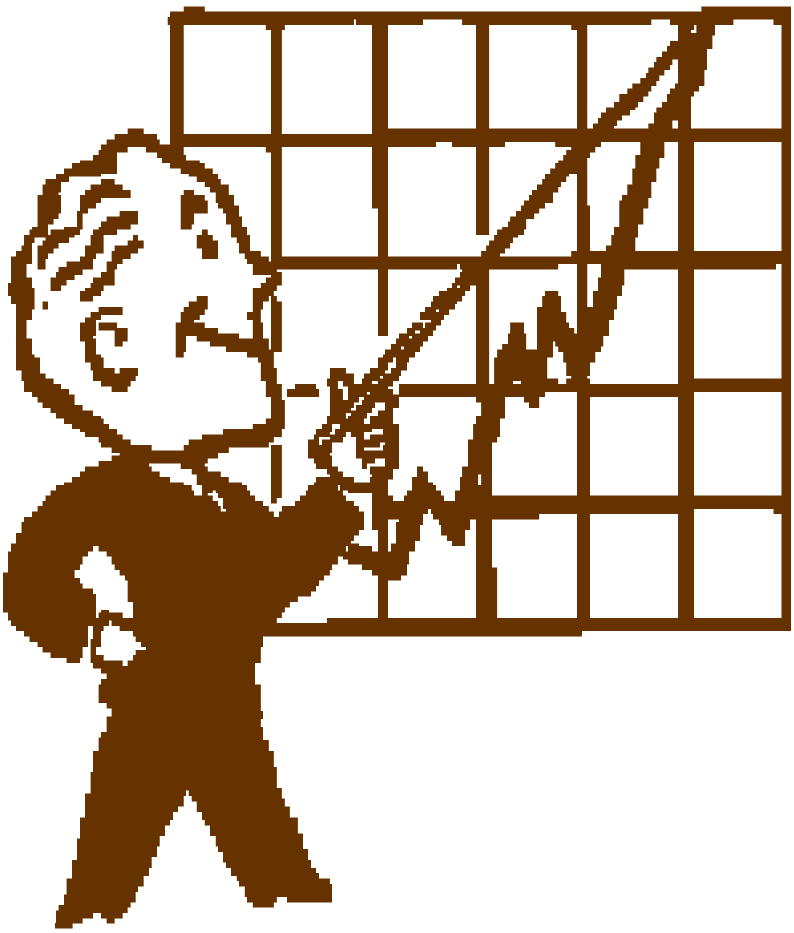
UNIDAD IV

¿Qué es predicción en el modelo lineal?

UNIDAD IV

¿Qué es la predicción en el modelo lineal?



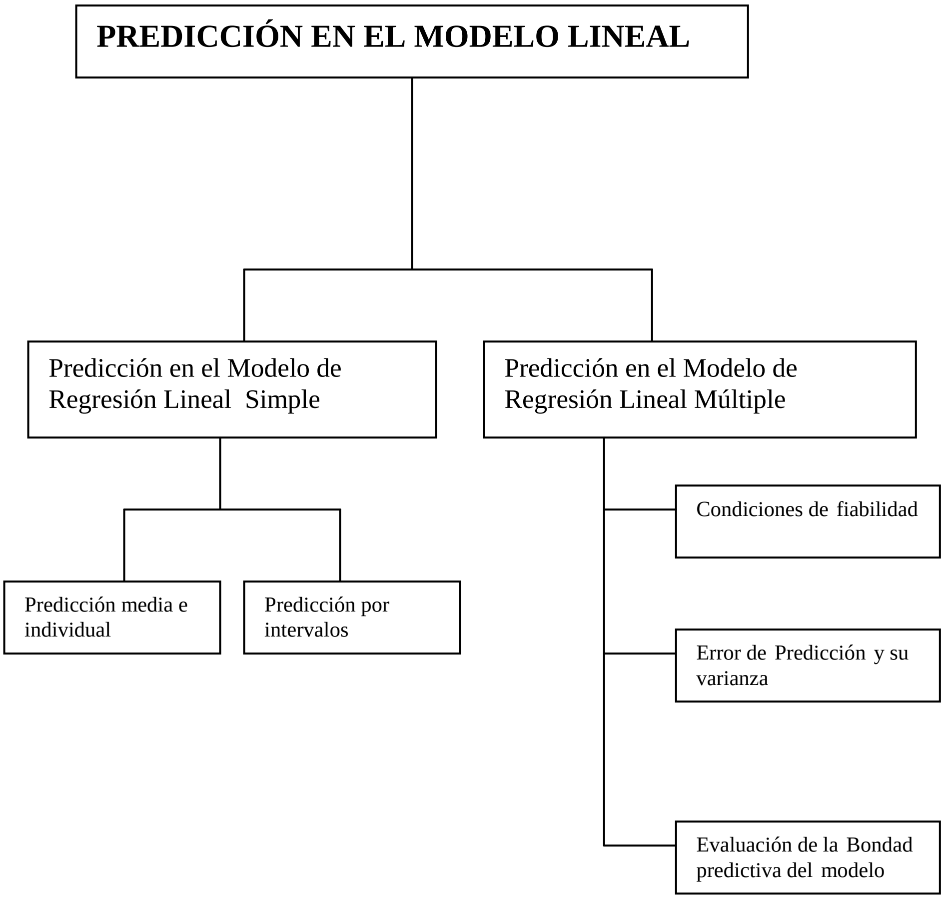
“La economía es el estudio del modo en que la sociedad gestiona sus recursos escasos”

Gregory Mankiw

- **¿Qué es predicción media e individual en el Modelo de Regresión Lineal Simple y en el Modelo de Regresión Lineal Múltiple?**
- **¿Cuáles son las condiciones de fiabilidad?**
- **¿Cómo se define el error de predicción y su varianza?**
- **¿Por qué se hace la Evaluación de la Bondad Predictiva del Modelo?**

PREDICCIÓN EN EL MODELO LINEAL

ESQUEMA CONCEPTUAL



COMPETENCIAS A LOGRAR

CONCEPTUAL	PROCEDIMENTAL	ACTITUDINAL
Explica que es una Predicción en el Modelo de Regresión Lineal Simple y Múltiple.	**Analiza la aplicación de la predicción en el modelo para proyectar el comportamiento de la variable dependiente.**	**Tiene una actitud crítica respecto a la predicción.**

CONCEPTOS –CLAVE

Predicción, Fiabilidad, Bondad Predictiva

LECCIÓN 1

PREDICCIÓN MEDIA E INDIVIDUAL EN EL MODELO DE REGRESIÓN LINEAL SIMPLE

1. PREDICCIÓN

Uno de los principales objetivos de la construcción de un modelo econométrico es la realización de predicciones, es decir, el pronóstico del valor que va a tomar la variable endógena para unos valores dados de las variables explicativas fuera de las observaciones muestrales

Siempre debe de tenerse presente que una predicción es fiable si el modelo está correctamente especificado, además las relaciones causales establecidas en el modelo se mantienen para las observaciones fuera de la muestra conociéndose con exactitud lo que valen las variables explicativas en el periodo de predicción.

La importancia de poder predecir el futuro con algún grado de exactitud es inapreciable. En el mundo de la empresa, la capacidad de prever y predecir sucesos y tendencias futuras refuerza en gran medida la probabilidad del éxito. Por consiguiente, no es de extrañar que las empresas inviertan gran cantidad de tiempo y esfuerzo en conseguir una predicción exacta de la tendencia y la evolución económica futura. Para elaborar predicciones útiles se pueden usar numerosas herramientas cuantitativas. Con ayuda de estas herramientas podrán construir su propia bola de cristal y emplearla para escrudiñar el futuro. (Allen Webster, 1998).

El éxito o fracaso de una gestión empresarial se juzga sobre la base de resultados obtenidos y la eficacia en la ejecución de los proyectos.
La toma de decisiones en asuntos relativos al ambiente económico ocupa un lugar relevante y se convierte en una actividad permanente. En ella intervienen algunos elementos perfectamente conocidos y otros manejables, así como también fenómenos o hechos de los cuales no se posee un conocimiento preciso, en muchos casos porque se derivan de acciones que sucederán en el futuro.
¿Cuál será la respuesta de la demanda ante una variación en el precio de venta de nuestro producto, si se decide mantener la calidad y se presume que nuestros competidores no nos van a seguir? ¿Reaccionarán nuestros competidores frente una agresiva política de precios de venta de nuestro producto? ¿Aumentará la productividad de los empleados si se decide una modificación en el horario de trabajo? ¿Será rentable invertir en maquinaria que al cabo de tres años aumentará la capacidad productiva en un 50%, si las condiciones de política económica son relativamente estables? ¿Se debe considerar que la demanda de nuestro producto sufre variaciones de acuerdo a cambios estacionales? De ser positiva la respuesta, ¿habrá que considerar un efecto estacional en el proceso productivo? Las ventas de un artículo aparentemente complementario al nuestro ¿juegan un rol importante en nuestras ventas? ¿Subirán los precios de nuestros insumos lo suficiente como para justificar un exceso de inventarios? Un incremento en los gasto de publicidad, ¿tendrá un impacto en las ventas de una empresa? Estas son algunas de las muchas preguntas que enfrenta un administrador durante su gestión, cuyas respuestas no pueden ser precisas debido a que interviene un conjunto de factores cuyo devenir no se conoce con certeza.

Sin embargo, es necesario tomar decisiones sobre la base de las que se presume serán las respuestas, tratando de reducir los costos (o beneficios no ganados) en caso de error. En tal tarea, el responsable de la gestión (administrador) tendrá que proyectar y evaluar un posible escenario de los acontecimientos futuros, así como analizar su impacto sobre los eventos que afectan la operación de la empresa.
La experiencia del administrador juega un papel importante. Al tomar en consideración métodos basados únicamente en su experiencia tiene la ventaja de incorporar elementos propios del investigador tales como su conocimiento del medio y del transcurrir de los hechos recientes en el marco dentro del cual se desarrolla la empresa.

En el proceso de toma de decisiones, se puede recurrir adicionalmente a un conjunto de técnicas o instrumentos cuantitativos que otorgan elementos de medición para evaluar el impacto de las decisiones potenciales sobre las variables denominadas objetivo, incorporando la menor proporción de elementos subjetivos. Y se puede extraer conclusiones que aporten en la búsqueda de la decisión a adoptar.
La predicción analizada como una técnica cumple un gran papel, y la estadística proporciona un vehículo útil para el administrador.
La predicción como una técnica estadística define cuál de los valores posibles futuros de la variable es más probable. Dicho de otro modo: se asigna una probabilidad de ocurrencia a cada uno de los posibles valores de la variable objetivo, con lo que se podrá hacer una lista priorizada sobre la base de su factibilidad probabilística. (Cortez Cumpa, 1994).

Ejemplo ilustrativo
Suponga que se espera un PBI real de $9,000 (mil millones) en el 2004, ¿Cuál es su pronóstico de gasto de consumo en dicho año? Cuya función consumo es:

$$\hat{Y}_t = -213.8 + 0.7194X_t$$

$$\hat{Y}_t = -213.8 + 0.7194X_t$$

Si se reemplaza el valor de 9000 se obtendrá un pronóstico a partir de reemplazar:

$$\hat{Y}_{2004} = -213.8 + 0.7194(9000) = 6260.8$$

2. PREDICCIÓN MEDIA E INDIVIDUAL

Predicción del valor medio condicional de Y correspondiente a un determinado $X/X=X_{n+1}$ que es el punto de la línea de regresión poblacional.

Predicción Media de Y

$$E(Y/X_{n+1}) = \hat{\beta}_1 + \hat{\beta}_2 X_{n+1}$$

Predicción de un Valor Individual o Predicción por Punto

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_{n+1}$$

Donde: $E(Y/X_{n+1}) = \hat{Y}$

Es decir la predicción media o de un valor individual de Y para el caso puntual coinciden.

3. PREDICCIÓN POR INTERVALOS

Se trata de encontrar el intervalo de variación de los valores de predicción ($\hat{Y}_{n+1}$) de la variable explicada, conociendo un valor de predicción X_{n+1} de la variable explicativa y señalando un nivel de confianza estadística.

-**La predicción puntual** es: $\hat{Y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 X_{n+1}$

La **Varianza** es la siguiente:

$$Var(\hat{Y}_{n+1}) = s^2 \left[\frac{1}{n} + \frac{\left(X_{n+1} - \overline{X}\right)^2}{\sum \left(X_i - \overline{X}\right)^2} \right]$$

La **Desviación Estándar** es:

$$s_{\hat{Y}_{n+1}} = s \sqrt{\frac{1}{n} + \frac{\left(X_{n+1} - \overline{X}\right)^2}{\sum \left(X_i - \overline{X}\right)^2}}$$

Construyendo una t:

$$t = \pm \frac{\hat{Y}_{n+1} - Y_{n+1}}{s_{\hat{Y}_{n+1}}}$$

La cual posee (n-2) g.1.(grado de libertad) y con un nivel de significancia de $1 - \frac{\alpha}{2}$

De donde: la fórmula de **predicción por intervalos promedio** de Y_i es:

$$\hat{Y}_{n+1} - ts_{\hat{Y}_{n+1}} < Y_{n+1} < \hat{Y}_{n+1} + ts_{\hat{Y}_{n+1}}$$

Si se desea una **predicción por intervalos individual de** Y_i **, la fórmula es:**

$$\hat{Y}_{n+1} - ts \sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \overline{X})^2}{\sum (X_i - \overline{X})^2}} < Y_{n+1} < \hat{Y}_{n+1} + ts \sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \overline{X})^2}{\sum (X_i - \overline{X})^2}}$$

La cual posee (n-2) g.1. y con un nivel de significancia de $1-\frac{\alpha}{2}$

Ejemplo Ilustrativo:

Sea los datos sobre consumo privado y sus variables explicativas respectivas, para el periodo 1984-2002.

Año	**C Consumo Privado**	**X_1 YND**
1984	1785	2344
1985	1849	2401
1986	1909	2414
1987	1987	2669
1988	2122	2889
1989	2210	2807
1990	2244	2817
1991	2254	2755
1992	2082	2723
1993	2131	3139
1994	2236	3335
1995	2356	3442
1996	2376	3461
1997	2167	2925
1998	2209	3012
1999	2296	3060
2000	2593	3419
2001	2839	3800
2002	2501	3452

El modelo con las variables será: $C = \beta_1 + \beta_2 YND$

C = 528.78877 + 0.56 YND
t_c (2.84715) (9.16902) R^2 = 83.2
F = 84.05
X_{t+1}= 4000

$$(X'X)^{-1} = \begin{pmatrix} 3.009163208 & -0.000987868 \\ -0.000987868 & 0.00000033 \end{pmatrix}$$

$$C_{t+1} = \hat{\beta}_1 + \hat{\beta}_2(4000) = \begin{pmatrix} 1 & 4000 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 4000 \end{pmatrix}\begin{pmatrix} 528.7887 \\ 0.56393 \end{pmatrix}$$

C_{t+1} = 2784.5

El intervalo de confianza es:

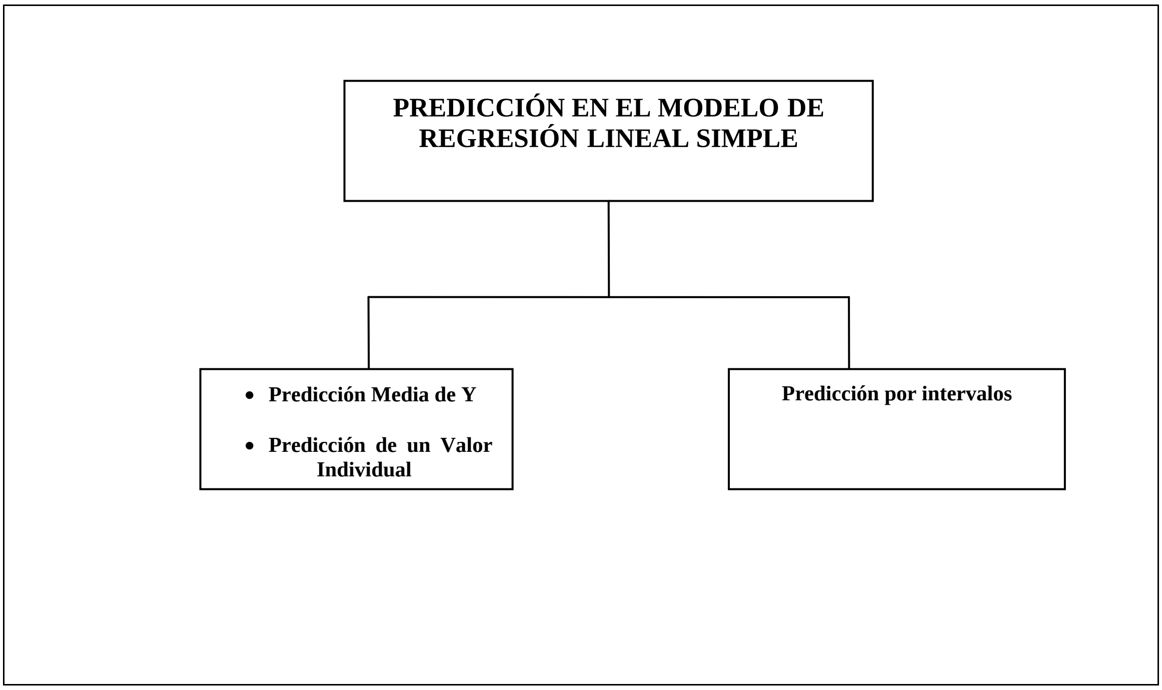
$$E(\hat{C}_{t+1}) \in \left\lfloor C_{t+1} \pm \sigma_{E(\hat{C}_{n+1})} t_{19-2,\alpha/2} \right\rfloor$$

Luego : $\hat{\sigma}_{E(\hat{C}_{t+1})} = \hat{\sigma}^2_{\hat{\beta}_C} = \sigma_{\mu} \sqrt{C'(X'X)^{-1}C}$ Dado: $t_{19-2,0.05/2} = 2.093$

$$\hat{\sigma}_{E(\hat{C}_{t+1})} = 107.0653 \sqrt{\begin{pmatrix} 1 & 4000 \end{pmatrix} \begin{pmatrix} 3.009163 & -0.000987 \\ -0.000987 & 0.00000 \end{pmatrix} \begin{pmatrix} 1 \\ 4000 \end{pmatrix}} = 78.6766$$

Entonces el intervalo de confianza será:

$$[2784.5 \pm 2.093(78.68)] = [261982, 294917]$$



LECCIÓN 2

PREDICCIÓN EN EL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

PREDICCIÓN

Es una de las aplicaciones del modelo de regresión por ejemplo, podría utilizarse para la simulación de medidas de política económica y también para estimar su impacto o para el diagnóstico. Predecir significa proyectar el comportamiento de la variable dependiente. En términos generales significa estimar el valor de "Y", mediante las variables independientes.

Modelo estimado: $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k = X'\beta$

a. La predicción puntual de Y_{n+1}

$$E(Y_{n+1}/X_{n+1}) = \hat{\beta}_1 + \hat{\beta}_2 X_{2,n+1} + ...\hat{\beta}_{k,n+1} X_{k,n+1}$$

Dado: $X'_{n+1} = \{1 \quad X_{2,n+1} \quad ....X_{k,n+1}\}$

Entonces $E(\hat{Y}_{n+1}) = E(X'_{n+1}\hat{\beta}) = X'_{n+1}E(\hat{\beta}) = X'_{n+1}\beta$

Así: $\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\beta_i})$ y como: $Y = X'_{n+1}\beta \Rightarrow \hat{Y} \sim N(Y, \sigma_Y)$

b. Intervalo de confianza de una predicción con un nivel de confianza α

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}^2_{\mu}\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \leq E(Y_{n+1}) \leq \hat{Y}_{n+1} + t_{\alpha/2}\sqrt{\hat{\sigma}^2_{\mu}\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

Es decir: $E(\hat{Y}_{n+1}) \in [X'_{n+1}\hat{\beta} \pm \sigma_{E(\hat{Y}_{n+1})} t_{n-k,\alpha/2}]$

Con n-k g.l. y con un nivel de significancia de $1 - \frac{\alpha}{2}$

.

LECCIÓN 3

CONDICIONES DE FIABILIDAD DE LA PREDICCIÓN

Para que la predicción obtenida sea fiable, se deben de cumplir las siguientes condiciones:

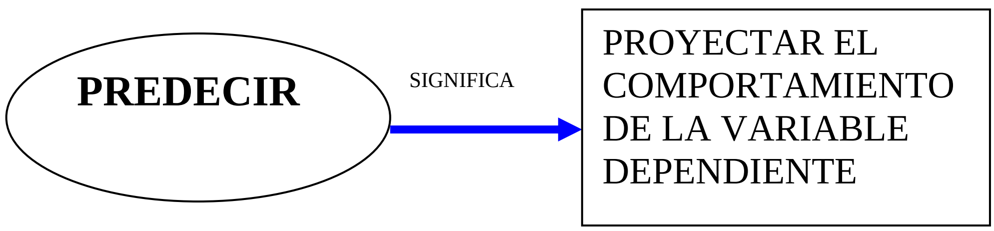
1. Que la relación lineal estimada entre Y_t, y las variables $X_2, X_3, \ldots, X_k$, se mantenga en el futuro

2. Que los coeficientes sean suficientemente estables como para que sus estimaciones obtenidas con la muestra actual sean una buena aproximación a los valores que se obtendrían al incorporar observaciones futuras.

3. Que se conozcan los valores futuros de las variables X_n, o que los modelos de predicción que utilicemos para dichas variables sean suficientemente fiables

4. Que el modelo lineal que se ha estimado esté correctamente especificado

5. Que el horizonte de predicción no sea muy lejano



LECCIÓN 4

ERROR DE PREDICCIÓN Y SU VARIANZA

1. ERROR DE PREDICCIÓN

El “error de predicción” de un periodo futuro, se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida:

$$\mu_{n+1} = Y_{n+1} - \hat{Y}_{n+1} = X_{n+1}\beta - X_{n+1}\hat{\beta} - \mu_{n+1}$$

Entonces, las fuentes del error de predicción serían:

a. El error en la estimación del vector β
b. El error en la predicción del vector X_{n+1}
c. El error estocástico inherente al modelo, μ_{n+1}

2. VARIANZA DE LA PREDICCION MEDIA

Se sabe que: $\hat{Y} = X'\hat{\beta} = X'(X'X)^{-1}X'Y$ por lo que: $\hat{Y}_{n+1} = X'_{n+1}\hat{\beta}$, y además los $\hat{\beta}$’s son estimadores insesgados: $E(\hat{Y}_{n+1}) = E(Y_{n+1} / X_{n+1}) = X'_{n+1}\hat{\beta}$

Con lo cual la varianza de la predicción es:

$$Var(\hat{Y}) = Var(X'_{n+1}\beta) = \hat{\sigma}^2_{\mu} X'_{n+1}(X'X)^{-1}X_{n+1}$$

Su desviación Standard es:

$$DS = \sqrt{\hat{\sigma}^2_{\mu} X'_{n+1}(X'X)^{-1}X_{n+1}}$$

3. VARIANZA DE LA PREDICCION INDIVIDUAL

Se desea predecir un valor individual correspondiente a $X = X_{n+1}$ su varianza es:

$$Var(Y / X_{n+1}) = \hat{\sigma}^2_{\mu}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]$$

Su desviación Standard es:

$$\hat{\sigma}_{Y_{n+1}} = \hat{\sigma}_{\mu}\sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}$$

4. - PREDICCIÓN POR INTERVALO :

Predicción del Valor Promedio

Construyendo en estadístico “t” para el valor promedio de la predicción se tendrá:

$$t = \frac{\hat{Y}_{n+1} - E(Y_{n+1})}{\hat{\sigma}_\mu \sqrt{X'_{n+1}(XX)^{-1}X_{n+1}}} \quad \sim \quad t_{n-k}$$

Luego:

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_\mu^2\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \le E(Y_{n+1}) \le \hat{Y}_{n+1} + t_{\alpha/2}\sqrt{\hat{\sigma}_\mu^2\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

Es decir: $E(\hat{Y}_{n+1}) \in [X'_{n+1}\hat{\beta} \pm \hat{\sigma}_{E(\hat{Y}_{n+1})} t_{n-k,\alpha/2}]$

Donde:

$$\hat{\sigma}_{E(\hat{Y}_{n+1})} = \hat{\sigma}_\mu \sqrt{X'_{n+1}(X'X)^{-1}X_{n+1}}$$

Predicción de un Valor en Particular

Construyendo un estadístico “t” para la predicción un valor particular Y_{n+1} se tendrá:

$$t = \frac{\hat{Y}_{n+1} - Y_{n+1}}{\hat{\sigma}_\mu \sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}} \quad \sim \quad t_{n-k}$$

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_\mu^2\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \le E(Y / X_{n+1}) \le \hat{Y}_{n+1} + t_{\alpha/2}\sqrt{\hat{\sigma}_\mu^2\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

Luego: $Y_{n+1} \in [X'_{n+1}\beta \pm \hat{\sigma}_{Y_{n+1}} t_{n-k,\alpha/2}]$

Donde : $\hat{\sigma}_{Y_{n+1}} = \hat{\sigma}_\mu \sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}$

LECCIÓN 5

EVALUACIÓN DE LA BONDAD PREDICTIVA DEL MODELO

La evaluación de la capacidad predictiva del modelo se hace en función de los datos pasados, tomando en cuenta los valores observados como estimados de la variable dependiente "Y". Suponiendo n periodos adicionales sobre los cuales se realiza la predicción, se puede evaluar la bondad de la predicción a través de las siguientes fórmulas:

1. RAIZ CUADRATICA MEDIA (rms):

Se determina mediante:

$$\text{rms} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2}$$

Donde: $\hat{Y}_t$: Valor estimado de Y_t

Y_t : Valor real observado

n : Número de periodos

La raíz cuadrática media, mide que tan bueno es el modelo para predecir. La rms, debe ser lo más pequeño posible para que el modelo sea bueno para la predicción.

2. COEFICIENTE DE THEIL (U)

Se define como:

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t\right)^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t\right)^2}}$$

Donde: $\hat{Y}_t$: Valor estimado de Y_t

Y_t : Valor real observado

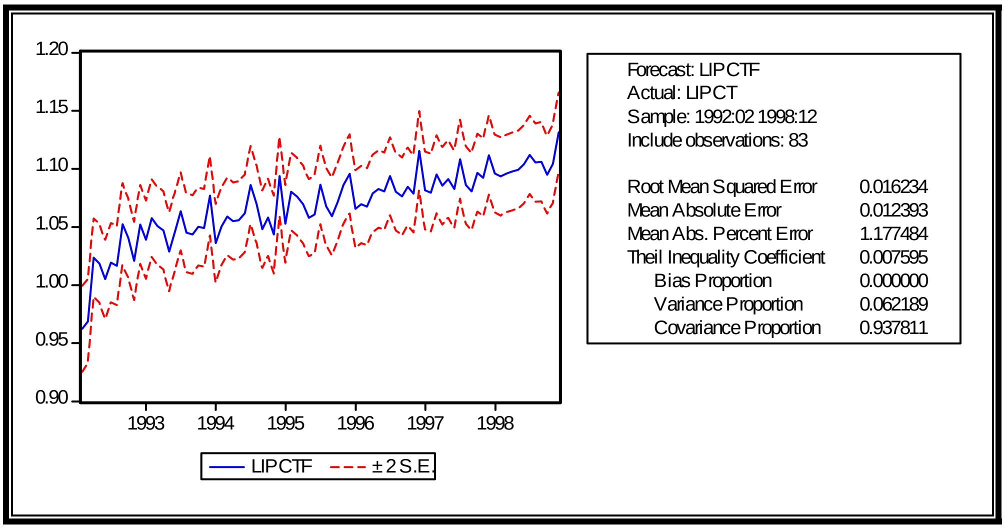
n : Número de períodos

El coeficiente de Theil, mide la calidad del modelo para predecir. Este coeficiente siempre caerá entre 0 y 1. Si U = 0, existe un ajuste perfecto y el modelo es bueno para predecir. En efecto, cuando los valores estimados de Y se aproximen más a los valores observados el valor del numerador se acercará a cero. Si U = 1, el modelo es muy malo para predecir.

Ejemplo Ilustrativo:

Volviendo al modelo de Inflación trabajado en la unidad anterior, se analizará la capacidad predictiva del mismo:

A continuación realizamos la predicción del modelo de inflación, con el fin de visualizar la bondad de estimación así como la fiabilidad de los pronósticos.



Sobre los datos históricos, encontramos que tienen un Coeficiente de Desigualdad de Theil = 0.007595. Este indicador trata de acotar el valor del indicador de bondad de predicción entre [0,1] y muestra si la correlación entre los valores observados en una predicción es alta o baja. En este caso el U es próximo a cero, con lo cual el modelo es útil al momento de realizar las prediciones

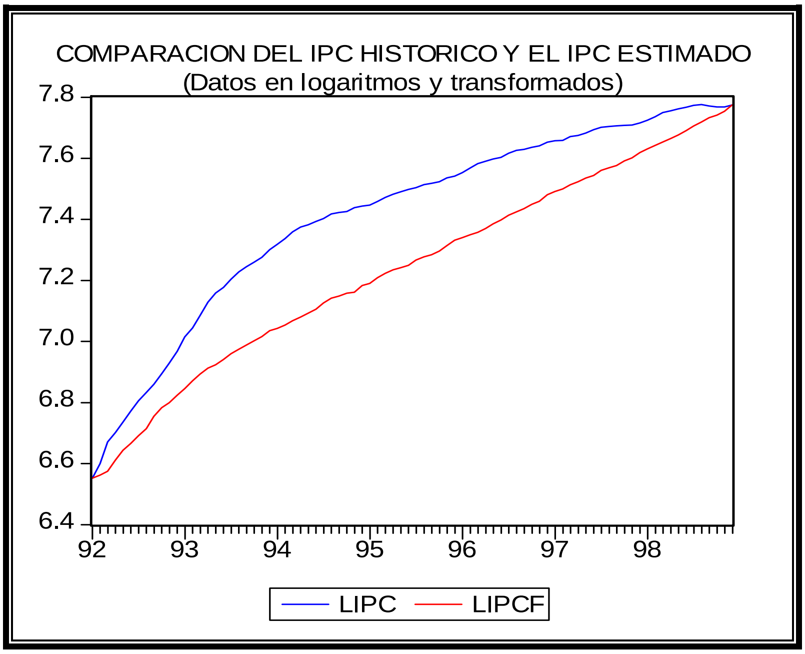
El Coeficiente de Theil se puede descomponer en:

Bias Proportion=0.000; El Bias proportion es usado para reconocer la presencia de algún error sistemático. En el modelo, no existe error sistemático.

Variance Proportion=0.062189; La Variance Proportion indica que tan bueno es el pronostico en representar la variabilidad de la variable real observada. Si es muy alto tiene una menor capacidad de replicar el comportamiento de la serie. En nuestro caso este indicador es bastante bajo.

Covariance Proportion=0.937811; Permite medir la parte no sistemática de los errores de predicción y su valor deberá ser la mayor parte del error total cometido. En nuestro caso, se cumple con lo anteriormente dicho por lo que su valor es considerado aceptable.

En el siguiente gráfico, se compara el IPC Histórico (LIPC) y el IPC Estimado (LIPCF) (para los datos en logaritmos y transformados), para el Modelo Final de la Inflación



En este caso se puede advertir la considerable aproximación que se ha logrado con los valores estimados.

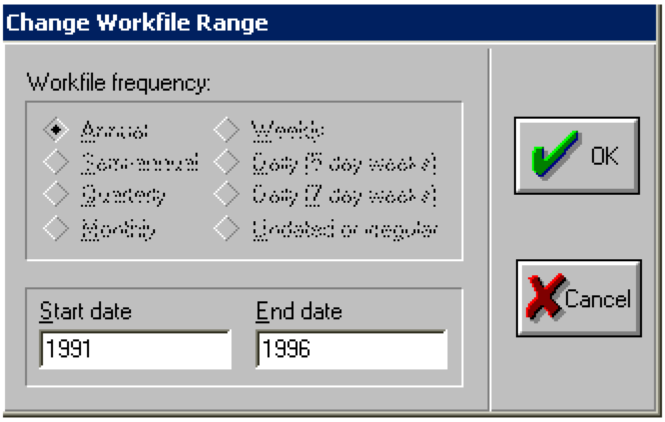
ANEXO

LA PREDICCIÓN Y EL E-VIEWS

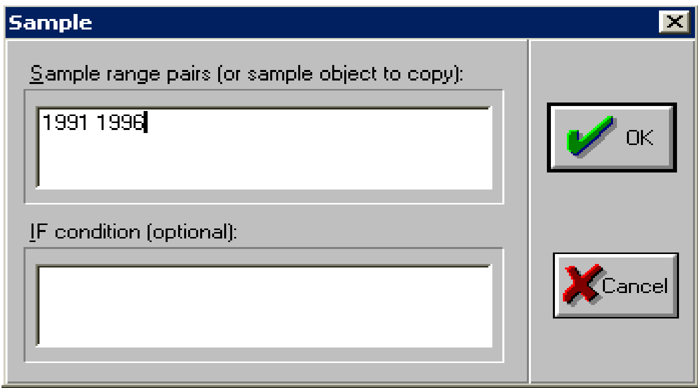
A. Predicción:

Hallar la predicción de la variable Y en el período t+1 sabiendo que en el período t+1 las variables X_1 y X_2 adoptan los valores de 6 y 8 respectivamente.

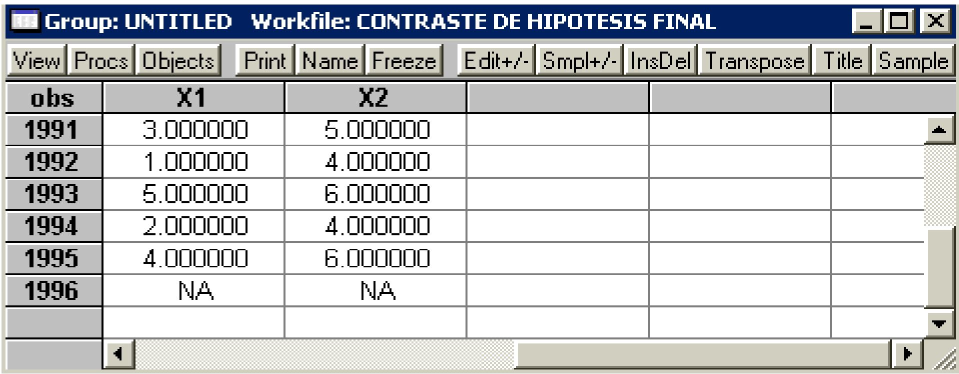
- El primer paso para predecir en el EVIEWS es expandir el rango del período de análisis. Estando en el workfile hacer clic en Procs, y escoger, haciendo clic, la alternativa Change Workfile Range, aparecerá la siguiente caja de dialogo (en donde se debe poner el nuevo rango)



- Luego se debe cambiar el tamaño de la muestra. Estando en el workfile hacer clic en el botón Sample (muestra). Aparecerá la siguiente caja de dialogo, en la cual se debe poner el nuevo tamaño de la muestra.

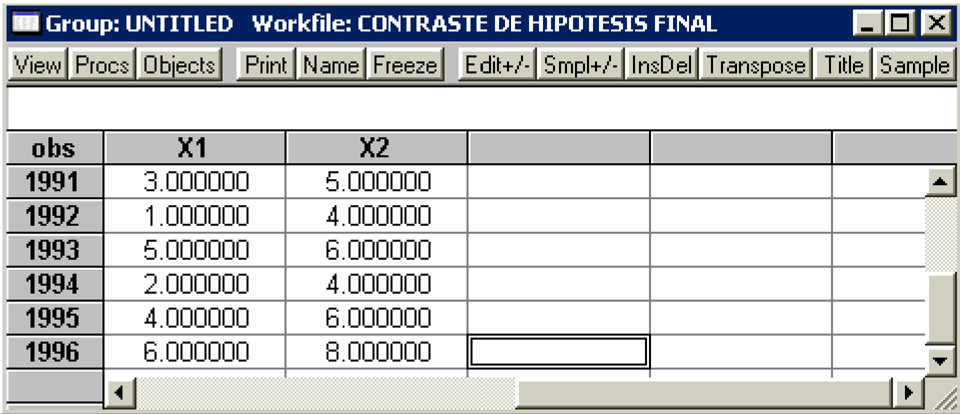


- El paso siguiente es rellenar los datos de las variables exógenas. En el workfile seleccionar las variables exógenas (haciendo clic conjuntamente con la tecla CONTROL), luego hacer clic derecho y escoger la alternativa Open/ as Group, aparecerá lo siguiente:



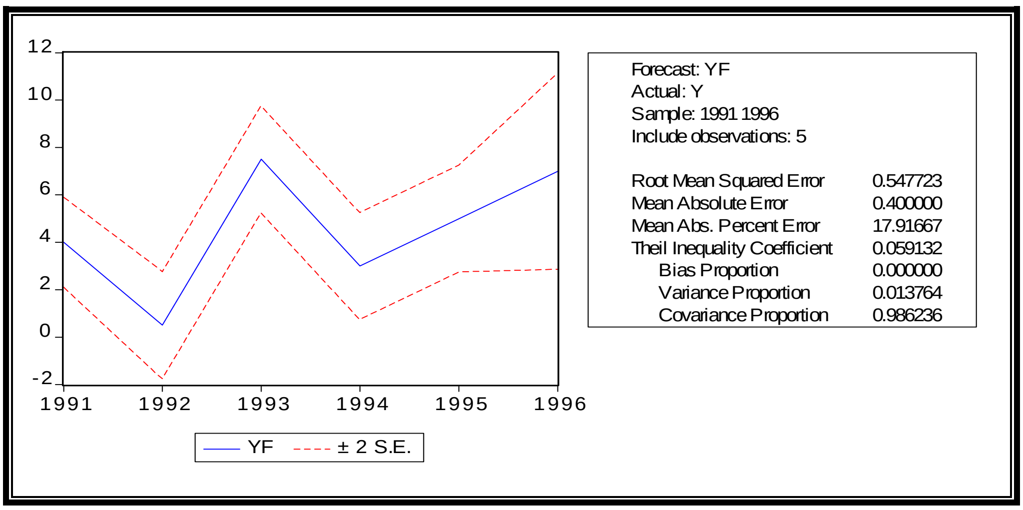
obs	X1	X2
1991	3.000000	5.000000
1992	1.000000	4.000000
1993	5.000000	6.000000
1994	2.000000	4.000000
1995	4.000000	6.000000
1996	NA	NA

- Hacer clic en el botón Edit +/- y escribir los valores de las variables exógenas para los años que se va a predecir (1996 para nuestro ejemplo):



obs	X1	X2
1991	3.000000	5.000000
1992	1.000000	4.000000
1993	5.000000	6.000000
1994	2.000000	4.000000
1995	4.000000	6.000000
1996	6.000000	8.000000

- Luego del llenado de los datos, hacer clic en el botón X de la esquina superior del cuadro anterior y hacer clic en YES.
- El paso siguiente es regresar al output de la regresión inicial y hacer clic en el botón Forecast y luego en OK. Aparecerá la siguiente gráfica, donde se observa el valor proyectado de la variable dependiente:



- Automáticamente se grabará los valores de la variable dependiente con un nuevo nombre (en nuestro caso será YF). Para ver los valores de la predicción hacer doble clic en YF, aparecerá los valores estimados y proyectados:

Last updated: 07/31/02 - 09:46	
Modified: 1991 1996 // eq01.forecast yf	
1991	4.000000
1992	0.500000
1993	7.500000
1994	3.000000
1995	5.000000
1996	7.000000

B. Construcción de un Intervalo de Confianza para el Predictor

a. El valor promedio de Y se encuentra en el intervalo comprendido entre:

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\sigma_{\mu}^{2} X'_{n+1}(X'X)^{-1}X_{n+1}} \leq E(Y/X_{n+1}) \leq \hat{Y}_{n+1} + t_{\alpha/2}\sqrt{\sigma_{\mu}^{2} X'_{n+1}(X'X)^{-1}X_{n+1}}$$

Donde $Y_{n+1} = 7$

Con 3 grados de libertad y un nivel de significancia del 5% se tiene que $t_{\alpha/2} = 3.182$ además:

$$DS = \sqrt{\hat{\sigma}_{\mu}^{2} X'_{n+1}(X'X)^{-1}X_{n+1}} = 1.8775$$

Reemplazando los datos, el valor promedio de Y se encuentra comprendido en el intervalo: [1.0258, 12.974]

b. El intervalo de confianza al 95% **para la predicción puntual** se calcula mediante la siguiente fórmula:

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \leq E(Y/X_{n+1}) \leq \hat{Y}_{n+1} + t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

Donde $Y_{n+1} = 7$

Con 3 grados de libertad y un nivel de significancia del 5% se tiene que $t_{\alpha/2} = 3.182$ además:

$$DS = \sqrt{\hat{\sigma}_{\mu}^{2}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]} = 2.06761$$

Reemplazando datos, tenemos que el intervalo de confianza para la predicción individual es: [0.42087, 13.579]

LABORATORIO

Ejercicio Aplicativo1

Caso Harina de Pescado, se tiene los siguientes modelos y los siguientes outputs:

MODELO 1: $VOLUMENX_t = \beta_1 + \beta_2 PRECIOX_t + \beta_3 IPCX_t + \mu_t$

MODELO 2: $VOLUMENX_t = \beta_1 + \beta_2 PRECIOX_t + \beta_3 IPCX_t + \beta_4 IPCPBI_t + \mu_t$

Donde:

VOLUMENX : Volumen de exportación de harina de pescado,
PRECIOX : Precio de exportación de la harina de pescado,
IPCX : Índice de precios de las exportaciones,
IPCPBI : Índice de precios del PBI

PERU: EXPORTACIONES Y PRECIOS DE HARINA DE PESCADO Y PRECIOS

AÑO	Harina de pescado MM US$	Precio US$/tm	IPC EXPORT	IPC PBI	Volumen MMTm
1970	303	162	0.000000643	0.000000359	1873
1971	267	153	0.000000588	0.000000384	1750
1972	219	144	0.000000592	0.000000414	1524
1973	138	395	0.000000842	0.00000047	348
1974	202	321	0.000001068	0.000000542	629
1975	168	216	0.000000996	0.000000668	781
1976	168	284	0.000001408	0.000000862	592
1977	184	422	0.00000224	0.00000117	436
1978	196	405	0.000004123	0.000001859	483
1979	256	390	0.000008366	0.000003225	657
1980	195	469	0.00001267	0.000005279	417
1981	141	448	0.000017008	0.000009029	315
1982	202	329	0.000025912	0.000015138	616
1983	80	387	0.000060568	0.000031389	205
1984	137	342	0.000127313	0.000066829	401
1985	118	233	0.00035821	0.000178604	508
1986	206	288	0.000457197	0.000308954	716
1987	229	308	0.000743838	0.0005632	741
1988	356	439	0.006599323	0.004107135	812
1989	411	372	0.14	0.11	1103
1990	339	310	9.44	6.75	1093
1991	441	393	33.3	31.9	1123
1992	427	430	55	53.9	993
1993	545	348	81.8	79.3	1568
1994	713	321	100	100	2221
1995	712	392	113.8	112.9	1816
1996	835	519	124.2	124.8	1610
1997	1031	535	136.1	134.3	1926
1998	392	588	127.8	142.9	666
1999	533	360	139.1	148.6	1482
2000	873	371	148.7	154	2352

1.- Analice el R^2 y el R^2 ajustado para ambos modelos ¿Qué se puede afirmar?

Resumiendo la bondad del ajuste para ambos modelos:

	Modelo 1	**Modelo 2**
R^2	0.732639	0.808625
R^2 ajustado	0.713542	0.787361

Podemos observar que como era de esperar el coeficiente de determinación ha crecido ya que se ha incluido a la variable Índice de precios del PBI, por lo que para tener una mejor idea acerca de la capacidad explicativa a partir del uso de las variables independientes consideradas es conveniente considerar el **R^2** ajustado, en este caso se puede apreciar que su valor no varia mucho de un modelo a otro por lo que la inclusión de la variable IP del PBI contribuye al entendimiento del volumen de exportaciones de harina de pescado, por lo tanto se puede afirmar que con la inclusión de esta última variable se logra una mejor explicación del modelo.

Resultados Modelo 1

Dependent Variable: VOLUMENX Method: Least Squares Sample: 1970 2000 Included observations: 31				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1920.016	220.8601	8.693359	0.0000
PRECIOX	-3.543931	0.642381	-5.516871	0.0000
IPCX	10.73698	1.249634	8.592100	0.0000
R-squared	0.732639	Mean dependent var		1024.419
Adjusted R-squared	0.713542	S.D. dependent var		615.0496
S.E. of regression	329.1857	Akaike info criterion		14.52289
Sum squared resid	3034170.	Schwarz criterion		14.66166
Log likelihood	-222.1047	F-statistic		38.36362
Durbin-Watson stat	1.013287	Prob(F-statistic)		0.000000

Resultados Modelo 2

Dependent Variable: VOLUMENX				
Method: Least Squares				
Date: 09/10/02 Time: 10:24				
Sample: 1970 2000				
Included observations: 31				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1835.397	192.0336	9.557687	0.0000
PRECIOX	-3.356909	0.556396	-6.033306	0.0000
IPCX	69.13566	17.86850	3.869137	0.0006
IPCPBI	-56.81575	17.35257	-3.274197	0.0029
R-squared	0.808625	Mean dependent var		1024.419

Adjusted R-squared	0.787361	S.D. dependent var	615.0496
S.E. of regression	283.6168	Akaike info criterion	14.25304
Sum squared resid	2171840.	Schwarz criterion	14.43807
Log likelihood	-216.9221	F-statistic	38.02798
Durbin-Watson stat	0.987144	Prob(F-statistic)	0.000000

2.- Pruebe la significancia de incorporar la variable IPCPBI al modelo (mediante un análisis de varianza parcial).

Para determinar la significancia de la inclusión de la nueva variable se realizará un análisis de varianza parcial. La dócima a realizar es:

H_0 : La variable IPCPBI no mejora el modelo,
H_1 : La inclusión de la variable IPCPBI mejora el modelo.

Modelo1: (n= 31)

SCT = $615.0496^2 \times (31\text{-}1)$	11348580.31
SCR	3034170
SCE = SCT – SCR	8314410.31

Modelo 2: (n=31)

SCT = $615.0496^2 \times (31\text{-}1)$	11348580.31
SCR	2171840
SCE = SCT – SCR	9176740.31

De los 2 modelos hallamos:

SCE(2) – SCE(1) = 9’176,740.31 – 8’314,410.31 = 862330

SCR(2) = 2’171,840

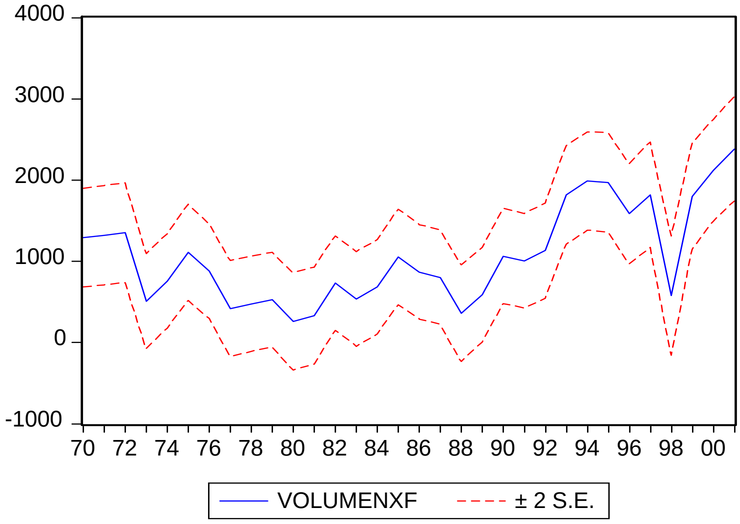
Además:
K = 3; s = 1; n= 31

$$F_C = \frac{862{,}330 \;/1}{2171{,}840 \;/27} = \frac{862{,}330}{80{,}438.51} = 10.72$$

La F de tablas ($F_{1,27,05}$) es 4.215, y como el F calculado es mayor al F de tablas, se concluye que se rechaza la hipótesis nula y por lo tanto la inclusión de la nueva variable mejora el modelo.

3.- Si las variables PRECIOX, IPCX y IPCPBI, aumentan en el año 2001 en 10%, 8% y 5% respectivamente. Hallar la predicción de la variable VOLUMENX para ese año.

Para obtener esos resultados se ha hecho uso del software E-Views, el cual permite obtener el gráfico siguiente:



Forecast: VOLUMENXF	
Actual: VOLUMENX	
Sample: 1970 2001	
Include observations: 31	
Root Mean Squared Error	264.6873
Mean Absolute Error	211.6487
Mean Abs. Percent Error	28.92377
Theil Inequality Coefficient	0.112648
Bias Proportion	0.000000
Variance Proportion	0.053055
Covariance Proportion	0.946945

4.- ¿Es el modelo bueno para predecir?

Al realizar la simulación sobre los datos históricos, encontramos un Coeficiente de Desigualdad de Theil = 0.112648.

Este indicador trata de acotar el valor del indicador de bondad de predicción entre [0,1]. Muestra si la correlación entre los valores observados en una predicción es alta o baja. En nuestro caso el U no es próximo a cero, con lo cual el modelo no puede ser utilizado para predecir.

El Coeficiente de Theil se puede descomponer en:

El Bias proportion; indica algún error sistemático. En nuestro caso su valor es cero (Bias Proportion=0.000;), por lo que no existe error sistemático.

La Variance Proportion; indica la habilidad del pronostico para replicar la variabilidad de la variable real observada. Si es muy alto tiene una menor capacidad de replicar el comportamiento de la serie. En nuestro caso este indicador es bastante bajo (Variance Proportion=0.053055).

El Covariance Proportion; Es usado para medir la parte no sistemática de los errores de predicción, su valor deberá ser la mayor del error total cometido, para este caso se observa un valor elevado con lo que cumple con lo anteriormente dicho (Covariance Proportion=0.946945).

Ejemplo Aplicativo:

Sean los datos siguientes sobre consumo privado, siendo las explicativas el ingreso nacional disponible, los precios relativos y las tasas de interés. Además considere que para los próximos 10 años las explicativas presentarán un incremento con relación a su año respectivo (disminución en el caso de la tasa de interés) del orden del 5%.

Año	Consumo Privado	YND	Precios Relativos	Tasas de Interés
1982	1785	2344	98.4	3.5
1983	1849	2401	98.4	1.6
1984	1909	2414	98.4	4.9
1985	1987	2669	100.5	-3.9
1986	2122	2889	104.0	-8.3
1987	2210	2807	104.1	-11.9
1988	2244	2817	103.0	-2.3
1989	2254	2755	101.4	-13.8
1990	2082	2723	101.7	-30.0
1991	2131	3139	100.0	-17.0
1992	2236	3335	99.7	-14.6
1993	2356	3442	100.2	-6.1
1994	2376	3461	96.7	-6.5
1995	2167	2925	96.4	-27.6
1996	2209	3012	94.4	-21.2
1997	2296	3060	92.0	-34.7
1998	2593	3419	95.5	-39.9
1999	2839	3800	97.4	-41.5
2000	2501	3452	97.8	-88.4

C : Consumo privado (Y)
YND: Ingreso nacional disponible(X_2)
PR : Precios relativos (X_3)
TI : Tasas de interés (X_4)

Modelo 1: $C = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

Pruebas Iniciales de Estabilidad

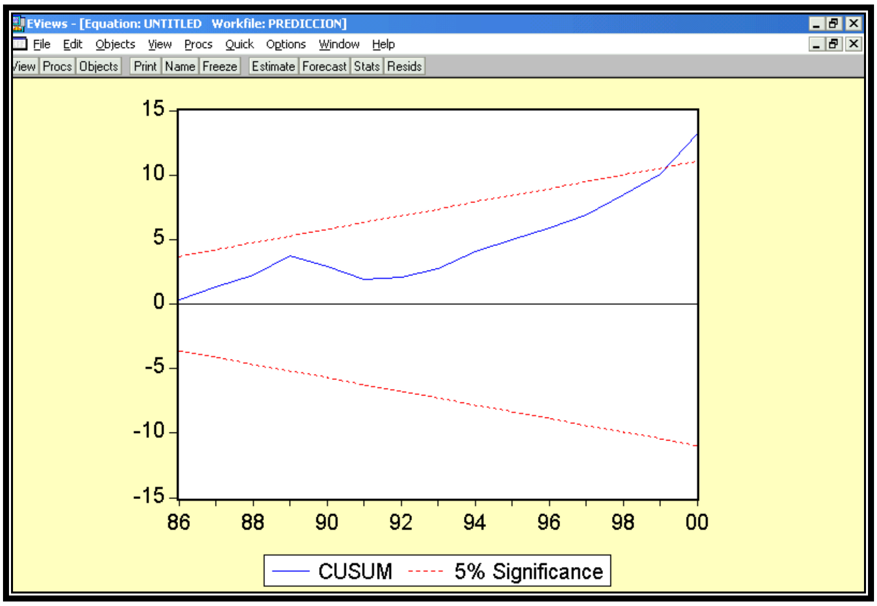
Para este modelo se procederá a verificar si se produce un cambio estructural, y contrastar la hipótesis de estabilidad de los parámetros en le periodo 1982-2000.

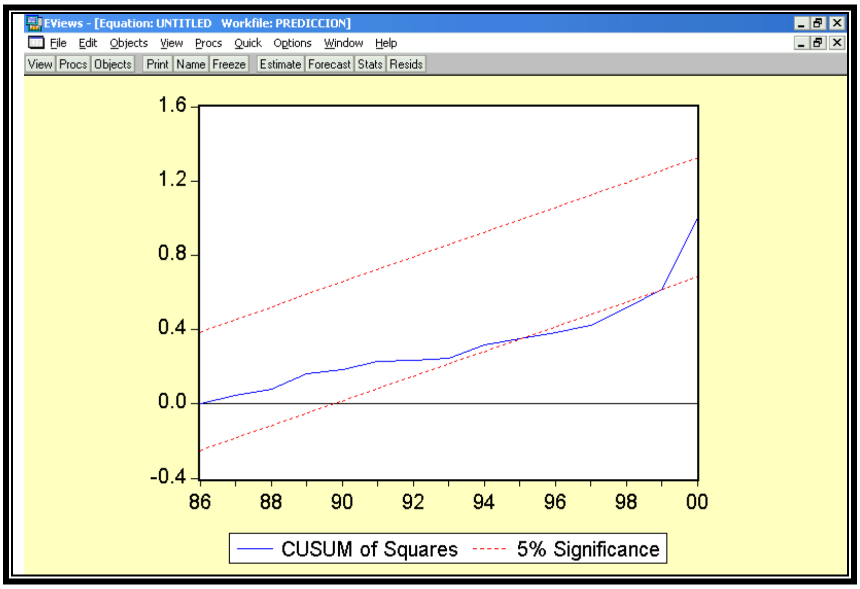
Mediante el Test de Chow, podemos verificar la presencia de cambio estructural, además de la estabilidad de parámetros.

Chow Breakpoint Test: 1995			
F-statistic	4.3224	Probability	0.0242
Log likelihood ratio	17.9475	Probability	0.0013

La Probabilidad asociada a la prueba de Chow, indica que no se puede rechazar la hipótesis nula de ausencia de cambio estructural en el año 1995, es decir que los parámetros no son estables.

Además de este test, se puede verificar la estabilidad de parámetros mediante el test de CUSUM y $CUSUM^2$, en donde ambos contrastan la hipótesis de que los parámetros son estables en el período de análisis.





Observamos en las gráficas que a continuación se muestran que la línea del estadístico CUSUM y $CUSUM^2$ salen fueran de las bandas permisibles. Entonces podemos afirmar que a un 95% de confianza, que los parámetros no son estables en el período de análisis, pudiéndose advertir además que a partir de 1994 se produce un crecimiento considerable en ambos gráficos, lo cual es otro indicio de inestabilidad del modelo.

Elaboración del Pronóstico del Modelo

A partir del modelo estimado se puede observar en la tabla siguiente que en general (significación conjunta) existe un buen ajuste del modelo, cosa que no ocurre para el caso individual a excepción del ingreso nacional disponible.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_2	0.506021	0.075663	6.687808	0.000
X_3	3.961643	8.403367	0.471435	0.6441
X_4	-2.212993	1.44228	-1.534372	0.1458
C	270.1128	887.1681	0.304466	0.765
R-squared	0.85504	Mean dependent var		2218.211
Adjusted R-squared	0.826049	S.D. dependent var		254.8705
S.E. of regression	106.3001	Akaike info criterion		12.35507
Sum squared resid	169495.6	Schwarz criterion		12.5539
Log likelihood	-113.3732	F-statistic		29.49238
Durbin-Watson stat	1.022951	Prob(F-statistic)		0.000002

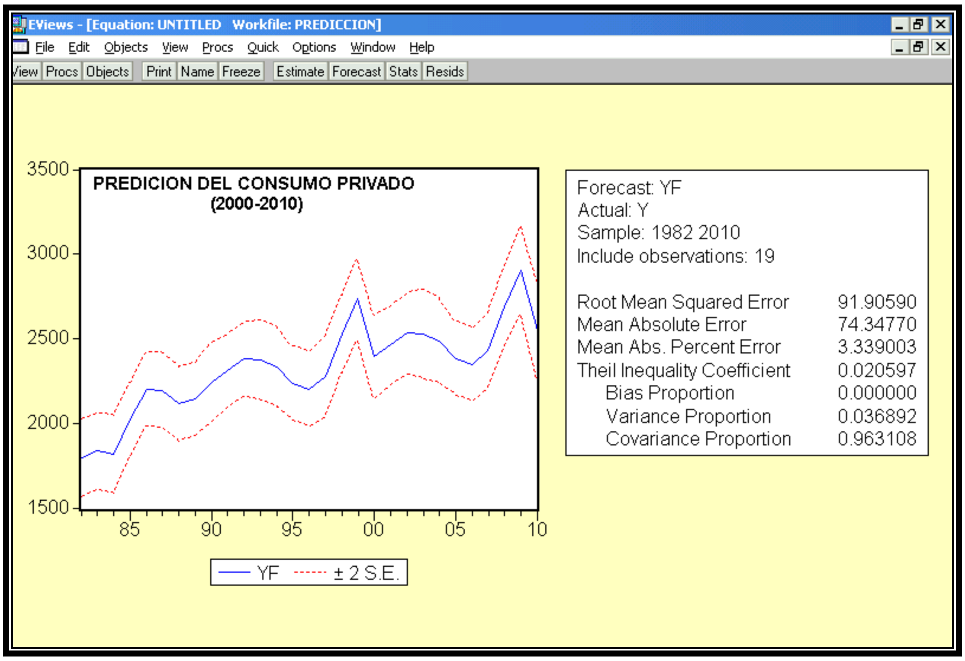
Para conocer la evolución futura del consumo, se dispone información acerca del comportamiento de las explicativas las cuales como se ha mencionado se incrementarán en un 5% en el transcurso de los 10 años siguientes, por ejemplo para el año 2001 los valores obtenidos corresponden a un incremento del 5% con relación al 1991. Los valores para estos 10 años siguientes serán:

Año	X_2	X_3	X_4
2001	3295.95	105	-17.85
2002	3501.75	104.685	-15.33
2003	3614.1	105.21	-6.405
2004	3634.05	101.535	-6.825
2005	3071.25	101.22	-28.98
2006	3162.6	99.12	-22.26
2007	3213	96.6	-36.435
2008	3589.95	100.275	-41.895
2009	3990	102.27	-43.575
2010	3002.108	112.1243	-33.075

Cálculo de las predicciones para el periodo 2000-2010:

Año	Y
2001	2393.40729
2002	2490.72176
2003	2529.90213
2004	2526.36767
2005	2289.35996
2006	2312.39422
2007	2359.28352
2008	2576.67014
2009	2790.72517
2010	2306.63335

Estos resultados obtenidos pueden representarse gráficamente, obteniéndose el gráfico siguiente:



Para analizar la capacidad predictiva del modelo se procederá a comparar los resultados obtenidos con las predicciones obtenidas con un modelo alternativo:

$$C = \beta_1 + \beta_2 X_2$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_2	0.549756	0.081411	6.75283	0.000
C	590.0311	243.1407	2.426706	0.0266
R-squared	0.728438	Mean dependent var		2218.211
Adjusted R-squared	0.712463	S.D. dependent var		254.8705
S.E. of regresión	136.6678	Akaike info criterion		12.77228
Sum squared resid	317527.4	Schwarz criterion		12.8717
Log likelihood	-119.3367	F-statistic		45.60071
Durbin-Watson stat	0.449732	Prob(F-statistic)		0.000003

De los resultados de la tabla anterior se concluye que al considerar únicamente al ingreso nacional disponible se concluye que se presenta una mejoría en la bondad de ajuste, ya que se observa que aunque posea un menor R^2 es más cercano a su R^2 ajustado respectivo en comparación con el otro modelo, lo que permite concluir que la incorporación de las variables precios relativos y la tasa de interés no mejoran la explicación del modelo.

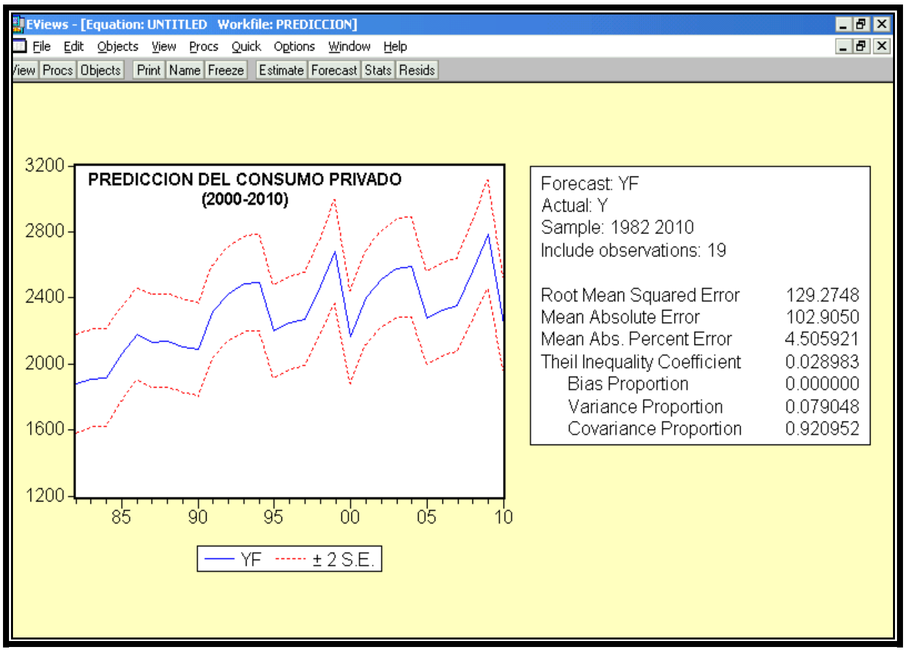
Para poder obtener una predicción a partir de este nuevo modelo será necesario tener los valores de los regresores para el periodo 2001-2010, razón por la cual se considerará el supuesto hecho inicialmente acerca del incremento en las explicativas en 5%.Los valores para estos 10 años siguientes serán:

Año	X_2
2001	3295.95
2002	3501.75
2003	3614.1
2004	3634.05
2005	3071.25
2006	3162.6
2007	3213
2008	3589.95
2009	3990
2010	3002.108

Predicciones para este modelo:

Año	Y
2001	2401.99978
2002	2515.13959
2003	2576.90469
2004	2587.87232
2005	2278.46958
2006	2328.6898
2007	2356.39751
2008	2563.62808
2009	2783.55801
2010	2240.45807

Estos resultados obtenidos pueden representarse gráficamente, obteniéndose el gráfico siguiente:



Comparando los estadísticos siguientes, se puede observar inicialmente que el más adecuado es el primer modelo pues se observa que le coeficiente de theil es cercano a cero, sin embargo esta situación también ocurre en el otro modelo lo que obliga a tomar en cuenta los elementos que lo constituyen.

Los dos modelos poseen una misma cantidad de la proporción del sesgo lo que nos permite advertir la no existencia de algún error sistemático.

En este caso se observa que el modelo1 es el que representa de mejor manera la variabilidad en el consumo privado, aunque en ambos casos sus valores son considerablemente bajos

En ambos casos se observa que sus Covariance Proporcion corresponden a la mayor parte del error total cometido, siendo el mayor en el modelo1 por lo que indica que a partir de éste es posible recoger una mayor parte del error no sistemático cometido.

En lo referente a la descomposición del error cuadrático, sin embargo, las conclusiones no son las mismas, ya que recae la mayor parte en la proporción de la varianza en el segundo modelo.

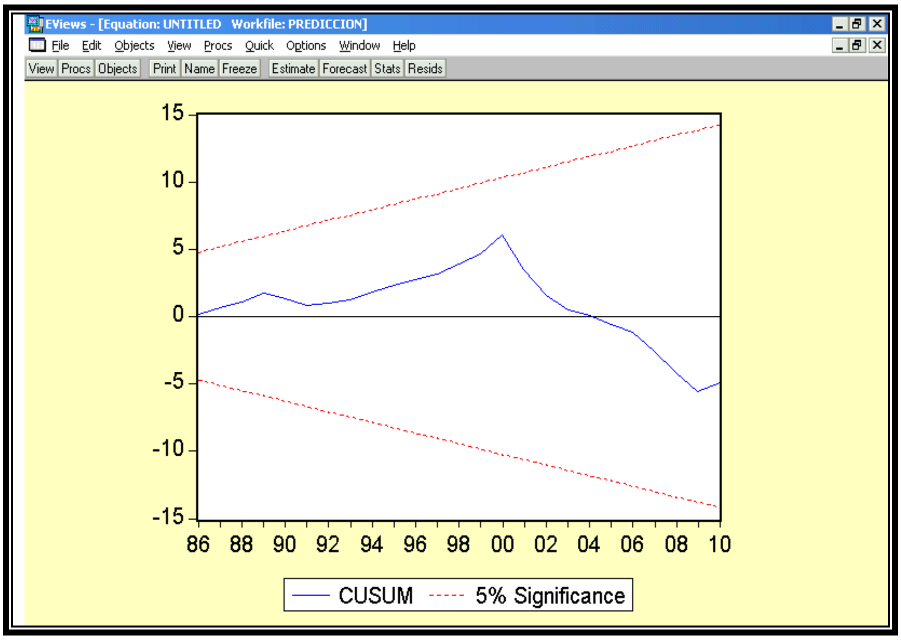
	Modelo 1	Modelo 2
Root Mean Squared Error	91.9059	129.2748
Mean Absolute Error	74.3477	102.9050
Mean Abs. Percent Error	3.33900	4.505921
Theil Inequality Cefficient	0.02060	0.028983
Bias Proporcion	0.00000	0.000000
Variante Proporcion	0.036892	0.079048
Covariance Proporcion	0.963108	0.920952

Pruebas de Estabilidad luego del Pronóstico

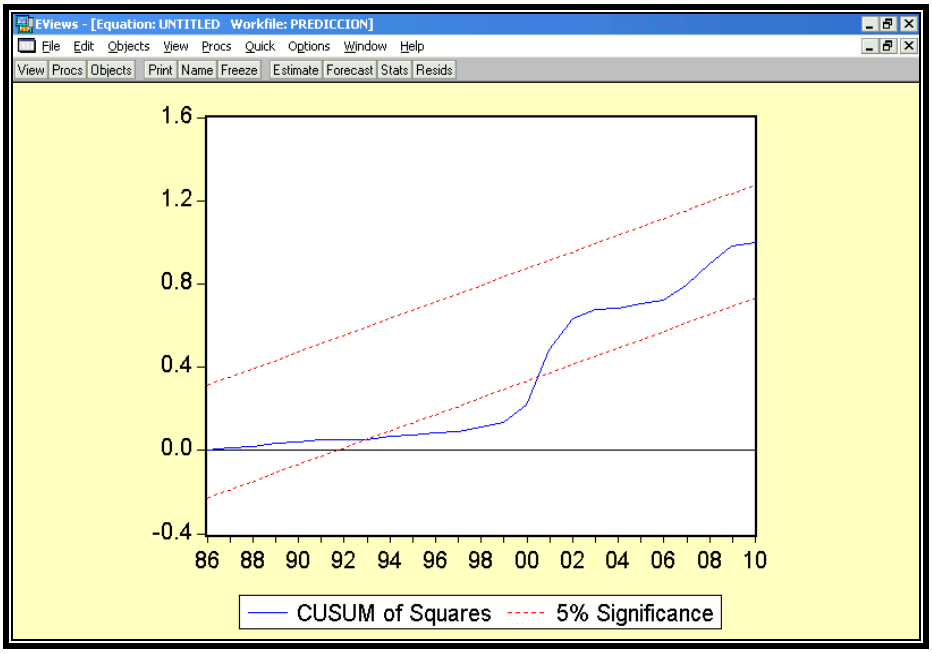
Debido a las consideraciones iniciales (incremento en 5% de las variables explicativas) el comportamiento de los modelos no cambiará en los pronósticos, ya que presentará una estabilidad en los parámetros.

A partir de las pruebas de CUSUM , CUSUM^2 y Chow se corroborará tal situación.

Prueba CUSUM



Prueba CUSUM^2



Prueba de Predicción de Chow

Los resultados mostrados en la tabla indican que se rechaza la hipótesis nula de estabilidad estructural, concluyéndose que las últimas observaciones muestrales disponibles no presentan cambios respecto a las anteriores.

Chow Forecast Test: Forecast from 2005 to 2010			
F-statistic	1.457301	Probability	0.245362
Log likelihood ratio	10.97864	Probability	0.089039

Ejercicio de autoconocimiento

¿Porqué hacer una predicción en un Modelo lineal?

	SI	NO	NO SÉ
1. Porque es útil para predecir valores futuros de la variable dependiente			
2. Porque permitirá dar un buen pronóstico de las variables de decisión			
3. Porque analizando el pasado se podrá predecir el futuro de la empresa			
4. Para que exista una eficacia en la ejecución de los proyectos			
5. Para predecir el comportamiento de los agentes económicos			
6. Porque permite hacer recomendaciones para el comportamiento de las variables dependientes			
7. Para definir cuál de los valores futuros de la variable es más probable			
8. Para que el responsable de la gestión empresarial pueda cumplir una meta trazada			
9. Para resolver problemas y solucionarlos			
10. Porque al usar la predicción se podrá tomar la precauciones			

CALIFICACION

Puntuar con un punto cada respuesta "SI".
Si obtienes de 1 - 3 puntos tienes pocas expectativas de hacer una buena predicción empresarial.
Si tienes entre 4 – 7, tienes buenas expectativas de hacer una buena predicción empresarial.
Y si tienes entre 8 – 10, denotas excelentes expectativas de hacer una buena predicción empresarial.

RESUMEN

La predicción como una técnica estadística define cuál de los posibles valores futuros de la variable es más probable.

Predicción del valor medio condicional de Y correspondiente a un determinado $X/X=X_{n+1}$ que es el punto de la línea de regresión poblacional.

Predicción Media de Y: $E(Y/X_{n+1}) = \hat{\beta}_1 + \hat{\beta}_2 X_{n+1}$

Predicción de un Valor Individual: $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_{n+1}$

La predicción media o de un valor individual de Y para el caso puntual coinciden:
$E(Y/X_{n+1}) = \hat{Y}$

La fórmula de **predicción por intervalos promedio** de Y_i es:

$$\hat{Y}_{n+1} - ts_{\hat{Y}_{n+1}} < Y_{n+1} < \hat{Y}_{n+1} + ts_{\hat{Y}_{n+1}}$$

La fórmula de **predicción por intervalos individual de** Y_i **, la fórmula es:**

$$\hat{Y}_{n+1} - ts\sqrt{1+\frac{1}{n}+\frac{(X_{n+1}-\bar{X})^2}{\sum(X_i-\bar{X})^2}} < Y_{n+1} < \hat{Y}_{n+1} + ts\sqrt{1+\frac{1}{n}+\frac{(X_{n+1}-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$$

El “error de predicción” de un periodo futuro, se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida:

$$\mu_{n+1} = Y_{t+a} - \hat{Y}_{n+1} = X_{n+1}\beta + X_{n+1}\hat{\beta} + \mu_{n+1}$$

La varianza de la predicción media es:

$$V(\hat{Y}) = V(X'_{n+1}\beta) = \sigma^2_{\mu} X'_{n+1}(X'X)^{-1}X_{n+1}$$

Su desviación Standard es:

$$DS = \sqrt{\hat{\sigma}^2_{\mu} X'_{n+1}(X'X)^{-1}X_{n+1}}$$

La varianza de predicción individual es:

$$\mathrm{var}(Y/X_{n+1}) = \hat{\sigma}^2_{\mu}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]$$

Su desviación Standard es:

$$\sigma_{Y_{n+1}} = \sigma_{\mu}\sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}$$

Predicción por intervalo:

- Del valor Promedio

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \le E(Y_{n+1}) \le \hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

$$\hat{\sigma}_{E(\hat{Y}_{n+1})} = \sigma_{\mu}\sqrt{X'_{n+1}(X'X)^{-1}X_{n+1}}$$

- De un valor en Particular

$$\hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]} \le E(Y/X_{n+1}) \le \hat{Y}_{n+1} - t_{\alpha/2}\sqrt{\hat{\sigma}_{\mu}^{2}\left[1 + X'_{n+1}(X'X)^{-1}X_{n+1}\right]}$$

$$\sigma_{Y_{n+1}} = \sigma_{\mu}\sqrt{1 + X'_{n+1}(X'X)^{-1}X_{n+1}}$$

La evaluación de la capacidad predictiva del modelo predicción se hará a través de las siguientes fórmulas:

Raíz Cuadrática Media **(rms):** $rms = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2}$

La rms debe ser lo más pequeño posible para que el modelo sea bueno para la predicción.

Coeficiente De Theil **(U):**

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t\right)^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t\right)^2}}$$

Este coeficiente siempre caerá entre 0 y 1. Si U = 0, existe un ajuste perfecto y el modelo es bueno para predecir. Si U = 1, el modelo es muy malo para predecir.

EXPLORACIÓN ON LINE

1. Predicción en regresión lineal simple. Estimación de las medias condicionadas. Predicción de una observación.
   http://www.udc.es/dep/mate/estadistica2/indice.html

2. La predicción que realiza el modelo lineal. Ejemplos aplicativos
   http://ftp.medprev.uma.es/libro/node42.htm

3. Fuentes de predicción
   http: //uhu.es/45132/ficheros_datos

LECTURA

Predicción en regresión lineal simple

Hay dos objetivos básicos en el ajuste de un modelo de regresión:

- Conocer la relación existente entre la variable respuesta y las variables regresoras. En el caso de la regresión lineal simple se estima la mejor recta de regresión que relaciona la variable Y con la variable X y se cuantifica la importancia de dicha relación por medio del coeficiente de correlación, r.

- Utilizar el modelo de regresión ajustado para "predecir" el valor de la variable respuesta Y cuando la variable regresora toma un valor determinado, $X = x_t$.

En esta sección se estudia este segundo objetivo. Esto es, estimada la recta de regresión, ¿cómo predecir el valor de Y sabiendo que la variable regresora toma el valor $X = x_t$?

Ante esta pregunta, se deben distinguir dos situaciones diferentes:

Estimar la media de la distribución condicionada de $Y/X = x_t$: $E(Y / X = x_i) = m_t$

Se quiere responder a preguntas del tipo: "¿cuál es el gasto medio en material informático de las empresas que tienen unos ingresos globales de 300 millones anuales?".

Predecir el valor de la variable respuesta en un individuo de la población en estudio del que se sabe que $X = x_t$. Esto es, predecir un valor de la variable condicionada $Y/X=x_t$

Se quiere responder a preguntas del tipo: "La empresa MEGA tiene unos ingresos anuales de 300 millones, ¿cuál será el gasto en material informático de esta empresa?".

Fuente: Juan Vilar Fernández 2003

ACTIVIDADES

1. ¿Qué es predicción en el modelo de regresión lineal?

2. Para el caso puntual ¿Qué sucede con la predicción media o de un valor individual de Y?

3. ¿Qué trata de encontrar la predicción por intervalos?

4. Defina con sus propias palabras el "error de predicción" de un periodo futuro

5. Mencione las fórmulas que sirven para evaluar la bondad de la predicción

AUTOEVALUACIÓN

Encierra en un círculo la letra que contenga la alternativa correcta.

1. Si el modelo escogido confirma la hipótesis, se puede utilizar para predecir los valores futuros de la variable ....Y, o de pronóstico, con base en el valor futuro conocido o esperado de la variable X...................

   a. Explicativa-independiente
   b. Dependiente-explicativa
   c. Dependiente-dependiente
   d. Explicativa-explicativa

2. Predecir significa proyectar el comportamiento de la ....

   a. Variable independiente
   b. Variable aleatoria
   c. Variable dependiente
   d. Ninguna variable

3. El ....de un periodo futuro, se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida

   a. Predicción
   b. Comportamiento
   c. Modelo
   d. Error de predicción

4. La raíz cuadrática media, mide que tan bueno es el modelo para predecir La rms debe ser ....para que el modelo sea bueno para la predicción.

   a. lo más pequeño posible
   b. lo más grande posible
   c. igual a 0
   d. menor que 1

5. El coeficiente de Theil, mide la calidad del modelo para predecir; si U = 0, existe un ajuste perfecto y el modelo es ....

   a. Malo para predecir
   b. Regular para predecir
   c. Bueno para predecir
   d. Ninguna de las anteriores

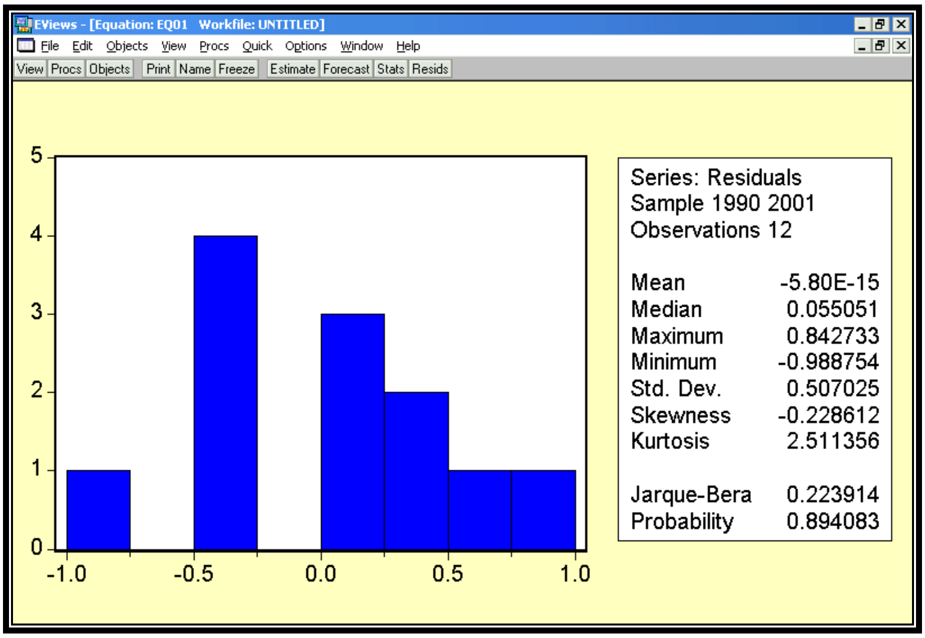
6. Los datos siguientes corresponden al IPC (base de 1994 = 100), los cuales se han obtenido principalmente a partir de los índices de precios del sector agropecuario (IP agro), referido a la gasolina (IP gasolina) y de la maquinaria y equipo importada (IPME).

Considere que para los próximos 5 años las explicativas del IPC presentarán un incremento con relación a su año respectivo del orden del 3%.

A continuación se mostrarán cálculos realizados con el fin de hallar un pronóstico del IPC para el periodo 2002-2006.

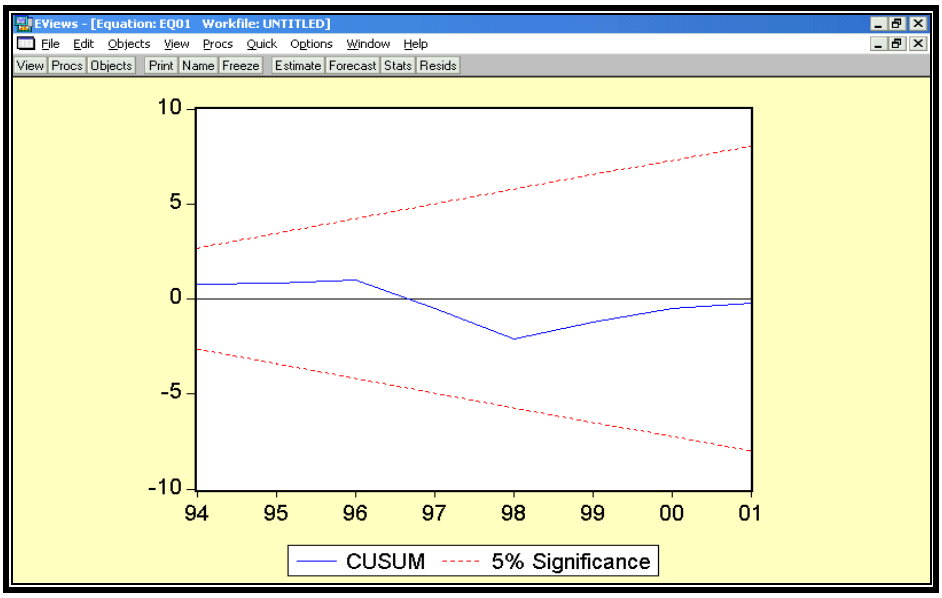
obs	IP agro	IP gasolina	IPME	IPC
1990	1974.150	1679.180	1216.140	139.2130
1991	1998.310	1678.370	1236.470	140.9317
1992	2050.350	1679.870	1243.010	142.7904
1993	2045.300	1678.230	1248.620	143.6630
1994	2067.410	1681.250	1255.890	144.5156
1995	2086.560	1693.240	1268.440	145.2821
1996	2107.280	1699.060	1274.180	146.1945
1997	2145.680	1695.860	1282.690	146.5811
1998	2119.610	1709.730	1306.470	145.7925
1999	2084.570	1729.090	1316.770	145.3043
2000	2064.420	1731.280	1329.070	145.3485
2001	2050.260	1718.130	1344.370	146.2498

Prueba de Normalidad de Residuales

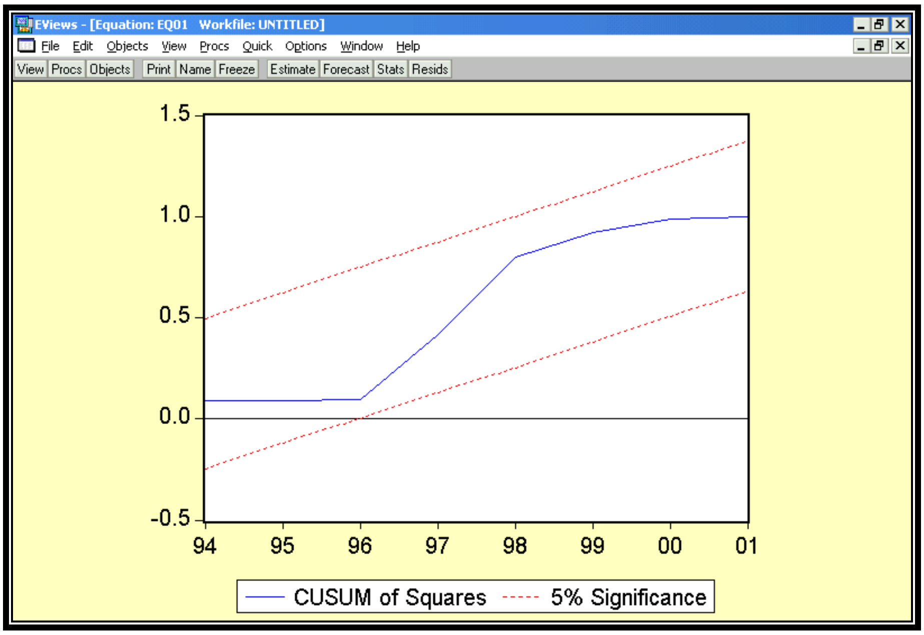


Pruebas de Cambio Estructural

CUSUM



CUSUM^2



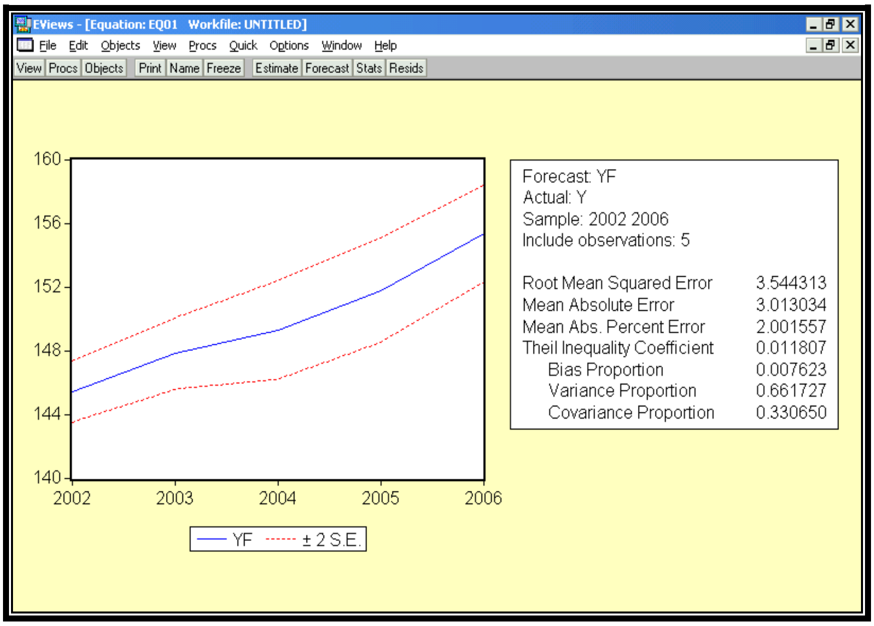
Estimación del Modelo

Dependent Variable: IPC				
Method: Least Squares				
Date: 10/13/04 Time: 09:55				
Sample: 1990 2001				
Included observations: 12				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
IP agro	0.029805	0.004395	6.780784	0.0001
IP gasolina	-0.05007	0.025055	-1.998403	0.0807
IPME	0.049434	0.01325	3.730817	0.0058
C	104.6292	29.86548	3.50335	0.008
R-squared	0.951114	Mean dependent var		144.3222
Adjusted R-squared	0.932782	S.D. dependent var		2.293126
S.E. of regresión	0.594525	Akaike info criterion		2.059094
Sum squared resid	2.82768	Schwarz criterion		2.220729
Log likelihood	-8.354562	F-statistic		51.88237
Durbin-Watson stat	1.296708	Prob(F-statistic)		0.000014

Prueba de Estabilidad de Chow

Chow Breakpoint Test: 1996			
F-statistic	6.936682	Probability	0.043625
Log likelihood ratio	24.85794	Probability	0.000054

Gráfico del Pronóstico y su Evaluación de su capacidad Predictiva



Identifique la alternativa que considere correcta:

a. El modelo utilizado para la estimación del pronóstico no es adecuado.
b. Los errores poseen una distribución normal.
c. De acuerdo a las pruebas CUSUM y CUSUM^2 y Chow, el modelo planteado no es estable en el periodo de 1990-2001,
d. Se observa una estabilidad en pronóstico del IPC para el periodo 2002-2006.

7. Para el periodo 2001-2010, el banco central de reserva del Perú desea estimar el valor de la emisión primaria, para lo cual ha considerado conveniente considerar al circulante (X_1) y las reservas bancarias (X_2).

   A continuación se muestran los cálculos realizados con el fin de hallar un pronóstico para el periodo 2001-2010.

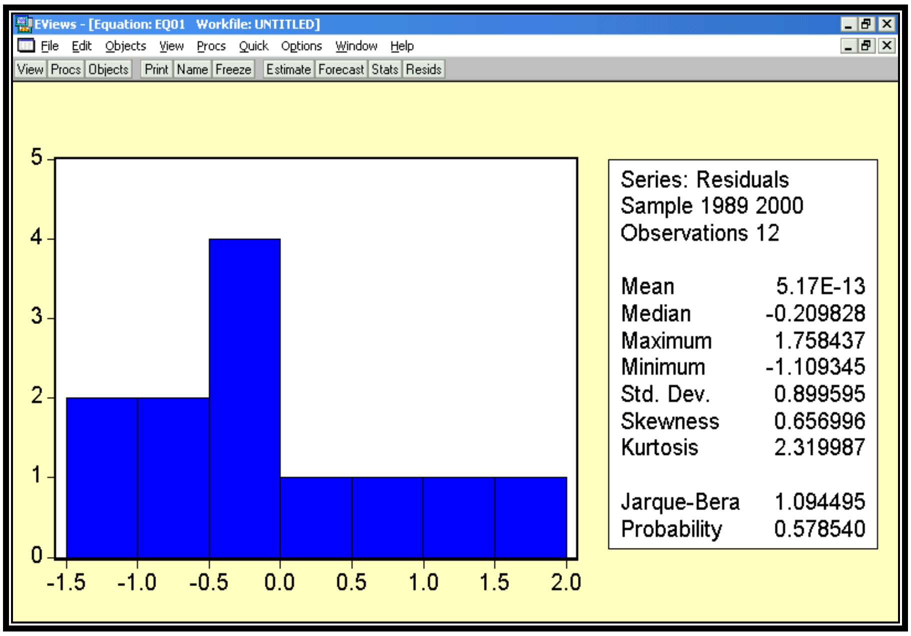
   Estimación del Modelo

Dependent Variable: Y
Sample: 1989 2000
Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X_1	1.000815	0.001631	613.6934	0.000
X_2	1.008059	0.005051	199.5635	0.000
C	6.017994	5.004424	1.202535	0.2598

R-squared	0.999982	Mean dependent var	3494.25
Adjusted R-squared	0.999978	S.D. dependent var	213.3572
S.E. of regression	0.994539	Akaike info criterion	3.039244
Sum squared resid	8.901978	Schwarz criterion	3.160471
Log likelihood	-15.23546	F-statistic	253119.5
Durban-Watson stat	1.912998	Prob(F-statistic)	0.0000

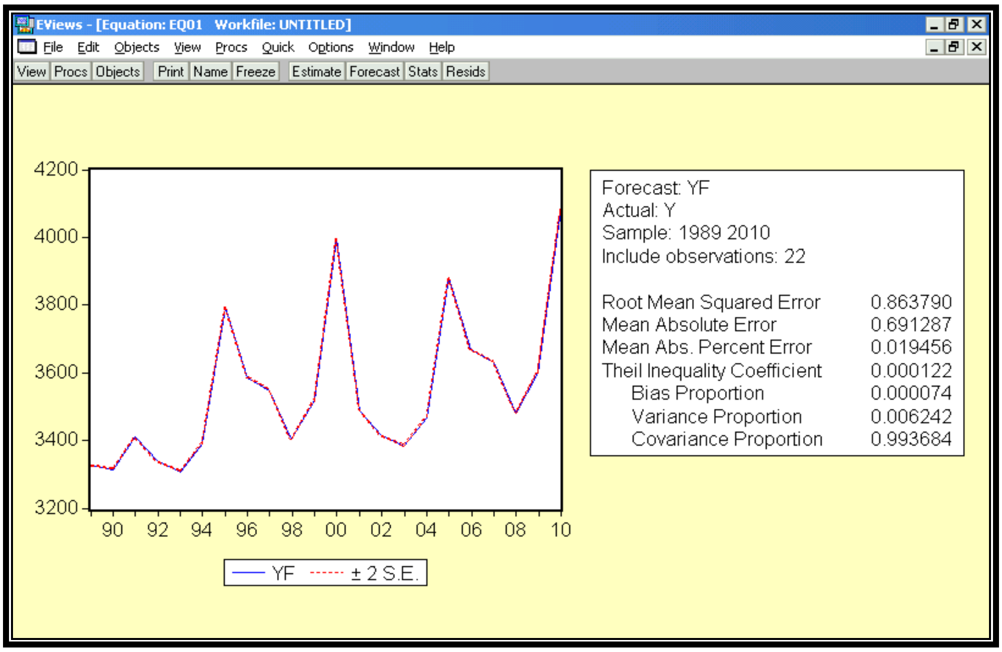
   Prueba de Normalidad de Residuales



Prueba de Estabilidad de Chow

Chow Breakpoint Test: 1995			
F-statistic	0.646615	Probability	0.613046
Log likelihood ratio	3.361613	Probability	0.339159

Gráfico del Pronóstico y su Evaluación de su capacidad Predictiva



Identifique la alternativa que considere correcta:

a. El modelo utilizado para la estimación del pronóstico no es adecuado.
b. Los errores poseen una distribución normal.
c. De acuerdo a las prueba de Chow, el modelo planteado es estable en el periodo de 2001-2010.
d. No Se observa una estabilidad en pronóstico para el periodo 2001-2010.

RESPUESTAS DE CONTROL

1. b, 2. c, 3. d, 4. a, 5. c, 6. d, 7.d