



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**REGRESIÓN EN ANÁLISIS DE SERIES DE TIEMPO
APLICADO A LOS
PRECIOS DEL COMBUSTIBLE EN
EUA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

ALDO ARTURO ZUNZUNEGUI LÓPEZ



**DIRECTOR DE TESIS:
MAT. MARGARITA ELVIRA CHÁVEZ CANO**

Ciudad Universitaria, CD. MX. 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno

Zunzunegui

López

Aldo Arturo

5510057779

Universidad Nacional Autónoma de

México

Facultad de Ciencias

Actuaría

310334932

2. Datos del tutor

Mat.

Margarita Elvira

Chávez

Cano

3. Datos del sinodal 1

M. en C.

Ismael

Rivero

Guzmán

4. Datos del sinodal 2

Act.

Gabriel

Ordaz

Rivera

5. Datos del sinodal 3

Act.

Jesús

Castillo

Gómez

6. Datos del sinodal 4

M. en C.

Jaime

Vázquez

Alamilla

7. Datos del trabajo escrito.

Regresión en Análisis de Series de Tiempo

Aplicado a los Precios del Combustible en

EUA

105 pág.

2022

*We don't do it for the glory
We don't do it for the recognition
We do it because it needs to be done
Because if we don't, no one else will
And we do it even if no one knows what we've done
Even if no one knows we exist
Even if no one remembers we ever existed*

Agradecimientos

Me gustaría comenzar agradeciendo a mi asesora, la Mat. Margarita Elvira Chávez Cano por todo su tiempo, paciencia y orientación para la realización de este trabajo.

A mis sinodales por su tiempo y apoyo, especialmente durante este tiempo de contingencia sanitaria.

Por supuesto mi familia, porque sin ellos esto no hubiera sido posible.

Quiero agradecer a la Universidad Nacional Autónoma de México, por ser una institución pública y autónoma que permite a los mexicanos soñar en grande, porque ser estudiante de esta universidad ha sido uno de mis mayores logros en la vida, y lo será para muchos otros, por muchos años más, para jóvenes que quieren hacer grande a este país y no sólo a sí mismos.

Por último, pero no menos importante, quiero agradecerme a mí, por nunca rendirme, por mi coraje de seguir adelante cuando pensé que era difícil, porque si no lo hubiera hecho, nadie más lo hubiera hecho por mí.

Índice general

Introducción	V
1. Análisis a Series de Tiempo sin Retraso	1
1.1. Modelo de Regresión de Series de Tiempo	2
1.1.1. Proceso Autorregresivo de Primer Orden	8
1.1.2. Función de Autocorrelación (FAC)	11
1.1.3. Prueba de Hipótesis	13
1.1.3.1. Prueba de Durbin-Watson	13
1.2. Método de Cochrane-Orcutt	14
1.2.1. Estimación de ψ	15
1.3. Método de Máxima Verosimilitud	15
1.4. Otros Procesos	17
1.4.1. Proceso Autorregresivo AR(p)	18
1.4.2. Modelo de medias móviles MA(q)	19
1.4.3. Modelo ARMA(p,q)	20
1.5. Otras Pruebas de Autocorrelación	21
1.5.1. Prueba de Breusch-Godfrey	21
1.5.2. Prueba de Ljung-Box	22
1.6. Proceso de Identificación	23
1.6.1. Función de Autocorrelación Parcial (FAP)	24
1.6.2. Cálculo de Otros estimadores	26
1.7. Regresión Múltiple	32
1.7.1. Prueba de Significancia Global	37
1.7.2. Prueba de Significancia Individual	39
1.7.3. Intervalos de Confianza	40
2. Análisis a Series de Tiempo con Retraso	43
2.1. Modelos con retrasos	43
2.1.1. Retrasos de Variables Dependientes	45
2.1.1.1. Pruebas de Autocorrelación	49
2.1.1.2. Estimación de Parámetros	51
3. Pronósticos	57
3.1. Error de pronóstico	58
3.2. Pronóstico	62

4. Aplicación	66
4.1. Caso 1	66
4.2. Caso 2	75
4.3. Caso 3	81
Conclusión	86
A. Apéndice: Tabla Durbin-Watson	89
B. Anexo: Códigos en R	90
B.1. Código para el Caso 1	90
B.2. Código para el Caso 2	93
B.3. Código para el Caso 3	95
Bibliografía	98

Introducción

Los métodos de regresión han sido una parte integral del análisis de series de tiempo durante mucho tiempo, remontándose al menos cien años al trabajo de Schuster (1898). El trabajo de Schuster sobre regresión sinusoidal se aplicó en la estimación de “periodicidades ocultas” y condujo a la invención de periodograma.

Los modelos de regresión estructural para series de tiempo han existido durante muchos años y han figurado de manera prominente en la literatura de econometría y finanzas. Tratados rigurosamente por Anderson (1971) y Fuller (1996) entre otros, estos modelos estructurales se han utilizado durante años en predicción y descomposición de series de tiempo en componentes de “tendencia”, “estacionalidad” e “irregulares”. Otro ejemplo distintivo es la clase de modelos autorregresivos integrados de media móvil (ARIMA) que se asociaron con Box y Jenkins (1979).

En este trabajo se pretende utilizar este tipo de modelos para analizar el comportamiento con datos reales, en este caso de los precios del combustible en Estados Unidos, analizar las variables que influyen en el y su ajuste a cada uno modelos ARIMA para concluir si es factible pronosticar con alguno de estos modelos.

Capítulo 1

Análisis a Series de Tiempo sin Retraso

Naturaleza de las series de tiempo

Las series de tiempo son registros numéricos realizados en diversos intervalos de tiempo creando una secuencia de observaciones sobre algún fenómeno observado en intervalos regulares. Estos intervalos pueden corresponder a años, trimestres, meses, días, etc. Los valores registrados varían en el tiempo, por lo que un estudio de las series de tiempo es precisamente analizar la naturaleza de las variaciones.

Las variaciones pueden ser sistemáticas o aleatorias; las variaciones sistemáticas son las que ocurren con frecuencia y nos permiten modelar la serie, estas variaciones pueden presentarse en periodos cortos y otras en periodos largos mostrando comportamientos crecientes o decrecientes, incluso podrían ser variaciones cíclicas. Para poder analizar estos comportamientos es necesario tomar en cuenta un periodo de tiempo relativamente extenso para establecer una tendencia lineal significativa, considerando que la relación es lineal, por otro lado, las variaciones aleatorias son provocadas por eventos poco frecuentes o extraños, pero posibles y son difíciles de predecir.

Los métodos de este capítulo son apropiados cuando se trata de Series de tiempo posiblemente no estacionarias. Además, si se hace hincapié en predecir los valores futuros, entonces el problema se trata fácilmente como un problema de regresión.

Esencialmente existen dos tipos de modelos de series de tiempo: con retraso y sin retraso. Un modelo sin retraso es en donde las variables endógenas y las exógenas son observadas al mismo tiempo, sin tener que observar valores anteriores de ningún tipo de variable, es decir, no están correlacionadas, por otro lado, existen los modelos con retraso en los que los valores de las variables endógenas están relacionados con los valores pasados de las variables endógenas y/o exógenas. En este capítulo nos enfocaremos en los modelos sin retraso.

1.1. Modelo de Regresión de Series de Tiempo

Gran parte del análisis de series de tiempo aplicado comienza con la siguiente premisa: X y Y son dos variables, que representan alguna población, y estamos interesados en explicar a Y en términos de X , o en estudiar cómo Y varía con cambios en X .

Una forma de ver las series de tiempo es mediante el contexto de regresión lineal, utilizando datos de series de tiempo particulares. Una colección de datos en un orden temporal especificado $\{X_t\}_{t=1,2,\dots,N}$ con intervalos entre X_t y X_{t+1} fijos y constantes es considerada una serie de tiempo.

Sea Y_t para $t = 1, \dots, N$, una variable influenciada por series independientes $\{X_{ti}\}$. Esta relación puede ser expresada a través del modelo de regresión lineal.

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_k X_{tk} + e_t$$

Donde β_0, \dots, β_k son los coeficientes de regresión fijos, y e_t es un error aleatorio.

Cuando se observa una aparente tendencia en los datos como en la Figura 1.1, puede aplicarse regresión lineal simple para estimar dicha tendencia mediante el ajuste del modelo.

Este gráfico (Figura 1.1) se denomina diagrama de dispersión. Esta pantalla sugiere claramente una relación entre la variable dependiente Y y la variable independiente X , la impresión es que los datos en general, pero no exactamente, caen a lo largo de una línea recta.

La diferencia entre el valor observado de Y_t y la recta se considera un error e_t para cada t . Es conveniente pensar en e_t como un error estadístico; es decir, es una variable aleatoria que explica el fallo del modelo para ajustar los datos exactamente.

En el análisis de modelos de regresión de series de tiempo está relacionado con términos estocásticos, que son los que incluyen términos de error. El modelo más simple se conoce como el *modelo de serie de tiempo simple* y se expresa como una línea recta con un término de error, una variable aleatoria a tiempo t que comúnmente se considera normal con media cero y varianza constante, pero no necesariamente, este modelo se expresa de la forma siguiente:

$$Y_t = a + bX_t + e_t \tag{1.1}$$

En donde se puede observar que Y_t es la variable endógena o dependiente, X_t la variable

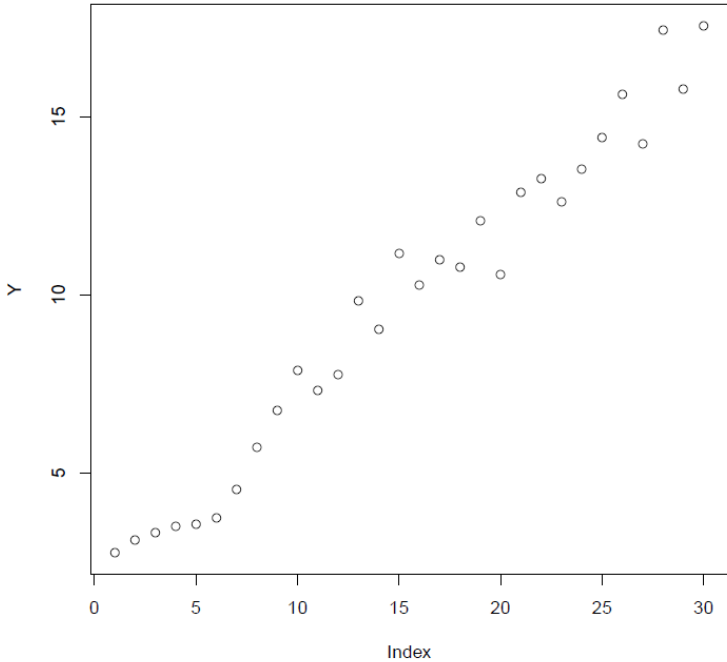


Figura 1.1: Diagrama de dispersión

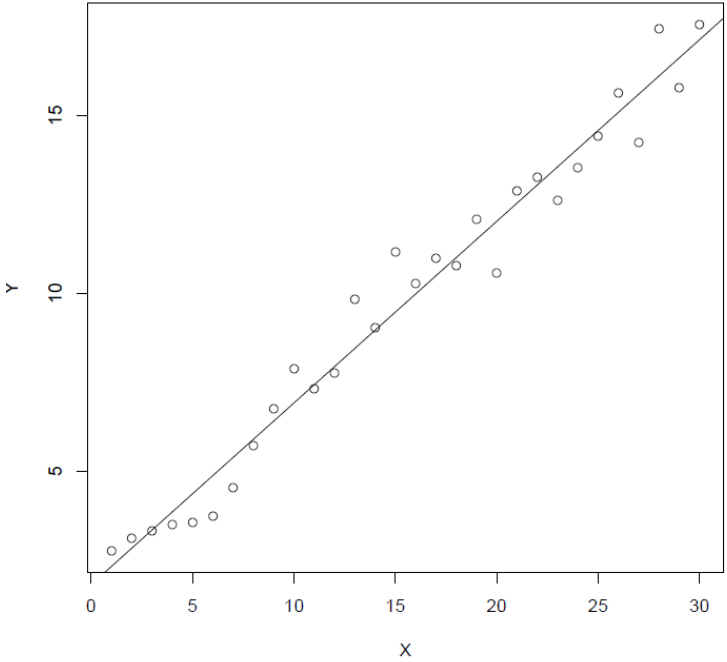


Figura 1.2: Datos ajustados a una línea recta por mínimos cuadrados.

exógena o independiente, a y b son los parámetros desconocidos, b es la pendiente y puede interpretarse como el cambio en la media de Y por un cambio de unidad en X , si el rango de datos en X incluye $X = 0$, entonces la intersección a es la media de la distribución de la respuesta Y cuando $X = 0$, e_t es el término de perturbación aleatorio, el cual no puede ser observado, por lo que debe ser distinguido del residual \hat{e}_t , y t indica el tiempo dentro del intervalo en el que están dadas las observaciones de Y_t y X_t , es decir, $t \in [1, N]$. Dado este modelo se consideran los siguientes supuestos:

1. Linealidad: la relación entre Y_t y X_t es lineal
2. X es no estocástico: $\mathbb{E}[e_t X_t] = 0$
3. Media zero: $\mathbb{E}[e_t] = 0$.
4. Varianza Constante: $\mathbb{E}[e_t^2] = \sigma^2$
5. No correlacionado: $\mathbb{E}[e_t e_{t-m}] = 0$ para $m \neq 0$

Siempre que se cumplan estos supuestos será posible estimar de manera eficiente los estimadores del modelo \hat{a} y \hat{b} mediante el principio de mínimos cuadrados minimizando la suma de cuadrados de los residuales, para eso se considera la siguiente expresión.

$$\sum_{t=1}^N \hat{e}_t^2 = \sum_{t=1}^N (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^N (Y_t - \hat{a} - \hat{b}X_t)^2$$

Utilizando herramientas de calculo para encontrar el mínimo de la expresión anterior se procede a derivar con respecto a los coeficientes \hat{a} y \hat{b} e igualando a cero se obtienen las ecuaciones normales:

$$N\hat{a} + \hat{b} \sum_{t=1}^N X_t = \sum_{t=1}^N Y_t$$

$$\hat{a} \sum_{t=1}^N X_t + \hat{b} \sum_{t=1}^N X_t^2 = \sum_{t=1}^N Y_t X_t$$

Teniendo un sistema de ecuaciones de dos por dos, se obtienen los estimadores de \hat{a} y \hat{b} :

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}, \quad \hat{b} = \frac{\sum_{t=1}^N (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^N (X_t - \bar{X})^2}$$

Se representan de esta manera utilizando sus medias muestrales $\bar{X} = \frac{1}{N} \sum_{t=1}^T X_t$, $\bar{Y} = \frac{1}{N} \sum_{t=1}^N Y_t$ para una expresión más compacta. Estos estimadores son los óptimos para el modelo simple ya que cumplen con varias propiedades importantes bajo los supuestos ya mencionados, para eso es necesario calcular sus esperanzas, varianzas y su covarianza.

$$\begin{aligned} \mathbb{E}[\hat{b}] &= \mathbb{E} \left[\sum_{t=1}^N \frac{(X_t - \bar{X})Y_t}{\sum_{t=1}^N (X_t - \bar{X})^2} \right] \\ &= \frac{\sum_{t=1}^N (X_t - \bar{X})\mathbb{E}[Y_t]}{\sum_{t=1}^N (X_t - \bar{X})^2} \\ &= \frac{\sum_{t=1}^N (X_t - \bar{X})(a + bX_t)}{\sum_{t=1}^N (X_t - \bar{X})^2} \\ &= a \frac{\sum_{t=1}^N (X_t - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} + b \frac{\sum_{t=1}^N (X_t - \bar{X})(X_t)}{\sum_{t=1}^N (X_t - \bar{X})^2} \\ &= b \end{aligned}$$

Para el caso de \hat{a}

$$\begin{aligned} \mathbb{E}[\hat{a}] &= \mathbb{E}[\bar{Y} - \hat{b}\bar{X}] = \frac{1}{N} \sum_{t=1}^N \mathbb{E}[Y_t] - \mathbb{E}[\hat{b}]\bar{X} \\ &= \frac{1}{N} \sum_{t=1}^N (a + bX_t) - b\bar{X} \\ &= a + b\bar{X} - b\bar{X} \\ &= a \end{aligned}$$

Lo que indica que estos estimadores son insesgados, por otro lado es importante calcular las varianzas de estos estimadores para un mejor análisis del modelo.

$$Var(\hat{a}) = Var(\bar{Y} - \hat{b}\bar{X}) = Var\left(\bar{Y} - \frac{\sum_{t=1}^N \bar{X}(X_t - \bar{X})Y_t}{\sum_{t=1}^N (X_t - \bar{X})^2}\right)$$

Ahora por el supuesto de independencia y la no aleatoriedad de X_t esta varianza se simplifica a sólo la varianza de Y_t , que es constante para toda t , por una constante en función de los valores de X_t .

$$\begin{aligned} Var\left(\bar{Y} - \frac{\sum_{t=1}^N \bar{X}(X_t - \bar{X})Y_t}{\sum_{t=1}^N (X_t - \bar{X})^2}\right) &= \sum_{t=1}^N \left(\frac{1}{N} - \frac{\bar{X}(X_t - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2}\right)^2 Var(Y_t) \\ &= \sigma^2 \sum_{t=1}^N \left(\frac{1}{N} - \frac{\bar{X}(X_t - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2}\right)^2 \\ &= \sigma^2 \left(\frac{1}{N} + \frac{\bar{X}^2}{\sum_{t=1}^N (X_t - \bar{X})^2}\right) \end{aligned}$$

De manera similar con la varianza de \hat{b}

$$\begin{aligned} Var(\hat{b}) &= Var\left(\sum_{t=1}^N \frac{(X_t - \bar{X})Y_t}{\sum_{t=1}^N (X_t - \bar{X})^2}\right) \\ &= \sum_{t=1}^N \left(\frac{X_t - \bar{X}}{\sum_{t=1}^N (X_t - \bar{X})^2}\right)^2 Var(Y_t) \\ &= \sigma^2 \left(\frac{1}{\sum_{t=1}^N (X_t - \bar{X})^2}\right) \end{aligned}$$

Estos estimadores son eficientes en sentido de varianza mínima y consistentes, es decir, entre más grande sea la muestra estos convergerán al verdadero valor de los parámetros.

Esta afirmación se hace por el teorema de Gauss-Markov.

Un problema que se presenta en los estimadores es precisamente sus varianzas, ya que estas dependen de la varianza de Y_i que por lo general en la práctica no se conoce, para eso es preciso calcular la estimación de σ^2 .

Una estimación para σ^2 puede ser el estimador máximo verosímil:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N \hat{e}_t^2$$

Aunque este estimador desafortunadamente no es insesgado, esto se puede solucionar. Para esto primero se necesita tomar en cuenta el valor esperado de $\sum_{t=1}^N \hat{e}_t^2$.

$$\begin{aligned} & \mathbb{E} \left[\sum_{t=1}^N \hat{e}_t^2 \right] \\ &= \mathbb{E} \left[\sum_{t=1}^N (\hat{a} - a)^2 + \sum_{t=1}^N (\hat{b} - b)^2 X_t^2 + \sum_{t=1}^N e_t^2 + 2(\hat{a} - a)(\hat{b} - b) \sum_{t=1}^N X_t - 2(\hat{a} - a) \sum_{t=1}^N e_t - 2(\hat{b} - b) \sum_{t=1}^N e_t \right] \\ &= N\mathbb{E}(\hat{a} - a)^2 + \sum_{t=1}^N X_t^2 \mathbb{E}(\hat{b} - b)^2 + \mathbb{E}e_t^2 + 2 \sum_{t=1}^N X_t \mathbb{E}[(\hat{a} - a)(\hat{b} - b)] \\ &\quad - 2\mathbb{E} \left[(\hat{a} - a) \sum_{t=1}^N e_t \right] - 2\mathbb{E} \left[(\hat{b} - b) \sum_{t=1}^N e_t \right] \\ &= \sigma^2 + \frac{N\bar{X}\sigma^2}{\sum_{t=1}^N (X_t - \bar{X})^2} + \sigma^2 \frac{\sum_{t=1}^N X_t^2}{\sum_{t=1}^N (X_t - \bar{X})^2} + N\sigma^2 \\ &\quad - 2N\sigma^2 \frac{\bar{X}}{\sum_{t=1}^N (X_t - \bar{X})^2} - 2\sigma^2 - 2\sigma^2 \\ &= \sigma^2 \frac{\sum_{t=1}^N X_t^2 - N\bar{X}^2}{\sum_{t=1}^N (X_t - \bar{X})^2} + (N - 3)\sigma^2 = (N - 2)\sigma^2 \end{aligned}$$

Por lo tanto un estimador insesgado para σ^2 está dado por la siguiente expresión:

$$\Rightarrow \frac{1}{(N-2)} \sum_{t=1}^N \hat{e}_t^2 = \hat{\sigma}^2 \quad (1.2)$$

Una propiedad interesante del estimador es que en la expresión del denominador $(N-2)$, el número 2 corresponde al número de parámetros en el modelo. De modo que si se tuviesen k parámetros en el modelo, el denominador de la estimación de la varianza sería $(N-k)$.

1.1.1. Proceso Autorregresivo de Primer Orden

Un supuesto fundamental en este caso es la no correlación, es decir, que los términos de perturbación e_t que son los términos aleatorios en el modelo, no deben de estar relacionados entre si mismos para cuales quiera tiempos distintos. Sin este supuesto el modelo no funcionaría correctamente, por lo tanto, es natural preguntarse que tanta correlación existe entre estos términos.

Un posible modelo se conoce como *Proceso autorregresivo de primer orden* que se expresa de la siguiente forma:

$$e_t = \psi e_{t-1} + v_t \quad -1 < \psi < 1 \quad (1.3)$$

Donde e_t y e_{t-1} son los términos de error a tiempo t y $t-1$ respectivamente, ψ es un coeficiente de regresión, y v_t es una variable aleatoria que cumple con las siguientes propiedades para toda t y $m \neq 0$:

$$\begin{aligned} \mathbb{E}[v_t] &= 0 \\ \mathbb{E}[v_t^2] &= \sigma_v^2 \\ \mathbb{E}[v_t v_{t-m}] &= 0 \\ \mathbb{E}[v_t e_{t-1}] &= 0 \end{aligned}$$

Este proceso se define de manera recursiva, cada término de perturbación puede expresarse en términos de los anteriores de manera lineal.

$$e_t = \psi e_{t-1} + v_t$$

$$\begin{aligned}
&= \psi(\psi e_{t-2} + v_{t-1}) + v_t \\
&\quad \cdot \\
&\quad \cdot \\
&\quad \cdot \\
&= \psi^m e_{t-m} + \psi^{m-1} v_{t-m-1} + \dots + \psi v_{t-1} + v_t
\end{aligned} \tag{1.4}$$

$$\begin{aligned}
&\quad \cdot \\
&\quad \cdot \\
&\quad \cdot \\
&= \psi^t e_0 + \psi^{t-1} v_1 + \psi^{t-2} v_2 + \dots + \psi v_{t-1} + v_t
\end{aligned} \tag{1.5}$$

De esta forma se expresan cada error como combinación lineal de v_t 's con sólo un término inicial e_0 . Ahora es preciso especificar el proceso generado por e_0 , para mantener los supuestos, la esperanza $\mathbb{E}[e_0] = 0$ y su varianza se puede deducir de la siguiente manera:

$$\begin{aligned}
\text{Var}(e_1) &= \mathbb{E}[e_1^2] \\
&= \mathbb{E}[(\psi e_0 + v_1)^2] \\
&= 2\psi \mathbb{E}[e_0 v_1] + \psi^2 \mathbb{E}[e_0^2] + \mathbb{E}[v_1^2] \\
&= \psi^2 \mathbb{E}[e_0^2] + \sigma_v^2 \\
\implies \mathbb{E}[e_1^2] &= \psi^2 \mathbb{E}[e_0^2] + \sigma_v^2
\end{aligned}$$

Para cumplir con el supuesto de varianza constante para toda t se puede asumir $\mathbb{E}[e_1^2] = \mathbb{E}[e_0^2]$

$$\implies \mathbb{E}[e_0^2] = \frac{\sigma_v^2}{1 - \psi^2}$$

Partiendo de la expresión (1.4) en la que se representa cada e_t con respecto a todos los términos anteriores, es necesario verificar si esta no viola ninguno de los supuestos.

$$\mathbb{E}[e_t] = \psi^t \mathbb{E}[e_0] + \psi^{t-1} \mathbb{E}[v_1] + \psi^{t-2} \mathbb{E}[v_2] + \dots + \psi \mathbb{E}[v_{t-1}] + \mathbb{E}[v_t] = 0$$

Dado que todas las variables tienen esperanza nula, la esperanza de e_t es cero para toda t . Después nótese que el supuesto de la varianza constante se sigue cumpliendo.

$$\begin{aligned}
Var(e_t) &= \psi^{2t}Var(e_0) + \psi^{2(t-1)}Var(v_1) + \dots + Var(v_t) \\
&= \psi^{2t}[\sigma_v^2/(1 - \psi^2)] + \psi^{2(t-1)}\sigma_v^2 + \dots + \psi^2\sigma_v^2 + \sigma_v^2 \\
&= \psi^{2t}[\sigma_v^2/(1 - \psi^2)] + \sigma_v^2[\psi^{2(t-1)} + \dots + \psi^2 + 1] \\
&= \sigma_v^2[\psi^{2t}/(1 - \psi^2) + (1 - \psi^{2t})/(1 - \psi^2)] \\
&= \sigma_v^2/(1 - \psi^2)
\end{aligned}$$

De esta manera todos los términos de error tienen varianza constante. Nótese que se utilizó el supuesto de $-1 < \psi < 1$, ya que si este fuera mayor o igual a 1 en valor absoluto, la varianza a medida que N tienda a infinito esta también y provocaría un modelo inservible con una perturbación gigantesca.

Por último, la covarianza entre los errores debe de ser cero o muy cercana a 0:

$$\begin{aligned}
Cov(e_t e_{t-m}) &= \mathbb{E}[e_t e_{t-m}] - \mathbb{E}[e_t]\mathbb{E}[e_{t-m}] \\
&= \mathbb{E}[e_t e_{t-m}]
\end{aligned}$$

Utilizando la expresión (1.4) donde se expresa a e_t de manera lineal y multiplicando por e_{t-m} , esta esperanza queda de la siguiente forma:

$$\begin{aligned}
&= \psi^m \mathbb{E}[e_{t-m}^2] + \psi^{m-1} \mathbb{E}[e_{t-m} v_{t-m+1}] + \dots + \psi \mathbb{E}[e_{t-m} v_{t-1}] + \mathbb{E}[e_{t-m} v_t] \\
&= \psi^m \mathbb{E}[e_{t-m}^2] \\
&= \psi^m Var(e_{t-m}) \\
&= \psi^m \sigma_e^2
\end{aligned}$$

En donde la varianza de e_{t-m} es distinta de cero y mientras ψ sea pequeño, no se estarían violando los supuestos.

1.1.2. Función de Autocorrelación (FAC)

Resulta importante considerar la función de autocorrelación, la cual representa la correlación entre dos errores en distintos puntos de tiempo, esta función se define a continuación como:

$$\begin{aligned}\rho(m) &= \frac{Cov[e_t e_{t-m}]}{\sqrt{Var(e_m)} \sqrt{Var(e_{t-m})}} \\ &= \frac{\psi^m \sigma_e^2}{\sigma_e^2} \\ &= \psi^m\end{aligned}$$

Este resultado indica teóricamente la correlación entre los términos de perturbación m periodos de diferencia respecto a t bajo el proceso autorregresivo de primer orden. De esta forma $\rho(m)$ se encuentra también entre -1 y 1. En el caso de $\psi = 0$ se siguen cumpliendo los supuestos ya que los e_t 's tienen media cero y varianza constante. En caso contrario con una fuerte correlación puede observarse gráficamente mediante un correlograma, como a continuación en la figura 1.3, una correlación positiva.

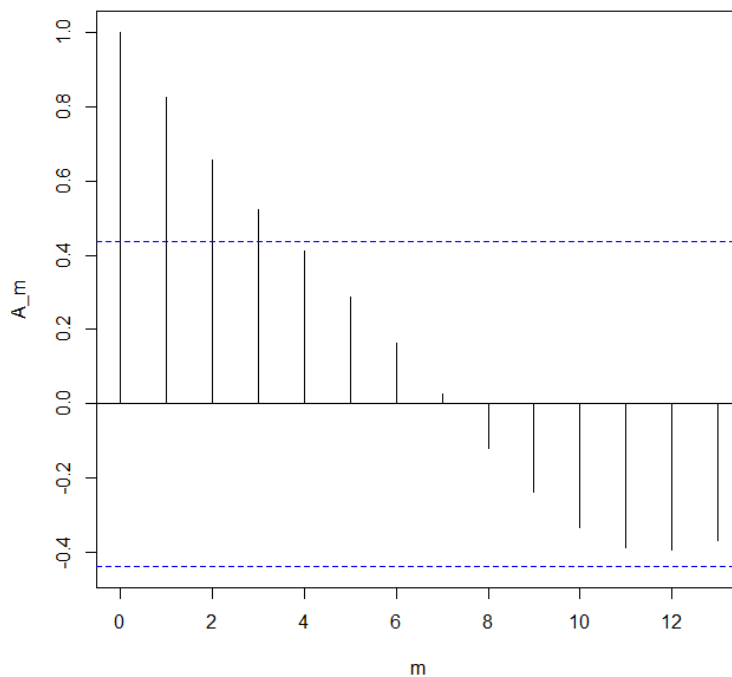


Figura 1.3: Correlograma generado en R con $\psi = 0.85$

Cunado el coeficiente de regresión es alto puede notarse a simple vista en el correlograma como el termino de perturbación e_t está fuertemente correlacionado con los términos

inmediatos anteriores y va disminuyendo, el término e_{t-1} se encuentra muy correlacionado con el término e_{t-2} y así sucesivamente, a veces incluso podrían estar mayor correlacionados unas veces que otras. En otros casos la correlación puede ser negativa, es decir, cuando $-1 < \psi < 0$. La función de autocorrelación $\rho(m)$ oscilaría como en la gráfica 1.4

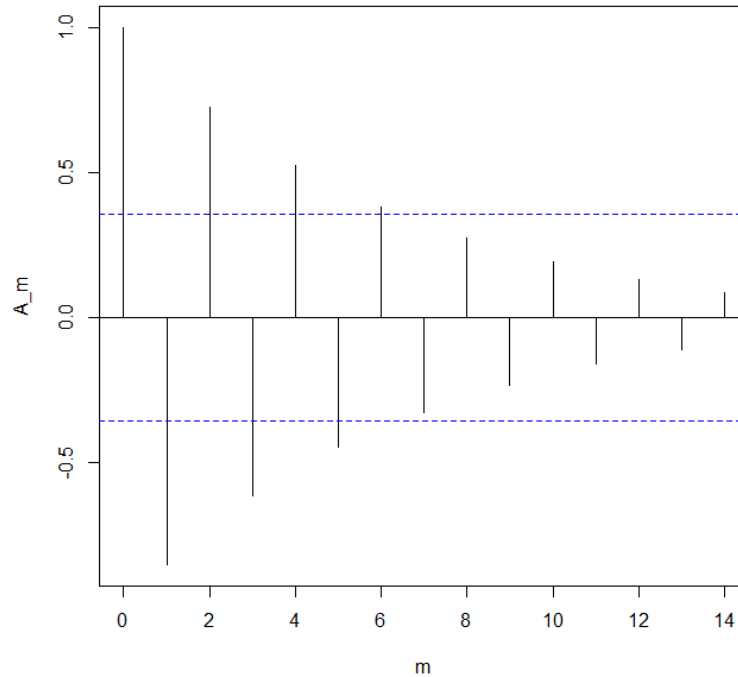


Figura 1.4: Correlograma generado en R con $\psi = -0.85$

Igual que en el primer caso la correlación es muy alta entre los términos de perturbación, pero cuando m es impar la función de autocorrelación es negativa, y positiva cuando m es par, por eso el comportamiento de la gráfica es oscilatorio.

Estos son los casos que se presentan cuando no se cumplen los supuestos de autocorrelación usando el proceso autorregresivo de primer orden, aunque este no es el único.

Por otro lado, generalmente no se puede calcular tan fácilmente las funciones de autocorrelación, por lo que se suele usar una estimación, la estimación comúnmente se realiza de siguiente forma:

$$\hat{\rho}(m) = \frac{C_m}{C_0}$$

Donde C_i es la estimación de la covarianza entre e_{t+i} y e_t

$$C_i = \frac{1}{N} \sum_{t=1}^N (e_t - \bar{e})(e_{t+i} - \bar{e})$$

1.1.3. Prueba de Hipótesis

La función de autocorrelación es muy útil pero para esto primero se necesita realizar una prueba de Hipótesis para ψ pues se desea que este valor sea lo más pequeño posible. Lo que se quisiera es que el valor de ψ sea prácticamente cero para decir que los errores no son autorregresivos, entonces es fácil pensar que la prueba de hipótesis adecuada es la siguiente:

$$H_0 : \psi = 0 \quad vs \quad H_A : \psi \neq 0$$

1.1.3.1. Prueba de Durbin-Watson

Para probar que no existe correlación mediante la prueba de hipótesis para el caso de modelo de regresión simple, existen varios test, en la literatura comúnmente se utiliza la prueba de Durbin-Watson(1951).

En general la correlación es positiva, por esta razón la prueba de Durbin-Watson considera a la hipótesis alternativa como $H_A : \psi > 0$, el otro caso, cuando la correlación es negativa se discutirá después. Para la prueba se considera la estadística d que se define como sigue:

$$d = \frac{\sum_t (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_t \hat{e}_t^2} \quad (1.6)$$

Es claro que si existe una correlación positiva entonces el numerador será pequeño en comparación al denominador i.e, $\sum_t (\hat{e}_t - \hat{e}_{t-1})^2 < \sum_t \hat{e}_t^2$, así la estadística d será pequeña. En caso de que la correlación sea negativa entonces pasaría lo contrario.

Para tomar una decisión con respecto a la hipótesis de si se rechaza o no H_0 . Durbin y Watson establecieron los límites para d .

si $d < d_l$ se rechaza H_0

si $d_u < d$ No se rechaza H_0

si $d_l < d < d_u$ No se puede concluir

Donde d_u se define como el valor crítico superior y a d_l como el valor crítico inferior. Estos valores dependen del tamaño de la muestra, del nivel de significancia y del número de coeficientes de regresión y se pueden obtener de la Tabla (Figura A.1), siempre y cuando los errores no estén retrasados por variables dependientes.

La estadística d cuando se aproxima lo suficiente a 2 implica que el valor de autocorrelación es cero, es decir, $\psi = 0$. Esto se puede ver con una aproximación de la estadística en la expresión (1.6) tomando en cuenta el estimador de ψ el cual se presenta más adelante.

$$d = \frac{\sum_t (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_t \hat{e}_t^2} = \frac{\sum_{t=2} \hat{e}_t^2 + \sum_{t=2} \hat{e}_{t-1}^2 - 2 \sum_{t=2} \hat{e}_t \hat{e}_{t-1}}{\sum_t \hat{e}_t^2} \approx 2(1 - \psi)$$

Cuando la hipótesis alternativa es $H_A : \psi < 0$, es decir, la correlación es negativa, dado que d oscila entre 0 y 4, ya que ψ se encuentra entre -1 y 1, simplemente se toma la regla de decisión de la siguiente manera:

si $4 - d_l < d$ *se rechaza* H_0

si $d < 4 - d_u$ *No se rechaza* H_0

si $4 - d_u < d < 4 - d_l$ *No se puede concluir*

Si no existe evidencia de una significativa autocorrelación entonces se pueden aceptar los estimadores de mínimos cuadrados, mientras que si se rechaza H_0 tal vez se pueda recurrir a otras técnicas de estimación.

1.2. Método de Cochrane-Orcutt

Cuando existe una correlación entre los datos y se tienen variaciones sistemáticas en los errores, provoca que la estimación de la varianza no sea la adecuada y el análisis no sea correcto. Si la autocorrelación aparente resulta de predictores faltantes y si estos predictores faltantes pueden ser identificados e incorporados en el modelo, el problema de la supuesta autocorrelación puede ser eliminado, cuando aun añadiendo las nuevas variables predictoras no se elimina el problema, es necesario tener en cuenta explícitamente la estructura autocorrelativa en el modelo y sus los estimadores, para esto se puede usar el método de Cochrane-Orcutt(1949).

Partiendo del modelo de serie de tiempo simple, este método consta en una transformación de la expresión (1.1) haciendo:

$$Y'_t = Y_t - \phi Y_{t-1} \tag{1.7}$$

y ahora sustituyendo

$$\begin{aligned}
Y'_t &= Y_t - \phi Y_{t-1} = a + bX_t + e_t - \phi(a + bX_{t-1} + e_{t-1}) \\
&= a(1 - \phi) + b(X_t - \phi X_{t-1}) + e_t - \phi e_{t-1}
\end{aligned} \tag{1.8}$$

Nótese que la expresión anterior depende de un parámetro desconocido ϕ , sin embargo, si se asume el proceso autorregresivo de primer orden, este se considera como ψ y puede ser estimado. De esta forma haciendo $a' = a(1 - \psi)$ y $X'_t = (X_t - \psi X_{t-1})$ la expresión anterior se simplifica.

$$\begin{aligned}
Y'_t &= a' + b'X'_t + e_t - \psi e_{t-1} \\
&= a' + b'X'_t + v_t
\end{aligned}$$

Obteniendo un modelo de serie de tiempo simple, ya que, v_t cumple con los supuestos básicos, por lo tanto, se le puede aplicar el método de mínimos cuadrados para calcular los estimadores de a' y b' y así, obtener los nuevos residuales para hacer la prueba de Durbin-Watson. Este procedimiento incluso se usa hasta una segunda vez para lograr que los errores se consideren no correlacionados.

Una pequeña desventaja de este método es que disminuye la cantidad de datos en uno.

1.2.1. Estimación de ψ

Hasta el momento se ha supuesto el modelo de mínimos cuadrados ordinarios (OLS-Ordinary least squares) el cual asume que se conocen las distribuciones de los errores y sus coeficientes de regresión. Es posible calcular los estimadores de los errores, en este caso el parámetro ψ , y después utilizar una transformación al modelo. Este proceso se conoce como Estimación Generalizada de Mínimos Cuadrados (EGLS).

Para encontrar el estimador de ψ se toman en cuenta los valores de los residuales \hat{e}_t

$$\hat{\psi} = \frac{\sum_{t=2}^N \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^N \hat{e}_t^2}$$

1.3. Método de Máxima Verosimilitud

Otra alternativa es utilizar estimadores calculados mediante el método de máxima verosimilitud. Este método tiene sus ventajas, ya que puede utilizarse si se aplica a una

estructura más elaborada que un simple modelo autorregresivo de primer orden, aunque la desventaja es que se toma por echo que la distribución de los errores e_i 's, es conocida, sin embargo, generalmente estos se consideran que se distribuyen normal con media cero y varianza constante.

Ejemplo: Para poder calcular los estimadores se necesita encontrar su función de verosimilitud $L(\Theta)$ en donde Θ representa el vector de todos los parámetros desconocidos. En el caso del modelo autorregresivo de primer orden se puede expresar el término v_t partiendo de la transformación $Y'_t = Y_t - \phi Y_{t-1}$:

$$\begin{aligned} Y_t - \phi Y_{t-1} &= a(1 - \phi) + b(X_t - \phi X_{t-1}) + e_t - \phi e_{t-1} \\ \Rightarrow v_t &= Y_t - \phi Y_{t-1} - a(1 - \phi) - b(X_t - \phi X_{t-1}) \end{aligned} \quad (1.9)$$

De esta forma, el término aleatorio v_t queda expresado respecto a todos los parámetros. Ahora por el supuesto de que las variables $\{v_t\}$'s son iid, es fácil calcular la distribución conjunta.

$$\begin{aligned} f(v_2, v_3, \dots, v_N) &= \prod_{t=2}^N \frac{1}{\sigma_v \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{v_t}{\sigma_v} \right)^2 \right\} \\ &= \left(\frac{1}{\sigma_v \sqrt{2\pi}} \right)^{N-1} \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=2}^N v_t^2 \right\} \end{aligned}$$

Sustituyendo el valor de v_t de la expresión(1.9), se obtiene la función de verosimilitud.

$$L(a, b, \phi) = \left(\frac{1}{\sigma_v \sqrt{2\pi}} \right)^{N-1} \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=2}^N (Y_t - \phi Y_{t-1} - a(1 - \phi) - b(X_t - \phi X_{t-1}))^2 \right\}$$

Para facilitar la tarea de maximizar esta función, por lo general se le aplica la función logaritmo, de este modo se obtiene la función de log-verosimilitud:

$$\ln(L(a, b, \phi))$$

$$\begin{aligned}
&= (N - 1)[0 - \ln(\sigma_v \sqrt{2\pi})] - \frac{1}{2\sigma_v^2} \sum_{t=2}^N (Y_t - \phi Y_{t-1} - a(1 - \phi) - b(X_t - \phi X_{t-1}))^2 \\
&= (1 - N)[\ln(\sigma_v) + \frac{1}{2}\ln(2\pi)] - \frac{1}{2\sigma_v^2} \sum_{t=2}^N (Y_t - \phi Y_{t-1} - a(1 - \phi) - b(X_t - \phi X_{t-1}))^2
\end{aligned}$$

Para maximizar la función de log-verosimilitud se tiene que minimizar la expresión siguiente:

$$\sum_{t=2}^N (Y_t - \phi Y_{t-1} - a(1 - \phi) - b(X_t - \phi X_{t-1}))^2$$

Esta expresión no es más que la suma de los errores cuadrados, por lo que los estimadores máximo verosímil de Θ , en este caso son los mismos que los estimadores por mínimos cuadrados.

Por el momento únicamente se ha analizado el comportamiento de los términos de error con un modelo de primer orden, cada uno relacionado con el anterior inmediato. Este razonamiento es válido porque se vio en los correlogramas anteriores (figura 1.3 y 1.4) que la correlación es más fuerte al principio, sea positiva o negativa, y esta va disminuyendo paulatinamente hasta un punto del tiempo en el cual, los errores e_{t-m} y e_t prácticamente ya no están correlacionados, aunque claro se asumió un procesos de autoregresión de primer orden, pero este puede no ser el correcto en todos los casos, no se puede deliberadamente asumir tal proceso.

El problema se presenta al preguntarse cual proceso es el que mejor ajusta a los errores. Lo que se puede hacer es utilizar los residuales para realizar un correlograma además de generar errores de diferentes procesos que pueden ser adecuados y compararlos, así ver cual es el que mejor ajusta.

1.4. Otros Procesos

En este capítulo se analizarán algunas opciones de procesos de correlación de los errores bajo un supuesto importante, que un proceso de serie de tiempo sea estacionario tanto en su media como en su varianza, que son parecidas a los supuestos que normalmente se toman para el análisis de regresión. Existen dos definiciones de estacionalidad, la estricta y la débil. A continuación se hablará de estacionariedad refiriéndose a estacionariedad

débil ya que la estacionariedad estricta es más complicada de probar y generalmente se utiliza la segunda.

Una serie de tiempo se dice que es estrictamente estacionaria si sus propiedades no se ven afectadas por un cambio en el origen temporal. Es decir, si la distribución de probabilidad conjunta de las observaciones $Y_r, Y_{r+1}, \dots, Y_{r+n}$ es exactamente igual a la distribución de probabilidad conjunta de las observaciones $Y_{t+k}, Y_{r+k+1}, \dots, Y_{r+k+n}$ entonces la serie temporal es estrictamente estacionaria.

Definición: Estacionariedad débil

Una serie de tiempo X_t se dice débilmente estacionaria si:

- Tiene media μ_t constante para toda t
- $Cov(x_t, x_{t-h})$ no depende de t , únicamente del retraso h .

1.4.1. Proceso Autorregresivo AR(p)

El caso anterior en donde $e_t = \psi e_{t-1} + v_t$ se conoce como el proceso autoregresivo de primer orden AR(1), el cual es una caso particular de un proceso autoregresivo AR(p), donde p representa el orden de este.

$$e_t = \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots + \psi_p e_{t-p} + v_t \quad (1.10)$$

De esta forma se considera muy similar al proceso autoregresivo de segundo orden, pero en vez de que los errores estén relacionados únicamente con el error anterior inmediato como en el AR(1), se consideran el primero y segundo anterior, este proceso se denota comúnmente como AR(2) indicando entre paréntesis el segundo orden y se representa de la forma siguiente:

$$e_t = \psi_1 e_{t-1} + \psi_2 e_{t-2} + v_t \quad (1.11)$$

con

$$\begin{aligned} \psi_1 + \psi_2 &< 1 \\ \psi_2 - \psi_1 &< 1 \\ -1 &< \psi_2 < 1 \\ \mathbb{E}[e_t] = \mathbb{E}[v_t] = \mathbb{E}[e_{t-i}v_t] &= 0 \quad i \neq 0 \\ \mathbb{E}[v_t^2] &= \sigma^2 \\ \mathbb{E}[v_t v_{t-i}] &= 0 \quad i \neq 0 \end{aligned}$$

Con una función de autocorrelación :

$$\rho(h) = \begin{cases} \psi_1/(1 - \psi_2) & \text{si } h = 1 \\ \psi_2 + (\psi_1^2/(1 - \psi_2)) & \text{si } h = 2 \\ \psi_1\rho(h - 1) + \psi_2\rho(h - 2) & \text{si } h > 2 \end{cases}$$

Se puede observar que esta función de autocorrelación es recursiva como la del AR(1), de echo se puede mostrar que para cualquier AR(p) su función de autocorrelación es de esta forma:

$$\rho(h) = \psi_1\rho(h - 1) + \psi_2\rho(h - 2) + \dots + \psi_p\rho(h - p)$$

Generalmente los modelos AR de orden mayores a dos son muy extraños en la practica, ya que se ajustan mejor otros procesos como lo son los Medias Móviles.

1.4.2. Modelo de medias móviles MA(q)

A diferencia de un AR un proceso de medias móviles MA utiliza una representación lineal de las variables v_t de q tiempos anteriores.

$$e_t = v_t + v_{t-1}\theta + v_{t-2}\theta + \dots + v_{t-q}\theta \quad (1.12)$$

En este modelo los coeficientes θ_i pueden ser negativas o positivas, por lo que en la literatura pueden encontrarse como $-\theta_i$, por comodidad se utilizará simplemente como θ_i .

En el caso particular de MA(1):

$$e_t = v_t + v_{t-1}\theta$$

Se tiene la siguiente función de correlación:

$$Cov(e_t, e_{t-h}) = \begin{cases} (\theta^2 + 1)\sigma_v^2 & \text{si } h = 0 \\ \theta\sigma_v^2 & \text{si } h = 1 \\ 0 & \text{si } h > 1 \end{cases}$$

Mientras que su función de autocorrelación $\rho(h)$ queda de la siguiente forma:

$$\rho(h) = \begin{cases} \frac{\theta}{(\theta^2+1)} & \text{si } h = 1 \\ 0 & \text{si } h > 1 \end{cases}$$

Estos modelos tienen una representación infinita de un AR, para eso se necesita el modelo MA(1) que sea invertible y escribirlo como sigue:

$$v_t = -\theta v_{t-1} + e_t$$

Siguiendo de manera recursiva sustituyendo los valores de v_{t-i}

$$\begin{aligned} &= -\theta(-\theta v_{t-2} + e_{t-1}) + e_t \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ &= \sum_{i=0}^{\infty} (-\theta)^i e_{t-i} \end{aligned}$$

De la misma forma un AR(1) puede representarse como una serie infinita de un MA, es decir, existe una relación interesante aunque algunas veces cuando el orden autoregresivo es mayor dos es más común utilizar un MA(q) mientras que cuando es de orden 1 o 2 es mejor utilizar un AR(p).

Otro dato importante de los MA es su autocorrelograma, como se definió $\rho(h)$ anteriormente se puede observar cuando $h > 1$ esta vale 0, esto puede observarse en el autocorrelograma.

1.4.3. Modelo ARMA(p,q)

Considerando los procesos AR y MA es posible crear otro proceso combinando, estos dos representando el término e_t con términos de e_{t-j} anteriores y añadiendo también las variables v_{t-i} creando, así, un modelo autoregresivo de medias móviles ARMA(p,q) en donde p representa el orden de un AR(p) y q el orden de un MA(q).

$$e_t = \psi_1 e_{t-1} + \dots + \psi_p e_{t-p} + v_t + \theta_1 v_{t-1} + \dots + \theta_q v_{t-q}$$

Nótese que si $q = 0$ se obtiene un AR(p) y al contrario si $p = 0$ queda un MA(q). Un caso particular muy común es el modelo ARMA(1,1):

$$e_t = \psi_1 e_{t-1} + v_t + \theta_1 v_{t-1}$$

En este modelo en particular se puede calcular su función de autocorrelación como:

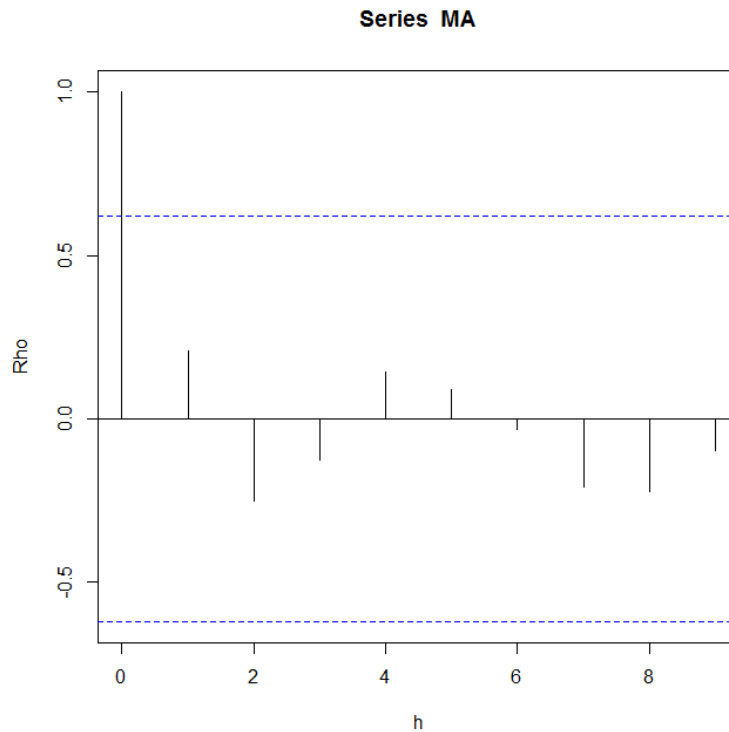


Figura 1.5: Correlograma MA(1) generado en R con ruido blanco y $\theta = 0.3$

$$\rho(h) = \begin{cases} \frac{(1+\psi_1\theta_1)(\psi_1+\theta_1)}{(1+\theta_1^2+2\psi_1\theta_1)} & \text{si } h = 1 \\ \psi_1\rho(h-1) & \text{si } h > 1 \end{cases}$$

1.5. Otras Pruebas de Autocorrelación

Uno esperaría poder probar si existen formas más generales de correlación, y no sólo la forma de un proceso autorregresivo de primer orden, es decir, poder tomar en cuenta los parámetros $\psi_2, \psi_3, \dots, \psi_p$ de un $AR(p)$ o de los parámetros $\theta_1, \theta_2, \dots, \theta_q$ de un $MA(q)$.

1.5.1. Prueba de Breusch-Godfrey

La prueba de Breusch-Godfrey utiliza los residuales de mínimos cuadrados ordinarios para probar la correlación entre los errores de los primeros h tiempos anteriores, a diferencia de la prueba de Durbin-Watson, esta también funciona en presencia de variables dependientes con retraso y se puede aplicar para modelos $AR(h)$ y $MA(h)$ o una mezcla de ellos.

Las hipótesis de esta prueba se definen de la siguiente forma a fin de probar la autocorrelación.

H_0 : No existe correlación serial de cualquier orden sobre h , $\theta_1 = \theta_2 = \dots = \theta_h = 0$

VS

H_A : Los errores provienen de un $AR(h)$ o un $MA(h)$

Esta prueba está basada en la prueba de Multiplicador de Lagrange aplicándola a la función de log-verosimilitud de la suma de cuadrados de los errores, para un modelo de regresión lineal. La estadística obtenida es nR^2 , la cual que se distribuye Chi-cuadrada con h grados de libertad.

Donde R^2 es el coeficiente de determinación:

$$R^2 = \frac{\sum_{i=1}^h \hat{e}_{t-i}^2}{\sum_{t=1}^N (Y_t - \bar{Y})^2}$$

y

$$n = N - h$$

La regla de decisión depende de si el valor obtenido nR^2 es muy grande, en ese caso se puede rechazar la hipótesis nula. Esta prueba se puede implementar en programas como R mediante el código `bgtest(datos, order=h)` el cual nos arroja el valor de la estadística y el p -value con el que se puede tomar una mejor decisión.

1.5.2. Prueba de Ljung-Box

Una prueba alternativa a la de Durbin-watson, que es un poco limitante por su aplicación a modelos autorregresivos, podría ser la de Box-Pierce, esta prueba tiene como finalidad probar si los términos de perturbación en el modelo se consideran ruido blanco, es decir, los errores son completamente aleatorios y cuya estadística se distribuye chi-cuadrada, pero posteriormente se probó que mediante una pequeña modificación a la estadística se podría obtener mejores resultados. Dando a conocer la prueba de Ljung-Box.

La prueba de hipótesis se presenta formalmente como:

$$H_0 : e_t \sim N(0, \sigma^2)$$

VS

$$H_A : e_t \text{ no es ruido blanco}$$

La estadística de prueba Q se define a continuación como:

$$Q = N(N + 2) \sum_{t=1}^h \frac{\hat{\psi}_t^2}{N - t} \sim \chi_h^2$$

Donde h es el orden del modelo autorregresivo AR(h) o de medias móviles MA(h) que se este probando, $\hat{\psi}_t$ es la autocorrelación estimada entre e_t y e_{t-h} .

$$\hat{\psi}_t = \frac{\sum_{t=h+1}^N \hat{e}_t \hat{e}_{t-h}}{\sum_{t=1}^N \hat{e}_t^2}$$

Mientras N sea lo suficientemente grande y se cumpla H_0 , la estadística Q se distribuirá chi-cuadrada con h grados de libertad.

Para la regla de decisión se toma en cuenta el cuantil $\chi_{h,\alpha}^2$ de la distribución chi-cuadrada con h grados de libertad al nivel de significancia α .

Por lo tanto, la regla de decisión se toma de la siguiente forma:

$$\text{si } Q < \chi_{h,\alpha}^2 \text{ NO se rechaza } H_0$$

$$\text{si } Q > \chi_{h,\alpha}^2 \text{ se rechaza } H_0$$

Nótese que para este tipo de pruebas es necesario tomar en cuenta el tamaño de la muestra para probar determinados número de retrasos, ya que entre mayor sea el retraso a probar, menor será el tamaño de los datos y como resultado una mayor variabilidad, por eso se propone que el retraso h no sea mayor a $N/4$ o $N/5$.

1.6. Proceso de Identificación

Una vez aplicado las pruebas anteriores y determinado si existe dependencia en el proceso, el problema se traduce a identificar a que proceso se ajustan mejor la muestra. Para esta tarea se pueden utilizar los residuales por mínimos cuadrados ordinarios para

calcular la función de autocorrelación, para ello se puede utilizar la función empírica $\hat{\psi}_t^2$ para compararla con las funciones de autocorrelación teóricas e identificar el modelo más apropiado.

1.6.1. Función de Autocorrelación Parcial (FAP)

Además de la función de autocorrelación se puede tomar en cuenta la función de autocorrelación parcial $\alpha(h)$ para realizar un mejor análisis, esta función se considera como la correlación entre e_1 y e_{h+1} ajustada por las observaciones intermedias e_2, e_3, \dots, e_h y se define de la siguiente forma:

$$\alpha(1) = \text{Corr}(e_{t+1}, e_t)$$

y

$$\alpha(h) = \text{Corr}(e_{h+1} - P_{sp\{1, e_2, \dots, e_h\}} e_{h+1}, e_1 - P_{sp\{1, e_2, \dots, e_h\}} e_1) \quad \text{para } h > 1$$

En donde $P_{sp\{1, e_2, \dots, e_h\}} e_{h+1}$ es la proyección de e_{h+1} sobre las demás e_i intermedias.

$$P_{sp\{1, e_2, \dots, e_h\}} e_{h+1} = \sum_{t=1}^h \alpha_t e_t \tag{1.13}$$

y las $\alpha_1, \alpha_2, \dots, \alpha_h$ satisfacen la siguiente igualdad:

$$\sum_{t=1}^h \alpha_t \mathbb{E}[e_i e_j] = \mathbb{E}[e_{h+1} e_j]$$

El teorema de proyección garantiza que la solución para las α_i existe y sustituyendo los valores de las α_i en la ecuación 1.13 esta se conoce como la mejor predicción lineal de e_{h+1} en términos de los demás e'_i s

A diferencia de la función de autocorrelación la FAP no puede ser estimada mediante una formula simple.

Una definición equivalente de la función de autocorrelación parcial es considerando un modelo de serie de tiempo estacionario e_t que no es necesariamente un proceso de AR, además se considera para cualquier valor fijo de k , las ecuaciones para la FAC de un proceso AR (p) que se expresan como:

$$\rho(j) = \sum_{i=1}^k \Phi_{1k} \rho(j-i)$$

con $j = 1, 2, \dots, k$

se pueden expresar de forma matricial como sigue:

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix} = \begin{pmatrix} \rho(0) & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & \rho(0) & \dots & \rho(k-2) \\ \vdots & \vdots & \dots & \vdots \\ \rho(k-1) & \rho(k-2) & \dots & \rho(0) \end{pmatrix} \begin{pmatrix} \Phi_{1k} \\ \Phi_{2k} \\ \vdots \\ \Phi_{kk} \end{pmatrix}$$

O expresado como:

$$\rho_k = \mathbb{P}_k \Phi_k$$

Así resolviendo para Φ_k

$$\Phi_k = \mathbb{P}_k^{-1} \rho_k \quad (1.14)$$

De esta forma la función de autocorrelación parcial estimada $\alpha(\hat{k})$ se define siempre que $e_j \neq e_i$ para alguna i y j , como:

$$\alpha(\hat{k}) = \hat{\Phi}_k$$

Donde $\hat{\Phi}_k$ es determinado por (1.14) con cada $\rho(j)$ reemplazada por la correspondiente autocorrelación estimada $\hat{\rho}(j)$.

También hay que tomar en cuenta que no existe un método para determinar exactamente el proceso por el que se generan los errores, pero, por medio de las funciones de autocorrelación y autocorrelación parcial es posible obtener información de gran apoyo para conocer mejor el comportamiento de los errores cuando no se tiene información a priori sobre el proceso. Identificando los patrones de autocorrelación que siguen los datos para seleccionar los modelos tentativos apropiados.

Para seleccionar un mejor modelo se toma en cuenta el modelo que mejor ajusta a las funciones de autocorrelación analizando si las autocorrelaciones son significativamente distintas de cero, para eso el error estándar de las funciones de autocorrelación $\rho(h)$ está dado por la siguiente expresión:

$$S_\rho = \sqrt{N^{-1}(1 + 2 \sum \hat{r}_k^2)}$$

Mientras que el error estándar de la función de autocorrelación parcial es estimado por la expresión:

$$S_{FAP} = \sqrt{N^{-1}}$$

Ya que se determine la significancia de la autocorrelación se puede identificar el proceso. Existen varias consideraciones para decidir el proceso adecuado, por lo que se pueden seguir las siguiente reglas.

a) Los residuales son aleatorios si $\hat{r}_k^2 = 0$ para toda k

b) Los residuales son generados por un AR(p) si la FAC va disminuyendo y la FAP tiene picos significativos en los p retrasos.

c) Los residuales son generados por un proceso MA(q) si la FAC tiene picos en los retrasos 1 hasta q y después disminuye completamente, mientras que la FAP decrece lentamente.

e) Los residuales son generados por un proceso ARMA(1,1) si las dos funciones decrecen exponencialmente después del primer retraso.

Puede ser difícil tomar una decisión, sobre todo cuando se trabaja con muestras pequeñas, ya que no puede realizarse pruebas para un proceso con cantidades grandes de retrasos, además de que para estos procesos se complica el cálculo de los estimadores. Actualmente existen programas como R en los que se puede realizar el calculo de los estimadores.

1.6.2. Cálculo de Otros estimadores

Una forma muy práctica de calcular los estimadores de modelos ARMA(p,q), es utilizando el lenguaje de programación R, este a su vez contiene un paquete para este tipo de análisis, el paquete *tseries*.

El paquete *tseries* contiene la función *arma* y dependiendo el modelo que se dese probar, esta devuelve los valores de los estimadores, con su grado de significancia y las funciones de autocorrelación y autocorrelación parcial mediante el código de la figura 1.6:

Ejemplo:

En donde la función *summary* regresa toda la información como se muestra en la figura 1.7:

En la columna *Estimate* se encuentran las estimaciones de ψ_1 — del AR(1), la estimación de θ_1 que es el estimador de la parte del MA(1) y el *intercept* sería el estimador de a si es que no se estuviesen analizando los residuales.

```

data(Y)
r <- diff(X)

summary(r.arma <- arma(r, order = c(1, 1)))

plot(r.arma)

```

Figura 1.6: Código en R para ajustar un ARMA(1,1)

```

Call:
arma(x = r, order = c(1, 1))

Model:
ARMA(1,1)

Residuals:
      Min       1Q   Median       3Q      Max
-1.5529214 -0.1189386  0.0001515  0.1337559  1.3663775

Coefficient(s):
      Estimate Std. Error t value Pr(>|t|)
ar1      -0.202320   0.074961  -2.699  0.00695 **
ma1       0.658111   0.056927  11.561 < 2e-16 ***
intercept 0.006799   0.018195   0.374  0.70866
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fit:
sigma^2 estimated as 0.06731,  Conditional Sum-of-Squares = 37.35,  AIC = 83.62

```

Figura 1.7: resumen del modelo ARMA(1,1) en R

Mientras que la función *plot* gráfica la serie, las funciones de autocorrelación y autocorrelación parcial (ACF Y PACF respectivamente por sus iniciales en Ingles).

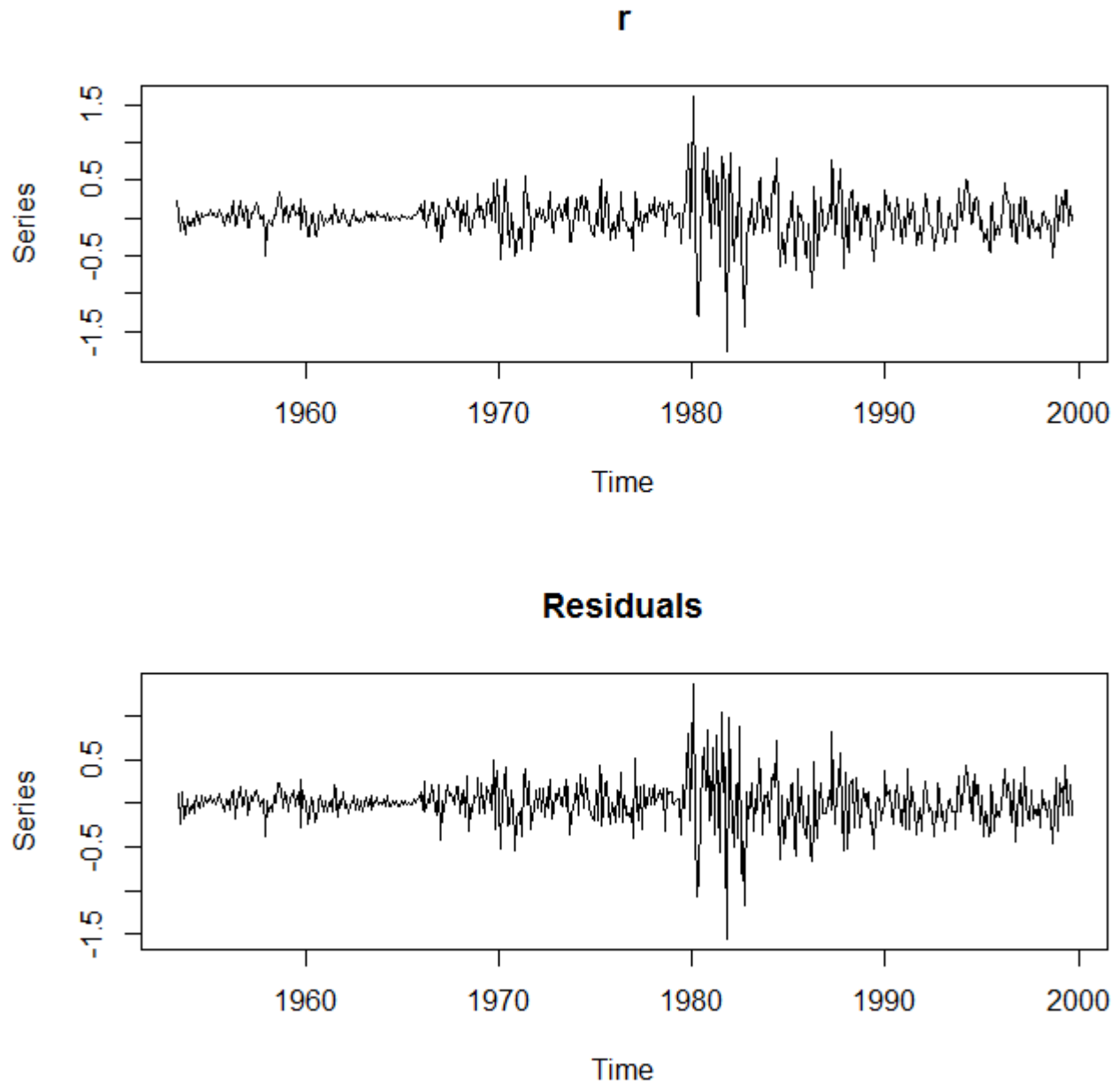


Figura 1.8: Serie que se desea ajustar

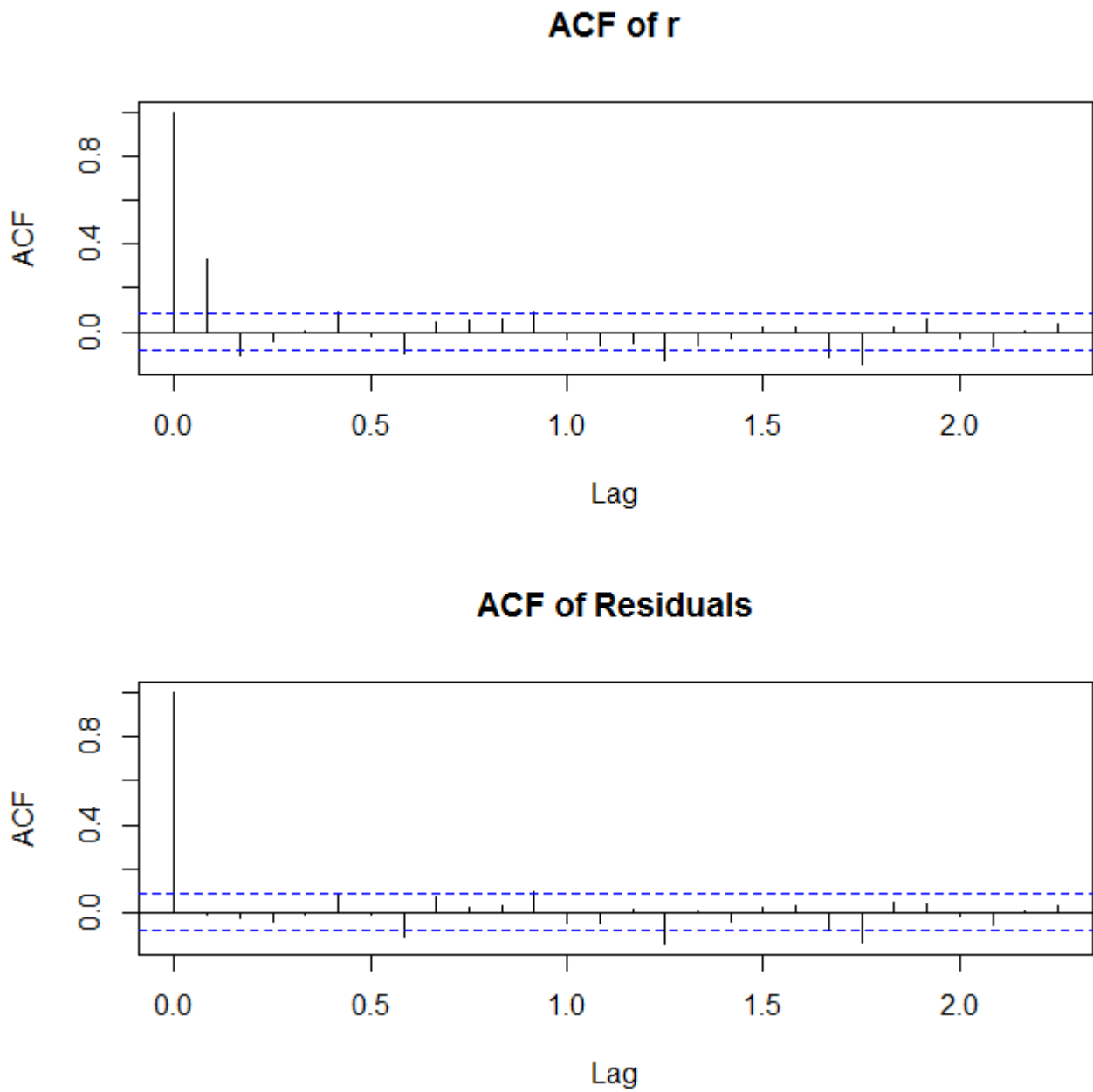


Figura 1.9: Funciones de autocorrelación

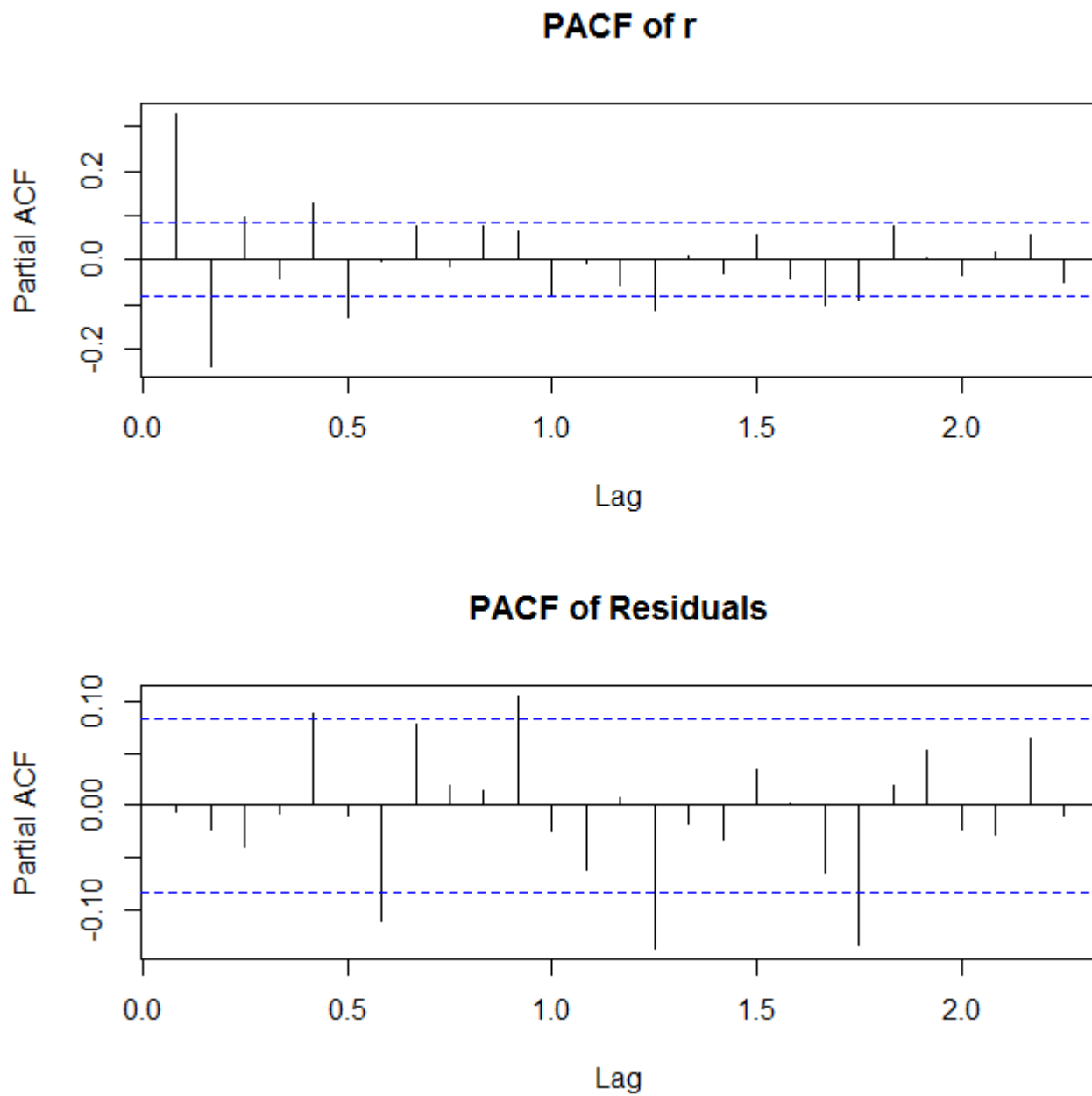


Figura 1.10: Funciones de Autocorrelación Parcial

1.7. Regresión Múltiple

Los modelos de regresión a menudo incluyen más de una variable predictora o regresiva. Si hay n predictores, el modelo de regresión lineal múltiple es la siguiente:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_k X_{tk} + e_t \quad (1.15)$$

Donde β_0, \dots, β_k son los coeficientes de regresión fijos, y e_t es un error aleatorio.

Si se siguen tomando todos los supuestos del modelo básico y además suponiendo no multicolinealidad y que haya menos cantidad de regresores que cantidad de observaciones, la fórmula de estimación cambia para reflejar la inclusión de variables explicativas adicionales. Existen $k + 1$ parámetros β_i . Para calcular dichos estimadores es común utilizar la notación matricial.

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (1.16)$$

Donde \mathbf{Y} es el vector de las observaciones en la variable dependiente Y_i , \mathbf{X} son las observaciones independientes $1, X_{t1}, X_{t2}, \dots, X_{tk}$ para cada t , β es el vector columna de los parámetros β_i y ϵ es el vector columna de los errores aleatorios e_t , con dimensiones $(N \times 1)$, $(N \times (k+1))$, $((k+1) \times 1)$ y $(N \times 1)$ respectivamente, y donde N es el tamaño de la muestra.

Es decir

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ 1 & X_{31} & \dots & X_{3k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & X_{N2} & \dots & X_{Nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix}$$

El método de mínimos cuadrados sigue siendo útil para calcular los estimadores de los coeficientes de regresión $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, es decir, el vector $\hat{\beta}$ tal que la suma de cuadrados de los residuales (SS_E) sea mínima.

Donde SS_E puede expresarse de la siguiente forma:

$$SS_E = \hat{\epsilon}'\hat{\epsilon} = \sum_{t=1}^N \hat{\epsilon}_t^2$$

Donde

$$\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$$

es decir, cada residual es de la forma:

$$\hat{\epsilon}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_{1t} - \hat{\beta}_2 X_{2t} - \dots - \hat{\beta}_k X_{kt}$$

por lo tanto para calcular los coeficientes de regresión se puede expresar la suma de cuadrados de residuales como:

$$\begin{aligned} SS_E &= \hat{\epsilon}'\hat{\epsilon} = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = (\mathbf{Y}' - \mathbf{X}'\hat{\beta}')(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\beta} - \mathbf{X}'\hat{\beta}'\mathbf{Y} + \mathbf{X}'\hat{\beta}'\mathbf{X}\hat{\beta} \end{aligned}$$

Para minimizar esta expresión al igual que en el caso del modelo simple se puede realizar derivando respecto a $\hat{\beta}$

$$\begin{aligned} \frac{\partial SS_E}{\partial \hat{\beta}} &= -\mathbf{Y}'\mathbf{X} - \mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned}$$

Igualando a cero y despejando:

$$\Rightarrow \mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\hat{\beta}$$

$$\Rightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Para que la solución exista es importante que la matriz inversa de $(\mathbf{X}'\mathbf{X})$ exista, es decir, que no haya una dependencia lineal entre las variables independientes.

Para considerar los supuestos de no autocorrelación se pueden seguir utilizando las mismas pruebas que para el proceso autoregresivo.

Además de calcular los estimadores de regresión, es necesario estimar la varianza del modelo σ^2 . para ello se puede demostrar que $\mathbb{E}[SS_E] = (N - k - 1)\sigma^2$. por lo que necesitaremos probar primero los siguientes resultados.

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{X}\beta + \epsilon]$$

$$\mathbb{E}[\mathbf{X}\beta] + \mathbb{E}[\epsilon] = \mathbf{X}\beta + \mathbb{E}[\epsilon]$$

Dado que se asume que los errores se distribuyen normal iid. con media cero y varianza constante entonces $\mathbb{E}[\epsilon] = 0$

$$\Rightarrow \mathbb{E}[\mathbf{Y}] = \mathbf{X}\beta$$

Para la varianza, de la misma forma se tiene que considerar que los errores se distribuyen normal iid. con media cero y varianza σ^2 , por lo que la $Var[\epsilon]$ es una matriz de varianzas y covarianzas, en donde la diagonal son las varianzas de las variables aleatorias, mientras que los elementos fuera de la diagonal son las covarianzas $Cov(e_i, e_j) = 0$.

Por lo tanto

$$Var(\mathbf{Y}) = Var(\mathbf{X}\beta + \epsilon) = Var(\epsilon) = \sigma^2 I_N$$

Teorema 1.1: Sea Y un vector aleatorio de tamaño N , sea A una matriz simétrica de tamaño $(N \times N)$. Si $\mathbb{E}[Y] = \mu$ y $Var(Y) = \Sigma$

Entonces,

$$\mathbb{E}[Y'AY] = tr(A\Sigma) + \mu' A \mu$$

Demostración:

$$\mathbb{E}[Y'AY] = \mathbb{E}[(Y - \mu)'A(Y - \mu) + \mu'AY + Y'A\mu - \mu'A\mu]$$

Al ser A simétrica $(Y'A\mu)' = \mu'AY$ y $\mathbb{E}[\mu'AY] = \mu'A\mathbb{E}[Y] = \mu'A\mu$

Entonces

$$\mathbb{E}[Y'AY] = \mathbb{E}[(Y - \mu)'A(Y - \mu)] + \mu'A\mu$$

$$= \sum_i \sum_j a_{ij} \mathbb{E}[(Y_i - \mu_i)'(Y_j - \mu_j)] + \mu'A\mu$$

$$\sum_i \sum_j a_{ij} \sigma_{ij} + \mu'A\mu$$

$$= tr(A\Sigma) + \mu'A\mu$$

■

Ahora, para encontrar un estimador insesgado para σ^2 partamos de SS_E , la cual se puede expresar de la siguiente forma:

$$SS_E = \hat{\epsilon}'\hat{\epsilon} = (Y - \hat{Y})'(Y - \hat{Y})$$

Sustituyendo $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y$

$$= (Y - X(X'X)^{-1}X'Y)'(Y - X(X'X)^{-1}X'Y)$$

Donde $\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' = H$ se conoce como la matriz sombrero, y cumple con las siguientes propiedades:

1. H y $(I - H)$ son matrices simétricas e idempotentes
2. $\text{rango}(I - H) = \text{tr}(I - H) = N - k - 1$
3. $(I - H)\mathbb{X} = 0$.

Entonces

$$\begin{aligned} SS_E &= \mathbb{Y}'(I - H)'(I - H)\mathbb{Y} \\ &= \mathbb{Y}'(I - H)^2\mathbb{Y} = \mathbb{Y}'(I - H)\mathbb{Y} \end{aligned}$$

Utilizando el teorema 1.1

$$\begin{aligned} \mathbb{E}[SS_E] &= \mathbb{E}[\mathbb{Y}'(I - H)\mathbb{Y}] \\ &= \text{tr}(\sigma^2(I - H)) + (\mathbb{X}\beta)'(I - H)\mathbb{X}\beta \end{aligned}$$

por la propiedad 3

$$= \text{tr}(\sigma^2(I - H))$$

$$= \sigma^2(N - k - 1)$$

$$\Rightarrow \mathbb{E}[SS_E] = \sigma^2(N - k - 1)$$

$$\Rightarrow \frac{SS_E}{(N - k - 1)} = \hat{\sigma}^2$$

En caso de utilizar Estimación de mínimos cuadrados generalizada, se estima la expresión 1.16 para obtener los residuales y de esta forma calcular la estimación de ψ y aplicar la siguiente transformación:

$$\begin{aligned} Y'_1 &= \sqrt{1 - \psi^2} Y_1 & t = 1 \\ X'_{1i} &= \sqrt{1 - \psi^2} X_{1i} & t = 1; i = 1, 2, \dots, k \\ Y'_t &= Y_t - \psi Y_{t-1} & t = 2, 3, \dots, N \\ X'_{ti} &= X_{ti} - \psi X_{t-1,i} & t = 2, 3, \dots, N; i = 1, 2, \dots, k \end{aligned}$$

Donde las estimaciones por el método de Estimación Generalizada de Mínimos Cuadrados se obtienen aplicando la regresión de Y'_t sobre las X'_{ti} .

1.7.1. Prueba de Significancia Global

Resulta importante probar si efectivamente existe una relación lineal entre las variables independientes y la variable dependiente, para ello se puede aplicar la siguiente prueba de hipótesis. Cabe mencionar que esta prueba considera los errores e_i ; en el modelo con distribución normal, independientes e idénticamente distribuidos con media cero y varianza σ^2 .

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_A &: \beta_j \neq 0 \text{ para alguna } j \in \{1, 2, \dots, k\} \end{aligned}$$

El procedimiento de la prueba implica un análisis de varianza, partiendo la suma total de cuadrados en la suma de cuadrados de regresión y la suma de cuadrados de residuales.

$$SS_N = \sum_{t=1}^N (Y_t - \bar{Y})^2 = \sum_{t=1}^N (\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^N (Y_t - \hat{Y}_t)^2 = SS_R + SS_E$$

donde las sumas de cuadrados se pueden escribir de la siguiente forma, utilizando la notación matricial:

$$SS_N = \mathbf{Y}'\mathbf{Y} - \frac{(\sum_{t=1}^N Y_i)^2}{N}$$

$$SS_R = \hat{\beta}'\mathbf{X}'\mathbf{Y} - \frac{(\sum_{t=1}^N Y_i)^2}{N}$$

$$SS_E = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$$

Ahora si la hipótesis nula es verdadera y los errores del modelo se distribuyen normal e independientes con la varianza constante, entonces:

$$\frac{SS_R}{\sigma^2} \sim \chi_{(k)}^2$$

$$\frac{SS_E}{\sigma^2} \sim \chi_{(N-k-1)}^2$$

Haciendo el cociente entre estas dos distribuciones y dividiendo cada una entre sus grados de libertad se obtiene la siguiente estadística:

$$\frac{\frac{SS_R}{k\sigma^2}}{\frac{SS_E}{(N-k-1)\sigma^2}} = \frac{SS_R/k}{SS_E/(N-k-1)} = F_0 \sim F_{(k, N-k-1)}$$

Se rechaza H_0 si se rebasa el cuantil superior F_α de la distribución F con k grados de libertad del numerador y N-k-1 grados de libertad del denominador.

Otra alternativa es utilizar el enfoque de P-Value para la prueba de hipótesis y, por lo tanto, rechazar la hipótesis nula si el valor P para la estadística F_0 es menor que α .

Este análisis de significancia generalmente se resume en el cuadro de análisis de varianza (ANOVA, cuadro 1.1), este casi siempre se realiza en un paquete software.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	Estadística F_0
Regresión	SS_R	k	$\frac{SS_R}{k}$	F_0
Errores	SS_E	$N - k - 1$	$\frac{SS_E}{(N - k - 1)}$	
Total	SS_N	$N - 1$		

Cuadro 1.1: Análisis de Varianza (ANOVA).

El lenguaje de R realiza este análisis al modelo de regresión que se trate con el siguiente código:

$$anova(prop.model)$$

```

R Console
> anova(prop.model)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
X       1  599.49   599.49  900.38 < 2.2e-16 ***
Residuals 28   18.64     0.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Figura 1.11: Análisis de varianza en R

1.7.2. Prueba de Significancia Individual

En ocasiones es de interés probar la hipótesis sobre los coeficientes de regresión de manera individual para determinar el valor o la contribución de cada variable predictora en el modelo de regresión, ya que, el modelo podría ser más efectivo con la inclusión de variables adicionales o quizás con la eliminación de una o más de las variables que ya están consideradas en el modelo.

Agregar una variable al modelo de regresión siempre causa que la suma de cuadrados de regresión SS_R aumente y la suma de cuadrados del error SS_E disminuya, por lo que, debemos decidir si el aumento en la suma de cuadrados de regresión es suficiente para garantizar el uso de la variable adicional en el modelo. Además, agregar una variable sin

importancia al modelo en realidad puede aumentar el error cuadrático medio, disminuyendo así la utilidad del modelo.

la prueba de hipótesis para probar algún coeficiente de regresión individualmente es la siguiente:

$$H_0 : \beta_j = 0$$

VS

$$H_A : \beta_j \neq 0$$

Si la hipótesis nula no se rechaza entonces quiere decir que la variable predictora X_j puede ser eliminada del modelo.

La estadística t_0 para esta prueba se define como a continuación:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

Donde C_{jj} es el elemento de la diagonal de la matriz $(\mathbb{X}'\mathbb{X})^{-1}$ correspondiente al elemento β_j .

Se rechaza H_0 si el valor absoluto de la estadística es mayor al cuantil $\alpha/2$ de la distribución t con $N - k - 1$ grados de libertad, es decir:

$$|t_0| > t_{\alpha/2, N-k-1}$$

1.7.3. Intervalos de Confianza

Una parte importante en el análisis de series de tiempo es construir intervalos de confianza para los estimadores de regresión, y así, tener un mejor panorama. El procedimiento para obtener estos intervalos de confianza requiere que supongamos que los errores del modelo se distribuyen normal e independientes con media cero y varianza σ^2 , los mismos supuestos hechos en las dos secciones anteriores sobre pruebas hipótesis.

El estimador de mínimos cuadrados $\hat{\beta}$ es una combinación lineal de las observaciones, por lo que se deduce que $\hat{\beta}$ se distribuye normal con el vector de esperanzas $\mathbb{E}[\hat{\beta}] = \beta$ y

la matriz de varianzas y covarianzas $Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Por lo tanto

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}} \sim N(0, 1)$$

sabemos que

$$\frac{SS_E}{\sigma^2} \sim \chi^2_{(k)}$$

entonces

$$\begin{aligned} & \frac{\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 C_{jj}}}}{\sqrt{\frac{SS_E}{\sigma^2(N - k - 1)}}} \\ &= \frac{\hat{\beta} - \beta}{\sqrt{\frac{SS_E}{(N - k - 1)} C_{jj}}} \\ &= \frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim t_{(N-k-1)} \end{aligned}$$

Por otro lado se define como el error estándar de $\hat{\beta}$ como $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2 C_{jj}}$

$$= \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 se(\hat{\beta})}} \sim t_{(N-k-1)}$$

Entonces se pueden calcular los intervalos con el $(1 - \alpha) \times 100\%$ de confianza.

$$\mathbb{P} \left[-t_{N-k-1}^{1-\frac{\alpha}{2}} \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 se(\hat{\beta})}} \leq t_{N-k-1}^{1-\frac{\alpha}{2}} \right] = 1 - \alpha$$

Donde $t_{N-k-1}^{1-\frac{\alpha}{2}}$ es el cuantil $1 - \frac{\alpha}{2}$ de una distribución t con $N - k - 1$ grados de libertad, por lo tanto, los intervalos para β_j son de la siguiente manera:

$$\hat{\beta}_j \pm t_{N-k-1}^{1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2 se(\hat{\beta})}$$

Capítulo 2

Análisis a Series de Tiempo con Retraso

En el capítulo anterior se estudió el caso de series de tiempo sin retraso en donde sólo se involucran variables exógenas que influyen en cada variable endógena, sin embargo, se puede sugerir evaluar modelos de diferentes tipos y en muchos casos se tiene que recurrir a utilizar modelos con más variables, incluyendo en el modelo variables tanto exógenas como endógenas de tiempos anteriores, ese es el caso de los modelos con retraso.

Este tipo de modelos pueden enfrentarse a problemas más complicados de estimación dada su estructura, que esencialmente se dividen en dos tipos ya sea por los retrasos de variables endógenas o por retrasos de variables exógenas.

2.1. Modelos con retrasos

En algunas ocasiones las variables que afectan a la variable dependiente Y_t , no sólo es X_t , sino también X_{t-1} , por lo que esta variable también tendría que considerarse en el modelo. Una representación del modelo sería de la forma siguiente:

$$Y_t = a + b_0X_t + b_1X_{t-1} + e_t \quad (2.1)$$

Considerando esta expresión se pueden calcular los estimadores, aplicar una prueba de autocorrelación y en caso de detectar una posible correlación significativa, aplicar las medidas necesarias para un mejor ajuste al modelo. El problema de este tipo de modelos es que usualmente no es posible especificar el retraso de tiempo en el que X_t afecta a Y_t , por lo que si no se tiene ninguna razón convincente para asumir un tiempo de respuesta dado, tal vez se debiera asumir que el retraso se distribuya en varios períodos de tiempo sucesivos, sobre k periodos con efecto parcial en cada periodo de retraso, es decir, de la

siguiente forma:

$$Y_t = a + b_0X_t + b_1X_{t-1} + \dots + b_kX_{t-k} + e_t \quad (2.2)$$

Este modelo se conoce comúnmente como *Modelo con retardos distribuidos* o en inglés *distributed lag model*. Para este modelo usualmente no se pueden aplicar directamente el calculo de los estimadores mediante mínimos cuadrados. El problema se presenta porque las variables independientes están fuertemente correlacionadas entre si, provocando así, problemas con la multicolinealidad, haciendo difícil separar el efecto de las variables independientes X_t sobre la variable dependiente Y_t .

Otro problema es cuando k es grande, muy cercano al valor total de la muestra N , se tendrían que calcular más estimadores, si además el tamaño de la muestra no es muy grande, se caería en un posible error de estimación, esto limitaría las herramientas para un correcto análisis.

Para mitigar estos problemas se ponen algunas restricciones a los porcentajes b_0, b_1, \dots, b_k de los retrasos. Un enfoque es que los valores de b_0, b_1, \dots, b_k deben de poder estimarse mediante un polinomio de grado r , tal que $r < k$

Otra opción sería imponer restricciones *a priori* en los parámetros b_0, b_1, \dots, b_k , como asumir que estos valores siguen un patrón de decremento exponencial, de la siguiente forma:

$$Y_t = a + b_0(X_t + cX_{t-1} + c^2X_{t-2} + \dots) + e_t \quad (2.3)$$

En donde claramente se reduce el número de parámetros y la dependencia de Y_t sobre las X_t se extiende indeterminadamente, aunque aún hay una gran cantidad de regresores, pero este detalle se puede tratar mediante la transformación de Koyck. Esta transformación consiste en utilizar la expresión 2.3 pero con Y_{t-1} , multiplicada por c y restándosela a la expresión 2.3, como se muestra a continuación.

$$cY_{t-1} = ca + b_0c(X_{t-1} + cX_{t-2} + c^2X_{t-3} + \dots) + ce_{t-1} \quad (2.4)$$

y

$$= ca + b_0(cX_{t-1} + c^2X_{t-2} + c^3X_{t-3} + \dots) + ce_{t-1}$$

$$\Rightarrow Y_t - cY_{t-1} = a - ac + b_0(X_t) + e_t - ce_{t-1}$$

$$\Rightarrow Y_t = a(1 - c) + b_0(X_t) + cY_{t-1} + e_t - ce_{t-1} \quad (2.5)$$

De esta manera se puede notar que en la expresión 2.5, los términos se encuentran serialmente correlacionados, aunque de esta también se tienen menos parámetros y variables dependientes, así, el modelo con retraso se puede expresar de dos formas, es decir, se puede pasar de la caracterización de retrasos distribuidos a un modelo con valores de variable de retraso de variables dependientes como del tipo de modelo siguiente:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 Y_{t-1} + e_t \quad (2.6)$$

Este modelo presenta una menor cantidad de problemas de estimación.

2.1.1. Retrasos de Variables Dependientes

Cuando se encuentran retrasos de variables dependientes en el modelo cambia el proceso de estimación, por lo que es importante analizar el modelo más sencillo, que es el siguiente:

$$y_t = \beta_0 y_{t-1} + e_t \quad (2.7)$$

Donde y_t es la desviación de la media, es decir, $y_t = Y_t - \bar{Y}$.

Los problemas y procedimientos de estimación dependen de los supuestos que se tomen sobre los errores e_t . Se pueden tomar dos conjuntos de supuestos de la siguiente manera:

1) los errores e_t consideran los supuestos de esperanza cero, varianza constante y que no estén autocorrelacionados, es decir:

a) Media zero: $\mathbb{E}[e_t] = 0$.

b) Varianza Constante: $\mathbb{E}[e_t^2] = \sigma^2$

c) No correlacionado: $\mathbb{E}[e_t e_{t-m}] = 0$ para $m \neq 0$

2) los errores se distribuyen de la forma: $e_t = \psi e_{t-1} + v_t$ con $-1 < \psi < 1$, y las variables v_t consideran los supuestos de esperanza cero, varianza constante y que no estén autocorrelacionadas.

a) Media zero: $\mathbb{E}[v_t] = 0$.

b) Varianza Constante: $\mathbb{E}[v_t^2] = \sigma_v^2$

c) No correlacionado: $\mathbb{E}[v_t v_{t-m}] = 0$ para $m \neq 0$

Es decir, los términos e_t siguen un proceso autorregresivo, por lo que no se consideran aleatorios.

CASO 1.

Dado que en este caso se considera que los errores no están correlacionados entre si, el problema se concentra en el uso de la variable dependiente y como variable explicativa. Al aplicar directamente en el modelo 2.7 los estimadores por mínimos cuadrados ordinarios, el estimador de β_0 queda de la forma:

$$\hat{\beta}_0 = \frac{\sum_t^N y_t y_{t-1}}{\sum_t^N y_{t-1}^2}$$

Tomando en cuenta que el modelo asume los siguientes supuestos

1. $\beta_0 < 1$
2. β_0 es pequeño en valor absoluto
3. y_t es una variable aleatoria
4. N es suficientemente grande

Entonces se puede demostrar que el valor esperado de $\hat{\beta}_0$ se puede expresar de la siguiente forma:

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 [1 - (2/N)]$$

De esta forma se puede notar que mientras N sea más grande, el estimador $\hat{\beta}_0$ tenderá a ser insesgado. de esta forma el modelo autorregresivo simple produce estimaciones con

sesgo aunque consistentes y relativamente eficientes del coeficiente de la variable dependiente con retraso, siempre y cuando se cumpla la condición de que los errores no sean correlacionados.

Para probar si los errores están correlacionados o no, la prueba de Durbin-Watson ya no es apropiada para este modelo, por lo tanto se requiere de otro test para probar la autocorrelación de los errores.

CASO 2.

Cuando los errores están autocorrelacionados y además se cuenta con retraso de la variable dependiente y , los estimadores por mínimos cuadrados ordinarios no son consistentes. Esto es porque la variable y_t depende directamente del error e_t , mientras que el valor de e_t se encuentra relacionado con e_{t-1} , lo que provoca que el valor de y_{t-1} este relacionado con e_t de la siguiente forma:

$$\begin{aligned} y_{t-1}e_t &= (\beta_0 y_{t-2} + e_{t-1})e_t \\ &= (\beta_0 y_{t-2} + e_{t-1})(\psi e_{t-1} + v_t) \\ &= \beta_0 \psi y_{t-2} e_{t-1} + \psi e_{t-1}^2 + \beta_0 y_{t-2} v_t + e_{t-1} v_t \end{aligned}$$

Por lo que la covarianza entre e_t y y_{t-1} se queda como sigue:

$$\begin{aligned} \mathbb{E}[y_{t-1}e_t] &= \mathbb{E}[\beta_0 \psi y_{t-2} e_{t-1} + \psi e_{t-1}^2 + \beta_0 y_{t-2} v_t + e_{t-1} v_t] \\ &= \beta_0 \psi \mathbb{E}[y_{t-2} e_{t-1}] + \psi \mathbb{E}[e_{t-1}^2] \\ &= \beta_0 \psi \mathbb{E}[y_{t-1} e_t] + \psi \mathbb{E}[e_t^2] \\ \Rightarrow \quad \mathbb{E}[y_{t-1} e_t] (1 - \beta_0 \psi) &= \psi \mathbb{E}[e_t^2] \end{aligned}$$

$$\Rightarrow \quad \mathbb{E}[y_{t-1}e_t] = \frac{\psi\sigma^2}{(1 - \beta_0\psi)}$$

De esta forma, bajo los supuestos del proceso autorregresivo de primer orden, esta covarianza es distinta de cero, lo que significa que la variable dependiente está correlacionada con los términos de perturbación, afectando así, a que los estimadores no sean consistentes.

Si se procede a calcular el estimador de β_0 del modelo que representa la ecuación 2.7, se caería en el error de omitir variables. Para darse cuenta de esto, se puede seguir el siguiente desarrollo (Hibbs 1974). Utilizando la ecuación 2.7, pero con y_{t-1} multiplicada por ψ se tiene lo siguiente:

$$\psi y_{t-1} = \psi(\beta_0 y_{t-2} + e_{t-1})$$

y después despejando ψe_{t-1}

$$\Rightarrow \quad \psi e_{t-1} = \psi y_{t-1} - \psi \beta_0 y_{t-2}$$

Por lo tanto, el modelo se puede escribir de la siguiente forma:

$$\begin{aligned} y_t &= \beta_0 y_{t-1} + e_t \\ &= \beta_0 y_{t-1} + \psi e_{t-1} + v_t \\ &= \beta_0 y_{t-1} + \psi y_{t-1} - \psi \beta_0 y_{t-2} + v_t \\ &= (\beta_0 + \psi) y_{t-1} - \psi \beta_0 y_{t-2} + v_t \end{aligned} \tag{2.8}$$

En este caso, el verdadero coeficiente de y_{t-1} es $(\beta_0 + \psi)$, mientras que el coeficiente de la variable omitida es $(-\beta_0\psi)$

Ahora, si se calcula el valor esperado del estimador calculado mediante mínimos cuadrados ordinarios del modelo de la ecuación 2.7, queda como sigue:

$$E[\hat{\beta}_0] = \beta_0 + \psi - \frac{\psi\beta_0 \sum_t^N y_t y_{t-1}}{\sum_t^N y_{t-1}^2}$$

Se puede demostrar, que este sesgo no se elimina a medida que el tamaño de la muestra se haga más grande. El sesgo asintótico, que es el sesgo en el estimador cuando el tamaño de la muestra se vuelve muy grande está dado por:

$$\frac{\psi(1 - \beta_0^2)}{1 + \psi\beta_0}$$

Por lo tanto, es probable sobrestimar el impacto de y_{t-1} cuando ψ es positiva. Además los residuales ya no proporcionan una aproximación precisa sobre las verdaderas perturbaciones porque los valores con retraso de y_t tienden a absorber el impacto sistemático de los disturbios. En resumen, en este caso, los residuales estarían sesgados y por ende la estadística de Durbin-Watson también. por lo tanto, es necesario desarrollar otras técnicas para medir la correlación y para calcular los estimadores.

2.1.1.1. Pruebas de Autocorrelación

Debido a los problemas de estimación se tiene que recurrir a otra prueba para medir la autocorrelación en presencia de retrasos por variables dependientes. Durbin presentó dos pruebas para este tipo de casos y una de estas es un caso particular de la prueba de Breusch-Godfrey.

Prueba- h de Durbin

Esta prueba no es muy utilizada, pero es fácil de aplicar cuando se usan los estimadores por mínimos cuadrados. Esta prueba consiste en identificar autocorrelación entre las variables dependientes utilizando la prueba de hipótesis:

H_0 : No existe autocorrelación

VS

H_A : Si existe autocorrelación entre las variables y_t

La estadística h , este esta definida mediante la siguiente expresión.

$$h = \hat{\psi} \sqrt{\frac{N}{1 - N\hat{v}\hat{a}r(\hat{\beta}_0)}}$$

Tomando en cuenta que la varianza de β_0 es la varianza muestral y N es el tamaño de la muestra. No importa cuantos valores con retraso de y_t se incluyan en una ecuación, se debe usar la varianza del coeficiente de y_1 , mientras que el estimador de ψ es calculado de la siguiente forma:

$$\hat{\psi} = 1 - \frac{d}{2}$$

Donde d es la estadística de la prueba de Durbin-Watson.

La estadística h se distribuye como una variable aleatoria normal estándar bajo la hipótesis nula de no correlación, por eso es necesario una muestra mayor a 30, de este modo la regla de decisión se define como sigue:

$$\text{si } h > Z_{.95} \quad \text{se rechaza } H_0$$

$$\text{si } h \leq Z_{.95} \quad \text{No se rechaza } H_0$$

Donde $Z_{.95}$ es el percentil .95 de la distribución normal estándar, es decir, $Z_{.95} = 1.64$

Otra Prueba Alternativa

Durbin sugirió otra prueba alternativa, la prueba- m . Esta prueba se utiliza cuando el valor de $N\hat{v}\hat{a}r(\hat{\beta}_0) > 1$, que generalmente se presenta cuando el tamaño de la muestra no es muy grande. Si se este caso se presenta en la prueba anterior, la estadística tendría una raíz de un número negativo.

Esta prueba consiste en generar la regresión por mínimos cuadrados ordinarios para calcular los residuales y correr la regresión de e_t con todas las variables de lado derecho

de la ecuación original más \hat{e}_{t-1} , es decir, la siguiente regresión:

$$\hat{e}_t = \psi_1 Y_{t-1} + \psi_2 X_t + \psi_3 \hat{e}_{t-1}$$

La correlación se puede medir con el P- value del coeficiente ψ_3 . La ventaja de esta prueba es que puede extenderse para contrastar con la hipótesis de correlación de orden superior incluyendo otros residuos con retraso y probando la hipótesis conjunta (Que todos los coeficientes ψ_i son iguales a cero). Sin embargo, el enfoque utilizado en la prueba m es similar a las funciones de autocorrelación y de autocorrelación parcial, y éstas se utilizan con más frecuencia, también se puede notar que esta prueba es un caso particular de la prueba de Breush-Godfrey.

De esta forma la prueba de Breush-Godfrey puede ser utilizada para medir la significancia de los coeficientes \hat{e}_{t-m} no solo para procesos autorregresivos AR(p), sino también para procesos de medias móviles MA(q).

2.1.1.2. Estimación de Parámetros

En el caso donde no se incluyen variables de retraso pero si con errores autocorrelacionados ya ha sido presentado, y este es aplicable a modelos con variables de retraso pero considerando el modelo siguiente:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + e_t \tag{2.9}$$

con

$$e_t = \psi e_{t-1} + v_t$$

Donde v_t asumen los supuestos de esperanza cero, varianza constante y que no estén autocorrelacionadas.

- a) Media zero: $\mathbb{E}[v_t] = 0$.
- b) Varianza Constante: $\mathbb{E}[v_t^2] = \sigma_v^2$
- c) No correlacionado: $\mathbb{E}[v_t v_{t-m}] = 0$ para $m \neq 0$

Además con ψ conocida se puede hacer la siguiente transformación a la expresión (2.9):

$$Y_t - \psi Y_{t-1} = \beta_0(1 - \psi) + \beta_1(X_t - \psi X_{t-1}) + \beta_2(Y_{t-1} - \psi Y_{t-2}) + v_t \quad (2.10)$$

Posteriormente se aplica directamente el método por mínimos cuadrados. Los estimadores resultarán sesgados pero serán consistentes y asintóticamente eficientes, siempre y cuando se conozca el valor de ψ , de lo contrario se tendrá que determinar la naturaleza del proceso generando los errores, calcular los estimadores y después transformar los datos, es decir, se puede aplicar la estimación generalizada de mínimos cuadrados.

Variabes Instrumentales (VI)

La presencia de variables Y_t con retraso significa que los métodos usuales para calcular los estimadores arrojarán estimadores inconsistentes, por esta razón la Estimación Generalizada de Mínimos Cuadrados se presenta en dos pasos. Primero, es necesario utilizar una técnica de variables instrumentales para generar estimaciones consistentes en los residuales. Después los residuales estimados son utilizados para identificar el modelo de los errores y determinar los estimadores $\hat{\psi}_i$.

En general se presentan dos problemas con respecto a la ecuación (2.9), uno de ellos es que Y_{t-1} está correlacionada con el término de error, el otro problema es que e_t está autocorrelacionado. La técnica de variables instrumentales resuelve el segundo problema, y por lo tanto se obtienen estimadores de regresión consistentes, el problema radica en que Y_{t-1} está correlacionada con e_t . El método de variables instrumentales se puede aplicar cuando se puede encontrar una nueva variable Z_t que no esté correlacionada con el término de error e_t , pero correlacionada con la variable Y_t .

En este caso suele utilizarse a Z_t como X_{t-1} , ya que claramente este no está correlacionado con el término de error pero usualmente si está correlacionada con la variable Y_t .

Mínimos Cuadrados en Dos Etapas

Una forma de utilizar las variables instrumentales es mediante mínimos cuadrados en dos etapas (TSLS - Two stage least squares). ya que la variable Z_t por uno de los supuestos, está correlacionada con la variable Y_t , se puede realizar la siguiente regresión:

$$Y_t = \pi_0 + \pi_1 Z_t + v_t$$

Sustituyendo Z_t y mediante mínimos cuadrados se pueden calcular los estimadores para π_0 y π_1

$$\hat{Y}_t = \hat{\pi}_0 + \hat{\pi}_1 X_{t-1}$$

Una vez obtenida la estimación de \hat{Y}_t como se muestra en la ecuación anterior, en la segunda etapa se procede a calcular los estimadores de la expresión (2.9) pero con la diferencia de que se utiliza la estimación \hat{Y}_t como se muestra a continuación:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 \hat{Y}_{t-1} + e_t$$

Las variables instrumentales provocarán que se obtengan estimadores consistentes, ya que las variables X_t y \hat{Y}_t dejan de estar correlacionadas con el término de error e_t .

Los estimadores siguen siendo ineficientes ya que los residuales están autocorrelacionados. La característica clave de las estimaciones por variables instrumentales es que proporciona estimadores consistentes de los residuales, así al utilizar estos residuales es posible identificar el comportamiento del proceso dependiente del tiempo subyacente al término de error mediante las funciones de Autocorrelación (FAC y FACP) de los residuales obtenidos.

Una vez identificado el orden del proceso $AR(p)$ o del proceso $MA(q)$ es necesario desarrollar un estimador inicial apropiado para ψ 's o θ 's, esto se puede lograr, en términos de un $AR(1)$, estimando ψ de la siguiente expresión:

$$\hat{e}_t = \psi \hat{e}_{t-1} + v_t$$

El método de variables instrumentales (VI) proporciona un lugar conveniente para comenzar un algoritmo de estimación iterativo. En este sentido, si se opta por utilizar una técnica iterativa de Cochrane-Orcutt, se indica que puede conducir a estimadores inconsistentes si comienza su búsqueda en $\psi = 0$, ya que puede converger en un máximo local. Para garantizar la coherencia es recomendable comenzar con el estimador $\hat{\psi}$ obtenido por el método de VI.

A diferencia de los modelos sin retraso que son relativamente sencillos de interpretar, en los modelos con retrasos se llega a complicar un poco más, por lo que una vez que se haya ajustado un modelo y calculado sus estimadores, el problema ahora es interpretar como es que las variables influyen en el fenómeno observado.

En el modelo de la forma:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + e_t$$

El coeficiente de Y_{t-1} afecta en la forma en que los valores X_t influyan sobre Y_t . Para notar el comportamiento de estos valores se puede considerar el siguiente proceso.

Tomando en cuenta la expresión:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1}$$

Se puede describir de tal forma que no esté presente explícitamente el retraso Y_{t-1} , sustituyendo a cada Y_{t-i} como se muestra a continuación:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} \\ &= \beta_0 + \beta_1 X_t + \beta_2(\beta_0 + \beta_1 X_t + \beta_2 Y_{t-2}) \\ &= \beta_0(1 + \beta_2) + \beta_1(X_t + \beta_2 X_{t-1}) + \beta_2^2 Y_{t-2} \end{aligned}$$

Así sucesivamente y haciendo a α igual a todas las constantes se puede expresar de la siguiente forma:

$$Y_t = \alpha + \beta_1 \sum_{i=0}^t \beta_2^i X_{t-i}$$

Si para cada término $\beta_1 \beta_2^i$ se le asigna una constante γ_i , de tal forma que Y_t se exprese como a continuación:

$$\beta_1 \beta_2^i = \gamma_i$$

Entonces

$$Y_t = \alpha + \gamma_0 X_t + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \dots$$

Donde

γ_0 es el multiplicador de impacto.

γ_i para $i \neq 0$ son los multiplicadores intermedios, y

$\sum_{i=0}^{\infty} \gamma_i$ es el multiplicador acumulativo y dado que las γ_i están entre -1 y 1 por el supuesto de que los parámetros β_i también lo están, se puede expresar de la siguiente forma:

$$\begin{aligned} \sum_{i=0}^{\infty} \gamma_i &= \beta_0 \sum_{i=0}^{\infty} \beta_2^i \\ &= \beta_0 \left(\frac{1}{1 - \beta_2} \right) \end{aligned}$$

Con esta representación se puede obtener más información acerca del comportamiento de la serie que se esté evaluando. Si X se mantiene en algún nivel \bar{X} durante n periodos, entonces el valor de equilibrio de \bar{Y} se podrá ver como:

$$\bar{Y} = \alpha + \sum_{i=0}^{\infty} \gamma_i \bar{X}$$

Si después vuelve a cambiar la media \bar{X} en algún tiempo t y permanece constante, entonces la media \bar{Y} cambiará gradualmente a un nuevo equilibrio con cambios de la forma $(\gamma_0 \Delta X)$ a tiempo t , $(\gamma_1 \Delta X)$ a tiempo $t+1$, $(\gamma_2 \Delta X)$ a tiempo $t+2$ y así, sucesivamente

El impacto multiplicativo indica que tanto cambio habrá en \bar{Y} , mientras que los multiplicadores intermedios indican el impacto de los periodos subsecuentes y el multiplicador acumulativo representa el cambio total de una unidad de cambio en la variable exógena sobre la variable endógena.

Entonces es posible determinar cómo es que son los cambios que afectan al fenómeno que se esté observando, es decir, la forma en que afectan los cambios de las variables endógenas a la variable exógena.

Capítulo 3

Pronósticos

Una vez estudiado el comportamiento de la serie de tiempo y haber observado cual es el modelo que mejor explica el comportamiento de los datos a través del tiempo, lo natural es preguntarse cómo se comportará en un futuro próximo. Mediante el uso de técnicas de regresión es simple hacer pronósticos a partir de las observaciones de la muestra.

Los pronósticos realizados no sólo son útiles para hacer predicciones, sino también se pueden utilizar para evaluar el modelo y de esta forma saber si sus pronósticos pueden ser considerados precisos.

Un modelo de series de tiempo proporcionará pronósticos de nuevas observaciones futuras que se pueden contrastar con lo que realmente se observa. Si existe una buena relación, se argumentará que esto proporciona una verificación más convincente del modelo que el ajuste en la muestra.

Otra función es que mediante los pronósticos se puede hacer inferencia de cierto tipo de comportamientos. Ya que el modelo sea evaluado y viable para predecir se pueden realizar pronósticos condicionando sobre ciertos tipos de comportamientos y determinar las implicaciones de tales comportamientos.

Con el fin de utilizar pronósticos de esta manera, los datos involucrados deben dividirse en dos conjuntos; la "muestra" que conforma las N observaciones y la "pos-muestra" la cual está conformada por n observaciones.

Normalmente las series de tiempo no se dan en periodos cortos o puede que provengan de algún experimento en donde conseguir más datos para compararlos con los pronósticos sería un grave problema, por lo que parecería prudente separar algunos datos para evaluar el modelo resultante. Si el modelo pasa su evaluación de la prueba pronosticada, entonces las n observaciones guardadas deben ser incorporadas al conjunto total de observaciones para re-estimar el modelo y poder generar las estimaciones futuras. En caso de que el modelo no sea lo suficientemente bueno, se tendría que volver desde cero y considerar otro modelo.

Por esto es importante primero probar la precisión del modelo, por lo tanto, primero se generan los primeros n pronósticos para posteriormente compararlas con la n observaciones que no se incluyeron en las primeras N .

Es importante notar que aunque en los primeros n pronósticos sean suficientemente buenos no se puede afirmar que los pronósticos que se realicen con las $N+n$ sean también buenos, entonces antes de enfocarse en cómo se tiene que proyectar es necesario saber a qué tanto error se está expuesto.

3.1. Error de pronóstico

Por supuesto ninguna predicción es perfecta, todas están expuestas a un error, lo importante es que el error sea el mínimo posible, este error de pronóstico no es más que la diferencia entre el valor verdadero de la serie ("pos-muestra") y el pronóstico:

$$Y_{N+h} - \hat{Y}_{N+h}$$

Donde \hat{Y}_{N+h} es el pronóstico a h unidades de tiempo posteriores a N .

En el caso en que se presente en el "modelo de serie simple":

$$Y_t = a + bX_t + e_t$$

Utilizando la historia del comportamiento de la serie es posible extrapolar hacia el futuro. Observe en particular que el pronóstico óptimo de Y_{N+h} es el pronóstico que minimiza el valor esperado del error de pronóstico cuadrático

$$\mathbb{E}[(Y_{N+h} - \hat{Y}_{N+h})^2]$$

El pronóstico de h tiempos después que minimiza el valor esperado del error de pronóstico cuadrado, es simplemente la esperanza condicional de Y_{N+h} con respecto a los valores observados $(Y_1, Y_2, Y_3, \dots, Y_N)$ y por consecuencia también de (X_1, X_2, \dots, X_N)

$$\hat{Y}_{N+h} = \mathbb{E}[Y_{N+h} | Y_1, Y_2, Y_3, \dots, Y_N, X_1, X_2, \dots, X_N] \quad (3.1)$$

Que puede escribirse de manera más compacta como:

$$= \mathbb{E}[Y_{N+h}|\Omega_N] \quad (3.2)$$

En donde Ω_N representa toda la información observada a tiempo N

$$= \hat{a} + \hat{b}X_{N+h} \quad (3.3)$$

Este pronóstico óptimo se conoce como la Proyección Mínima del Error Cuadrado Medio. Además este pronóstico es insesgado porque

$$\mathbb{E}[Y_{N+h} - \hat{Y}_{N+h}] = \mathbb{E} \left[\mathbb{E}[Y_{N+h} - \hat{Y}_{N+h}|\Omega_N] \right] = \hat{Y}_{N+h} - \hat{Y}_{N+h} = 0$$

Este pronóstico es sólo una estimación de lo que podría ser en el futuro, aunque claramente este diferirá del valor real, de esta manera la diferencia se define como el error de pronóstico.

$$Y_{N+1} - \hat{Y}_{N+1} = (a + bX_{N+1} + e_{N+1}) - (\hat{a} + \hat{b}X_{N+1})$$

Este error se distribuye normal con media cero y varianza S_F^2 .

$$Var(Y_{N+1} - \hat{Y}_{N+1}) = Var(Y_{N+1} - (\hat{a} + \hat{b}X_{N+1}))$$

Ya que Y_{N+1} y \hat{Y}_{N+1} son normales porque son combinación lineal de normales, las varianzas se pueden separar.

$$= Var(Y_{N+1}) + Var(\hat{a} + \hat{b}X_{N+1})$$

sustituyendo a $\hat{a} = \bar{Y} - \hat{b}\bar{X}$

$$= \sigma^2 + Var(\bar{Y} - \hat{b}\bar{X} + \hat{b}X_{N+1})$$

$$= \sigma^2 + Var(\bar{Y} + \hat{b}(X_{N+1} - \bar{X}))$$

sustituyendo \hat{b}

$$= \sigma^2 + Var \left[\bar{Y} + \frac{\sum_{t=1}^N (X_t - \bar{X}) Y_t}{\sum_{t=1}^N (X_t - \bar{X})^2} (X_{N+1} - \bar{X}) \right]$$

$$= \sigma^2 + \sum_{t=1}^N \left[\frac{1}{N} + \frac{(X_t - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} (X_{N+1} - \bar{X}) \right]^2 Var(Y_t)$$

$$= \sigma^2 + \sum_{t=1}^N \left[\frac{1}{N^2} + 2 \frac{1}{N} \frac{(X_t - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} (X_{N+1} - \bar{X}) + \frac{(X_t - \bar{X})^2}{(\sum_{t=1}^N (X_t - \bar{X})^2)^2} (X_{N+1} - \bar{X})^2 \right] \sigma^2$$

$$= \sigma^2 + \left[\frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum_{t=1}^N (X_t - \bar{X})^2} \right] \sigma^2$$

ya que

$$\begin{aligned} \sum_{t=1}^N (X_t - \bar{X})(X_{N+1} - \bar{X}) &= \sum_{t=1}^N (X_t X_{N+1} - X_t \bar{X} - \bar{X} X_{N+1} + \bar{X}^2) \\ &= N \bar{X} X_{N+1} - N \bar{X} \bar{X} - N \bar{X} X_{N+1} + N \bar{X}^2 = 0 \end{aligned}$$

$$\therefore Var(Y_{N+1} - \hat{Y}_{N+1}) = \left[1 + \frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum_{t=1}^N (X_t - \bar{X})^2} \right] \sigma^2$$

Como σ^2 generalmente es desconocida, esta se sustituye por su estimador máximo verosímil, de esta forma la varianza queda expresada como a continuación:

$$S_F^2 = S^2 \left[1 + \frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum_t (X_t - \bar{X})^2} \right]$$

Donde S_F^2 es la varianza de pronóstico, S^2 es la varianza muestral. Se puede notar que entre mayor sea el tamaño de la muestra, una mayor dispersión de la variable explicativa y una menor distancia entre X_t y la media, menor será el error de pronóstico.

Por lo tanto, se puede decir que los el pronósticos son mejores si se hacen dentro del rango de experiencia, por lo que entre más lejos se quiera estimar de \bar{X} , menos realista será dicho pronóstico.

Dado que el error de pronóstico se distribuye normal con media cero y varianza σ^2 , es posible construir un intervalo de confianza de la siguiente forma:

$$\frac{(Y_{N+1} - \hat{Y}_{N+1}) - 0}{\sqrt{\left[1 + \frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum_{t=1}^N (X_t - \bar{X})^2} \right] \sigma^2}} \sim N(0, 1)$$

Por otro lado,

$$\frac{\sum_{t=1}^N (Y_t - \bar{Y})^2}{\sigma^2} \sim \chi_{(N-2)}^2$$

Haciendo una transformación con estas dos distribuciones para eliminar el término desconocido σ^2 como a continuación:

$$\frac{(Y_{N+1} - \hat{Y}_{N+1})}{\sqrt{\text{Var}(Y_{N+1} - \hat{Y}_{N+1})}} \sqrt{\frac{\sum_{t=1}^N (Y_t - \bar{Y})^2}{\sigma^2}} \sqrt{\frac{N-2}{N-2}}$$

$$\begin{aligned}
&= \frac{(Y_{N+1} - \hat{Y}_{N+1})}{\sqrt{1 + \frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum_{t=1}^N (X_t - \bar{X})^2}} \sqrt{\frac{\sum_{t=1}^N (Y_t - \bar{Y})^2}{T - 2}}} \\
&= \frac{(Y_{N+1} - \hat{Y}_{N+1})}{S_F} \sim t_{(N-2)}
\end{aligned}$$

$$\mathbb{P} \left[-t_{N-2} < \frac{(Y_{N+1} - \hat{Y}_{N+1})}{S_F} < t_{N-2} \right] = 1 - \alpha$$

Por lo que el intervalo de confianza al rededor del pronóstico que contendrá el verdadero valor de Y_{N+1} con el $(1-\alpha)*100\%$ de confianza, es el siguiente:

$$(\hat{Y}_{N+1} - t_{N-2}S_F, \hat{Y}_{N+1} + t_{N-2}S_F)$$

Donde t_{N-2} es el cuantil $\frac{\alpha}{2}$ de una distribución t con $(N - 2)$ grados de libertad.

Por lo tanto, es posible generar pronósticos. Estos pronósticos contendrán una cierta cantidad de errores procedentes de dos fuentes distintas. La magnitud del error puede ser estimada y los intervalos de pronóstico pueden ser construidos

Una vez construido el intervalo, se puede concentrar en generar y evaluar los el pronósticos.

3.2. Pronóstico

La mejor forma de hacer pronósticos es usando los estimadores por mínimos cuadrados:

$$Y_{N+h} = \hat{a} + \hat{b}X_{N+h} + e_{N+h}$$

Donde \hat{a} y \hat{b} son los mejores estimadores bajo nuestros supuestos, y e_{N+h} son los términos aleatorios con media cero, varianza constante e independientes, De esta forma Y_{N+h} es el mejor pronóstico lineal, es decir, cualquier otro pronóstico lineal de Y tiene

un error mayor de pronóstico.

Es importante notar que una excepción de este modelo lineal se produce cuando los términos de error no son serialmente independientes, ya que al haber una correlación serial, las predicciones serán afectadas y serían ineficientes con varianzas muy grandes. Cuando existe correlación, el problema de pronosticar se puede tratar como enseguida se muestra, si e_{N+i} está correlacionado con e_{N+h-1} , cualquier pronóstico que omita el error puede ser mejorado.

Por ejemplo utilizando el modelo simple:

$$Y_{N+h} = a + bX_{N+h} + e_{N+h} \quad (3.4)$$

con

$$e_{N+h} = \psi e_{N+h-1} + v_{N+h} \quad (3.5)$$

Donde v_{N+h} cumple con los supuestos usuales. Para pronosticar con términos de error serialmente correlacionados existen dos alternativas.

Una consiste en sustituir en cada periodo el termino de perturbación e_{N+h} , despejando de la ecuación (3.4)

$$e_{N+h} = Y_{N+h} - a - bX_{N+h}$$

por lo tanto, al sustituir el termino de perturbación e_{N+h-1} en e_{N+h} queda expresado de la siguiente forma:

$$e_{N+h} = \psi e_{N+h-1} + v_{N+h}$$

$$e_{N+h} = \psi(Y_{N+h-1} - a - bX_{N+h-1}) + v_{N+h}$$

De esta manera en la expresión(3.4) al sustituir el término e_{N+h} queda como a continuación:

$$Y_{N+h} = a + bX_{N+h} + e_{N+h}$$

$$\begin{aligned}
&= a + bX_{N+h} + \psi(Y_{N+h-1} - a - bX_{N+h-1}) + v_{N+h} \\
&= a(1 - \psi) + b(X_{N+h} - \psi X_{N+h-1}) + \psi Y_{N+h-1} + v_{N+h}
\end{aligned}$$

Por lo que para poder realizar los pronósticos, se puede asumir que v_{T+i} toma su valor esperado de la siguiente forma:

$$Y_{N+h} = a(1 - \psi) + b(X_{N+h} - \psi X_{N+h-1}) + \psi \hat{Y}_{N+h-1}$$

Nótese que además se está estimado la variable endógena, ya que usualmente no se conocen sus valores.

$$\hat{Y}_{N+h-1} = a(1 - \psi) + b(X_{N+h-1} - \psi X_{N+h-2}) + \psi \hat{Y}_{N+h-2}$$

Por lo tanto, adicionalmente de las fuentes habituales de error de pronóstico se tiene que considerar la variabilidad en Y_{N+i-j} , haciendo el cálculo del error de pronóstico más complicado. Por supuesto los problemas de estimar los pronósticos de las variables endógenas no se presenta cuando es a un periodo, ya que las variables anteriores son conocidas, el problema surge cuando se generan pronósticos a más de un periodo hacia el futuro.

La otra alternativa para generar los pronósticos es simplemente no tener en cuenta la correlación serial en el período posterior a la muestra, es decir:

$$Y_{N+h} = \hat{a} + \hat{b}X_{N+h}$$

Suponiendo que el término de error toma el valor esperado $e_{N+h} = 0$

El pronóstico del método anterior implica el uso de retrasos más largos que en el caso no correlacionado en serie, por lo que puede haber acumulación de errores debido a incorporar suposiciones sobre el comportamiento del término de error, por lo que no hay

evidencia de que algún método sea el mejor.

Por otro lado es razonable pensar que el primer método es mejor, ya que incluye una consideración sobre la correlación.

Capítulo 4

Aplicación

En esta sección se aplican los modelos propuestos en este documento mediante el uso del software R, los códigos utilizados se anexan al presente trabajo.

Los datos utilizados son los precios de la gasolina convencional en EUA, los datos fueron obtenidos de la pagina de U.S. Energy Information Administration:

https://www.eia.gov/oil_gas/petroleum/data_publications/wrgp/mogas_history.html

4.1. Caso 1

Para el primer caso se analizaron los precios de la gasolina en general del territorio americano, las cifras son del precio en dólares por galón de gasolina y los precios de la gasolina en la sección de *EastCoast*, lo que se plantea es cómo se comportan los precios del combustible en Estados Unidos considerando las variaciones de los precios en *EastCoast*.

Los datos se consideran quincenales a partir de Enero 2017 a Septiembre 2018, ya que en este intervalo de tiempo se comportan de forma creciente y lineal, por lo que es un excelente ejemplo para este modelo.

Para ello se propone el siguiente modelo lineal:

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad (4.1)$$

Donde

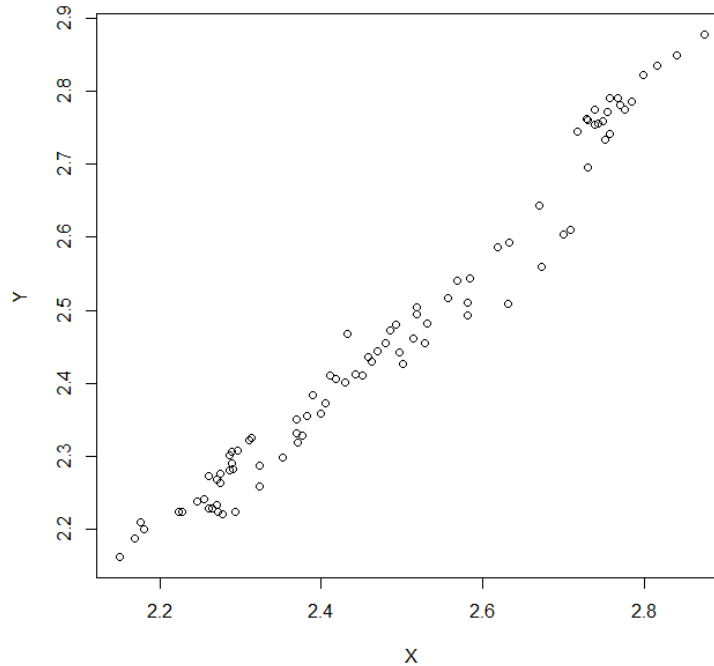
Y_t : Representa los precios del combustible en todo Estados Unidos.

X_t : Representa los precios de la gasolina únicamente en *EastCoast*.

e_t : son los términos de perturbación.

Este es el modelo de regresión simple al cual se le pueden obtener los estimadores de β_0 y β_1 mediante el método de mínimos cuadrados. Los resultados obtenidos son los siguientes:

Figura 4.1: Precios de la Gasolina en EUA



$$\hat{\beta}_0 = -0.03358$$

$$\hat{\beta}_1 = 1.00578$$

La recta ajustada a los datos mediante mínimos cuadrados se observa en la figura 4.2

Una vez ajustada la recta se debe analizar el comportamiento de los residuales para poder considerar los errores independientes o ajustar los errores a un ARMA, para ello se calcula la función de correlación y un correlograma.

La función de Autocorrelación (Figura 4.3) y el correlograma (Figura 4.4) sugieren que los residuales se correlacionan entre si con un proceso Autorregresivo de orden 4, ya que la FAC evaluada en 1,2,3 y 4 representa mayor correlación, en el correlograma la correlación va decreciendo gradualmente, a partir del quinto retraso la correlación deja de ser significativa.

Por otro lado, utilizando la función de autocorrelación parcial PACF (Figura 4.5) que elimina el efecto de los residuales intermedios entre e_t y e_{t-h} , indica que la correlación se debe al residual inmediato anterior e_{t-1} mientras que los demás residuales no representan

Figura 4.2: Recta ajustada a precios de la Gasolina en EUA

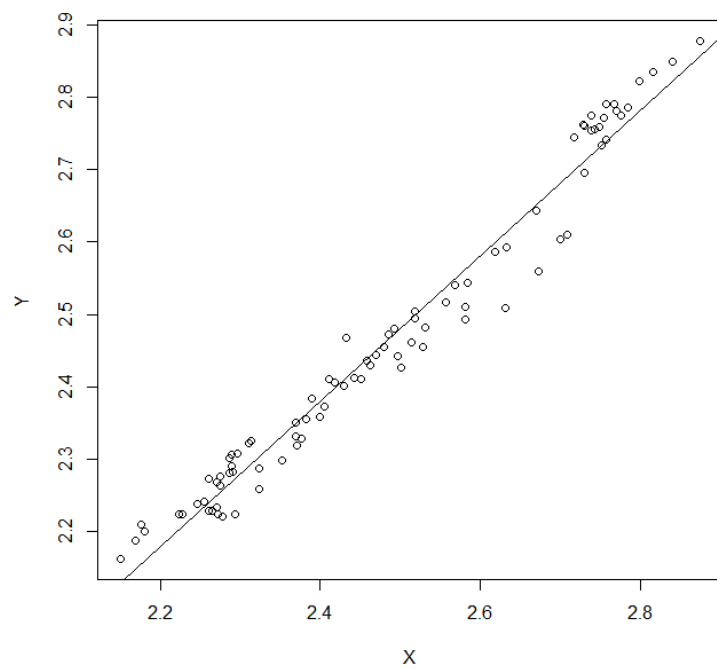


Figura 4.3: Función de Autocorrelación, FAC

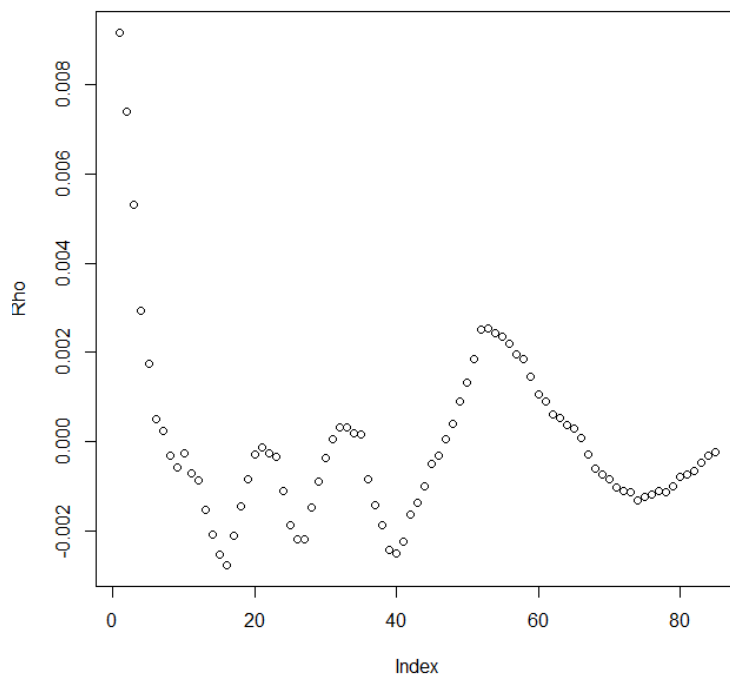
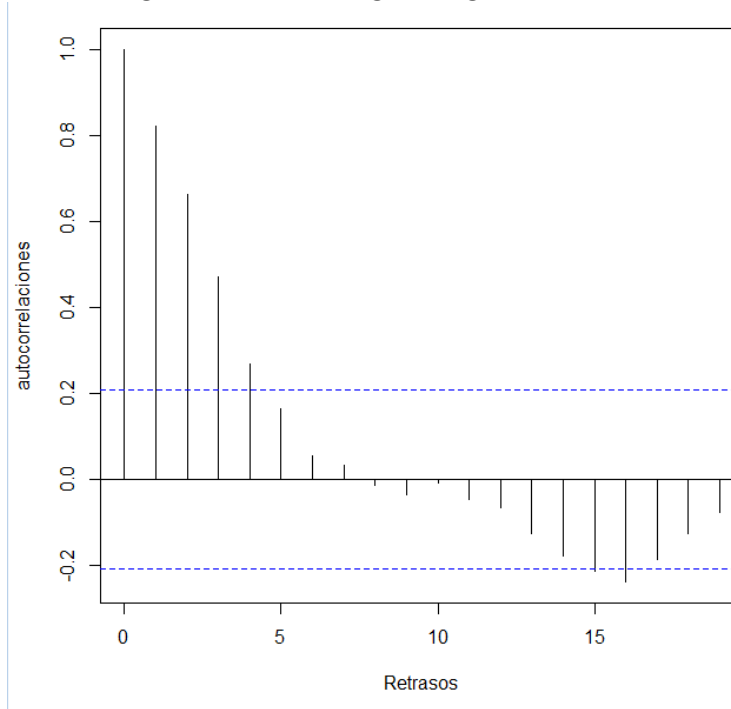


Figura 4.4: Correlograma generado en R



una correlación significativa.

La prueba de Durbin-Watson nos confirma que los residuales se encuentran correlacionados, programando la prueba en R la estadística d resulta ser:

$$d = 0.3416182$$

$$d_l = 1.65$$

$$d_u = 1.69$$

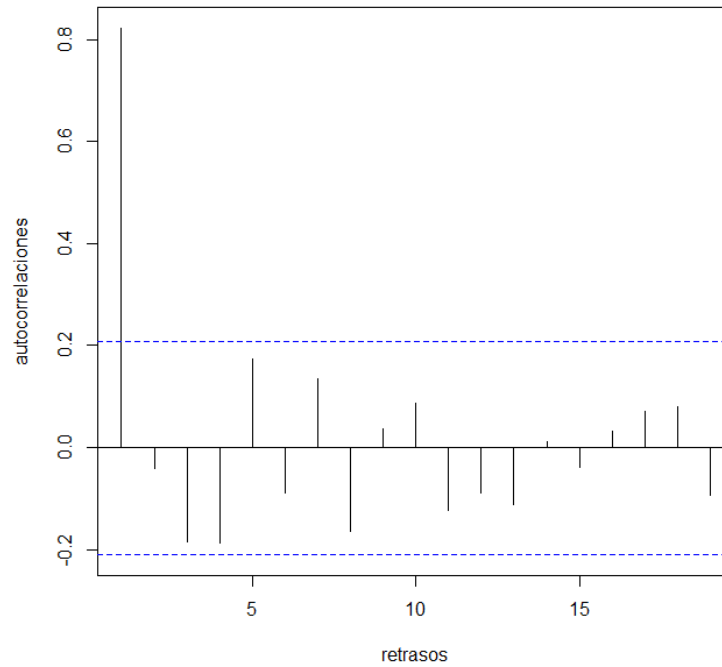
entonces $d < d_l$, por lo tanto se rechaza H_0 , es decir, si existe una correlación positiva entre los residuales.

Ahora la pregunta es a cual de los modelos ARMA se ajusta mejor, si las funciones de correlación nos indican que los residuales se comportan como un proceso autoregresivo de orden 4 (AR(4)) ya que el correlograma va decreciendo gradualmente hasta el retraso 5, agregando la información que nos arroja la PACF la mejor opción sería un proceso AR(1).

Si se considera un proceso autoregresivo AR(1) el modelo sería el siguiente:

$$e_t = \psi e_{t-1} + v_t$$

Figura 4.5: Función de Autocorrelación Parcial (PACF)



para calcular el estimador de ψ puede tratarse como un modelo de regresión lineal y utilizar el método de mínimos cuadrados, calculando el estimador en R;

$$\hat{\psi} = 0.8267$$

Para tratar de eliminar la correlación del primer retraso se puede utilizar el método de Conchrane-Orcutt.

$$Y'_t = Y_t - \Phi Y_{t-1}$$

En este caso $\hat{\Phi} = \hat{\psi} = 0.8267$

Realizando la transformación queda la siguiente expresión

$$Y'_t = a + bX'_t + v_t \quad (4.2)$$

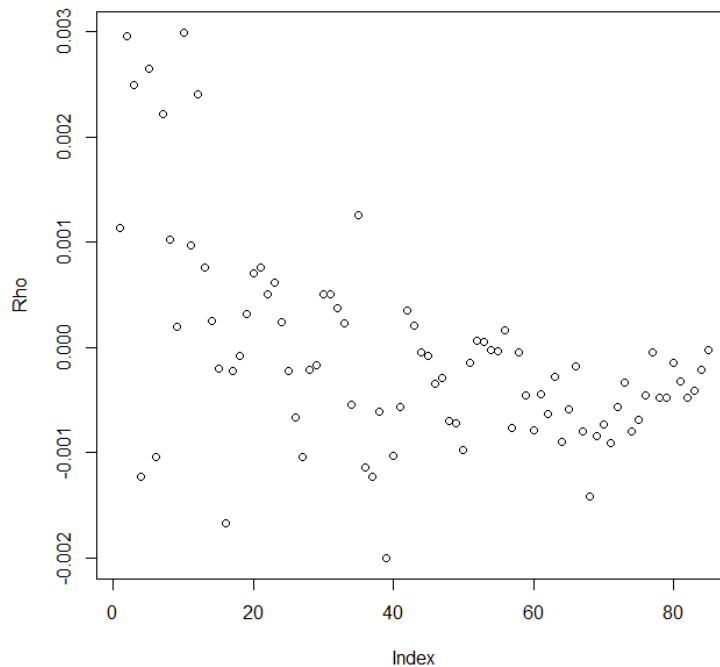
Donde $X'_t = X_t - \psi X_{t-1}$, $v_t = e_t - \psi e_{t-1}$ y los estimadores de a y b se calculan mediante mínimos cuadrados:

$$\hat{a} = 0.06119, \hat{b} = 0.85236$$

Para identificar si la correlación fue eliminada, pueden observarse las las funciones de autocorrelación y autocorrelación parcial.

Sin duda pude observarse que la correlación es menor, en el correlograma(Figura 4.7) después del primer retraso las correlaciones son menos significativas, sin embargo, en la función de autocorrelación parcial se observan picos en los retrasos 2, 4 y 7, para comprobar si estas observaciones son considerables o si se logró eliminar la correlación en los residuales, se debe de considerar la prueba Durbin-Watson.

Figura 4.6: Función de Autocorrelación



Haciendo la prueba de Durbin-Watson la estadística $d = 1.807009$, por lo que no se puede concluir con base a esta prueba, sin embargo las funciones de autocorrelación indican que la autocorrelación es positiva, es decir, que sigue habiendo una correlación entre los residuales.

A diferencia del primer caso donde no se considera la transformación Conchran-Orcutt, en este segundo caso se complica un poco más el identificar un $ARIMA(p,q)$ que se ajuste a los residuales, debido a que las funciones de autocorrelación y autocorrelación parcial descartan los $AR(P)$ y tampoco es evidente que provengan de un $MA(q)$, lo cual, en vez de simplificar el problema, lo agranda.

Figura 4.7: Correlograma en R

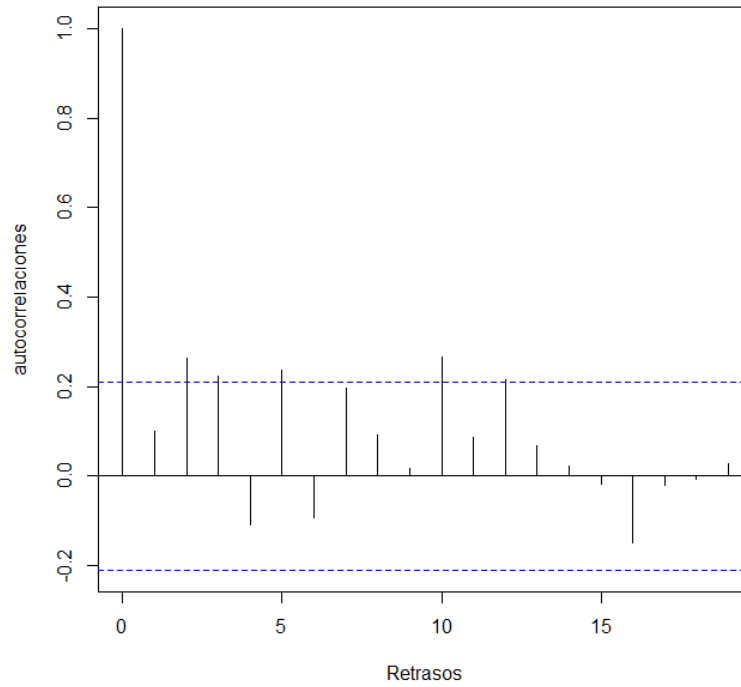
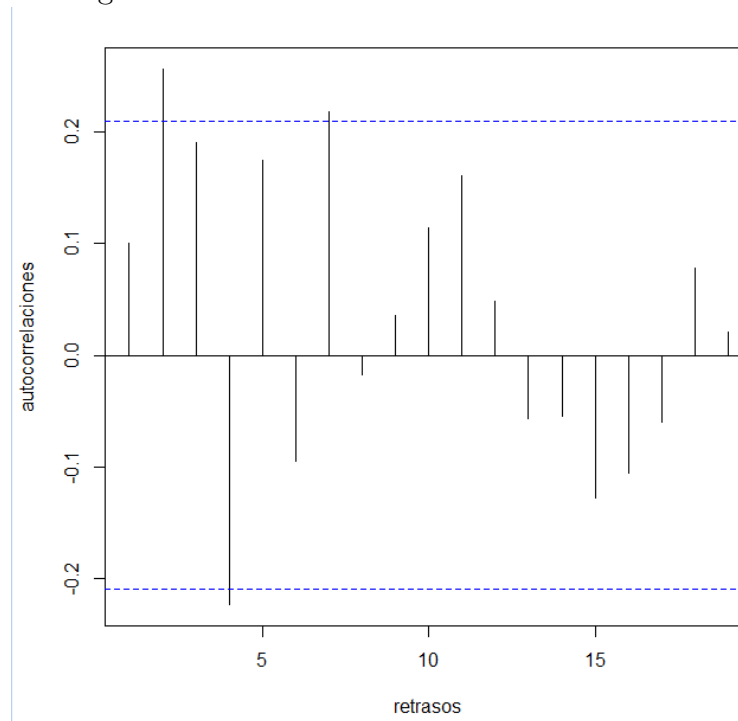


Figura 4.8: Función de Autocorrelación Parcial



Por tanto, en este caso la mejor opción es partir de la expresión 4.1 y considerar un AR(1) para el efecto por la correlación de los residuales y pronosticar las \hat{Y}_{N+h} , para ello se utiliza la siguiente expresión:

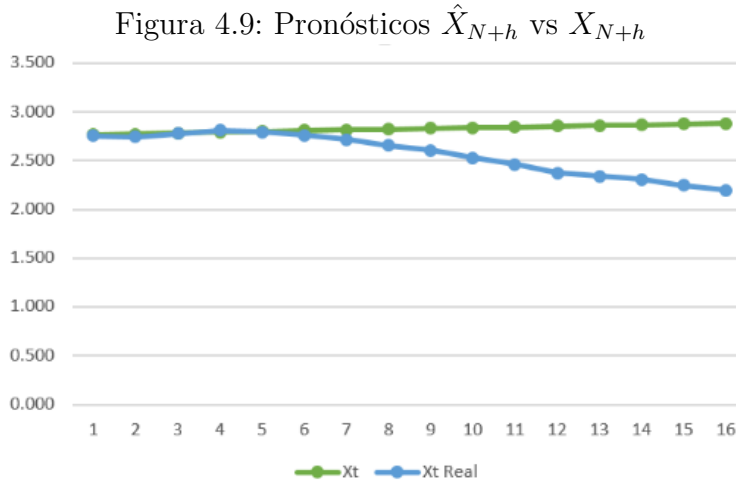
$$\begin{aligned}\hat{Y}_{N+h} &= \hat{\beta}_0 + \hat{\beta}_1 X_{N+h} + e_{N+h} \\ &= \hat{\beta}_0(1 - \hat{\psi}) + \hat{\beta}_1(X_{N+h} - \hat{\psi}X_{N+h-1}) + \hat{\psi}\hat{Y}_{N+h-1}\end{aligned}$$

Nótese que el pronóstico también depende de la variable X a tiempo $N+h$, por lo que esta también debería de estimarse, puede estimarse de la misma forma que la variable Y_t mediante mínimos cuadrados haciendo que X_t dependa del tiempo como se expresa a continuación:

$$X_t = a + bt + e$$

con $\hat{a} = 2.190466$ y $\hat{b} = 0.006608$

Calculando los residuales suponiendo un AR(1) se realizan los pronósticos para X_{N+h} , en la figura 4.9 se observa la comparación de los pronósticos vs los datos reales:

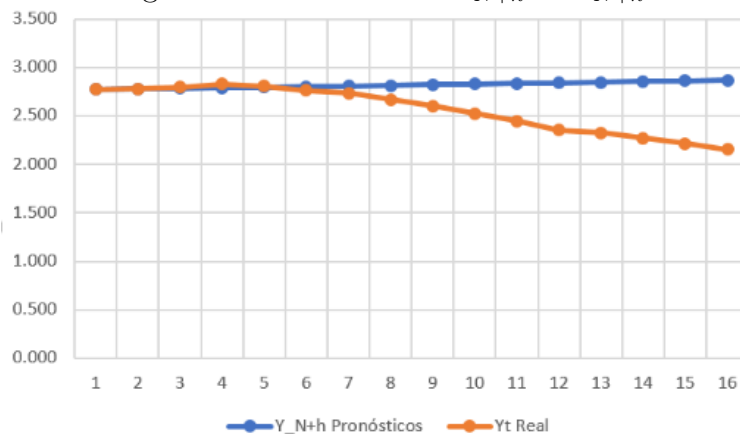


Ahora utilizando los pronósticos \hat{X}_{N+h} se pueden generar los pronósticos para Y_{N+h} de la siguiente forma:

$$\hat{Y}_{N+h} = \hat{\beta}_0(1 - \hat{\psi}) + \hat{\beta}_1(\hat{X}_{N+h} - \hat{\psi}\hat{X}_{N+h-1}) + \hat{\psi}\hat{Y}_{N+h-1}$$

Figura 4.10: Pronósticos \hat{Y}_{N+h} vs Y_{N+h}

Pronósticos			Precios Reales	
h	X_N+h	Y_N+h Pronósticos	Yt Real	variaciones
1	2.765957	2.770	2.771	-0.02%
2	2.775417	2.776	2.779	-0.10%
3	2.784454	2.782	2.796	-0.50%
4	2.793131	2.788	2.829	-1.44%
5	2.8015	2.794	2.806	-0.41%
6	2.809608	2.801	2.763	1.37%
7	2.817494	2.807	2.7332	2.71%
8	2.825189	2.814	2.67	5.39%
9	2.832724	2.820	2.604	8.31%
10	2.840121	2.827	2.525	11.96%
11	2.847401	2.834	2.445	15.89%
12	2.854581	2.840	2.35	20.86%
13	2.861676	2.847	2.325	22.45%
14	2.868699	2.854	2.268	25.82%
15	2.87566	2.860	2.216	29.07%
16	2.882569	2.867	2.156	32.98%

Figura 4.11: Pronósticos \hat{Y}_{N+h} vs Y_{N+h} 

Comparando los pronósticos \hat{Y}_{N+h} con los datos reales Y_{N+h} en la figura 4.11 se observa que los primeros pronósticos son muy parecidos a los datos reales, sin embargo, a partir de $h = 8$ los precios reales del combustible empiezan a disminuir, lo que provoca que la diferencia sea cada vez mas grande.

Debe notarse que $\hat{\beta}_1$ es practicante 1, esto quiere decir a que la variable Y_t es totalmente dependiente de X_t . Por cada unidad que cambie X_t , Y_t cambiara la misma unidad. Esto puede confirmarse observando las gráficas de las figuras 4.9 y 4.11, que son muy similares, solo que la gráfica 4.9 corresponde a el comportamiento de X_t y la gráfica 4.11 corresponde al comportamiento de las Y_t .

El caso de la expresión 4.2 donde se considera la transformación de Cochrane-Orcutt no eliminó la correlación entre los residuales, no obstante, aun quedan más casos a considerar, ya que se pueden identificar variables que afecten el precio general de la gasolina en Estados Unidos, tal vez se pueda mejorar el modelo y la estimación de los pronósticos.

4.2. Caso 2

Una forma para eliminar el error del modelo y de los pronósticos es identificar otras variables que contribuyan al comportamiento de la variable endógena que no se estén considerando y agregarlas al modelo.

En el primer caso se analizaron los precios de la gasolina en general del territorio americano, respecto de los precios dela gasolina en la sección de *EastCoast*, ahora se agregan los precios de *NewEngland*

Para ello se propone el siguiente modelo lineal:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + e_t \quad (4.3)$$

Donde

Y_t : Representa los precios del combustible en todo Estados Unidos.

$X_{1,t}$: Representa los precios de la gasolina en *EastCoast*.

$X_{2,t}$: Representa los precios de la gasolina en *NewEngland*.

e_t : son los términos de perturbación.

La única diferencia con el modelo de regresión simple, es que este contiene una variable adicional, por lo que se le pueden obtener los estimadores de β_0 , β_1 y β_2 mediante el método de mínimos cuadrados. Los resultados obtenidos son los siguientes:

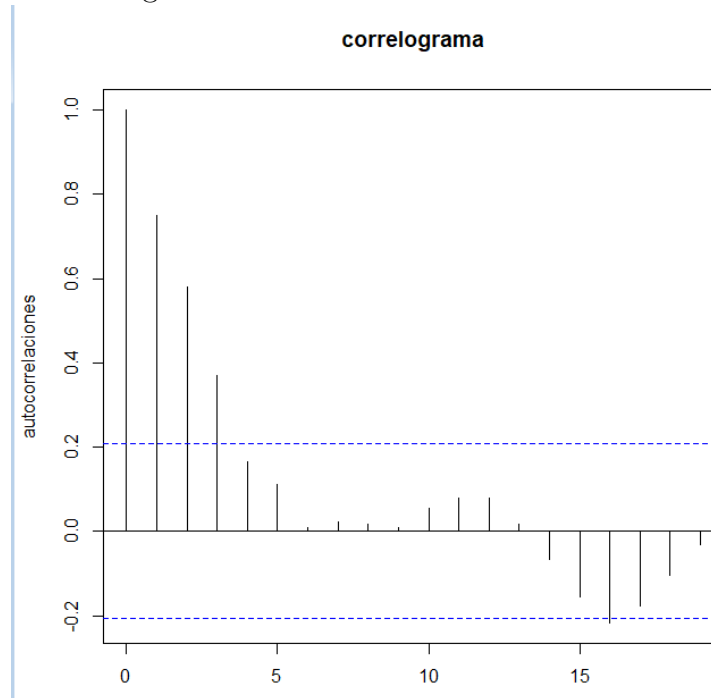
$$\hat{\beta}_0 = 0.04272$$

$$\hat{\beta}_1 = 0.32017$$

$$\hat{\beta}_2 = 0.63379$$

Una vez obtenidos los estimadores se pueden calcular los residuales y con ellos sus funciones de autocorrelación y autocorrelación parcial.

Figura 4.12: Función de Autocorrelación



Analizando las funciones de autocorrelación FAC (Figura 4.12) y autocorrelación Parcial FACP (Figura 4.13), se observa como la correlación no disminuye respecto al modelo de regresión simple, ya que la FAC decrece lentamente y en la FACP podemos confirmar que el retraso 1 es significativo, por lo que nos indica que un AR(1) sería el más adecuado para modelar los términos de error.

Si se realiza la prueba de Durbin-Watson y la de Breusch-Godfrey nos arrojan p-values del 0.6056 y de 0.7374 respectivamente, por lo que no se descarta que los términos de error se consideren como ruido blanco, por lo que se pueden desarrollar ambos casos; considerando que las e_t se comporten como un AR(1) y oro donde los términos de error no están correlacionadas.

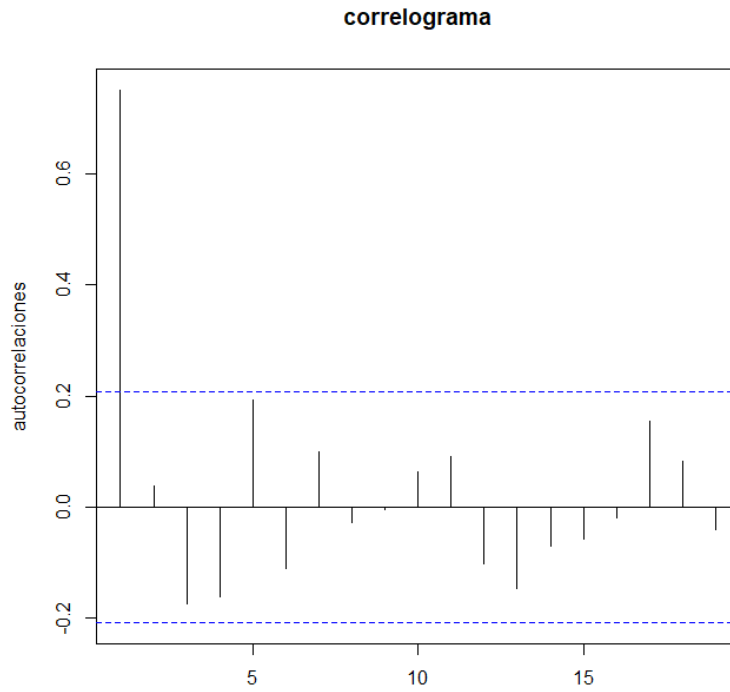
Caso 2.1: Considerando un AR(1).

$$e_t = \psi e_{t-1} + v_t \quad (4.4)$$

Calculando el estimador para ψ nos arroja $\hat{\psi} = 0.7518$

Para calcular ahora los pronósticos de Y_{N+h} considerando el comportamiento de los errores primero se tiene que observar que a partir de la expresión (4.3), el término de

Figura 4.13: Función de Autocorrelación Parcial



perturbación e_{N+h} puede expresarse como a continuación :

$$e_{N+h} = Y_{N+h} - \beta_0 - \beta_1 X_{1,N+h} - \beta_2 X_{2,N+h} \quad (4.5)$$

Si se sustituye de esta forma en la expresión 4.4 entonces e_t quedaría de la siguiente forma:

$$\begin{aligned} e_{N+h} &= \psi_1 e_{N+h-1} + v_{N+h} \\ &= \psi_1 (Y_{N+h-1} - \beta_0 - \beta_1 X_{1,N+h-1} - \beta_2 X_{2,N+h-1}) + v_{N+h} \end{aligned}$$

por lo tanto \hat{Y}_{N+h} se estimará de la siguiente forma:

$$\begin{aligned} \hat{Y}_{N+h} &= \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1,N+h} + \hat{\beta}_2 \hat{X}_{2,N+h} + e_{N+h} \\ &= \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1,N+h} + \hat{\beta}_2 \hat{X}_{2,N+h} + \hat{\psi}_1 (\hat{Y}_{N+h-1} - \hat{\beta}_0 - \hat{\beta}_1 \hat{X}_{1,N+h-1} - \hat{\beta}_2 \hat{X}_{2,N+h-1}) \end{aligned} \quad (4.6)$$

La expresión anterior depende de las variables endógenas y exógenas por lo que estas variables también deben de estimarse, en el caso anterior se estimó $X_{1,N+h}$, de la misma

forma se estima para $X_{2,N+h}$ haciendo:

$$X_{2,t} = a + bX_{2,t-1} + v_t$$

con $\hat{a} = 2.229248$ y $\hat{b} = 0.007592$ y considerando los términos de error como un AR(1) el estimador $\psi_{X_2} = 0.8692$.

Al realizar los pronósticos para $\hat{X}_{2,N+h}$ se obtienen los siguientes resultados (Figura 4.14 y 4.15):

Figura 4.14: Pronósticos $\hat{X}_{2,N+h}$ vs $X_{2,N+h}$

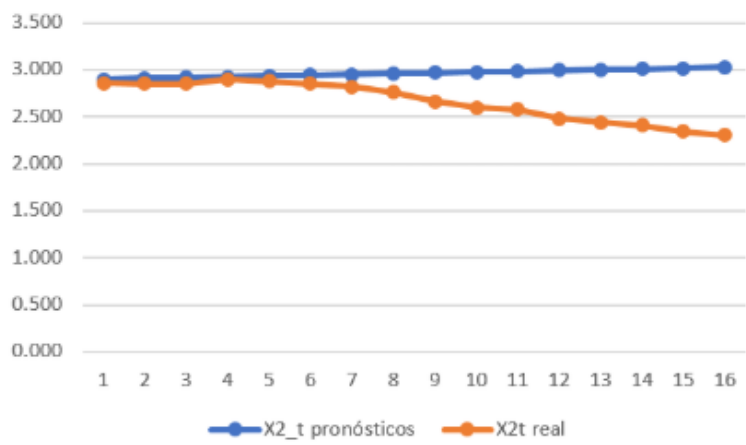
X2_t pronósticos	X2t real	variaciones
2.899	2.86	1.35%
2.908	2.851	2.00%
2.917	2.854	2.22%
2.926	2.895	1.08%
2.935	2.875	2.09%
2.944	2.849	3.32%
2.952	2.817	4.80%
2.960	2.76	7.26%
2.969	2.663	11.48%
2.977	2.597	14.63%
2.985	2.575	15.92%
2.993	2.479	20.74%
3.001	2.441	22.94%
3.009	2.408	24.96%
3.017	2.34	28.92%
3.025	2.302	31.39%

Utilizando la expresión 4.6 para generar los pronósticos de Y_{N+h} , se obtienen los siguientes resultados:

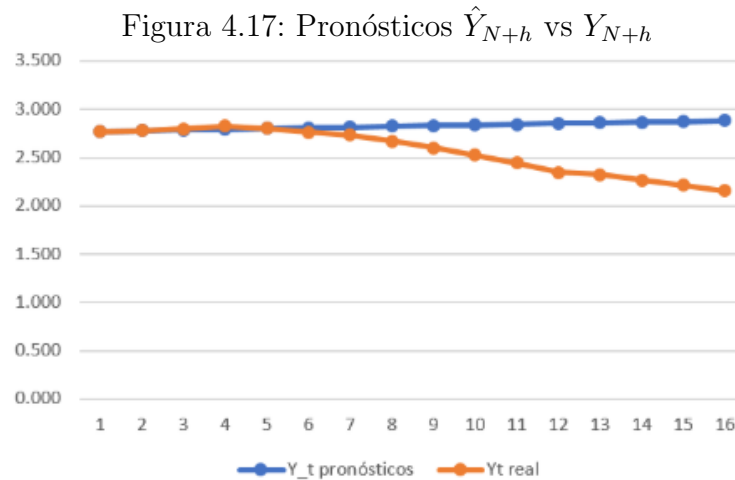
En este caso se puede observar que los pronósticos para Y_{N+h} se comportan de la misma forma que en el caso 1, aunque se presentan ligeramente más desviados respecto a las Y_t reales.

Caso 2.2 Por otro lado, se puede desarrollar el caso en el que los términos de error se comportan como ruido blanco, entonces los pronósticos para Y_{N+h} se pueden generar de la siguiente forma:

$$\hat{Y}_{N+h} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1,N+h} + \hat{\beta}_2 \hat{X}_{2,N+h} \quad (4.7)$$

Figura 4.15: Pronósticos $\hat{X}_{2,N+h}$ vs $X_{2,N+h}$ Figura 4.16: Pronósticos \hat{Y}_{N+h} vs Y_{N+h}

Y_t pronósticos	Yt real	variaciones
2.772	2.771	0.0%
2.779	2.779	0.0%
2.787	2.796	-0.3%
2.794	2.829	-1.2%
2.802	2.806	-0.1%
2.809	2.763	1.7%
2.817	2.732	3.1%
2.824	2.67	5.8%
2.832	2.604	8.8%
2.839	2.525	12.4%
2.847	2.445	16.4%
2.854	2.35	21.4%
2.861	2.325	23.1%
2.868	2.268	26.5%
2.876	2.216	29.8%
2.883	2.156	33.7%

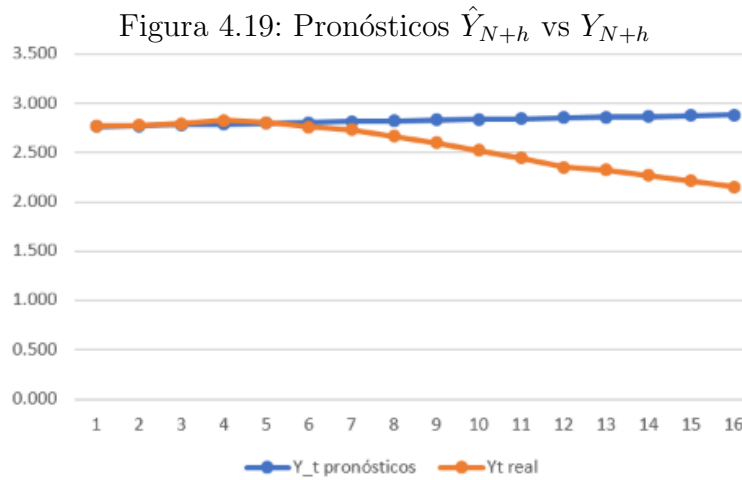


Utilizando los resultados previos para $\hat{X}_{1,N+h}$ y $\hat{X}_{2,N+h}$ los pronóstico obtenidos para \hat{Y}_{N+h} son los siguientes:

Figura 4.18: Pronósticos \hat{Y}_{N+h} vs Y_{N+h}

Y_t pronósticos	Yt real	variaciones
2.765	2.771	-0.20%
2.774	2.779	-0.16%
2.783	2.796	-0.46%
2.792	2.829	-1.32%
2.800	2.806	-0.22%
2.808	2.763	1.62%
2.816	2.732	3.07%
2.824	2.67	5.75%
2.831	2.604	8.73%
2.839	2.525	12.43%
2.846	2.445	16.41%
2.854	2.35	21.43%
2.861	2.325	23.05%
2.868	2.268	26.47%
2.875	2.216	29.76%
2.883	2.156	33.70%

Comparando los resultados, podemos concluir en este caso que el hecho de considerar un modelo para los términos de error no tiene una mejor aproximación que si no se considerase, ya que el comportamiento de los pronósticos fue el mismo, sin embargo, la correlación de los errores sigue estado presente, quizás se pueda mejorar la estimación si se modela como un proceso con retraso.



4.3. Caso 3

En los casos anteriores se logró identificar que los errores de los precios de la gasolina se encuentran autocorrelacionados, es decir cuentan con retrasos, los cuales se comportan como procesos auto-regresivos, sin embargo, la variable Y_t de igual forma puede tratarse como un proceso con retraso, es decir, puede estar correlacionada con las variables X_{t-i} por lo que se propone la siguiente expresión para Y_t :

$$Y_t = a + b_0X_{1,t} + d_0X_{2,t} + b_1X_{1,t-1} + d_1X_{2,t-1} + b_2X_{1,t-2} + d_2X_{1,t-2} \dots \quad (4.8)$$

El número de retrasos depende del tiempo en el que se ubique t y del tamaño de la muestra. Para reducir la cantidad de términos y para generalizar la expresión 4.6 podemos poner restricciones a priori en los parámetros b_i y d_i y aplicar la transformación de Koyck como en la expresión (2.5), por lo que se tendría la siguiente expresión:

$$Y_t = \beta_0 + \beta_1X_{1,t} + \beta_2X_{2,t} + \beta_3Y_{t-1} + e_t \quad (4.9)$$

Donde

Y_t : Representa los precios del combustible en todo Estados Unidos.

$X_{1,t}$: Representa los precios de la gasolina en *EastCoast*.

$X_{2,t}$: Representa los precios de la gasolina en *NewEngland*.

e_t : son los términos de perturbación.

Como puede observarse en la expresión (4.7) la variable Y_t se encuentra expresada con los términos Y_{t-1} , lo que afectaría los resultados si se procede a calcular los estimadores para β_0 , β_1 , β_2 y β_3 utilizando el método de mínimos cuadrados ordinarios, sin embargo, en esta situación tendremos que proceder calculando los estimadores utilizando el método de mínimos cuadrados en dos etapas, considerando variables instrumentales para eliminar la correlación entre Y_{t-1} y e_t .

Para implementar las variables instrumentales, se expresa a Y_t en términos de una variable que no este correlacionada con e_t pero si con Y_t como se muestra a continuación:

$$Y_t = \pi_1 + \pi_2 X_{t-1} + v_t \quad (4.10)$$

La expresión 4.8 es un modelo de regresión simple donde v_{t-1} es el término de perturbación, por lo que se puede utilizar el método de mínimos cuadrados para calcular las estimaciones de π_1 y π_2 .

$$\hat{\pi}_1 = 0.03619$$

$$\hat{\pi}_2 = 0.97962$$

Una vez obtenidos los estimadores de π_1 y π_2 se realizan las estimaciones para las \hat{Y}_{t-1} , las cuales se utilizarán en la expresión (4.7).

$$\hat{Y}_{t-1} = \hat{\pi}_1 + \hat{\pi}_2 X_{t-2}$$

Sustituyendo en la expresión 4.7:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 \hat{Y}_{t-1} + e_t \quad (4.11)$$

De esta forma se pueden calcular los estimadores de β_0 , β_1 , β_2 y β_3 por el método de mínimos cuadrados:

$$\hat{\beta}_0 = 0.06647$$

$$\hat{\beta}_1 = 0.31470$$

$$\hat{\beta}_2 = 0.68045$$

$$\hat{\beta}_3 = -0.05264$$

Al calcular \hat{Y}_t y con ello los residuales \hat{e}_t podemos observar las funciones de autocorrelación y autocorrelación parcial (figura 4.19 y 4.20) las cuales indican que los residuales se comportan como un proceso autorregresivo de orden AR(1), ya que la FAC va decreciendo lentamente y la FACP sólo tiene picos en 1.

$$e_t = \psi e_{t-1} + v_t$$

Figura 4.20: Función de Autocorrelación

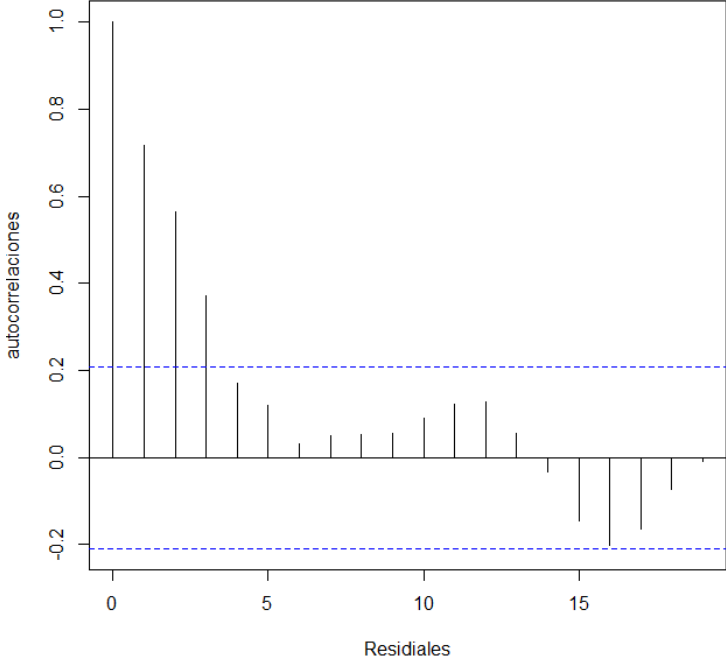
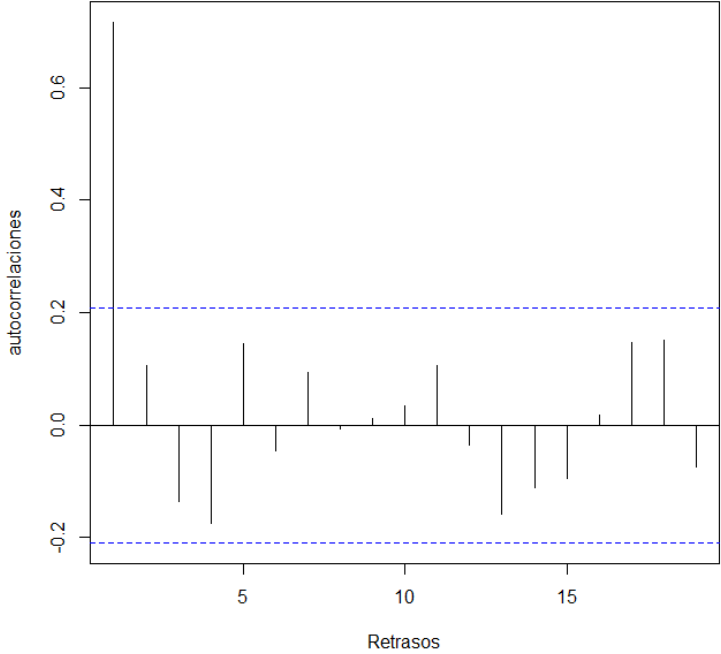


Figura 4.21: Función de Autocorrelación Parcial



Donde el estimador $\hat{\psi} = 0.74289$

Para comprobar si este modelo es viable, comprobamos mediante la prueba de Breusch-Godfrey, la cual nos arroja un P -value de 0.3089, por lo tanto no se rechaza H_0 es decir, no se tiene evidencia para asumir que e_t es un AR(1).

Por otro lado, otra opción es utilizar la expresión 4.7 considerando que los errores se distribuyen como ruido blanco, para pronosticar la variable Y_t de la siguiente forma:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1,t} + \hat{\beta}_2 \hat{X}_{2,t} + \hat{\beta}_3 \hat{Y}_{t-1} \quad (4.12)$$

donde $\hat{X}_{1,t}$ y $\hat{X}_{2,t}$ son estimadas mediante mínimos cuadrados, como en el caso anterior:

$$\hat{X}_{1,t} = a_1 + b_1 t + e$$

$$\hat{X}_{2,t} = a_2 + b_2 t + e$$

Al comparar los resultados de los pronósticos vs los datos reales de la variable $Y_t + h$ notamos que los resultados son muy parecidos a los de los casos anteriores, no obstante este modelo no considera el efecto de los elementos de perturbación, debido a que se consideraron como ruido blanco. Esto quiere decir que el hecho de considerar el precio del combustible en Estados Unidos se comporta como un modelo de regresión lineal con retraso, funciona para generar pronósticos sin necesidad de modelar los elementos de perturbación.

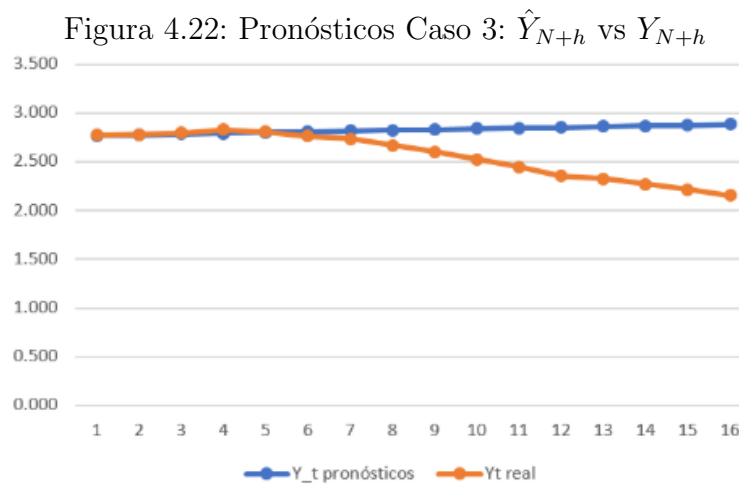


Figura 4.23: Pronósticos Caso3: \hat{Y}_{N+h} vs Y_{N+h}

Y_t pronósticos	Yt real	variaciones
2.766	2.771	-0.19%
2.775	2.779	-0.16%
2.783	2.796	-0.46%
2.792	2.829	-1.32%
2.800	2.806	-0.22%
2.808	2.763	1.62%
2.816	2.732	3.06%
2.823	2.67	5.74%
2.831	2.604	8.71%
2.838	2.525	12.41%
2.846	2.445	16.39%
2.853	2.35	21.41%
2.860	2.325	23.03%
2.868	2.268	26.44%
2.875	2.216	29.73%
2.882	2.156	33.67%

Conclusión

Se concluye que los precios del combustible de Estados Unidos, dado su comportamiento fue posible modelarlo mediante procesos de regresión lineal. En este documento se analizaron tres casos:

- El primero se ajustó un modelo de regresión lineal simple y un proceso autoregresivo de orden 1 AR(1) para los términos de perturbación
- El segundo se agregó una variable endógena $X_{2,t}$ la cual resultó ser significativa en el comportamiento de las Y_t , donde los cambios en $X_{2,t}$ resultan de mayor impacto que los de $X_{1,t}$ debido a que el estimador $\hat{\beta}_2$ resultó mayor que $\hat{\beta}_1$, sin embargo, estas dos series tuvieron el mismo comportamiento.
- Por último en el tercer caso se consideró como un modelo con retraso, es decir que también se consideraron en las variables endógenas de los tiempos anteriores X_{t-m} . En este caso se utilizó el método de mínimos cuadrados en dos etapas y no se descartó el caso de considerar que los términos de error son independientes.

En cada uno de estos casos los resultados fueron muy similares, tanto que prácticamente los pronósticos se podría decir que fueron los mismos, todos tuvieron las mismas tendencias, considerando o sin considerar los términos de perturbación, con o sin retraso. Esto quiere decir que a pesar de incluir otras variables, el modelo simple no pudo ser mejorado, por lo que se podría utilizar cualquier modelo. No se detectaron diferencias significativas para concluir lo contrario.

Si bien los pronósticos se van desviando de los datos reales, se logró identificar en los 3 casos, cómo es que Y_t es afectada por las variables endógenas. Si se conocieran los datos reales de $X_{1,t}$ y $X_{2,t}$ sabemos como los cambios en estas variables afectan a Y_t . En la figura 4.24 se puede apreciar como serían los pronósticos de Y_t con los precios reales de X_t . En cualquiera de los 3 casos la gráfica es similar.

Esto se debe justo a la dependencia lineal que existe entre los precios de la gasolina en todo el territorio americano y los precios por localidad, como se observa en el diagrama de dispersión la figura 4.25. misma que se presenta en el caso 1.

Por lo que si se desean hacer mejores pronósticos se requeriría modelar mejor las X_t respecto al tiempo, o identificar otras variables que justifiquen el cambio de tendencia

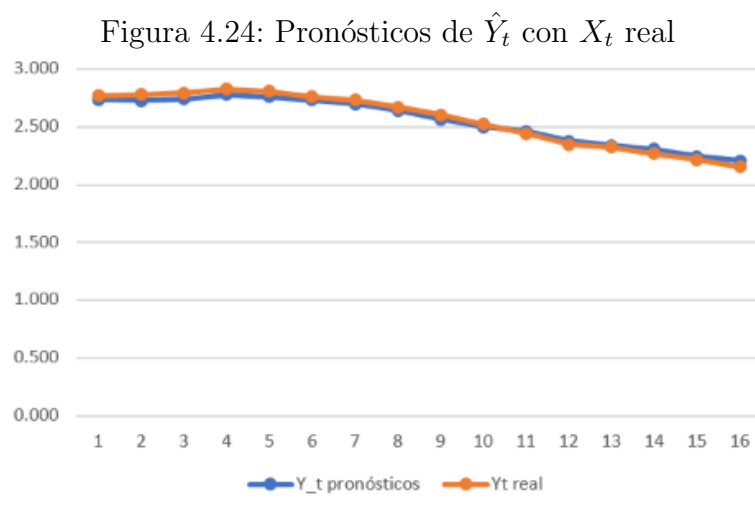
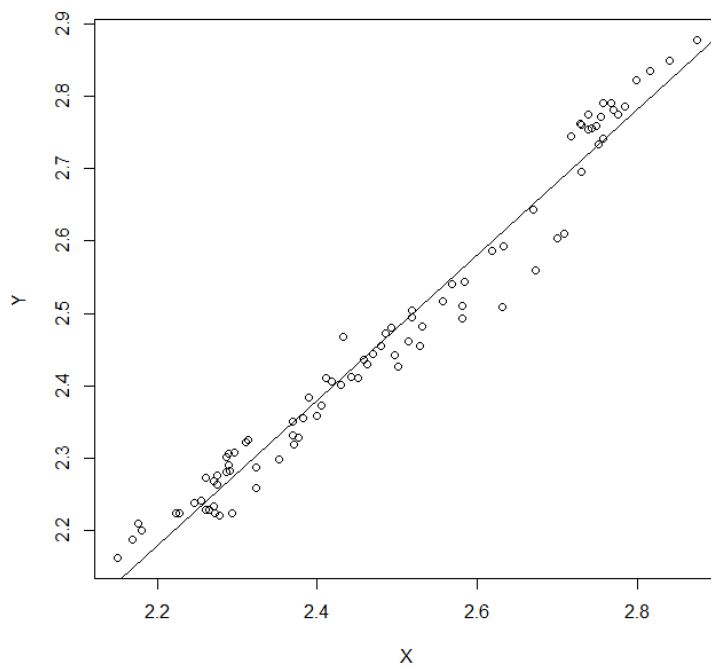
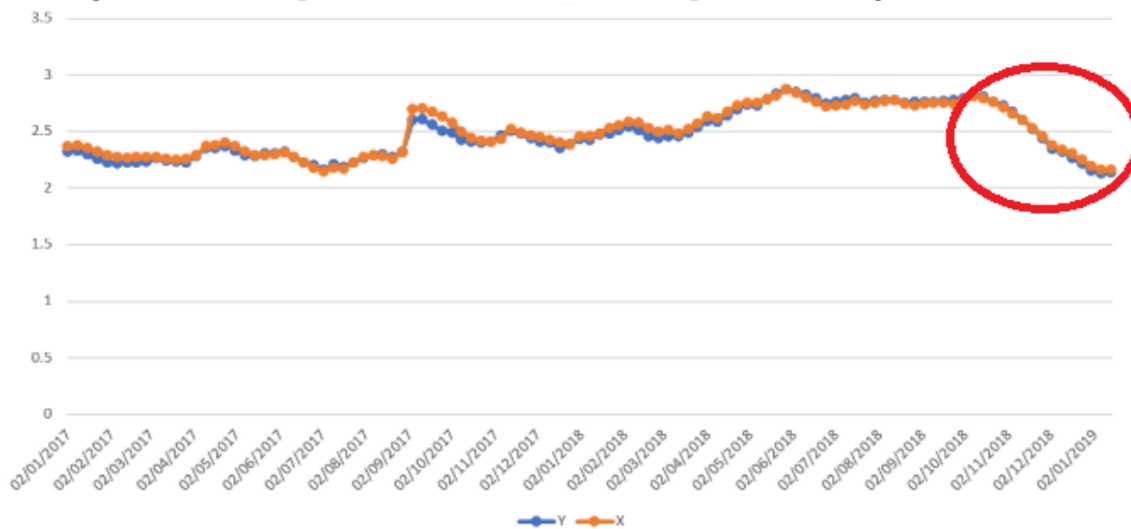


Figura 4.25: Diagrama de dispersión de los precios de la Gasolina en EUA y los precios de East Coast



que sufrieron los precios de la gasolina.

Figura 4.26: Comportamiento histórico de los precios de la gasolina en EU



Por otro lado, en este documento se analizaron los precios de un mismo país, sin embargo, se podrían utilizar los precios de otros países para identificar como se afectarían los cambios del combustible de un país a los precios de otro, o no solo combustibles sino otras variables que pudieran ser de interés.

Apéndice A

Apéndice: Tabla Durbin-Watson

Figura A.1: Tabla de Durbin-Watson

Sample Size	Probability in Lower Tail (Significance Level = α)	k = Number of Regressors (Excluding the Intercept)									
		1		2		3		4		5	
		d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	.01	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
	.025	.95	1.23	.83	1.40	.71	1.61	.59	1.84	.48	2.09
	.05	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
20	.01	.95	1.15	.86	1.27	.77	1.41	.63	1.57	.60	1.74
	.025	1.08	1.28	.99	1.41	.89	1.55	.79	1.70	.70	1.87
	.05	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99
25	.01	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
	.025	1.13	1.34	1.10	1.43	1.02	1.54	.94	1.65	.86	1.77
	.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89
30	.01	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
	.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	.98	1.73
	.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
	.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
	.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
	.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
80	.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
	.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
100	.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
	.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Anexos B

Anexo: Códigos en R

librería en R

```
install.packages("rriskDistributions")  
library(tseries)  
library(lmtest)
```

B.1. Código para el Caso 1

```
Datos<-read.csv("Datos.csv",header=T)
```

```
n<-length(Datos[, "X"])  
X<-c()  
Y<-c()  
for(i in 1:n){  
    Y[i]=Datos[i, "Y"]  
    X[i]=Datos[i, "X"]  
}
```

```
prop.model<-lm(Y~X)  
summary(prop.model)
```

```
plot(X,Y)  
abline(prop.model)
```

```
alfa<- -0.03358  
beta<-1.00578
```

```

Y2<-c()#Regresion
e<-c()#residuales
e2<-c()#residualesal cuadrado
for(i in 1:n-1){
    Y2[i]=alfa + beta*X[i]
    e[i]=(Y[i]-Y2[i])
    e2[i]=e[i]^2
}

pe<-mean(e)#media de residuales
Ci<-c()
Rho<-c()#funcion de autocorrelacion
Sumae<-c()#auxiliar
Sumae[1]<-(e[1]-pe)*(e[1]-pe)
for(t in 2:(n-2)){Sumae[t]<-(e[t]-pe)^2 + Sumae[t-1]}
C0=Sumae[(n-2)]

for(i in 1:(n-4)){
Sumae[1]<-(e[1]-pe)*(e[1+i]-pe)
for(t in 2:(n-2-i)){Sumae[t]<-(e[t]-pe)*(e[t+i]-pe)+ Sumae[t-1]}
Ci[i]<-Sumae[(n-2-i)]/n
Rho[i]=Ci[i]/C0
}

plot(Rho)
acf(e, ci=0.95, xlab="Retrasos", ylab="autocorrelaciones", main="correlograma",
pacf(e, ci=0.95, xlab="retrasos", ylab="autocorrelaciones", main="PACF", plot=TRUE)

#dwtest#
w <-0
g <-0
for(i in 1:(n-1)){
    w <-(e[i+1]-e[i])^2 + w
    g <- e2[i] + g
}
d <-w/g #Estadistica Durbin-Watson

#dwtest de R#
e_1<-e[-c(1)]
e<-e[-c(89)]
dwtest(e_1 ~ 0+e)

```

```
ModelAR1<-lm(e_1 ~ 0+e)
```

```
psi <-0.8277 #parametro
```

```
#####Conchrane-Orcutt#####
```

```
Yn<-c()
```

```
Xn<-c()##nueva variable X'=Xi-(psi)Xi-1
```

```
for(i in 1: n-1){
```

```
  Xn[i]<-X[i+1]-psi*X[i]
```

```
  Yn[i]<-Y[i+1]-psi*Y[i]
```

```
}
```

```
CO<-lm(Yn ~ Xn)
```

```
a<-0.06119
```

```
b<-0.85236
```

```
plot(Xn, Yn)
```

```
abline(CO)
```

```
Y2<-c()#Regresion
```

```
e<-c()#residuales
```

```
e2<-c()#residualesal cuadrado
```

```
for(i in 1:n-1){
```

```
  Y2[i]=a + b*Xn[i]
```

```
  e[i]=(Yn[i]-Y2[i])
```

```
  e2[i]=e[i]^2
```

```
}
```

```
###FAC y FACP ##
```

```
acf(e, ci=0.95, xlab="Retrasos", ylab="autocorrelaciones", main="correlograma"
```

```
pacf(e, ci=0.95, xlab="retrasos", ylab="autocorrelaciones", main="PACF", plot=T
```

```
#####dwtest###
```

```
w<-0
```

```
g<-0
```

```
for(i in 1:(n-2)){
```

```
  w <-(e[i+1]-e[i])^2 + w
```

```
  g <- e2[i] + g
```

```
}
```

```
d<-w/g
```

```
e_1<-e[-c(1)]
```

```
e<-e[-c(88)]
```

```

dwtest(e_1 ~ 0+e)

#####Fin Conchrane-Orcutt#####

#modelo para Xt
t<-c()
for(i in 1:n){
  t[i]=i
}

prop.model<-lm(X~t)
a=2.190466
b=0.006608

X2<-c()#Regresion
e<-c()#residuales
e2<-c()#residuales al cuadrado
for(i in 1:n){
  X2[i]=a + b*t[i]
  e[i]=(X[i]-X2[i])
  e2[i]=e[i]^2
}

e_1<-e[-c(1)]
e<-e[-c(88)]
Ex<-lm(e_1 ~ 0+e)

psi_x <- 0.8516

```

B.2. Código para el Caso 2

```

Datos2<-read.csv("Datos2.csv",header=T)

n<-length(Datos2[, "X"])
X<-c()
Y<-c()
X2<-c()
for(i in 1:n){
  Y[i]=Datos2[i, "Y"]
  X[i]=Datos2[i, "X"]
  X2[i]=Datos2[i, "X2"]
}

```

```
prop.model<-lm(Y~X + X2)
summary(prop.model)
```

```
alfa<-0.04272
beta<-0.32017
beta2<-0.63379
```

```
Y2<-c()#Regresion
e<-c()#residuales
e2<-c()#residualesal cuadrado
for(i in 1:n-1){
  Y2[i]=alfa + beta*X[i]+ beta2*X2[i]
  e[i]=(Y[i]-Y2[i])
  e2[i]=e[i]^2
}
```

```
#####FUNCION DE AUTOCORELACION
pe<-mean(e)#media de residuales
```

```
Ci<-c()
Rho<-c()
Sumae<-c()#auxiliar
Sumae[1]<-(e[1]-pe)*(e[1]-pe)
for(t in 2:(n-2)){Sumae[t]<-(e[t]-pe)^2 + Sumae[t-1]}
C0=Sumae[n-2]
```

```
for(i in 1:n-1){
Sumae[1]<-(e[1]-pe)*(e[1+i]-pe)
for(t in 2:(n-2-i)){Sumae[t]<-(e[t]-pe)*(e[t+i]-pe)+ Sumae[t-1]}
Ci[i]<-Sumae[n-2-i]/n
Rho[i]=Ci[i]/C0
}
```

```
plot(Rho)
acf(e, ci=0.95, xlab="Residiales", ylab="autocorrelaciones", main="correlogram")
pacf(e, ci=0.95, xlab="Retrasos", ylab="autocorrelaciones", main="correlograma")
```

```
#####dwttest###
w<-0
g<-0
```

```

for (i in 1:(n-1)){
  w <- (e[i+1]-e[i])^2 + w
  g <- e2[i] + g
}
d<-w/g
#0.4907414

e<-e[-c(89)]
dwtest(e_1 ~ 0+e)
E<-lm(e_1 ~ 0+e)

psi <-0.7518

###modelo para X2, t
t<-c()
for (i in 1:n){
  t[i]=i
}

prop.model<-lm(X2~t)
a=2.229248
b=0.007592

Xs<-c()#Regresion
e<-c()#residuales
e2<-c()#residuales al cuadrado
for (i in 1:n){
  Xs[i]=a + b*t[i]
  e[i]=(X2[i]-Xs[i])
  e2[i]=e[i]^2
}

e_1<-e[-c(1)]
e<-e[-c(89)]
Ex<-lm(e_1 ~ 0+e)
psi_x2<- 0.8692

```

B.3. Código para el Caso 3

```

Datos2<-read.csv("Datos2.csv", header=T)

n<-length(Datos2[, "X"])
X<-c()

```

```

Y<-c()
X2<-c()
for(i in 1:n){
  Y[i]=Datos2[i,"Y"]
  X[i]=Datos2[i,"X"]
  X2[i]=Datos2[i,"X2"]
}

```

##primero Variables instrumentales

```

Y<-Y[-1]
X_1<-c()
X_1<-X[-89]

```

```

prop.model<-lm(Y~1+X_1)
summary(prop.model)

```

```

p1<-0.03619
p2<-0.97962
Yp<-c()
for(i in 1:(n-1)){
  Yp[(i)]= p1 + p2*X_1[i]
}

```

```

###
Y_1<-c()## Y retrasada por 1
Y_1<-Yp[-c(88)] # en 3 a 88
Y<-Y[-c(1)] # para que empiece en 3
X<-X[-c(1,2)] # para que empiece en 3
X2<-X2[-c(1,2)] # para que empiece en 3

```

```

prop.model<-lm(Y~X + X2 + Y_1)
summary(prop.model)

```

```

alfa<-0.06647
beta<-0.31470
beta2<- 0.68045
beta3<- -0.05264

```

aqui ya puedo usar otra vez toda la muestra t:1,89


```

Y2<-c()#Regresion
Y2[1]=Y[1]
e<-c()#residuales
e2<-c()#residualesal cuadrado
for(i in 1:(n-2)){
  Y2[(i+1)]=alfa + beta*X[(i+1)]+ beta2*X2[(i+1)] + beta3*Y2[i]
  e[i]=(Y[i]-Y2[i])
  e2[i]=e[i]^2
}

acf(e, ci=0.95, xlab="Residiales",ylab="autocorrelaciones",main="correlograma")
pacf(e, ci=0.95, xlab="Retrasos",ylab="autocorrelaciones",main="correlograma")

e_1<-e[-c(1)]
e<-e[-c(88)]

E <-lm(e_1 ~ 0+e)
summary(E)

psy<-0.74289

bgtest(e_1 ~ 0+ e) ##test Breusch-Godfrey con p-value = 0.3089
[8] [5] [9] [6] [2] [4] [7] [1] [3]

```

Bibliografía

- [1] ANGRIST, J. D., AND IMBENS, G. W. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association* 90, 430 (1995), 431–442.
- [2] GODFREY, L. G. *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*. No. 16. Cambridge University Press, 1991.
- [3] HARRIS, R., AND SOLLIS, R. *Applied time series modelling and forecasting*. Wiley, 2003.
- [4] KLEIBER, C., AND ZEILEIS, A. *Applied econometrics with R*. Springer Science & Business Media, 2008.
- [5] LJUNG, G. M., AND BOX, G. E. On a measure of lack of fit in time series models. *Biometrika* (1978), 297–303.
- [6] MONTGOMERY, D. C., PECK, E. A., AND VINING, G. G. *Introduction to linear regression analysis*. John Wiley & Sons, 2015.
- [7] NIKOLOPOULOS, K. *Introduction to econometrics: Christopher dougherty*, oxford university press, 2002, paperback, 424 pages. isbn: 0198776438, £ 27.99, 2004.
- [8] OSTROM, C. W. *Time series analysis: Regression techniques*, vol. 9. Sage, 1990.
- [9] SHUMWAY, R. H., AND STOFFER, D. S. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.