

GUÍA DE MANEJO DEL PROGRAMA E-VIEWS

CAPÍTULO II

TRATAMIENTO BÁSICO DE SERIES ESTADÍSTICAS. INTRODUCCIÓN AL PROCEDIMIENTO DE REGRESIÓN.

Ramón Mahía
Departamento de Economía Aplicada
Universidad Autónoma de Madrid

CAPÍTULO II

Tratamiento básico de Series Estadísticas. Introducción al Procedimiento de Regresión.

II.A. - OBJETIVOS Y CONTENIDOS DE ESTE SEGUNDO CAPÍTULO

El objetivo de este segundo capítulo consiste en abordar progresivamente algunos de los procedimientos de transformación de series de datos y análisis estadístico más sencillos, toda vez que se suponen aprendidos los conceptos mínimos necesarios para “arrancar” con el programa E-VIEWS.

Dentro de las múltiples operaciones de “manipulación” de datos que pueden realizarse cómodamente en E-VIEWS, vamos a prestar especial atención en este capítulo a aquellas que permiten especificar y estimar correctamente un Modelo Básico de Regresión Lineal.

Con total seguridad, el analista que se enfrenta a la etapa de estimación de un modelo básico de regresión lineal deberá “cocinar” las series de datos elegidas y recolectadas en bruto de las fuentes estadísticas, durante la anterior etapa de especificación. Así, probablemente el analista se verá en la obligación de generar alguna magnitud monetaria “en constantes” a partir de deflatores y datos “en corrientes”, deberá calcular las tasas de variación a partir de series “en niveles” o encontrará útil un examen gráfico preliminar de la evolución temporal de las series consideradas. Por tanto, difícil será evitar un análisis previo que incluirá, la mayor parte de las ocasiones, sencillos procedimientos de análisis gráfico de las series, su transformación en tasas, logaritmos, etc, o, por ejemplo, un simple análisis previo de correlaciones entre variables explicativas y endógena:

- ◆ El análisis gráfico permite estudiar visualmente la evolución de las variables consideradas en el período histórico detectando comportamientos cíclicos, tendenciales, estacionales o, en muchas ocasiones, puntos atípicos (puntos de valor extraordinariamente grande o pequeño con relación al valor promedio de la serie).
- ◆ La transformación de variables será necesaria en muchas ocasiones ya que pocas veces la especificación seleccionada para nuestro modelo nos permitirá utilizar los datos en su estado natural.
- ◆ El análisis previo de correlaciones puede servir, por una parte, como tosca aproximación previa a la posible existencia de relación entre la variable a estudiar y las explicativas elegidas y, por otra, para la detección prematura de posibles problemas de multicolinealidad, es decir, de excesiva y perjudicial relación entre las variables exógenas incluidas en la especificación (*más adelante se detallará el significado de este concepto y los procedimientos necesarios para su detección*).

Estas y otras transformaciones similares que caracterizan todo proceso previo a la estimación de los parámetros de un MBRL no podrán realizarse a menos que se tengan una serie de ideas claras en cuanto a la forma que debe utilizarse E-VIEWS para operar con una o varias series.

A recordar....

- El capítulo abordará los procedimientos de análisis simple y transformaciones básicas de series necesarios para “poner en marcha” la correcta especificación y estimación de un MBRL. Para su correcto aprovechamiento es imprescindible haber asimilado el manejo básico del programa comentado en el capítulo anterior.

II.B. - GRÁFICOS DE SERIES EN E-VIEWS

En E-VIEWS resulta extremadamente sencillo obtener información básica preliminar de una serie así transformar una serie original según distintos procedimientos. Debe recordarse que, en términos de E-Views, **una serie es un OBJETO**. Por tanto, para obtener información sobre una serie de datos o realizar alguna transformación o procedimiento sencillo sobre la misma, parece lógico editar (*abrir*) el objeto a analizar (*la serie*) haciendo *doble clic* sobre el nombre de la misma en el Work File.

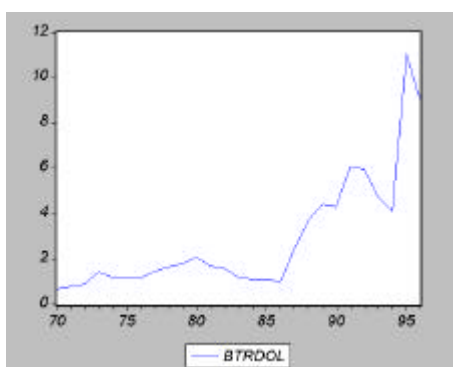
Una vez editada una serie, el menú de objeto que se despliega en su parte superior permite realizar toda clase de operaciones sencillas.

II.B.1.- Cómo representar gráficos de series individuales

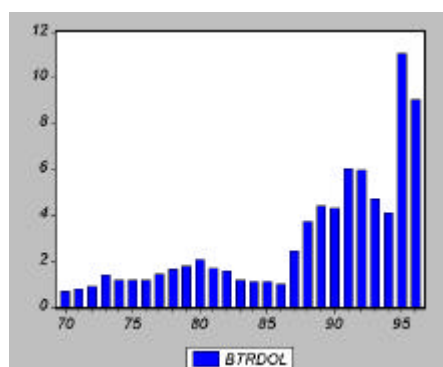
El menú (*views*) de la serie editada, permite observar la serie en distintos formatos: tipo hoja de cálculo, representaciones gráficas diversas ,etc. Los gráficos de EVIEWS disponibles para representar series individuales son, en principio, de dos tipos: de línea o de barras (aunque veremos más adelante que. En la ilustración siguiente se muestran ambos para una determinada serie:

Diferentes vistas (Views) para una misma serie de datos

Formato de LÍNEA



Formato de BARRAS



Una vez seleccionado uno de los tipos de gráfico para una determinada serie, **puede cambiarse de un tipo a otro** pulsando, en la ventana activa del gráfico, los botones *Bar/Line* de la barra de tareas.

Los gráficos, una vez en pantalla, no pueden salvarse como objetos ya que, en realidad ya lo son: se trata de series en vista de gráfico. Existe, no obstante, la posibilidad de guardarlos como un "**objeto congelado**". La utilidad de esta operación es la de conservar una "fotografía del gráfico", de modo que este gráfico pueda conservarse invariable, independientemente de que la serie representada cambie alguno de sus valores posteriormente. Esta operación se realiza con el **comando "freeze"**, especificando después un nombre para el nuevo objeto (congelado) creado. El comando freeze puede, en general, aplicarse del mismo modo a cualquier otro objeto abierto.

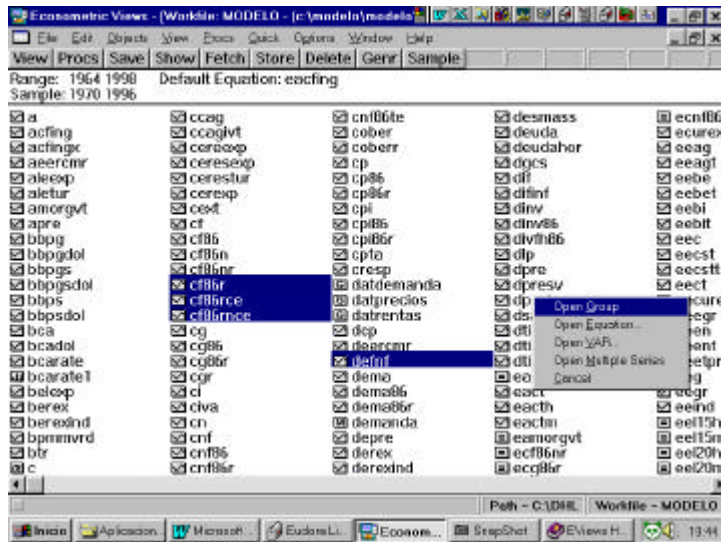
II.B.2.- Cómo representar gráficos de varias series

Evidentemente, el EVIEWS permite representar simultáneamente más de una serie. Suele ser habitual, en efecto, representar conjuntamente dos o más series en un mismo gráfico, casi siempre con el objeto de comparar lo similar o diferente de su evolución temporal. Así, por ejemplo, y de modo sólo aproximativo, la evolución comprada de la variable endógena y las explicativas puede ser suficientemente expresiva como para revelar la causalidad entre unas variables y otras, anticipando resultados en el análisis previo de un procedimiento de regresión.

Para representar dos o más series deben seleccionarse primero las series de interés en la ventana del *workfile*. Para ello, deben marcarse las series, una a una, manteniendo pulsada la tecla **CTRL** y, una vez

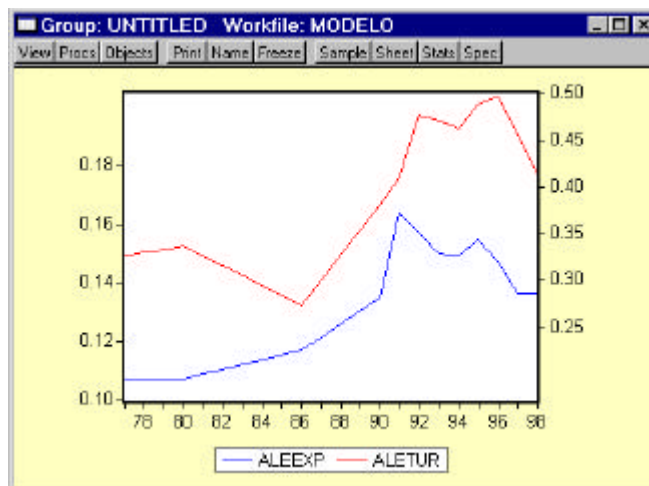
marcadas, abrir el menú contextual del ratón (*botón derecho*). En ese menú contextual encontraremos la a opción *Open Group*.

4 Series marcadas simultáneamente dispuestas a ser abiertas con la opción Open Group



Una vez abierto el grupo de series, usaremos de nuevo la opción **View – Graph** de la barra de herramientas de la ventana del grupo activa, consiguiéndose así un gráfico simultáneo en el que cada serie aparecerá representada por un color. La opción **Multiple Graph** del mismo menú, permitirá ver en una misma pantalla, los gráficos separados de las series seleccionadas

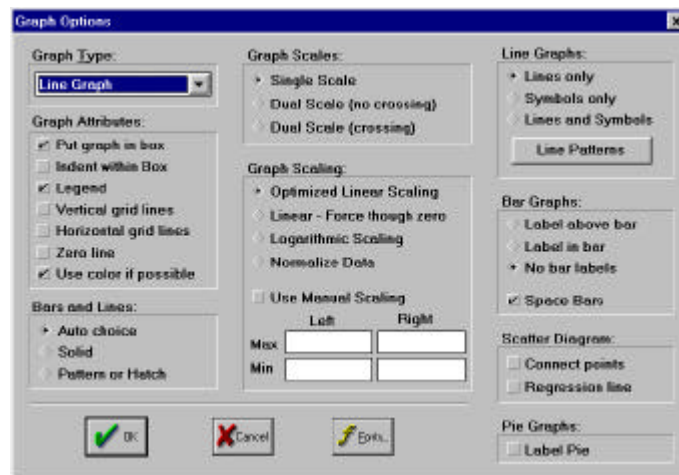
2 Series representadas simultáneamente



II.B.3.- Opciones de gráficos

Existen muchas opciones que permiten modificar el aspecto de cada gráfico de modo individual. Para activarlas, basta con hacer *doble clic* en cualquier parte del área del gráfico que se está visualizando en un momento dado.

Menú de opciones de gráfico



La modificación más habitual, consiste, por ejemplo, en cambiar el tipo de gráfico representado. Las opciones son:

- Line graph* - Gráfico de **línea**: en el que cada punto se une al siguiente con una línea.
- Stacked Line* – **Línea acumulada**: en el que cada línea representa los valores de una serie acumulados sucesivamente (*el último punto representado es la suma de todos los valores de la serie*).
- Bar Graph*- Gráfico de **barras**: en el que cada valor de la serie se representa con una barra vertical
- Stacked Bar* – **Barras acumuladas**: en el que cada barra representa los valores de una serie acumulados sucesivamente (*la última barra representa la suma de todos los valores de la serie*).
- Mixed Bar & Line* – Gráfico **mixto de barras y líneas**: en el que la primera serie se representa como una barra y las siguientes como líneas.
- Scatter Diagram* – Gráfico **de Dispersión**: en el que el eje horizontal del tiempo es sustituido por el de la primera serie, de forma que cada punto del gráfico representa una coordenada formada por los valores de las dos series representadas. Este tipo de gráfico resulta de gran utilidad cuando quiere representarse la clásica "nube de puntos de regresión" que aproxima la eventual relación entre dos variables para las que se supone una determinada relación.
- Pie Graph* – Gráfico de **Tarta**: en el que cada observación es representada como una tarta, siendo los sectores de cada una de esas tartas los valores de cada serie para esa observación. El valor de cada serie se representa, por tanto, como la aportación proporcional de esa serie al valor suma de todas ellas.

Además del tipo de gráfico, **pueden cambiarse algunos atributos de los gráficos** en la columna izquierda del cuadro de diálogo. Cada atributo puede ser activado o desactivado marcando la casilla correspondiente con el ratón.

- Puede quitarse el marco gris (*box*) que rodea el gráfico y poner el mismo dentro de ella, puede colocarse una leyenda del nombre de cada serie representada, colocar línea verticales u horizontales, representar la vertical del valor cero o usar el color, cuando las opciones gráficas lo permitan, para distinguir una serie de otra..

- Así mismo, puede **cambiarse la apariencia de las línea o barras** en un pequeño cuadro de diálogo que aparece en la parte inferior.

En la parte central del menú de opciones se agrupan las siguientes opciones:

- En ocasiones interesará **un segundo eje de ordenadas** para representar dos o más series. Si se elige la opción de doble eje, el EViews representará la primera de las series en el primero de los ejes y el resto en el segundo eje. Esta opción puede resultar interesante en el caso de que las series vengan representadas por valores medidas en escalas muy diferentes.
- Dentro de la opción de doble eje, ambos ejes pueden configurarse de forma que las dos series representadas no se crucen, o que sí lo hagan.
- Las opciones de **escala del gráfico** permiten alterar ciertos parámetros del eje izquierdo como forzar a que aparezca el valor cero en el eje, transformar los valores del eje a una escala logarítmica o normalizar los datos de las series representadas (*opción muy útil para representar dos series con medidas muy diferentes en un solo eje vertical*).
- Además podemos especificar manualmente el menor y el mayor valor de cada eje (*incluso en un segundo eje de ordenadas*).

Una serie de opciones sencillas controlan, en la parte derecha de este cuadro de diálogo, la forma en que aparecerán representadas líneas, barras, gráficos de dispersión o secciones del gráfico de tarta:

- En las **opciones de línea del gráfico** puede añadirse, algún atributo especial o símbolo entre observaciones, a la línea con que se representa cada serie.
- En las **opciones de barra del gráfico** puede añadirse la leyenda del valor de la serie sobre o dentro de cada barra representada, así como eliminarse el espacio que separa por defecto unas barras de otras.
- En las **opciones para el diagrama de dispersión** puede unirse cada punto del diagrama con una línea o representarse sobre la nube de puntos del diagrama la línea de regresión hipotética. Esta línea, vendría a ilustrar la recta de regresión lineal estimada por Mínimos Cuadrados Ordinarios de la variable representada en el eje de ordenadas en función de la variable representada en el eje de abscisas.
- En las **opciones disponibles para el gráfico de tarta** puede añadirse, cerca de cada sector del gráfico, una leyenda conteniendo el valor de la serie representado.

A recordar....

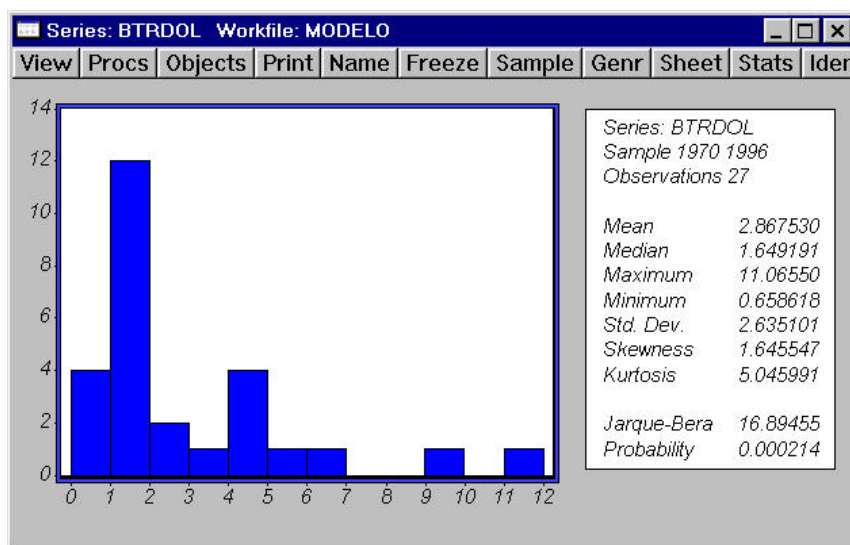
- El análisis gráfico de las series involucradas en un análisis cuantitativo es siempre recomendable como etapa de "toma de contacto" con la información. Este análisis permite observar "a vista de pájaro" el comportamiento global de las series así como algunas peculiaridades puntuales que escaparían a simple vista al observar los datos de las mismas.
- El E-Views no es un programa de enorme potencial para la representación gráfica pero, al menos, permite la representación gráfica de una o más series en gráficos individuales o simultáneos.
- Las opciones que, por defecto, definen el aspecto de un gráfico en E-Views pueden alterarse con sencillez en una única pantalla de opciones de gráfico.

II.C.- OBTENCIÓN DE ESTADÍSTICOS BÁSICOS DE UNA SERIE

II.C.1.- Estadísticos básicos de series individuales

Obtener los estadísticos descriptivos básicos de cualquier serie de datos es muy simple en E-Views. Así, una de las opciones de vista (*view*) de cualquier serie individual es, precisamente, la de histograma y estadísticos (*Histogram & Stats*). Si ejecutamos esta opción podremos observar en una misma pantalla el histograma de frecuencias de la serie y algunos estadísticos básicos:

Histograma de frecuencias y estadísticos básicos de una serie (*BTRDOL en el ejemplo*)



Como es bien sabido, el **histograma de frecuencias** representa gráficamente la distribución de frecuencias de los valores de la serie. Esta representación tiene en general poco sentido para series temporales en las que, habitualmente, no suelen repetirse valores en distintos períodos a lo largo de la muestra considerada. Por ello, el EVIEWS genera unos intervalos de igual extensión y representa con barras el número de ocasiones en las que la serie toma valores comprendidos en cada uno de estos intervalos. Este histograma "por intervalos" admite las modificaciones comentadas en el apartado general referidas a las opciones de gráfico exceptuando que sólo puede representarse en "tipo barras".

Los estadísticos descriptivos ofrecidos a la derecha del histograma son conceptualmente muy sencillos y fáciles de interpretar. Aparecen en primer lugar dos **medidas de tendencia central de la serie**:

- La **media (mean)** de la serie, calculada como promedio aritmético.
- La **mediana (median)** de la serie; es decir, aquel valor que separa los valores de la serie en dos conjuntos de igual densidad de frecuencias. En el caso de una serie temporal representada por intervalos, se toma como mediana la marca de clase del intervalo mediano, convenientemente desplazada según la mayor o menor carga a izquierdas y a derechas del mismo.

A continuación se muestran dos **aproximaciones a la dispersión de la serie** respecto a sus valores centrales:

- El valor **máximo y mínimo** de la serie
- La **desviación típica (Std.Dev)** de la serie (*raíz de la varianza de la serie*)

Más adelante se muestran algunos **cálculos que ayudan a valorar la normalidad estadística de la serie**:

- La **simetría de la serie (Skewness)** con respecto a su media es un cálculo sencillo (*que utiliza el momento de orden tres con respecto a la media así como el cubo de la desviación típica de la serie*). Esta expresión toma el valor cero en el caso de una distribución perfectamente simétrica (*con igual densidad de frecuencias a izquierda y derecha de la media*). Valores positivos indican asimetría a derechas respecto a la normal y valores negativos asimetría a izquierdas.

$$S = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n}}{s^3}$$

- La **curtosis (Kurtosis) de una serie** indica si su distribución de frecuencias es más aplanada o más apuntada que una distribución normal, es decir, si alrededor de la media se concentran más o menos valores que en el caso de una normal y por tanto sus "colas" son más o menos estrechas. La curtosis de una serie puede calcularse utilizando el momento de orden cuatro con respecto a la media y la potencia cuarta de la desviación típica de la serie. Para una distribución normal el valor de este coeficiente de curtosis es 3 (*distribución mesocúrtica*). Valores superiores a 3 indican un apuntamiento mayor que el de una distribución normal (*distribución leptocúrtica*) y valores inferiores a 3, un apuntamiento menor (*distribución platocúrtica*).

$$K = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^4}{n}}{s^4}$$

- A partir de ambas medidas, simetría y curtosis, puede elaborarse un **contraste paramétrico de normalidad de la serie que se denomina contraste de Jarque Bera**. La idea del cálculo se apoya en comparar simultáneamente los valores obtenidos para los coeficientes de simetría y curtosis con los de referencia para una normal. La ventaja del cálculo es que su resultado permite contrastar la hipótesis nula de que la serie se distribuye como una Normal ya que esta expresión (*JB*) se distribuye como una χ^2 con dos grados de libertad.

$$JB = \frac{n - k}{6} \left[S^2 + \frac{1}{4} (K - 3)^2 \right]$$

El término "k" minúscula se refiere al número de variables exógenas utilizadas, cuando este test se usa para contrastar la normalidad de los residuos de una determinada regresión y toma el valor "0" cuando contrastamos la normalidad de una serie individual.

El valor de la probabilidad ofrecido por Eviews, se entiende como el nivel de significación asociado al rechazo de la nula: valores pequeños para esa probabilidad (*inferiores a 0,05 ó 0,1*) indicarían, por tanto, ausencia de normalidad en la distribución de valores de la variable analizada.

Los estadísticos de normalidad de una serie son comúnmente utilizados en el contexto de la validación de un modelo básico de regresión, ya que la normalidad de las variables (*en especial de la perturbación aleatoria*) es una de las hipótesis básicas sobre la que se apoyan las buenas propiedades de los procedimientos de estimación e inferencia.

Una vez más debe recordarse que esta salida, histograma y estadísticos, pueden ser almacenados de modo independiente de la serie como un objeto congelado utilizando el comando *freeze*.

I.C.2.- Estadísticos para grupos de series

Además de la información suministrada individualmente para cada serie, podemos solicitar información que afecte a dos o más series simultáneamente. Para ello, debemos primero crear el grupo de variables a analizar, o bien simplemente marcar las distintas series y abrirlas simultáneamente, tal y como se describió en el anterior apartado II.B.2.

Una vez abierto el grupo, tenemos cuatro opciones básicas de vista (*view*) que ofrecen información estadística:

- Descriptive Stats, o sea estadísticos descriptivos** para cada una de las series del grupo. No se incluye el histograma de frecuencias como en el caso del análisis individual. El EVIEWS permite obtener los estadísticos de las series sólo para aquellas observaciones en las que todas las series tengan valores (*Common Sample*) o aprovechando individualmente el número de observaciones que tenga cada una de ellas (*Individual Sample*). Con cualquiera de las opciones aparecen los mismos estadísticos comentados en el anterior apartado para el caso del análisis individual.
- Test de igualdad (Test of Equality)**
- Análisis de tabulación cruzada (N-Way Tabulation)**
- Análisis simple de correlación (Correlations)**
- Análisis de covarianzas (Covariances)**
- Test de Causalidad de Granger (Granger Casuality Test)**

Estas últimas posibilidades se comentan con más detalle en el apartado siguiente.

A recordar....

- El análisis preliminar de las series a utilizar en cualquier ejercicio cuantitativo puede completarse en E-Views con el cálculo de sus estadísticos básicos de tendencia, dispersión y normalidad.
- El cálculo de estadísticos básicos puede realizarse para una serie individual o para un grupo de series.

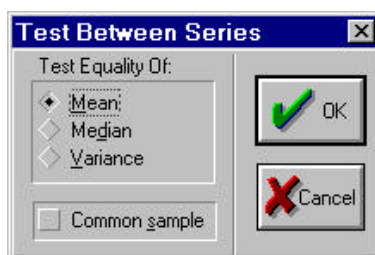
II.D.- ALGUNAS OPERACIONES PREVIAS CON SERIES DE DATOS DE CARA A LA REALIZACIÓN DE UN ANÁLISIS DE REGRESIÓN

Las opciones enunciadas en la parte final del apartado anterior tienen una clara utilidad. Todas ellas inciden en el conocimiento de la fortaleza y calidad de la relación existente entre dos o más variables por lo que resultan de gran ayuda de cara a ir recogiendo información relevante para un análisis de regresión posterior entre las mismas.

- Test de igualdad (Test of Equality)**. Esta clase de tests, muy habituales en estadística, nos permiten contrastar la igualdad en media, mediana o varianza de un grupo determinado de series respondiendo a preguntas del tipo: ¿tendrán hombres y mujeres el mismo salario medio?, ¿presentarán el PIB español y alemán una media de crecimiento similar en el periodo analizado?, ¿presentarán el IBEX – 35 y el DOW JONES una dispersión similar a lo largo del periodo de análisis?.

Para realizar un contraste de igualdad de medias, medianas o varianzas, basta con marcar en un mismo grupo las series de interés y, una vez abierto el grupo, solicitar el análisis en la opción *View – Test of Equality*.

Menú del Test de Igualdad



- ◆ Para testar **la igualdad en media** E-Views aplica un contraste F-ANOVA muy simple. Este test, se basa en comparar primero las medias particulares de cada variable con la media global para todas las series generando así la llamada Variación Entre - Grupal (*between*). Una vez computada la variación Entre - Grupal, se genera la variación Intra - Grupal (*within*) comparando los valores de cada serie con la media de esta. En la medida en que las medias de las series son iguales, la variación entregupal tiende a ser mínima en tanto que la intragrupal crece. El test *F* presentado por E-Views contrasta la H_0 de igualdad en medias por lo que, valores elevados del test (o pequeños del nivel de significación) invitan al rechazo de la nula.

Resultado para un análisis de igualdad de medias para el Crecimiento de la Población Activa Masculina y Femenina 70- 00:

Los valores de los estadísticos muestran cómo en todos los casos No puede rechazarse la nula de igualdad de medias

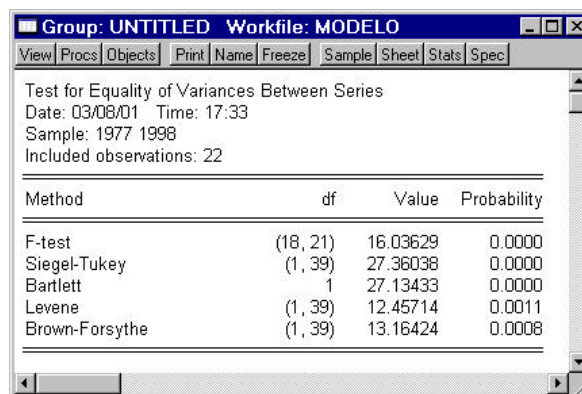
Method	df	Value	Probability
t-test	42	1.024630	0.3114
Anova F-statistic	(1, 42)	1.049866	0.3114

- ◆ Para testar **la igualdad en mediana** (*algo menos interesante*), E-Views ofrece varios tests no paramétricos sobre la H_0 de que todas las series consideradas presenten igual mediana frente a la alternativa de que al menos una de ellas tenga mediana diferente. Dada la escasa utilidad de un contraste de mediana en nuestro contexto específico, enumeraremos simplemente los métodos de análisis: Mann-Whitney U-test y su aproximación normal a una distribución U-Statistic, Kruskal-Wallis one-way ANOVA por rangos y su aproximación normal a una χ^2 (*generalización del Mann-Whitney test para más de dos series*), Test Van der Waerden (*análogo a Kruskal-Wallis test aunque levemente mejorado contra la presencia de puntos atípicos*), Test ANOVA χ^2 para la mediana (*también conocido como Test Mediana se basa en la comparación del número de observaciones por encima y debajo de la mediana y se presenta con corrección de continuidad de Yates y sin ella*)
- ◆ Para testar **la igualdad en varianza**, E-Views propone algunos tests de entre los más conocidos. El **test F** sólo se muestra para el caso de dos series y se basa en computar el ratio entre la varianza mayor de las dos (L) y la menor (S), ratio que se distribuye como una F con n_L-1, n_S-1 grados de libertad para la nula de igualdad de varianza. El **test Siegel-Tukey** también se muestra sólo para el caso de dos series y se computa según el procedimiento Mann-Whitney comentado para testar la igualdad en mediana aunque con diferente asignación de rangos. El **test de Bartlett** (*y su ajuste por ausencia de normalidad*) compara el logaritmo de la varianza media ponderada con la suma ponderada de logaritmos de las

varianzas; bajo la hipótesis nula de igualdad de varianzas el test es aproximadamente una χ^2 . El **test de Levene** se basa en un análisis ANOVA sobre las diferencias de las medias en valor absoluto. El **test Brown-Forsythe** (o *Levene modificado*) es una modificación del test de Levene en el que se reemplaza el valor absoluto de la diferencia de medias con el de las diferencias de las medianas.

Resultado para un análisis de igualdad de varianzas:

Los valores de los estadísticos muestran cómo en todos los casos debe rechazarse la nula de igualdad de varianzas



Method	df	Value	Probability
F-test	(18, 21)	16.03629	0.0000
Siegel-Tukey	(1, 39)	27.36038	0.0000
Bartlett	1	27.13433	0.0000
Levene	(1, 39)	12.45714	0.0011
Brown-Forsythe	(1, 39)	13.16424	0.0008

A recordar....

- El E-Views permite realizar tests de igualdad de medias, medianas y varianzas para un conjunto de series determinado. Estos test resultarán de utilidad al econométra para asegurar el cumplimiento de ciertas hipótesis de partida sobre homogeneidad en media y/o varianza de las series involucradas en el análisis.

Análisis de tabulación cruzada (N-Way Tabulation): Este análisis, clásico entre las técnicas multivariantes más sencillas, sirve:

- ◆ Para organizar información relativa a dos factores considerados conjuntamente; un factor es una variable cualitativa, es decir, discreta, medida por tanto generalmente en escala nominal.
- ◆ Por otro lado, la técnica permite además contrastar la dependencia o independencia entre los factores considerados.

Podrá intuirse que este procedimiento no tiene especial sentido en el caso de disponer de variables de tipo continuo, es decir, medidas en escala de intervalo o razón como las que generalmente manejaremos en el análisis de regresión. La razón es que, en esos casos, existen técnicas más perfeccionadas para la exploración conjunta de la relación existente entre ambas variables. Por esta razón, en caso de solicitar al E-VIEWS un análisis *N-Way* para variables continuas, éste dividirá los valores de las variables en intervalos a fin de "transformarlas" en variables categóricas.

Por la razón comentada en el párrafo anterior, no se explicará en este texto el detalle del análisis *N-Way* realizado por E-Views. Baste decir, básicamente, que E-Views presenta en los resultados:

- ◆ una tabla de contingencia de doble entrada que muestra la distribución conjunta de los valores de cada par de variables analizadas

- ◆ dos tests (*Pearson χ^2* y *Likelihood Ratio G^2*) para el contraste de la H_0 de independencia entre las variables seleccionadas.
 - ◆ tres medidas de asociación es decir, de intensidad de la eventual dependencia entre las variables (*Coefficiente Phi*, *V de Cramer* y *Coefficiente de Contingencia*)
- **Análisis simple de correlación (*Correlations*).** El E-Views permite observar el coeficiente de correlación simples para cada par de variables seleccionadas en el grupo. Este coeficiente varía entre -1 (correlación perfecta y negativa) y $+1$ (correlación perfecta y positiva). Un análisis preliminar de estos coeficientes resulta interesante para determinar tres cuestiones previas a un análisis de regresión multivariante:
- ◆ Los coeficientes de correlación simple entre endógena y cada una de las exógenas permiten anticipar si la variable endógena guarda una relación, a priori, con las variables que van a utilizarse en el análisis de regresión. Si las correlaciones son altas, esto supone una cierta garantía de éxito de la posterior regresión múltiple en la medida en que toda relación de causalidad entre dos variables suele manifestarse con elevadas correlaciones entre ellas. No obstante, debe señalarse (1) que, si bien un bajo coeficiente de correlación suele indicar "ausencia de causalidad" un coeficiente de correlación elevada NO implica causalidad segura y (2) que, incluso en el caso de ausencia de correlación individual entre una exógena y la endógena, no debemos rechazar esa variable como explicativa, ya que la naturaleza multivariante de la regresión considera la relación múltiple entre el conjunto de variables explicativas y la variable endógena, una relación múltiple que, en ciertos casos, puede ser elevada aún cuando las correlaciones individuales no lo sean.
 - ◆ Los coeficientes de correlación simple entre endógena y cada una de las exógenas permiten anticipar si los signos de la relación simple entre la variable endógena y cada una de las explicativas es, a priori acorde con lo esperado desde el punto de vista conceptual. En cierto modo, un signo incorrecto en la correlación entre una variable explicativa y la endógena debe representar una llamada de alerta. Sin embargo, conviene recordar, una vez más, que lo que interesa a la regresión múltiple es el signo de cada variable en el contexto de la causalidad múltiple y ese signo puede no coincidir, para cada variable, con el signo exhibido en la relación individual.
 - ◆ Los coeficientes de correlación simple entre endógena y las exógenas consideradas, permiten detectar problemas posteriores de multicolinealidad en la estimación de la ecuación.
- **Análisis de covarianzas (*Covariances*)** *Covariances* o *covarianzas*, entre cada par de variables del grupo. La covarianza es un concepto similar a la correlación pero afectada por las unidades de medida de las variables consideradas. (*a mayores unidades de medición de las variables consideradas, mayor valor de la covarianza*). Sin embargo, en múltiples ocasiones el dato de la covarianza entre series es necesario para desarrollar cálculos de interés en el contexto de un análisis cuantitativo.
- **Test de causalidad de Granger (*Granger Causality Test*).** El test de causalidad de Granger es ampliamente conocido en la práctica econométrica ya que permite identificar de forma rápida relaciones de causalidad entre las variables explicativas y la variable a explicar. El objetivo teórico de este test es determinar si una variable X causa a otra variable Y. El procedimiento que se utiliza es sencillo, especificada la variable X y la variable Y se realiza la regresión de la variable endógena Y_t sobre su propio pasado, es decir, Y_{t-1} , Y_{t-2} , Y_{t-3} , sobre la variable X_t y una serie de valores retrasados de la misma, es decir, X_{t-1} , X_{t-2} , X_{t-3} , etc. Una vez realizada esta la regresión, se determina si resulta más fácil predecir el futuro de la variable Y con este instrumento de lo que resultaría estimado Y_t exclusivamente en función de su pasado sin conocer su relación con X: dicho de otro modo, se analiza si la variable X actual y pasada aporta información valiosa para explicar el futuro de Y (se dice, en ese caso que X es causa Granger de Y)

Para realizar este test en E-Views, debe seleccionarse la opción *Granger Causality* del menú Views de la ventana del grupo activo. EViews nos preguntará el número de retardos de la endógena y

¹ Como se comentará más adelante no es tan fácil asimilar el concepto de causalidad Granger a la existencia real de dependencia entre las variables.

exógena que queremos incluir en la regresión. A este respecto, la elección debe realizarse atendiendo a criterios puramente conceptuales procurando, eso sí, no quedarse demasiado corto, ya que, siempre que se cuente con un número elevado de observaciones, el test de Causalidad de Granger es más “defendible” cuanto mayor se el número elevado de retardos incorporados.

Una vez seleccionado el número de retardos a considerar, E- Views realiza automáticamente las distintas regresiones que completan la idea del test expuesta anteriormente y muestra el resultado. Debe observarse que E- Views muestra el test de causalidad en ambas direcciones *X como causa Granger de Y* e *Y como causa Granger de X*

La hipótesis que se contrasta es que los coeficientes de las regresiones de Y sobre X así como los de X sobre Y son nulos para la variable de apoyo, es decir, que la variable X no aporta información para explicar a Y o bien que Y no aporta información para explicar X.

Si el valor del estadístico de referencia “F” supera el valor tabulado se rechazará la hipótesis nula y por tanto se aceptará que X causa a Y o viceversa. En la ilustración que aparece en este ejemplo, X causaría a Y según el concepto de causalidad de Granger aunque la relación inversa no se produciría.

Resultado para un análisis de Causalidad de Granger:

El valor de los estadísticos muestran que X es causa Granger de Y (al menos al 9,6% de nivel de significación) mientras que y no es causa Granger de X

Null Hypothesis:	Obs	F-Statistic	Probability
VARX does not Granger Cause VARY	22	2.69269	0.09641
VARY does not Granger Cause VARX		0.39667	0.57862

El problema de este test, es que la correlación no implica necesariamente causalidad en el sentido en el que se conoce generalmente. Son comunes los ejemplos de altas correlaciones de carácter espurio o “casual” pero no “causal”. El test de causalidad de Granger utiliza un concepto ampliado de correlaciones para encontrar causalidades. Por tanto, a pesar de un resultado positivo del test de Granger, nunca debe concluirse que si X causa a Y la variable Y sea el efecto de la variable X.

A recordar....

- La exploración de las relaciones entre las variables implicadas en un análisis cuantitativo es de una indudable utilidad: permite aproximar el signo y la fortaleza de las potenciales relaciones de causalidad y anticipa problemas de multicolinealidad entre exógenas.
- El E-Views incorpora algunos procedimientos encaminados al análisis de relaciones entre variables: N-Way Tests, Correlaciones, Covarianzas y Test de Causalidad de Granger. El conocimiento de estos procedimientos nos ayudará a caracterizar la naturaleza del conjunto de variables utilizadas en un procedimiento de regresión facilitando su utilización posterior.
- El manejo solvente de cada uno de los procedimientos antes mencionados pasa necesariamente por el conocimiento de su formulación y propiedades estadísticas. No obstante, el E-Views permite, de manera sencilla, la utilización práctica aproximada de todos ellos: basta, en realidad, con conocer la hipótesis nula utilizada en cada contraste para aplicar las nociones mínimas de contraste de hipótesis y obtener conclusiones preliminares.

II.E.- INTRODUCCIÓN A LA ESTIMACIÓN DE ECUACIONES CON EIEWS

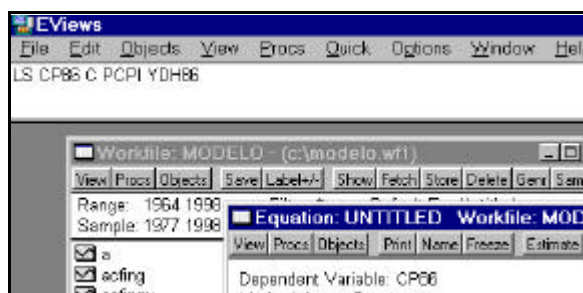
Sin perjuicio de que, en el siguiente capítulo, se analicen con detalle los principales procedimientos que conlleva un análisis de regresión completo, vamos a repasar ahora el procedimiento básico para la creación de una ecuación y algunos de los resultados más inmediatos.

II.E.1.- Creación de un “objeto” ecuación

Una ecuación es, al igual que una serie o un gráfico, un objeto de EIEWS. Una vez creado por vez primera tendrá, como el resto de objetos, varios formatos de vistas y podrá ser sometida, como el resto de objetos, a una serie de procedimientos particulares. Una ecuación se crea la primera vez que se estima, una vez creada. Para crear por primera vez una ecuación, o sea, para estimar una ecuación con EIEWS existen, en principio, varios procedimientos operativos alternativos:

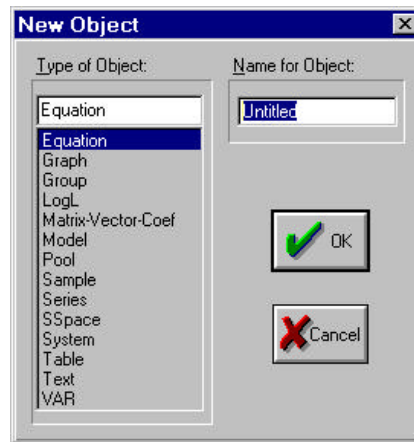
- El primero de ellos consiste en **crear un grupo con las variables**, exógenas y endógena, involucradas en la regresión. Una vez creado el grupo, la regresión será uno de los procedimientos básicos accesibles en la barra de herramientas del grupo.
- El segundo consiste en **utilizar, en la pantalla de comandos, la sintaxis de la vieja versión** de E-Views (Micro TSP) para generar la ecuación de regresión. Por ejemplo, para ordenar una regresión del Consumo Privado en pesetas constantes del año 86 (CP86) en función de un término independiente “C”, el Deflactor del Consumo Privado (PCPI) y la Renta Disponible de los Hogares en pesetas constantes (YDH86) la sintaxis de Micro TSP sería:

Sección de la ventana de sintaxis de comandos en E-Views



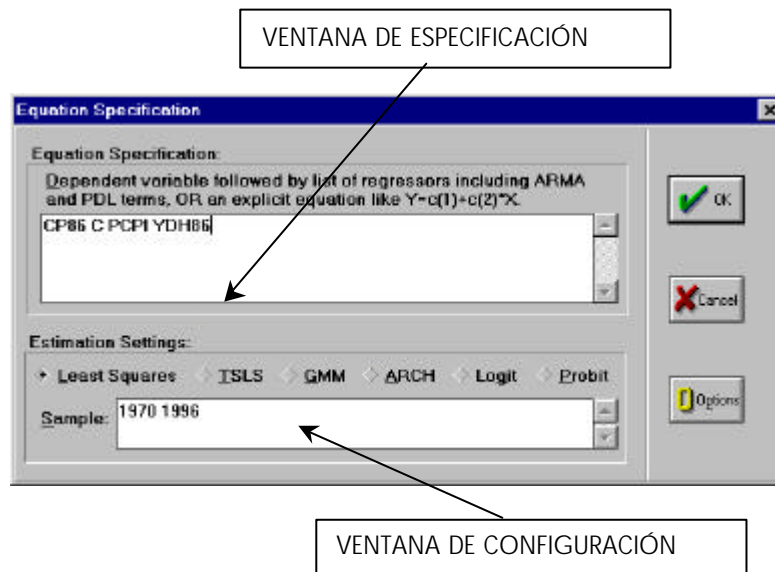
- El tercero, y más habitual, es **crear un objeto nuevo** utilizando los comandos *Object/New Object/Equation*. Una vez ejecutado este comando, el E-Views nos solicita un nombre para el nuevo objeto (*para la ecuación*) que en nuestro ejemplo puede ser QCP86.

Ventana para la creación de un nuevo objeto



II.E.2.- Ventana de estimación de ecuaciones

Una vez que hemos dado nombre a la nueva ecuación, accederemos a un cuadro de diálogo en el que el programa EViews ofrece diversos parámetros relacionados con la estimación de ecuaciones. Ese cuadro de diálogo es el que se muestra a continuación:

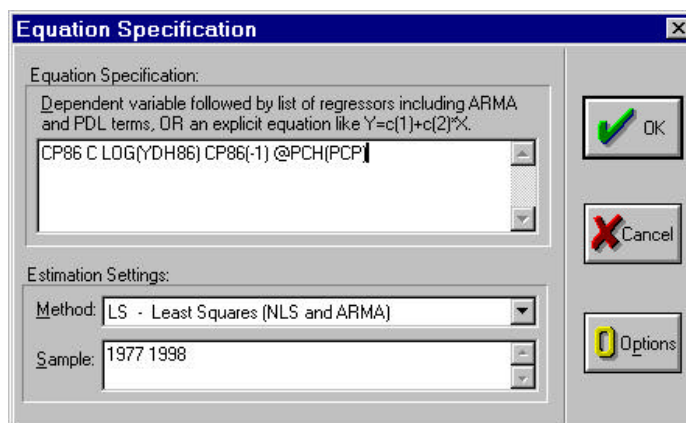


En la primera ventana, **la ventana de especificación de la ecuación**, es donde debe especificarse el conjunto de variables explicativas a utilizar. A este respecto debemos señalar los siguientes aspectos:

- ◆ El orden en el que deben introducirse las variables es: en primer lugar el nombre de la endógena seguido de la lista de explicativas separadas, unas de otras, por un espacio en blanco.
- ◆ El **término independiente** se especifica con la letra "C" por tanto no puede asignarse este nombre a ninguna otra variable
- ◆ **Pueden incluirse fórmulas** u operadores sencillos en la ventana de especificación. Esto permite estimar las ecuaciones con transformaciones sencillas de la variables (como tasas o logaritmos) sin tener que generar las series independientemente como objetos nuevos. Por ejemplo, en la siguiente

ilustración, la variable CP86 se explica en función del logaritmo de la variable YDH86, de la propia endógena de consumo privado retardada un periodo CP86(-1) y la tasa de variación porcentual (*Percental Change*) del deflactor del consumo (@PCH(PCP)).

Estimación de una ecuación utilizando transformaciones de las variables



- ◆ Por último, debe observarse que esta ventana puede invocarse para una ecuación ya creada en cualquier momento pulsando el botón *Estimate* una vez abierta la ecuación.

En la ventana inferior del cuadro de diálogo, la **ventana de configuración**, se especifica el método de estimación de los parámetros de la ecuación, así como la muestra de datos a considerar en esta estimación. Los métodos alternativos de estimación que se ofrecen son:

- ◆ **Least Squares** ó Mínimos Cuadrados ordinarios
- ◆ **TOLS** ó Mínimos Cuadrados en dos etapas
- ◆ **GMM** ó Método generalizado de momento
- ◆ **ARCH** ó Modelos autorregresivos para la modelización de la varianza (*GARCH, TGARCH, EGARCH, C-ARCH*)
- ◆ **Modelos de elección discreta** (*Lógit, Probit o de Valor Extremo*)
- ◆ **Modelos censurados** (*Censored data – Tobit*)
- ◆ **Modelos para variables discretas** (*Integer count dependent variable*)

Salvo que se seleccione el método de Mínimos Cuadrados Ordinarios, cada uno de los otros métodos de estimación requerirá completar una segunda ventana de diálogo en la que deberá especificarse con detalle alguna información necesaria para ejecutar una estimación de este tipo. Así, por ejemplo, el método TOLS obliga a la selección de las variables que serán utilizadas como instrumentos o el método ARCH permite seleccionar distintos tipos de especificación ARCH y distintos regresores para la varianza,

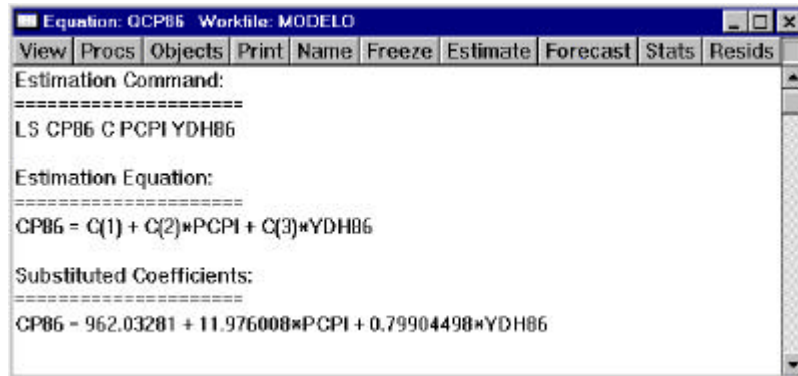
Así mismo, el **botón de opciones** que aparece en todas las pantallas de estimación abre siempre un nuevo cuadro de diálogo para especificar aún más algunos parámetros de estimación (*por ejemplo, el tratamiento de una eventual heterocedasticidad o el procedimiento de iteración que a seguirse en cada caso cuando el procedimiento de estimación lo requiera*).

II.E.3.- Vistas de una ecuación

Una vez que la ecuación se ha estimado, como el resto de objetos, ésta admite distintas formas de “ser vista”. Aunque el significado de cada una de estas vistas sólo puede comprenderse plenamente cuando se tiene cierta experiencia en estimación de ecuaciones, podemos ahora anticipar una breve descripción del sentido de cada una de ellas:

- □ **View/Representations** muestra la ecuación en tres formas diferentes: como una lista de series (formato tradicional de especificación), como una expresión algebraica con coeficientes simbólicos y como la ecuación ya estimada con el valor de los coeficientes resultantes.

Pantalla de representaciones de una determinada ecuación



- □ **View/Estimation Output** ofrece, a la forma tradicional los principales resultados obtenidos en relación con parámetros, test de significación y contraste de hipótesis obtenidos.

Vista de Resultados de la Estimación de una determinada ecuación

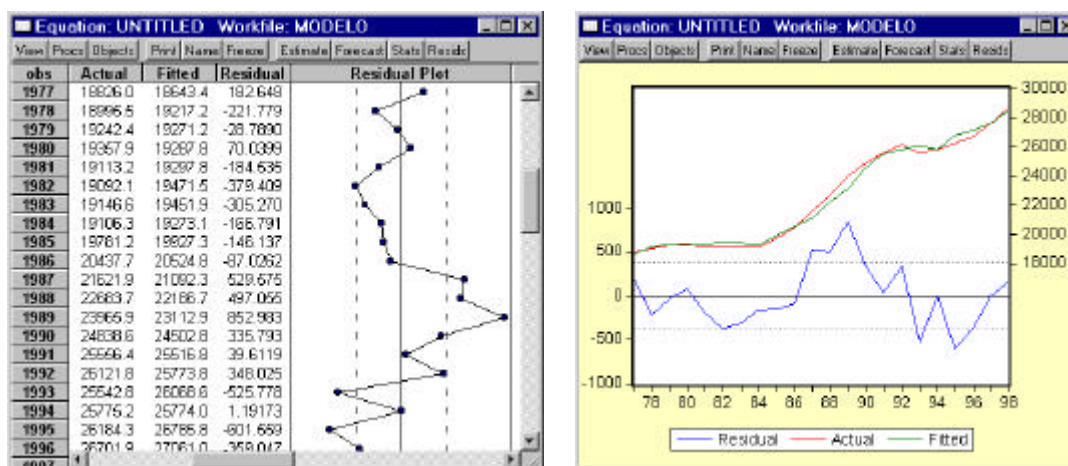
Variable	Coefficien	Std. Error	t-Statistic	Prob.
C	962.0328	837.2761	1.149003	0.2619
PCPI	11.97601	3.625558	3.303218	0.0030
YDH86	0.799045	0.047217	16.92268	0.0000

R-squared	0.992792	Mean dependent var	20535.30
Adjusted R-squared	0.992191	S.D. dependent var	3886.537
S.E. of regression	343.4394	Akaike info criterion	11.78246
Sum squared resid	2830815.	Schwarz criterion	11.92644
Log likelihood	-194.3746	F-statistic	1652.825
Durbin-Watson stat	1.063393	Prob(F-statistic)	0.000000

- □ **View/Actual, Fitted, Residual/Table** muestra los valores reales de la variable dependiente, los valores ajustados o estimados para esta variable con la ecuación y la diferencia entre ambos, es decir, el residuo de la estimación. La forma en que se puede mostrar esta información es doble: en forma de tabla o en forma de gráfico. Esta representación es una de las más útiles de cara al análisis posterior de los resultados obtenidos ya que, de un solo vistazo, puede apreciarse gráficamente si el ajuste es adecuado o queda lejos de la realidad.

Valor actual, estimado y residuo en formato "tabla"

Valor actual, estimado y residuo en formato "gráfico"



- **View/Covariance Matrix** muestra la matriz de varianzas y covarianzas de los **parámetros** estimados necesaria, entre otras cosas, para evaluar la precisión de la estimación.

Matriz de Varianzas y Covarianzas de los parámetros estimados

Coefficient Covariance Matrix				
	C	PCPI	YDH86	
C	1516504.	5680.391	-83.43176	
PCPI	5680.391	27.39782	-0.339525	
YDH86	-83.43176	-0.339525	0.004719	

Esta matriz no debe confundirse nunca con la matriz de varianzas y covarianzas de las variables incluidas en la regresión. Recordemos que, para el caso de la estimación con MCO en el Modelo Básico de Regresión Lineal, esta matriz de varianzas y covarianzas de los parámetros estimados responde a la expresión:

$$V(\hat{b}) = \hat{S}^2 (X'X)^{-1}$$

siendo S^2 la varianza de la perturbación aleatoria (o una estimación insesgada de la misma) y X la matriz de observaciones de las variables exógenas.

- **View/Coefficient Tests, Residual Tests, & Stability Tests** muestra menús adicionales específicos para la puesta en marcha de test de distinta naturaleza; desde test de cambio estructural (stability test ó coefficient test) como Wald, Chow o Ramsey, a tests sobre las propiedades de los residuos de la estimación (residual tests) y de diagnóstico de la ecuación estimada.

Finalizaremos con algo obvio pero necesario: una vez que, por uno u otro método, se ha estimado la ecuación, esta **debe guardarse como un objeto** con nombre propio. Por tanto, si no se ha utilizado el comando *Object - New Object* debe entonces usarse el comando *Name* de la barra de herramientas si se quiere conservar en el futuro. Suele ser un buen procedimiento, utilizar como nombres de ecuaciones el nombre de la variable endógena precedido por la letra "Q". En este caso, por ejemplo, el nombre de la ecuación podría ser *QCP86*. Así, logramos que todas las ecuaciones aparezcan juntas en el listado de objetos contenidos en un workfile .

II.E.4.- Otros botones de la ventana de ecuaciones

Además de estas operaciones básicas para estimar y ver en sus distintos formatos una ecuación, la pantalla de ecuaciones ofrece algunas operaciones adicionales que repasaremos a continuación.

- El botón **Procs** ofrece un submenú en el que las dos primeras opciones conciden con los comandos Estimación (*Estimate*) y Predicción (*Forecast*) que veremos más adelante. La tercera de las opciones (*Make Regressor Group*) es muy interesante, ya que permite crear un grupo de variables en el que quedarán guardadas las variables explicativas y la endógena utilizadas en la estimación de la ecuación.
- El menú **Objects** despliega una serie de opciones básicas de copiado y pegado. Además, incluye la opción *freeze* ya comentada anteriormente así como el comando *Store*. Este comando, permite almacenar la ecuación independientemente del workfile, es decir, al ejecutar este comando, la ecuación queda almacenada como un fichero independiente con la extensión *.DBE*. Si no se ha nombrado la ecuación, el EVIEWS permitirá especificar el nombre en ese momento.
- El botón **Print** imprime la vista activa en ese momento de la ecuación.
- El botón **Name** permite guardar con un nombre la ecuación estimada. Esta quedará como un objeto en el directorio del workfile precedida por un icono gráfico con el símbolo =.
- El botón **Freeze** copia la vista activa de la ecuación como un objeto independiente (*tabla o gráfico*) que no se modificará aunque lo haga la estimación original e la ecuación.
- El botón **Estimate** abre el cuadro de diálogo de la estimación de forma que pueda alterarse la especificación adicional y reestimarse los parámetros.
- El botón **Forecast** calcula la predicción a futuro de la variable endógena según la estimación realizada (*siempre y cuando las exógenas tengan especificados sus valores a futuro*).
- El botón **Stats** abre directamente la tabla de resultados de la estimación.
- El botón **Resids** muestra el gráfico o la tabla de valores reales, estimados y residuos.