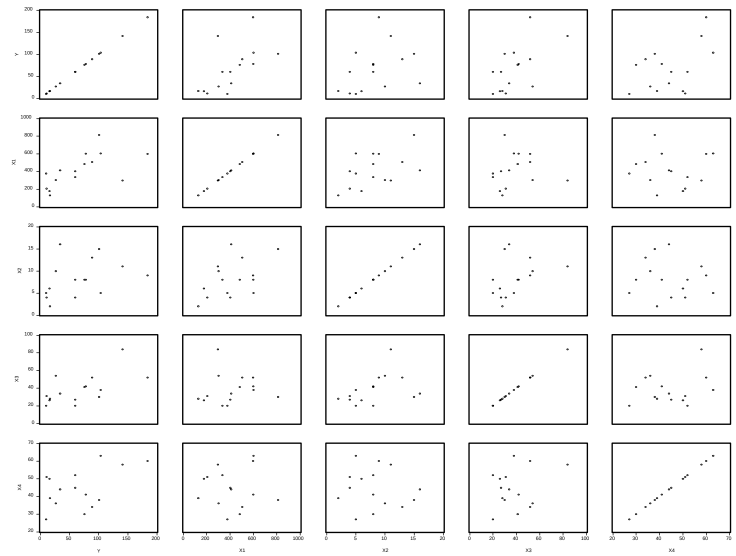

Figura 2:

En nuestro caso, los gráficos de la Figura 1 muestran una clara relación positiva lineal entre el grado de estrés (Y) y cada una de las variables consideradas, si bien parece que la relación es más fuerte con el tamaño de la empresa (X_1) y el salario anual (X_3).

Una visión más general de la relación entre el conjunto de variables consideradas se obtendría mediante **View/Multiple Graphs/Scatter/Matrix of all pairs** (**SCATMAT**).En este caso la salida adopta la forma de una matriz en la que cada elemento recoge la representación gráfica de cada par de variables, incluidas también las variables explicativas (Figura 2).

El resumen numérico de la información que suministran estos gráficos

se recoge en la matriz de correlaciones entre las variables, que se obtiene usando **View/Correlations**, y cuyo resultado para el ejemplo que estamos siguiendo es:

	Y	X_1	X_2	X_3	X_4
Y	1.000000	0.621569	0.355443	0.614634	0.485174
X_1	0.621569	1.000000	0.501089	0.088364	-0.018855
X_2	0.355443	0.501089	1.000000	0.384154	-0.113456
X_3	0.614634	0.088364	0.384154	1.000000	0.260760
X_4	0.485174	-0.018855	-0.113456	0.260760	1.000000

De esta matriz de correlaciones vemos que tanto $X1$ como $X3$ son las variables que presentan mayor correlación lineal con el nivel de estrés y que existe una cierta correlación lineal entre $X1$ y $X2$, así como entre $X2$ y $X3$.

Las características descriptivas básicas de cada serie se obtienen con **View/Descriptive Stats/Common Sample.** Los resultados para nuestro ejemplo son:

	Y	X_1	X_2	X_3	X_4
Mean	67.20000	415.7333	8.266667	38.60000	44.53333
Median	60.00000	401.0000	8.000000	34.00000	44.00000
Maximum	184.0000	812.0000	16.00000	84.00000	63.00000
Minimum	10.00000	127.0000	2.000000	20.00000	27.00000
Std. Dev.	51.16388	187.5132	4.148436	16.74515	10.94706
Skewness	0.755415	0.353083	0.431821	1.311686	0.143565
Kurtosis	2.869032	2.527598	2.237346	4.549391	1.981337
Jarque-Bera	1.437350	0.451147	0.829700	5.801683	0.700074
Probability	0.487398	0.798058	0.660439	0.054977	0.704662
Observations	15	15	15	15	15

Una vez realizado el análisis descriptivo y confirmada la sensatez de realizar un modelo de regresión lineal, pasamos a estimar los parámetros del modelo.

Para ello, los resultados numéricos de la estimación del modelo propuesto pueden obtenerse de varias maneras:

- Como un procedimiento (**Procs**) del "objeto grupo". Para llevar a cabo la estimación tendríamos que pulsar el botón **Procs/Make Equation** de la ventana del objeto grupo.
- Abriendo las series que intervienen como un "objeto ecuación". Para ello, situados en la ventana del Workfile, tendríamos que seleccionar las series que intervienen en el modelo mediante la tecla CTRL y el botón izquierdo del ratón, hacer doble clic en el área sombreada por la selección y pulsar **Open Equation** en el cuadro resultante.
- Seleccionando el mandato **Quick/Estimate Equation** en el menú principal del programa.

Los tres caminos conducen al mismo cuadro de diálogo de la especificación del modelo.

La primera información que debemos suministrar es la ecuación del modelo de regresión. Con respecto a esta cuestión existen varias opciones. Una primera opción es introducir una lista que incluya el nombre de la variable dependiente seguida de los nombres de las variables independientes o regresores, teniendo en cuenta que el término independiente va asociado a una variable que el programa siempre nombra C. Una segunda opción es introducir la ecuación propiamente dicha nombrando a los coeficientes con $C(1), C(2)$...En nuestro ejemplo, deberíamos escribir:

$$Y = C(1) + C(2) * X1 + C(3) * X2 + C(4) * X3 + C(5) * X4$$

Por defecto, EViews utiliza la primera opción manteniendo el orden de la selección de las variables en el grupo e introduciendo siempre el término independiente al final de las mismas.

Una vez establecida la ecuación a estimar existen varias posibilidades en cuanto al método de estimación a emplear. Por defecto, Eviews utiliza mínimos cuadrados ordinarios, **LS-Least Squares (NLS and ARMA)**, aunque cuenta con otros métodos de estimación.

Establecida la ecuación a estimar y el método de estimación, sólo queda señalar la muestra de datos a emplear. Eviews utiliza la que esté activa en ese momento.

A este respecto, uno de los procedimientos del Workfile es la posibilidad de cambiar la muestra con la que vamos a trabajar. Para efectuarlo seleccionamos **Procs/Sample** en su Barra de Herramientas o hacemos doble clic en la línea del Workfile que nos indica la muestra actual, dos líneas por debajo de dicha barra. La opción **Sample** también aparece como otro botón en la barra de herramientas.

Cualquiera que sea el camino que tomemos, aparece un cuadro de diálogo en el que se presenta por defecto, en la parte superior, la muestra existente en ese momento, que, de no haber sido modificada, coincidirá con el rango especificado al inicio de la sesión de ordenador. En dicha ventana podemos cambiar la muestra escribiendo la observación inicial y final del intervalo de trabajo que queramos (siempre, por supuesto, que estemos dentro del rango).

La muestra seleccionada no tiene por qué ser un intervalo continuo, sino que también podemos escoger varios intervalos o bien aquellos elementos de la muestra que cumplan una determinada condición.

Para escoger varios intervalos no solapados dentro de la muestra disponible, indicaremos en esa misma ventana las observaciones iniciales y finales de cada uno de los intervalos elegidos, con la única condición de que guarden entre sí un orden de menor a mayor.

Por ejemplo, supongamos que los datos son anuales y que el rango especificado al inicio de la sesión fue de

1964 1986

Ahora, si escribimos

1965 1968 1970 1975 1980 1986

estaríamos seleccionando una muestra que comprendería tres intervalos de 1965 a 1968, de 1970 a 1975 y de 1980 a 1986. Todas las observaciones que no pertenezcan a tales intervalos no se incluirían dentro de la muestra. Si alguno de estos intervalos hubiera estado formado solamente por una observación, la forma de seleccionarlo sería repitiendo la misma fecha como observación inicial y final. Por ejemplo, si en el caso anterior el primer intervalo fuera únicamente el año 1965, las dos primeras fechas deberían ser

1965 1965.

Eviews presenta ciertas abreviaturas que pueden utilizarse a la hora de describir el intervalo. Así, si queremos volver al total del rango desde el ejemplo anterior, en lugar de escribir de nuevo 1964 1986 podemos escribir **@all**. Otras abreviaturas posibles son **@first** y **@last** para la primera y última observación, respectivamente, del rango total. Por ejemplo

@first 1967 1980 **@last**

seleccionaría los intervalos que van desde 1964 a 1967 y de 1980 a 1986.

Igualmente, si queremos reservar un número concreto de observaciones a partir de una determinada, se puede indicar escribiendo la primera fecha como fecha de inicio y la primera fecha + el número de observaciones que completan el intervalo deseado como fecha final. Así, escribiendo 1970 1970+12 escogeríamos los 13 datos que hay entre 1970 y 1982.

Si adicionalmente queremos que los elementos de la muestra seleccionada cumplan una determinada condición, podemos utilizar la ventana inferior del cuadro de diálogo de **Sample**. En este caso, la muestra resultante serán la intersección del conjunto de observaciones indicadas en la ventana superior y del conjunto de observaciones definidas por la condición incluida en la ventana inferior. Por ejemplo, si una de las variables se denomina $X1$, por ejemplo, entonces si escribimos respectivamente en cada ventana:

1964 1986
$X1 > 0$

estaríamos seleccionando como muestra las observaciones de todas las variables comprendidas entre 1964 y 1986 para las que la variable $X1$ es positiva.

De igual forma, si queremos seleccionar aquellas observaciones para las que $X1$ ha decrecido con respecto al año anterior las órdenes de cada ventana serían:

1964 1986
$X1 < X1(-1)$

También pueden construirse condiciones más complejas utilizando las órdenes **AND** y **OR**. Por ejemplo,

1964 1986
$X1 > 0$ and $X2 >= 1$

selecciona las observaciones para las que $X1$ es positiva y $X2$ es mayor o igual que 1.

Como otro ejemplo,

1965 1975 1980 **@last**
($X1 >= -10$ **and** $X1 <= 100$) **or** $X2 >= 1$

selecciona las observaciones comprendidas entre los años 1965 a 1975 y 1980 a 1986 para las que $X1$ esté comprendida entre -10 y 100 o $X2$ sea mayor o igual a 1.

Si dispusiéramos de más observaciones y, por lo tanto, quisiéramos aumentar el rango fijado, el procedimiento a seguir sería seleccionar en la barra de herramientas del Workfile **Procs/Change workfile Range** o de forma más rápida, hacer doble clic sobre la palabra **Range** que aparece debajo de la barra de herramientas del Workfile.

Volviendo al ejemplo sobre el estrés laboral, como resultado de la estimación aparecen los siguientes datos:

Dependent Variable: Y

Method: Least Squares

Sample: 1 15

Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-126.5053	32.28107	-3.918871	0.0029
X1	0.176293	0.040095	4.396907	0.0013
X2	-1.562948	2.012053	-0.776793	0.4553
X3	1.574538	0.445674	3.532933	0.0054
X4	1.629285	0.628717	2.591444	0.0269

R-squared 0.842423	Mean dependent var 67.20000
Adjusted R-squared 0.779393	S.D. dependent var 51.16388
S.E. of regression 24.03109	Akaike info criterion 9.457775
Sum squared resid 5774.932	Schwarz criterion 9.693792
Log likelihood -65.93331	F-statistic 13.36529
Durbin-Watson stat 2.437614	Prob(F-statistic) 0.000506

En la parte superior aparece el nombre de la variable dependiente, el método de estimación empleado, el periodo muestral y el número de observaciones que abarca el mismo.

En segundo lugar, encontramos la estimación de los coeficientes asociados a cada una de las variables explicativas (**Coefficient**) junto con su error estándar o desviación típica estimada (**Std Error**), el estadístico t de significación individual (**t-Statistic**) y el correspondiente p-valor para contrastar la hipótesis nula de no significatividad (**Prob.**). Prob proporciona dos veces el área que el valor absoluto del estadístico t deja a su derecha e indica la probabilidad de cometer el error de Tipo I. Estos valores están calculados a partir de la distribución t-Student con $n-(K+1)$ grados de libertad.

En la notación del curso, tendríamos:

Coefficient: $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, $\widehat{\beta}_4$

Std. Error: s_{β_0}, s_{β_1}, s_{β_2}, s_{β_3}, s_{β_4}

t-Statistic: $Z_0^{'}$ para cada coeficiente, $H_0 : \beta_j = 0$, $H_a : \beta_j \neq 0$, $j = 0, 1, 2, 3, 4$.

En tercer lugar, en la zona inferior, hay un bloque de estadísticos que nos permiten evaluar parcialmente la regresión realizada.

R-Squared es el coeficiente de determinación (R^2), medida estadística que sirve para valorar el éxito de la regresión para predecir los valores de la variable dependiente dentro del período muestral.

Adjusted R-squared es el coeficiente de determinación corregido ($\overline{R}^2$). Permite comparar la capacidad explicativa de modelos referidos a una misma muestra de la misma variable endógena con distinto número de variables explicativas.

S.E. of regression: error estándar de la regresión ($s = \sqrt{\frac{1}{n-K-1}\sum_{i=1}^{n}\widehat{u}_i^2}$),

Sum squared resid: suma de los cuadrados de los residuos $\left(\sum_{i=1}^{n}\widehat{u}_i^2\right)$.

Mean dependent var: Media muestral de la variable dependiente $\left(\overline{Y}\right)$.

S.D. dependent var: desviación típica muestral de la variable dependiente $\left(S_{\overline{Y}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}\right)$.

Log Likelihood: valor del logaritmo de la función de verosimilitud evaluada en los estimadores de máxima verosimilitud. En este sentido, cuando las variables se supone que siguen una distribución normal, los estimadores MV del modelo de regresión lineal clásico $\left(\widehat{\beta}_{MV}, s_{MV}^2\right)$se demuestran que son iguales a:

$$\widehat{\beta}_{MV} = \widehat{\beta}_{MCO}, \qquad s_{MV}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^2.$$

Durbin-Watson stat: estadístico de Durbin-Watson. Sirve para contrastar la hipótesis nula de perturbaciones no correlacionadas frente a la alternativa de perturbaciones correlacionadas (más detalles de este contraste en notas posteriores).

Akaike info criterion (AIC): criterio de información de Akaike. El criterio es elegir entre un grupo de modelos a aquel que tenga menor AIC

Schwarz criterion (SIC): criterio de Schwarz. El criterio es elegir entre un grupo de modelos a aquel que tenga menor SIC.

AIC y SIC son dos estadísticos que sirven para analizar la capacidad explicativa de un modelo y permiten realizar comparaciones a este respecto entre modelos anidados.

F-statistic: estadístico que contrasta la hipótesis nula de significatividad global, es decir, la hipótesis nula de que todos los parámetros (excepto término constante) son cero frente a la hipótesis alternativa de que al menos

uno de ellos es significativo. Permite pues contrastar la capacidad explicativa conjunta de las variables introducidas en el modelo.

Prob(F-statistic): al igual que con el estadístico t, este valor mide la probabilidad de cometer error tipo I. Para calcular este p-valor se utiliza la distribución F con K grados de libertad en el numerador y $n-(K+1)$ grados de libertad en el denominador.

En la barra de la ventana que contiene los resultados de la estimación, tenemos las representaciones y procedimientos que Eviews tiene definidos de forma específica para una ecuación.

Concretamente, en **View** tenemos:

Representations: muestra la especificación del modelo y el modelo finalmente estimado.

En el ejemplo sobre el estrés laboral obtenemos:

Estimation Command:

=====================

LS Y C X1 X2 X3 X4

Estimation Equation:

=====================

Y = C(1) + C(2)*X1 + C(3)*X2 + C(4)*X3 + C(5)*X4

Substituted Coefficients:

=====================

Y=-126.5053228+0.1762933627*X1-1.562947623*X2

+1.574537804*X3+1.629285252*X4

Estimation Output: presenta la salida de la estimación, ya recogida en Representations. Esta representación puede mostrarse más rápidamente con el botón **Stats** de la barra de herramientas de la ventana de la ecuación.

Covariance Matrix: muestra la matriz de varianzas y covarianzas estimada de los estimadores de los coeficientes del modelo. La raíz cuadrada de los elementos de la diagonal principal de ésta son los errores estándar (**Std Error**) de los estimadores (**Coefficient**) que aparecen en los resultados generales de la estimación.

En el ejemplo sobre el estrés laboral obtenemos:

	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$
$\widehat{\beta}_0$	1042.067	-0.3274	-15.1088	-1.2603	-15.5817
$\widehat{\beta}_1$	-0.3274	0.00161	-0.04138	0.00275	-0.00236
$\widehat{\beta}_2$	-15.1088	-0.04138	4.048357	-0.3989	0.31979
$\widehat{\beta}_3$	-1.2603	0.00275	-0.3989	0.19863	-0.0955
$\widehat{\beta}_4$	-15.5817	-0.00236	0.31979	-0.0955	0.3953

En la notación del curso, la **matriz de varianzas y covarianzas** de los estimadores por mínimos cuadrados se denota:

$$Var\left(\widehat{\beta}\right) = \sigma^2 \left(\sum_{i=1}^{n} X_i X_i^{'}\right)^{-1},$$

donde

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \qquad X_i = \begin{pmatrix} X_{0i} \\ X_{1i} \\ \vdots \\ X_{Ki} \end{pmatrix},$$
$$X_{0i} = 1, \qquad i = 1, 2, ..., n.$$

La **matriz de varianzas y covarianzas estimada** de $\widehat{\beta}$ sería:

$$\widehat{Var}\left(\widehat{\beta}\right) = s^2 \left(\sum_{i=1}^{n} X_i X_i^{'}\right)^{-1},$$

que es lo que EViews denomina **Covariance Matrix.**

Actual, Fitted, Residual contiene varias opciones que permiten obtener tanto gráficos como tablas de los valores observados (**Actual**) y estimados (**Fitted**) de la variable dependiente y de los errores (**Residual**) que se cometen al realizar la estimación.

Las distintas posibilidades son:

Actual, Fitted, Residual Table: los gráficos de los residuos (**Residual Plot**) presenta dos líneas discontinuas que vienen dadas por $\pm 2s$, donde s es el error estándar de la regresión. Estas bandas permiten analizar la presencia de datos atípicos o la existencia de residuos significativamente distintos de cero.

Actual, Fitted, Residual Graph: Representa en el eje de ordenadas de la izquierda los residuos y en el derecho los valores de la variable dependiente observados y estimados. Obsérvese que el modelo de regresión será tanto mejor cuanto mejor reproduzca el comportamiento de la variable dependiente. El gráfico de los residuos permite ver si existe un comportamiento sistemático en los mismos que nos haga sospechar de algún error de especificación. Las bandas en la parte inferior, correspondiente a los errores, vienen dadas por $\pm 2s$. Este gráfico puede también obtenerse pulsando sobre el botón **Resids** de la barra de herramientas de la ventana de la ecuación.

Finalmente Eviews presenta también la posibilidad de realizar sólo el gráfico de los errores o residuos (**Residual Graph**) y de los errores estandarizados (**Standardized Residual Graph**). Estos últimos se obtienen dividiendo entre s cada error, pues al tener término independiente el modelo, la media de los errores es cero.

Por último, toda la información se perderá si realizamos nuevas operaciones. Para conservar el trabajo realizado, debemos realizar las siguientes operaciones:

Name: se encuentra en la barra de herramientas de la ventana de la ecuación. Sirve para darle un nombre a la ecuación y poder recuperarla posteriormente.

Procs/Make Residual Series: para darle un nombre a la serie de los residuos y poder recuperarlos posteriormente. Por defecto, el programa los denomina RESID01.

Procs/Forecast: para darle un nombre a la serie de valores ajustados de la variable dependiente y poder recuperarlos posteriormente. Por defecto el programa los denomina YF, donde Y es el nombre de la variable dependiente y F es de Forecast (predecir, en inglés). Más rápidamente, podemos usar **Forecast** en la barra de herramientas de la ventana de la ecuación.

Procs/Make Regressor Group: para almacenar las variables del modelo.

NOTA:

Los estadísticos que aparecen en los resultados de la estimación se almacenan temporalmente en @-funciones y pueden ser utilizados a la hora de generar series, escalares, vectores o matrices. Hay dos tipos de @-funciones, la que devuelven un escalar y las que contienen un vector o matriz. En cuanto a estimación, las @-funciones que contienen escalares son (véase lista completa en el Help del programa):

@regobs: número de observaciones incluídas.
@coefs(i): valor del estimador MCO del parámetro i del modelo
@stderrs(i): Std error del estimador MCO del parámetro i del modelo
@tstats(i): t-statistic del estimador MCO del parámetro i del modelo
@rs: coeficiente de determinación
@rbar2: coeficiente de determinación corregido

@se: error estándar de la regresión
@ssr: Suma cuadrados de los residuos
@logl: Log de la función de verosimilitud
@dw: estadístico Durbin-Watson
@meandep: media muestral de la variable dependiente
@sddep: desviación típica muestral de la variable dependiente
@aic: Criterio de Akaike
@schwarz: Criterio de Schwarz
@f: F-statistic.

Además, @cov(i,j) indica el elemento i,j de la matriz de varianzas y covarianzas estimada de los estimadores MCO, y @ncoef recoge el número de parámetros incluidos en la ecuación del modelo.

Las @-funciones que permiten obtener un vector o matriz son:

@coefs: da el vector con las estimaciones de los parámetros
@stderrs: proporciona un vector con los errores estándar de los estimadores MCO
@tstats: genera un vector con los estadísticos t asociados a cada parámetro
@cov: devuelve la matriz de varianzas y covarianzas estimada de los estimadores MCO.

Por ejemplo, si quisiéramos generar un objeto escalar, llamado chi, que recoja el número de observaciones multiplicado por el coeficiente de determinación de esta ecuación, escribiríamos en la ventana de comandos:

scalar chi=@regobs*@r2

y pulsamos INTRO. El resultado sería un nuevo icono escalar, en la ventana del Workfile seguido del nombre “chi”. Para ver su contenido hacemos doble clic sobre el mismo y **su valor se muestra en la parte izquierda de la línea de estado.**

Eviews, como siempre, solamente almacena el valor más reciente de estas @-funciones. No obstante, podemos darle nombre a la ecuación y guarda. Por ejemplo, se le llamamos eq01, tendríamos que escribir

s
calar chi=eq01.@regobs*eq01.@r2

y pulsar INTRO. Evidentemente, si tenemos dos o más ecuaciones guardadas, podemos combinar las órdenes.