

# Eviews tutorial<sup>1</sup>

*Universidad de San Andrés*

## Previo a la sesión de Eviews:

La estrategia para aprender a utilizar Eviews es la siguiente: primero vamos a construir un modelo arbitrariamente, para luego poder estimarlo por Eviews. Este procedimiento nos permitirá estimar un modelo del cual conocemos totalmente sus características, y podemos así evaluar la calidad de las estimaciones del modelo.

Lo primero que haremos es preparar nuestros datos<sup>2</sup>: Creamos cuatro series aleatorias<sup>3</sup> de 100 observaciones cada una:

- $X_1$  es una variable distribuida uniformemente en el intervalo  $[0,10]$
- $X_2$  es una variable distribuida uniformemente en el intervalo  $[-2,2]$
- $X_3$  es una variable distribuida uniformemente en el intervalo  $[-3,0]$
- $\varepsilon$  es el error del modelo, es un variable distribuida normalmente con media cero y varianza uno.

Con las variables explicativas generadas, estamos listos para generar las observaciones de la variable a explicar. Utilizamos el siguiente modelo lineal:

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \varepsilon$$

Para poder construir la variable Y, fijamos los parámetros del modelo a valores arbitrarios:  $\alpha = 10$ ,  $\beta_1 = 2$ ,  $\beta_2 = -1$  y  $\beta_3 = 0$  y, entonces, tenemos que Y es lo siguiente:

$$Y = 10 + 2 \cdot X_1 - X_2 + \varepsilon$$

Una vez generadas las variables del modelo, en formato Excel, estamos en condiciones de iniciar nuestra tutorial de Eviews.

## Sesión de Eviews:

### Creando el workfile:

El workfile es el área de trabajo en Eviews. El workfile es el lugar en donde guardaremos nuestra data y nuestros resultados. Un nuevo workfile se crea de la siguiente manera:

*File → New → Workfile*

Inmediatamente después, el Eviews nos pregunta sobre la frecuencia de nuestro dataset. Como dijimos antes, imaginaremos que nuestro dataset corresponde a datos anuales, para los años 1902-2001.

<sup>1</sup> Esta es una versión adaptada (para los fines de este curso) de la orginial creada por Federico Andres Bugni.

<sup>2</sup> Los datos se encuentran disponibles en la página web: [www.geocities.com/bugnif/](http://www.geocities.com/bugnif/)

<sup>3</sup> Los datos se crearon en S-Plus.

*Anual → Start date: 1902 ; End date: 2001 → Ok*

El Eviews puede trabajar con datos time series o con datos cross section.

Si los datos son time series, se pueden clasificar en *Anual, Semi-anual, Quarterly, Monthly, Weekly* ó *Daily*. Es importante fijar correctamente la frecuencia de los datos debido a que ésta puede ser utilizada para generar automáticamente variables explicativas.

Si los datos son cross section, utilizaremos la frecuencia *undated or irregular*, indicando en *End Observation* el total de número de observaciones.

<b>Frecuencia</b>	<b>Start</b>	<b>End</b>
Anual	Año inicial	Año final
Semi-anual	Año inicial: semestre inicial	Año inicial: semestre final
Quarterly	Año inicial: cuarto inicial	Año final: cuarto final
Monthly	Año inicial: Mes inicial	Año final: mes final
Weekly	Año inicial	Año final
Daily (7 days weeks)	Año inicial	Año final
Daily (5 days weeks)	Año inicial	Año final
Undated or irregular	Observación inicial	Observación final

Nuestro workfile: untitled, contiene solamente dos variables,

- **C:** es la constante del modelo.
- **Resid:** son los residuos estimados del modelo.

Por el momento, C tiene valor nulo y Resid se encuentra no disponible debido a que no hemos corrido ninguna regresión. Ambas variables se actualizarán automáticamente al estimar regresiones.

#### **Salvando/Abriendo el workfile:**

Para salvar un workfile en formato Eviews (.wf1), utilizamos el comando:

*File → Save As...*

de esta manera le daremos un nombre al workfile.

Un workfile en formato Eviews puede ser abierto utilizando el comando:

*File → Open → Workfile*

### Importando data:

Para importar el dataset en formato Excel a nuestro workfile, utilizamos el comando

*File → Import → Read Lotus-Text-Excel...*

Ahí le debemos indicar algunas características de nuestro dataset en formato Excel:

*By observation - series in columns (clickeado)*  
*Upper-left data cell: A2*  
*Excell 5+ sheet name:*  
*Names for the series or the number of series if names in file: 4*  
*Sample to import: 1902 2001*

Entonces aparecen 4 variables nuevas, con los nombres originales que tenían en el Excel: Y, X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>.

### Creando una serie:

El Eviews puede crear series a partir de la información incluida en el workfile, utilizando el comando:

*Quick → Generate Series...*

Luego, se debe insertar la fórmula de la serie a generar. Para ello se puede utilizar los nombres de las variables del workfile y operaciones algebraicas conocidas.

Operador	Función
Y+X	Adición de series
Y-X	Substracción de series
Y*X	Producto de series
Y/X	División de series
Y^n	Elevar a la n una serie
abs(Y)	Valor absoluto de una serie
Y(-n)	Rezaga la variable Y, n periodos
D(Y)	Toma diferencias de una serie

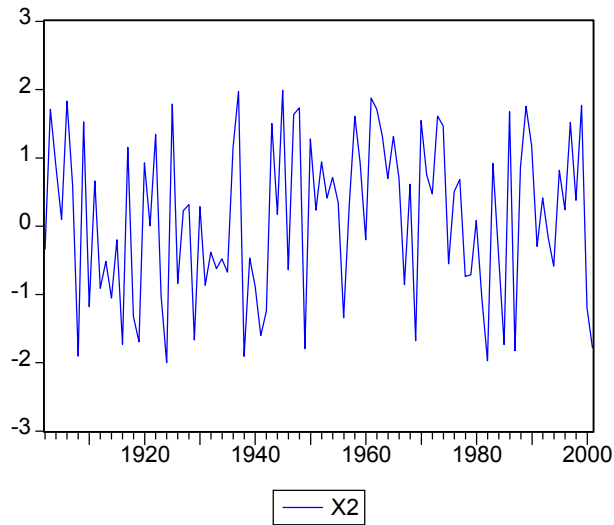
### Estudiando una serie:

Podemos analizar cualquier variable en forma individual haciendo doble click en la misma. Esto abrirá una nueva ventana con el spreadsheet de nuestra variable.

Estudiemos, por ejemplo, la variable X<sub>2</sub>, una de nuestras variables explicativas, de la cual conocemos que es tiene una distribución uniforme en el intervalo [-2,2]. Para generar un gráfico de línea de la variable usamos:

*View → Line graph*

Produce el siguiente gráfico:

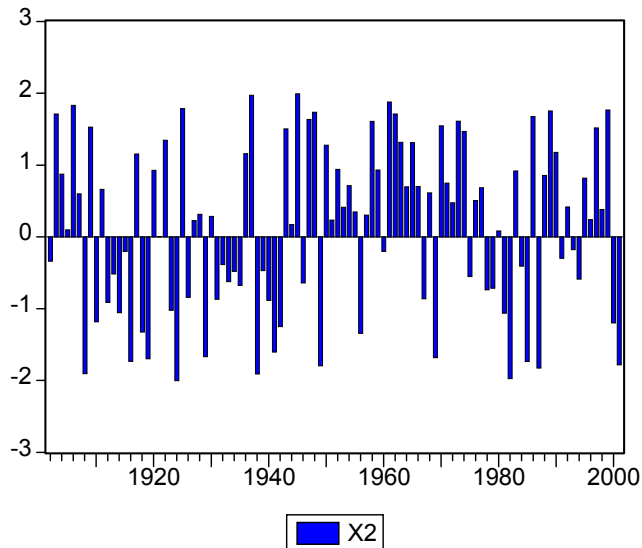


Haciendo doble click dentro del gráfico se pueden cambiar las opciones del mismo.

Para producir un gráfico similar al anterior de barras, usamos:

*View → Bar graph*

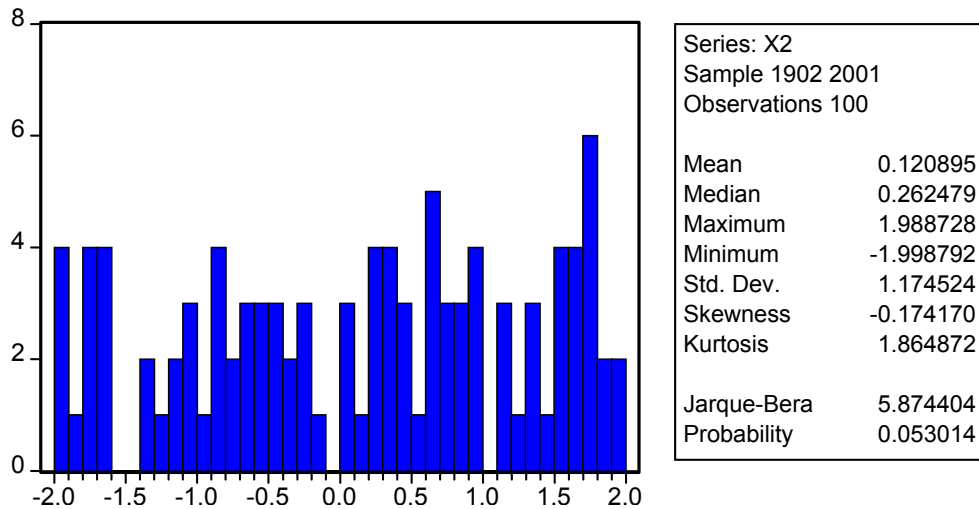
Que genera un gráfico similar anterior, pero en barras:



Para poder ver estadísticas descriptivas de la variable, usamos:

*View → Descriptive Statistics → Histogram and Stats*

Que produce lo siguiente:



Mean es la media muestral (promedio), Median es la mediana muestral<sup>4</sup>, y Std. Dev. es la raíz de la medida insesgada de varianza muestral (dividida por n-1).

Skewness es una medida de la asimetría de la distribución de  $X_2$  alrededor de la media. Para una serie normalmente distribuida, Skewness es cero. Skewness positiva indica que la distribución tiene una cola derecha relativamente larga y Skewness negativa indica que la distribución tiene una cola izquierda relativamente larga.

Kurtosis mide si la distribución tiene pico o si es relativamente chata. Para una distribución normal, la Kurtosis es 3. Si la Kurtosis excede el valor de 3, la distribución tiene pico (relativo a la normal), y si la Kurtosis tiene valor menor a 3, la distribución es chata (relativo a la normal).

El estadístico Jarque-Bera corresponde al siguiente test de la normalidad:

$H_0$ :  $X_2$  está normalmente distribuido.

$H_a$ :  $X_2$  no está normalmente distribuido.

Debido a que la variable  $X_2$  fue creada por nosotros con distribución uniforme en el intervalo  $[-2, 2]$ , ya sabemos que la hipótesis nula es falsa.

Dado el Pvalor del test, diremos que para un nivel de significatividad del 1% o 5%, se rechaza la normalidad de la distribución, pero a un nivel de significatividad del 10% no se rechaza.

Se pueden realizar diferentes tests de hipótesis sobre la distribución de la variable elegida, usamos:

*View → Test for Descriptive Stats → Simple Hipotesis Tests*

Entonces debemos introducir la Hipótesis nula de nuestro test sobre los parámetros de la distribución de la variable elegida. Por ejemplo, realicemos el siguiente test:

$H_0$ :  $E(X_2) = 0$

$H_a$ :  $E(X_2) \neq 0$

<sup>4</sup> Si el número de observaciones (n) es par, la mediana se calcula como el promedio simple de las dos observaciones centrales.

Debido a que  $X_2$  es uniforme en el intervalo  $[-2,2]$ , sabemos que la hipótesis nula es verdadera. El resultado del test es el siguiente:

Hypothesis Testing for X2		
Date: 06/22/02 Time: 15:38		
Sample: 1902 2001		
Included observations: 100		
Test of Hypothesis: Mean = 0.000000		
Sample Mean = 0.120895		
Sample Std. Dev. = 1.174524		
<u>Method</u>	<u>Value</u>	<u>Probability</u>
t-statistic	1.029307	0.3058

Un Pvalor de 0.3058 nos lleva a no rechazar la hipótesis nula para niveles de significatividad razonables. Eviews no se equivoca.

Para volver al spreadsheet de la variable, usamos:

*View → SpreadSheet*

### Creando un grupo:

Podemos analizar un conjunto de variables del workfile conformando un grupo. Utilizando el mouse, y manteniendo apretada la tecla Control, seleccionamos el conjunto de variables que conformarán nuestro grupo. Luego de definidas la/s variables que formarán nuestro grupo, utilizamos el comando:

*View → Open Selected → One Window → One Group*

En una nueva ventana aparecerá nuestro grupo.

El menú view ahora nos permite encontrar estadísticas descriptivas de las variables de las variables del grupo:

*View → Descriptive Statistics → Common Sample*

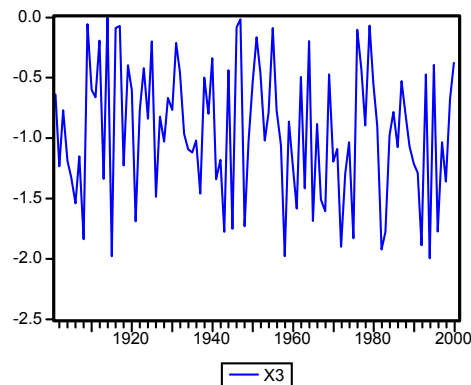
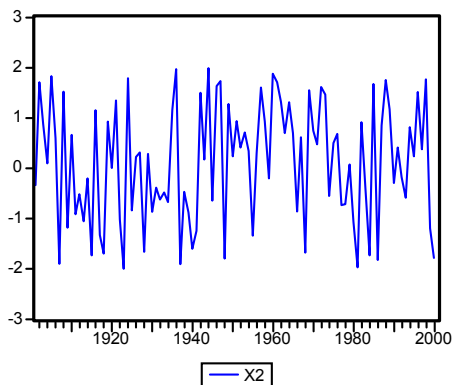
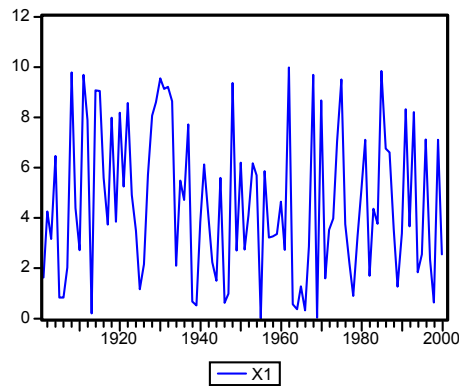
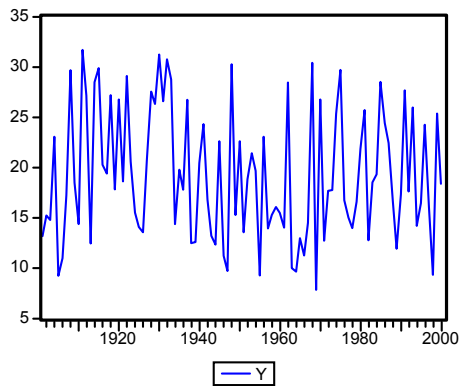
que produce la siguiente tabla de estadísticos descriptivos:

	Y	X1	X2	X3
Mean	19.26857	4.593447	0.120895	-0.943967
Median	17.80627	3.914813	0.262479	-0.930699
Maximum	31.70575	9.982131	1.988728	-0.000393
Minimum	7.838590	0.010946	-1.998792	-1.995487
Std. Dev.	6.431214	2.977881	1.174524	0.554390
Skewness	0.294067	0.283181	-0.174170	-0.141815
Kurtosis	1.921392	1.862446	1.864872	2.075214
Jarque-Bera	6.288738	6.728310	5.874404	3.898648
Probability	0.043094	0.034591	0.053014	0.142370
Observations	100	100	100	100

Se puede graficar hacer gráficos de línea de todas las variables:

*View → Multiple Graphs → Line*

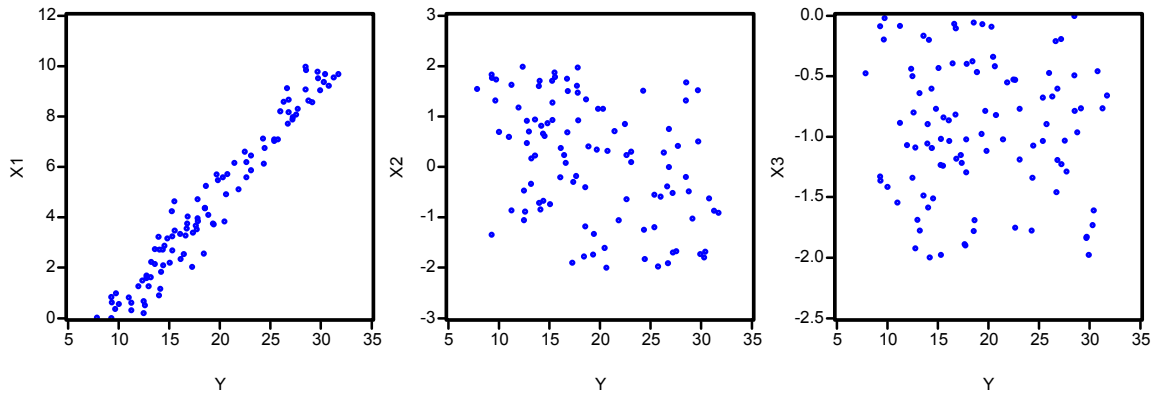
que produce los siguientes gráficos:



Se puede realizar un gráfico de puntos de la primera variable (Y) en función del resto de las variables del modelo:

*View → Multiple Graphs → Scatter → First Series Against All*

que produce los siguientes gráficos de puntos:



También podemos mirar el coeficiente de correlación entre las variables del grupo:

*View → Correlations*

Es importante verificar que existe baja correlación entre las variables del modelo  $X_1$ ,  $X_2$ ,  $X_3$ , para evitar el problema de la multicolinealidad (cercana).

	Y	X1	X2	X3
Y	1.000000	0.975532	-0.400602	-0.043193
X1	0.975532	1.000000	-0.247301	-0.056818
X2	-0.400602	-0.247301	1.000000	-0.102874
X3	-0.043193	-0.056818	-0.102874	1.000000

Esto nos muestra que las correlaciones muestrales de todas las variables explicativas del modelo ( $X_1$ ,  $X_2$ ,  $X_3$ ) son (en valor absoluto) menores al 0.25, lo cual asegura que no vamos a caer en el problema de multicolinealidad cercana.

### Regresión por OLS:

Dentro del menú asociado al grupo elegimos:

*Procs → Specify/Estimate*

Inmediatamente después debemos indicar algunas características de la estimación que deseamos realizar.

*Equation Specification: Y X1 X2 X3 C*  
*Method: LS - Least Squares (NLS and ARMA)*  
*Sample: 1901 2001*

Esto último indica que estamos haciendo una regresión en donde la variable explicada es Y y las variables explicativas observables son  $X_1$ ,  $X_2$ ,  $X_3$ , una constante y, por supuesto un residuo no observable. Entonces, el modelo a regresar es:

$$Y = \alpha + \beta_1.X_1 + \beta_2.X_2 + \beta_3.X_3 + \varepsilon$$

El método de regresión es mínimos cuadrados ordinarios y que la muestra que se utiliza para la regresión son los 100 años comprendidos en los años 1901 - 2001.

La siguiente es la salida de la regresión:

Dependent Variable: Y				
Method: Least Squares				
Date: 06/22/02 Time: 16:49				
Sample(adjusted): 1901 2000				
Included observations: 100 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	2.014702	0.033268	60.56046	0.0000
X2	-0.934666	0.084659	-11.04031	0.0000
X3	-0.089898	0.174068	-0.516450	0.6067
C	10.04228	0.235831	42.58258	0.0000
R-squared	0.978769	Mean dependent var	19.26857	
Adjusted R-squared	0.978106	S.D. dependent var	6.431214	
S.E. of regression	0.951602	Akaike info criterion	2.777838	
Sum squared resid	86.93240	Schwarz criterion	2.882044	
Log likelihood	-134.8919	F-statistic	1475.264	
Durbin-Watson stat	1.853140	Prob(F-statistic)	0.000000	

Analicemos la salida de Eviews paso a paso, para las cosas que usaremos:

Dependent Variable: Y

Eviews muestra que la variable dependiente del modelo es Y,

Method: Least Squares

que el modelo es estimado por OLS,

Sample(adjusted): 1901 2000

Included observations: 100 after adjusting endpoints

En las estimación se utilizaron las 100 observaciones disponibles, que corresponde a las observaciones de todas las variables desde el año 1901 hasta el año 2001.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	2.014702	0.033268	60.56046	0.0000
X2	-0.934666	0.084659	-11.04031	0.0000
X3	-0.089898	0.174068	-0.516450	0.6067
C	10.04228	0.235831	42.58258	0.0000

A continuación se nos indica que las variables explicativas del modelo son:  $X_1$ ,  $X_2$ ,  $X_3$  y la constante C. Por lo tanto, el modelo a estimar por OLS fue el siguiente:

$$Y = C + C(1)X_1 + C(2)X_2 + C(3)X_3$$

La estimación de Eviews de los coeficientes del modelo a través de OLS es:

$$Y = 10.04 + 2.01.X_1 - 0.93.X_2 - 0.08.X_3$$

Lo cual es una excelente predicción del modelo real.

Para cada variable explicativa, el Eviews reporta el coeficiente estimado, el error estándar del coeficiente estimado, el t-estadístico asociado al test de no significatividad individual del coeficiente y el Pvalor correspondiente a dicho t-estadístico. Recordar que el t-estadístico se conforma como el cociente del coeficiente estimado y del error estándar en la estimación del mismo.

Por ejemplo, para el coeficiente de la variable  $X_1$ , el test de significatividad individual es el siguiente:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

El t-estadístico del test es 60.56, cuyo Pvalor asociado es (casi) nulo. Esto nos permite rechazar la hipótesis nula (casi) para cualquier nivel de significatividad.

Debido a que nosotros "construimos" la variable dependiente, sabemos que la hipótesis de no significatividad es falsa para los coeficientes de  $X_1$ , de  $X_2$  y la constante, y es verdadera para el coeficiente de  $X_3$ . Los tests de hipótesis individuales no se equivocan: los coeficientes de las variables  $X_1$ ,  $X_2$  y la constante son significativos casi para cualquier nivel de significatividad, mientras que el coeficiente de la variable  $X_3$  es no significativo para niveles de significatividad razonables.

El resto de la tabla se lee en dos columnas:

R-squared	0.978769
-----------	----------

El  $R^2$  de la regresión es una medida de la bondad de ajuste del modelo: indica cual es el cociente entre la suma de cuadrados explicados y la suma de cuadrados totales de la variable independiente, es decir, que porcentaje de la variabilidad de la variable dependiente es explicada a través de la variabilidad de las variables independientes del modelo.

Este valor no suele ser de utilidad en forma individual, sino que es útil para comparar entre diferentes modelos.

En este caso, con las variables explicativas del modelo estamos explicando un 97% de la variabilidad en la variable dependiente.

Adjusted R-squared	0.978106
--------------------	----------

El  $R^2$  tiene la propiedad de no decrecer con el número de variables explicativas de un modelo, es decir, no penaliza por la cantidad de variables explicativas del modelo. Por lo tanto, el criterio de elegir el modelo con mayor  $R^2$  nos llevará a incluir todas las variables explicativas posibles. Para corregir este "vicio" del  $R^2$  miramos el  $R^2$  ajustado, que es una medida de bondad de ajuste que crece con el porcentaje de la variabilidad de la variable dependiente explicado por las variables explicativas, pero decrece con el número de variables incluidas en el modelo. Es decir, el  $R^2$  ajustado penaliza por el número de variables explicativas del modelo. Nuevamente, este valor no suele ser de utilidad en forma individual, sino que es útil sólo para comparar entre diferentes modelos.

En este caso, el  $R^2$  y el  $R^2$  ajustado son bastantes similares reflejando que en el modelo hay pocas variables irrelevantes, que el  $R^2$  ajustado penaliza.

S.E. of regression	0.951602
--------------------	----------

El S.E. of regression es nuestra estimación del desvío estándar de los residuos del modelo. Se calcula como la raíz de  $S^2$ , siendo  $S^2$  la suma de los cuadrados de los residuos de la regresión divididos por  $n-k$ . El desvío estándar de la regresión es una medida del error del desvío estándar que tendrá nuestro modelo estimado para pronosticar. Cuanto mayor sea el  $S^2$ , mayor será el error de pronóstico del modelo.

Sum squared resid	86.93240
-------------------	----------

El Sum squared resid es la suma de los cuadrados de los residuos del modelo, o sea, la suma de cuadrados no explicados en el modelo.

Mean dependent var	19.26857
--------------------	----------

Mean dependent var es la media muestral (promedio) de la variable dependiente de la regresión,  $Y$ . En este caso, podemos comprobar que coincide con la media muestral de los estadísticos descriptivos mostrados anteriormente.

S.D. dependent var	6.431214
--------------------	----------

S. D. dependent var es el desvío estándar estimado de la variable dependiente de la regresión,  $Y$ . Se calcula como la raíz de la medida insesgada de varianza muestral (dividida por  $n-1$ ).

Akaike info criterion	2.777838
Schwarz criterion	2.882044

Akaike Info Criterion (AIC) y Schwarz Criterion (SC) son dos medidas del error de pronóstico que tiene nuestro modelo estimado, al igual que el  $S^2$ . Estos dos medidas penalizan el uso de variables explicativas en el modelo (al igual que el  $S^2$ ).

Si ordenamos en forma creciente las medidas de error de pronóstico en términos del grado de penalidad por cada variable explicativa adicional diremos que la que menos penaliza es  $S^2$ , luego viene AIC y luego viene SC.

Al igual que el  $S^2$ , el uso de AIC o de SC no tiene demasiado sentido en forma individual, sino que son de utilidad para comparar entre modelos alternativos.

F-statistic	1475.264
Prob(F-statistic)	0.000000

Eviews nos presenta el F-estadístico y el Pvalor correspondientes al test de no significatividad conjunta de todas las variables explicativas del modelo (salvo la constante). Es decir corresponde al test cuya hipótesis nula es que todas las variables explicativas (salvo la constante) tienen asociados coeficientes nulos, mientras que la hipótesis alternativa es que alguna o algunas de las variables explicativas (salvo la constante) tienen asociados coeficientes significativamente distintos de cero. Bajo la hipótesis nula, el mejor predictor de la variable dependiente es una constante.

En el ejemplo, el test de significatividad conjunta es:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

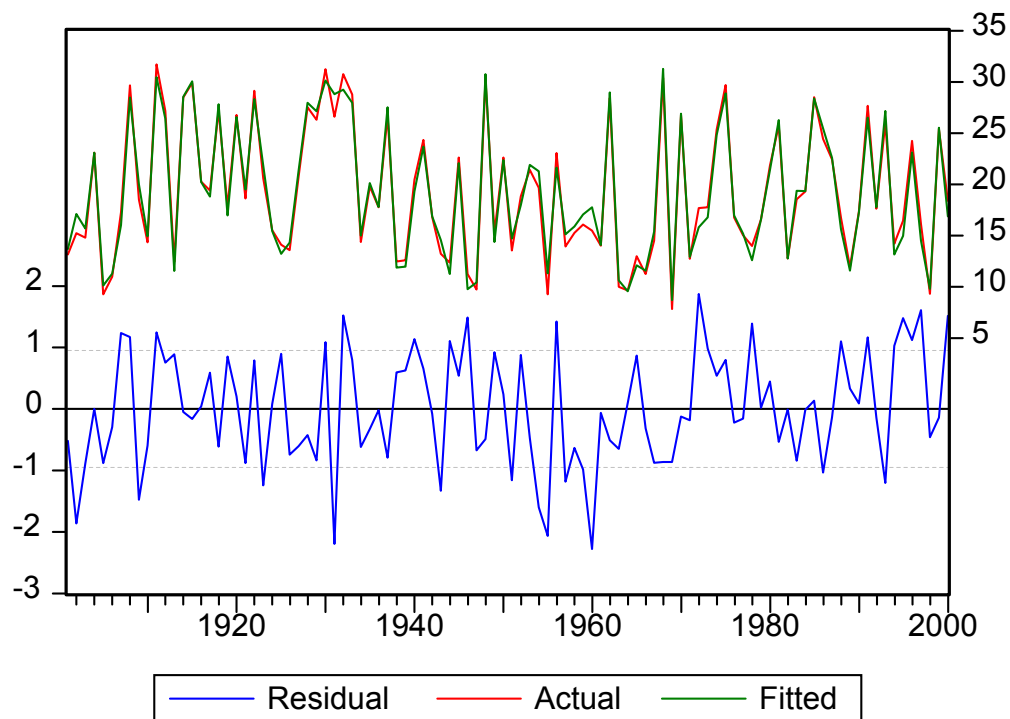
$$H_a: \beta_1 \neq 0 \text{ ó } \beta_2 \neq 0 \text{ ó } \beta_3 \neq 0$$

El Pvalor asociado casi nulo permite rechazar la hipótesis nula en forma terminante para (casi) cualquier nivel de significatividad. Sabemos de antemano que la hipótesis nula es falsa, dado que  $\beta_1$  y  $\beta_2$  son no nulos.

Luego de entender la salida, resulta siempre interesante hacer un gráfico de línea de nuestra variable dependiente, de nuestro pronóstico de la misma y de los residuos. Esto se puede hacer a través del comando:

<i>View → Coefficient Test → Actual, Fitted, Residual → Actual, Fitted, Residual Graph</i>
--------------------------------------------------------------------------------------------

Que produce el siguiente gráfico:



#### Test de Wald:

Si llamamos  $\beta$  al vector columna de variables explicativas (incluyendo la constante),  $\beta \in \mathbb{R}^{k+1}$ , el test de Wald puede realizar el siguiente test:

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

$$H_a: R \cdot \beta \neq r$$

donde  $R$  es una matriz de constantes,  $R \in \mathbb{R}^{q \times k}$  y  $r$  es un vector de constantes  $r \in \mathbb{R}^{q \times 1}$ . Entonces, el test de Wald nos permite realizar tests sobre cualquier sistema de ecuaciones lineales de los parámetros. Para realizar un test de Wald recurrimos al siguiente comando dentro del menú de grupo:

*View → Coefficient Test → Wald/Coefficient Restrictions...*

Un ejemplo de test de Wald son, por lo tanto, los tests de significatividad individual. Otro ejemplo posible de test de Wald es el siguiente:

$$H_0: -\beta_1 = 2 \cdot \beta_2 \text{ y } \beta_3 = 0$$

$$H_a: -\beta_1 \neq 2 \cdot \beta_2 \text{ ó } \beta_3 \neq 0 \text{ ó ambas}$$

El resultado es el siguiente:

Wald Test:			
Equation: EJEMPLO			
Null Hypothesis:	C(1)=-2*C(2) C(3)=0		
F-statistic	0.519233	Probability	0.596638
Chi-square	1.038465	Probability	0.594977

Antes de realizar el test, sabemos por construcción que la hipótesis nula es verdadera. El resultado del test indica que no rechazamos la hipótesis nula casi para niveles de significatividad razonables.

## Pronosticando con nuestro modelo:

Eviews nos permite construir automáticamente un pronóstico de la variable dependiente a partir del modelo estimado. El comando para ello es:

*Procs → Forecast...*

Instantáneamente aparecerá un cuadro para especificar las características de nuestro pronóstico. Elegimos pronosticar la variable para los 15 años de los datos: 1985 - 2000.

*Forecast Name: YF*  
*S.E. (optional): SE*  
*Sample Range For Forecast: 1985 2000*

Esta estimación genera dos series nuevas en nuestro worksheet: YF y SE, siendo YF nuestro pronóstico de la variable dependiente en el intervalo 1985 - 2000 y SE la estimación del error estándar de la estimación en el intervalo 1985 - 2000.

Suponiendo que nuestro modelo está correctamente especificado y que los residuos del modelo son normales entonces, con una probabilidad del 95% el verdadero valor de la variable dependiente debería caer en el intervalo de confianza dado por  $YF \pm 2 \cdot SE$ . Para verificar ello, construimos la cotas superiores o inferiores del intervalo de confianza, con el comando:

*Quick → Generate Series...*

E insertamos la fórmula de nuestras nuevas variables:

*Enter equation: UPPER = YF + 2\*SE*  
*Enter equation: LOWER = YF - 2\*SE*

Para mostrar la bondad de la estimación generamos un grupo con las variables YF, Y, UPPER y LOWER.

*View → Open Selected → One Window → One Group*

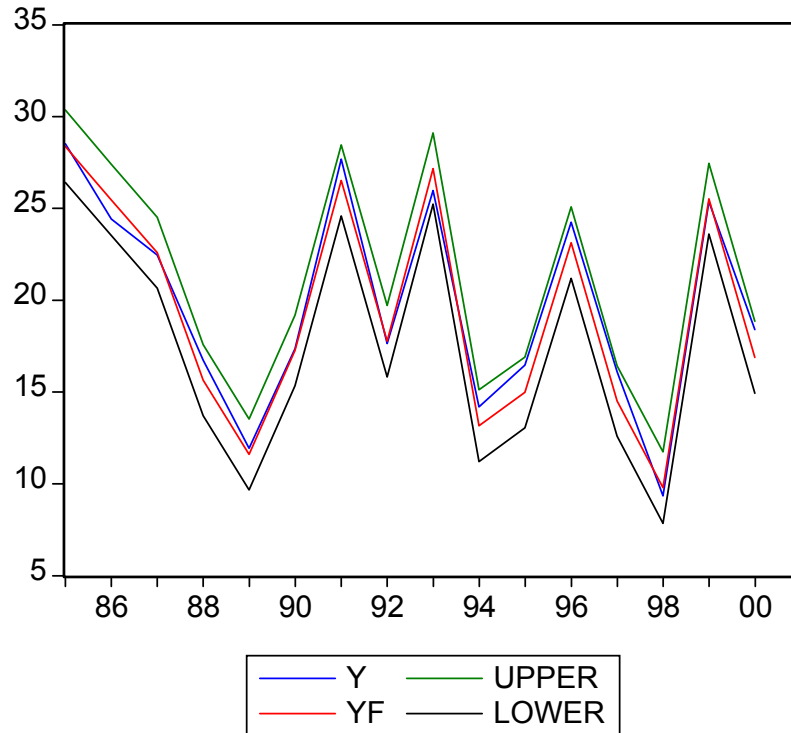
Restringimos los valores de los series al período 1985-2000

*Sample → Sample range pairs (or sample object to copy): 1985 - 2000*

Luego, graficamos las cuatro series:

*View → Graphs → Line*

Para producir el siguiente gráfico:



Como podemos ver, el ajuste es muy bueno: la serie real se ubica siempre dentro del intervalo de confianza del 95% generado por LOWER y UPPER.

### **Menú de Ayuda:**

El programa Eviews cuenta con un excelente menú de ayuda, el cual puede ser consultados sobre comandos y sobre conceptos de estadística. Se accede a él a través del comando:

*Help → EViews Help Topics...*

Una vez ingresado en el índice de ayuda, se puede tipear el nombre del comando o concepto estadístico sobre el cuál se requiere ayuda.