Escuela de Gobierno y Gestión Pública

Universidad de Chile

**Guía 5**

**Estadística Aplicada**

*Profesor: Francisco Ormazabal*

*Ayudantes: Ignacio Monge*

*Daniela Paz*

**Autocorrelación**

Esto puede ser definido como “correlación entre miembros de series de observaciones ordenadas en el tiempo”. Simbólicamente podriamos expresarlo como:

E (µiµj) = 0 con ij

Esto quiere decir que se supone que el error relacionado con una observación cualquiera no se ve influenciado por la perturbación relacionado con otra observación. Sin embargo, si la relación es distinta de 0, existe autocorrelación, es decir, los errores estarán relacionados, y con ello los estimadores dejan de ser eficientes. Y eso es lo que tenemos que arreglar, de modo en que podamos tener un modelo bien definido.

Causas de la autocorrelación

* Inercia: Existen tendencias que influyen en los valores futuros
* Sesgo de especificación: Se escoge mal la función o se omiten variables, lo que puede provocar un comportamiento sistemático en lo aleatorio de las otras variables.
* Tiempo de ajuste: Los agentes económicos deben analizar la información, y eso toma tiempo. Por ende, eso puede afectar en periodos futuros.
* Preparación de los datos: Al ordenar los datos con respecto a una variable, se podría “apreciar” un proceso de autocorrelación.

Problemas de la autocorrelación:

* Estimadores pocos eficientes
* Invalidez de las pruebas de contraste

Formas de determinar la autocorrelación:

1.- Método gráfico

2.- Test de rachas (prueba de Geary)

Podemos definir una racha como una sucesión ininterrumpida de un símbolo, en este caso + y -. La longitud de la racha es cuántas veces se presenta este símbolo.

Entones, sea:

N = Número total de observaciones (N1+N2)

N1 = Cantidad de +

N2 = Cantidad de –

R = Cantidad de rachas

Media = \begin{displaymath}\mu_R=\frac{2\,n_1 n_2}{n_1+n_2} + 1
\end{displaymath}

Varianza = \begin{displaymath}\sigma_R^2=\frac{2\,n_1n_2 (2\,n_1n_2 -n_1-n_2)}{(n_1+n_2)^2
(n_1+n_2-1)}
\end{displaymath}

Entonces, se establece un IC en función de R (+- media x valor z). Comúnmente este será del 95%.

Regla de decisión: No se debe rechazar la hipótesis nula al valor Z si R está en el IC anterior, se rechaza si la R estimada está fuera de esos límites ( Si R pertenece al IC se acepta la hipótesis nula, en caso contrario, se acepta la alternativa)

Ejemplo:

(+++)(- -)(+)(-)(++)(-)

Aplicando las fórmulas, tenemos:

E(R)= 5.8

Var(R)= 2.02

De(R)= 1.42

IC al 95%

5.8 - 1.96 x 1.42 = 3.01 // 5.8 + 1.96 x 1.42 = 8.58

Entonces, se acepta la hipótesis nula. No existe autocorrelación

3.- Durbin Watson

Estadístico *d*, definido como

d = {\sum_{t=2}^T (e_t - e_{t-1})^2 \over {\sum_{t=1}^T e_t^2}}. lo que corresponde a la suma de las diferencias al cuadrado de los residuos sucesivos sobre el SCR. Posee un n-1 ya que se pierde una observación al obterner estas diferencias.

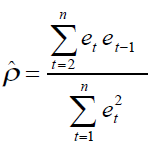
Se basa en los siguientes supuestos:

* Se incluye el término intersección, por lo cual, si no está presente como en el caso de que comience en el origen, es necesario realizar nuevamente la regresión para obtener la SCR.
* Las variables explicativas x, son constantes en un muestreo repetitivo
* Las perturbaciones se generan mediante el esquema autorregresivo de primer orden
* Se supone que los errores están distribuidos normalmente
* El modelo no incluye valores rezagados (t-1)
* No faltan observaciones en los datos

Podemos interpretar que:

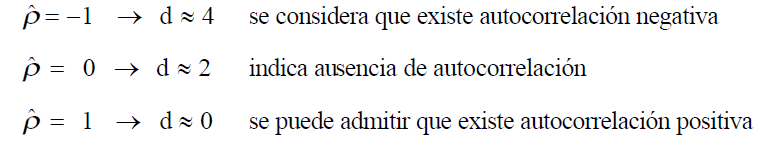
* Si existe una autocorrelación positiva, los valores tendrán distancias muy pequeñas, por lo que será cercano a 0
* Si es negativa, los valores tendrán una gran diferencia y el estadístico será cercano al límite superior, que en este caso, es 4
* Si no existe autocorrelación, la distancia será bastante heterogénea, por lo cual el estadístico tomará un valor intermedio.

Entonces:

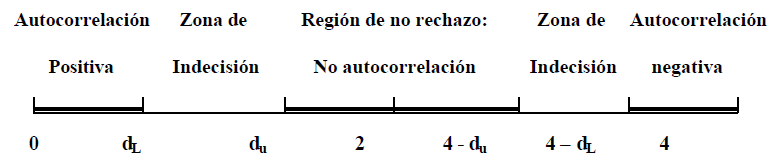
donde la sumatoria de los errores al cuadrado es SCR

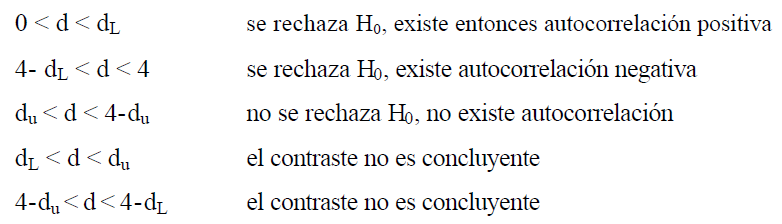
Por otra parte, tenemos que 

La variación de rho es , y con ello, se puede deducir que conforme a la variación del estadístico  *d* y el signo de la autocorrelación

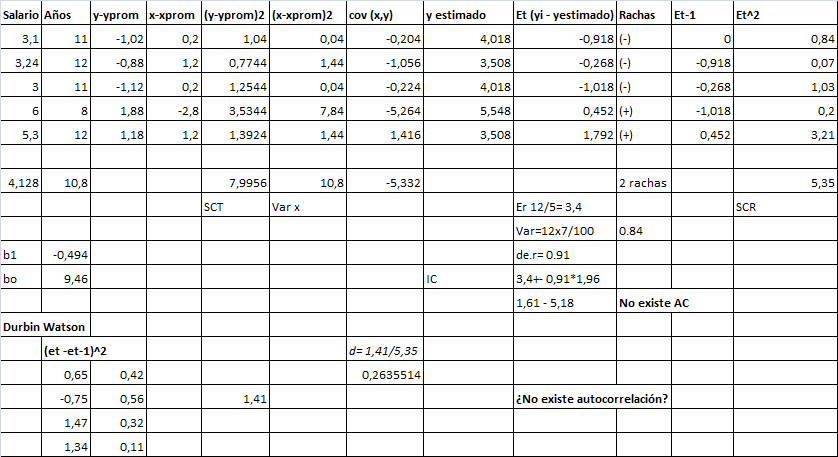


Gráficamente podemos verlo de la siguiente manera:





Ejemplo

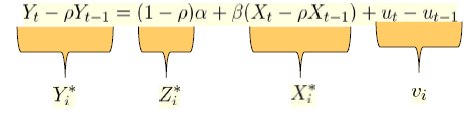


¿Cómo corregirlo?

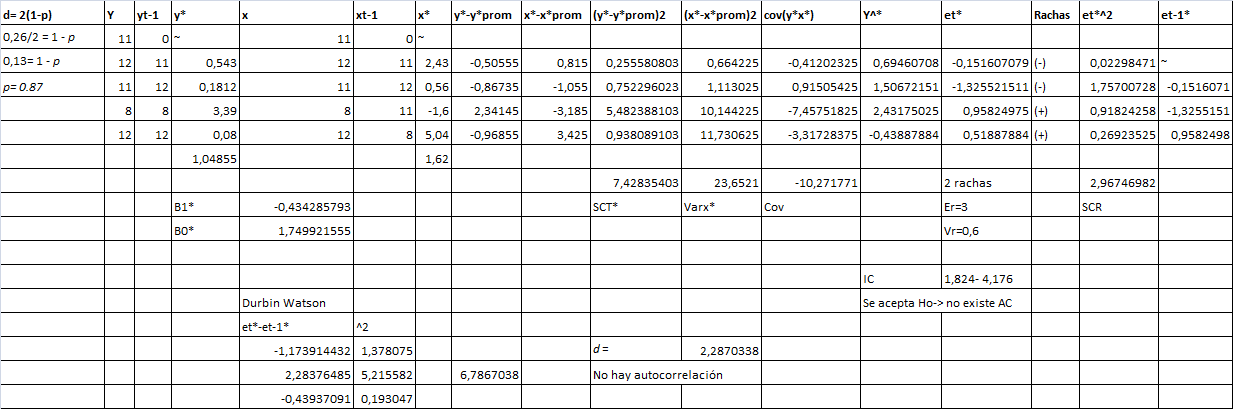
Mínimos cuadrados generalizados (MCG)

Al conocer rho, podemos decir que MCG se expresa 

Donde:

donde Z\*= β\*1 y vi= β\*2=β2

Ahora, retomando el ejemplo anterior…



Para dejar de hacer modekos, debe caer en la región de aceptación en el Durbin Watson.

Si cae en indecisión se aplica rachas -> se obtiene rho -> MCG

¿Qué pasa si desconozco rho?

Método de la primera diferencia: Dado que rho está en dos posiciones extremas [-1,1], se puede suponer que rho= 0, donde no hay correlación y del otro extremo, considerar rho como +- 1, osea, una autocorrelación positiva o negativa perfecta, donde:

Yt - Yt-1 = β1 (Xt - Xt - 1) + (µt - µt - 1)