



CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS  
CIMAT

**MODELOS DE REGRESIÓN LINEAL  
CON PROBLEMAS DE  
AUTOCORRELACIÓN**

**GUILLERMO ANTONIO DE LIRA ARENAS**

Tesina para obtener el grado de  
Maestro en Ciencias con Orientación en Estadística Oficial

Profesor Supervisor:  
**DR. JOAQUÍN ORTEGA SÁNCHEZ**

Guanajuato Gto, mayo de 2011

© 2011, Guillermo Antonio de Lira Arenas





CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS  
CIMAT

## **MODELOS DE REGRESIÓN LINEAL CON PROBLEMAS DE AUTOCORRELACIÓN**

**GUILLERMO ANTONIO DE LIRA ARENAS**

Tesina presentada a la Comisión integrada por los profesores:

**DRA. GRACIELA MARÍA DE LOS DOLORES GONZÁLEZ FARÍAS**

**DR. ROGELIO RAMOS QUIROGA**

**DR. JOAQUÍN ORTEGA SÁNCHEZ**

Para completar las exigencias del grado de  
Maestro en Ciencias en Estadística Oficial

Guanajuato Gto, mayo de 2011



El presente trabajo está dicado especialmente a mi hijo  $\neq$  y esposa, que siempre me apoyó.



## **AGRADECIMIENTOS**

Quiero agradecer el apoyo que siempre recibí por mi esposa y mis padres quienes me alentaron para seguir adelante con esta aventura académica, también a mis catedráticos el Dr. Rogelio y el Dr. Joaquín, quienes me orientaron en el proceso de aprendizaje por último a mis compañeros por su apoyo en tiempos difíciles.





## ÍNDICE GENERAL

	Pág.
DEDICATORIA.....	iii
AGRADECIMIENTOS .....	v
ÍNDICE DE TABLAS .....	viii
1. <i>Introducción</i> .....	1
2. <i>Antecedentes</i> .....	4
3. <i>MCG en los procedimientos de C-O y P-W</i> .....	8
4. <i>Aplicaciones</i> .....	17
5. <i>Conclusiones</i> .....	31
BIBLIOGRAFIA.....	32
A N E X O S.....	33
Anexo A: <i>Glosario de definiciones</i> .....	35
Anexo B: <i>Código de cálculos y gráfica</i> .....	37
ANEXO c: <i>Distribución del estadístico de contraste de Durbin y Watson</i> .....	43

## ÍNDICE DE TABLAS

	Pág.
Tabla 1.1: <i>Indicadores económicos para modelo de la aplicación 1</i> .....	17
Tabla 1.2: <i>Resultados de la aplicación 1</i> .....	20
Tabla 1.3: <i>Valores de la correlación en cada iteración (Aplicación 1)</i> .....	21
Tabla 2.1: <i>Datos para la aplicación 2</i> .....	23
Tabla 2.2: <i>Resultados de la aplicación 2</i> .....	28
<i>ANEXO C Tabla 1: Distribución de estadístico de Durbin-Watson</i> .....	43

## 1. *Introducción*

Es conocido que la técnica de mínimos cuadrados ordinarios es ampliamente aplicada en varias ramas del conocimiento, algunas ocasiones utilizada como herramienta descriptiva y otras para hacer inferencia. En estas aplicaciones generalmente los supuestos básicos del modelo comúnmente no son verificados aunque hay quienes consideran que no es del todo necesario si se conoce bien la naturaleza de los datos, sin embargo es importante tener en cuenta que se pueden cometer algunos errores indeseables si se realiza inferencia sobre un modelo obtenido de la violación de alguno de los supuestos. En tal caso se deben tomar las precauciones necesarias para no distorsionar la aplicación de la técnica y caer en conclusiones erróneas.

Por ejemplo, para fenómenos temporales los datos obtenidos comúnmente presentan ciclos y tendencias, los cuales al realizar una regresión lineal son captados por los residuales del modelo causando que estos estén relacionados con el valor del residual inmediato anterior. A lo anterior se le conoce como residuales correlacionados o autocorrelacionados.

Aún cuando esta situación no es exclusiva para este tipo de datos, es importante y se recomienda que después de estimar un modelo de regresión lineal, se le realice una prueba para descartar residuales autocorrelacionados ya que esto acarrea una serie de problemas que hacen inapropiada la inferencia sobre este tipo de modelos.

De esta manera cuando se detectan problemas de autocorrelación en los residuales de un modelo de regresión lineal, la mayoría de los textos sugieren usar procedimientos en dos etapas o iterativos para solucionar esta situación, no siempre dejando claro algunos puntos clave de la utilización de las técnicas y procedimientos, tales como:

- Las diferencias que existen en los resultados obtenidos por procedimientos en dos etapas, los cuales pueden diferir dramáticamente de lo obtenido por procesos iterativos.
- Detalles explícitos para obtener los resultados que ofrecen la mayoría de los paquetes de regresión, ya que estos implementan procedimientos iterativos, cuya metodología por sí misma no es expresada de una manera sencilla para los estudiantes e investigadores inexpertos.
- La importancia que tiene el factor de corrección del error para pronósticos en modelos con residuales autocorrelacionados, ya que en ocasiones éste factor es omitido en los cálculos.
- El cálculo adecuado de  $R^2$  del modelo resultante de Mínimos Cuadrados Generalizados, ya que la estimación que reportan algunos paquetes de regresión no es comparable con la obtenida por Mínimos Cuadrados Ordinarios.
- Dada la variedad de métodos en dos etapas e iterativos en la literatura no siempre es claro cuál método es mejor.

Este trabajo resume brevemente los pasos para implementar procedimientos iterativos, explícitamente para modelos con residuales autocorrelacionados con los cuales se pretende generar pronósticos, en donde además surge la necesidad de hacer comparaciones apropiadas en relación al ajuste dado por  $R^2$ .

Es importante señalar que gran parte de los modelos en que aparecen problemas de autocorrelación, presentan procesos autorregresivos de primer orden. Estos se definen como primer orden de autocorrelación ( $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ ) y es el único caso que se estudiará en este trabajo es la estimación de modelos de regresión lineal homocedásticos con problemas de autocorrelación de primer orden.

Es necesario señalar que existen órdenes mayores de autocorrelación ( $\varepsilon_t = \rho_1\varepsilon_{t-1} + \dots + \rho_k\varepsilon_{t-k} + v_t$ ) e incluso casos en que los residuales no pueden ser modelados por un AR(p) y es necesario utilizar MA(q) o modelos mixtos, pero estos temas están fuera del alcance de este documento.

La existencia de tendencia y ciclos en los datos son causas frecuentes de autocorrelación, esto es, si la variable endógena del modelo tiene una tendencia creciente o presenta un comportamiento cíclico que no es explicado por las variables exógenas, entonces el término de error (los residuales) captará ese ciclo o tendencia. La manipulación de datos, puede ser causa de autocorrelación. Por ejemplo si se cuenta con información mensual de  $n$ -años y se pretende obtener información trimestral manipulando los datos mediante promedios u otra medida de estimación, es preferible extraer los datos originales de cada trimestre. Otra causa común de la autocorrelación es la no estacionalidad en una serie de datos.

Por lo general los estudiantes aprenden que cuando los residuales de una regresión lineal son autocorrelacionados, surge una serie de problemas derivados de dicha correlación o autocorrelación. Si la autocorrelación es ignorada, los estimadores de *Mínimos Cuadrados Ordinarios (MCO)* resultan ser imparciales, consistentes, normales y asintóticamente distribuidos, sin embargo al ser obtenidos en presencia de autocorrelación, los estimadores dejan de ser eficientes, esto es dejan de ser de varianza mínima.

La consecuencia más grave de la autocorrelación en los residuales de un modelo de regresión es que los estimadores de *MCO* dejan de ser eficientes, además si se trabaja bajo el supuesto de que los residuales son homocedásticos pero están correlacionados, la matriz de covarianzas resultaría ser positiva y simétrica, pero no diagonal, con lo cual el estimador de la varianza sería sesgado, por esta razón no es válida la inferencia realizada bajo estas condiciones y la aplicación de las pruebas  $t$ ,  $F$  y  $\chi^2$  resultan inapropiadas y puede conducir a conclusiones erróneas.

Para corregir los problemas generados por la presencia de la autocorrelación, a menudo se utiliza *Mínimos Cuadrados Generalizados (MCG)* en procedimientos iterativos tales como *Cochrane-Orcutt* o *Prais-Winsten*, los cuales son basados en un estimador particular del

coeficiente de correlación (Greene, 2003). Adicional a los procedimientos que se verán en este trabajo existen técnicas de reciente aparición como *cointegración* para el análisis de series temporales. El concepto de cointegración lo introdujo Granger (1981) y posteriormente, se estudió en mayor profundidad en Engle y Granger (1987), Engle y Yoo (1987) y Johansen (1988), entre otros. Esta técnica surge como procedimiento que permite discriminar relaciones reales de relaciones espurias entre variables. Aún cuando el tema no está dentro de los alcances de este trabajo es importante conocer su existencia.

Es importante definir de una manera clara los objetivos y alcances de este trabajo, los cuales son:

*Primero*, ofrecer los elementos suficientes para que los estudiantes e investigadores principiantes tomen las precauciones necesarias al elegir procedimientos en dos etapas (*two-step*) o procedimientos iterativos cuando se corrige la autocorrelación, ya que estos pueden producir resultados que difieran bastante unos de otros, esto se ilustrará en las aplicaciones presentadas en éste trabajo.

*Segundo*, expresar de manera clara y precisa cada uno de los pasos que conforman los procesos iterativos ya que la metodología por sí misma no es expresada de una manera clara en la mayoría de los textos. La importancia de esto radica en que la mayoría de los paquetes de regresión tienen implementados procesos iterativos automáticos que tienen que ser interpretados por los estudiantes.

*Tercero*, ofrecer sugerencias para tratar de responder a la pregunta ¿Cuál método es mejor? Ya que existen varias metodologías disponibles para corregir la autocorrelación en los residuales de un modelo de regresión, en este trabajo se dará un breve resumen de los procedimientos iterativos.

*Cuarto*, expresar de una manera clara y sencilla la transformación adecuada para hacer comparaciones apropiadas entre los ajustes obtenidos por *MCO* y *MCG* (a partir de los procedimientos que se verán más adelante) ya que estos ajustes no son directamente comparables, además de mencionar el error que se comete comúnmente al calcular y reportar  $R^2$ .

*Quinto*, mientras la teoría de pronósticos con la presencia de autocorrelación ha estado bien documentada, los estudiantes a menudo tienen dificultad con la aplicación formal de esta técnica, por lo que se detallará cada uno de los pasos para calcular pronósticos en base a los residuales en observaciones anteriores, esto es, se detallará el cálculo de pronósticos para el tiempo  $T + 1$  utilizando la descripción de los residuales del modelo.

Como se mencionó, para corregir el problema de la autocorrelación en el modelo de regresión lineal es necesario realizar una transformación del modelo que lleva al uso de la técnica de *MCG* en los procedimientos de *Cochrane-Orcutt* o *Prais-Winsten*, por lo que en el *Capítulo 3* se mostrará un resumen del procedimiento de *MCG* en estos procedimientos.

## 2. *Antecedentes*

Antes de ver los procedimientos para corregir el problema de autocorrelación, se ilustrarán las consecuencias de omitir la autocorrelación en los residuales de un modelo, para ello se considerará un conjunto de datos de control definido a continuación.

Como ya se mencionó, si un modelo de regresión lineal tiene residuales autocorrelacionados surge una serie de problemas, en donde resulta inadecuada la inferencia sobre estos modelos. Entre esos problemas se encuentra la subestimación ó sobre estimación de la variabilidad, lo cual afecta a las pruebas de hipótesis y por otro lado los estimadores del modelo no son de varianza mínima y los intervalos de confianza no serán adecuados. Sin pérdida de generalidad se mostrará un ejemplo en este sentido donde bastará ver los intervalos de confianza para ilustrar la consecuencia de omitir los problemas de la autocorrelación.

Considere el siguiente conjunto de datos control, sean las variables predictoras dadas por  $x_1$  y  $x_2$  y una variable de respuesta  $y$  afectada por un ruido autorregresivo de primer orden con  $\rho = 0.7$  de la siguiente manera:

$$y = b_0 + b_1x_1 + b_2x_2 + \text{sim.ar}$$

Donde  $b_0 = 5$ ,  $b_1 = 1$ ,  $b_2 = 1$

```
x1 <- c(8.1, -1.2, 1.3, 7.9, 1.9, 0.8, 0.7,-0.4, 1.5, 3.3, 2.8, 0.7, 1.7 ,1.0, 1.0,
1.3, 1.3, 1.6, 2.9, 3.1, 4.2, 5.5, 5.7, 4.4, 3.2, 6.2, 11.0, 9.1, 5.8, 6.5, 7.6,
11.3, 13.5, 10.3, 6.2, 3.2, 4.3, 3.6, 1.9, 3.6, 4.1, 4.8, 5.4 ,4.2, 3.0, 3.0, 2.6,
2.8, 3.0)
```

```
x2 <- c(-4.6,-0.2, 1.1,-1.9, 0.4, 1.7, 0.3, 0.8,-0.9,-0.8, 0.6, 2.6,-0.1, 0.6, 1.3,
0.8, 0.9, 0.2, 0.5, 1.1, 2.9,-0.3, 0.3, 2.1, 2.0, 1.1, 0.4, 3.4, 4.2, 2.7, 2.7,
1.6, 2.7, 2.6, 4.0, 6.0, 4.8, 5.1, 5.0, 3.2, 3.1, 2.8, 3.9, 4.5, 4.7, 3.9, 2.9,
2.2, 1.4)
```

El conjunto de residuales autocorrelacionados con  $\rho = 0.7$  está dado por:

```
set.seed(159)
nn <- length(x2)
sim.ar <- as.vector(arima.sim(list(ar=c(0.7)),n=nn))

y <- b0 + b1*x1 + b2*x2 + sim.ar
```

Entonces basta ver los intervalos de confianza para los estimadores de los modelos de *MCO* y *MCG*, así pues tomando los datos control para calcular los modelos y sus respectivos intervalos de confianza se tiene lo siguiente:

### *Cálculos para MCO*

```
mmco <- lm(y~x1+x2)
summary(mmco)
```

```

a <- as.vector(mmco$coef) # b0 = 4.419085    b1 = 0.953944    b2 = 0.802752
std.errs <- c(0.40995,0.07045,0.11057)
t.s <- rep(qt(0.975,46),3)

liminf <- a - (t.s * std.errs) # 3.5938982 0.8121356 0.5801861
limsup <- a + (t.s * std.errs) # 5.244271 1.095753 1.025318

#Amplitud de los intervalos para MCO 1.6503731 0.2836170 0.4451317

```

### *Cálculos para MCG*

```

fml <- gls( y ~ 1 + x1 + x2 , correlation =corAR1(0.7))
summary(fml)

ga <- as.vector(fml$coef) # b0 = 3.8947533    b1 = 1.0547616    b2 = 0.9150603
gstd.errs <- c(0.8192541,0.0818098,0.1351378)
gt.s <- rep(qt(0.975,46),3)

ga - (gt.s * gstd.errs) # 2.245680 0.890087 0.643042
ga + (gt.s * gstd.errs) # 5.543826 1.219436 1.187079

#Amplitud de los intervalos para MCG 3.2981459 0.3293492 0.5440366

```

Con este ejercicio se puede observar que los resultados de *MCO* presentan intervalos de confianza de menor amplitud que los intervalos de *MCG*, esto es, se subestima la variabilidad en los parámetros del modelo lo que puede generar un falso optimismo acerca de los resultados obtenidos y por otra parte, los pronósticos que se realicen se ven seriamente afectados ya que además del problema de la variabilidad, no sería tomado en cuenta el término de error.

Entonces, ¿Cómo se puede detectar el problema de autocorrelación oportunamente? Por lo general se aplica la prueba de contraste para residuales propuesta por *Durbin* y *Watson* en 1950, esta prueba está diseñada para detectar autocorrelación de primer orden.

**Estadístico *Durbin Watson* (DW):** La prueba de *Durbin-Watson* es la prueba más conocida para detectar la autocorrelación de primer orden. El estadístico  $d$  propuesto por *Durbin* y *Watson* en 1950 se basa en los residuos del modelo estimado por *MCO* y los siguientes supuestos:

- El modelo de regresión incluye el término  $\beta_0$
- Las variables explicativas no son estocásticas
- Los términos de error estocástico  $\varepsilon_j$  se generan a través de un esquema autorregresivo de primer orden, i.e.  $\varepsilon_j = \rho\varepsilon_{j-1} + v_j$ ;  $|\rho| < 1$
- El modelo de regresión no es autorregresivo, es decir no es del tipo:

$$Y = \beta_0 + \beta_1 Y_{j-1} + \varepsilon_j$$

- No faltan observaciones en los datos

Luego la hipótesis nula es:

$$\begin{aligned} H_0 : \rho &= 0 \\ H_1 : \varepsilon_j &= \rho\varepsilon_{j-1} + v_j \end{aligned}$$

donde  $v_j$  es ruido blanco, así pues el estadístico propuesto es:

$$d = \frac{\sum_{j=2}^n (e_j - e_{j-1})^2}{\sum_{j=1}^n e_j^2}$$

Note que:

$$\begin{aligned} d &= \frac{\sum_{j=2}^n (e_j - e_{j-1})^2}{\sum_{j=1}^n e_j^2} = \frac{\sum_{j=2}^n (e_j^2 - 2e_j e_{j-1} + e_{j-1}^2)}{\sum_{j=1}^n e_j^2} \\ &= \frac{\sum_{j=2}^n e_j^2}{\sum_{j=1}^n e_j^2} + \frac{\sum_{j=2}^n e_{j-1}^2}{\sum_{j=1}^n e_j^2} - 2 \frac{\sum_{j=2}^n e_j e_{j-1}}{\sum_{j=1}^n e_j^2} \end{aligned}$$

Observe que:

$$\frac{\sum_{j=2}^n e_j^2}{\sum_{j=1}^n e_j^2} \approx 1 \qquad \frac{\sum_{j=2}^n e_{j-1}^2}{\sum_{j=1}^n e_j^2} \approx 1$$

Además  $\frac{\sum_{j=2}^n e_j e_{j-1}}{\sum_{j=1}^n e_j^2}$  es una buena aproximación de la correlación entre  $e_j, e_{j-1}$ , que a su vez es una buena estimación para  $\rho$ , entonces se obtiene que:

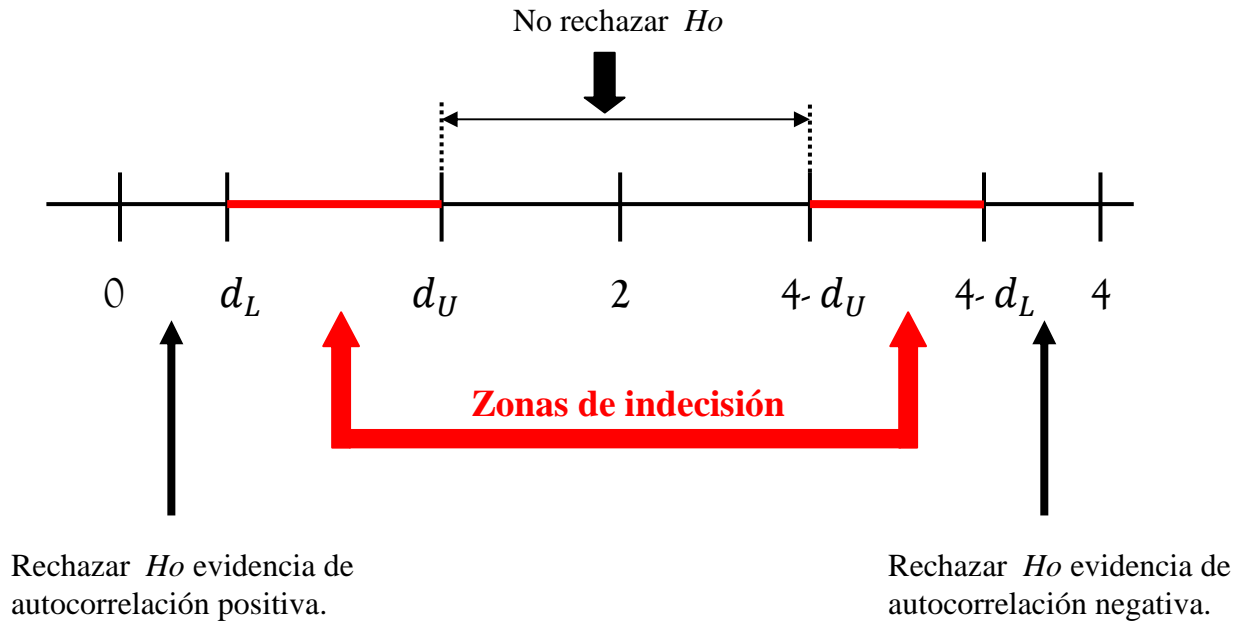
$$d \approx 2 - 2\hat{\rho} = 2(1 - \hat{\rho})$$

Recuerde que  $-1 \leq \rho \leq 1$ , con lo que se puede analizar los valores que tomará el estadístico de DW. Si  $\rho = -1$ , entonces existe autocorrelación negativa en los residuales y  $d \approx 4$ ; en el mismo sentido si  $\rho = 1$ , entonces existe autocorrelación positiva y  $d \approx 0$ , por lo tanto  $0 \leq d \leq 4$ , pero ¿cómo saber cuándo rechazar o no  $H_0$ ?

Si  $\rho = 0$  entonces  $d \approx 2$ , de lo que se intuye que cuando  $d$  es cercano a 2 no hay evidencia suficiente para rechazar  $H_0$  mientras que si  $d$  es cercano a 0 ó 4 hay evidencia de residuales autocorrelación ya sea de manera positiva o negativa, pero ¿hasta qué punto existe evidencia? para responder a ello existen tablas del estadístico de *Durbin y Watson*, que de forma aproximada, permiten establecer áreas de certeza alrededor de los valores límites  $d_L$  (*límite inferior*) y  $d_U$  (*límite superior*) los cuales ayudarán para realizar el contraste para rechazar o



no la hipótesis nula, donde cabe la posibilidad de caer en una zona de indecisión como se muestra en la figura siguiente:



Si existe evidencia de autocorrelación es necesario utilizar procedimientos alternos que permitan remediar el problema generado por los residuales autocorrelacionados, estos procedimientos serán detallados en el capítulo siguiente.

### 3. *MCG en los procedimientos de C-O y P-W*

El procedimiento de *MCG* en los procedimientos de *C-O* y *P-W*, parte de la siguiente estructura de datos:

$$X = \begin{bmatrix} 1 & X_{1,1} & \dots & X_{1,(k-1)} & X_{1,k} \\ 1 & X_{2,1} & \dots & X_{2,(k-1)} & X_{2,k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & X_{T-1,1} & \dots & X_{T-1,(k-1)} & X_{T-1,k} \\ 1 & X_{T,1} & \dots & X_{T,(k-1)} & X_{T,k} \end{bmatrix}_{T \times (k+1)}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T \times 1} ; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} ; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

Donde:  $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ .

Se quiere estimar el modelo:

$$Y = X\beta + \varepsilon$$

Los residuales del modelo se distribuyen  $\varepsilon \sim N(0, \Sigma)$  lo cual es precisamente el caso de *MCG* ya que  $\Sigma = \sigma^2\Omega$  donde  $\Omega$  es positiva definida pero no necesariamente la identidad, en este caso está dada por:

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{bmatrix}_{T \times T}$$

Además el estimador *MCG* para los parámetros del modelo es:

$$\tilde{\beta} = (X'\Omega^{-1}X)(X'\Omega^{-1}Y)$$

y la matriz  $\Omega^{-1}$  está dada por:

$$\Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \dots & \rho^{k-2} \\ 0 & -\rho & 1 + \rho^2 & \dots & \vdots \\ \vdots & 0 & -\rho & \ddots & \vdots \\ \vdots & \vdots & \vdots & 1 + \rho^2 & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}_{T \times T}$$

En este caso particular sea la matriz  $P$  tal que  $P'P = \Omega^{-1}$  la cual puede escribirse explícitamente como:

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & 0 & -\rho & \ddots & \vdots \\ \vdots & \vdots & \vdots & 1 & 0 \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}_{T \times T}$$

Cuando  $\rho$  es conocida la transformación del modelo puede ser determinada por:

$$PY = PX\beta + P\varepsilon$$

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

Por otro lado, al omitir la primera observación se tiene la transformación de *Cochrane-Orcutt*, cuya matriz es la siguiente:

$$\begin{bmatrix} -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & 0 & -\rho & \ddots & \vdots \\ \vdots & \vdots & \vdots & 1 & 0 \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}_{(T-1) \times T}$$

En otras palabras, se asume que los residuales de un modelo de regresión lineal son idénticamente pero no necesariamente independientemente distribuidos. Entonces el modelo queda de la siguiente manera:

$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t$	(1)
---	-----

donde los residuales son como en la *ecuación (2)*

$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, -1 < \rho < 1$	(2)
--	-----

Además  $v_t$  es independiente e idénticamente distribuida (*i.i.d.*), los residuales  $\varepsilon_t$  no lo son para  $\rho \neq 0$ . Este es el caso simple de autocorrelación, llamado primer orden de autocorrelación.

Si  $\rho$  es conocida, el método de diferencias generalizadas puede ser aplicado al modelo para transformarlo en uno donde los errores sean independientes. Para describir el procedimiento amigablemente, se utiliza el hecho de que el modelo en la *ecuación (1)* se mantiene para todos los periodos de tiempo de la siguiente manera:

$$Y_{t-1} = \beta_0 + \beta_1 X_{1,t-1} + \dots + \beta_k X_{k,t-1} + \varepsilon_{t-1} \quad (3)$$

Ahora multiplicando (3) por  $\rho$  y restando de (1), se obtiene un modelo con una estructura de errores como en (2):

$$\begin{aligned} Y_t - \rho Y_{t-1} &= (\beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t) - \rho(\beta_0 + \beta_1 X_{1,t-1} + \dots + \beta_k X_{k,t-1} + \varepsilon_{t-1}) \\ &= (\beta_0 - \rho\beta_0) + (\beta_1 X_{1,t} - \rho\beta_1 X_{1,t-1}) + \dots + (\beta_k X_{k,t} - \rho\beta_k X_{k,t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}) \\ &= \beta_0(1 - \rho) + \beta_1(X_{1,t} - \rho X_{1,t-1}) + \dots + \beta_k(X_{k,t} - \rho X_{k,t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}) \end{aligned}$$

Lo que resulta en la siguiente transformación:

$$Y_t^* = \beta_0^* + \beta_1^* X_{1,t}^* + \dots + \beta_k^* X_{k,t}^* + v_t \quad (4)$$

donde:

$$\begin{aligned} Y_t^* &= Y_t - \rho Y_{t-1} \\ \beta_0^* &= \beta_0(1 - \rho) \\ \beta_j^* &= \beta_j \quad \forall j \neq 1 \\ X_{j,t}^* &= X_{j,t} - \rho X_{j,t-1} \\ v_t &= \varepsilon_t - \rho\varepsilon_{t-1} \end{aligned}$$

Por construcción la transformación de la ecuación (4) tiene modificados los residuales que son *i.i.d.* resultando que los parámetros de los estimadores para MCO sean eficientes, entonces aplicando MCO a la ecuación (4) produce una estimación para MCG.

$$\tilde{Y}_t^* = \tilde{\beta}_0^* + \tilde{\beta}_1^* X_{1,t}^* + \dots + \tilde{\beta}_k^* X_{k,t}^* \quad (5)$$

(Nota: el símbolo de gorrito “^” indica estimación por MCO mientras que tilde “~” denota estimación por MCG.)

La generalización del procedimiento de diferencias es muy útil si se conoce a priori el valor de  $\rho$ . Usualmente no es el caso y la estimación de MCG está basada en una estimación de  $\rho$ , a menudo obtenida por los procedimientos de *Cochrane-Orcutt* o *Prais-Winsten* (en dos etapas ó iterativos). Bajo el supuesto de normalidad estos procedimientos proporcionan estimadores que convergen en probabilidad a un máximo estimador probable, el cual es consistente y asintóticamente eficiente.

Las ecuaciones anteriores son suficientes para mostrar los procedimientos desarrollados por C-O y P-W en dos etapas e iterativos.

Así pues el **procedimiento de Cochrane-Orcutt** es el siguiente:

**Paso 1.** Mínimos cuadrados ordinarios se usa para estimar la *ecuación (1)*.

**Paso 2.** Se calculan los residuales de la *ecuación (1)* que están dados por:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t$$

donde:  $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \dots + \hat{\beta}_k X_{k,t}$

Se define un nuevo conjunto de residuales  $\hat{\varepsilon}_{t-1}$  de tamaño T-1 aplicando un rezagando a los residuales  $\hat{\varepsilon}_t$  los cuales son utilizados para obtener  $\hat{\rho}$ , mediante la aplicación *MCO*.

Recuerde que:  $v_t = \varepsilon_t - \rho\varepsilon_{t-1} \Rightarrow \varepsilon_t = \rho\varepsilon_{t-1} + v_t$

además  $v_t$  son *i.i.d.* y además se utiliza *MCO* para estimar  $\rho$  a partir de los residuales  $\varepsilon_t$  y los residuales rezagados  $\varepsilon_{t-1}$ .

**Paso 3.** La *ecuación (4)* es estimada usando  $\hat{\rho}$  en lugar de  $\rho$ . La estimación se hace aplicando *MCO* en la *ecuación (4)* lo cual produce una estimación de *MCG*.

Las estimaciones del proceso hasta el **paso 3** es generalmente llamado el procedimiento *Cochrane-Orcutt* en dos etapas.

Los estimadores calculados en el **paso 3** son ahora estimaciones válidas de los parámetros para la *ecuación (1)*. La mayoría de los textos mencionan que las estimaciones para el intercepto del **paso 3** pueden ser ajustados dividiendo entre  $1 - \hat{\rho}$ , éste mismo ajuste, puede ser realizado para el error estándar del intercepto.

El procedimiento *Cochrane-Orcutt* de dos etapas provee solamente un estimador de  $\rho$ , mientras que el procedimiento iterativo estima  $\rho$  iterativamente, esto es, por aproximación sucesiva, tomando un valor inicial de  $\rho$  (dado en el *paso 2*) terminando con una condición de paro (que será dada en el *paso 5*), calculando  $\rho$  tantas veces como iteraciones se calculen en el proceso. Así pues para aplicar el procedimiento iterativo *Cochrane-Orcutt* se debe continuar en el **paso 4** dado a continuación.

**Paso 4.** Las estimaciones de  $\tilde{\beta}$  obtenidas en el **paso 3** son utilizadas para obtener un nuevo conjunto de residuales. Hay que ser cautelosos ya que este es un momento donde la intuición miente. Los residuales pueden ser calculados usando:

$\tilde{\varepsilon}_t = Y_t - \tilde{Y}_t$	(6)
---	-----

donde  $\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,t} + \dots + \tilde{\beta}_k X_{k,t}$

$$\tilde{\beta}_0 = \frac{\tilde{\beta}_0^*}{1-\hat{\rho}}$$

$$\tilde{\beta}_j = \tilde{\beta}_j^*$$

Comúnmente se comete el error de calcular los residuales de la forma  $\tilde{Y}_t^* - \tilde{Y}_t$  a partir de la ecuación (4). De hecho estos pasos no son explícitamente indicados en cualquier libro de texto.

**Paso 5.** Se repite el *paso 2*. La recomendación general es parar cuando las estimaciones para  $\rho$  sean salidas que difieran por una cantidad pequeña, por decir menor que 0.0001.

### **Fin del procedimiento de Cochrane-Orcutt.**

Cuando se cumple la condición de paro del *paso 5* se terminan las iteraciones y se obtiene un modelo en donde los estimadores son eficientes y además son una estimación de MCG.

Cuando se opta por usar el procedimiento *Cochrane-Orcutt* de “dos etapas” o iterativo, se pueden obtener diferencias en los resultados, esto se debe a la estimación de  $\rho$  ya que en el procedimiento iterativo se parte de la misma estimación que en “dos etapas” y puede terminar en un valor significativamente diferente, donde lo que impactará en la transformación, de ahí las diferencias.

Por otro lado se tienen diferencias con respecto al procedimiento de *Prais-Winsten* debidas a la pérdida de la primera observación, lo que no siempre es tomado muy en cuenta, pero tiene relevancia para casos donde se cuenta con pocos datos ya que para muestras pequeñas, está bien documentado que conforme la primer observación está presente o se omite puede obtenerse una diferencia substancial en los resultados de la regresión. Más aún se marcan esas diferencias si la primera observación difiere bastante de la segunda ó si se tiene la sospecha que ésta es un punto extremo, mientras que la pérdida de una observación en muestras grandes tiende a ser intrascendente.

De lo anterior se puede intuir que el procedimiento de *Cochrane-Orcutt* debe funcionar bien para regresiones donde se cuenta con una muestra no pequeña, aún, si es el caso en que se cuenta con pocos datos, se tiene la ventaja de poder comparar los resultados ofrecidos por este procedimiento con una mejora del mismo ofrecida por *Prais-Winsten* la cual preserva la primera observación.

Nótese que la matriz de transformación de *P-W* es de tamaño  $T \times T$  mientras la de *C-O* es de tamaño  $T-1 \times T$ . Desde las matrices se nota que la de *P-W* preserva todas las observaciones.

Entonces a continuación se detallará el **procedimiento de Prais-Winsten** el cual es:

**Paso 1.** Al igual que en *C-O* el método de mínimos cuadrados ordinarios se usa para estimar la ecuación (1).

**Paso 2.** Se calculan los residuales de la ecuación (1) son  $\hat{\varepsilon}_t = Y_t - \hat{Y}_t$   
donde:  $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \dots + \hat{\beta}_k X_{k,t}$

Se define un nuevo conjunto de residuales  $\hat{\varepsilon}_{t-1}$  de tamaño T-1 aplicando un rezago a los residuales  $\hat{\varepsilon}_t$  los cuales son utilizados para obtener  $\hat{\rho}$ , mediante la aplicación MCO, tal como en el **paso 2** del procedimiento de C-O.

**Paso 3.** La ecuación (4) es estimada usando  $\hat{\rho}$  en lugar de  $\rho$ . La estimación se hace aplicando MCO en la ecuación (4) tomando como datos el resultado de la transformación de Prais-Winsten, la cual es:

$$Y_1^* = Y_1 \sqrt{1 - \hat{\rho}^2} \quad ; \quad Y_t^* = Y_t - \rho Y_{t-1} \quad \forall t > 1$$

$$X_{j,1}^* = X_{j,1} \sqrt{1 - \hat{\rho}^2} \quad ; \quad X_{j,t}^* = X_{j,t} - \rho X_{j,t-1} \quad \forall t > 1$$

Esto produce una estimación del modelo para MCG.

Las estimaciones del proceso hasta el **paso 3** es generalmente llamado el procedimiento de Prais-Winsten en dos etapas.

Para continuar con el procedimiento de Prais-Winsten iterativo, los pasos a seguir son los mismos **paso 4** y **paso 5** del procedimiento de Cochrane-Orcutt (C-O).

### **Fin del procedimiento de Prais-Winsten.**

Es importante mencionar que la estimación de la correlación  $\hat{\rho}$  para los residuales que se planteó en el **paso 2** en ambos procedimientos C-O y P-W no es la única disponible en la literatura por ejemplo; el estadístico de Durbin Watson (DW) es un estimador conveniente, este estadístico puede ser aproximado por:

$$DW = 2(1 - \rho) \Rightarrow \hat{\rho}_{DW} = 1 - \frac{DW}{2}$$

Así pues existen otras alternativas para obtener estimadores de la correlación de los residuales  $\rho$  tales como Theil-Nagar (1961) entre otros.

Lo mencionado hasta este momento tiene una gran relevancia e importancia a la hora de la práctica, ya que al trabajar con modelos de regresión con residuales autocorrelacionados se debe tomar en cuenta que se debe optar por mejor modelo posible para las predicciones sean confiables y se debe elegir entre:

- (a) Método de dos etapas y método iterativo.
- (b) Procedimientos de estimación Cochrane-Orcutt o Prais-Winsten.
- (c) Varios estimadores de  $\rho$  usado en la estimación del modelo MCG.

Algo que ayuda a tomar la decisión al elegir un modelo es tomar en cuenta una serie de estadísticos útiles tales como los ajustes de los modelos, la significancia de los estimadores de los parámetros del modelo, así como el estadístico de *Durbin-Watson* para descartar la persistencia de la autocorrelación y la validez del modelo y su inferencia.

Continúa la pregunta ¿Cuál metodología es la mejor? Se resumen las siguientes generalidades como sigue (ver Greene, 2003 p. 274):

Para muestras pequeñas con  $\rho < 0.3$ , actualmente se prefiere usar *MCO* en lugar de *MCG* esto evita la variación adicional del muestreo causado por el estimado  $\hat{\rho}$ . Es preferible emplear el método iterativo sobre el método en dos etapas incluso si la marginal consigue ser mínima. La aproximación de *Prais-Winsten* en muchos casos es mejor a pesar de la simplicidad lograda por la pérdida de la primera observación en el procedimiento de *Cochrane-Orcutt*, y por su parte el cálculo para estimar  $\hat{\rho}$  a partir de la regresión de los residuales contra ellos mismos con un rezago (como en el **paso 2** de los procedimientos vistos) generalmente trabaja bien.

Lo que siempre se desea en el caso de procedimientos iterativos es que el punto de partida no padeciera del problema de convergencia local, en cuyo caso no necesariamente tenga convergencia al mínimo global. Si se piensa en el caso de la regresión por *MCO* en base al procedimiento de *C-O*, se está pensando en un valor inicial por defecto de  $\hat{\rho} = 0$ . Éste valor ésta secuencialmente actualizado en cada una de las iteraciones del procedimiento. Es aconsejable (más no indispensable) calcular los procedimientos iterativos con varios valores iniciales diferentes de  $\hat{\rho}$ , para asegurar que la convergencia se dé en algún valor posible de  $\hat{\rho}$ . Por otro lado si los estimadores tienen un comportamiento oscilatorio, se recomienda usar el método de búsqueda propuesto por Hildreth-Lu (vea Greene 2003). La idea es hacer una malla para  $\rho$  con pasos pequeños entre (-1,1) y buscar el valor que minimice la suma de cuadrados de los residuales correspondiente a la ecuación transformada para la estimación de *MCG*.

Por otro lado sobre los pronósticos que se pueden obtener del modelo hay que recordar que si la autocorrelación es ignorada, la inferencia sobre el modelo no es válida por lo cual las estimaciones de pronósticos no resultarían adecuadas ya que podrían ser ineficientes, esto es con grandes varianzas. Para poder obtener pronósticos eficientes, los estimadores de los parámetros del modelo deben ser eficientes y a partir de ese momento se estará en condiciones de proponer una estructura adecuada para los pronósticos, al emplear los procedimientos de *Cochrane-Orcutt* o *Prais-Winsten* pueden resultar estimadores que son en ambos casos objetivos y eficientes.

Los pronósticos resultan ser eficientes desde que se logra explicar explícitamente la autocorrelación de los residuales del modelo, de manera que en la *ecuación (5)* sustituyendo  $T + 1$  en  $t$  se tiene lo siguiente:

$\tilde{Y}_{T+1}^* = \beta_0^* + \tilde{\beta}_1^* X_{1,T+1}^* + \dots + \tilde{\beta}_k^* X_{k,T+1}^* \quad (7)$
---



$T$  representa el número de observaciones en la muestra, además:

$$\tilde{Y}_{T+1}^* = \tilde{Y}_{T+1} - \tilde{\rho}Y_T \quad \text{y} \quad X_{j,T+1}^* = X_{j,T+1}^o - \tilde{\rho}X_{j,T}$$

El valor de  $X_{j,T+1}^o$  es en algunos casos conocido (o estimado), éste es el valor de  $X_j$  para el periodo  $T + 1$ .

Sustituyendo y resolviendo para  $\tilde{Y}_{T+1}$  se tiene la estructura del pronóstico para un periodo  $T + 1$ :

$\tilde{Y}_{T+1} = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T+1}^o + \dots + \tilde{\beta}_k X_{k,T+1}^o + \tilde{\rho}\tilde{\epsilon}_T \tag{8}$
--

donde:

$$\tilde{\epsilon}_T = Y_T - \tilde{Y}_T$$

$$\tilde{\epsilon}_T = Y_T - (\tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T} + \dots + \tilde{\beta}_k X_{k,T})$$

Siguiendo esta lógica, los pronósticos para  $Y_{T+s}$  son obtenidos como:

$\tilde{Y}_{T+s} = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T+1}^o + \dots + \tilde{\beta}_k X_{k,T+1}^o + \tilde{\rho}^s \tilde{\epsilon}_T \tag{9}$
---

En el cálculo de pronósticos a menudo se ignora el factor de corrección  $\tilde{\rho}^s \tilde{\epsilon}_T$ , lo cual sesga en gran medida los pronósticos, resultando estos inadecuados. Un pronóstico puede ser calculado usando la *ecuación* (8) como se verá en las aplicaciones, donde se detallará de una manera clara el uso de estas ecuaciones.

Es verdad que la mayoría de los paquetes estadísticos tienen implementados los procedimientos iterativos *Cochrane-Orcutt* o *Prais-Winsten*, que además son calculados por medio de comandos simples, lo que facilita su aplicación al evitar los tediosos procesos iterativos que suelen confundir a los estudiantes. Sin embargo en algunos paquetes es reportada una  $R^2$  que no es comparable con los correspondientes estimados para *MCO* (“modelo original”). Una manera apropiada de calcular  $R^2$  y que sea comparable con los métodos alternativos es la siguiente:

$$R^2 = \left( \text{correlación}(Y_t, \tilde{Y}_t) \right)^2$$

$Y_t$  representa el valor original de la variable dependiente,

$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,t} + \dots + \tilde{\beta}_k X_{k,t}$  es el modelo resultante después de aplicar cualquiera de los procedimientos antes vistos.

Algunos paquetes erróneamente calculan  $R^2$  como  $R^2 = \left( \text{correlación}(Y_t^*, \tilde{Y}_t^*) \right)^2$

donde  $Y_t^* = Y_t - \tilde{\rho}Y_t$  es el valor transformado de  $Y_t$ .

En las aplicaciones se mostrará a detalle el uso de cada una de las ecuaciones que se utilizan para realizar los pronósticos con la finalidad de despejar cualquier duda al respecto y con lo cual quedarán cubiertos los objetivos que se plantearon.

Para realizar los cálculos se utilizará el software estadístico **R**, este cuenta con la función “`gls()`” que calcula el modelo regresión por *MCG*, pero se necesita una aproximación de la matriz de varianzas-covarianzas, que es precisamente el dato con que no se cuenta, por lo que se recomienda la aplicación de *MCG* mediante los procedimientos vistos anteriormente, además para facilitar la aplicación de los procedimientos se desarrolló una función en **R** que realiza los cálculos siguiendo cada paso para *C-O* y otra para *P-W* en ambos casos para procedimientos en *dos etapas* e *iterativo* (*Ver código anexo*). De igual manera se generaron gráficas para visualizar los residuales de los modelos calculados en términos de las variables originales así como del efecto que se obtiene al aplicarles la transformación, haciendo posible una comparación directa que ayude a madurar los conceptos vistos.

Por otro lado si en la práctica se cuenta con una cantidad importante de variables y se desea considerar solo aquellas que sean significativas para el modelo pero se tienen problemas de autocorrelación en los residuales, lo primero es considerarlas todas para estimar el modelo por *MCO*, luego aplicar *MCG* en los procedimientos de *C-O* y *P-W* para solucionar los problemas que acarrear los residuales autocorrelacionados y después de eso se pueden aplicar las pruebas necesarias para seleccionar las variables.

Con la finalidad de ilustrar la utilidad y la manera correcta de la aplicación de *MCG* utilizando los procedimientos *C-O* y *P-W* en dos etapas e iterativos se mostrarán dos aplicaciones, una para un modelo de regresión múltiple con presencia de autocorrelación de primer orden y en la segunda aplicación con un modelo de regresión lineal simple con presencia de autocorrelación de orden mayor, con la intención de establecer las limitaciones de los procedimientos.

## 4. Aplicaciones

### Aplicación 1

En esta aplicación se cuenta con un modelo en dos variables, ahora se analizarán los modelos resultantes de aplicar MCG por los procedimientos antes vistos, para la relación entre la tasa de inflación, el déficit y la tasa de interés a corto plazo. El modelo que se estimará es el siguiente:  $rate_t = \beta_0 + \beta_1 inflation_t + \beta_2 deficit_t + \varepsilon_t$

Donde *rate* es la tasa de interés trimestral anualizada, *inflation* es la tasa de inflación anual basada en el índice de precios al consumidor (CPI) y *deficit* es el déficit del presupuesto federal como un porcentaje de GDP. Es común en la literatura encontrar el uso de tasas de interés trimestrales como una aproximación para las tasas de interés de corto plazo (vea Hoelscher, 1983, Makin, 1983, and Mascaro and Meltzer, 1983). Datos de 1948 a 1996 se obtuvieron de un reporte económico de 2004 y son presentados en la Tabla 1.1.

Tabla 1.1 *Indicadores económicos para modelo de la aplicación 1*

<b>Año</b>	<b>Rate</b>	<b>inflation</b>	<b>Déficit</b>	<b>Año</b>	<b>rate</b>	<b>Inflation</b>	<b>Déficit</b>
1948	1.04	8.1	-4.6	1973	7.041	6.2	1.1
1949	1.102	-1.2	-0.2	1974	7.886	11	0.4
1950	1.218	1.3	1.1	1975	5.838	9.1	3.4
1951	1.552	7.9	-1.9	1976	4.989	5.8	4.2
1952	1.766	1.9	0.4	1977	5.265	6.5	2.7
1953	1.931	0.8	1.7	1978	7.221	7.6	2.7
1954	0.953	0.7	0.3	1979	10.041	11.3	1.6
1955	1.753	-0.4	0.8	1980	11.506	13.5	2.7
1956	2.658	1.5	-0.9	1981	14.029	10.3	2.6
1957	3.267	3.3	-0.8	1982	10.686	6.2	4
1958	1.839	2.8	0.6	1983	8.63	3.2	6
1959	3.405	0.7	2.6	1984	9.58	4.3	4.8
1960	2.928	1.7	-0.1	1985	7.48	3.6	5.1
1961	2.378	1	0.6	1986	5.98	1.9	5
1962	2.778	1	1.3	1987	5.82	3.6	3.2
1963	3.157	1.3	0.8	1988	6.69	4.1	3.1
1964	3.549	1.3	0.9	1989	8.12	4.8	2.8
1965	3.954	1.6	0.2	1990	7.51	5.4	3.9
1966	4.881	2.9	0.5	1991	5.42	4.2	4.5
1967	4.321	3.1	1.1	1992	3.45	3	4.7
1968	5.339	4.2	2.9	1993	3.02	3	3.9
1969	6.677	5.5	-0.3	1994	4.29	2.6	2.9
1970	6.458	5.7	0.3	1995	5.51	2.8	2.2
1971	4.348	4.4	2.1	1996	5.02	3	1.4
1972	4.071	3.2	2	<u>1997</u>	<u>5.07</u>	<u>2.3</u>	<u>0.3</u>

Con fines ilustrativos para del análisis se removerá la última observación de la serie de datos de la *Tabla 1.1* lo cual será útil para comparar los resultados obtenidos en los pronósticos que se realizarán con los modelos obtenidos de los procedimientos de *C-O* y *P-W* en dos etapas e iterativos.

Así pues el modelo que se estimará es el siguiente:

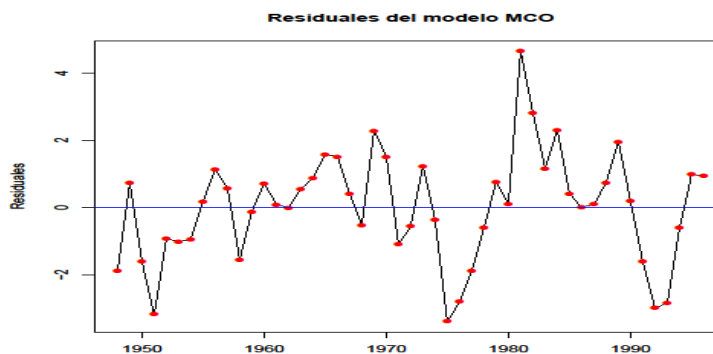
$$rate = \beta_0 + \beta_1 inflation + \beta_2 deficit$$

los estimadores de *MCO* para el modelo son:

$$rate = -1.2517 + 0.6104 inflation + 0.7103 deficit$$

Para el modelo los estimadores de los coeficientes resultan altamente significativos al igual que el intercepto, en donde el coeficiente de determinación  $R^2 = 0.6995$ , indica que el 70% de la variabilidad en la tasa de interés (*rate*) es explicada por la variabilidad de las variables independientes (*inflation* y *deficit*). Ahora es necesario analizar los residuales, ya que si esto es omitido se corre el riesgo de los estimadores del modelo no sean eficientes, por lo que no resultaría adecuada la inferencia. Si es necesario se procederá a la aplicación de ambos métodos en dos etapas e iterativos y se compararán los estadísticos de cada uno de los modelos obtenidos para tomar una decisión sobre la selección del “mejor” modelo para realizar los pronósticos.

Observe los residuales obtenidos por *MCO* antes de aplicar *Durbin Watson*:

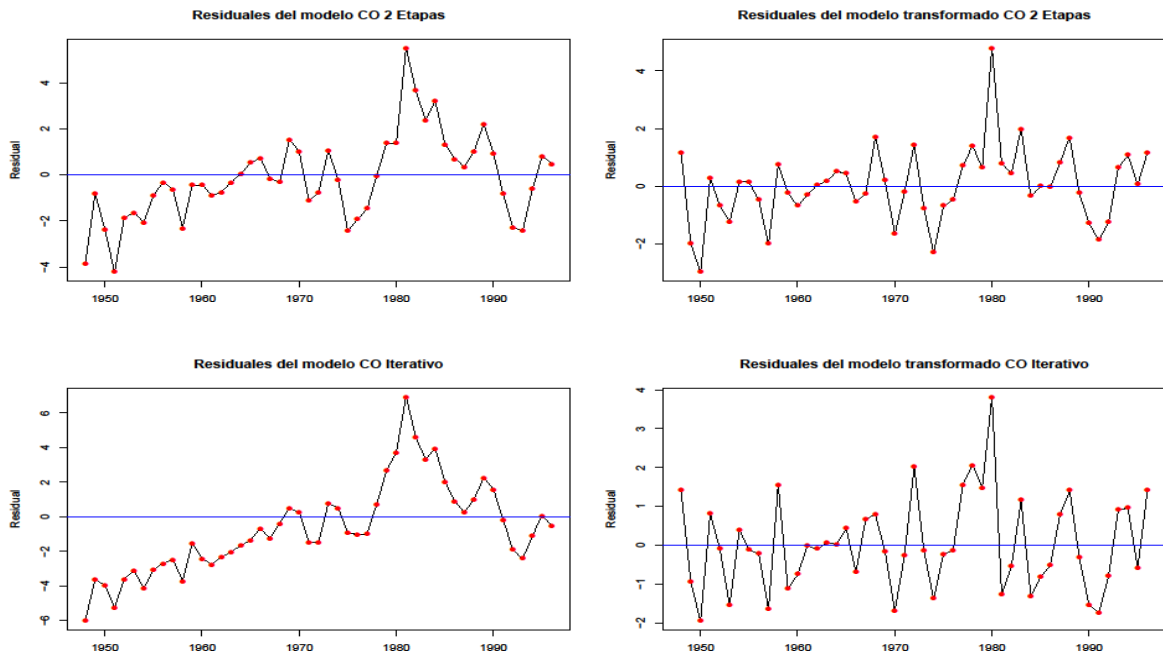


Gráfica 1.1

De la gráfica es claro que los residuales no son ruido blanco, por el contrario existe una relación temporal entre ellos, una cierta tendencia y algunos ciclos, esto es un indicio de que los residuales están correlacionados, lo cual se puede corroborar calculando el estadístico de *Durbin-Watson*,  $DW = 0.927$ ;  $p\_value = 7.333e - 06$  lo cual comprueba la existencia de autocorrelación positiva. Por lo cual es prudente continuar el análisis usando los procedimientos de *C-O* y *P-W*. La *Tabla 1.2* muestra un resumen de los resultados obtenidos

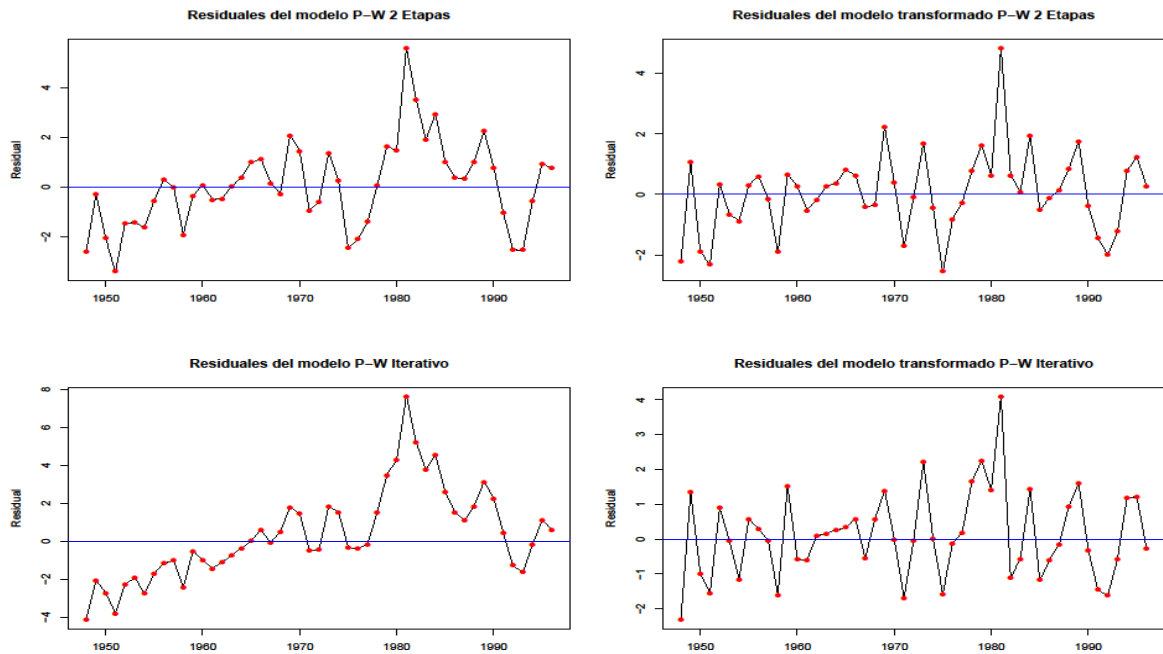
por los procedimientos en *dos etapas e iterativos*, pero antes de comentar algo acerca de los resultados.

Observe las siguientes gráficas de residuales.



Grafica 1.2 (*Residuales obtenidos por C-O*)

Note que los residuos aparentemente se comportan igual en todos los modelos.



Grafica 1.3 (*Residuales obtenidos por P-W*)

Al igual que en los modelos obtenidos por los procedimientos de *C-O* los residuales obtenidos por los procedimientos de *P-W* no son i.i.d. como se puede apreciar en los gráficos.

Así pues se obtiene la siguiente tabla de resultados.

Tabla 1.2 *Resultados de la aplicación 1*

Resultados de la Regresión sobre los efectos de <i>Inflation</i> y <i>Deficits</i> sobre las tasas de interés ( <i>Rates</i> )					
Variable dependiente: The Three-month T-bill interest rate, annualized (1948-1996)					
	MCO	C-O 2-Etapas	C-O Iterativo	P-W 2-Etapas	P-W Iterativo
<b>Intercepto</b>	1.2517	2.5775***	4.9846***	2.0803**	3.5011**
<b>Error-st</b>		0.6701	1.3150	0.6456	1.1339
<b>Inflation</b>	0.6104	0.4883***	0.22018*	0.4859***	0.2568*
<b>Error-st</b>		0.0856	0.0965	0.0885	0.0996
<b>Deficit</b>	0.7103	0.3572*	-0.0605	0.51601***	0.09197
<b>Error-st</b>		0.1564	0.1707	0.1403	0.1660
<b>Ajuste <math>R^2</math></b>	0.6995	0.6739	0.3441	0.6984	0.5948
<b>Durbin-Watson</b>	0.9273***	1.4745	1.771	1.5722	1.7885
$\hat{\rho}$	n/a	0.5188	0.8403	0.5188	0.8275
$\tilde{Y}_{T+1}$	n/a	4.06	5.02	3.75	4.64
Nivel de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '					

Al aplicar *MCO* se obtuvo que los residuales son autocorrelacionados, con un valor del estadístico de  $DW=0.9273$  el cual es significativo para  $d_L = 1.38$  y  $d_U = 1.54$  e indica que hay evidencia suficiente para rechazar  $H_0 : \rho = 0$ , por lo que los estimadores dejan de ser eficientes y la inferencia sobre el modelo no es adecuada, en la **Tabla 1.2** se concentran los resultados de los cálculos para los procesos de *C-O* y *P-W* en *dos etapas* e *iterativos*.

Para esta aplicación, los interceptos tienen valores más grandes en los procedimientos iterativos que en dos etapas, esta diferencia también está presente para los coeficientes del modelo, pero estos tienen menor impacto en el modelo al aplicar los procedimientos iterativos, de hecho existe un cambio de signo en la aplicación del *procedimiento iterativo* de *C-O* comparado con *dos etapas*, ya que la variable *Deficit* pasa de  $0.3572$  a  $-0.0605$ , por otro lado en *P-W iterativo* ésta misma variable no cambia de signo, pero varía al pasar de  $0.51601$  en *dos etapas* a  $0.09197$  en el procedimiento *iterativo*.

El ajuste  $R^2$  del modelo también disminuye después de la aplicación de los procesos iterativos, la menor pérdida se presenta para  $P-W$ .

También se encontró que hay poca diferencia en los estimadores de la correlación  $\tilde{\rho}$  para los procedimientos iterativos de  $C-O$  y  $P-W$ , pero si hay una diferencia significativa con respecto a los procedimientos en “*dos etapas*”, esta diferencia debido a las iteraciones ya que en ambos casos se parte de la misma estimación, aun que hay autores que minimizan la necesidad de iterar, por ejemplo en el *Greene 2003*(p. 273) se menciona que “*desde que el estimador es asintóticamente eficiente en cada iteración, nada se obtiene por hacer eso*”. Pero como se mencionó y se pudo observar en esta aplicación, los resultados que pueden diferir bastante después de iterar y obtener cambios en la estimación de  $\rho$ , ésto es, entre los procedimientos en dos etapas y los iterativos, además mejora a la estimación de los pronósticos al iterar. Como ilustración se presenta la *Tabla 1.3* en la que se muestran los cambios en la estimación de  $\rho$  para cada iteración para los procedimientos de  $C-O$  y  $P-W$ .

Tabla 1.3 *Valores de la correlación en cada iteración (Aplicación 1)*

<b>Iteración</b>	<b>C-O <math>\hat{\rho}</math></b>	<b>P-W <math>\hat{\rho}</math></b>
1	0.51879	0.51879
2	0.66415	0.62776
3	0.76016	0.69794
4	0.81204	0.74864
5	0.83174	0.78311
6	0.83782	0.80407
7	0.83958	0.81566
8	0.84008	0.82167
9	0.84022	0.82468
10	0.84026	0.82617
11	0.84027	0.82690
12	0.84027	0.82725
13	0.84027	0.82742
14	0.84027	0.82750
15	0.84027	0.82754

Para esta aplicación  $C-O$  converge más rápido que  $P-W$ , esto no quiere decir que es mejor. La finalidad de mostrar esta tabla es que se pueda observar que las estimaciones de  $\tilde{\rho}$  no oscilan, por lo que no es necesario utilizar otro tipo de búsquedas para estimar  $\rho$ . Note que en ambas columnas de la *Tabla 1.3* la diferencia es notable entre la estimación de la primera iteración y la última.

Según los resultados obtenidos en la **Tabla 1.2**, el modelo que se seleccionaría para realizar inferencia y realizar los pronósticos para el año 1997, sería el obtenido por el procedimiento iterativo de *P-W* ya que tiene en mejor ajuste en  $R^2$  y no existe evidencia suficiente para rechazar la hipótesis nula de que  $\rho = 0$ , esto es que no persiste el problema de la autocorrelación.

A partir de la selección del modelo se pueden realizar los pronósticos ya que los estimadores son eficientes, así pues usando la *ecuación* (8) y los valores de la inflación y el déficit para el año 1997 (recuerde que estos valores pueden ser conocidos ó estimados), es decir usando el modelo obtenido después de utilizar el procedimiento iterativo de *Prais-Winsten*, el pronóstico para el tiempo T+1 es calculado como sigue:

Note que la ecuación (8) está dada por:

$$\tilde{Y}_{T+1} = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T+1}^o + \dots + \tilde{\beta}_k X_{k,T+1}^o + \tilde{\rho} \tilde{\varepsilon}_T \quad (8)$$

recuerde que:  $\tilde{\varepsilon}_T = Y_T - (\tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T} + \dots + \tilde{\beta}_k X_{k,T})$

para este caso k=2 y el modelo obtenido por el procedimiento iterativo de *P-W* es:

$$rate_{T+1} = 3.5011 + 0.2568 \text{ infl}_{T+1} + 0.09197 \text{ deficit}_{T+1} + \tilde{\rho} \tilde{\varepsilon}_T$$

donde  $\tilde{\varepsilon}_T = \tilde{\varepsilon}_{1996} = 5.02 - (3.5011 + 0.2568 (3) + 0.09197(1.4)) = 0.6197$

entonces

$$\begin{aligned} rate_{1997} &= 3.5011 + (0.2568 * \text{infl}_{1997}) + (0.09197 * \text{deficit}_{1997}) + (0.8275 * \tilde{\varepsilon}_{1996}) \\ &= 3.5011 + (0.2568 * 2.3) + (0.09197 * 0.3) + (0.8275 * 0.6197) = 4.64 \end{aligned}$$

Solo como práctica usando el modelo obtenido por el procedimiento iterativo de *Cochrane-Orcutt* se tiene lo siguiente:

$$\begin{aligned} rate_{1997} &= 4.9846 + (0.220 * \text{infl}_{1997}) - (0.0605 * \text{deficit}_{1997}) + (0.8403 * \tilde{\varepsilon}_{1996}) \\ &= 4.9846 + (0.220 * 2.3) - (0.0605 * 0.3) + (0.8403 * -0.546) = \mathbf{5.02} \end{aligned}$$

Si se omitiera el término de error que es el que explícitamente explica la autocorrelación de los errores, se podrían obtener estimaciones o pronósticos para *C-O* y *P-W*, de 5.48 y 4.12 respectivamente, de aquí la importancia de no pasar por alto el término de ajuste del error.



## Aplicación 2

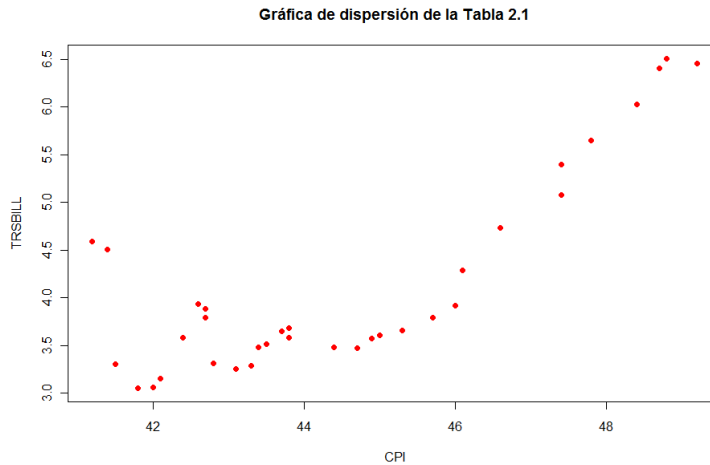
La segunda aplicación presenta un modelo que involucra la tasa de interés a corto plazo TRSBILL como variable dependiente y el índice de precios al consumidor CPI como variable explicatoria, ambas medidas de manera mensual entre enero de 1971 y diciembre de 1973 en Canadá. Los datos se muestran en la siguiente tabla.

Tabla 2.1 *Datos para la aplicación 2*

Año	Mes	TRSBILL	CPI	Año	Mes	TRSBILL	CPI	
1971	Ene	4.59	41.2	1972	Jul	3.48	44.4	
	Feb	4.51	41.4		Ago	3.47	44.7	
	Mar	3.30	41.5		Sep	3.57	44.9	
	Abr	3.05	41.8		Oct	3.57	44.9	
	May	3.06	42.0		Nov	3.61	45.0	
	Jun	3.15	42.1		Dic	3.66	45.3	
	Jul	3.58	42.4		1973	Ene	3.79	45.7
	Ago	3.88	42.7			Feb	3.92	46.0
	Sep	3.93	42.6			Mar	4.29	46.1
	Oct	3.79	42.7			Abr	4.73	46.6
	Nov	3.31	42.8			May	5.08	47.4
	Dic	3.25	43.1			Jun	5.40	47.4
1972	Ene	3.29	43.7	Jul		5.65	47.8	
	Feb	3.48	43.4	Ago		6.03	48.4	
	Mar	3.51	43.5	Sep		6.41	48.7	
	Abr	3.65	43.7	Oct		6.51	48.8	
	May	3.68	43.8	Nov		6.46	49.2	
	Jun	3.58	43.8	Dic		6.38	49.5	

Con fines ilustrativos dentro del análisis se removerá la última observación la cual será útil para comparar los resultados obtenidos en los pronósticos que se realizarán con los modelos obtenidos de los procedimientos de *C-O* y *P-W* en dos etapas e iterativos.

Así pues observe la gráfica de dispersión de los datos, donde se puede observar que el modelo que se ajuste por *MCO* no será del todo bueno, esto debido a naturaleza de los datos, pero *MCO* es el punto de partida para el análisis.

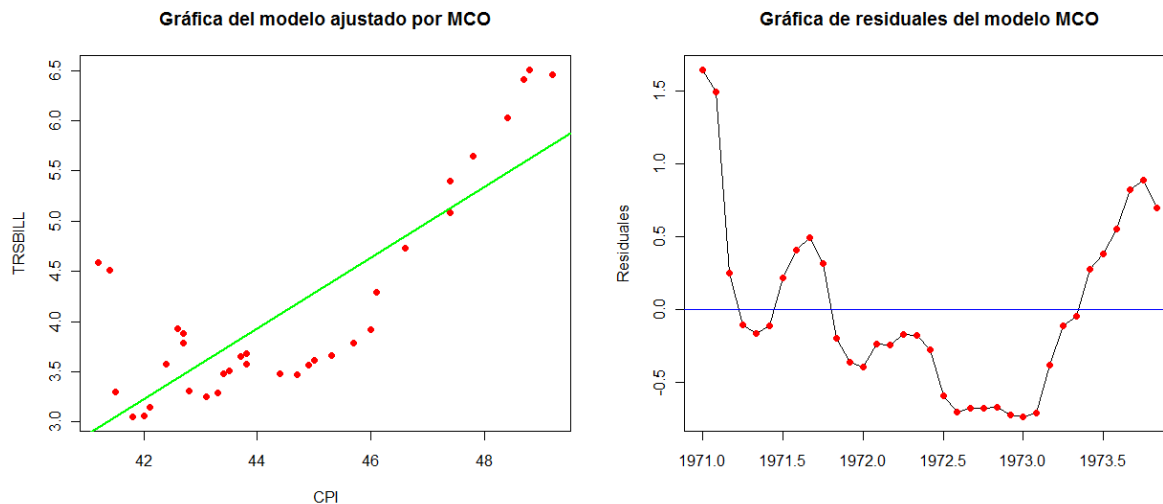


*Gráfica 2.1*

El modelo de *MCO* resulta ser:

$$TRSBILL = -11.5724 + 0.3523 \text{ CPI}$$

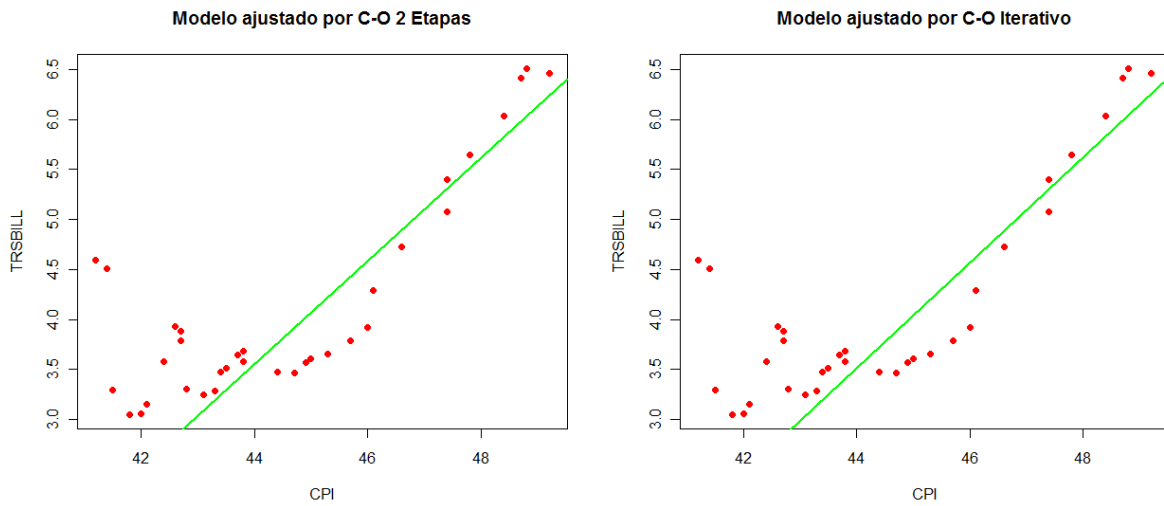
Con una  $R^2 = 0.6466$ , en la siguiente gráfica se pueden observar los datos con el modelo ajustado por *MCO* y sus residuales:



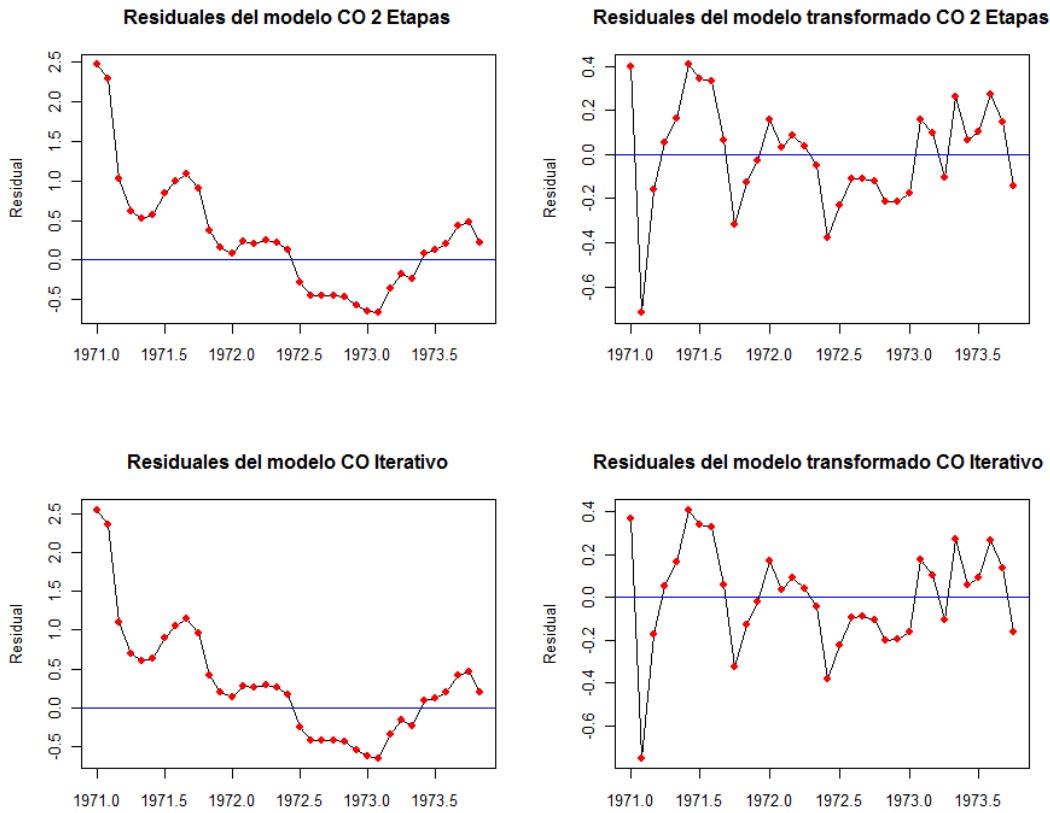
*Gráfica 2.2*

De la gráfica de residuales (Gráfica 2.2) se puede observar que los residuales no son ruido blanco (no son *i.i.d.*), por el contrario parece existir una relación entre ellos, por lo que se puede sospechar que los residuales están correlacionados, luego para dispersar la duda se debe calcular el estadístico de *Durbin-Watson*, el cual para este modelo resulta ser  $DW = 0.2199$ ,  $p\text{-value} = 3.074e-15$ , lo que indica una fuerte evidencia de correlación positiva, por lo tanto existen las condiciones necesarias para aplicar *MCG* con los procedimientos de *C-O* y *P-W*.

A continuación se muestran las gráficas de los datos con los modelos ajustados por el procedimiento de C-O en dos etapas e iterativo, además de sus respectivos residuales.



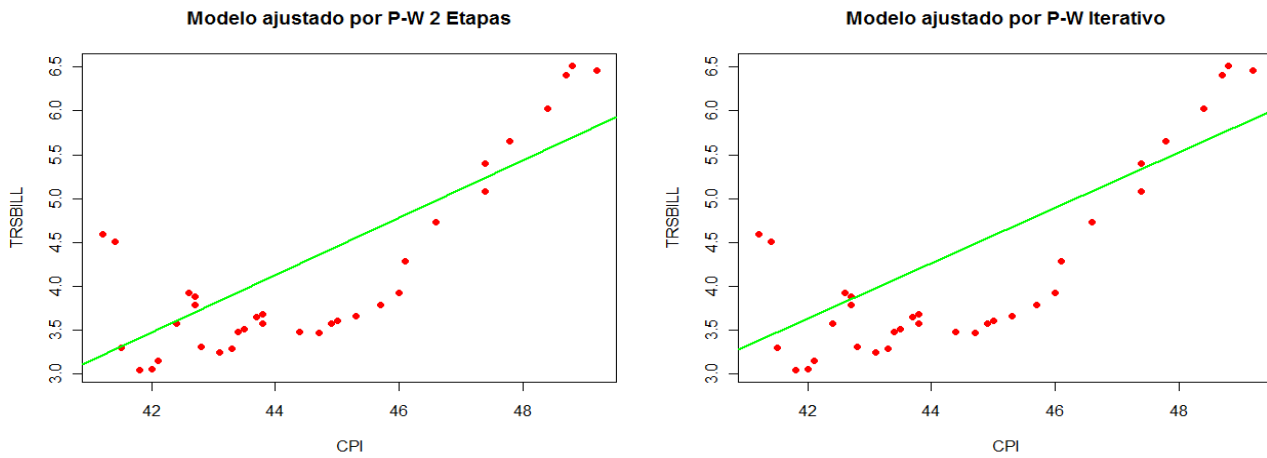
Gráfica 2.3 (Modelos ajustados por C-O)



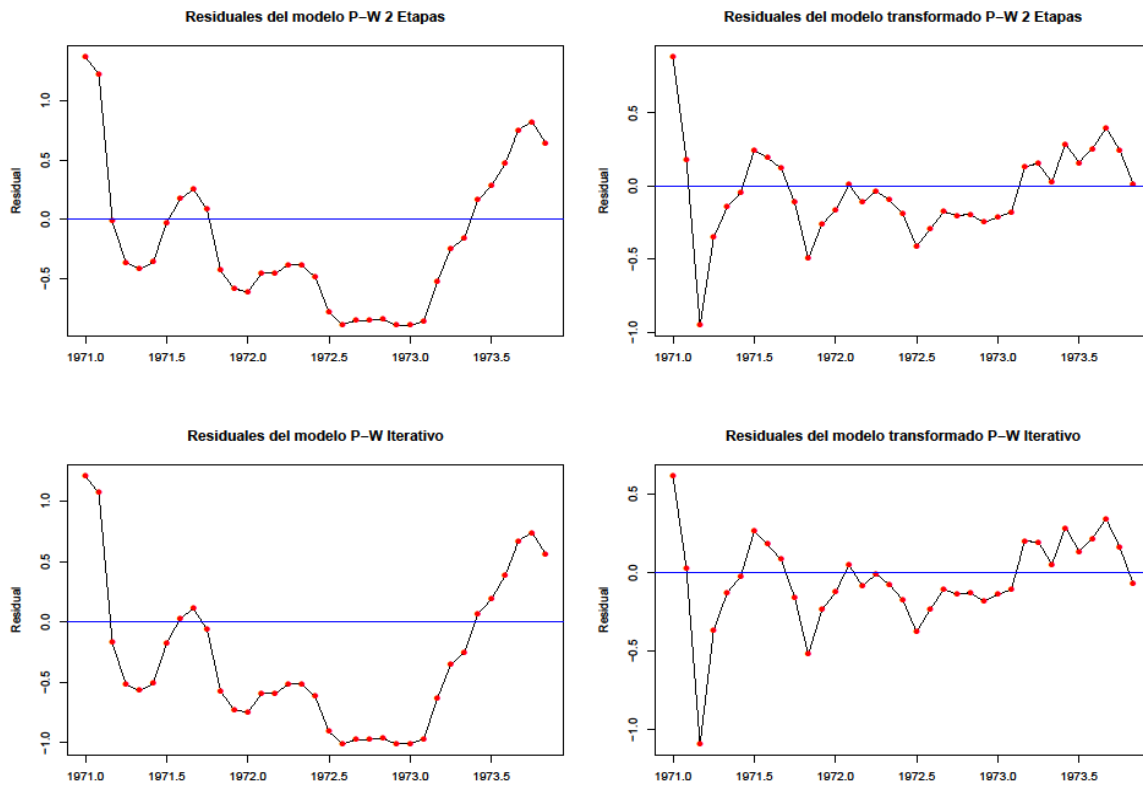
Gráfica 2.4 (Residuales obtenidos por C-O)

Como se puede observar no existe una diferencia marcada entre el ajuste del modelo en dos etapas y el iterativo lo que se refleja en los residuales. Para la transformación de *C-O* los residuales se muestran como material de apoyo en sentido exploratorio para observar cambio que sufren los residuales con dicha transformación.

Por otro lado las gráficas de los modelos ajustados por *P-W* se muestran a continuación:



Gráfica 2.5 (Modelos ajustados por *P-W*)



Gráfica 2.6 (Residuales obtenidos por *P-W*)

No hay que perder de vista los detalles como los que se observan en la *Gráfica 2.5*, donde se puede apreciar una ligera diferencia en los ajustes, tal vez causadas por el valor del intercepto ya que en el procedimiento iterativo de *P-W* parece ser mayor que para dos etapas, este efecto es menos perceptible en las gráficas para el procedimiento de *C-O*. Por otro lado, también hay que observar la información de los residuales, porque se debe tener en mente el cálculo de los pronósticos en base al modelo seleccionado y estos tienen que ser lo mejor posible. De aquí la necesidad de aplicar nuevamente pruebas de autocorrelación, para descartar la presencia del problema.

En las gráficas es posible ver las diferencias de los modelos obtenidos por *MCO* y los procedimientos de *C-O* y *P-W*. En este caso particular la primera observación parece jugar un papel importante en el modelo resultante; esto se puede observar en las *gráficas (2.3, 2.3 y 2.5)* donde la inclinación de los modelos ajustados cambia, aún cuando se calcularon inicialmente con la misma  $\rho$ , aquí es donde se aprecia que procedimientos similares pueden producir resultados diferentes.

En cuanto a los residuales, son importantes las *gráficas (2.4 y 2.6)* que contienen los residuales de los ajustes en las unidades originales de las variables y para los valores transformados con lo que se puede comparar la mejora que ofrecen los procedimientos en dos etapas e iterativos. Estas gráficas son de ayuda para observar si el problema no persiste en los residuales de los modelos obtenidos por los procedimientos de *C-O* y *P-W* además ver el comportamiento que se tiene al incluir la descripción de la autocorrelación en el modelo.

Recuerde que los ajustes de los modelos obtenidos por *C-O* y *P-W* no son directamente comparables con el de *MCO*, en ese sentido se debe tener cuidado con el cálculo del ajuste y seguir la recomendación que se describió en el capítulo anterior. Por otro lado la estimación de los pronósticos obtenidos de la estimación de *MCG* en los procedimientos vistos contemplan el término de error descrito en la *ecuación (8)*. Una vez que se realizaron los cálculos se obtuvo la tabla de resultados siguiente:

Tabla 2.2 Resultados de la aplicación 2

Resultados de la Regresión sobre los efectos del IPC sobre las tasas de interés (TRSBILL)					
Variable Dependiente: T-bill interest rate (TRSBILL), month (Jan1971 - Dic 1973 in Canada)					
	MCO	C-O 2-Etapas	C-O Iterativo	P-W 2-Etapas	P-W Iterativo
Intercepto	-11.5724	-19.1054***	-19.6691***	-10.1894**	-9.61767*
Error-st		3.11282	3.34917	3.27736	4.16177
IPC	0.3523	0.5151***	0.5269***	0.3254***	0.31538**
Error-st		0.06845	0.07349	0.07326	0.09282
Ajuste $R^2$	0.6466	0.646609	0.646609	0.646609	0.6466
$R^2$ Modelo Transformado	n/a	0.9527	0.9449	0.9323	0.8709
Durbin-Watson	0.2199 ***	1.3913	1.4254	0.9022***	1.0376
$\hat{\rho}$	n/a	0.7633	0.7838	0.7633	0.8597
$\bar{Y}_{T+1}$	n/a	6.4080	6.4162	6.3086	6.3813
Nivel de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ''					

Al aplicar *MCO* se obtuvo que los residuales son autocorrelacionados, con un valor del estadístico de  $D-W=0.2199$  el cual es significativo, lo que implica que los estimadores del modelo de *MCO* no son eficientes, esto es no son de varianza mínima, por lo cual las pruebas de hipótesis no son adecuadas al igual que la inferencia. La **Tabla 2.2** se tienen los resultados de los cálculos para los procesos de *C-O* y *P-W* en 2 etapas e iterativos.

Observe el renglón de interceptos, los que son calculados para *C-O* tanto en dos etapas como en el iterativo son claramente diferentes a los obtenidos por *P-W*. Esto es debido a la pérdida de la primera observación ya que ésta se genera de acuerdo al peso del valor en el modelo. Esta diferencia también está presente en el coeficiente del IPC, que en ambos procedimientos tiene un efecto positivo sobre TRSBILL, pero es más fuerte en *C-O*.

El ajuste comparable con *MCO* es el mismo para todos los modelos, si fuera necesario seleccionar un modelo usado solamente este estadístico, no sería claro por cual decidirse, en el mismo sentido se cuenta con la información presentada por el estadístico de  $D-W$  en este caso con límites dados por  $d_L = 1.30$  y  $d_U = 1.42$ , lo que deja en zona de indecisión al modelo *C-O* en dos etapas, mientras que para *C-O* iterativo no hay evidencia suficiente para rechazar  $H_0 : \rho = 0$ , lo que no es tan malo. Por otro lado los estadísticos de  $D-W$  para los modelos obtenidos por *P-W* hay evidencia suficiente para rechazar la hipótesis nula, esto es muestran presencia de autocorrelación positiva, entonces esto ayuda a tomar una decisión, ya que al menos un modelo resulta adecuado según la prueba de *Durbin Watson*. Este es un caso

raro, donde se debe sospechar de un grado mayor de autocorrelación, esto será aclarado más adelante.

El renglón del ajuste  $R^2$  del modelo transformado es el ajuste obtenido de la estimación de *MCG*, el cual no es comparable con lo obtenido del modelo original de *MCO*. La intención de incluirlo es mostrar el efecto que tiene la transformación sobre los modelos resultantes.

En algunos casos en la práctica, si se encuentran complicaciones como las encontradas en este ejemplo, es posible que el modelo tenga un grado mayor de autocorrelación en los residuales, lo que se puede corregir agregando una variable exógena al modelo, aunque hay que ser conscientes que en la práctica no siempre es posible agregarla por no contar con la información, en cuyo caso se debe explorar la existencia de autocorrelación de orden mayor.

En este momento si se tuviera que seleccionar uno de los modelos obtenidos los de *P-W* estarían descartados por presentar evidencia de presencia de autocorrelación. Por otro lado, *C-O iterativo* no tiene evidencia suficiente para rechazar  $H_0$  en la prueba de Durbin Watson, por lo que sería el modelo seleccionado para realizar el pronóstico.

Así pues, utilizando la *ecuación (8)*, considerando el modelo obtenido por el procedimiento iterativo de *C-O* se tiene lo siguiente:

Note que la *ecuación (8)* está dada por:

$\tilde{Y}_{T+1} = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T+1}^o + \dots + \tilde{\beta}_k X_{k,T+1}^o + \tilde{\rho}\tilde{\epsilon}_T \quad (8)$
--

recuerde que:  $\tilde{\epsilon}_T = Y_T - (\tilde{\beta}_0 + \tilde{\beta}_1 X_{1,T} + \dots + \tilde{\beta}_k X_{k,T})$

Para este caso en particular  $k = 1$  entonces para el modelo obtenido por el procedimiento iterativo de *C-O* se tiene:

$$TRSBILL_{T+1} = -19.6691 + 0.5269 CPI_{T+1}^o + \tilde{\rho}\tilde{\epsilon}_T$$

donde:  $T = 35$ ;  $CPI_{T+1}^o = 49.5$   $\tilde{\rho} = 0.7838$

$$\tilde{\epsilon}_T = TRSBILL_T - (-19.6691 + 0.5269 CPI_T)$$

$$= 6.46 - (-19.6691 + 0.5269(49.2)) = 6.46 - 6.25 = 0.21$$

$$\therefore TRSBILL_{T+1} = -19.6691 + 0.5269 CPI_{T+1}^o + \tilde{\rho}\tilde{\epsilon}_T$$

$$= -19.6691 + 0.5269 (49.5) + (0.7838)(0.21)$$

$$= 6.25737 + 0.158829 = 6.416199$$

$$\therefore TRSBILL_{T+1} = 6.416199$$

En este punto del análisis podría caerse en un exceso de optimismo ya que el pronóstico parece bueno, pero, ¿En realidad lo es?. En la práctica no se tiene la posibilidad comparar el pronóstico, hasta que se conoce el dato real y comienza nuevamente la necesidad de calcular un nuevo pronóstico para el periodo siguiente. La confianza en un modelo debe estar respaldada las pruebas que se le aplican para saber que es un buen modelo y a partir de ese momento construir los pronósticos. En particular para el caso del ejemplo, no son satisfactorios los resultados de la prueba de *Durbin-Watson* para los residuales en los modelos obtenidos por los procedimientos de *C-O* y *P-W*. Por lo cual el pronóstico no es adecuado, pues hay que ser conscientes de los resultados obtenidos en la *Tabla 2.2* los cuales muestran indicios de problemas en los residuales, lo que quiere decir que no pueden ser modelados por un *AR(1)*, lo que viola el supuesto del esquema autorregresivo de primer orden para los residuales en la prueba de *D-W* y surge la necesidad de aplicar otros métodos y procedimientos que resuelvan el problema.

Por lo cual, si en la práctica se presenta la misma situación, es decir, los procedimientos vistos son aplicados y no es resuelto el problema en los residuales, es necesario explorar otras técnicas, tomando como punto de partida la prueba de *Breush-Godfrey* la cual es útil para descartar autocorrelación de orden mayor y otro tipo de estructuras como un *MA(q)* o modelos mixtos para los residuales. Esto marca las limitaciones de la estimación de *MCG* en los procedimientos de *C-O* y *P-W* con lo que se pretende orientar al lector en que momento será necesario dejar de lado los pronósticos ya que en estas condiciones no serán adecuados y será necesario comenzar un análisis con la prueba de *Breush-Godfrey*.



## 5. Conclusiones

Como ya se sabe ignorar la presencia de autocorrelación puede acarrear problemas graves para calcular intervalos de confianza y las pruebas hipótesis usuales  $t$  y  $F$  pueden dejar de ser válidas, por lo que pueden conducir a conclusiones erróneas. En estos casos debe utilizarse  $MCG$  en lugar de  $MCO$ , aun cuando los estimadores obtenidos por  $MCO$ , sean insesgados y consistentes ya que no son eficientes (no son de varianza mínima), para lo que son muy útiles los procedimientos de  $C-O$  y  $P-W$  cuando  $\Omega$  es desconocida.

En este documento se vieron estadísticos que son de ayuda para seleccionar un modelo en el que los estimadores sean eficientes, estos los podemos resumir en los siguientes puntos:

- Los procedimientos iterativos resultan en mejores pronósticos que los de dos etapas, esto por la búsqueda del mejor estimador para  $\rho$ . En este sentido puede existir una diferencia amplia la cual impacta en los cálculos y la inferencia que se puede realizar con los modelos resultantes.
- Si bien los modelos obtienen una  $R^2$  que no es comparable con la obtenida por  $MCO$  es importante para ver el efecto que tiene la transformación sobre el modelo de  $MCG$ , pero además de eso es importante calcular adecuadamente el ajuste para hacerlo comparable con  $MCO$ .
- El estadístico de *Durbin Watson* es importante tanto para  $MCO$  como para los modelos obtenidos por los procedimientos vistos, esto para descartar la persistencia de autocorrelación. Si persiste la autocorrelación es indicio de que existe un orden mayor de autocorrelación para  $MCO$ .
- Una vez que se seleccionó un modelo se debe tener cuidado de no omitir el término de error en los cálculos de los pronósticos, tal como pudo verse en este trabajo ya que de nada serviría tener un buen modelo si este término es omitido.

Es importante contar con gráficas de los residuales, en estas se pueden detectar tendencias o ciclos que son los principales indicios de que los residuales son autocorrelacionados para lo cual se utilizan las pruebas desarrolladas para tal fin.

Por último si no se logra obtener un buen modelo según los cálculos que se vieron (como en el *Ejemplo 2*), debe analizarse si los datos presentan problemas de autocorrelación de orden mayor, aunque las pruebas para detectar ese problema pueden realizarse desde el principio del análisis.

## BIBLIOGRAFIA

Greene, William. 2003. *Econometric Analysis* (fifth edition).  
Upper Saddle River, New Jersey: Prentice Hall.

Alan Pankratz, 1991, *Forecasting with dynamic regression models*  
John Wiley and Sons, New York

Johnston, J., and John DiNardo, 1997. *Econometric Methods*, 4th ed.  
New York: McGraw Hill,

Jeffrey M. Wooldridge, 2005, *Introductory Econometrics a Modern Approach*  
Hardcover

Theil, H. and Nagar, A.L., 1961, *Testing the Independence of Regression Disturbances*  
J. Amer. Statist. Assoc.

Sanjiv Jaggia, 2008, *Practical Considerations when Estimating in the Presence of Autocorrelation*  
<http://www.bentley.edu/csbiggs/documents/Jaggia.pdf>

Ángel Alejandro Juan Pérez, *Autocorrelación*  
<http://www.uoc.edu/in3/emath/docs/Autocorrelacion.pdf>

Ramón Mahía, 2010, *Conceptos básicos sobre la autocorrelación en el modelo básico de regresión lineal*  
[http://www.uam.es/personal\\_pdi/economicas/rmc/econometria/pdf/HPAutocorr\\_2010.pdf](http://www.uam.es/personal_pdi/economicas/rmc/econometria/pdf/HPAutocorr_2010.pdf)

## **ANEXOS**



## ANEXO A: Glosario de definiciones

**Autocorrelación:** La autocorrelación surge cuando los términos de error del modelo no son independientes entre sí, es decir, cuando:

$$E(v_i, v_j) \neq 0 \quad \forall i \neq j; v_t = \varepsilon_t - \rho\varepsilon_{t-1} \text{ donde } \varepsilon_t \text{ es } i.i.d.$$

Entonces los errores estarán vinculados entre sí. Los estimadores mínimos cuadrados ordinarios (MCO) obtenidos bajo esta circunstancia, dejan de ser eficientes.

**Convergencia en probabilidad:** La variable aleatoria  $x_n$  converge en probabilidad a una constante  $c$  si  $\lim_{n \rightarrow \infty} P(|x_n - c| > \varepsilon) = 0$  para cualquier  $\varepsilon > 0$ . Esta definición indica que se hace cada vez más improbable que  $x_n$  tome valores distintos de  $c$ , a medida que  $n$  (tamaño de la muestra) aumenta.

**Estimadores:**

- **Consistente:** El estadístico  $T$  es un "estimador consistente" del parámetro  $\theta$  si y sólo si para cualquier constante positiva  $c$  se cumple que  $\lim_{n \rightarrow \infty} P(|T - \theta| \geq c) = 0$  donde  $n$  es el tamaño de muestra.
- **Eficientes:** Un estimador es más **eficiente** o más **preciso** que otro estimador, si la varianza del primero es menor que la del segundo. Por ejemplo, si  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son ambos estimadores de  $\theta$  y  $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ .
- **Insesgado o Imparcial:** Un estadístico  $T$  es un estimador insesgado del parámetro  $q$  si y solo si  $E(T) = q$ .
- **Asintóticamente eficiente:** Se dice que un estimador consistente  $\hat{\theta}$  es asintóticamente eficiente si para cualquier otro estimador  $\theta^*$ , la varianza de la  $\hat{\theta}$  es menor que la varianza de  $\theta^*$ .

**Variable exógena:** Las variables exógenas están determinadas fuera del modelo. El valor de las variables exógenas está determinado por factores o variables que no son incluidas en el modelo que se está utilizando y se toman como datos para analizar el comportamiento de las variables endógenas.

**Variable aleatoria ó estocástica:** Se denomina variable estocástica a una cantidad cuyo valor se determina como resultado de un experimento. A tal fin se define una variable aleatoria (estocástica)  $X$  como una función cuyo dominio es un espacio de probabilidad  $S$  y sus valores  $X(s)$  pertenecen a los reales  $\mathbb{R}$ . Simbólicamente se escribe  $X: S \rightarrow \mathbb{R}$ .



## ANEXO B: Código de cálculos y gráfica

### Código en R

```
library(zoo);library(car);library(lmtest)

# Procedimiento de Cochrane Orcutt

it.cochrane.orcutt <- function(mod,k=1)
{
  rhoo <- c(0); m <- 1
  cc <- length(coef(mod))
  X <- model.matrix(mod); xo <- X
  y <- model.response(model.frame(mod)); yo <- y
  e <- residuals(mod)
  n <- length(e)
  names <- colnames(X)
  # Comienza el proceso iterativo #
  while (m < k+1)
  {
    rho <- sum(e[1:(n-1)]*e[2:n])/sum(e^2)
    rhoo[m] <- rho

    y <- yo[2:n] - rho * yo[1:(n-1)]
    X <- xo[2:n,] - rho * xo[1:(n-1),]

    model <- lm(y ~ X - 1)
    b <- coef(model)
    y_e <- b%*%t(xo)
    e <- yo - y_e
    m <- m+1
  }
  # Termina el proceso iterativo #
  # presentación de resultados #
  R2iCO <- cor(yo ,t(y_e))^2

  dw <- dwtest(model)

  result <- list()
  result$coefficients <- coef(model)
  result$residu <- t(e)
  names(result$coefficients) <- names
  summary <- summary(model, corr = F)
  result$cov <- (summary$sigma^2) * summary$cov.unscaled
  dimnames(result$cov) <- list(names, names)

  result$eiCO <- residuals(model)
  result$sigma <- summary$sigma
  result$rho <- rho
  result$suma <- summary(model)
  result$R_cuadr <- R2iCO
  result$DW <- dw
  result$rho.iter <- t(rhoo)
  result$dif.rhok <- abs(rhoo[k]-rhoo[k-1])
  class(result) <- 'it.cochrane.orcutt'
  result
}

# Procedimiento de Prais-Winsten

it.prais.winsten <- function(mod1,k=1)
{
  rhoo <- c(0); m <- 1
  cc <- length(coef(mod1))
  X <- model.matrix(mod1) ; xo <- X
  y <- model.response(model.frame(mod1)); yo <- y
```

```

e <- residuals(mod1)
n <- length(e)
names <- colnames(X)

while (m < k+1)
{
rho <- sum(e[1:(n-1)]*e[2:n])/sum(e^2)
rhoo[m] <- rho

y <- c(yo[1] * (1 - rho^2)^0.5, yo[2:n] - rho * yo[1:(n-1)])
X <- rbind(xo[1,] * (1 - rho^2)^0.5, xo[2:n,] - rho * xo[1:(n-1),])

model <- lm(y ~ X - 1)
b <- coef(model)
y_e <- b%*%t(xo)
e <- yo - y_e
m <- m+1
}

R2iPW <- cor(yo ,t(y_e))^2

dw <- dwtest(model)

result <- list()
result$coefficients <- coef(model)
result$residu <- t(e)
names(result$coefficients) <- names
summary <- summary(model, corr = F)
result$cov <- (summary$sigma^2) * summary$cov.unscaled
dimnames(result$cov) <- list(names, names)

result$eiPW <- residuals(model)
result$sigma <- summary$sigma
result$rho <- rho
result$suma <- summary(model)
result$R_cuadr <- R2iPW
result$DW <- dw
result$rho.iter <- t(rhoo)
result$dif.rhok <- abs(rhoo[k]-rhoo[k-1])
class(result) <- 'prais.winsten'
result
}

```

## Gráficas y cálculos:

### *Aplicación 1*

```

dats <- read.table('D:/DATOS_txs/Dts_Tna_P2.txt', header = T)
attach(dats)
Rate[50]
mco2 <- lm( Rate[-50] ~ Inflac[-50] + Deficit[-50] )
summary(mco2)
e2 <- ts(residuals(mco2),start = 1948, end = 1996, frequency = 1)
n2 <- length(residuals(mco2))

windows()
ts.plot(e2,main = 'Residuales del modelo MCO',ylab='Residuales', xlab = ' ')
points(e2, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')
dw_mco1 <- dwtest(mco2)

```

### *Gráficas y cálculos para CO Ejemplo 1*

```

(C_O <- it.cochrane.orcutt(mco2))

```



```

(C_O$dif.rhok < 0.0001)
  C_O$coefficients
  C_O$rho
  C_O$R_cuadr
  C_O$suma

windows()
par(mfrow=c(2,2))
eCO <- ts(C_O$residu,start = 1948, end = 1996, frequency = 1)
ts.plot(eCO,main = 'Residuales del modelo CO 2 Etapas',xlab=' ',ylab='Residual')
points(eCO, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

eCOT <- ts(C_O$eiCO,start = 1948, end = 1996, frequency = 1)
ts.plot(eCOT,main = 'Residuales del modelo transformado CO 2 Etapas',ylab='Residual',xlab=' ')
points(eCOT, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

  eTco <- C_O$residu[n2]
  (y_Tmlco <- sum(C_O$coefficients * c(1,Inflac[50], Deficit[50])) + (C_O$rho * eTco))

(iC_O <- it.cochrane.orcutt(mco2,15))
(iC_O$dif.rhok < 0.0001)

  iC_O$coefficients
  iC_O$rho
  iC_O$R_cuadr
  iC_O$suma

windows()
par(mfrow=c(1,2))

eCOi <- ts(iC_O$residu,start = 1948, end = 1996, frequency = 1)
ts.plot(eCOi, main = 'Residuales del modelo CO Iterativo',xlab=' ',ylab='Residual')
points(eCOi, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

eCOit <- ts(iC_O$eiCO,start = 1948, end = 1996, frequency = 1)
ts.plot(eCOit,main = 'Residuales del modelo transformado CO Iterativo',xlab=' ',ylab='Residual')
points(eCOit, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

  eTcoi <- iC_O$residu[n2]
  (y_Tmlcoi <- sum(iC_O$coefficients * c(1,Inflac[50], Deficit[50])) + (iC_O$rho * eTcoi))

```

### *Gráficas y cálculos para PW Ejemplo 1*

```

(P_W <- it.prais.winsten(mco2))
(P_W$dif.rhok < 0.0001)
  P_W$coefficients
  P_W$rho
  P_W$R_cuadr
  P_W$suma

windows()
par(mfrow=c(2,2))

ePW <- ts(P_W$residu,start = 1948, end = 1996, frequency = 1)
ts.plot(ePW,main = 'Residuales del modelo P-W 2 Etapas',xlab=' ',ylab='Residual')
points(ePW, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

ePWt <- ts(P_W$eiPW,start = 1948, end = 1996, frequency = 1)
ts.plot(ePWt,main = 'Residuales del modelo transformado P-W 2 Etapas',xlab=' ',ylab='Residual')
points(ePWt, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

```

```

eTpw <- P_W$residu[n2]
(y_Tmlpw <- sum(P_W$coefficients * c(1,Inflac[50], Deficit[50])) + (P_W$rho * eTpw))

(iP_W <- it.prais.winsten(mco2,15))
(iP_W$dif.rhok < 0.0001)
iP_W$coefficients
iP_W$rho
iP_W$R_cuadr
iP_W$suma

windows()
par(mfrow=c(1,2))

ePWi <- ts(iP_W$residu,start = 1948, end = 1996, frequency = 1)
ts.plot(ePWi,main = 'Residuales del modelo P-W Iterativo',xlab=' ',ylab='Residual')
points(ePWi, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

ePWit <- ts(iP_W$eiPW,start = 1948, end = 1996, frequency = 1)
ts.plot(ePWit,main = 'Residuales del modelo transformado P-W Iterativo',xlab=' ',ylab='Residual')
points(ePWit, pch=16, col = 'red')
abline(a=0, b=0, col = 4)

eTpwi <- P_W$residu[n2]
(y_Tmlpwi <- sum(iP_W$coefficients * c(1,Inflac[50], Deficit[50])) + (iP_W$rho * eTpwi))

rc <- iC_O$rho.iter
rp <- iP_W$rho.iter

matrix(cbind(rc,rp),15,2)

```

## Aplicación 2

```

datos <- read.table('D:/DATOS_txs/Dts_Tna_Pl.txt', header = T)
attach(datos)
windows()
plot(CPI[-36],TRSBILL[-36], main = 'Gráfica de dispersión de la Tabla 2.1',ylab = 'TRSBILL',
xlab = 'CPI',col='red',pch=16)
(mco <- lm( TRSBILL[-36] ~ CPI[-36])) ; summary(mco)
summary(mco)

windows()
par(mfrow=c(1,2))
plot(CPI[-36],TRSBILL[-36], main = 'Gráfica del modelo ajustado por MCO', ylab = 'TRSBILL',
xlab = 'CPI',pch= 16,col='red')
abline(coef = coef(mco),col='green', lwd=2)
n <- length(residuals(mco))
re <- ts(residuals(mco), frequency=12,start=c(1971,1))
ts.plot(re,main = 'Gráfica de residuales del modelo MCO', ylab = 'Residuales', xlab = '')
points(re, pch= 16, col = 'red')
abline(a=0, b=0, col = 'blue')

dw02 <- dwtest(mco)

```

## Gráficas y cálculos para CO Ejemplo 2

```

(CO1 <- it.cochrane.orcutt(mco))
CO1$coefficients
CO1$rho
CO1$R_cuadr
CO1$suma

```

```

eTco <- CO1$residu[n]
(yTmlco <- sum(CO1$coefficients * c(1,CPI[35])) + (CO1$rho * eTco))

windows()
plot(CPI[-36],TRSBILL[-36], main ='Modelo ajustado por C-0 2 Etapas', ylab = 'TRSBILL', xlab
= 'CPI',pch= 16,col='red')
abline(coef = CO1$coefficients,col='green', lwd=2)

windows()
par(mfrow=c(2,2))
co <- ts(CO1$residu, frequency=12,start=c(1971,1))
ts.plot(co,main ='Residuales del modelo CO 2 Etapas',xlab=' ',ylab='Residual')
points(co, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')

tco <- ts(CO1$eiCO, frequency=12,start=c(1971,1))
ts.plot(tco,main ='Residuales del modelo transformado CO 2 Etapas',xlab=' ',ylab='Residual')
points(tco, pch= 16,col='red')
abline(a=0, b=0, col = 'blue')

(iCO1 <- it.cochrane.orcutt(mco,15))
(iCO1$dif.rhok < 0.0001)
iCO1$coefficients
iCO1$rho
iCO1$R_cuadr
iCO1$suma

ieTco <- iCO1$residu[n]
(iyTmlco <- sum(iCO1$coefficients * c(1,CPI[35])) + (iCO1$rho * ieTco))

windows()
plot(CPI[-36],TRSBILL[-36], main ='Modelo ajustado por C-0 Iterativo',ylab = 'TRSBILL', xlab
= 'CPI',pch= 16,col='red')
abline(coef = iCO1$coefficients,col='green', lwd=2)

windows()
par(mfrow=c(1,2))
coi <- ts(iCO1$residu, frequency=12,start=c(1971,1))
ts.plot(coi,main ='Residuales del modelo CO Iterativo',xlab=' ',ylab='Residual')
points(coi, pch=16, col = 'red')
abline(a=0,b=0, col = 'blue')

tcoi <- ts(iCO1$eiCO, frequency=12,start=c(1971,1))
ts.plot(tcoi,main ='Residuales del modelo transformado CO Iterativo ',xlab='
',ylab='Residual')
points(tcoi, pch=16, col = 'red')
abline(a=0, b=0,col = 'blue')

```

## Gráficas y cálculos para PW Ejemplo 2

```

(P_W1 <- it.prais.winsten(mco))
P_W1$coefficients
P_W1$rho
P_W1$R_cuadr
P_W1$suma

eTpw <- P_W1$residu[n]
(yTmlpw <- sum(P_W1$coefficients * c(1,CPI[35])) + (P_W1$rho * eTpw))

windows()
par(mfrow=c(1,2))
plot(CPI[-36],TRSBILL[-36], main ='Modelo ajustado por P-W 2 Etapas', ylab = 'TRSBILL', xlab
= 'CPI',pch= 16,col='red')
abline(coef = P_W1$coefficients,pch= 16,col='green',lwd=2)

```

```

windows()
par(mfrow=c(2,2))
pw <- ts(P_W1$residu, frequency=12,start=c(1971,1))
ts.plot(pw,main = 'Residuales del modelo P-W 2 Etapas',xlab=' ',ylab='Residual')
points(pw, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')

pwt <- ts(P_W1$eiPW, frequency=12,start=c(1971,1))
ts.plot(pwt,main = 'Residuales del modelo transformado P-W 2 Etapas',xlab=' ',ylab='Residual')
points(pwt, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')

(iP_W1 <- it.prais.winsten(mco,15))
(iP_W1$dif.rhok < 0.0001)

    iP_W1$coefficients
    iP_W1$rho
    iP_W1$R_cuadr
    iP_W1$suma

    ieTpw <- iP_W1$residu[n]
    (iyTmlpw <- sum(iP_W1$coefficients * c(1,CPI[35])) + (iP_W1$rho * ieTpw))

windows()
par(mfrow=c(1,2))

plot(CPI[-36],TRSBILL[-36], main = 'Modelo ajustado por P-W Iterativo', ylab = 'TRSBILL', xlab = 'CPI',pch= 16,col='red')
abline(coef = iP_W1$coefficients,pch= 16,col='green',lwd=2)

windows()
par(mfrow=c(1,2))
ipw <- ts(iP_W1$residu, frequency=12,start=c(1971,1))
ts.plot(ipw,main = 'Residuales del modelo P-W Iterativo',xlab=' ',ylab='Residual')
points(ipw, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')

ipwt <- ts(iP_W1$eiPW, frequency=12,start=c(1971,1))
ts.plot(ipwt,main = 'Residuales del modelo transformado P-W Iterativo',xlab=' ',ylab='Residual')
points(ipwt, pch=16, col = 'red')
abline(a=0, b=0, col = 'blue')

#### fin ####

```

**ANEXO C: Distribución del estadístico de contraste de Durbin y Watson**

$n$	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
16	0.98	1.24	0.86	1.40	0.75	1.59	0.64	1.80	0.53	2.03
17	1.01	1.25	0.90	1.40	0.79	1.58	0.68	1.77	0.57	1.98
18	1.03	1.26	0.93	1.40	0.82	1.56	0.72	1.74	0.62	1.93
19	1.06	1.28	0.96	1.41	0.86	1.55	0.76	1.73	0.66	1.90
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.72	0.70	1.87
21	1.10	1.30	1.01	1.41	0.92	1.54	0.83	1.69	0.73	1.84
22	1.12	1.31	1.04	1.42	0.95	1.54	0.86	1.68	0.77	1.82
23	1.14	1.32	1.06	1.42	0.97	1.54	0.89	1.67	0.80	1.80
24	1.16	1.23	1.08	1.43	1.00	1.54	0.91	1.66	0.83	1.79
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
26	1.19	1.35	1.12	1.44	1.04	1.54	0.96	1.65	0.88	1.76
27	1.21	1.35	1.12	1.44	1.06	1.54	0.99	1.64	0.91	1.75
28	1.22	1.37	1.15	1.45	1.08	1.54	1.01	1.64	0.93	1.74
29	1.24	1.38	1.17	1.45	1.10	1.54	1.03	1.63	0.96	1.73
30	1.25	1.38	1.80	1.46	1.12	1.54	1.05	1.63	0.98	1.73
31	1.26	1.39	1.20	1.47	1.13	1.55	1.07	1.63	1.00	1.72
32	1.27	1.40	1.21	1.47	1.15	1.55	1.08	1.63	1.02	1.71
33	1.28	1.41	1.22	1.48	1.16	1.55	1.10	1.63	1.04	1.71
34	1.29	1.41	1.24	1.48	1.17	1.55	1.12	1.63	1.06	1.70
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
36	1.31	1.43	1.25	1.49	1.20	1.56	1.15	1.63	1.09	1.70
37	1.32	1.43	1.27	1.49	1.21	1.56	1.16	1.62	1.10	1.70
38	1.33	1.44	1.28	1.50	1.23	1.56	1.17	1.62	1.12	1.70
39	1.34	1.44	1.29	1.50	1.24	1.56	1.19	1.63	1.13	1.69
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.53	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.54	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.63	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
75	1.53	1.58	1.50	1.61	1.47	1.64	1.45	1.67	1.42	1.70
80	1.54	1.59	1.52	1.63	1.49	1.65	1.47	1.67	1.44	1.70
85	1.56	1.60	1.53	1.63	1.51	1.66	1.49	1.68	1.46	1.71
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
95	1.58	1.62	1.56	1.65	1.54	1.67	1.52	1.69	1.50	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72