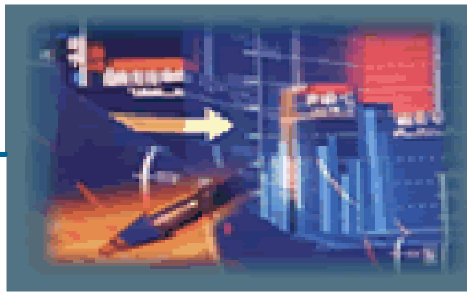
## UNIDAD V

# Levantamiento de Supuestos

**UNIDAD V**

# ¿Qué es levantamiento de supuestos?



**"La Estadística demuestra que lo "bien hecho" siempre es mejor que "lo bien dicho"**

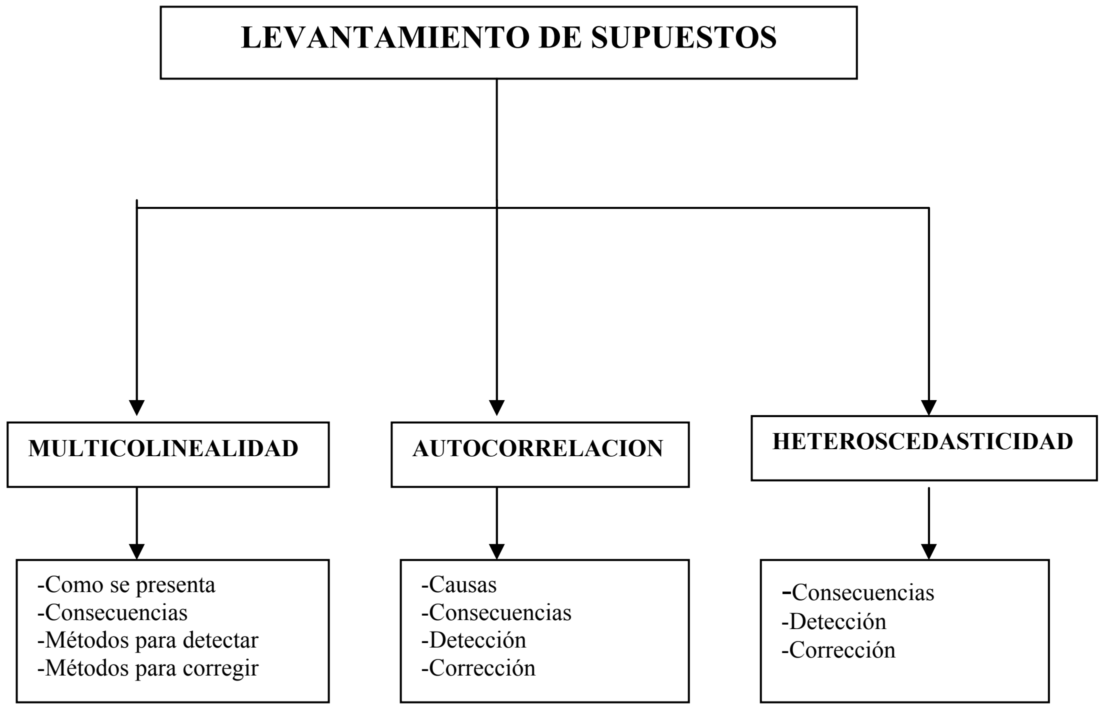
**I Ching**

- **¿Qué es Multicolinealidad? ¿Cómo se presenta? ¿Cuáles son sus consecuencias y métodos para detectarlo? ¿Cuáles son los métodos para corregirla?**
- **¿Qué es la autocorrelación? ¿Cuáles son las causas de su procedencia y sus consecuencias? ¿Cuáles son los métodos para detectarlo?**
- **¿Qué es heteroscedasticidad, sus consecuencias y su detección?**

## LECCIÓN 1

# INFERENCIA EN EL MODELO LINEAL

## ESQUEMA CONCEPTUAL




## COMPETENCIAS A LOGRAR

| CONCEPTUAL | PROCEDIMENTAL | ACTITUDINAL |
|---|---|---|
| Explica los conceptos de multicolinealidad, autocorrelación y heterocedasticidad, como se detecta y los métodos para corregirlos | Identifica y aplica los métodos para corregir la multicolinealidad, la autocorrelación y la heterocedasticidad en los modelos. | Asume como importante el uso de estos métodos para resolver problemas en los estudios de los modelos. |

**CONCEPTOS –CLAVE**
Multicolinealidad, Autocorrelación y Heterocedasticidad

# LECCIÓN 1

## MULTICOLINEALIDAD

La multicolinealidad es un problema que aparece cuando las variables explicativas de un modelo están correlacionadas entre sí. En la práctica sucede cuando uno introduce variables que siendo importantes para explicar el modelo se encuentran relacionadas entre sí, por lo que no será necesario incluir a todas ellas para explicar el comportamiento de la variable dependiente, bastará con incluir a unas cuantas explicativas las cuales representarán el comportamiento de las otras. Por ejemplo el modelo del consumo en función de ingreso y la riqueza, estas variables igualmente están relacionadas pero es importante en algunos casos incluir las dos.

### 1. PROBLEMAS DE LA MULTICOLINEALIDAD

- Los parámetros se encuentran distorsionados debido a que los valores de los mismos se han mezclado, luego no se pueden utilizar para medir el verdadero impacto de las variables explicativas en el modelo.
- Los estadísticos "t" no son significativos porque las desviaciones estándar son innecesariamente más grandes que aquellas que se obtendrían con una previa corrección de la multicolinealidad.

### 2. COMO SE PRESENTA LA MULTICOLINEALIDAD

Supongamos en el siguiente modelo:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3$$

El cual no presenta el problema de multicolinealidad; si se incorpora la variable $X_4$, donde $Cov(X_4, X_2) \neq 0$.

O si no: $X_4 = f(X_2, X_3)$, $X_2 = f(X_4, X_3)$, $X_3 = f(X_4, X_2)$

Por lo tanto, con la inclusión de $X_4$ al modelo estaremos ante la presencia de **multicolinealidad**.

Lo anteriormente mencionado puede servir para entender el significado de lo que es la multicolinealidad, sin embargo no es tan fácil detectarla. Una manera de hacerlo sería a partir de los estadísticos "t" los cuales ante la presencia de multicolinealidad tienden a no ser significativos.

### 3. CLASES DE MULTICOLINEALIDAD

a. Multicolinealidad Perfecta $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu_i$ $\quad X_3 = kX_2$

b. Multicolinealidad Imperfecta $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu_i$ $\quad X_3 = kX_2 + \mu_i$

## 4. CONSECUENCIAS DE LA MULTICOLINEALIDAD

### a. Multicolinealidad Perfecta

Se presenta cuando un modelo presenta la forma:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu_i \qquad X_3 = kX_2$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (X'X) = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 \end{bmatrix}$$

Reemplazando: $X_3 = k\, X_2$ tendremos:

$$(X'X) = \begin{bmatrix} n & \sum X_{i2} & k\sum X_{i2} \\ \sum X_{i2} & \sum X_{i2}^2 & k\sum X_{i2}^2 \\ \sum X_{i3} & \sum X_{i2}X_{i3} & k^2\sum X_{i2}^2 \end{bmatrix}$$

De donde se obtendrá que det(X'X)= 0, por lo tanto no podrá estimarse $\hat{\beta}$ ya que (X'X) no tiene inversa.

### b. Multicolinealidad Imperfecta

En este caso el det(X'X) es muy próximo a cero (nunca es cero), por lo que los parámetros se pueden estimar. Sin embargo los valores de la matriz se vuelven muy pequeños, lo que hace que las desviaciones estándar de los estimadores se vuelvan pequeñas. Entonces, la situación a darse será:

$R^2$ alto (en E-Views se observa un R-squared considerablemente alto).

Al analizar la significancia de las variables en forma conjunta en el modelo se tendrá que $F_c > F_t$ (en E-Views se representa este hecho por el valor de "Prob (F-statistic)" menor que el nivel de significancia considerado).

Al analizar la significancia de cada una de las variables usadas en el modelo se tendrá que $t_c < t_{n-k,\alpha/2}$ (tabla), donde: $t_c = \dfrac{\hat{\beta}_i}{\hat{\sigma}_{\beta i}}$ (en E-Views el valor de las columnas "Prob" en todas las variables son mayores que el nivel de significancia considerado).

## 5. MÉTODOS PARA DETECTAR LA MULTICOLINEALIDAD

### a. Regla práctica

Comparando con un mismo nivel de significancia se observa:

- $R^2$ alto
- $F_c$ es > del $F_t$,
- Las pruebas "t" de significancia individual no son significativos
- Los parámetros van cambiando de signo al incorporar nuevas variables

**b. Evaluación con el coeficiente de determinación $R^2$**

Sea el modelo: $Y = \beta_1 + \beta_2X_2 + \beta_3X_3 + ....... + \beta_iX_i + ....... \beta_kX_k + \mu$

Si el $R^2_{\text{con } X_i} \cong R^2_{\text{sin } X_i}$ entonces se pueden dar dos posibilidades:

- $X_i$ no explica a Y.
- El efecto de $X_i$ ya esta incorporado en las otras variables.

En el primer caso habría que hacer alguna regresión $Y = f(X_i)$ y observar el $R^2$:

Si el $R^2$ es bajo entonces, efectivamente el $X_i$ no explica Y, es decir $Y \neq f(X_i)$.

Si el $R^2$ es alto entonces, el efecto de $X_i$ ha sido incorporado en el modelo.

**c. Cálculo de la matriz de correlación**

A partir del cálculo de la matriz de correlaciones de las variables independientes se podrá tener una idea inicial acerca del grado de asociación lineal entre cada par de regresores, en el caso de que alguno de sus valores sea cercano a ±1 será un indicativo de la presencia de una multicolinealidad imperfecta.

**d. Significación individual y conjunta de los regresores**

Por lo general cuando se presenta un problema de multicolinealidad aproximada los regresores de manera individual no son significativos, pero al considerarlos de manera conjunta son significativos, esto se debe a que la multicolinealidad aproximada influye en la significación individual de los regresores pero no en su significación conjunta.

## 6. METODOS PARA CORREGIR LA MULTICOLINEALIDAD

**a. Devolver $X_i$ al término de perturbación.**
Omitir del modelo una de las variables colineales; si es que el modelo tiene muchas variables explicativas. Pero al omitir una variable se podría incurrir en un sesgo de especificación o error de especificación (especificación incorrecta del modelo).

**b. Datos nuevos o ampliando el tamaño de la muestra.**
A veces aumentando el tamaño de la muestra, se puede atenuar el problema de la colinealidad (relación entre variables).

**c. Utilizando información a priori.**
Supóngase que se considera el modelo: $Y_i = \beta_1 + \beta_2X_{2i} + \beta_3X_{3i} + \mu_i$

Donde:
Y: Consumo; $X_2$: Ingreso; $X_3$: Riqueza

Como se sabe las variables: ingreso y riqueza; tienden a ser altamente colineales. Pero supóngase que a priori, se cree que $\beta_3 = 0.10\beta_2$. Se puede efectuar entonces la siguiente regresión:

$$Y_i = \beta_1 + \beta_2 X_{2i} + 0.10\beta_2 X_{3i} + \mu_i$$
$$Y_i = \beta_1 + \beta_2 X_i + \mu_i$$

Donde: $X_i = X_{2i} + 0.10X_{3i}$. Una vez que se ha obtenido $\hat{\beta}_2$, se puede estimar $\hat{\beta}_3$ a partir de la relación postulada entre $\beta_2$ y $\beta_3$.

Se tiene el siguiente modelo: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \mu_i$, el cual presenta el problema de multicolinealidad. Para corregir se desea mantener 2 variables explicativas, se intuye la correlación lineal $\beta_2 = 3\beta_3$, por lo tanto $\beta_3 = 0.33\beta_2$.

Realizando las operaciones necesarias, la ecuación de regresión será:

$$Y_i = \beta_1 + \beta_2 X_{2i} + 0.33\beta_2 X_{3i} + \mu_i$$
$$Y_i = \beta_1 + \beta_2 X_i$$

Donde $X_i = X_{2i} + 0.33X_{3i}$

**d. Transformación de Variables.**

Supóngase que se tiene información de series de tiempo sobre el gasto de consumo, el ingreso y la riqueza. Una razón para la alta multicolinealidad entre el ingreso y la riqueza es que en el tiempo las dos variables tienden a moverse en la misma dirección. Una forma de minimizarla, es proceder de la siguiente manera.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \mu_t \qquad \text{... (1)}$$

Si la relación anterior se cumple en el periodo "t", también debe cumplirse en el periodo "t-1" puesto que el origen del tiempo es arbitrario. Por consiguiente se tiene:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,\,t-1} + \beta_3 X_{3,\,t-1} + \mu_{t-1} \qquad \text{... (2)}$$

Si se resta (1) – (2), se obtiene:

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2,\,t-1}) + \beta_3 (X_{3t} - X_{3,\,t-1}) + \varepsilon_t \qquad \text{... (3)}$$

Donde: $\varepsilon_t = \mu_t - \mu_{t-1}$

A esta última ecuación se le conoce como la forma en primeras diferencias porque no se está corrigiendo la regresión sobre las variables originales, sino sobre las diferencias de los valores sucesivos de dichas variables.

Con este procedimiento se reduce la severidad de la multicolinealidad. Sin embargo, la transformación crea algunos problemas adicionales. El término de error $\varepsilon_t$ puede no satisfacer uno de los supuestos del modelo clásico de regresión lineal, pues las perturbaciones están serialmente correlacionadas.

Tres transformaciones alternativas son:

**Realizar las regresiones entre las variaciones en términos relativos**.
Las variaciones porcentuales nos muestran la dinámica en el corto plazo, y no necesariamente podrían tener un patrón de multicolinealidad

**Transformar a logaritmo natural cada una de las variables**.
El comportamiento se transforma a una forma de tipo geométrica o exponencial a partir del cual manteniendo las relaciones entre las variables, no necesariamente pueden tener multicolinealidad las variables explicativas.

**Combinando las variables que ocasionan la multicolinealidad**, a través de la elaboración de indicadores.

Por ejemplo una relación ingreso riqueza, es decir que porcentaje de la riqueza representa el ingreso, es igualmente una alternativa.

**Ejercicio Ilustrativo 1:**

En el modelo planteado anteriormente para la inflación[1]: $LIPC = \beta_1 + \beta_2 CIRPROM + \beta_3 LTC$, para detectar el denominado "caso clásico" de multicolinealidad, debemos observar simultáneamente lo siguiente:

- "t" estadísticos no significativos.
- $R^2$ alto.
- Signos de los parámetros contrarios a los esperados.

Un indicador adicional de la posible existencia de multicolinealidad es la existencia de un coeficiente de correlación muy alto entre las variables exógenas.

Los resultados de nuestra estimación nos dicen que:

- Todas las pruebas "t" de significancia individual son "significativas".
- El $R^2$ es alto. ($R^2$ =0.9937)
- Los signos de los parámetros son los esperados.

Dado que no se satisfacen los tres requisitos simultáneamente, concluimos que no se presenta, en este modelo, el caso clásico de multicolinealidad.

Un segundo método consiste en hallar la matriz de correlaciones de las variables. Si la correlación entre dos variables es cercana a 1, entonces podría existir multicolinealidad.

[1] Ver Ejercicio Aplicativo 1 de la Unidad III

| | LCIRPROM | LTC |
|---|---|---|
| **LCIRPROM** | 0.000181 | -0.000330 |
| **LTC** | -0.000330 | 0.000708 |

De la matriz de correlaciones del modelo de inflación confirmamos la no presencia de multicolinealidad.

**Ejercicio Ilustrativo 2[2]:**

**Aplicación del Análisis multivariado en el análisis de regresión: la Elaboración de Indicadores Sintéticos a fin de corregir el problema de multicolinealidad.**

Supongamos que se quisiera relacionar los problemas de la salud en función a las condiciones del medio ambiente. ¿Como relacionar temas multidimensionales tanto para las variables explicativas como las variables dependientes?

Un trabajo realizado con información al año 2001 a nivel departamental para Perú permitió analizar las relaciones existentes entre variables asociadas al medio ambiente con aquellas referidas a la salud de la población.

Habiéndose en primer lugar, evaluado las **variables de salud** se consideró finalmente la incorporación de las siguientes 5 variables:

- Prevalencia de la diarrea en niños menores de 5 años.
- Tos y respiración rápida en las 2 últimas semanas (%).
- Tasa de Mortalidad Infantil por cada mil niños menores de 1 año.
- IRA (por cada mil niños menores de 5 años).
- Malaria

Se decidió elegir estas variables porque la **matriz de correlaciones entre sus indicadores simples fue significativa**. En efecto, de las correlaciones entre las variables se pueden deducir importantes conclusiones respecto a los indicadores. Se puede apreciar que la "Prevalencia de la diarrea en niños menores de 5 años" presenta correlación positiva con la "Tasa de Mortalidad Infantil por cada mil niños menores de 1 año" (0.548), es decir, que en las ciudades donde aumenta el la prevalencia de diarrea en menores de 5 años, esto es por cada mil niños, también aumentaría la tasa de mortalidad infantil.

Es de mencionar que el indicar sintético de salud se elaboró de tal forma que cuanto menor sea éste, mejor era la situación, en el campo de la salud, de un determinado departamento.

Habiéndose evaluado las **variables de medio ambiente** se consideró finalmente la incorporación de las siguientes variables:

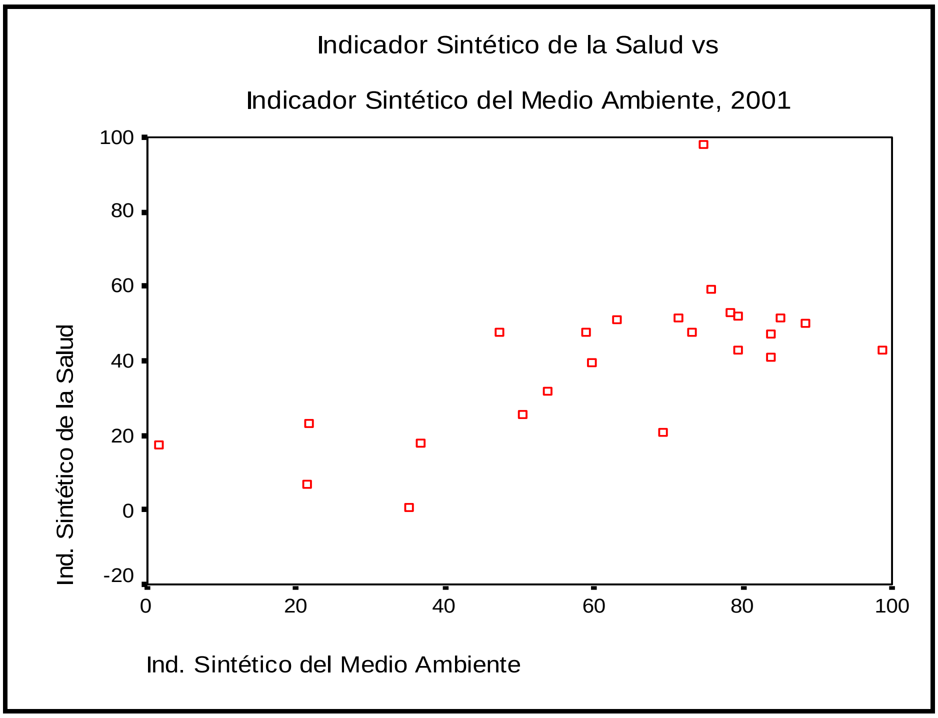
[2] Análisis de la Relación entre Indicadores de Salud y Medio Ambiente

- Vivienda con hacinamiento.
- Parque automotor por cada 10 000 habitantes.
- Densidad Poblacional (personas por km2).
- Población Rural %.
- Uso de Kerosene
- Uso de Leña y Carbón
- Vivienda sin servicio higiénico
- Vivienda sin servicio de agua por red publica

La decisión de elegir estas variables, se tomó porque la matriz de correlaciones entre sus indicadores simples fue significativa. En efecto, de las correlaciones entre las variables se pueden deducir importantes conclusiones respecto a los indicadores. Se puede apreciar que el "Uso de leña y carbón" presenta correlación positiva con "Viviendas con hacinamiento" (0.614), es decir, que en las ciudades donde existe mayor presencia de viviendas con hacinamiento, el uso de la leña y carbón aumenta. Otra correlación importante se aprecia entre el "Uso de kerosene" y la "Población rural (%)" (-0.65), es decir que en ciudades donde la población rural aumenta, el uso de kerosene disminuiría. Por otro lado tenemos la correlación entre el "Uso de la leña y carbón" y "La población rural (%)" (0.88), es decir que a mayores ciudades con población rural (%), el consumo de leña y carbón aumenta.

Con los indicadores sintéticos de ambos ítems se realizaron un estudio de sus correlaciones, y se pudo apreciar que ambos indicadores tenían una correlación directa.

En el gráfico se aprecia una asociación lineal directa entre el indicador sintético de la salud y el indicador sintético del medio ambiente, cuyo coeficiente de correlación es igual a 0.66, es decir cuando el indicador del medio ambiente tienda a subir entonces el indicador de la salud tendrá una tendencia ascendente.



La visualización gráfica de los datos departamentales nos muestra una relación positiva entre los indicadores sintéticos de la salud y el medio ambiente. Es decir, cuanto mayor es el Indicador de medio ambiente, mayor será el indicador de la salud.

A continuación se corre el modelo entre 2 variables que sintetiza el tema de la salud y el medio ambiente:

| **Departamentos** | **Indicador Sintético de la Salud** | **Indicador Sintético del Medio Ambiente** | **Departamentos** | **Indicador Sintético de la Salud** | **Indicador Sintético del Medio Ambiente** |
|---|---|---|---|---|---|
| Lima 1/ | 17.3 | 1.6 | S. Martin. | 51.7 | 71.3 |
| Tacna. | 7.1 | 21.5 | Puno | 47.6 | 73.1 |
| Arequipa. | 23.0 | 21.6 | Loreto. | 98.1 | 74.6 |
| Moquegua. | 0.5 | 35.2 | Ucayali. | 59.1 | 75.7 |
| Ica | 17.9 | 36.6 | Cusco. | 53.0 | 78.3 |
| Lambayeque. | 47.8 | 47.2 | Apurimac. | 42.9 | 79.3 |
| La Libertad. | 25.5 | 50.3 | Pasco. | 52.2 | 79.4 |
| Tumbes. | 31.8 | 53.7 | Amazonas. | 41.0 | 83.6 |
| Junin. | 47.6 | 58.9 | Ayacucho. | 47.2 | 83.7 |
| Ancash. | 39.6 | 59.6 | Cajamarca. | 51.3 | 85.1 |
| Piura. | 51.2 | 63.1 | Huanuco. | 50.0 | 88.4 |
| M. Dios | 20.6 | 69.2 | Huancavelica. | 42.7 | 98.6 |

El modelo de regresión hallado, utilizando estos indicadores es:
$\hat{Y}_i = 6.571 + 0.543 * X_i$ Donde:

Y: Indicador Sintético de la Salud
X: Indicador Sintético del Medio Ambiente

Analizando la significancia del modelo, se obtuvo que la probabilidad asociada a la constante es mayor al nivel de significancia de 5% (0.466). Por lo tanto no es significativo para el modelo. En cambio, el indicador sintético de la salud es significativo para el modelo, ya que su probabilidad asociada (0.000) es inferior al 5%. Además, se tiene que el $R^2$ del modelo es bajo (0.405) y a la vez el F-estadístico (0.000) indica la significancia conjunta del modelo.

Evaluando nuevamente el modelo, pero sin el intercepto, se obtiene el siguiente modelo:
$\hat{Y}_i = 0.635 * X_i$

Las salidas de este nuevo modelo muestran que el modelo es significativo, ya que la prueba F (0.000) es menor al nivel de significancia de 5%, aunado a una prueba t que indica la significancia de la variable, así como un $R^2$ alto (0.887).

En conclusión, el estado de salud de las personas está influenciado fuertemente por un conjunto de factores relacionados con el medio ambiente.

**Ejercicio Ilustrativo 3:**

Según el siguiente modelo tenemos a las cantidades demandadas de productos lácteos en el Perú en función de los precios e ingreso:

$Imp = \beta_1 * \beta_2 PP * \beta_3 IP * \beta_4 ICV * \beta_5 YNN$

Siendo:

Imp : Cantidades importadas por Perú, en miles de toneladas
PP : Precios promedios al detalle en Perú de productos lácteos importados de Argentina.
NIP : Números índices de los precios promedios al detalle en Perú de productos lácteos no importados de Argentina
ICV : Índice del costo de vida
YNN: Ingreso nacional neto a costo de factores

La transformación conveniente para establecer la ecuación logarítmica, es:

$$\textbf{Log Imp} = \beta_1 + \beta_2 \textbf{ Lg PP} + \beta_3 \textbf{ Lg NIP} + \beta_4 \textbf{ Lg ICV} + \beta_5 \textbf{ LgYNN}$$

Los datos transformados son los siguientes:

| **AÑO** | **LImp** | **LPP** | **LNIP** | **LICV** | **LYNN** |
|---|---|---|---|---|---|
| 1984 | 3.1565 | 1.1062 | 3.0000 | 2.3034 | 3.5682 |
| 1985 | 3.1724 | 1.1697 | 3.0055 | 2.2814 | 3.5848 |
| 1986 | 3.1738 | 1.1345 | 3.0082 | 2.2840 | 3.5932 |
| 1987 | 3.2039 | 1.1361 | 3.0241 | 2.2840 | 3.5999 |
| 1988 | 3.1928 | 1.1245 | 2.9894 | 2.2765 | 3.5926 |
| 1989 | 3.22269 | 1.1059 | 2.9805 | 2.2637 | 3.6175 |
| 1990 | 3.2143 | 1.1294 | 2.9987 | 2.2610 | 3.6184 |
| 1991 | 3.2567 | 1.1249 | 2.9899 | 2.2558 | 3.6209 |
| 1992 | 3.3192 | 1.0522 | 2.9284 | 2.2367 | 3.5974 |
| 1993 | 3.3596 | 0.9404 | 2.8768 | 2.2081 | 3.5642 |
| 1994 | 3.4002 | 0.9249 | 2.8470 | 2.1962 | 3.5524 |
| 1995 | 3.4630 | 0.8468 | 2.7882 | 2.1872 | 3.5714 |
| 1996 | 3.4847 | 0.8185 | 2.7497 | 2.1903 | 3.5889 |
| 1997 | 3.4605 | 0.8600 | 2.7709 | 2.1962 | 3.6137 |
| 1998 | 3.4742 | 0.9117 | 2.8109 | 2.2081 | 3.6423 |
| 1999 | 3.4956 | 0.9411 | 2.8482 | 2.2284 | 3.6643 |
| 2000 | 3.4469 | 0.9686 | 2.8561 | 2.2342 | 3.6694 |

**1. Regresión del modelo y evaluación de los parámetros.**

| Dependent Variable: LIMP | | | | |
|---|---|---|---|---|
| Method: Least Squares | | | | |
| Date: 11/03/04 Time: 09:34 | | | | |
| Sample: 1984 2000 | | | | |
| Included observations: 17 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| LPP | -0.701950 | 0.368778 | -1.903449 | 0.0812 |
| LNIP | 0.029628 | 0.481688 | 0.061509 | 0.9520 |
| LICV | -1.098483 | 0.442209 | -2.484080 | 0.0287 |
| LYNN | 1.172611 | 0.206303 | 5.683918 | 0.0001 |
| C | 2.187357 | 1.556175 | 1.405598 | 0.1852 |
| R-squared | 0.977628 | Mean dependent var | | 3.323352 |
| Adjusted R-squared | 0.970171 | S.D. dependent var | | 0.129914 |
| S.E. of regression | 0.022437 | Akaike info criterion | | -4.516242 |
| Sum squared resid | 0.006041 | Schwarz criterion | | -4.271180 |
| Log likelihood | 43.38806 | F-statistic | | 131.0986 |
| Durbin-Watson stat | 1.657679 | Prob(F-statistic) | | 0.000000 |

Se observa del resultado anterior que algunos parámetros no son significativos con un 95% de confianza como es el caso de LPP y LNIP.

El estadístico F es bastante alto con lo cual la variabilidad de la variable endógena queda bien explicada por el modelo, esto es corroborado por el $R^2$ de la regresión el cual ha alcanzando un 97%, por lo cual el modelo considerado es aceptable.

## 2. Evaluación de los Signos de los Parámetros

$\hat{\beta}_1 = 2.187357$, indica el nivel promedio de las cantidades importadas por Perú de productos lácteos de Argentina (consumo mínimo o autónomo), cuando el resto de parámetros son cero.

$\hat{\beta}_2 = -0.701950$, este resultado nos muestra la elasticidad de la cantidad demandada respecto al precio, pues si el precio promedio varía en 1 %, la cantidad importada varía en 70.1950%

$\hat{\beta}_3 = 0.029628$, este resultado nos muestra la elasticidad de la cantidad de demandada respecto al índice de precio promedio; Pues si éste varía en 1%, la cantidad importada varía en 2.9628%.

$\hat{\beta}_4 = -1.098483$, este resultado nos muestra la elasticidad de la cantidad demandada respecto al índice de costo de vida; Pues si éste varía en 1%, la cantidad importada disminuirá en 9.8483%

$\hat{\beta}_5 = 1.172611$, este resultado nos muestra la elasticidad de la cantidad demandada respecto al Ingreso Nacional Neto. Si este varía en 1%, la cantidad importada aumentará en 17.2611%.

## 3. Evaluación de Multicolinealidad en el Modelo

Si existe multicolinealidad porque:

- Tres de los coeficientes no son significativos a un nivel de significancia de 5%.
- El $R^2$ es casi igual a 1 lo cual nos indica un alineamiento casi perfecto de los datos.
- Existen parámetros con signo cambiado al que se esperaba, como por ejemplo el parámetro LNIP debería resultar con signo negativo y sin embargo tiene signo positivo.

Otra forma es por el método de la matriz de correlaciones de las variables:

| | LICV | LNIP | LPP | LYNN |
|---|---|---|---|---|
| **LICV** | 1 | 0.937526 | 0.938392 | 0.002873 |
| **LNIP** | 0.937526 | 1 | 0.988619 | -0.10719 |
| **LPP** | 0.938392 | 0.988619 | 1 | -0.0258 |
| **LYNN** | 0.002873 | -0.10719 | -0.0258 | 1 |

Se puede observar que existe una alta correlación entre las variables LICV-LNIP, LICV-LPP además de LNIP-LPP. De donde se observa que LNIP-LPP presenta casi un 99% de asociación lineal.

Al regresionar el modelo resulta un $R^2$ de 97.7%, el cual es bastante alto lo que indica un considerable grado de correlación existente entre estas variables.

## 4. Efectos de la multicolinealidad

Los parámetros se encuentran distorsionados (estadísticamente no significativos) dado que existe una combinación lineal entre las variables explicativas.

La colinealidad está referida a la existencia de una sola relación lineal entre las variables explicativas y, por lo tanto, la multicolinealidad se refiere a la existencia de más de una relación lineal. Es importante anotar que la multicolinealidad se refiere sólo a relaciones lineales entre las variables independientes y no a cualquier otro tipo de relación.

## 5. Predicción en Presencia de Multicolinealidad

El problema de multicolinealidad no es teórico sino de grado, en todo modelo econométrico tenderá a presentarse este problema, en algunos casos la severidad del problema será mayor. Esto quiere decir que existe interdependencia entre las variables económicas por lo cual no se puede eliminar.

El grado, se refiere a la severidad de la correlación entre las variables explicativas, la correlación entre una o más variables independientes puede ser perfecta. Se puede afirmar que para el caso de series de tiempo económicas, siempre se tendrá un grado relativamente alto de multicolinealidad.

## 6. Modelo Recomendado en Presencia de multicolinealidad

Las variables INIP y LPP están altamente correlacionadas por lo tanto se irá eliminando del modelo hasta obtener valores de $R^2$ más bajos y que los coeficientes sean más significativos.

De preferencia se recomienda trabajar con el modelo más sencillo posible pero que sea altamente explicativo, con el fin de evitar el problema de multicolinealidad.

Eliminamos del modelo la variable LPP y regresionamos:

Dependent Variable: LIMP
Method: Least Squares
Date: 11/03/04 Time: 09:40
Sample: 1984 2000
Included observations: 17

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| LNIP | -0.822996 | 0.194184 | -4.238226 | 0.0010 |
| LICV | -1.161413 | 0.483418 | -2.402505 | 0.0319 |
| LYNN | 0.974265 | 0.195191 | 4.991337 | 0.0002 |
| C | 4.810229 | 0.792733 | 6.067903 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.970874 | Mean dependent var | 3.323352 |
| Adjusted R-squared | 0.964152 | S.D. dependent var | 0.129914 |
| S.E. of regression | 0.024597 | Akaike info criterion | -4.370044 |
| Sum squared resid | 0.007865 | Schwarz criterion | -4.173994 |
| Log likelihood | 41.14538 | F-statistic | 144.4446 |
| Durbin-Watson stat | 1.503959 | Prob(F-statistic) | 0.000000 |

Como podemos ver todos los t-statistics son significativos al 95% de confianza; por lo tanto, tiene un mejor ajuste y la variable LNIP ahora está con signo negativo que es como se esperaba, además tenemos un $R^2$ de 97% de éxito del modelo estimado y el F-statistics es bastante alto por lo tanto se tendrá que las variables independientes explican al modelo en su conjunto.

Conclusión: el modelo con que se recomienda trabajar es:

**LIMP = 4.810 – 0.823 LNIP – 1.1614 LICV + 0.9742 LYNN**

# LECCIÓN 2

## MÍNIMOS CUADRADOS GENERALIZADOS

Cuando se estima un modelo con el método de mínimos cuadrados ordinarios uno de los supuestos se refiere a la matriz de varianzas y covarianzas del término de perturbación, que es una matriz cuya diagonal principal es un escalar. Ello expresa en resumen la presencia de 2 supuestos: El supuesto de **homocedasticidad**, donde la varianza del término de perturbación permanece constante. El segundo supuesto se refiere a la ausencia de **autocorrelación** entre dichas variables. Al cumplimiento simultáneo de estos supuestos también se le reconoce como aceptar que las perturbaciones "u" son esféricas[3].

No obstante, en la práctica puede darse situaciones diferentes.

Una de estas situaciones se produce cuando la varianza del término de perturbaciones es diferente de unas observaciones a otras. A esta situación se le conoce con el nombre de **Heteroscedasticidad.**

$$\mathrm{Var}(\mu_i) = \sigma_i^2$$

Una segunda situación ocurre cuando los términos de error correspondientes a diferentes observaciones no son independientes entre sí, es decir,

$$\mathrm{Cov}\ (\mu i,\ \mu j) \neq 0, \text{ para } i \neq j$$

Ello hace que los elementos fuera de la diagonal principal en la matriz no serán iguales a cero. A esta situación se le denomina **Autocorrelación**

Supongamos lo siguiente:

$$Y = X\beta + \hat{u} \qquad E(\hat{u}) = 0 \qquad V(\hat{u}) = \sigma^2\Omega$$

$\Omega$ es una matriz simétrica, positiva definida y conocida.

Donde:

$$\Omega = \begin{bmatrix} a_1 \ldots\ldots.. 0 \\ 0 \ldots\ldots\ldots. a_n \end{bmatrix}$$

A continuación se debe encontrar una matriz $P_{n,n}$, tal que:

$$\Omega = PP' \quad \Rightarrow \quad P^{-1}\Omega = P' \quad \Rightarrow \quad P^{-1}\Omega(P^{-1})' = I$$

Luego:

$$Y^* = P^{-1}Y = P^{-1}X\beta + P^{-1}\mu \qquad X^* = P^{-1}X \qquad \mu^* = P^{-1}\mu$$

[3] Esta denominación se debe a que se considera como parte del supuesto a la variable aleatoria de cada una de las unidades de observación "$u_i$" como una variable que se distribuye normalmente con media "0" y varianza "1".

Se pasa a un nuevo modelo $\Rightarrow$ $Y^* = X^* \beta + \mu^*$ $\quad V(\mu^*) = \sigma^2 I$

$E(\mu^*) = 0$ $\quad \mu^* \sim N$

$$E(\mu^*) = E(P^{-1}\mu) = 0$$
$$V(\mu^*) = (P^{-1})(\sigma^2\Omega)(P^{-1})' = \sigma^2 I$$

## 1. ESTIMADOR MÍNIMOS CUADRADOS GENERALIZADOS

Los estimadores lineales e insesgados de mínima varianza (dentro de una clase) de los parámetros $\beta i$ dado por:

$$\hat{\beta} = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^*$$
$$\hat{\beta} = \left(X'(P^{-1})'P^{-1}X\right)^{-1} X'(P^{-1})'P^{-1}Y$$

Como $\Omega = PP'$, entonces: $\Omega^{-1} = (P^{-1})'P^{-1}$

$\hat{\beta}_{MCG} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$ Estimadores Mínimos Cuadrados Generalizados

## 2. VARIANZAS Y COVARIANZAS

No obstante si se estimara el modelo con presencia de algunos de los problemas antes mencionados con la fórmula a partir de los MCO. Es decir con: $\hat{\beta} = (X'X)^{-1}X'Y$
La fórmula correcta para estimar la varianza sería la siguiente:

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X'(P^{-1})'P^{-1}X)^{-1} = \sigma^2 (X'\Omega^{-1}X)^{-1}$$

Si se estima con la fórmula habitual correspondiente a los MCO se produciría un sesgo que induciría a un error de interpretación.

Luego planteado los problemas de heterocedasticidad o autocorrelación las fórmulas adecuadas corresponden a las variables transformadas.

$\hat{\beta} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$ Estimadores Mínimos Cuadrados Generalizados

$$\mathrm{Var}(\beta_{MCG}) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$$

## LECCIÓN 3

### AUTOCORRELACIÓN

La autocorrelación es la relación que se da entre las variables perturbadoras ($\mu$), contraviniendo uno de los supuestos para estimar el modelo a partir de la independencia que debería existir entre estas variables. Este problema se presenta fundamentalmente cuando se realizan estudios econométricos de series históricas. En este caso, en cada periodo, las variables independientes deberían ser las únicas que expliquen el modelo, y no lo que halla sucedido en periodos anteriores. Es decir, ni los fenómenos ocurridos anteriormente, o la tendencia o ciclo de la variable debería afectar, en dicho periodo, el comportamiento de la variable dependiente.

La matriz de varianza covarianza de las perturbaciones, en presencia de autocorrelación, tiene la forma siguiente:

$$E(\mu\mu') = \sigma_\mu^2 \begin{pmatrix} 1 & \rho & \rho^2 & . & \rho^{n-1} \\ \rho & . & . & . & . \\ \rho^2 & . & . & . & . \\ . & . & . & . & . \\ \rho^{n-1} & . & . & . & 1 \end{pmatrix}$$

La cual como se observa posee "n" columnas y una diagonal principal, la misma que contiene la varianza de cada perturbación, mientras que los valores que están afuera de la diagonal constituyen las covarianzas asociadas a su correspondiente perturbación. El término $\sigma_\mu^2$ representa la varianza de cada residual, la cual como lo planteado inicialmente para un modelo lineal general es constante.

En el contexto de la regresión, el modelo clásico de regresión supone que no existe tal autocorrelación en las perturbaciones, lo cual se representa simbólicamente como:

$$E(\mu_i\mu_j) = 0$$

Significando que el término de perturbación no se encuentra influenciado por otra de éstas; en el caso de que tal dependencia exista se está ante una autocorrelación la cual se representa simbólicamente por:

$$E(\mu_i\mu_j) \neq 0$$

Entonces, en esta lección se considerará que la matriz de varianza covarianza del término de perturbación no es escalar; buscándose analizar la naturaleza y las complicaciones asociadas a esta presencia. Asimismo, se mostrará las estrategias necesarias para poder enfrentar este problema.

Antes de analizar las causas frecuentes de la presencia de este tipo de problema, se mostrará algunos patrones de la autocorrelación y no autocorrelación. La figura siguiente (A-1) nos permite determinar la existencia o no de la autocorrelación; en las figuras (a) y (d) se observa la presencia de un patrón distinguible entre las perturbaciones, siendo en la figura (a) un patrón cíclico mientras que en la figura (d) se puede advertir la presencia de términos de tendencia lineal y cuadrática. En las otras 2 figuras se tiene una presencia lineal hacia el origen (b) y hacia abajo (c) de las perturbaciones.

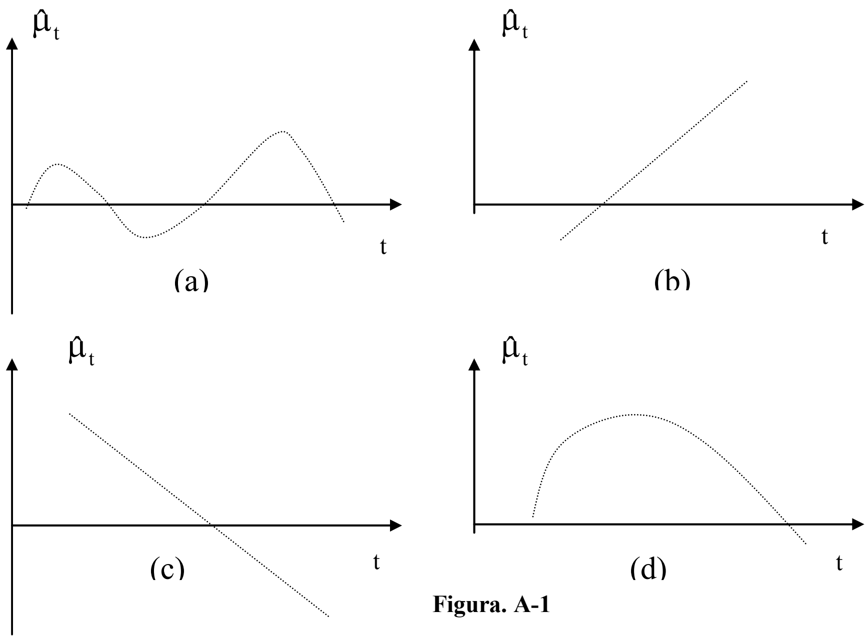



**Figura. A-1**

## 1. CAUSAS DE LA PROCEDENCIA DE AUTOCORRELACION

**a.** La inercia de los procesos económicos, es decir, la variable debe crecer en el tiempo.

**b.** El sesgo de especificación, se da cuando se excluye a una variable importante o la forma funcional del modelo no es la correcta. En el primer caso:

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$, es difícil especializar el fenómeno.

**c.** Rezagos en los modelos:

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + \mu_t$$
$$C_{t-1} = \beta_1 + \beta_2 Y_{t-1} + \beta_3 C_{t-2} + \mu_{t-1}$$

Cuando se habla de rezagos puede haber autocorrelación.

**d.** Existencia de ciclos y tendencias, si la autocorrelación es positiva, un valor alto de $\mu_t$ genera un valor de $Y_t$, por encima de su promedio, y en consecuencia una elevada probabilidad debe ir seguido de un valor alto de $\mu_{t+1}$, y de un valor de $Y_{t+1}$ por encima de su promedio; lo mismo ocurriría para valores de $Y_t$ por debajo de su promedio.

e. La autocorrelación positiva está asociada a la existencia de rachas de valores altos y bajos de $Y_t$. (Fig. A -2)

f. El conjunto de variables explicativas del modelo no explican adecuadamente dicho comportamiento, entonces el término de error incorporará dicha tendencia, y esto conduce a la existencia de autocorrelación positiva. Un conjunto de residuos negativos seguida por otro conjunto de residuos positivos. (Fig. A -3)

g. Manipulación de datos, en especial cuando se realiza la interpolación (es obtener un valor dentro de la muestra) y extrapolación (es un valor fuera de la muestra) de la información muestral.

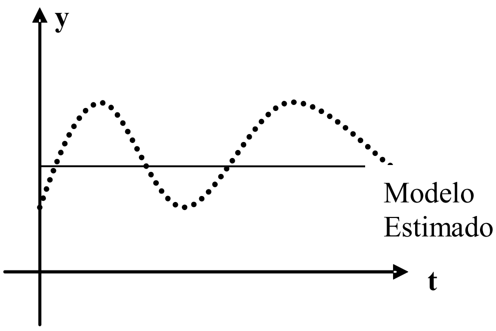



**Figura. A-2**. Autocorrelación producida por un ciclo

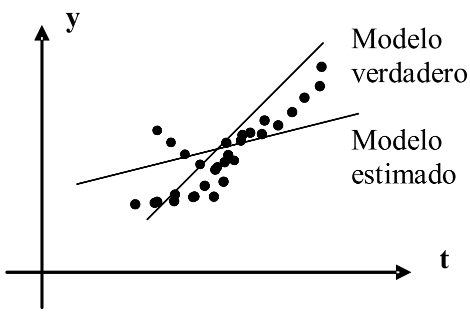



**Figura. A-3**. Autocorrelación producida por una tendencia

## 2. CONSECUENCIAS DE LA AUTOCORRELACION

Los estimadores de Mínimos Cuadrados Ordinarios (MCO) tendrán las siguientes propiedades:

**a.** Son insesgados en muestras repetidas sus valores medios son iguales a las verdaderos valores reales.
**b.** Son consistentes a medida que el tamaño de la muestra se aproxime al tamaño de la población.
**c.** Como en el caso de heterocedasticidad ya no son eficientes (mínima varianza) ni para muestras pequeñas ni grandes.

Por lo tanto ante la presencia de autocorrelación se debe de tener presente que:

- Aunque se tenga en cuenta la correlación serial en los estimadores mínimos cuadrados ordinarios (MCO) y se utilice la fórmula de varianza con autocorrelación los estimadores serán menos eficientes. En consecuencia los intervalos de confianza son más anchos de lo necesario y la prueba de significancia pierden fuerza.

  Si se ignora el problema de autocorrelación y se aplica la fórmula clásica de MCO (bajo el supuesto de ausencia de correlación) las consecuencias son todavía más serias y la varianza residual tiende a subestimar la varianza verdadera.

- Incluso si la varianza residual $(\sigma_{\mu})$ estuviera subestimada las varianzas y los errores estándar de los estimadores MCO $(\sigma_{\hat{\beta}}$ y $\sigma^2_{\hat{\beta}})$ tienden a subestimar las verdaderas varianzas de los parámetros estimados.

- En consecuencia las pruebas de significancia pierden validez y si se aplican tienden a dar conclusiones erróneas.

Planteadas las ideas iniciales, un modelo que presente autocorrelación, tendrá la forma siguiente:

$$Y = X\beta + \mu$$

Siendo $E(\mu) = 0$ y $E(\mu\mu') = \sigma^2_{\mu}\Omega$

En la práctica, usualmente se supone que las $\mu_t$ siguen el esquema autorregresivo de primer orden[4] :

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t$$

Donde $|\rho| < 1$ y donde las $\varepsilon_t$ siguen los supuestos MCO de valor esperado cero, varianza constante y no autocorrelación, es decir:

$$E(\varepsilon_t) = 0$$
$$Var(\varepsilon_t) = \sigma^2$$
$$Cov(\varepsilon_{t,}\varepsilon_{t+s}) = 0,\ s \neq 0$$

Además si se observa de manera individual, se tendrá que los diferentes valores de las perturbaciones $\mu$ para cada instante de tiempo originaran una varianza igual a:

$$Var(\mu_t) = \sigma^2_{\mu} = E(\mu^2_t) = \frac{\sigma^2_{\varepsilon}}{1-\rho^2}$$

Algo similar ocurre con la covarianza, la cual cambia entre los términos de error que están a "s" periodos distantes:

$$cov(\mu_t\mu_{t+s}) = (E(\mu_t\mu_{t-s}) = \rho^s \frac{\sigma^2_{\varepsilon}}{1-\rho^2} = \rho^s\sigma^2_{\mu}$$

## 3. DETECCIÓN DE AUTOCORRELACION

El supuesto de no autocorrelación del modelo clásico se relaciona con las perturbaciones poblacionales $\mu_t$, las cuales no pueden ser observadas directamente. En su lugar, se dispone de sus valores aproximados, los residuales $\hat{\mu}_t$ que pueden obtenerse a partir del procedimiento usual MCO. Aunque las $\hat{\mu}_t$ no son lo mismo que las $\mu_t$, con

[4] La autocorrelación será de orden "p" cuando el subíndice de $\mu$ sea "t-p"

mucha frecuencia un examen visual de las $\hat{\mu}_t$ nos da algunas claves sobre la posible presencia de autocorrelación en las $\mu_t$.

**a. Prueba de Durbin Watson**

Se construye la dócima de hipótesis:

$H_0$: $\rho = 0$ (no existe autocorrelación de orden 1)
$H_1$: $\rho \neq 0$ (existe autocorrelación de orden 1)

Se utiliza el estadístico de **Durbin Watson**:

$$d = \frac{\sum_{t=2}^{n}(\hat{\mu}_t - \hat{\mu}_{t-1})^2}{\sum_{t=2}^{n}\hat{\mu}_t^2}$$

Con lo cual: $$d = 2(1-\hat{\rho}) \Rightarrow \hat{\rho} = \frac{2-d}{2}$$

Por lo que:

$\hat{\rho} = -1 \Rightarrow d = 4$ Autocorrelación negativa
$\hat{\rho} = 1 \Rightarrow d = 0$ Autocorrelación positiva
$\hat{\rho} = 0 \Rightarrow d = 2$ No existe autocorrelación.

**Ejercicio Ilustrativo 1:**
En el modelo de inflación, especificado anteriormente, probamos la existencia de autocorrelación. De las salidas de la regresión tenemos que:

| | |
|---|---|
| Durbin-Watson stat | 0.628449 |

Esto nos indica que **existe autocorrelación positiva de orden 1.** Sin embargo, podría darse el caso que exista autocorrelación de mayor orden, por lo que este test no es el adecuado, pues sólo permite identificar autocorrelación de orden 1.

**b. Detección de Autocorrelación cuando el Modelo tiene una Variable Endógena Rezagada ($Y_{t-1}$)**

Cuando se tiene un modelo con una variable rezagada entre las explicativas, esta variable rezagada generalmente se refiere a la variable dependiente.
En este otro caso, para detectar autocorrelación, el estadístico propuesto por Durbin Watson, es el estadístico h, y se calcula como:

$$h = \hat{\rho}\sqrt{\frac{n}{1-n\hat{\sigma}^2_{\hat{\beta}_1}}}$$

Donde:

$\hat{\beta}_1$ : Estimador de la variable endógena rezagada (del primer retardo). Es decir es el coeficiente de $Y_{t-1}$ en la regresión mínimo-cuadrática (MCO):

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ..... + u_t$$

$\hat{\sigma}^2_{\beta_1}$ : Es la estimación de la varianza muestral de $\hat{\beta}_1$.

Es utilizable sólo para $n\hat{\sigma}^2_{\hat{\beta}_1} \leq 1$, siendo $\hat{\rho}$ el coeficiente estimado de autocorrelación de primer orden de los residuos ($\rho$), calculable a partir del valor de:

$$\hat{\rho} \sim 1 - \frac{1}{2}d$$

$H_0 : \hat{\rho} = 0$

$H_1 : \hat{\rho} < 0$ ó $\hat{\rho} > 0$

$$d = \frac{\sum_{t=2}^{T}(\mu_t - \mu_{t-1})^2}{\sum_{t=1}^{T}\mu_t{}^2}$$ es el estadístico de contraste.

El contraste de $h$ se realiza a partir de su consideración de variable distribuida normalmente con parámetros (0,1). Así, con un nivel de significación del 5%, si $|h| > 1.645$ debe rechazarse la hipótesis de autocorrelación nula (Fig. A-4).

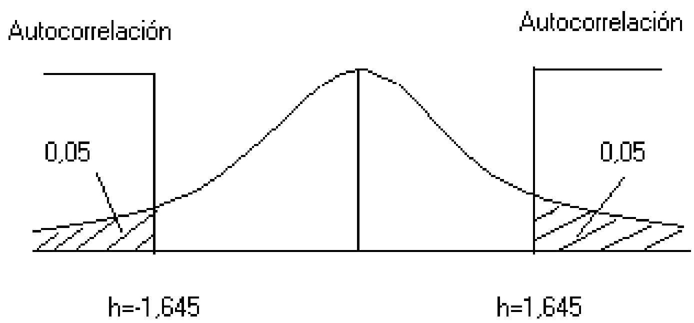



**Figura. A-4**

**Ejercicio Ilustrativo 2:**
En un modelo que contiene una variable explicativa rezagada, se ha detectado que el d = 1.75, n =50, $\hat{\sigma}_{\hat{\beta}_1} = 0.15$. Se pide determinar si el modelo presenta problemas de autocorrelación.

Entonces, se debe hallar el "h": $\hat{\rho} = 1 - \frac{1}{2}d = 1 - \frac{1}{2}(1.75) = 0.125$

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\hat{\sigma}^2_{\hat{\beta}_1}}}$$

Entonces: $h = 0.125\sqrt{\frac{50}{1 - 25(0.0225)}} = 1.89$

Luego, para un nivel de significancia del 5%, tenemos:

Como h = 1.89, entonces se cumple la desigualdad, $|h| > 1.645$. Por lo tanto, tenemos evidencias para rechazar la hipótesis nula de ausencia de autocorrelación.

### c. Contraste de Autocorrelación en Base al Multiplicador de Lagrange (LM)

El procedimiento del multiplicador de Lagrange es usado para la detección de autocorrelación entre los residuos de una ecuación. El contraste es adecuado para los casos de indeterminación al usar el estadístico "d" de Durbin-Watson, o imposibilidad de obtención del "h" de Durbin[5]. Sin embargo, es importante señalar que su uso no debe limitarse a dichos casos. Para su aplicación se debe considerar un estadístico de prueba LM = $nR^2$, donde n es el tamaño muestral el cual sigue aproximadamente una distribución Chi-cuadrado con "p" grados de libertad ($\chi^2_p$), es decir:

$M \sim \chi^2_p$ ; p grados de Libertad (autocorrelación de orden p)

La regla de decisión para una autocorrelación de orden p es:

Si LM > $\chi^2_p$ : Existe autocorrelación significativa de orden p.
Si LM < $\chi^2_p$ : No existe autocorrelación significativa de orden p.

**Ejercicio Ilustrativo 3:**
En el modelo de inflación aplicamos el test LM. El paquete E-Views calcula este test, el cual permite identificar la presencia de autocorrelación de cualquier orden.

En este caso se calculará la autocorrelación de orden 2, obteniendo los siguientes resultados:

[5] Sin embargo, es importante que el lector no interprete que su uso se limita a dichos casos. El contraste LM se puede y debe quizás aplicar como complemento a los demás.

| Breusch-Godfrey Serial Correlation LM Test: | | | | |
|---|---|---|---|---|
| F-statistic | 33.50958 | Probability | | 0.000000 |
| Obs*R-squared | 38.55391 | Probability | | 0.000000 |

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 09/22/02 Time: 09:38

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.060917 | 0.062686 | 0.971787 | 0.3341 |
| LCIRPROM | -0.009815 | 0.010168 | -0.965219 | 0.3374 |
| LTC | 0.016273 | 0.020029 | 0.812500 | 0.4189 |
| RESID(-1) | 0.694486 | 0.115892 | 5.992536 | 0.0000 |
| RESID(-2) | 0.016642 | 0.117288 | 0.141892 | 0.8875 |

| | | | |
|---|---|---|---|
| R-squared | 0.458975 | Mean dependent var | 1.50E-15 |
| Adjusted R-squared | 0.431581 | S.D. dependent var | 0.025635 |
| S.E. of regression | 0.019327 | Akaike info criterion | -4.996966 |
| Sum squared resid | 0.029509 | Schwarz criterion | -4.852275 |
| Log likelihood | 214.8726 | F-statistic | 16.75479 |
| Durbin-Watson stat | 1.896173 | Prob(F-statistic) | 0.000000 |

La hipótesis nula es: no existe autocorrelación de orden (p). En este caso, la hipótesis nula es que no existe autocorrelación de orden 2. Analizando los resultados obtenidos, rechazamos la hipótesis nula al 99% de confianza ($\alpha$=0.01). Entonces, concluimos que existe autocorrelación de orden 2.

### d. El Contraste de Ljung y Box

El estadístico de Ljung y Box (1978), al igual que el contraste anterior, sirve para contrastar esquemas generales de autocorrelación en la perturbación. La Hipótesis nula, indica la ausencia de autocorrelación serial. El estadístico es el siguiente:

$$Q^* = n(n+2)\left[\frac{\rho_1^2}{(n-1)} + \frac{\rho_2^2}{(n-2)} + \ldots + \frac{\rho_p^2}{(n-p)}\right] \sim \chi_p^2$$

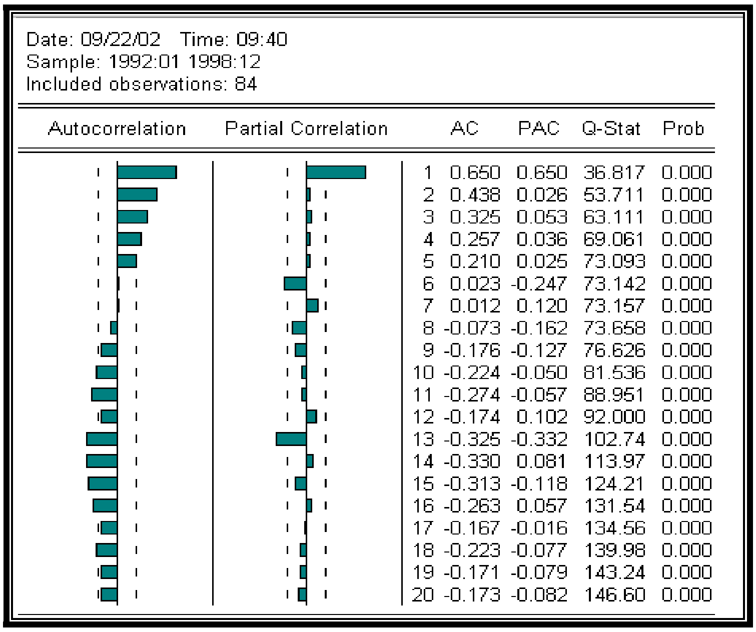
donde $\rho_1, \rho_2, \ldots \rho_p$ son los coeficientes de autocorrelación de orden 1,2 y p, cuyas expresiones son conocidas.

La hipótesis de autocorrelación significativa de orden p es aceptada si el valor calculado del estadístico supera al valor crítico de la tabla con p grados de libertad.

Un aspecto importante, es que los estadísticos de contraste Ljung y Box detectan el grado de autocorrelación de orden p, de manera discontinua. Esto permite afirmar que si por ejemplo $\hat{\rho}_4$ es significativamente elevado eso no implica que $\hat{\rho}_1, \hat{\rho}_2$ y $\hat{\rho}_3$ también lo sean, a no ser que de manera complementaria se demuestre lo contrario.

**Ejercicio Ilustrativo 4:**
En el modelo de inflación, probamos la existencia de autocorrelación mediante el Contraste de Ljung y Box, hasta un orden establecido con este

La hipótesis nula es que no existe autocorrelación hasta el orden (n). Del paquete E-Views, se obtiene el siguiente gráfico:


Date: 09/22/02 Time: 09:40
Sample: 1992:01 1998:12
Included observations: 84

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.650 | 0.650 | 36.817 | 0.000 |
| | | 2 | 0.438 | 0.026 | 53.711 | 0.000 |
| | | 3 | 0.325 | 0.053 | 63.111 | 0.000 |
| | | 4 | 0.257 | 0.036 | 69.061 | 0.000 |
| | | 5 | 0.210 | 0.025 | 73.093 | 0.000 |
| | | 6 | 0.023 | -0.247 | 73.142 | 0.000 |
| | | 7 | 0.012 | 0.120 | 73.157 | 0.000 |
| | | 8 | -0.073 | -0.162 | 73.658 | 0.000 |
| | | 9 | -0.176 | -0.127 | 76.626 | 0.000 |
| | | 10 | -0.224 | -0.050 | 81.536 | 0.000 |
| | | 11 | -0.274 | -0.057 | 88.951 | 0.000 |
| | | 12 | -0.174 | 0.102 | 92.000 | 0.000 |
| | | 13 | -0.325 | -0.332 | 102.74 | 0.000 |
| | | 14 | -0.330 | 0.081 | 113.97 | 0.000 |
| | | 15 | -0.313 | -0.118 | 124.21 | 0.000 |
| | | 16 | -0.263 | 0.057 | 131.54 | 0.000 |
| | | 17 | -0.167 | -0.016 | 134.56 | 0.000 |
| | | 18 | -0.223 | -0.077 | 139.98 | 0.000 |
| | | 19 | -0.171 | -0.079 | 143.24 | 0.000 |
| | | 20 | -0.173 | -0.082 | 146.60 | 0.000 |

Al analizar el estadístico Q tenemos que:

- Rechazamos la hipótesis nula de no autocorrelación hasta el orden 1 al 99% de confianza. (nivel de significación $\alpha$ =0.01)
- Rechazamos la hipótesis nula de no autocorrelación hasta el orden 2 al 99% de confianza. Es decir, puede existir autocorrelación de orden 1 ó 2.
- Rechazamos la hipótesis nula de no autocorrelación hasta el orden 3 al 99% de confianza. Es decir, puede existir autocorrelación de orden 1, 2 ó 3.

Este proceso continúa hasta:

- Rechazar la hipótesis nula de no autocorrelación del orden 20 al 99% de confianza. Es decir, puede existir autocorrelación de orden 1, 2, 3, 4,..., 19 ó 20.

Para saber cuál es el orden de autocorrelación analizamos el comportamiento del los coeficientes de autocorrelación parcial (Partial Correlation). El orden de autocorrelación estará determinado por el orden del último coeficiente que esta fuera de las bandas de confianza. En este caso, se trata del coeficiente de autocorrelación parcial de orden 13 ya que es el último que está fuera de las bandas. Por lo tanto, concluimos que existe autocorrelación de orden 13.

## 4. CORRECCIÓN DE AUTOCORRELACION

**Forma General**

Puesto que las perturbaciones $\mu_t$ no son observables, la naturaleza de la correlación serial es frecuentemente un asunto de especulación o de exigencias prácticas. En la práctica, usualmente se supone que las $\mu_t$ siguen el esquema autorregresivo de primer orden, es decir:

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t \qquad \ldots(1)$$

Donde $|\rho| < 1$ y donde las $\varepsilon_t$ siguen los supuestos MCO: de valor esperado cero, varianza constante y no autocorrelación, es decir:

$$E(\varepsilon_t) = 0$$
$$Var(\varepsilon_t) = \sigma^2$$
$$Cov(\varepsilon_t, \varepsilon_{t+s}) = 0,\ s \neq 0$$

Si se supone la validez de (1), el problema de correlación serial puede ser resuelto satisfactoriamente si se conoce $\rho$, el coeficiente de autocorrelación. Analizando para un modelo con dos variables[2].

$$Y_t = \beta_1 + \beta_2 X_t + \mu_t \qquad \ldots(2)$$

Si (2) es cierta en el tiempo t, también es cierta en el tiempo t-1. Por lo tanto,

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \mu_{t-1} \qquad \ldots(3)$$

Multiplicando (3) por $\rho$ a ambos lados, se obtiene

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho\mu_{t-1} \qquad \ldots(4)$$

Restando (4) de (2) se obtiene

$$(Y_t - \rho Y_{t-1}) = (\beta_1 - \rho\beta_1) + (\beta_2 X_t - \rho\beta_2 X_{t-1}) + (\mu_t - \rho\mu_{t-1})$$

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \qquad \ldots(5)$$

donde en el último paso, se hace uso de la relación (1) : $\varepsilon_t = \mu_t - \rho\mu_{t-1}$

Se puede expresar (5) como

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \qquad \ldots(6)$$

[2] No importa si el modelo tiene más de una variable explicativa porque la autocorrelación es una propiedad de los $\mu_i$

Donde:

$\beta_1^* = \beta_1(1-\rho)$, $Y_t^* = (Y_t - \rho Y_{t-1})$ y $X_t^* = (X_t - \rho X_{t-1})$

Puesto que $\varepsilon_t$ satisface todos los supuestos MCO, se puede proceder a aplicar MCO sobre las variables transformadas Y* y X* y obtener estimadores con todas las propiedades óptimas.

La regresión (5) se conoce por el nombre de **ecuación en diferencia generalizada.** Esta consiste en regresar Y sobre X, no en la forma original, sino en forma de diferencia, lo cual se logra restando una proporción (de valor igual a "ρ") del valor de una variable en el período de tiempo anterior de su valor en el período de tiempo actual. En este procedimiento de diferenciación se pierde una observación puesto que la primera observación no tiene precedente. Para evitar esta pérdida de una observación, la primera observación sobre Y y X es transformada por una de las siguientes formas[3]:

$$\sqrt{1-\rho^2} \quad , \quad Y_1\sqrt{1-\rho^2} \quad y \quad X_1\sqrt{1-\rho^2}$$

Esta transformación es conocida como la **transformación de Prais-Winsten**.

**a. Procedimiento de Durbin –Watson**

Este procedimiento está basado en la relación entre los estadísticos d y $\rho$ según Durbin y Watson.

Dicha relación es la siguiente: $d \sim 2(1-\hat{\rho})$

Puede despejarse $\hat{\rho}$ en función de una metodología estimada, y así obtener de manera indirecta, y también aproximada el valor de $\hat{\rho}$:

$$\hat{\rho}_{dw} \sim 1 - d/2$$

Es decir, consiste en obtener el grado de autocorrelación habiendo obtenido previamente el estadístico de Durbin-Watson de detección de la misma, a partir de la relación aproximadamente señalada. En síntesis, el procedimiento completo es el siguiente:

**1º Estimación MCO del modelo, y cálculo del estadístico $d$ de Durbin y Watson.**
**2º Obtención de $\hat{\rho}$ mediante la relación aproximada entre $\hat{\rho}$ y $d$.**
**3º Aplicación de la ecuación de diferencias generalizadas, considerando el valor de $\hat{\rho}$ estimado, en el paso 2.**

[3] La pérdida de una observación puede no ser muy grave en una muestra grande, pero puede constituir una diferencia sustancial en los resultados en una muestra pequeña.

Los supuestos en los cuales se basa son:

- El modelo de regresión incluye el término de intercepto. Si dicho término no está presente, como es el caso de la regresión a través del origen, es esencial efectuar nuevamente la regresión incluyendo el término del intercepto para obtener la SCR.
- Las variables explicativas X, son no estocásticas, es decir, son fijas en muestreo repetido.
- Las perturbaciones $\mu_t$ del modelo original se generan mediante un modelo que relaciona el error de un periodo temporal con el siguiente (esquema autorregresivo de primer orden):

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t$$

  donde $\rho$ es la correlación de los errores del modelo original y $\varepsilon_t$ es el término de error aleatorio en la predicción de los errores, es decir, mide el error que se comente cuando se trata de estimar el error en el modelo original a partir del modelo planteado.

- El modelo de regresión no incluye valor(es) rezagado(s) de la variable dependiente como una de las variables explicativas. Por tanto, la prueba es inaplicable a modelos del siguiente tipo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ...... + \beta_k X_{kt} + \gamma Y_{t-1} + \mu_t$$

  donde $Y_{t-1}$ es el valor de Y rezagado un período.
  Tales modelos se conocen como **modelos autoregresivos**.

- No hay observaciones faltantes en los datos.

**b. Procedimiento Iterativo de Cochran – Orcutt**

Sin duda alguna, se trata del procedimiento más usado y uno de los más importantes para el tratamiento de la autocorrelación, dado que está basado en la estimación de $\rho$ como coeficiente de autocorrelación propiamente dicho, en base a la expresión:

$$\rho = \frac{\sum \mu_t \mu_{t-1}}{\sum \mu_t^2}$$

Asimismo, se puede estimar con el modelo $\mu_t = \rho\mu_{t-1} + \varepsilon_t$

Los procedimientos para estimar "$\rho$" son los siguientes:

**Primera Etapa**
Estimación del modelo

$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_k X_{kt} + \mu_t$ (1)

y obtención del estimador de la autocorrelación de Cochran-Orcutt, con el fin de aplicar la ecuación de diferencias, de la siguiente forma:

Retrasando un período y multiplicando por $\rho$:

$$\rho Y_{t-1} = \beta_1\rho + \beta_2\rho X_{2t-1} + ... + \beta_k \rho X_{kt-1} + \rho\mu_{t-1} \qquad (2)$$

restando (2) de (1) :

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 (X_{2t} - \rho X_{2t-1}) + ... + \beta_k( X_{kt} - \rho X_{kt-1}) + (\mu_t - \rho\mu_{t-1})$$

**Segunda Etapa**

A partir de los residuos de la ecuación de diferencias de la primera etapa, obtener un nuevo estimador de Cochran-Orcutt, tal como:

$$\hat{\rho}^* = \frac{\sum \mu_t^* \mu_{t-1}^*}{\sum \mu_t^{*2}}$$

el mismo que se reemplaza en una nueva ecuación de diferencias:

$$Y_t - \hat{\rho}^* Y_{t-1} = \beta_1 (1 - \hat{\rho}^*) + \beta_2 (X_{1t} - \hat{\rho}^* X_{1t-1}) + \cdots + \varepsilon_t^*$$

siendo $\varepsilon_t$ el error que se comente cuando se trata de estimar el error en el modelo de la segunda etapa a partir del modelo obtenido por diferencias al igual que en la primera etapa.

Alternativamente se estima “$\rho^*$” en el modelo:

$$\mu^*{}_t = \rho^* \mu^*{}_{t-1} + \varepsilon^*{}_t$$

**Tercera Etapa y Etapas Sucesivas**

De los residuos de la ecuación de diferencias anterior (segunda etapa), puede lograrse una nueva estimación del coeficiente de autocorrelación, tal como:

$$\hat{\rho}^{**} = \frac{\sum \mu_t^{**} \mu_{t-1}^{**}}{\sum \mu_t^{**2}}$$

Forma alternativa para estimar $\rho^{**}$:

$$\mu^{**}{}_t = \rho^{**} \mu^{**}{}_{t-1} + \varepsilon^{**}{}_t$$

De esta manera se reemplaza en una nueva ecuación de diferencias, y así sucesivamente. El proceso puede repetirse hasta que las estimaciones sucesivas de $\rho$ no difieran significativamente entre sí, pudiendo lograrse 2, 5, 10 ó más iteraciones.

Entre las ventajas de este procedimiento destacan las vinculadas a las propiedades estimativas, dado que estas descansan sobre sólidas bases metodológicas al facultarse la estimación del grado de autocorrelación como el coeficiente de correlación entre residuos sucesivos, lo cual es acorde con la definición de autocorrelación de primer orden.

Sin embargo, también debe tenerse en cuenta que muchas veces un elevado número de iteraciones, cercano al número de observaciones, transforma la información original, dificultando notablemente la interpretación lógica de las relaciones entre las variables.

### c. Tratamiento de Autocorrelación de Orden "p" (Cochrane – Orcutt)

Ante la existencia de autocorrelación en un orden mayor al primero, pueden generalizarse las metodologías.

Es factible obtener los diversos grados de autocorrelación ante los diversos órdenes de la misma. Así, se obtiene el $\hat{\rho}_p$

$$\hat{\rho}_p = \frac{\Sigma \hat{\mu}_t \hat{\mu}_{t-p}}{\Sigma \hat{\mu}_t^2} \qquad \text{grado de autocorrelacion de orden p}$$

Entonces, utilizando el procedimiento iterativo de Cochrane – Orcutt y partiendo de una ecuación de tipo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \mu_t$$

En el cual existe un proceso de autocorrelación de orden p como el siguiente:

$$\mu_t = \rho_1 \mu_{t-1} + \rho_1 \mu_{t-2} + \ldots + \rho_1 \mu_{t-p} + \varepsilon_t \qquad AR(p)$$

puede aplicarse una ecuación de diferencias generalizadas como la que sigue:

$$\begin{aligned}(Y_t - \rho_1 Y_{t-1} - \ldots - \rho_p Y_{t-p}) = \beta_1(1 - \rho_1 - \ldots - \rho_p) + \beta_2(X_{2t} - \rho_1 X_{2t-1} - \ldots - \rho_p X_{2t-p}) + \ldots \\ + \beta_k(X_{1t} - \rho_1 X_{kt-1} - \ldots - \rho_1 X_{kt-p}) + (\mu_t - \rho_1 \mu_{t-1} - \ldots - \rho_1 \mu_{t-p})\end{aligned}$$

donde $\rho_1$ $\rho_2$ … $\rho_p$ serán estimados mediante el proceso iterativo explicado con anterioridad.

## 5. PREDICCIÓN CON AUTOCORRELACION O COVARIANZAS NO NULAS

Como se recordará, cuando se detecta este problema se puede corregir aplicando mínimos cuadrados ordinarios al modelo original utilizando la información modificada (es decir MCG):

$$N^*{}_t = N_t - \hat{\rho} N_{t-1}$$

en donde $\hat{\rho}$ podría obtenerse a través de la fórmula $\hat{\rho} = \dfrac{\sum_{t=2}^{n} \hat{\mu}_t \hat{\mu}_{t-1}}{\sum_{t=2}^{n} \hat{\mu}_{t-1}^2}$ .

La técnica permite conseguir estimadores de los parámetros $\beta_i$ con las características deseadas.

Para utilizar el modelo estimado por MCG en predicción se requerirá ajustar los valores proyectados de todas las variables exógenas $X_{t+1}$, a los valores corregidos proyectados de ellas mismas .Con esos valores se podrá obtener el predictor $\hat{Y}^*{}_{t+1}$:

$$X_{t+1}^* = X_{t+1} - \hat{\rho} X_t \qquad \hat{Y}_{t+1}^* = \hat{Y}_{t+1} - \hat{\rho} Y_t$$

El esquema del predictor puntual del valor futuro de $Y_t$:

$$\hat{Y}_{t+1}^* = \hat{\beta}_1^* + \hat{\beta} * X_{t+1}^*$$

$$\hat{Y}_{t+1} = \hat{\beta}_1^* + \hat{\beta} * \left( X_{t+1} - \hat{\rho} X_t \right) + \hat{\rho} Y_t$$

Esto significa que en la predicción por intervalo, los límites que son fijados por:

$$\hat{Y}_{t+1} \pm t_{tab} \sqrt{\hat{\sigma}_{\mu MCG}^2}$$

Dependerán del nivel de confianza y del número de grados de libertad de la estimación por MCG. Para determinar el valor $\sigma_{\mu MCG}^2$ se utilizará la varianza de los estimadores obtenidos con la información muestral modificada para la aplicación de MCG.

**Ejercicio Ilustrativo 5:**

Se presenta un modelo cuya variable dependiente son las importaciones en función de la formación bruta de capitales a valores constantes de 1979, del periodo de 1979 – 1998.

**M=C(1)+C(2)*FBK**

Se desea analizar el problema de autocorrelación para el modelo anteriormente descrito, para el cual se realizará:

1. Detección de autocorrelación método Durbin Watson.
2. Corrección con el procedimiento Durbin Watson.
3. Contraste de autocorrelación con el multiplicador de Lagrange.
4. Corrección con el procedimiento de Cochran-Orcutt

Resultados del Eviews:

| Dependent Variable: M | | | | | |
|---|---|---|---|---|---|
| Method: Least Squares | | | | | |
| Sample: 1979 1998 | | | | | |
| Included observations: 20 | | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. | |
| C | -72.10892 | 75.26001 | -0.958131 | 0.3507 | |
| FBK | 1.002990 | 0.079313 | 12.64599 | 0.0000 | |
| R-squared | 0.898832 | Mean dependent var | | 846.7770 | |
| Adjusted R-squared | 0.893211 | S.D. dependent var | | 268.2676 | |
| S.E. of regression | 87.66604 | Akaike info criterion | | 11.87959 | |
| Sum squared resid | 138336.0 | Schwarz criterion | | 11.97916 | |
| Log likelihood | -116.7959 | F-statistic | | 159.9210 | |
| Durbin-Watson stat | 0.783378 | Prob(F-statistic) | | 0.000000 | |

**1. Prueba de Durbin Watson**

El modelo que estamos analizando presenta una $R^2$ del 90% aproximadamente lo que le da una buena bondad de ajuste al modelo, con un buen nivel de significancia del parámetro de la FBK y con una prueba F que asegura su significancia global; es decir tenemos un modelo que explica de una buena manera las importaciones en función de la formación bruta de capital.

Para la prueba de Durbin Watson utilizáremos un nivel de significancia del 5%, con un grado de libertad igual a 1.

**dl=1.201 4-dl=2.589**
**du=1.411 4-dl=2.79**

La conclusión de esta prueba es que el modelo muestra autocorrelación positiva (D-W: 0.783378 < 1.201).

**2. Corrección Durbin Watson.**

Hallamos ρ a través del valor de Durbin Watson:

**ρ = 1- (Durbin Watson/2) =0.608311**

Corregimos las variables con el coeficiente de autocorrelación hallado, el cual se muestra en la tabla siguiente:

| Obs | M1 | FBK1 | Obs | M1 | FBK1 |
|---|---|---|---|---|---|
| 1979 | 528.7780 | 600.6792 | 1989 | 94.47736 | 57.60925 |
| 1980 | 462.4510 | 570.4242 | 1990 | 259.4790 | 315.2691 |
| 1981 | 477.9764 | 615.1930 | 1991 | 336.8414 | 339.0263 |
| 1982 | 416.1886 | 395.8117 | 1992 | 360.7952 | 284.7985 |
| 1983 | 98.00196 | 3.732050 | 1993 | 313.8694 | 371.2941 |
| 1984 | 151.7961 | 206.0035 | 1994 | 496.0415 | 552.5263 |
| 1985 | 180.5833 | 178.1658 | 1995 | 623.5556 | 600.4738 |
| 1986 | 321.8671 | 405.8461 | 1996 | 483.1665 | 425.6563 |
| 1987 | 351.5099 | 503.8133 | 1997 | 603.6439 | 605.4949 |
| 1988 | 225.7588 | 285.9143 | 1998 | 530.7962 | 492.1347 |

Luego hallamos el siguiente modelo: **M1=C(1)+C(2)*FBK1**

| | | | | |
|---|---|---|---|---|
| Dependent Variable: M1 | | | | |
| Method: Least Squares | | | | |
| Sample: 1979 1998 | | Included observations: 20 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 47.62736 | 33.48209 | 1.422473 | 0.1720 |
| FBK1 | 0.814999 | 0.077968 | 10.45302 | 0.0000 |
| R-squared | 0.858564 | Mean dependent var | | 365.8789 |
| Adjusted R-squared | 0.850706 | S.D. dependent var | | 161.2517 |
| S.E. of regression | 62.30537 | Akaike info criterion | | 11.19661 |
| Sum squared resid | 69875.27 | Schwarz criterion | | 11.29619 |
| Log likelihood | -109.9661 | F-statistic | | 109.2656 |
| Durbin-Watson stat | 1.072147 | Prob(F-statistic) | | 0.000000 |

Efectuada la corrección se nota que aún tenemos un test de Durbin Watson que presenta autocorrelación positiva. Efectuemos la corrección de nuevo.

**ρ = 1- (Durbin Watson/2) = 0.570718**

| obs | M2 | FBK2 | obs | M2 | FBK2 |
|---|---|---|---|---|---|
| 1979 | 434.2040 | 493.2454 | 1989 | -34.36723 | -105.5672 |
| 1980 | 160.6679 | 227.6057 | 1990 | 205.5591 | 282.3905 |
| 1981 | 214.0472 | 289.6417 | 1991 | 188.7520 | 159.0966 |
| 1982 | 143.3989 | 44.71001 | 1992 | 168.5538 | 91.31010 |
| 1983 | -139.5244 | -222.1648 | 1993 | 107.9571 | 208.7545 |
| 1984 | 95.86459 | 203.8735 | 1994 | 316.9105 | 340.6221 |
| 1985 | 93.95055 | 60.59586 | 1995 | 340.4558 | 285.1371 |
| 1986 | 218.8049 | 304.1637 | 1996 | 127.2921 | 82.95506 |
| 1987 | 167.8145 | 272.1896 | 1997 | 327.8921 | 362.5652 |
| 1988 | 25.14577 | -1.621064 | 1998 | 186.2857 | 146.5679 |

El modelo: **M2=C(1)+C(2)*FBK2**

| | | | | |
|---|---|---|---|---|
| Dependent Variable: M2 | | | | |
| Method: Least Squares | | | | |
| Sample: 1979 1998 | | Included observations: 20 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 42.01530 | 17.46538 | 2.405634 | 0.0271 |
| FBK2 | 0.711659 | 0.072289 | 9.844579 | 0.0000 |
| R-squared | 0.843363 | Mean dependent var | | 167.4833 |
| Adjusted R-squared | 0.834661 | S.D. dependent var | | 131.3406 |
| S.E. of regression | 53.40551 | Akaike info criterion | | 10.88834 |
| Sum squared resid | 51338.68 | Schwarz criterion | | 10.98792 |
| Log likelihood | -106.8834 | F-statistic | | 96.91574 |
| Durbin-Watson stat | 1.897048 | Prob(F-statistic) | | 0.000000 |

Con este nuevo modelo la autocorrelación queda resuelta, pues observamos que queda un valor de Durbin Watson mayor que el valor crítico d, razón por la cual la autocorrelación no existe.

## 3. Contraste de Autocorrelación de Lagrange

### De Primer Orden

Para el contraste debemos analizar el modelo en el que la variable independiente serán los residuos del modelo inicial **M=C(1)+C(2)*FBK** y que se describe como un modelo autorregresivo; el modelo a regresionar es el siguiente:

**RES0=C(1)+C(2)*FBK+C(3)*RES0(-1)**

Dependent Variable: RES0
Method: Least Squares
Date: 10/02/00 Time: 15:48
Sample(adjusted): 1980 1998
Included observations: 19 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 23.72669 | 65.89525 | 0.360067 | 0.7235 |
| FBK | -0.020123 | 0.068673 | -0.293031 | 0.7733 |
| RES0(-1) | 0.629883 | 0.214244 | 2.940033 | 0.0096 |

| | | | |
|---|---|---|---|
| R-squared | 0.350797 | Mean dependent var | 1.091793 |
| Adjusted R-squared | 0.269647 | S.D. dependent var | 87.52240 |
| S.E. of regression | 74.79727 | Akaike info criterion | 11.61138 |
| Sum squared resid | 89514.11 | Schwarz criterion | 11.76050 |
| Log likelihood | -107.3081 | F-statistic | 4.322801 |
| Durbin-Watson stat | 1.493070 | Prob(F-statistic) | 0.031553 |

RES0 : Residual del modelo inicial.
RES0(-1) : Residual del modelo inicial rezagado un periodo.

**Lagrange : n(numero de datos) * $R^2$.**
**Chi cuadrada : (1 gl, 5%)**

| Obs | LAGR1 | CHI1 |
|---|---|---|
| 1979 | 7.015940 | 0.301084 |

El test de lagrange nos muestra que existe autocorrelación en el modelo, pues el estadístico tiene un valor mucho mayor al valor de Chi cuadrado con

### Segundo Orden

Para el contraste de segundo orden tendremos el modelo anterior con dos rezagos de los residuos.

**RES0=C(1)+C(2)*FBK+C(3)*RES0(-1) +C(4)*RES0(-2)**

| | | | | |
|---|---|---|---|---|
| Dependent Variable: RES0 | | | | |
| Method: Least Squares | | | | |
| Date: 10/02/00 Time: 16:05 | | | | |
| Sample(adjusted): 1981 1998 | | | | |
| Included observations: 18 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -26.91815 | 71.54782 | -0.376226 | 0.7124 |
| FBK | 0.035988 | 0.074240 | 0.484750 | 0.6354 |
| RES0(-1) | 0.815376 | 0.248688 | 3.278712 | 0.0055 |
| RES0(-2) | -0.404311 | 0.286234 | -1.412519 | 0.1796 |
| R-squared | 0.440875 | Mean dependent var | | 6.377633 |
| Adjusted R-squared | 0.321063 | S.D. dependent var | | 86.88314 |
| S.E. of regression | 71.58966 | Akaike info criterion | | 11.57291 |
| Sum squared resid | 71751.10 | Schwarz criterion | | 11.77077 |
| Log likelihood | -100.1562 | F-statistic | | 3.679712 |
| Durbin-Watson stat | 2.037373 | Prob(F-statistic) | | 0.038296 |

RES0 : Residual del modelo inicial.
RES0(-1) : Residual del modelo inicial rezagado un periodo.
RES0(-2) : Residual del modelo inicial rezagado dos periodo.

| Obs | LAGR2 | CHI2 |
|---|---|---|
| 1979 | 8.817500 | 0.147621 |

Al igual que en la prueba de primer orden en la salida de segundo orden se nota que existe autocorrelación.

## 4. Corrección con el procedimiento de Cochran-Orcutt

El primer paso consiste en hallar el modelo y buscar si existe autocorrelación; para este modelo ya se ha demostrado la existencia de autocorrelación, por lo que pasaremos a hallar el coeficiente de autocorrelación según la definición dada para el término$\rho$.

Modelo estimado :M=C(1)+C(2)*FBK+$\mu_t$

Modelo planteado :M*=C(1)+C(2)*FBK

Por lo tanto : **$\mu_t$= M – M***

El cálculo del $\rho$ será a partir de:

$$\rho = \frac{\sum_{t=1}^{20} \mu_t \mu_{t-1}}{\sum_{t=1}^{20} \mu_t^2}$$

Según esto para hallar $\rho$ debemos de calcular el modelo siguiente:

**RES0=C(1)*RES0(-1)**

| | | | | |
|---|---|---|---|---|
| Dependent Variable: RES0 | | | | |
| Method: Least Squares | | | | |
| Date: 10/02/00 Time: 17:32 | | | | |
| Sample(adjusted): 1980 1998 | | | | |
| Included observations: 19 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| RES0(-1) | 0.619525 | 0.201704 | 3.071456 | 0.0066 |
| R-squared | 0.343768 | Mean dependent var | | 1.091793 |
| Adjusted R-squared | 0.343768 | S.D. dependent var | | 87.52240 |
| S.E. of regression | 70.90026 | Akaike info criterion | | 11.41162 |
| Sum squared resid | 90483.24 | Schwarz criterion | | 11.46133 |
| Log likelihood | -107.4104 | Durbin-Watson stat | | 1.516000 |

Por lo tanto el ρ será 0.619525

Corregimos las variables importaciones y formación bruta de capitales (M y FBK) con el coeficiente de autocorrelación hallado, luego se estima el modelo:

**M1=C(1)+C(2)*FBK1**

Donde M1 y FBK1 son los nuevos valores para M y FBK

| | | | | |
|---|---|---|---|---|
| Dependent Variable: M1 | | | | |
| Method: Least Squares | | | | |
| Date: 10/02/00 Time: 17:39 | | | | |
| Sample: 1979 1998 | | | | |
| Included observations: 20 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 43.12060 | 17.79588 | 2.423067 | 0.0262 |
| FBK1 | 0.738521 | 0.067116 | 11.00359 | 0.0000 |
| R-squared | 0.870577 | Mean dependent var | | 187.0030 |
| Adjusted R-squared | 0.863387 | S.D. dependent var | | 146.0555 |
| S.E. of regression | 53.98386 | Akaike info criterion | | 10.90989 |
| Sum squared resid | 52456.63 | Schwarz criterion | | 11.00946 |
| Log likelihood | -107.0989 | F-statistic | | 121.0791 |
| Durbin-Watson stat | 1.422102 | Prob(F-statistic) | | 0.000000 |

De esta corrección tenemos que la autocorrelación desaparece pues el estadístico Durbin Watson se encuentra en la zona de no autocorrelación.

**Ejercicio ilustrativo 6:**

De la siguiente información estadística, referido a las Importaciones, Producto Bruto Interno e Índice de Precios al Consumidor, para el Perú en el periodo 1970-2000:

| AÑO | PBI | IPC | IMPORT | AÑO | PBI | IPC | IMPORT |
|---|---|---|---|---|---|---|---|
| 1970 | 62,022.00 | 0.00000019 | 9,491.00 | 1986 | 99,267.00 | 0.00016300 | 10,607.00 |
| 1971 | 64,627.00 | 0.00000021 | 9,920.00 | 1987 | 107,208.00 | 0.00030200 | 12,182.00 |
| 1972 | 66,501.00 | 0.00000022 | 9,892.00 | 1988 | 97,881.00 | 0.00231900 | 11,091.00 |
| 1973 | 70,092.00 | 0.00000024 | 11,110.00 | 1989 | 86,429.00 | 0.08100000 | 8,287.00 |
| 1974 | 76,611.00 | 0.00000028 | 14,127.00 | 1990 | 81,983.00 | 6.15000000 | 9,270.00 |
| 1975 | 79,215.00 | 0.00000035 | 13,724.00 | 1991 | 83,760.00 | 31.35000000 | 11,130.00 |
| 1976 | 80,800.00 | 0.00000047 | 11,995.00 | 1992 | 83,401.00 | 54.39 | 12,113.00 |
| 1977 | 81,123.00 | 0.00000064 | 12,037.00 | 1993 | 87,375.00 | 80.82 | 12,572.00 |
| 1978 | 81,366.00 | 0.00000102 | 9,103.00 | 1994 | 98,577.00 | 100.00 | 15,922.00 |
| 1979 | 86,086.00 | 0.00000170 | 10,860.00 | 1995 | 107,039.00 | 111.13 | 20,232.00 |
| 1980 | 90,562.00 | 0.00000271 | 14,144.00 | 1996 | 109,709.00 | 123.97 | 20,259.00 |
| 1981 | 95,181.00 | 0.00000480 | 16,395.00 | 1997 | 117,110.00 | 134.56 | 22,724.00 |
| 1982 | 94,610.00 | 0.00000780 | 16,758.00 | 1998 | 116,485.00 | 144.32 | 23,251.00 |
| 1983 | 83,446.00 | 0.00001650 | 11,791.00 | 1999 | 117,590.00 | 149.33 | 19,724.00 |
| 1984 | 87,785.00 | 0.00003470 | 9,647.00 | 2000 | 121,267.00 | 154.94 | 20,428.00 |
| 1985 | 90,243.00 | 0.00009150 | 8,812.00 | | | | |

Se estima el modelo:
IMPORT = f(PBI, IPC)
IMPORT = $\beta_1 + \beta_1$ PBI + $\beta_1$ IPC

Siendo:
IMPORT = Importaciones. (En millones de nuevos soles de 1994)
PBI = Producto Bruto Interno. (En millones de nuevos soles de 1994)
IPC = Índice de Precios al Consumidor. (Base 1994 = 100)

**1. Cálculo e interpretación del modelo por M.C.O.**

**ls importaciones c pbi ipc**

Dependent Variable: IMPORT
Method: Least Squares
Date: 09/05/02 Time: 21:12
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 3427.817 | 3293.850 | 1.040672 | 0.3069 |
| IPC | 45.02326 | 11.06852 | 4.067687 | 0.0004 |
| PBI | 0.094182 | 0.039275 | 2.398012 | 0.0234 |

| | | | |
|---|---|---|---|
| R-squared | 0.755860 | Mean dependent var | 13535.42 |
| Adjusted R-squared | 0.738422 | S.D. dependent var | 4362.854 |
| S.E. of regression | 2231.369 | Akaike info criterion | 18.35038 |
| Sum squared resid | 1.39E+08 | Schwarz criterion | 18.48916 |
| Log likelihood | -281.4309 | F-statistic | 43.34425 |
| Durbin-Watson stat | 0.655933 | Prob(F-statistic) | 0.000000 |

***Ecuación estimada: IMPORT = $\beta_1 + \beta_1$*PBI + $\beta_1$*IPC***
IMPORT = 3427.817007 + 45.0232614*IPC + 0.0941819102*PBI

**Interpretación:**

- Los parámetros son significativos, a excepción del intercepto, a un nivel de significancia del 5%. El estadístico F refleja que la variable dependiente es explicado por el modelo en su conjunto y la bondad de ajuste es de 75.58%.
- Las Importaciones, autónoma del IPC y PBI asciende en promedio a 3427.817 millones de nuevos soles.
- Ante una variación de 1 millón de nuevos soles en el IPC, las importaciones se incrementará en un 45.0.23 millones de nuevos soles, manteniendo las demás variables constantes.
- Ante una variación de 1 millón de nuevos soles en el PBI, las importaciones se incrementaran en 0.09418191 millones de nuevos soles, manteniendo las demás variables constantes.

2. **Significancia individual y conjunta del Modelo.**

Individualmente el modelo es significativo demostrado mediante la prueba t, cuyas probabilidades son menores que 0.05 a excepción del intercepto.

En conjunto el modelo es significativo, demostrado por la prueba F al ser su probabilidad menor a 0.05. Sin embargo, la bondad de ajuste, medida a través del $R^2$ y $R^2$ ajustado, demuestra regular ajuste. (73% a 75%).

3. **Especificación del Modelo**

Test de Reset Ramsey

H0 : El Modelo esta correcta especificado.

H1 : El Modelo no esta correcta especificado.

| Ramsey RESET Test: | | | | |
|---|---|---|---|---|
| F-statistic | 1.988089 | Probability | | 0.157240 |
| Log likelihood ratio | 4.411501 | Probability | | 0.110168 |
| Test Equation: | | | | |
| Dependent Variable: IMPORT | | | | |
| Method: Least Squares | | | | |
| Date: 09/05/02 Time: 22:22 | | | | |
| Sample: 1970 2000 | | Included observations: 31 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 22336.85 | 10694.62 | 2.088607 | 0.0467 |
| IPC | -667.4942 | 420.3267 | -1.588037 | 0.1244 |
| PBI | -1.281869 | 0.825272 | -1.553268 | 0.1324 |
| FITTED^2 | 0.000976 | 0.000607 | 1.607575 | 0.1200 |
| FITTED^3 | -1.99E-08 | 1.30E-08 | -1.526744 | 0.1389 |
| R-squared | 0.788244 | Mean dependent var | | 13535.42 |
| Adjusted R-squared | 0.755666 | S.D. dependent var | | 4362.854 |
| S.E. of regression | 2156.563 | Akaike info criterion | | 18.33711 |
| Sum squared resid | 1.21E+08 | Schwarz criterion | | 18.56840 |
| Log likelihood | -279.2252 | F-statistic | | 24.19574 |
| Durbin-Watson stat | 0.721749 | Prob(F-statistic) | | 0.000000 |

Como la probabilidad asociada este estadístico es Prob. = 0.157240 > 0.05. Entonces se acepta la hipótesis nula. Es decir el Modelo esta correctamente especificado, con un nivel de significancia del 5%.

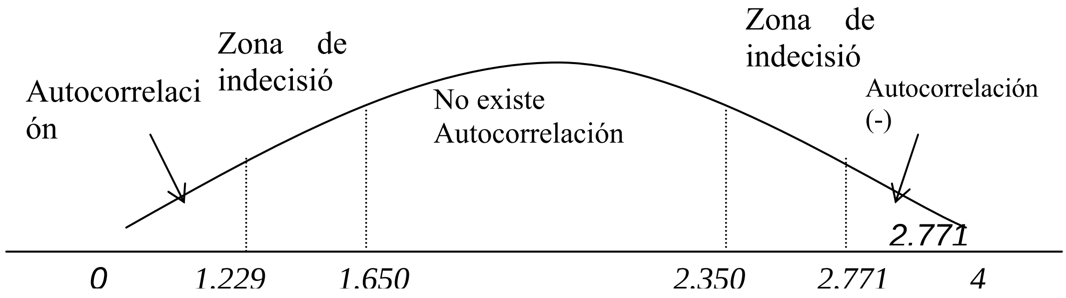
## 4. Detección de Autocorrelación

### a. Estadístico Durbin Watson

Como sabemos el estadístico Durbin Watson mide la autocorrelación de primer grado.

Del modelo:

$$\begin{array}{c} k = 3 \\ n = 31 \\ \alpha = 0.05 \end{array} \Rightarrow \begin{array}{l} D_L = 1.229 \\ 4 - D_L = 2.771 \\ D_U = 1.650 \\ 4 - D_U = 2.35 \end{array}$$




Del Output tenemos que el indicador Durbin Watson es 0.655933, por lo tanto podemos afirmar que existen problemas de autocorrelación positiva de primer orden con un nivel de significancia del 5%.

### b. Test de Lagrange

$H_0$ : El Modelo no tiene autocorrelación de orden p (p =2).
$H_1$ : El Modelo si tiene autocorrelación de orden p (p =2).

| Breusch-Godfrey Serial Correlation LM Test: | | | |
|---|---|---|---|
| F-statistic | 15.24058 | Probability | 0.000042 |
| Obs*R-squared | 16.72975 | Probability | 0.000233 |

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 09/06/02 Time: 03:49

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -482.4213 | 2323.977 | -0.207584 | 0.8372 |
| IPC | -1.393453 | 7.800520 | -0.178636 | 0.8596 |
| PBI | 0.005824 | 0.027703 | 0.210227 | 0.8351 |
| RESID(-1) | 0.944639 | 0.179381 | 5.266111 | 0.0000 |
| RESID(-2) | -0.407837 | 0.181036 | -2.252797 | 0.0329 |

| | | | |
|---|---|---|---|
| R-squared | 0.539669 | Mean dependent var | 2.57E-13 |
| Adjusted R-squared | 0.468849 | S.D. dependent var | 2155.707 |
| S.E. of regression | 1571.081 | Akaike info criterion | 17.70361 |
| Sum squared resid | 64175681 | Schwarz criterion | 17.93489 |
| Log likelihood | -269.4059 | F-statistic | 7.620288 |
| Durbin-Watson stat | 1.948734 | Prob(F-statistic) | 0.000334 |

Como Prob.= 0.000042 < 0.05. Entonces se rechaza la hipótesis nula. Es decir el Modelo tiene problemas de autocorrelación con una probabilidad del 95%.

### c. Test Estadístico Ljung y Box

$H_0$ : El Modelo no tiene autocorrelación hasta el orden p(p =2).
$H_1$ : El Modelo si tiene autocorrelación de orden p (p =2).

Date: 09/06/02 Time: 04:06
Sample: 1970 2000
Included observations: 31

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. |***** |</td><td>. |***** |</td><td>1</td><td>0.665</td><td>0.665</td><td>15.073</td><td>0.000</td></tr>
<tr><td>. |**. |</td><td>***| . |</td><td>2</td><td>0.217</td><td>-0.404</td><td>16.730</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>3</td><td>-0.040</td><td>0.055</td><td>16.790</td><td>0.001</td></tr>
<tr><td>. *| . |</td><td>. *| . |</td><td>4</td><td>-0.134</td><td>-0.073</td><td>17.472</td><td>0.002</td></tr>
<tr><td>. | . |</td><td>. |**. |</td><td>5</td><td>0.008</td><td>0.299</td><td>17.475</td><td>0.004</td></tr>
<tr><td>. |* . |</td><td>. *| . |</td><td>6</td><td>0.145</td><td>-0.104</td><td>18.333</td><td>0.005</td></tr>
<tr><td>. |* . |</td><td>. | . |</td><td>7</td><td>0.125</td><td>-0.027</td><td>19.004</td><td>0.008</td></tr>
<tr><td>. | . |</td><td>. *| . |</td><td>8</td><td>-0.018</td><td>-0.173</td><td>19.019</td><td>0.015</td></tr>
<tr><td>. *| . |</td><td>. | . |</td><td>9</td><td>-0.171</td><td>0.013</td><td>20.380</td><td>0.016</td></tr>
<tr><td>.**| . |</td><td>. *| . |</td><td>10</td><td>-0.240</td><td>-0.106</td><td>23.195</td><td>0.010</td></tr>
<tr><td>.**| . |</td><td>.**| . |</td><td>11</td><td>-0.296</td><td>-0.233</td><td>27.688</td><td>0.004</td></tr>
<tr><td>.**| . |</td><td>. *| . |</td><td>12</td><td>-0.303</td><td>-0.090</td><td>32.628</td><td>0.001</td></tr>
<tr><td>. *| . |</td><td>. |* . |</td><td>13</td><td>-0.169</td><td>0.160</td><td>34.248</td><td>0.001</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>14</td><td>-0.034</td><td>-0.012</td><td>34.319</td><td>0.002</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>15</td><td>0.057</td><td>0.051</td><td>34.525</td><td>0.003</td></tr>
<tr><td>. | . |</td><td>.**| . |</td><td>16</td><td>0.026</td><td>-0.198</td><td>34.571</td><td>0.005</td></tr>
</table>

Debido a que las probabilidades son menores a 0.05 se puede decir que existe autocorrelación de orden 1,2,3 ...15 o 16, con una probabilidad del 95%

## 5. Corrección de la Autocorrelación

### a. Estadístico Durbin Watson

Analizaremos la autocorrelación de primer orden

**Primer cambio de variable.**
El coeficiente estimado de la variable rezagada nos servirá para corregir a las variables
Así corregimos nuevamente con $d = 0.655933$
Además:
$d = 2(1-\rho) \Rightarrow \rho = 1 - d/2$
$\rho = 0.6720335$

Así generando nuevas variables:

1. importd = import-0.6720335*import(-1) , de 1971 –2000
2. pbid = pbi-0.6720335*pbi(-1) , de 1971 –2000
3. ipcd = ipc-0.6720335*ipc(-1) , de 1971 –2000

1. importd = ((1-0.6720335^2)^(1/2))*import , de 1970.
2. pbid = ((1-0.6720335^2)^(1/2))*pbi , de 1970.
3. ipcd = ((1-0.6720335^2)^(1/2))*ipc , de 1970.

Dependent Variable: IMPORTD
Method: Least Squares
Date: 09/19/02 Time: 10:17
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -2119.581 | 1413.029 | -1.500026 | 0.1448 |
| PBID | 0.207043 | 0.047158 | 4.390456 | 0.0001 |
| IPCD | 22.84478 | 15.22981 | 1.500004 | 0.1448 |

| | | | |
|---|---|---|---|
| R-squared | 0.598372 | Mean dependent var | 4802.570 |
| Adjusted R-squared | 0.569684 | S.D. dependent var | 2337.408 |
| S.E. of regression | 1533.303 | Akaike info criterion | 17.60000 |
| Sum squared resid | 65828514 | Schwarz criterion | 17.73877 |
| Log likelihood | -269.8000 | F-statistic | 20.85814 |
| Durbin-Watson stat | 1.327187 | Prob(F-statistic) | 0.000003 |

**Segundo Cambio de Variable**
El coeficiente estimado de la variable rezagada nos servirá para corregir a las variables

Así corregimos nuevamente con $d = 1.327187$

Además:
d = 2 ( 1-$\rho$ ) $\Rightarrow$ $\rho$ = 1 – d/2
$\rho$ = 0.3364065

Así generando nuevas variables:
1. importd1p = importd-0.3364065*importd(-1) , de 1971 –2000
2. pbid1 = pbid-0.3364065*pbid(-1) , de 1971 –2000
3. ipcd1 = ipcd-0.3364065*ipcd(-1) , de 1971 –2000

1. importd1 = ((1-0.3364065^2)^(1/2))*importd , de 1970.
2. pbid1 = ((1-0.3364065^2)^(1/2))*pbid , de 1970.
3. ipcd1 = ((1-0.3364065^2)^(1/2))*ipcd , de 1970.

Dependent Variable: IMPORTD1
Method: Least Squares
Date: 09/19/02 Time: 10:28
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -1355.666 | 854.7584 | -1.586022 | 0.1240 |
| PBID1 | 0.203359 | 0.040381 | 5.036036 | 0.0000 |
| IPCD1 | 23.06234 | 18.77573 | 1.228305 | 0.2296 |

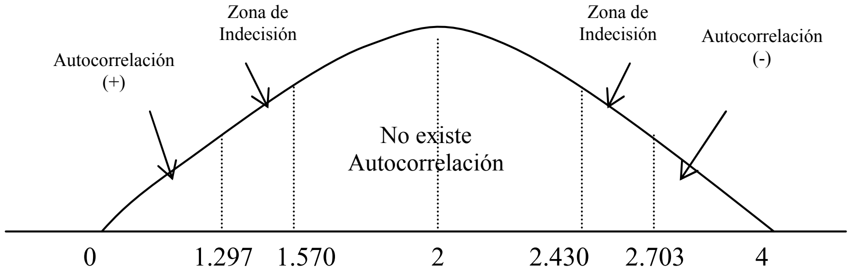
| | | | |
|---|---|---|---|
| R-squared | 0.566868 | Mean dependent var | 3251.578 |
| Adjusted R-squared | 0.535930 | S.D. dependent var | 2134.947 |
| S.E. of regression | 1454.383 | Akaike info criterion | 17.49432 |
| Sum squared resid | 59226466 | Schwarz criterion | 17.63309 |
| Log likelihood | -268.1619 | F-statistic | 18.32270 |
| Durbin-Watson stat | 1.947398 | Prob(F-statistic) | 0.000008 |

Midiendo la autocorrelación de primer orden:

Donde:
k` = numero de variables explicativas
(numero de parámetros menos 1)
n = numero de observaciones

$\Rightarrow$ $D_{k`;\, n;\, \alpha} = D_{2`;\, 31;\, 0.05}$ :

$D_L$=1.297
4-$D_L$=2.703
$D_U$=1.570
4-$D_U$=2.430



Se observa que estadístico Durbin Watson es 1.947398, encontrándose en la zona de no autocorrelación y tiene un valor cercano a 2 por lo tanto se puede afirmar que modelo ya no tiene problemas de autocorrelación. Sin embargo la variable IPCD1 pierde significancia individual, así como pierde bondad de ajuste, aun cuando el modelo en conjunto sigue explicando a la variable dependiente (Prueba F).

**b. Procedimiento Iterativo de Cochran Orcutt**

Mediante el proceso Cochran Orcutt lo que se pretende es obtener una estimación para la autocorrelación, el cual esta basado en la estimación del ρ como coeficiente de autocorrelación propiamente dicha.

$$\rho = \frac{\sum \mu_t \mu_{t-1}}{\sum \mu_t^2}$$

Asimismo se puede estimar con el modelo: $\mu_t = \mu_{t-1} + \varepsilon$

**Primera Etapa**: Regresionamos el modelo por MCO:

Dependent Variable: RESID1
Method: Least Squares
Date: 09/19/02 Time: 12:29
Sample(adjusted): 1971 2000
Included observations: 30 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -38.80981 | 302.4778 | -0.128306 | 0.8988 |
| RESID1(-1) | 0.674691 | 0.141307 | 4.774637 | 0.0001 |

| | | | |
|---|---|---|---|
| R-squared | 0.448788 | Mean dependent var | -7.394418 |
| Adjusted R-squared | 0.429102 | S.D. dependent var | 2192.159 |
| S.E. of regression | 1656.347 | Akaike info criterion | 17.72696 |
| Sum squared resid | 76817625 | Schwarz criterion | 17.82037 |
| Log likelihood | -263.9044 | F-statistic | 22.79716 |
| Durbin-Watson stat | 1.458345 | Prob(F-statistic) | 0.000051 |

El coeficiente estimado de la variable RESID1 rezagada es 0.674691 el cual aproximamos como un $\hat{\rho} = 0.674691$, cuyo valor servirá para corregir a las variables. Para ello debo generar nuevas variables:

Así generando nuevas variables:

1. import2 = import-0.674691*import(-1) , de 1971 –2000
2. pbi2 = pbi-0.674691*pbi(-1) , de 1971 –2000
3. ipc2 = ipc-0.674691*ipc(-1) , de 1971 –2000

1. import2 = ((1-0.674691^2)^(1/2))*import , de 1970.
2. pbi2 = ((1-0.674691^2)^(1/2))*pbi , de 1970.
3. ipc2 = ((1-0.674691^2)^(1/2))*ipc , de 1970.

Haciendo la regresión:

| Dependent Variable: IMPORT2<br>Method: Least Squares<br>Date: 09/19/02 Time: 12:40<br>Sample: 1970 2000<br>Included observations: 31 | | | | |
|---|---|---|---|---|
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -2117.422 | 1401.949 | -1.510342 | 0.1422 |
| PBI2 | 0.207486 | 0.047117 | 4.403617 | 0.0001 |
| IPC2 | 22.80412 | 15.27169 | 1.493228 | 0.1466 |
| R-squared | 0.597614 | Mean dependent var | | 4767.610 |
| Adjusted R-squared | 0.568873 | S.D. dependent var | | 2333.324 |
| S.E. of regression | 1532.067 | Akaike info criterion | | 17.59839 |
| Sum squared resid | 65722441 | Schwarz criterion | | 17.73716 |
| Log likelihood | -269.7750 | F-statistic | | 20.79250 |
| Durbin-Watson stat | 1.331629 | Prob(F-statistic) | | 0.000003 |

**Segunda Etapa**: Regresionamos el modelo por MCO:
Con los residuos obtenidos en el último modelo: u* t = ρ*u*t-1 + ε*t

| Dependent Variable: RESID2<br>Method: Least Squares<br>Date: 09/19/02 Time: 12:44<br>Sample(adjusted): 1971 2000<br>Included observations: 30 after adjusting endpoints | | | | |
|---|---|---|---|---|
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 4.803428 | 263.6973 | 0.018216 | 0.9856 |
| RESID2(-1) | 0.331815 | 0.178823 | 1.855545 | 0.0741 |
| R-squared | 0.109501 | Mean dependent var | | 12.52136 |
| Adjusted R-squared | 0.077698 | S.D. dependent var | | 1503.750 |
| S.E. of regression | 1444.150 | Akaike info criterion | | 17.45277 |
| Sum squared resid | 58395919 | Schwarz criterion | | 17.54618 |
| Log likelihood | -259.7915 | F-statistic | | 3.443048 |
| Durbin-Watson stat | 1.904826 | Prob(F-statistic) | | 0.074074 |

El coeficiente estimado de la variable RESID2 rezagada es 0.331815 el cual aproximamos como un $\hat{\rho} = 0.331815$, cuyo valor servirá para corregir a las variables. Para ello debo generar nuevas variables:
Así generando nuevas variables:

1.- import2a = import2-0.331815*import2(-1) , de 1971 –2000
2.- pbi2a = pbi2-0.331815*pbi2(-1) , de 1971 –2000
3.- ipc2a = ipc2-0.331815*ipc2(-1) , de 1971 –2000

1.- import2a = ((1-0.331815^2)^(1/2))*import2 , de 1970.
2.- pbi2a = ((1-0.331815^2)^(1/2))*pbi2 , de 1970.
3.- ipc2a = ((1-0.331815^2)^(1/2))*ipc2 , de 1970.

Haciendo la regresión:

| Dependent Variable: IMPORT2A<br>Method: Least Squares<br>Date: 09/19/02 Time: 13:12<br>Sample: 1970 2000<br>Included observations: 31 | | | | |
|---|---|---|---|---|
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -1359.692 | 855.4361 | -1.589473 | 0.1232 |
| PBI2A | 0.203597 | 0.040451 | 5.033229 | 0.0000 |
| IPC2A | 23.06025 | 18.77613 | 1.228168 | 0.2296 |
| R-squared | 0.566682 | Mean dependent var | | 3249.057 |
| Adjusted R-squared | 0.535730 | S.D. dependent var | | 2134.615 |
| S.E. of regression | 1454.470 | Akaike info criterion | | 17.49444 |
| Sum squared resid | 59233498 | Schwarz criterion | | 17.63321 |
| Log likelihood | -268.1638 | F-statistic | | 18.30881 |
| Durbin-Watson stat | 1.943394 | Prob(F-statistic) | | 0.000008 |

Mediante el estadístico Durbin Watson, que no existe Autocorrelación, pero el modelo presenta pierde significancia individual en la variable IPC 2A, así como disminuye ligeramente su bondad de ajuste.

Realizaremos un nuevo proceso para analizar los cambios en la percepción de autocorrelación, significancia y bondad de ajuste:

**Tercera Etapa:** Regresionamos el modelo por MCO:
Con los residuos obtenidos en el ultimo modelo: $u^{**}\,t = \rho^{**}u^{**}t\text{-}1 + \varepsilon^{**}t$

| Dependent Variable: RESID3<br>Method: Least Squares<br>Date: 09/19/02 Time: 13:15<br>Sample(adjusted): 1971 2000<br>Included observations: 30 after adjusting endpoints | | | | |
|---|---|---|---|---|
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 27.75120 | 263.9197 | 0.105150 | 0.9170 |
| RESID3(-1) | 0.021544 | 0.188060 | 0.114559 | 0.9096 |
| R-squared | 0.000468 | Mean dependent var | | 27.47374 |
| Adjusted R-squared | -0.035229 | S.D. dependent var | | 1420.679 |
| S.E. of regression | 1445.487 | Akaike info criterion | | 17.45462 |
| Sum squared resid | 58504108 | Schwarz criterion | | 17.54803 |
| Log likelihood | -259.8193 | F-statistic | | 0.013124 |
| Durbin-Watson stat | 1.947803 | Prob(F-statistic) | | 0.909612 |

El coeficiente estimado de la variable RESID3 rezagada es 0.021544 el cual aproximamos como un $\hat{\rho} = 0.021544$, cuyo valor servirá para corregir a las variables. Para ello debo generar nuevas variables:

Así generando nuevas variables:

1.- import2b = import2a-0.021544*import2a(-1) , de 1971 –2000
2.- pbi2b = pbi2a-0.021544*pbi2a(-1) , de 1971 –2000
3.- ipc2b = ipc2a-0.021544*ipc2a(-1) , de 1971 –2000

1.- import2b = ((1-0.021544^2)^(1/2))*import2a , de 1970.
2.- pbi2b = ((1-0.021544^2)^(1/2))*pbi2a , de 1970.
3.- ipc2b = ((1-0.021544^2)^(1/2))*ipc2a , de 1970.

Haciendo la regresión:

Dependent Variable: IMPORT2B
Method: Least Squares
Date: 09/19/02 Time: 13:23
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -1312.642 | 833.4782 | -1.574897 | 0.1265 |
| PBI2B | 0.202633 | 0.040064 | 5.057783 | 0.0000 |
| IPC2B | 23.23856 | 19.07484 | 1.218283 | 0.2333 |

| | | | |
|---|---|---|---|
| R-squared | 0.565187 | Mean dependent var | 3183.027 |
| Adjusted R-squared | 0.534129 | S.D. dependent var | 2131.303 |
| S.E. of regression | 1454.716 | Akaike info criterion | 17.49477 |
| Sum squared resid | 59253578 | Schwarz criterion | 17.63355 |
| Log likelihood | -268.1690 | F-statistic | 18.19773 |
| Durbin-Watson stat | 1.982754 | Prob(F-statistic) | 0.000009 |

Una conclusión preliminar respecto a la autocorrelación empleando esta metodología es que solo será necesario regresionar en dos etapas, puesto hasta esta etapa se superó el problema de autocorrelación, seguir otra etapa significaría reducir (ligeramente) la significancia de los parámetros.

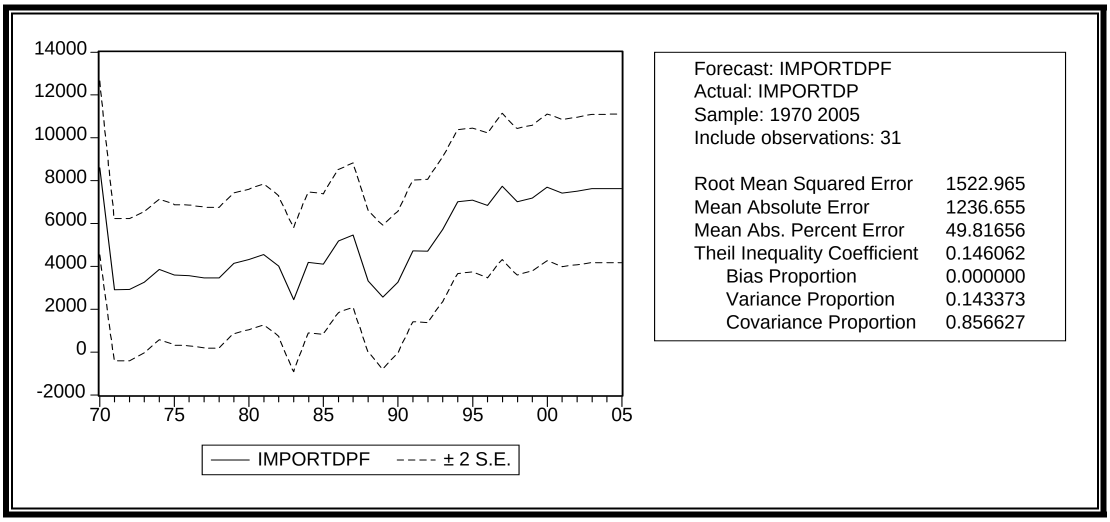
## 6. Predicción

**Corregida la autocorrelación precedemos a la predicción:**

En base a los siguientes pronósticos:

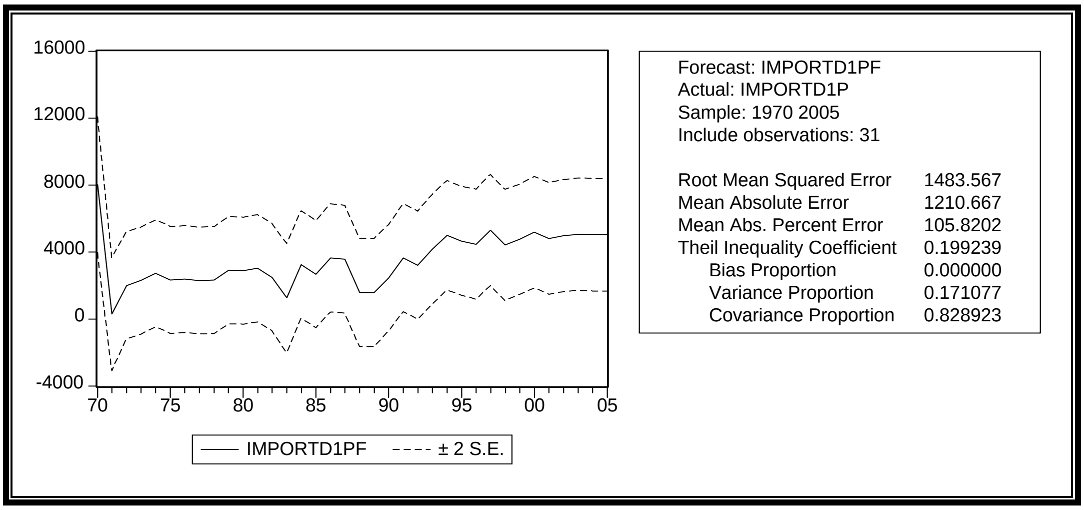
| Año | IMPORT | PBI | IPC |
|---|---|---|---|
| 2001 | | 121568 | 160 |
| 2002 | | 122036 | 165.02 |
| 2003 | | 122657 | 170.25 |
| 2004 | | 122975 | 174.36 |
| 2005 | | 123278 | 176.89 |

a. **Para el Modelo**: ls importd c pbid ipcd




| Año | importdp |
|---|---|
| 2001 | 7418.589 |
| 2002 | 7515.985 |
| 2003 | 7627.905 |
| 2004 | 7634.972 |
| 2005 | 7639.592 |

**b. Para el Modelo**: ls importd1 c pbid1 ipcd1




Se observa en las salidas de la predicción que el coeficiente de Theil del segundo modelo predicho (19.92%), es mayor al obtenido por el primero (14.60%), entonces se podría deducir que el primer modelo es mejor para la predicción.

| Año | Importd1p |
|---|---|
| 2001 | 4806.349 |
| 2002 | 4980.457 |
| 2003 | 5059.087 |
| 2004 | 5031.870 |
| 2005 | 5030.988 |

## 7. Corrección de Autocorrelación de Segundo Orden

Se ha observado que el modelo tiene problemas de autocorrelación de segundo orden por ello se ha propuesto la corrección de este mediante el modelo de Cochran Orcutt.

Regresionamos el modelo original:

Dependent Variable: RESID1
Method: Least Squares
Date: 09/19/02 Time: 15:52
Sample(adjusted): 1972 2000
Included observations: 29 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -19.05861 | 292.1253 | -0.065241 | 0.9485 |
| RESID1(-1) | 0.943024 | 0.179419 | 5.255986 | 0.0000 |
| RESID1(-2) | -0.407993 | 0.181040 | -2.253605 | 0.0329 |

| | | | |
|---|---|---|---|
| R-squared | 0.538820 | Mean dependent var | -21.63177 |
| Adjusted R-squared | 0.503345 | S.D. dependent var | 2229.550 |
| S.E. of regression | 1571.248 | Akaike info criterion | 17.65483 |
| Sum squared resid | 64189322 | Schwarz criterion | 17.79627 |
| Log likelihood | -252.9950 | F-statistic | 15.18856 |
| Durbin-Watson stat | 1.968664 | Prob(F-statistic) | 0.000043 |

Utilizar los coeficientes estimados de las variables explicadas resid1(-1) y resid1(-2), para transformar las variables:

Así generando nuevas variables:

1.-import2o = import-0.943024*import(-1)+ 0.407993*import(-2),de 1971 –2000
2.-pbi2o = pbi-0.943024*pbi(-1)+ 0.407993*pbi(-2) , de 1971 –2000
3.-ipc2o = ipc-0.943024*ipc(-1)+ 0.407993*ipc(-2) , de 1971 –2000

1.-import2o = ((1-0.407993^2)^(1/2))*import , de 1971.
2.-pbi2o = ((1-0.407993^2)^(1/2))*pbi , de 1971.
3.-ipc2o = ((1-0.407993^2)^(1/2))*ipc , de 1971.

Dependent Variable: IMPORT2O
Method: Least Squares
Date: 09/19/02 Time: 16:11
Sample(adjusted): 1971 2000
Included observations: 30 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -968.6563 | 1734.464 | -0.558476 | 0.5811 |
| PBI2O | 0.158657 | 0.041787 | 3.796818 | 0.0008 |
| IPC2O | 31.29594 | 12.85795 | 2.433976 | 0.0218 |

| | | | |
|---|---|---|---|
| R-squared | 0.636849 | Mean dependent var | 6552.616 |
| Adjusted R-squared | 0.609949 | S.D. dependent var | 2399.435 |
| S.E. of regression | 1498.544 | Akaike info criterion | 17.55702 |
| Sum squared resid | 60632136 | Schwarz criterion | 17.69713 |
| Log likelihood | -260.3552 | F-statistic | 23.67464 |
| Durbin-Watson stat | 1.831468 | Prob(F-statistic) | 0.000001 |

En conclusión superamos el problema de autocorrelación manteniendo la significancia individual de los coeficientes de las variables, igualmente la significancia colectiva y mejorando la bondad de ajuste.

En conclusión consideramos este modelo para la predicción.

## 8. Predicción (autocorrelación de orden 2)

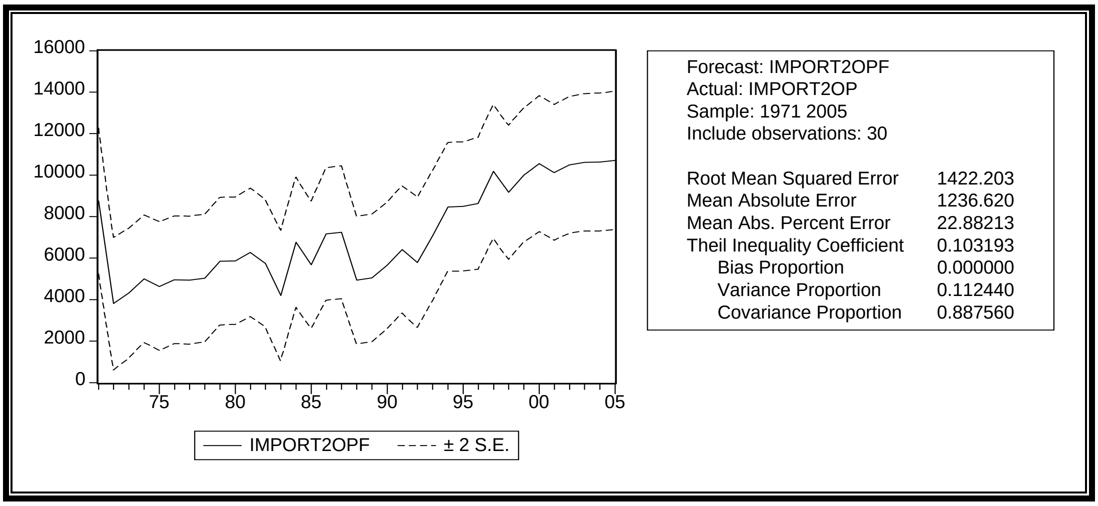
Expandiendo el rango generamos las siguientes variables:

1.-import2oP = import-0.943024*import(-1)+ 0.407993*import(-2) ,de 1971 –2005
2.-pbi2oP = pbi-0.943024*pbi(-1)+ 0.407993*pbi(-2) , de 1971 –2005
3.-ipc2oP = ipc-0.943024*ipc(-1)+ 0.407993*ipc(-2) , de 1971 –2005

1.-import2oP = ((1-0.407993^2)^(1/2))*import , de 1971.
2.-pbi2oP = ((1-0.407993^2)^(1/2))*pbi , de 1971.
3.-ipc2oP = ((1-0.407993^2)^(1/2))*ipc , de 1971.

Efectuando la regresión y la predicción respectiva obtenemos los siguientes resultados:

| Año | Import2o |
|---|---|
| 2001 | 10129.56 |
| 2002 | 10494.93 |
| 2003 | 10618.20 |
| 2004 | 10638.31 |
| 2005 | 10704.95 |




Entonces hallando la predicción del 2001:
$import2oP_{2001} = import_{2001} - 0.943024*import_{2000} + 0.407993*import_{1999}$
$import_{2001} = import2oP_{2001} + 0.943024*import_{2000} - 0.407993*import_{1999}$
$import_{2001} = (10129.56) + 0.943024*(20428) - 0.407993*(19724)$
$import_{2001} = 21346.40$

# LECCIÓN 4

## HETEROSCEDASTICIDAD

La utilización del modelo lineal general se refiere a que las varianzas de las perturbaciones sean las mismas (homocedasticidad). Hablar del problema de heterocedasticidad significa decir que las varianzas de las perturbaciones son diferentes, su presencia afecta a la elaboración de la banda de confianza para precisar la predicción de la variable endógena.

La heterocedasticidad se presenta generalmente en estudios de corte transversal y cuando se analizan modelos en los cuales las variables que intervienen tienen diferenciaciones como consecuencias de la estratificación de las unidades de observación. Por ejemplo: análisis de ventas, costos, utilidades, ahorro, niveles de productividad por tamaño de la empresa (número de personal ocupado - Valor del producto). Cuando se estudia el consumo, ahorro, etc. por estrato de ingreso del hogar

Planteamiento matricial de la presencia de heterocedasticidad:

$$E(\mu\mu') = \sigma^2\Omega = \sigma^2\begin{bmatrix} 1/\lambda_1 & 0 & ... & 0 \\ 0 & 1/\lambda_2 & ... & 0 \\ : & : & ... & : \\ 0 & 0 & ... & 1/\lambda_n \end{bmatrix}$$

### 1. CONSECUENCIAS DE LA HETEROCEDASTICIDAD

**a.** Los estimadores por MCO (Mínimos Cuadrados Ordinarios) siguen siendo insesgados consistentes, pero ya no son eficientes para un tipo de muestra, grande o pequeña.

**b.** Si se usa la fórmula de la varianza del modelo con heterocedasticidad en lugar del modelo lineal general, el intervalo para $\beta i$ es innecesariamente ancho y la prueba de significancia tiene menos fuerza.

**c.** El problema empeora si en lugar de utilizar la fórmula para heterocedasticidad (MCO), se usa la fórmula tradicional.

La fórmula correcta para estimar la varianza en presencia de heterocedasticidad (MCO) es la siguiente:

$$Var(\hat{\beta}_i) = \frac{\sum X_i^2\sigma_i^2}{(\sum X_i^2)^2} \text{ ; para un solo parámetro.}$$

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega(X'X)^{-1} \text{ ; para todos los parámetros.}$$

Como consecuencia de lo anterior, si en condiciones de heterocedasticidad continuamos aplicando equivocadamente la fórmula de MCO las pruebas t y F tienden a exagerar la significancia estadística de los parámetros estimados t y F.

## 2. DETECCIÓN DE HETEROSCEDASTICIDAD

### a. Test de Goldfeld – Quandt

Este método es aplicable si se supone que la varianza heterocedástica $\sigma_i^2$ está relacionada positivamente con una de las variables explicativas en el modelo de regresión. Por tanto se efectúa a partir de la siguiente hipótesis de trabajo:

- Si existe heterocedasticidad ella se debe a que la varianza de la $\mu_s$ es una función monótona de una de las $X_s$.
- Inicialmente es necesario realizar gráficos de los residuales con respecto a la variable Y, y a cada uno de las $X_s$.
- Si observa que el residual (el residual es un estimador del término de perturbación) tiene comportamientos ascendentes o decrecientes con respecto a Y se puede suponer que existe heterocedasticidad.
- Se puede suponer que si la variable X tiene este problema, es la que estaría ocasionando la heterocedasticidad, y es con esta con la cual se trabaja el test de Goldfeld – Quant.

Entonces el test consiste en verificar que la varianza de $\mu$ no es función monótona de las $X_s$, en cuyo caso no habría heterocedasticidad.

**Pasos:**

1) Ordenar las observaciones de la serie en orden creciente, tomando como referencia los datos de la variable exógena ($X_i$) que supuestamente produce el problema de heterocedasticidad.

2) Se toma dos grupos en los extremos estimándose para cada uno su respectiva varianza.

| Grupo I | Grupo II |
|---|---|
| $\hat{\sigma}^2_{\mu_I} = \dfrac{\mu'_I \mu_I}{n_1 - k}$ | $\hat{\sigma}^2_{\mu_{II}} = \dfrac{\mu'_{II} \mu_{II}}{n_2 - k}$ |

$$\frac{\sigma^2_{\mu_{II}}}{\sigma^2_{\mu_I}} = F_c \sim F_{(n_1 - k, n_2 - k; \alpha)}$$

3) Omitir c observaciones centrales, donde c se ha especificado a priori, y dividir las observaciones restantes (n-c) en dos grupos, cada uno de (n-2)/2 observaciones.

| | Grupo I | Grupo II |
|---|---|---|
| Número de observaciones | $n_1 = \frac{(n-c)}{2}$ | $n_2 = \frac{(n-c)}{2}$ |

4) Realizar regresiones separadas por MCO para cada grupo y se obtendrá:

$$\hat{\beta}_I = \left(X_I^{'}X_I\right)^{-1}X_I^{'}Y_I$$
$$\hat{\beta}_{II} = \left(X_{II}^{'}X_{II}\right)^{-1}X_{II}^{'}Y_{II}$$

La primera regresión para los datos asociados con variables bajos de X y la segunda asociada con valores altos de X.

Cada regresión implicará (n-c)/2 piezas de datos y ((n-c)/2)-k grados de libertad, c debe ser lo bastante pequeño para asegurar que se dispone de suficientes grados de libertad para permitir la estimación apropiada de cada una de las regresiones por separadas.

5) Obtener las respectivas sumas residuales de cuadrados ($\mu'_I\mu_I$ y $\mu'_{II}\mu_{II}$) para cada uno de los grupos:

| Grupo | Suma Residual de Cuadrados (SCR) |
|---|---|
| Grupo I | $\mu'_I\,\mu_I = SCR_I = Y'_I\,Y_I - \hat{\beta}_I X'_I\,Y_I$ |
| Grupo II | $\mu'_{II}\,\mu_{II} = SCR_{II} = Y'_{II}\,Y_{II} - \hat{\beta}_{II} X'_{II}\,Y_{II}$ |

Donde los grados de libertad para cada una de ellas, es:

| | Grupo I | Grupo II |
|---|---|---|
| Grados de libertad (g.l.) | $n_1 - k$ | $n_2 - k$ |

Además se sabe que k es el número de parámetros que deben ser estimados, incluyendo el intercepto. Por tanto:

$$\mu'_I\,\mu_I \quad \sim \quad \chi^2 \quad \text{con} \quad n_1 - k \quad \text{g. l.}$$
$$\mu'_{II}\,\mu_{II} \quad \sim \quad \chi^2 \quad \text{con} \quad n_2 - k \quad \text{g. l.}$$

6) A continuación se construye el estadístico $F_c$, del cociente de las dos distribuciones Chi – cuadrado, dividido por sus grados de libertad. Este será:

$$F_c = \frac{\mu'_{II}\,\mu_{II}\,/\,\text{g. l.}}{\mu'_I\,\mu_I\,/\,\text{g. l.}}$$

el cual, definido el nivel de significación, se contrasta con el Estadístico (F)

7) Dócima de Hipótesis
   $H_0$: No existe heterocedasticidad (Homocedasticidad)
   $H_1$: Existe Heterocedasticidad.
   Si en una aplicación el $F_c$ es superior al F al nivel de significancia seleccionado, se puede rechazar la hipótesis de homocedasticidad (hipótesis nula), es decir, se puede afirmar que la presencia de heterocedasticidad es muy probable.

   Simplificando tenemos, si:

$$F_c > F \quad \Rightarrow \quad \text{se rechaza } H_0 \text{(Existe Heterocedasticidad)}$$
$$F_c < F \quad \Rightarrow \quad \text{se acepta } H_0 \text{(No existe Heterocedasticidad)}$$

**Ejercicio ilustrativo 1:**

Una empresa de transporte interprovincial PEGASO desea estimar las ventas anuales de pasajes (Y), en función de los gastos realizados para la mejora en la calidad del servicio ($X_2$) y publicidad ($X_3$).

Para ello se cuenta con la siguiente información (miles de nuevos soles):

| **Y** | **$X_2$** | **$X_3$** |
|---|---|---|
| 119 | 10 | 1 |
| 133.2 | 11 | 5 |
| 148.1 | 12 | 7 |
| 174.7 | 15 | 6 |
| 204 | 18 | 4 |
| 230 | 20 | 9 |
| 288.4 | 25 | 16 |
| 319.9 | 30 | 3 |
| 284.3 | 26 | 2 |
| 182.1 | 16 | 9 |
| 392.2 | 32 | 10 |
| 388.4 | 36 | 12 |
| 371.6 | 34 | 8 |
| 418.3 | 40 | 3 |
| 453 | 42 | 9 |
| 501.4 | 48 | 6 |
| 550.4 | 52 | 5 |
| 615 | 60 | 1 |
| 612 | 58 | 6 |
| 636 | 62 | 2 |

Solución

Se tiene que n = 20 datos, además c = 6.

Ordenando los datos (en forma ascendente) en función a la variable $X_2$ (la cual supuestamente produce Heterocedasticidad, se tiene los datos para la primera y segunda regresión:

| Y | $X_2$ | $X_3$ | |
|---|---|---|---|
| 119 | 10 | 1 | Observaciones para la primera regresión: (n-c)/2 =7 |
| 133.2 | 11 | 5 | |
| 148.1 | 12 | 7 | |
| 174.7 | 15 | 6 | |
| 182.1 | 16 | 9 | |
| 204 | 18 | 4 | |
| 230 | 20 | 9 | |
| 288.4 | 25 | 16 | Observaciones omitidas = 6 |
| 284.3 | 26 | 2 | |
| 319.9 | 30 | 3 | |
| 392.2 | 32 | 10 | |
| 371.6 | 34 | 8 | |
| 388.4 | 36 | 12 | |
| 418.3 | 40 | 3 | Observaciones para la segunda regresión: (n-c)/2 =7 |
| 453 | 42 | 9 | |
| 501.4 | 48 | 6 | |
| 550.4 | 52 | 5 | |
| 612 | 58 | 6 | |
| 615 | 60 | 1 | |
| 636 | 62 | 2 | |

De la primera regresión, obtenemos:

Dependent Variable: $Y_I$
Method: Least Squares
Date: 07/02/02 Time: 09:50
Sample: 1 7
Included observations: 7

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 17.02864 | 5.260470 | 3.237095 | 0.0318 |
| $X_{2I}$ | 10.22867 | 0.432883 | 23.62918 | 0.0000 |
| $X_{3I}$ | 0.696960 | 0.566639 | 1.229990 | 0.2861 |

| | | | |
|---|---|---|---|
| R-squared | 0.995609 | Mean dependent var | 170.1571 |
| Adjusted R-squared | 0.993414 | S.D. dependent var | 39.49510 |
| S.E. of regression | 3.205136 | Akaike info criterion | 5.464913 |
| Sum squared resid | 41.09158 | Schwarz criterion | 5.441732 |
| Log likelihood | -16.12720 | F-statistic | 453.5277 |
| Durbin-Watson stat | 1.857907 | Prob(F-statistic) | 0.000019 |

$$\mu'_I \mu_I = \sum \mu_I^2 = 41.09158$$

De la segunda regresión, obtenemos:

| Dependent Variable: $Y_{II}$ | | | | |
|---|---|---|---|---|
| Method: Least Squares | | | | |
| Date: 07/02/02 Time: 10:16 | | | | |
| Sample: 14 20 | | | | |
| Included observations: 7 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0.517561 | 14.96996 | -0.034573 | 0.9741 |
| $X_{2II}$ | 10.21966 | 0.245228 | 41.67402 | 0.0000 |
| $X_{3II}$ | 2.818984 | 0.777417 | 3.626091 | 0.0222 |
| R-squared | 0.998195 | Mean dependent var | | 540.8714 |
| Adjusted R-squared | 0.997292 | S.D. dependent var | | 85.63264 |
| S.E. of regresión | 4.456230 | Akaike info criterion | | 6.124010 |
| Sum squared resid | 79.43195 | Schwarz criterion | | 6.100829 |
| Log likelihood | -18.43404 | F-statistic | | 1105.808 |
| Durbin-Watson stat | 2.132042 | Prob(F-statistic) | | 0.000003 |

$$\mu'_{II}\,\mu_{II} = \sum \mu_{II}^2 = 79.43195$$

Seguidamente se realiza la dócima de hipótesis:

Ho : Las varianzas en el modelo considerado son iguales (homocedasticidad)
$H_1$ : Las varianzas en el modelo considerado no son iguales

$$F_{cal} = \frac{\mu'_{II}\,\mu_{II} / g.\,l.}{\mu'_{I}\,\mu_{I} / g.\,l.} = \frac{\mu'_{II}\,\mu_{II}}{\mu'_{I}\,\mu_{I}} = \frac{79.43195}{41.09158} = 1.93304687$$

Se observa que el F calculado ($F_c$=1.933) es menor al F crítico (F = 6.39), entonces no se rechaza la hipótesis nula, es decir, no existe heterocedasticidad en el modelo al 95% de confianza.

**b. Test de White**

El test de White es un contraste general que no requiere la elección de una variable que explique la volatilidad de los residuos. En particular, esta prueba supone que dicha varianza es una función lineal de los regresores originales del modelo, sus cuadrados y productos cruzados.

$$\mu^2 = f\left(\beta, X_2,\ldots,X_n, X_2^2,\ldots,X_n^2, X_2X_3,\ldots,X_{n-1}X_n\right)$$

La aplicación del test involucra la estimación de una regresión auxiliar del cuadrado de los residuos sobre las variables originales del modelo y las transformaciones mencionadas líneas arriba.

De lo anterior se deduce que conforme mayor sea el grado de ajuste de esta regresión auxiliar, mayor será la posibilidad de rechazar la hipótesis nula de la homocedasticidad. En otras palabras, en la medida en que alguno de los regresores originales o su transformación expliquen adecuadamente a la varianza del error, no podemos concluir que ésta sea constante.

Formalmente, la prueba de hipótesis se realiza sobre la base del estadístico $nR^2 \sim \chi^2_{p-1}$ (donde n es el número de observaciones de la muestra, $R^2$ el coeficiente de bondad de ajuste de la regresión auxiliar y p el número de parámetros a estimar en dicha regresión).

En este test no es necesario especificar la forma que adopta la heterocedasticidad.

## 3. CORRECCIÓN DE HETEROCEDASTICIDAD

Después de haber detectado que el modelo tiene una variación en la dispersión por cada término de perturbación, es decir:

$$E(\mu'\mu) = \sigma^2_\mu \Omega$$

Donde:

$$\Omega = \begin{bmatrix} a_1 \ldots\ldots 0 \\ 0 \ldots\ldots a_n \end{bmatrix}$$

A continuación se debe encontrar una matriz $P_{n,n}$, tal que $P^{-1}\Omega(P^{-1})' = I$

Por ejemplo en el caso que un valor de la diagonal de la matriz $\Omega$ fuera igual a 4 ($a_i$=4), entonces $P_i^{-1} = \frac{1}{\sqrt{4}}$, porque $P^{-1}\Omega(P^{-1})' = I$

En general si cada elemento de $\Omega$ tuviera un valor $a_i$, entonces $P_i^{-1} = 1/\sqrt{a_i}$ donde $a_i$ puede ser $X_i, X_i^2, \sqrt{X_i}, \beta_1 + \beta_2 X_i$, etc.

Una vez hallado P, se procede a corregir los valores originales a partir de una transformación del modelo original:

$$Y^* = P^{-1}Y$$
$$X^* = P^{-1}X$$

Con lo cual para pasar al modelo transformado se tiene que dividir las observaciones de todas las variables por $\sqrt{a_i}$, es decir, por la desviación típica de las perturbaciones

$$\frac{Y_i}{\sqrt{a_i}} = \frac{\beta_1}{\sqrt{a_i}} + \beta_2 \frac{X_{2i}}{\sqrt{a_i}} + \ldots\ldots\ldots\ldots + \frac{\beta_k X_{ki}}{\sqrt{a_i}} + \frac{\mu_i}{\sqrt{a_i}}$$

Para obtener los nuevos estimados bastará con reemplazar los valores transformados, obteniéndose:

$$\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

$$\mathrm{Var}\left(\hat{\beta}_{MCG}\right)=\sigma_{\mu}^{2}\left(X'\Omega^{-1}X\right)^{-1} \qquad \sigma_{MCG}^{2}=\frac{\hat{\mu}_{MCG}^{'}\Omega^{-1}\hat{\mu}_{MCG}}{n-k}$$

Por ejemplo, se tiene el siguiente modelo, el cual presenta el problema de heterocedasticidad:

$$Y_i=\beta_1+\beta_2X_2+\beta_2X_3+.......+\beta_kX_k+\mu_i$$
$$E\left(\mu_i^2\right)=\sigma_\mu^2 a_i$$

Entonces:

$$\frac{Y_1}{\sqrt{a_1}}=\frac{\beta_1}{\sqrt{a_1}}+\beta_2\frac{X_{21}}{\sqrt{a_1}}+...............+\frac{\beta_kX_{k1}}{\sqrt{a_1}}+\frac{\mu_1}{\sqrt{a_1}}$$

$$\frac{Y_n}{\sqrt{a_n}}=\frac{\beta_1}{\sqrt{a_n}}+\beta_2\frac{X_{2n}}{\sqrt{a_n}}+..............+\beta_{kn}\frac{X_{kn}}{\sqrt{a_n}}+\frac{\mu_n}{\sqrt{a_n}}$$

En el E-Views se genera una nueva variable, dividiendo entre $\sqrt{a_i}$ .

Se tiene los datos de un modelo con 2 variables explicativas ($X_2$ y $X_3$):
También: $Y_i=\beta_1+\beta_2X_{2t}+\beta_3X_{3t}+\mu_i$ con $E(\mu_i^2)=\sigma_\mu^2 a_i$ y $a_i=10$

| **Y** | **$X_2$** | **$X_3$** |
|---|---|---|
| 119 | 10 | 1 |
| 113.2 | 11 | 5 |
| . | . | . |
| . | . | . |
| . | . | . |
| 636 | 62 | 2 |

| $Y_T$ | $X_{2T}$ | $X_{3T}$ | $P_T$ |
|---|---|---|---|
| $119/\sqrt{10}$ | $10/\sqrt{10}$ | $1/\sqrt{10}$ | $1/\sqrt{10}$ |
| $113.2/\sqrt{11}$ | $11/\sqrt{11}$ | $5/\sqrt{11}$ | $1/\sqrt{11}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $636/\sqrt{62}$ | $62/\sqrt{62}$ | $2/\sqrt{62}$ | $1/\sqrt{62}$ |

Ahora, con $Y_t=\beta_T+\beta_2X_{2T}+\beta_3X_{3T}$ ; $E\left(\mu^2\right)=\sigma_\mu^2X_2^2$ , se tendría lo siguiente:

| $Y_T$ | $X_{2T}$ | $X_{3T}$ | $P_T$ |
|---|---|---|---|
| 119/10 | 10/10 | 1/10 | 1/10 |
| 113.2/11 | 11/11 | 5/11 | 1/11 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 636/62 | 62/62 | 2/62 | 1/62 |

**Ejercicio ilustrativo 2 -Heterocedasticidad**

Dado el siguiente modelo:

$$INV = \gamma 1 + \gamma 2\ PBI + \gamma 3\ r + \gamma 4\ IPINV + \gamma 5\ PR + \mu$$

Donde:

INV = Inversión
PBI = Producto Bruto Interno
r = Tasa de interés real
IPINV = Índice de precios de las inversiones
PR = Precios relativos del PBI con respecto a las inversiones

| | | | | Tasa interés | | |
|---|---|---|---|---|---|---|
| **AÑO** | **IMPORT** | **FBK** | **PBI** | **real** | **IPINV** | **IPPBI** |
| | | **INV** | **PBI** | **r** | **IPINV** | **PR** |
| 1970 | 9,491 | 10,306 | 62,022 | 3.500 | 0.00000035 | 0.00000036 |
| 1971 | 9,920 | 11,630 | 64,627 | 1.600 | 0.00000039 | 0.00000038 |
| 1972 | 9,892 | 12,069 | 66,501 | 4.900 | 0.00000042 | 0.00000041 |
| 1973 | 11,110 | 16,176 | 70,092 | -3.900 | 0.00000043 | 0.00000047 |
| 1974 | 14,127 | 20,066 | 76,611 | -8.300 | 0.00000049 | 0.00000054 |
| 1975 | 13,724 | 20,572 | 79,215 | -11.900 | 0.00000064 | 0.00000067 |
| 1976 | 11,995 | 18,451 | 80,800 | -2.300 | 0.00000085 | 0.00000086 |
| 1977 | 12,037 | 16,413 | 81,123 | -13.800 | 0.00000119 | 0.00000117 |
| 1978 | 9,103 | 14,954 | 81,366 | -30.000 | 0.00000204 | 0.00000186 |
| 1979 | 10,860 | 17,103 | 86,086 | -17.000 | 0.00000357 | 0.00000323 |
| 1980 | 14,144 | 20,846 | 90,562 | -14.600 | 0.00000567 | 0.00000528 |
| 1981 | 16,395 | 24,205 | 95,181 | -6.100 | 0.00000961 | 0.00000900 |
| 1982 | 16,758 | 23,680 | 94,610 | -6.500 | 0.00001735 | 0.00001510 |
| 1983 | 11,791 | 16,545 | 83,446 | -27.600 | 0.00003857 | 0.00003140 |
| 1984 | 9,647 | 15,530 | 87,785 | -21.200 | 0.00009039 | 0.00006680 |
| 1985 | 8,812 | 13,777 | 90,243 | -34.700 | 0.00027008 | 0.00017900 |
| 1986 | 10,607 | 16,303 | 99,267 | -39.900 | 0.00039877 | 0.00030900 |
| 1987 | 12,182 | 19,338 | 107,208 | -41.500 | 0.00062688 | 0.00056300 |
| 1988 | 11,091 | 16,646 | 97,881 | -88.400 | 0.00529875 | 0.00410700 |
| 1989 | 8,287 | 13,349 | 86,429 | -20.370 | 0.14000000 | 0.10800000 |
| 1990 | 9,270 | 13,625 | 81,983 | -18.000 | 7.61000000 | 6.75000000 |
| 1991 | 11,130 | 13,929 | 83,760 | -2.000 | 31.70000000 | 31.86000000 |
| 1992 | 12,113 | 14,060 | 83,401 | -0.280 | 52.70000000 | 56.95000000 |
| 1993 | 12,572 | 15,627 | 87,375 | -0.262 | 81.10000000 | 82.16000000 |
| 1994 | 15,922 | 20,901 | 98,577 | -0.195 | 100.00000000 | 100.00000000 |
| 1995 | 20,232 | 25,468 | 107,039 | -0.505 | 113.80000000 | 112.80000000 |
| 1996 | 20,259 | 24,738 | 109,709 | -0.530 | 124.70000000 | 124.70000000 |
| 1997 | 22,724 | 28,518 | 117,110 | -0.403 | 131.20000000 | 134.20000000 |
| 1998 | 23,251 | 28,110 | 116,485 | 0.442 | 138.90000000 | 143.30000000 |
| 1999 | 19,724 | 24,984 | 117,590 | 1.439 | 153.50000000 | 148.58000000 |
| 2000 | 20,428 | 23,731 | 121,267 | 1.312 | 159.40000000 | 154.00000000 |

Regresionando el modelo:

**INV = γ1 + γ2 PBI + γ3 r + γ4 IPINV + γ5 PR + μ**

Dependent Variable: INV
Method: Least Squares
Date: 09/10/02 Time: 15:51
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -10954.63 | 4774.259 | -2.294519 | 0.0301 |
| PBI | 0.353642 | 0.062086 | 5.695976 | 0.0000 |
| R | 116.7644 | 36.94933 | 3.160122 | 0.0040 |
| IPINV | -270.8711 | 263.4286 | -1.028253 | 0.3133 |
| PR | 239.2287 | 263.2265 | 0.908832 | 0.3718 |

| | | | |
|---|---|---|---|
| R-squared | 0.760594 | Mean dependent var | 18440.32 |
| Adjusted R-squared | 0.723762 | S.D. dependent var | 4999.747 |
| S.E. of regression | 2627.784 | Akaike info criterion | 18.73236 |
| Sum squared resid | 1.80E+08 | Schwarz criterion | 18.96365 |
| Log likelihood | -285.3516 | F-statistic | 20.65050 |
| Durbin-Watson stat | 0.950988 | Prob(F-statistic) | 0.000000 |

$\hat{\gamma}_1 = -10954.63$, indica el nivel promedio de las inversiones, cuando el resto de parámetros son cero.

$\hat{\gamma}_2 = 0.353642$, este resultado nos muestra que si el PBI varía en 1%, la inversión varía en 35.3642%

$\hat{\gamma}_3 = 116.7644$, este resultado nos muestra que si la tasa de interés real varía en 1%, la inversión aumentará en 16.7644%

$\hat{\gamma}_4 = -270.871$, este resultado nos muestra que si el índice de precio de las inversiones varía en una unidad, la inversión disminuirá en –270.8711.

$\hat{\gamma}_5 = 239.2287$, este resultado nos muestra que si los precios relativos del PBI respecto de las inversiones varían en una unidad, la inversión aumentará en 239.2287.

1. **TEST DE WHITE** (considerando el efecto producido entre regresores)

Ho : Las varianzas en el modelo considerado son iguales (homocedasticidad)
$H_1$ : Las varianzas en el modelo considerado no son iguales

En este caso se observa que la probabilidad asociada al test de White es de 0.77, la cual es mayor al nivel de significancia escogido (0.05), por lo que se concluye que en un modelo donde no se ha considerado los efectos entre las variables no se presenta homocedasticidad a un nivel de confianza del 95%.

| White Heteroskedasticity Test: | | | |
|---|---|---|---|
| F-statistic | 0.596246 | Probability | 0.770678 |
| Obs*R-squared | 5.523690 | Probability | 0.700414 |

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/10/02 Time: 16:12
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -15199469 | 69035028 | -0.220170 | 0.8278 |
| PBI | 477.3539 | 1645.729 | 0.290056 | 0.7745 |
| PBI^2 | -0.002694 | 0.009590 | -0.280897 | 0.7814 |
| R | -67263.34 | 257532.2 | -0.261184 | 0.7964 |
| R^2 | -165.3252 | 2723.256 | -0.060709 | 0.9521 |
| IPINV | -3528052. | 2789103. | -1.264941 | 0.2191 |
| IPINV^2 | 12724.18 | 9988.453 | 1.273889 | 0.2160 |
| PR | 3523473. | 2737062. | 1.287319 | 0.2114 |
| PR^2 | -12835.23 | 9849.387 | -1.303150 | 0.2060 |

| | | | |
|---|---|---|---|
| R-squared | 0.178184 | Mean dependent var | 5791501. |
| Adjusted R-squared | -0.120659 | S.D. dependent var | 6579776. |
| S.E. of regression | 6965428. | Akaike info criterion | 34.58852 |
| Sum squared resid | 1.07E+15 | Schwarz criterion | 35.00484 |
| Log likelihood | -527.1220 | F-statistic | 0.596246 |
| Durbin-Watson stat | 1.099102 | Prob(F-statistic) | 0.770678 |

2. **TEST DE WHITE** (Términos cruzados :considerando el efecto producido entre regresores)

Ho: Modelo no presenta heterocedasticidad
$H_1$: Modelo presenta heterocedasticidad

En este caso se observa que la probabilidad asociada al test de White es de 0.89, la cual es mayor al nivel de significancia escogido (0.05), por lo que se concluye que un modelo donde se considera los efectos entre las variables no se presenta heterocedasticidad a un nivel de confianza del 95%.

| White Heteroskedasticity Test: | | | |
|---|---|---|---|
| F-statistic | 0.501012 | Probability | 0.894325 |
| Obs*R-squared | 8.587010 | Probability | 0.803371 |

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/10/02 Time: 16:17
Sample: 1970 2000
Included observations: 31

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -1.75E+08 | 1.73E+08 | -1.011368 | 0.3260 |
| PBI | 4660.078 | 4460.731 | 1.044689 | 0.3108 |
| PBI^2 | -0.029535 | 0.028199 | -1.047358 | 0.3096 |
| PBI*R | -25.58410 | 28.04803 | -0.912153 | 0.3744 |
| PBI*IPINV | 541.5411 | 906.8817 | 0.597146 | 0.5583 |
| PBI*PR | -511.7080 | 894.9488 | -0.571774 | 0.5750 |
| R | 2091313. | 2411013. | 0.867400 | 0.3978 |
| R^2 | -3064.552 | 4271.637 | -0.717419 | 0.4829 |
| R*IPINV | 340891.2 | 966929.2 | 0.352550 | 0.7288 |
| R*PR | -289683.8 | 977041.4 | -0.296491 | 0.7704 |
| IPINV | -37198665 | 57450567 | -0.647490 | 0.5260 |
| IPINV^2 | -176651.7 | 342416.2 | -0.515898 | 0.6126 |
| IPINV*PR | 167992.1 | 337804.6 | 0.497305 | 0.6253 |
| PR | 35194137 | 56709187 | 0.620607 | 0.5431 |

| | | | |
|---|---|---|---|
| R-squared | 0.277000 | Mean dependent var | 5791501. |
| Adjusted R-squared | -0.275882 | S.D. dependent var | 6579776. |
| S.E. of regression | 7432182. | Akaike info criterion | 34.78299 |
| Sum squared resid | 9.39E+14 | Schwarz criterion | 35.43060 |
| Log likelihood | -525.1363 | F-statistic | 0.501012 |
| Durbin-Watson stat | 1.187606 | Prob(F-statistic) | 0.894325 |

## 3. PRUEBA DE GOLDFELD-QUANDT

Se omitirán 11 observaciones centrales y 2 grupos de 10 observaciones cada uno (los datos están ordenados en orden creciente para las variables IPINV, PR).

Se obtiene el estadístico de los residuales:

Fc = 9706682 / 18308626 = 0.53016988 (F-Statistics)

Se realiza la hipótesis:

Ho : Las varianzas en el modelo considerado son iguales (homocedasticidad)
$H_1$ : Las varianzas en el modelo considerado no son iguales

El F obtenido de tablas es 5.05, y como su valor es mayor al calculado, se acepta la Ho, y se concluye que el modelo no presenta problemas de heterocedasticidad.

Regresión sobre el primer grupo:

| | | | | |
|---|---|---|---|---|
| Dependent Variable: INV | | | | |
| Method: Least Squares | | | | |
| Date: 09/10/02 Time: 17:26 | | | | |
| Sample: 1970 1979 | | | | |
| Included observations: 10 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| PBI | 0.398393 | 0.279491 | 1.425422 | 0.2134 |
| R | 26.34102 | 106.4741 | 0.247394 | 0.8144 |
| IPINV | -3.07E+10 | 3.94E+10 | -0.779222 | 0.4711 |
| PR | 3.31E+10 | 4.58E+10 | 0.721350 | 0.5030 |
| C | -14895.39 | 16213.61 | -0.918697 | 0.4004 |
| R-squared | 0.837135 | Mean dependent var | | 15774.00 |
| Adjusted R-squared | 0.706842 | S.D. dependent var | | 3534.210 |
| S.E. of regression | 1913.563 | Akaike info criterion | | 18.25817 |
| Sum squared resid | 18308626 | Schwarz criterion | | 18.40947 |
| Log likelihood | -86.29087 | F-statistic | | 6.425054 |
| Durbin-Watson stat | 1.177442 | Prob(F-statistic) | | 0.033108 |

Regresión sobre el segundo grupo

| | | | | |
|---|---|---|---|---|
| Dependent Variable: INV | | | | |
| Method: Least Squares | | | | |
| Date: 09/10/02 Time: 17:27 | | | | |
| Sample: 1991 2000 | | | | |
| Included observations: 10 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| PBI | 0.216057 | 0.153920 | 1.403694 | 0.2194 |
| R | -2255.303 | 1027.193 | -2.195597 | 0.0795 |
| IPINV | -290.1799 | 168.5920 | -1.721196 | 0.1458 |
| PR | 388.3896 | 159.3789 | 2.436895 | 0.0589 |
| C | -11470.34 | 9260.306 | -1.238657 | 0.2704 |
| R-squared | 0.965487 | Mean dependent var | | 22006.60 |
| Adjusted R-squared | 0.937877 | S.D. dependent var | | 5590.149 |
| S.E. of regression | 1393.319 | Akaike info criterion | | 17.62362 |
| Sum squared resid | 9706682. | Schwarz criterion | | 17.77491 |
| Log likelihood | -83.11809 | F-statistic | | 34.96833 |
| Durbin-Watson stat | 2.216423 | Prob(F-statistic) | | 0.000755 |

# LABORATORIO

Con la siguiente información elaboraremos un modelo que explique la relación entre las variables:

$PBI_{construccion} = f(VENTA_{cemento}, VENTA_{barra\ de\ acero})$

PBIconstrucion = PBIco
VENTAcemento = VTAcem
VENTAbarra de acero = VTAbarra

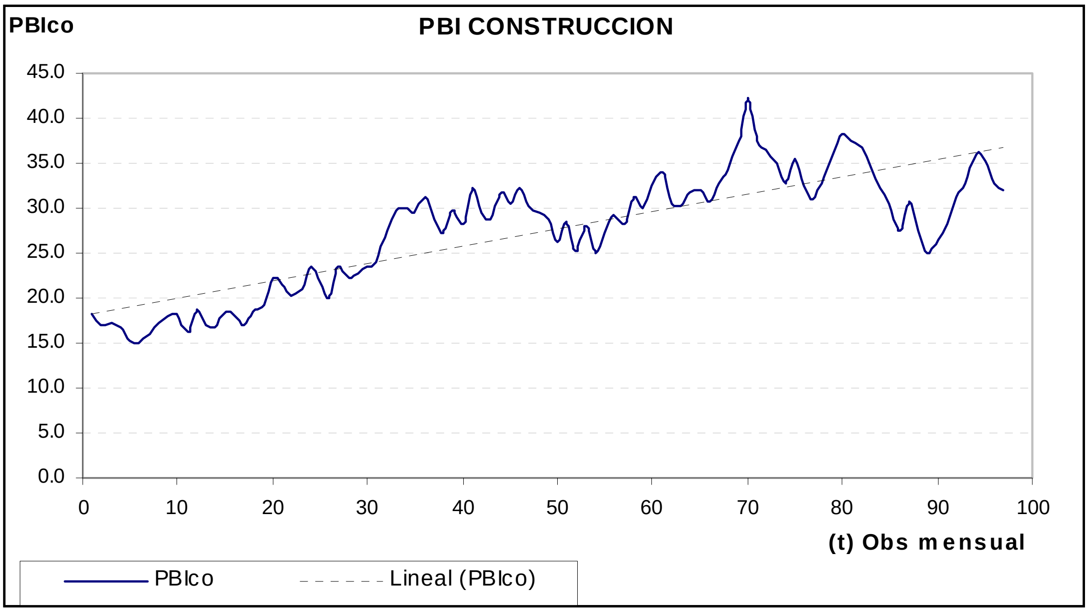
PBIco = f(VTAcem; VTAbarra)

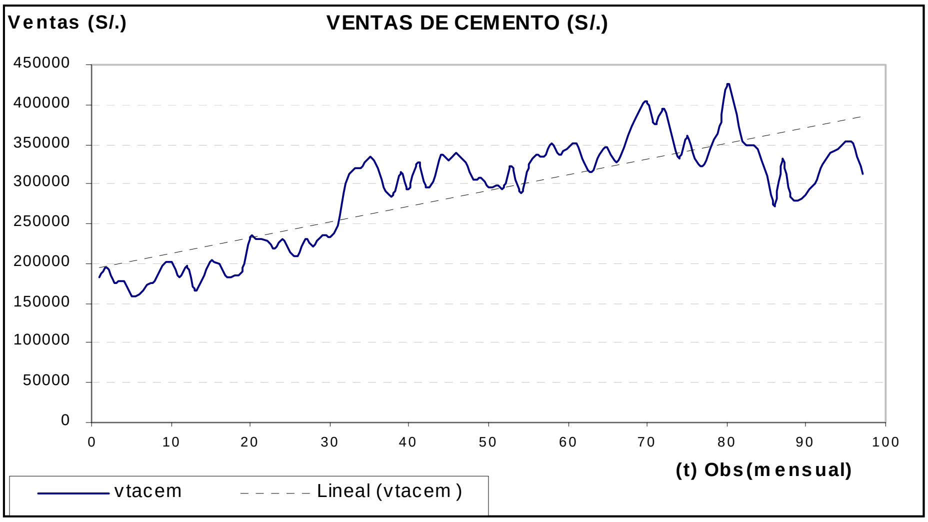
Utilizaremos la siguiente información estadística:

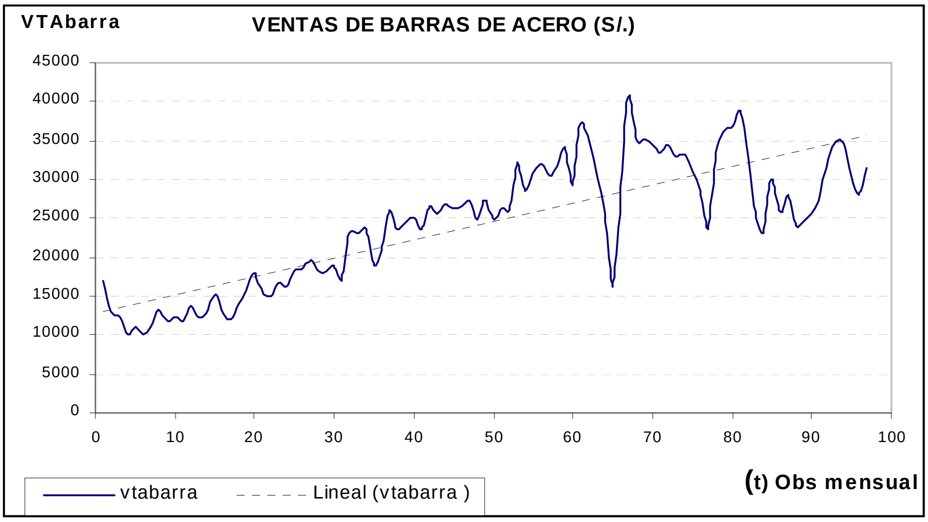
| *Obs* | Mes | vtacem | vtabarra | pbi construcción (soles) | Obs | mes | vtacem | vtabarra | pbi construcción (soles) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 199201 | 183213 | 16925 | 18.3 | 49 | 199601 | 307481 | 27176 | 28.8 |
| 2 | 199202 | 194484 | 12929 | 17.0 | 50 | 199602 | 296386 | 24772 | 26.3 |
| 3 | 199203 | 176178 | 12175 | 17.3 | 51 | 199603 | 298544 | 26345 | 28.5 |
| 4 | 199204 | 178918 | 10119 | 16.7 | 52 | 199604 | 295281 | 26058 | 25.3 |
| 5 | 199205 | 158561 | 11013 | 15.4 | 53 | 199605 | 321715 | 32333 | 28.1 |
| 6 | 199206 | 160630 | 10018 | 15.0 | 54 | 199606 | 289842 | 28467 | 24.9 |
| 7 | 199207 | 172372 | 11052 | 16.1 | 55 | 199607 | 325212 | 30775 | 27.3 |
| 8 | 199208 | 178754 | 13178 | 17.3 | 56 | 199608 | 336051 | 31945 | 29.3 |
| 9 | 199209 | 197417 | 11799 | 17.9 | 57 | 199609 | 334620 | 30488 | 28.4 |
| 10 | 199210 | 201507 | 12311 | 18.2 | 58 | 199610 | 351596 | 31664 | 31.3 |
| 11 | 199211 | 183455 | 11871 | 16.2 | 59 | 199611 | 335806 | 34078 | 29.9 |
| 12 | 199212 | 196976 | 13841 | 18.6 | 60 | 199612 | 345818 | 29146 | 32.6 |
| 13 | 199301 | 165602 | 12341 | 17.1 | 61 | 199701 | 351193 | 37124 | 34.0 |
| 14 | 199302 | 180878 | 12730 | 16.8 | 62 | 199702 | 328424 | 34671 | 30.5 |
| 15 | 199303 | 201011 | 15306 | 18.5 | 63 | 199703 | 314409 | 30290 | 30.2 |
| 16 | 199304 | 198593 | 12756 | 17.9 | 64 | 199704 | 336822 | 25696 | 31.8 |
| 17 | 199305 | 183761 | 12059 | 17.0 | 65 | 199705 | 346085 | 16278 | 32.1 |
| 18 | 199306 | 186465 | 14113 | 18.5 | 66 | 199706 | 326683 | 27268 | 30.7 |
| 19 | 199307 | 189744 | 15644 | 19.2 | 67 | 199707 | 346447 | 40558 | 32.8 |
| 20 | 199308 | 232517 | 17988 | 22.2 | 68 | 199708 | 373488 | 34893 | 34.1 |
| 21 | 199309 | 231768 | 15301 | 21.4 | 69 | 199709 | 391529 | 35262 | 37.5 |
| 22 | 199310 | 227776 | 15075 | 20.3 | 70 | 199710 | 403797 | 34510 | 42.3 |
| 23 | 199311 | 219757 | 16718 | 21.0 | 71 | 199711 | 374946 | 33476 | 37.5 |
| 24 | 199312 | 230442 | 16226 | 23.5 | 72 | 199712 | 395240 | 34404 | 36.5 |
| 25 | 199401 | 213194 | 18287 | 21.9 | 73 | 199801 | 361198 | 32896 | 35.1 |
| 26 | 199402 | 208455 | 18404 | 19.9 | 74 | 199802 | 332984 | 33305 | 32.8 |
| 27 | 199403 | 231596 | 19733 | 23.5 | 75 | 199803 | 359779 | 31045 | 35.4 |
| 28 | 199404 | 220376 | 18092 | 22.4 | 76 | 199804 | 331183 | 28486 | 32.4 |
| 29 | 199405 | 235017 | 18136 | 22.7 | 77 | 199805 | 321998 | 23684 | 31.1 |
| 30 | 199406 | 233380 | 19054 | 23.5 | 78 | 199806 | 344054 | 33549 | 33.5 |
| 31 | 199407 | 246724 | 17217 | 24.0 | 79 | 199807 | 372117 | 36077 | 36.1 |
| 32 | 199408 | 299957 | 23004 | 27.6 | 80 | 199808 | 425734 | 36791 | 38.2 |
| 33 | 199409 | 317009 | 23204 | 29.8 | 81 | 199809 | 397771 | 38757 | 37.5 |

| *Obs* | Mes | vtacem | vtabarra | pbi construcción (soles) | Obs | mes | vtacem | vtabarra | pbi construcción (soles) |
|---|---|---|---|---|---|---|---|---|---|
| 34 | 199410 | 320803 | 23613 | 30.1 | 82 | 199810 | 353001 | 32837 | 36.6 |
| 35 | 199411 | 333757 | 18895 | 29.5 | 83 | 199811 | 348144 | 25179 | 34.4 |
| 36 | 199412 | 318943 | 21024 | 31.2 | 84 | 199812 | 343260 | 23133 | 32.3 |
| 37 | 199501 | 292259 | 26109 | 28.8 | 85 | 199901 | 311203 | 29896 | 30.5 |
| 38 | 199502 | 285689 | 23562 | 27.2 | 86 | 199902 | 273058 | 25760 | 27.6 |
| 39 | 199503 | 315214 | 24692 | 29.8 | 87 | 199903 | 331499 | 27957 | 30.8 |
| 40 | 199504 | 293054 | 25080 | 28.2 | 88 | 199904 | 284956 | 24050 | 27.5 |
| 41 | 199505 | 326852 | 23499 | 32.3 | 89 | 199905 | 278223 | 24667 | 25.1 |
| 42 | 199506 | 296535 | 26479 | 29.4 | 90 | 199906 | 285707 | 25501 | 26.4 |
| 43 | 199507 | 302599 | 25651 | 28.7 | 91 | 199907 | 300335 | 27220 | 28.2 |
| 44 | 199508 | 335925 | 26872 | 31.8 | 92 | 199908 | 324273 | 32619 | 31.4 |
| 45 | 199509 | 328595 | 26406 | 30.6 | 93 | 199909 | 339146 | 34993 | 32.8 |
| 46 | 199510 | 338925 | 26564 | 32.1 | 94 | 199910 | 343501 | 34588 | 36.1 |
| 47 | 199511 | 327876 | 27194 | 30.2 | 95 | 199911 | 353612 | 30587 | 35.3 |
| 48 | 199512 | 306074 | 24807 | 29.5 | 96 | 199912 | 350593 | 27946 | 32.8 |
| | | | | | 97 | 200001 | 312532 | 31391 | 32.0 |

1. **En primer lugar se analiza la información estadística, mediante su evolución en el tiempo:**






Del análisis de la serie de las variables en el tiempo podemos comentar lo siguiente:

- El Producto Bruto interno del sector construcción y la Venta de Cemento muestran comportamientos similares en su evolución mensual desde 1992 hasta enero del 2001, es decir a primera vista existe una relación directa e importante entre las dos variables.
- Coinciden los periodos de mayores niveles registrados en el PBI Construcción y la venta de Cemento. Las cuales fueron en enero de 1997 y enero de 1998.
- La tendencia (Lineal) observada en las tres variables son crecientes
- Se puede observar en la serie de ventas de barras, que a partir de la observación 60 (principios de 1997), el comportamiento de dicha variable es bastante irregular pero que no se reflejaría en forma gravitante el comportamiento del PBI construcción, puesto a primera vista mantiene una mayor relación con la variable Venta de Cemento.

## 2. Estimación del Modelo Lineal

$$PBIco = f(VTAcem; VTAbarra)$$
$$PBIco = C(1) + C(2)*VTAcem + C(3)*VTAbarra$$

| | | | | |
|---|---|---|---|---|
| Dependent Variable: PBICO | | | | |
| Method: Least Squares | | | | |
| Date: 09/18/02 Time: 23:23 | | | | |
| Sample: 1 97 | Included observations: 97 | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 0.359737 | 0.667219 | 0.539158 | 0.5911 |
| VTACEM | 9.02E-05 | 4.84E-06 | 18.61325 | 0.0000 |
| VTABARRA | 4.21E-05 | 4.10E-05 | 1.026586 | 0.3072 |
| R-squared | 0.957725 | Mean dependent var | | 27.47113 |
| Adjusted R-squared | 0.956826 | S.D. dependent var | | 6.567644 |
| S.E. of regression | 1.364651 | Akaike info criterion | | 3.490113 |
| Sum squared resid | 175.0535 | Schwarz criterion | | 3.569744 |
| Log likelihood | -166.2705 | F-statistic | | 1064.777 |
| Durbin-Watson stat | 0.988372 | Prob(F-statistic) | | 0.000000 |

### 3. Especificación del modelo: Test de Reset Ramsey

| | | | | |
|---|---|---|---|---|
| Ramsey RESET Test: | | | | |
| F-statistic | 0.245591 | Probability | | 0.782753 |
| Log likelihood ratio | 0.516500 | Probability | | 0.772402 |
| Test Equation: | | | | |
| Dependent Variable: PBICO | | | | |
| Method: Least Squares | | | | |
| Date: 09/19/02 Time: 22:55 | | | | |
| Sample: 1 97 | | | | |
| Included observations: 97 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 0.740340 | 9.829252 | 0.075320 | 0.9401 |
| VTACEM | 9.16E-05 | 0.000108 | 0.846452 | 0.3995 |
| VTABARRA | 4.32E-05 | 6.65E-05 | 0.649589 | 0.5176 |
| FITTED^2 | -0.003105 | 0.044937 | -0.069095 | 0.9451 |
| FITTED^3 | 6.85E-05 | 0.000543 | 0.126078 | 0.8999 |
| R-squared | 0.957950 | Mean dependent var | | 27.47113 |
| Adjusted R-squared | 0.956122 | S.D. dependent var | | 6.567644 |
| S.E. of regression | 1.375736 | Akaike info criterion | | 3.526026 |
| Sum squared resid | 174.1239 | Schwarz criterion | | 3.658743 |
| Log likelihood | -166.0122 | F-statistic | | 523.9656 |
| Durbin-Watson stat | 1.016619 | Prob(F-statistic) | | 0.000000 |

La probabilidad asociada a este estadístico nos señala que a un nivel de significancia del 5% no se debe rechazar la $H_0$, es decir que el modelo esta correctamente especificado. Esta conclusión errónea que contradice el planteamiento inicial, nos conduce a especificar que hay problemas de multicolinealidad y lo primero que debe hacerse es plantear la solución del problema de multicolinealidad

## 4. Detectando la Multicolinealidad

Analizando la Matriz de Correlación entre las variables independientes:

| | VTACEM | VTABARRA |
|---|---|---|
| **VTACEM** | 1.000000 | 0.905848 |
| **VTABARRA** | 0.905848 | 1.000000 |

Se observa que las variables VTACEM están correlacionada con la variable VTABARRA

Probando los siguientes modelos:

Dependent Variable: PBICO
Method: Least Squares
Date: 09/19/02 Time: 00:23
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.081367 | 0.609807 | 0.133431 | 0.8941 |
| VTACEM | 9.47E-05 | 2.05E-06 | 46.12258 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.957251 | Mean dependent var | 27.47113 |
| Adjusted R-squared | 0.956801 | S.D. dependent var | 6.567644 |
| S.E. of regression | 1.365038 | Akaike info criterion | 3.480644 |
| Sum squared resid | 177.0161 | Schwarz criterion | 3.533731 |
| Log likelihood | -166.8112 | F-statistic | 2127.293 |
| Durbin-Watson stat | 1.048183 | Prob(F-statistic) | 0.000000 |

Dependent Variable: PBICO
Method: Least Squares
Date: 09/19/02 Time: 00:32
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 9.612641 | 0.958264 | 10.03131 | 0.0000 |
| VTABARRA | 0.000734 | 3.74E-05 | 19.61101 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.801915 | Mean dependent var | 27.47113 |
| Adjusted R-squared | 0.799830 | S.D. dependent var | 6.567644 |
| S.E. of regression | 2.938390 | Akaike info criterion | 5.014004 |
| Sum squared resid | 820.2430 | Schwarz criterion | 5.067091 |
| Log likelihood | -241.1792 | F-statistic | 384.5916 |
| Durbin-Watson stat | 0.940910 | Prob(F-statistic) | 0.000000 |

Se ratifica la apreciación inicial que la variable VTAcem es la que mejor se ajusta, pero existen problemas de autocorrelación que es necesario corregir; sin embargo ambas variables son significativas

Por lo tanto la ecuación estimada seria:

PBICO = C(1) + C(2)*VTACEM + C(3)*VTABARRA

PBICO = 0.3597366667 + 9.015568921E-05*VTACEM + 4.213246486E-05*VTABARRA

**Interpretación**:

El intercepto de la ecuación es 0.3597
Ante una variación de 1 nuevo sol en los PBIco, las ventas de cemento se incrementarán en 9.01556E-05 nuevos soles. y las ventas de barras de acero se incrementarían en 4.21324E-05 nuevos soles

**Significancia individual y conjunta del Modelo.**

Individualmente el modelo es significativo demostrado mediante la prueba T, cuyas probabilidades son menores que 0.05, tanto del coeficiente de ingresos como del intercepto.

En conjunto el modelo es significativo, demostrado por la prueba F al ser su probabilidad menor a 0.05. Asimismo, la bondad de ajuste, medida a través del $R^2$ y $R^2$ ajustado, demuestra un buen ajuste. (Alrededor del 95%).

## 5. Detección de la Heterocedasticidad.

### a. Contraste de White.

El Test de White es un contraste general que no requiere la elección de una variable que explique la volatilidad de los residuos, en particular esta prueba supone que dicha varianza es una función lineal de los reagresores originales del modelo, sus cuadrados y productos cruzados

$H_0$ : El Modelo es homoscedástico.
$H_1$ : El Modelo no es homoscedástico.

| White Heteroskedasticity Test: | | | | |
|---|---|---|---|---|
| F-statistic | 3.313768 | Probability | | 0.013903 |
| Obs*R-squared | 12.21549 | Probability | | 0.015819 |

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample: 1 97 Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 6.734537 | 5.976135 | 1.126905 | 0.2627 |
| VTACEM | -9.79E-05 | 7.00E-05 | -1.398090 | 0.1654 |
| VTACEM^2 | 1.58E-10 | 1.15E-10 | 1.369796 | 0.1741 |
| VTABARRA | 0.000623 | 0.000470 | 1.324816 | 0.1885 |
| VTABARRA^2 | -8.63E-09 | 8.56E-09 | -1.008019 | 0.3161 |

| | | | |
|---|---|---|---|
| R-squared | 0.125933 | Mean dependent var | 1.804675 |
| Adjusted R-squared | 0.087930 | S.D. dependent var | 3.080460 |
| S.E. of regression | 2.941912 | Akaike info criterion | 5.046167 |
| Sum squared resid | 796.2458 | Schwarz criterion | 5.178884 |
| Log likelihood | -239.7391 | F-statistic | 3.313768 |
| Durbin-Watson stat | 1.477165 | Prob(F-statistic) | 0.013903 |

Como Prob.= 0.010805 < 0.05. Entonces se rechaza la hipótesis nula. Es decir el Modelo es no homoscedástico, es decir existen problemas de heteroscedasticidad, con un nivel de confianza del 95%.

## b. Prueba de Goldfeld – Quandt

El Test consiste en verificar que la varianza de $\mu$ no es función monótona creciente o decreciente de las $x_s$, en cuyo caso no habría heteroscedasticidad.

Ordenar las observaciones de la serie en orden creciente, tomando como referencia los datos de la variable exógena ($x_i$) que supuestamente produce el problema de heteroscedasticidad.

Se toma dos grupos en los extremos estimándose para cada uno su respectiva varianza.

| *GRUPO I* | *GRUPO II* |
|---|---|
| $\hat{\sigma}^2_{\mu_I} = \frac{e'_I e_I}{n_1 - k}$ | $\hat{\sigma}^2_{\mu_{1II}} = \frac{e'_{II} e_{II}}{n_2 - k}$ |

$$\frac{\sigma^2_{\mu_{II}}}{\sigma^2_{\mu_I}} = F_c \sim F_{(n1-k,\ n2-k\ ;\ k)}$$

Si el Fc es mayor que el de tabla se rechaza la Ho de Homoscedasticidad

**En el Eviews:**

- En el Excel ordenar las observaciones de la serie en orden creciente, tomando como referencia los datos de la variable exógena.
- Omitir c = 33 observaciones centrales, y dividir las observaciones restantes (n-c) en dos grupos, cada uno de (n-33)/2 = 32 observaciones.
- En el Eviews hacer clic en Quick Empty Group (Edit series) y copiar las observaciones ordenadas en forma ascendente y ponerles nuevos nombres a las series (PBICOSORT e VTACEMSORT, VTABARRASORT).
- Regresionar el modelo ls pbicosort c vtacemsort vtabarrasort. En el Workfile se selecciona las variables pbicosord, vtacemsort, vtabarrasort. Hacer clic derecho en dicha selección y escoger OPEN/AS EQUATION, aparecerá la ecuación digitada en una caja de dialogo. En esta caja aparece la opción Sample cambiar la muestra: 1 32 y OK.
- Se obtiene los siguientes resultados:

Dependent Variable: PBICOSORT
Method: Least Squares
Sample: 1 32
Included observations: 32

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 2.280431 | 1.479442 | 1.541414 | 0.1341 |
| VTACEMSORT | 7.10E-05 | 1.25E-05 | 5.688875 | 0.0000 |
| VTABARRASORT | 0.000177 | 9.68E-05 | 1.831155 | 0.0774 |

| | | | |
|---|---|---|---|
| R-squared | 0.881644 | Mean dependent var | 19.38125 |
| Adjusted R-squared | 0.873482 | S.D. dependent var | 2.834401 |
| S.E. of regression | 1.008179 | Akaike info criterion | 2.943229 |
| Sum squared resid | 29.47634 | Schwarz criterion | 3.080642 |
| Log likelihood | -44.09166 | F-statistic | 108.0121 |
| Durbin-Watson stat | 1.240423 | Prob(F-statistic) | 0.000000 |

Regresionar el modelo ls pbicosort c vracemsort vtabarrasort. Con la muestra de 65 97.

Dependent Variable: PBICOSORT
Method: Least Squares
Sample: 65 97
Included observations: 33

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 4.874702 | 3.337960 | 1.460384 | 0.1546 |
| VTACEMSORT | 7.73E-05 | 1.11E-05 | 6.999134 | 0.0000 |
| VTABARRASORT | 6.24E-05 | 5.28E-05 | 1.182145 | 0.2464 |

| | | | |
|---|---|---|---|
| R-squared | 0.734387 | Mean dependent var | 34.11212 |
| Adjusted R-squared | 0.716680 | S.D. dependent var | 2.605494 |
| S.E. of regression | 1.386848 | Akaike info criterion | 3.578453 |
| Sum squared resid | 57.70045 | Schwarz criterion | 3.714499 |
| Log likelihood | -56.04447 | F-statistic | 41.47317 |
| Durbin-Watson stat | 2.071584 | Prob(F-statistic) | 0.000000 |

- Se obtiene las sumas residuales (SCR) para ambas regresiones y se obtiene el estadístico F.

$$F_{cal} = \frac{e'_{II}e_{II} / g.l.}{e'_{I}e_{I} / g.l.} = \frac{e'_{II}e_{II}}{e'_{I}e_{I}} = \frac{57.70045}{29.47634} = 1.955752$$

A continuación se realiza la hipótesis
Ho: No Existe heteroscedasticidad
$H_1$: Existe heteroscedasticidad

- El $F_{(29,5)}$ de tablas a un nivel del 5% de significancia es 3.33 , como el F de la tabla es mayor al F calculado, entonces se acepta la Ho. Se concluye que el modelo no presenta problemas de Heteroscedasticidad.

## 6. Detección de la Autocorrelación

Dependent Variable: PBICO
Method: Least Squares
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.359737 | 0.667219 | 0.539158 | 0.5911 |
| VTACEM | 9.02E-05 | 4.84E-06 | 18.61325 | 0.0000 |
| VTABARRA | 4.21E-05 | 4.10E-05 | 1.026586 | 0.3072 |

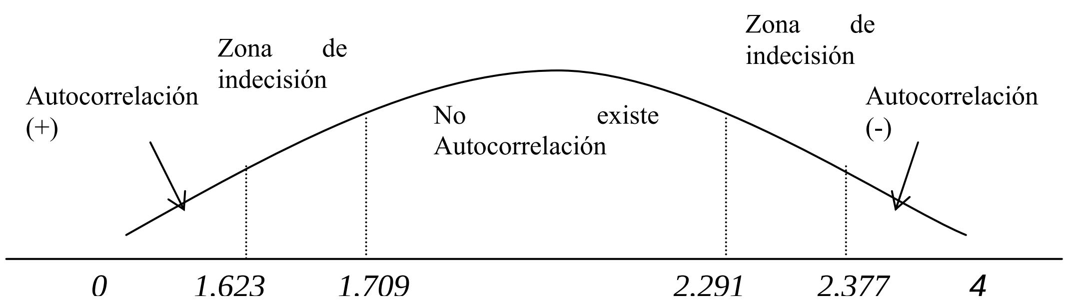
| | | | |
|---|---|---|---|
| R-squared | 0.957725 | Mean dependent var | 27.47113 |
| Adjusted R-squared | 0.956826 | S.D. dependent var | 6.567644 |
| S.E. of regression | 1.364651 | Akaike info criterion | 3.490113 |
| Sum squared resid | 175.0535 | Schwarz criterion | 3.569744 |
| Log likelihood | -166.2705 | F-statistic | 1064.777 |
| Durbin-Watson stat | 0.988372 | Prob(F-statistic) | 0.000000 |

### a. Estadístico DURBIN WATSON

Como sabemos el estadístico Durbin Watson mide la autocorrelación de primer grado.

Del modelo:

| | | |
|---|---|---|
| k = 3 | | $D_L$ = 1.623 |
| n =97 | $\Rightarrow$ | 4- $D_L$ = 2.377 |
| $\alpha$ = 0.05 | | $D_U$ = 1.709 |
| | | 4- $D_U$ = 2.291 |



Del Output tenemos que el indicador Durbin Watson es 0.988372, por lo tanto podemos afirmar que existen problemas de autocorrelación positiva de primer orden con un nivel de significancia del 5%.

### c. Test de LAGRANGE

H0 : El Modelo no tiene autocorrelación de orden p (p =2).

H1 : El Modelo si tiene autocorrelación de orden p (p =2).

| Breusch-Godfrey Serial Correlation LM Test: | | | | |
|---|---|---|---|---|
| F-statistic | 15.97282 | Probability | | 0.000001 |
| Obs*R-squared | 25.00070 | Probability | | 0.000004 |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 0.271809 | 0.584724 | 0.464850 | 0.6431 |
| VTACEM | -4.15E-06 | 4.32E-06 | -0.960251 | 0.3394 |
| VTABARRA | 3.86E-05 | 3.67E-05 | 1.052314 | 0.2954 |
| RESID(-1) | 0.495633 | 0.106093 | 4.671671 | 0.0000 |
| RESID(-2) | 0.051753 | 0.106989 | 0.483723 | 0.6297 |
| R-squared | 0.257739 | Mean dependent var | | 3.31E-15 |
| Adjusted R-squared | 0.225467 | S.D. dependent var | | 1.350361 |
| S.E. of regression | 1.188420 | Akaike info criterion | | 3.233296 |
| Sum squared resid | 129.9354 | Schwarz criterion | | 3.366013 |
| Log likelihood | -151.8149 | F-statistic | | 7.986412 |
| Durbin-Watson stat | 1.955360 | Prob(F-statistic) | | 0.000014 |

Como Prob.= 0.000042 < 0.05. Entonces se rechaza la hipótesis nula. Es decir el Modelo tiene problemas de autocorrelación con una probabilidad del 95%.

### d. Test Estadístico Ljung y Box

$H_0$ : El Modelo no tiene autocorrelación hasta el orden p(p =2).

$H_1$ : El Modelo si tiene autocorrelación de orden p (p =2).

Date: 09/20/02 Time: 10:37
Sample: 1 97
Included observations: 97

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. |**** |</td><td>. |**** |</td><td>1</td><td>0.491</td><td>0.491</td><td>24.143</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. | . |</td><td>2</td><td>0.275</td><td>0.045</td><td>31.800</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. |** |</td><td>3</td><td>0.325</td><td>0.230</td><td>42.611</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. | . |</td><td>4</td><td>0.239</td><td>-0.008</td><td>48.522</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. |*. |</td><td>5</td><td>0.199</td><td>0.069</td><td>52.669</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. | . |</td><td>6</td><td>0.153</td><td>-0.028</td><td>55.138</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. | . |</td><td>7</td><td>0.095</td><td>-0.018</td><td>56.106</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>.*| . |</td><td>8</td><td>0.041</td><td>-0.060</td><td>56.283</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. |*. |</td><td>9</td><td>0.111</td><td>0.115</td><td>57.625</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>.*| . |</td><td>10</td><td>0.008</td><td>-0.136</td><td>57.632</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. |*. |</td><td>11</td><td>0.026</td><td>0.099</td><td>57.709</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. |*. |</td><td>12</td><td>0.180</td><td>0.154</td><td>61.357</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>**| . |</td><td>13</td><td>-0.042</td><td>-0.247</td><td>61.563</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>**| . |</td><td>14</td><td>-0.202</td><td>-0.189</td><td>66.296</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>15</td><td>-0.111</td><td>0.008</td><td>67.751</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>.*| . |</td><td>16</td><td>-0.193</td><td>-0.171</td><td>72.160</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>. | . |</td><td>17</td><td>-0.192</td><td>0.042</td><td>76.565</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>18</td><td>-0.128</td><td>0.027</td><td>78.543</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>19</td><td>-0.145</td><td>0.028</td><td>81.117</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>20</td><td>-0.162</td><td>-0.025</td><td>84.384</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>21</td><td>-0.095</td><td>0.001</td><td>85.518</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>22</td><td>-0.060</td><td>0.061</td><td>85.974</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>23</td><td>-0.036</td><td>0.052</td><td>86.141</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>.*| . |</td><td>24</td><td>0.010</td><td>-0.077</td><td>86.154</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>.*| . |</td><td>25</td><td>-0.151</td><td>-0.123</td><td>89.189</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>.*| . |</td><td>26</td><td>-0.244</td><td>-0.107</td><td>97.216</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>.*| . |</td><td>27</td><td>-0.187</td><td>-0.118</td><td>102.02</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>28</td><td>-0.182</td><td>0.002</td><td>106.63</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>.*| . |</td><td>29</td><td>-0.208</td><td>-0.060</td><td>112.74</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>30</td><td>-0.164</td><td>-0.037</td><td>116.58</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>31</td><td>-0.140</td><td>0.008</td><td>119.44</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>32</td><td>-0.118</td><td>0.024</td><td>121.50</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. |*. |</td><td>33</td><td>0.027</td><td>0.165</td><td>121.61</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>34</td><td>0.029</td><td>-0.008</td><td>121.74</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>.*| . |</td><td>35</td><td>-0.064</td><td>-0.150</td><td>122.37</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>36</td><td>-0.048</td><td>-0.040</td><td>122.73</td><td>0.000</td></tr>
</table>

Debido a que las probabilidades son menores a 0.05 se puede decir que existe autocorrelación de orden 1,2,3 ...15 o 36. con una probabilidad del 95%

## 7. Corrección de la Autocorrelación

### a. Estadístico Durbin Watson

Analizaremos la autocorrelación de primer orden

PRIMER CAMBIO DE VARIABLE.
El coeficiente estimado de la variable rezagada nos servirá para corregir a las variables
Así corregimos nuevamente con d = 0.988372
Además:
$d = 2(1-\rho) \Rightarrow \rho = 1 - d/2$
$\rho = 0.505814$

Así generando nuevas variables:

1. pbicod = pbico-0.505814*pbico(-1) , de 1 - 97
2. vtacemd = vtacem-0.505814*vtacem(-1) , de 1 - 97
3. vtabarrad = vtabarra -0.505814* vtabarra(-1) , de 1 - 97

1. pbicod = ((1-0.505814^2)^(1/2))* pbico , de 1.
2. vtacemd = ((1-0.505814^2)^(1/2))* vtacem , de 1.
3. vtabarrad = ((1-0.505814^2)^(1/2))* vtabarra , de 1.

eqprimercambio

Dependent Variable: PBICOD
Method: Least Squares
Date: 09/20/02 Time: 11:04
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.658307 | 0.509521 | 1.292012 | 0.1995 |
| VTACEMD | 8.19E-05 | 4.99E-06 | 16.42719 | 0.0000 |
| VTABARRAD | 0.000103 | 3.85E-05 | 2.683052 | 0.0086 |

| | | | |
|---|---|---|---|
| R-squared | 0.886550 | Mean dependent var | 13.74272 |
| Adjusted R-squared | 0.884136 | S.D. dependent var | 3.549567 |
| S.E. of regression | 1.208231 | Akaike info criterion | 3.246630 |
| Sum squared resid | 137.2232 | Schwarz criterion | 3.326261 |
| Log likelihood | -154.4616 | F-statistic | 367.2785 |
| Durbin-Watson stat | 1.914361 | Prob(F-statistic) | 0.000000 |

Se observa que se ha corregido el problema de autocorrelación, por lo tanto ya no seria necesario realizar un segundo cambio de variable. El modelo es significativo individualmente y grupalmente sus probabilidades son menores que 0.05 nivel de significancia; así como, también su nivel de ajuste es bueno 88.86%.

SEGUNDO CAMBIO DE VARIABLE
El coeficiente estimado de la variable rezagada nos servirá para corregir a las variables
Así corregimos nuevamente con d = 1.914361

Además:
$d = 2(1-\rho) \Rightarrow \rho = 1 - d/2$

$\rho = 0.0428195$

Así generando nuevas variables:

1. pbicod1 = pbicod-0.0428195*pbicod(-1) , de 1 - 97
2. vtacemd1 = vtacemd-0.0428195*vtacemd(-1) , de 1 - 97
3. vtabarrad1 = vtabarrad -0.0428195* vtabarrad(-1) , de 1 - 97

1. pbicod1 = ((1-0.0428195^2)^(1/2))* pbicod , de 1.
2. vtacemd1 = ((1-0.0428195^2)^(1/2))* vtacemd , de 1.
3. vtabarrad1 = ((1-0.0428195^2)^(1/2))* vtabarrad , de 1.

eqsegundocambio

Dependent Variable: PBICOD1
Method: Least Squares
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.632366 | 0.503292 | 1.256460 | 0.2121 |
| VTACEMD1 | 8.17E-05 | 5.03E-06 | 16.24054 | 0.0000 |
| VTABARRAD1 | 0.000106 | 3.85E-05 | 2.765122 | 0.0068 |

| | | | |
|---|---|---|---|
| R-squared | 0.879736 | Mean dependent var | 13.16089 |
| Adjusted R-squared | 0.877177 | S.D. dependent var | 3.453524 |
| S.E. of regression | 1.210325 | Akaike info criterion | 3.250095 |
| Sum squared resid | 137.6994 | Schwarz criterion | 3.329725 |
| Log likelihood | -154.6296 | F-statistic | 343.8069 |
| Durbin-Watson stat | 1.990701 | Prob(F-statistic) | 0.000000 |

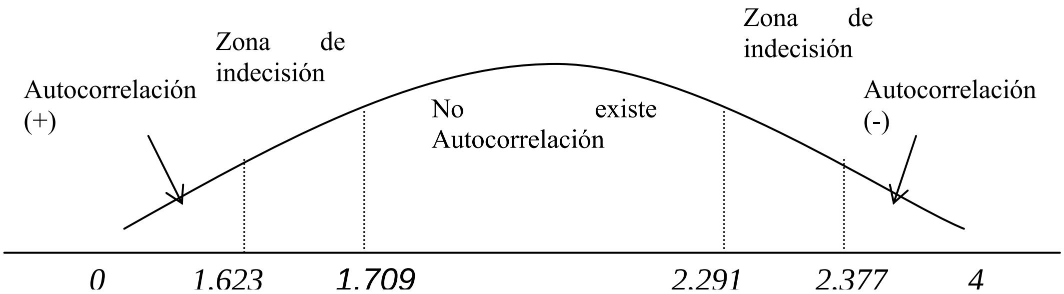
Se observa que ha mejorado mas aun la autocorrelación obteniéndose un valor de Durbin Watson alrededor de 2 las variables son significativas tanto individualmente y gradualmente. Lo que implica que ya no existen problemas de autocorrelación

Midiendo la autocorrelación de primer orden:

Donde:
k`= numero de variables explicativas
(numero de parámetros menos 1)
n = numero de observaciones

$D_{k`;\, n;\, \alpha} = D_{2`;\, 97;\, 0.05}$ : $\Rightarrow$

$D_L=1.297$
$4-D_L=2.703$
$D_U=1.570$
$4-D_U=2.430$



Se observa que estadístico Durbin Watson es 1.990701, encontrándose en la zona de no autocorrelación y tiene un valor cercano a 2 por lo tanto se puede afirmar que modelo ya no tiene problemas de autocorrelación. Sin embargo pierde ligeramente bondad de ajuste, aun cuando el modelo en conjunto sigue explicando a la variable dependiente (Prueba F).

**b. Procedimiento Iterativo De Cochran Orcutt**

Mediante el proceso Cochran Orcutt lo que se pretende es obtener una estimación para la autocorrelación, el cual esta basado en la estimación del ρ como coeficiente de autocorrelación propiamente dicha.

$$\rho = \frac{\sum \mu_t \mu_{t-1}}{\sum \mu_t^2}$$

Asimismo se puede estimar con el modelo:
$\mu_t = \rho\mu_{t-1} + \varepsilon_t$

PRIMERA ETAPA: Regresionamos el modelo por MCO:

Eqreisduo2

Dependent Variable: RESID1
Method: Least Squares
Sample(adjusted): 2 97
Included observations: 96 after adjusting endpoints

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.003863 | 0.120577 | 0.032035 | 0.9745 |
| RESID1(-1) | 0.504543 | 0.090485 | 5.575963 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.248549 | Mean dependent var | -0.007390 |
| Adjusted R-squared | 0.240555 | S.D. dependent var | 1.355476 |
| S.E. of regression | 1.181245 | Akaike info criterion | 3.191628 |
| Sum squared resid | 131.1619 | Schwarz criterion | 3.245052 |
| Log likelihood | -151.1981 | F-statistic | 31.09137 |
| Durbin-Watson stat | 1.949211 | Prob(F-statistic) | 0.000000 |

El coeficiente estimado de la variable RESID1 rezagada es 0.504543 el cual aproximamos como un $\hat{\rho} = 0.504543$, cuyo valor servirá para corregir a las variables. Para ello debo generar nuevas variables:

Así generando nuevas variables:

1. pbico2 = pbico-0.504543*pbico(-1) , de 1 - 97
2. vtacem2 = vtacem-0.504543*vtacem(-1) , de 1 - 97
3. vtabarra2 = vtabarra -0.504543* vtabarra(-1) , de 1 - 97

1. pbico2 = ((1-0.504543^2)^(1/2))* pbico , de 1.
2. vtacem2 = ((1-0.504543^2)^(1/2))* vtacem , de 1.
3. vtabarra2 = ((1-0.504543^2)^(1/2))* vtabarra , de 1.

Haciendo la regresión:

Eqcochran2

Dependent Variable: PBICO2
Method: Least Squares
Sample: 1 97
Included observations: 97

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 2.471690 | 0.846490 | 2.919926 | 0.0044 |
| VTACEM2 | 2.71E-05 | 3.98E-06 | 6.816105 | 0.0000 |
| VTABARRA2 | 0.000285 | 5.84E-05 | 4.888779 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.708297 | Mean dependent var | 13.75144 |
| Adjusted R-squared | 0.702091 | S.D. dependent var | 3.531647 |
| S.E. of regression | 1.927611 | Akaike info criterion | 4.180879 |
| Sum squared resid | 349.2743 | Schwarz criterion | 4.260510 |
| Log likelihood | -199.7727 | F-statistic | 114.1228 |
| Durbin-Watson stat | 2.065389 | Prob(F-statistic) | 0.000000 |

Se puede seguir haciendo una segunda da etapa pero el modelo esta significativamente bien representado por sus variables tanto individualmente cono colectivamente. Asimismo, el indicador de autocorrelación esta prácticamente tiene el valor de 2 y la bondad de ajuste es regular

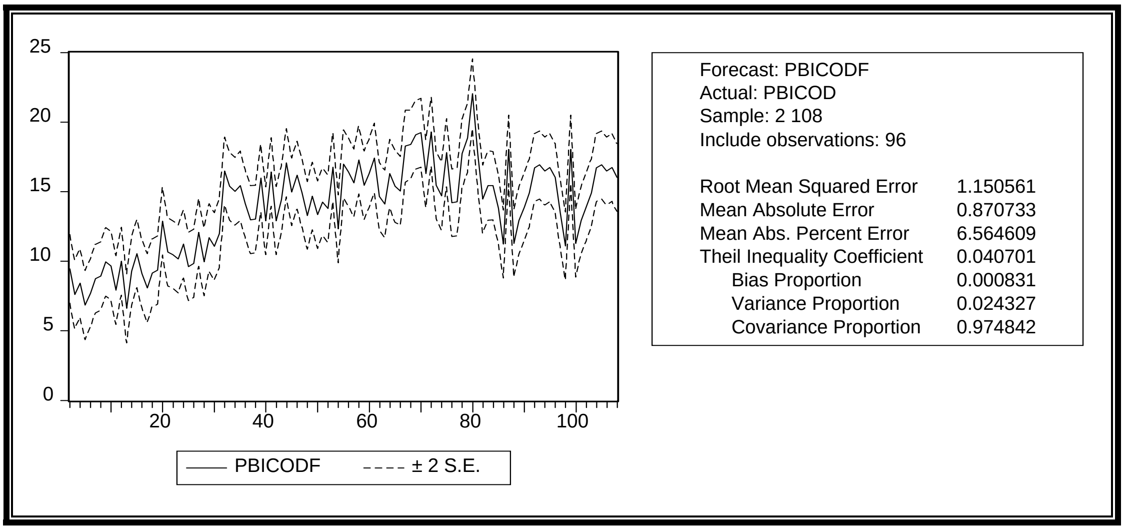
## 8. Predicción

De acuerdo a lo resultados de loa modelos podemos realizar la predicción. Como ejercicio podemos realizar una predicción para todo el año 2001 con los siguientes datos. Asumimos la proyección de ventas de cemento y barras de acero las mismas del año anterior en sus respectivos meses.

| ***Obs*** | **mes** | **vtacem** | **vtabarra** | **pbi construcción (soles)** |
|---|---|---|---|---|
| 98 | 200002 | 273058 | 25760 | |
| 99 | 200003 | 331499 | 27957 | |
| 100 | 200004 | 284956 | 24050 | |
| 101 | 200005 | 278223 | 24667 | |
| 102 | 200006 | 285707 | 25501 | |
| 103 | 200007 | 300335 | 27220 | |
| 104 | 200008 | 324273 | 32619 | |
| 105 | 200009 | 339146 | 34993 | |
| 106 | 200010 | 343501 | 34588 | |
| 107 | 200011 | 353612 | 30587 | |
| 108 | 200012 | 350593 | 27946 | |

ls pbicod c vtacemd vtabarrad

1. vtacemd = vtacem-0.505814*vtacem(-1) , de 1 - 108
2. vtabarrad = vtabarra -0.505814* vtabarra(-1) , de 1 - 108

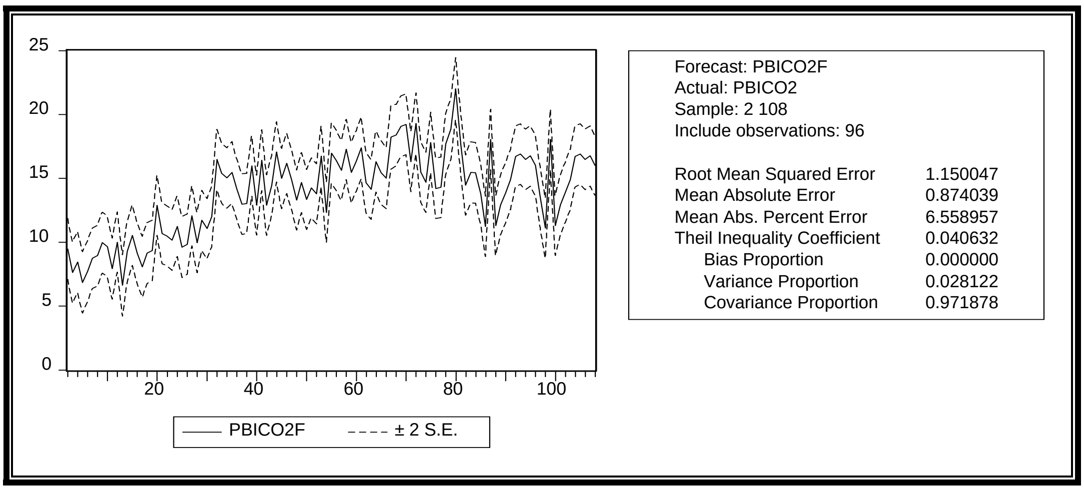



Se puede observar que el coeficiente de Thiel es 0.04701 un valor bajo por lo que podemos decir que este modelo es bueno par predecir

| **Obs** | **MES** | **PBICO (Proyectado)** |
|---|---|---|
| 98 | 200002 | 11.09940 |
| 99 | 200003 | 18.04491 |
| 100 | 200004 | 11.29099 |
| 101 | 200005 | 12.93600 |
| 102 | 200006 | 13.88217 |
| 103 | 200007 | 14.90450 |
| 104 | 200008 | 16.72722 |
| 105 | 200009 | 16.91694 |
| 106 | 200010 | 16.49161 |
| 107 | 200011 | 16.74780 |
| 108 | 200012 | 16.01764 |

Similarmente para el modelo para el modelo: ls pbico2 c vtacem2 vtabarra2

1. vtacem2 = vtacem-0.504543*vtacem(-1) , de 1 – 108
2. vtabarra2 = vtabarra -0.504543* vtabarra(-1) , de 1 – 108

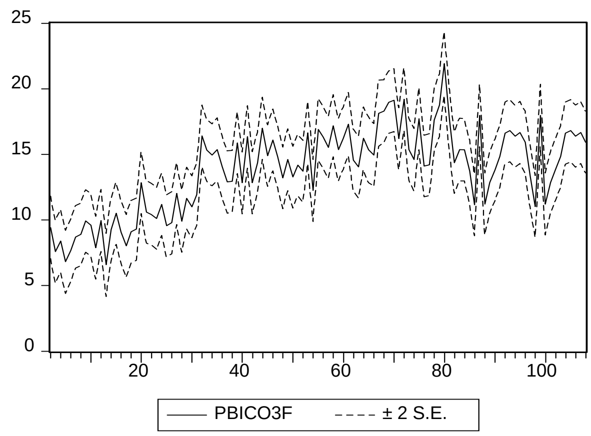


Se observa que el coeficiente de Thiel es de 0.40632 un valor bajo por lo tanto podemos decir también que es bueno para la predicción

| Obs | MES | PBICO (Proyectado) |
|---|---|---|
| 98 | 200002 | 11.11225 |
| 99 | 200003 | 18.03578 |
| 100 | 200004 | 11.30535 |
| 101 | 200005 | 12.93555 |
| 102 | 200006 | 13.87940 |
| 103 | 200007 | 14.89795 |
| 104 | 200008 | 16.70553 |
| 105 | 200009 | 16.89987 |
| 106 | 200010 | 16.48244 |
| 107 | 200011 | 16.75377 |
| 108 | 200012 | 16.02631 |

También podemos realizar para el siguiente modelo ls pbico3 c vtacem3 vtabarra3

1. vtacem3 = vtacem-0.5072557*vtacem(-1) , de 1 – 108
2. vtabarra3 = vtabarra -0.5072557* vtabarra(-1) , de 1 – 108




| | |
|---|---|
| Forecast: PBICO3F | |
| Actual: PBICO3 | |
| Sample: 2 108 | |
| Include observations: 96 | |
| Root Mean Squared Error | 1.149845 |
| Mean Absolute Error | 0.873703 |
| Mean Abs. Percent Error | 6.593306 |
| Theil Inequality Coefficient | 0.040843 |
| Bias Proportion | 0.000000 |
| Variance Proportion | 0.028356 |
| Covariance Proportion | 0.971644 |

Se observa que el coeficiente de Thiel es de 0.040843 un valor bajo por lo tanto podemos decir también que es bueno para la predicción

| Obs | MES | PBICO (Proyectado) |
|---|---|---|
| 98 | 200002 | 11.03173 |
| 99 | 200003 | 17.96265 |
| 100 | 200004 | 11.22143 |
| 101 | 200005 | 12.86271 |
| 102 | 200006 | 13.80746 |
| 103 | 200007 | 14.82383 |
| 104 | 200008 | 16.62779 |
| 105 | 200009 | 16.81514 |
| 106 | 200010 | 16.39360 |
| 107 | 200011 | 16.66291 |
| 108 | 200012 | 15.93456 |

Se observa que los valores son bastante semejantes por lo que existe buena consistencia en la predicción, podemos escoger el modelo que tiene el menor coeficiente de Thiel, en este caso segundo modelo.
Como las variables que se han mostrado son variables transformadas, entonces para realizar las proyecciones de la variable dependiente original simplemente se despeja de las ecuaciones siguientes

pbicodf = pbico-0.505814*pbico(-1)
pbicod2f = pbico-0.504543*pbico(-1)
pbicod3f = pbico-0.507256*pbico(-1)

Reemplazando y calculando podemos obtener las predicciones para cada observación

# Ejercicio de autoconocimiento

## ¿Porqué hacer un Levantamiento de Supuestos?

| | SI | NO | NO SÉ |
|---|---|---|---|
| 1. Considero útil para resolver problemas que se presentan cuando se hace un estudio. | | | |
| 2. Porque permitirá hacer un análisis de los modelos en cuanto a las variables que intervienen | | | |
| 3. Porque mediante su detección y su corrección permitirá dar buenas conclusiones. | | | |
| 4. Para que exista una eficacia en la evaluación de los coeficientes. | | | |
| 5. Para dar soluciones alternativas a los problemas. | | | |
| 6. Porque permite hacer recomendaciones para el comportamiento de las variables dependientes. | | | |
| 7. Para determinar si el modelo presenta problemas de autocorrelación, de multicolinealidad y/o de heteroscesdasticidad. | | | |
| 8. Para que el responsable de la gestión empresarial pueda corregir dichos problemas. | | | |
| 9. Para resolver problemas y solucionarlos | | | |
| 10. Porque al usar el levantamiento de supuestos se recomendará el modelo apropiado a trabajar. | | | |

**CALIFICACION**

Puntuar con un punto cada respuesta “SI”.
Si obtienes de 1 - 3 puntos tienes pocas expectativas de hacer un buen levantamiento de supuestos.
Si tienes entre 4 – 7, tienes buenas expectativas de hacer un buen levantamiento de supuestos.
Y si tienes entre 8 – 10, denotas excelentes expectativas de hacer un buen levantamiento de supuestos.

# RESUMEN

La multicolinealidad es un problema que aparece cuando las variables explicativas de un modelo están correlacionadas entre sí.

Hay dos de Clases de muticolinealidad:
-Multicolinealidad Perfecta $\quad Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu_i \quad X_3 = kX_2$
-Multicolinealidad Imperfecta $\quad Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu_i \quad X_3 = kX_2 + \mu_i$

La autocorrelación es la relación que se da entre las variables perturbadoras ($\mu$), contraviniendo uno de los supuestos para estimar el modelo a partir de la independencia que debería existir entre estas variables.
Este problema se presenta fundamentalmente cuando se realizan estudios econométricos de series históricas.

La heterocedasticidad es un caso particular que se presenta en un modelo cuando la varianza de error no es constante, es uno de los supuestos iniciales que se plantean al elaborar un modelo lineal general

# EXPLORACIÓN ON LINE

1. El Problema de la Multicolinealidad. Detección de la multicolinealidad.
   www.monografias.com/trabajos10/promu/promu.shtml - 54k

2. Multicolinealidad en el modelo de regresión lineal múltiple.

   www.udc.es/dep/mate/estadistica2/sec9_2.html - 57k -

3. Multicolinealidad. Autocorrelación

   www.stockssite.com/pf/17_Autocorrelacion.htm - 35k -

4. Autocorrelación**.** El método de Durbin Watson .

   www.stockssite.com/pf/17_Autocorrelacion.htm - 35k -

## LECTURA: Problema de Multicolinealidad

Uno de los supuestos del modelo de regresión lineal, es que no debe haber un alto grado de correlación entre las variables predeterminadas, pues esto, trae serias consecuencias que podemos resumir así:

Los estimadores por mínimos cuadrados ordinarios siguen siendo lineales, insesgados y óptimos pero las estimaciones tienen varianzas y covarianzas grandes.

Las razones t de uno o mas coeficientes tienden a ser estadísticamente no significativas, con lo que se pierde de perspectiva el análisis.

Aun cuando la razón t de uno o más coeficientes, es estadísticamente no significativa, el coeficiente de determinación tiende a ser elevado, con lo que se demuestra que no se puede separar el efecto individual de cada variable predeterminada hacia la endógena.

Luego entonces, es necesario que luego de estimado un modelo, tengamos que determinar la existencia o no de un alto grado de correlación entre las variables predeterminadas.

Los métodos de detección de multicolinealidad son:

**1.- Método de la relación entre t y $R^2$**

Mediante este método podemos determinar la existencia de multicolinealidad observando las razones t y si estas no son estadísticamente significativas y contamos con un coeficiente de determinación elevado (superior a 0.80), podemos estar ante un síntoma claro de multicolinealidad.

**2.- Método de la matriz de correlación**

Como el problema de multicolinealidad es un problema con las variables predeterminadas, establecemos una matriz de correlación entre aquellas

**3.- Método de la prueba F**

En un modelo de K-1 variables predeterminadas, es conveniente determinar cual de las mencionadas variables X esta correlacionada con las restantes para lo cual hay necesidad de hacer regresiones auxiliares de cada X con las restantes y obtener el $R^2$ correspondiente.

**4.- Método de los valores propios e índice de condición**

El tema de los valores propios es uno puramente matemático, que tiene que ver con el álgebra matricial y que de alguna manera ponemos de manifiesto en el apéndice y que son calculados por los paquetes econométricos y matemáticos del caso. En todo caso, partiendo de los valores propios de la matriz X'X, que es la que contiene las variables predeterminadas, se establece lo que se conoce como número de condición (K).

**Javier Uriol Chávez**

# ACTIVIDADES

1. ¿Cuál de los siguientes enunciados van de acuerdo a la definición de multicolinelidad?

   a. Se presentan con frecuencia en los problemas econométricos y no incumplen ninguna suposición del modelo lineal clásico.
   b. Un tamaño de muestra considerable (n>100) asegura la no existencia de multicolinealidad
   c. La varianza de los estimadores decrece a medida que el determinante de la matriz (X'X) este más alejado de 1.
   d. Los estimadores obtenidos son significativos ya que en presencia de multicolinelaidad se obtienen estimadores de varianza mínima.

2. Para explicar el comportamiento del ahorro neto familiar (ANF) se ha considerado conveniente considerar a la renta disponible familiar (RDF) y los impuestos directos pagados por las familias (IPF) Por lo que la función de ahorro familiar es:

$$ANF = \beta_1 + \beta_2 RDF + \beta_3 IPF$$

| Matriz de Correlación | | | |
|---|---|---|---|
| | AFN | RDF | IPF |
| AFN | 1 | 0.97284 | 0.96297 |
| RDF | 0.97284 | 1 | 0.99243 |
| IPF | 0.96297 | 0.99243 | 1 |

Dependent Variable: AFN
Method: Least Squares
Date: 11/11/04 Time: 11:51
Sample: 1982 2002
Included observations: 23

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| RDF | 0.0919 | 0.0268 | 3.4278 | 0.0017 |
| IPF | -0.0936 | 0.1870 | 0.2351 | 0.6202 |
| C | 29327.910 | 124724.6 | 0.2351 | 0.8156 |

| | | | |
|---|---|---|---|
| R-squared | 0.946834 | Mean dependent var | 1540511.0 |
| Adjusted R-squared | 0.943511 | S.D. dependent var | 1429436.0 |
| S.E. of regresión | 339740.7 | Akaike info criterion | 28.3916 |
| Sum squared resid | 3.69E+12 | Schwarz criterion | 28.5249 |
| Log likelihood | -493.8525 | F-statistic | 284.9423 |
| Durbin-Watson stat | 1.72674 | Prob(F-statistic) | 0.0000 |

Dependent Variable:Y AFN

| | | | | |
|---|---|---|---|---|
| Method: Least Squares | | | | |
| Date:09/29/04 Time: 11:59 | | | | |
| Sample: 1982 2002 | | | | |
| Included observations: 23 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| IPF | 0.5425 | 0.0264 | 20.5182 | 0.0000 |
| C | 368624.500 | 87373.6700 | 4.2189 | 0.0002 |
| R-squared | 0.927312 | Mean dependent var | | 1540511.0 |
| Adjusted R-squared | 0.925109 | S.D. dependent var | | 1429436.0 |
| S.E. of regresión | 391181.6 | Akaike info criterion | | 28.6472 |
| Sum squared resid | 5.05E+12 | Schwarz criterion | | 28.7361 |
| Log likelihood | -499.3256 | F-statistic | | 420.9957 |
| Durbin-Watson stat | 1.775222 | Prob(F-statistic) | | 0.0000 |

De acuerdo a las tablas mostradas, ¿El modelo propuesto inicialmente es adecuado para explicarla variabilidad del ahorro? Si no lo es ¿cuál cree que lo será?

3. Considere un modelo que trata de explicar el número de tarjetas de crédito por familia, para lo cual se propone:

   Y : Número de tarjetas de crédito por familia
   $X_1$ : Número de miembros por familia
   $X_2$ : Renta familiar
   $X_3$ : Número de vehículos por familia

   Identificar qué problema que se presenta por considerar las variables anteriormente mencionadas

4. Con relación a la heterocedasticidad que afirmación es correcta

   a. No se produce una pérdida de eficiencia en el estimador de mínimos cuadrados.
   b. Por lo general no se presentan en muestras cuyos datos son valores que se han obtenido agregando o promediando datos individuales.
   c. No es una violación una de las hipótesis sobre las que se asienta el modelo de regresión lineal clásico.
   d. Los contrastes estadísticos usados no plantean la hipótesis nula si un modelo presenta o no heterocedasticad, si no la presencia de homocedasticidad.

# AUTOEVALUACIÓN

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

1. La muticolinealidad aparece cuando las variables………….. de un modelo están correlacionadas entre sí.

   a. Independientes
   b. Dependientes
   c. Explicadas
   d. Explicativas

2. Las clases de multicolinealidad son:

   a. Simple y compuesta
   b. Variable y fija
   c. Perfecta e imperfecta
   d. Ninguna de las anteriores

3. Una de las reglas prácticas para detectar la multicolinealidad es:

   a. $R^2$ es alto, no existe significancia individual de los regresores pero sí conjunta.
   b. Existe significancia conjunta de todos los regresores
   c. Modelo no es significativo con un $R^2$ elevado
   d. Ninguna de las anteriores

4. Uno de los métodos para corregir la multicolinealidad sería:

   a. Anular dos variables explicativas
   b. Ampliando el tamaño la muestra
   c. Ampliando la población
   d. Ninguna de las anteriores

5. La autocorrelación se presenta fundamentalmente cuando se realizan estudios econométricos de:

   a. Series históricas
   b. Estudios de corte transversal
   c. Modelos significativos
   d. Ninguna de las anteriores

5. Para poder describir el ahorro neto familiar se ha considerado a la renta neta disponible (RNDFAM) y los impuestos directos pagados por las familias (TDFAM). Para tal fin se ha considerado el modelo siguiente:

$$SNF=\beta_1 + \beta_2 RNDFAM + \beta_3 TDFAM + \mu$$

| Matriz de Correlación |
|---|

| | SNFAM | RNDFAM | TDFAM |
|---|---|---|---|
| SNFAM | 1.0000 | 0.9728 | 0.9630 |
| RNDFAM | 0.9728 | 1.0000 | 0.9924 |
| TDFAM | 0.9630 | 0.9924 | 1.0000 |

Modelo de Regresión Inicial

Dependent Variable: SNFAM
Method: Least Squares
Sample: 1964 1998 Included observations: 35

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| RNDFAM | 0.091864 | 0.0268 | 3.42779 | 0.0017 |
| TDFAM | -0.093563 | 0.186988 | -0.500366 | 0.6202 |
| C | 29327.91 | 124724.6 | 0.235141 | 0.8156 |

| | | | |
|---|---|---|---|
| R-squared | 0.946834 | Mean dependent var | 1540511 |
| Adjusted R-squared | 0.943511 | S.D. dependent var | 1429436 |
| S.E. of regression | 339740.7 | Akaike info criterion | 28.39157 |
| Sum squared resid | 3.69E+12 | Schwarz criterion | 28.52489 |
| Log likelihood | -493.8525 | F-statistic | 284.9423 |
| Durbin-Watson stat | 1.72674 | Prob(F-statistic) | 0.00000 |

Modelo de Regresión usando a TDFAM como única regresora

Dependent Variable: SNFAM
Method: Least Squares
Sample: 1964 1998 Included observations: 35

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| TDFAM | 0.54 | 0.0264 | 20.5182 | 0.00 |
| C | 368624.50 | 87373.6700 | 4.2189 | 0.00 |

| | | | |
|---|---|---|---|
| R-squared | 0.927312 | Mean dependent var | 1540511.000 |
| Adjusted R-squared | 0.925109 | S.D. dependent var | 1429436.000 |
| S.E. of regression | 391181.6 | Akaike info criterion | 28.647 |
| Sum squared resid | 5.05E+12 | Schwarz criterion | 28.736 |
| Log likelihood | -499.3256 | F-statistic | 420.996 |
| Durbin-Watson stat | 1.775222 | Prob(F-statistic) | 0.000 |

Modelo de Regresión usando a RNDFAM como única regresora

Dependent Variable: SNFAM
Method: Least Squares
Sample: 1964 1998 Included observations: 35

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| RNDFAM | 0.079 | 0.003 | 24.143 | 0.000 |
| C | 75433.000 | 83098.950 | 0.908 | 0.371 |

| | | | |
|---|---|---|---|
| R-squared | 0.946418 | Mean dependent var | 1540511.000 |
| Adjusted R-squared | 0.944794 | S.D. dependent var | 1429436.000 |
| Sum squared resid | 3.72E+12 | Schwarz criterion | 28.431 |
| Log likelihood | -493.9889 | F-statistic | 582.875 |
| Durbin-Watson stat | 1.801663 | Prob(F-statistic) | 0.000 |

De acuerdo a los resultados siguientes que modelo considera que es el más conveniente para explicar las variaciones en el ahorro neto disponible

a. $SNF=\beta_1 + \beta_2 RNDFAM + \beta_3 TDFAM + \mu$
b. $SNF=\beta_1 + \beta_3 TDFAM + \mu$
c. $SNF=\beta_1 + \beta_2 RNDFAM + \mu$
d. $SNF= \beta_1 RNDFAM + \beta_2 TDFAM + \mu$

6. Se ha recogido información acerca de 50 empresas dedicadas a la producción y distribución de verduras, la información obtenida es concerniente a las ventas, sus recursos propios, personal que labora en la empresa y el número de puntos de venta. Con la información disponible se pretende explicar el comportamiento de las ventas en función las otras variables mencionadas, por tal motivo lo que se pretende es analizar al significación individual y conjunta de las variables. Un paso inicial fue elaborar las pruebas siguientes con fin de garantizar los resultados que se obtienen por mínimos cuadrados ordinarios:

| Dependent Variable: VENTAS | | | | |
|---|---|---|---|---|
| Method: Least Squares | | | | |
| Sample: 1 50 | Included observations: 50 | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| RECURSOS | 2.487487 | 0.20379 | 12.20612 | 0.0000 |
| EMPLEO | 13.25027 | 1.579404 | 8.389411 | 0.0000 |
| TIENDAS | 84.77498 | 17.34995 | 4.886181 | 0.0000 |
| C | 3330.98 | 4730.742 | 0.704114 | 0.4849 |
| R-squared | 0.956122 | Mean dependent var | | 80419.7000 |
| Adjusted R-squared | 9.53E-01 | S.D. dependent var | | 118268.6000 |
| S.E. of regression | 25568.84 | Akaike info criterion | | 23.2128 |
| Sum squared resid | 3.01E+10 | Schwarz criterion | | 23.3657 |
| Log likelihood | -576.3189 | F-statistic | | 334.1221 |
| Durbin-Watson stat | 2.026817 | Prob(F-statistic) | | 0.0000 |

Donde:

Ventas : Total de ventas producidas por cada empresa
Recursos : Recursos propios
Empleo : Personal que labora en la empresa
Tiendas : Número de puntos de venta

| White Heteroskedasticity Test: | | | |
|---|---|---|---|
| F-statistic | 6.137897 | Probability | 0.000022 |
| Obs*R-squared | 29.00066 | Probability | 0.000648 |

| Constraste Goldfed y Quandt | | | |
|---|---|---|---|
| | sample | S.E. of regression | |
| Primera muestra | 1 a 19 | 5286.5198 | |
| Primera muestra | 32 a 50 | 42723.3127 | |
| F-statistic | 65.3115 | Probability | 0.0000 |

De los resultados mostrados a partir del uso de E-Views, indicar la alternativa correcta:

a. No existe heterocedasticidad.
b. Existe heterocedasticidad.
c. No se puede afirmar nada al respecto.
d. Se posee información insuficiente.

7. Se desea construir un modelo el cual sea capaz de explicar la inversión en función del PBI a precios de 1990 y el grado de utilización de la capacidad productiva. Para cumplir con tal fin se dispone de información anual desde 1964 hasta 1998.

| | | | | |
|---|---|---|---|---|
| Dependent Variable: I80 | | | | |
| Method: Least Squares | | | | |
| Sample: 1964 1998 Included observations: 35 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| PIBCF80 | 0.278864 | 0.014243 | 19.57922 | 0.0000 |
| CU | 7563351 | 2028877 | 3.727852 | 0.0007 |
| C | -6445864 | 1742946 | -3.698258 | 0.0008 |
| R-squared | 0.928005 | Mean dependent var | | 3776977 |
| Adjusted R-squared | 0.923505 | S.D. dependent var | | 1163356 |
| S.E. of regression | 321758 | Akaike info criterion | | 28.2828 |
| Sum squared resid | 3.31E+12 | Schwarz criterion | | 28.41612 |
| Log likelihood | -491.9491 | F-statistic | | 206.2367 |
| Durbin-Watson stat | 0.338568 | Prob(F-statistic) | | 0.0000 |

| Durbin-Watson Test | |
|---|---|
| $\rho$ | 0.661432 |
| d | 0.338568 |
| $d_L$ | 1.584 |
| $d_U$ | 1.343 |

De los resultados mostrados, indicar la alternativa correcta:

a. No existe Autocorrelación
b. Existe Autocorrelación
c. Se posee información insuficiente.
d. Existe una Autocorrelación negativa.

RESPUESTAS DE CONTROL

1. d, 2. c, 3. a, 4. b, 5. a ,6.c 7.c 8.b