"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Programa de Licenciatura en Matemáticas

Series de Tiempo y el Problema de Regresión Espuria

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Valeria Cienfuegos Colunga

Director de Tesis:

Dra. Gudelia Figueroa Preciado

Hermosillo, Sonora, México, Noviembre 2016.

SINODALES

Dr. José Arturo Montoya Laos

Dra. María Teresa Robles Alcaraz

MC. José María Bravo Tapia

Dr. Francisco Cienfuegos Velasco

A mis padres y hermano
Alicia, Francisco y Pablo

Agradecimientos

La gratitud es el acto más humilde que alguien puede expresar por un beneficio recibido. Ya que no encuentro palabras que expresen mi agradecimiento por el apoyo que he tenido durante este tiempo, lo único que me resta por decir es un sincero ***gracias****.*

Gracias*, primeramente, al creador de la vida, y de este maravilloso arte, Dios. Gracias a todos mis maestros, por darme a conocer el increíble mundo de las matemáticas. En especial, gracias a la maestra María Teresa Robles Alcaráz, al Dr. Jose Arturo Montoya Laos, al Dr. Martin Gildardo García Alvarado, al Dr. Jorge Ruperto Vargas Castro, por las enseñanzas de cada uno de usedes dentro y especialmente fuera del aula, por sus consejos y por las palabras de ánimo que brindaron, y por ese ejemplo de cómo ser un matemático. Gracias a mis compañeros de generación, por los conocimientos compartidos y el apoyo brindado. Muchas gracias amigos, a todas esas personas que me acompaaron en este no tan corto viaje, y que junto conmigo, vieron nacer poco a poco, este trabajo.*

Especialmente, gracias a mi directora de tesis, Dra. Gudelia Figueroa Preciado, por sus consejos, por su tiempo, y su interminable paciencia.

Gracias a la Universidad de Sonora, a los bibliotecarios y todo el personal con el que alguna vez tuve contacto en este período. No hay persona menos importante que otra, pues todos tienen un papel fundamental en el crecimiento de un matemático.

Y lo mejor a lo último, gracias infinitas a mi familia, mis pilares, mi papás y mi hermano, gracias por su apoyo incondicional y constante; gracias por inyectar la energía de seguir en este camino no fácil, soy lo que soy por el ejemplo que he recibido de ustedes. ***Gracias****.*

Índice general

Capítulo 1

Introducción

Las series de tiempo son un conjunto de datos u observaciones generadas secuencialmente en el tiempo. Entre estas observaciones existe una dependencia que por lo general es más fuerte entre más cercanas se encuentren. En general supondremos que sus mediciones han sido tomadas en tiempos que están igualmente espaciados. A diferencia de muchos análisis estadísticos en los que el orden en que se selecciona la muestra puede ser irrelevante, en series de tiempo, ese orden resulta de gran importancia. Si los valores futuros de una serie de tiempo pueden ser determinados por medio de una función matemática, la serie se dice determinista; pero si éstos son descritos en términos de una distribución de probabilidad, la serie se conoce como no-determinista o serie de tiempo estadística.

Las observaciones de una serie de tiempo estadística, que en adelante llamaremos simplemente serie de tiempo, pueden considerarse como una realización de un proceso teórico llamado proceso estocástico. Uno de los objetivos principales del análisis de series de tiempo es el de realizar pronósticos, para ello es necesario desarrollar modelos matemáticos que proporcionen descripciones plausibles de los datos y así poder realizar inferencias acerca de las propiedades o características del proceso estocástico para el cual se propone un modelo, el cual se espera tenga propiedades similares a las del mecanismo que genera este proceso.

Para analizar un fenómeno mediante series de tiempo, o sea, para identificar la estructura de una serie, es necesario conocer la distribución conjunta de $X_1, \ldots, X_n$; sin embargo, esto no es muy realista en el caso de series de tiempo, por lo que cual se trabaja con su primer y segundo momento, que resumen en buena medida su distribución, y el análisis se enfoca generalmente en propiedades de la serie que dependan sólo de estos momentos; con ellos es posible obtener medidas y construir gráficas que permiten analizar el comportamiento de la serie. Todo esto es presentado en el Capítulo 2, junto con algunos ejemplos de series de tiempo y el proceso de ruido blanco, muy útil en la construcción de procesos estocásticos más complejos, como caminatas aleatorias, proceso de promedios móviles, procesos autorregresivos, etcétera.

En el Capítulo 3 se amplía el estudio de los procesos presentados en el Capítulo 2, construyendo, a partir de datos generados por medio de un proceso de ruido blanco, un proceso de promedios móviles, un autorregresivo y una caminata aleatoria. Para el análisis de estas series se utilizan las técnicas descriptivas presentadas en el Capítulo 2.

En muchos ámbitos, particularmente aquí consideramos el económico, las se-

ries de tiempo muestran generalmente una tendencia positiva o negativa y en ocasiones presentan una varianza no constante; ello origina que estas series sean no-estacionarias y esa característica dificulta describir el fenómeno en ciertos intervalos de tiempos, así como el realizar predicciones. El problema de no-estacionariedad ha sido uno de los grandes temas en econometría, ya que el ánalisis de series de tiempo está fundamentado en la teoría de estacionariedad y como la mayoría de las series económicas no lo son, es importante verificar la estacionariedad de éstas. Es por ello que abundan pruebas estadísticas que permiten detectar este tipo de series y evitar obtener *regresiones espurias*, tema central en el Capítulo 4.

Actualmente se cuenta con una amplia variedad de pruebas de estacionariedad, las cuales pueden diferir en supuestos, potencias, tamaño necesario para detectar no-estacionariedad, etcétera. Este trabajo se enfoca en estudiar el uso de las pruebas de Dickey-Fuller y Phillips-Perron, muy utilizadas en la literatura y práctica económica. Estas pruebas se presentan en el Capítulo 4 y se analizan indicadores sencillos que permiten detectar regresiones espurias. Se presentan varios ejemplos; dos de ellos para mostrar lo común de observar o construir regresiones espurias y otros donde se utilizan las pruebas de Dickey-Fuller y Phillips-Perron, se delimitan los modelos a probar y se efectúa una comparación de los resultados obtenidos de éstas. Es importante señalar que en la selección del modelo es fundamental el conocimiento sobre el problema de aplicación. Aunque existen muchas otras pruebas de estacionariedad, como Zivot & Andrews (1992), Elliott *et al.* (1996), Kwiatkowski *et al.* (1992), su uso es similar y pueden diferir, como se mencionó previamente, en los supuestos, potencia, así como en la especificación de la hipótesis nula y alternativa. En las pruebas que se utilizan en este trabajo, la hipótesis nula plantea que la serie de tiempo es no-estacionaria, a diferencia por ejemplo de la prueba KPSS donde la hipótesis nula plantea que la serie es estacionaria frente a una alternativa de que no lo es.

Finalmente, en el Capítulo 5 se presenta una serie de simulaciones, en donde se generaron conjuntos de 10000 muestras aleatorias provenientes de diferentes procesos autorregresivos, donde se hizo variar el valor ρ, tomando éste valores de 0.8, 0.9, 0.95, 0.99, 1, 1.02 y 1.05. Los tamaños de estas muestras fueron de 50, 100 y 200 datos. Las series se simularon a partir de un modelo autorregresivo de orden 1 sin intersección, y con intersección, coeficiente que se fijó en 0.1 y 0.5. En todas las series simuladas se aplicaron las pruebas de estacionariedad de Dickey-Fuller y Phillips-Perron, y se contabilizó el porcentaje de veces que se rechazó la hipótesis nula de no estacionariedad, utilizando un nivel de significancia de $\alpha = 0.05$. En este capítulo se muestra la importancia del valor que toma ρ y el tamaño de muestra, pues como se verá, en muchas de las series estacionarias simuladas, donde ρ es menor que uno pero muy cercano a éste, puede fácilmente concluirse que la serie proviene de un proceso no-estacionario. Por otra parte, al generar series con valores de ρ cercanos a uno, pero mayores que éste, se tendrán series de tiempo explosivas; en esos casos las hipótesis cambian, pues en la hipótesis alternativa se plantea que la serie es explosiva, lo cual debe establecerse claramente en el software estadístico que se esté utilizando, de otra manera los resultados obtenidos no serán confiables.

Dado pues, que generalmente las series de tiempo que se utilizan en el área de

economía son no-estacionarias, y éste es un aspecto que comúnmente se pasa por alto, resulta de interés el reconocer dichas series y transformarlas a estacionarias para no obtener resultados espurios. Conociendo también que cualquier resultado estadístico conlleva cierto grado de incertidumbre, resultó de interés estudiar cómo influye el tamaño de muestra, el valor de ρ y la prueba a utilizar, en la conclusión obtenida. Es importante mencionar que la selección del modelo a probar en estas pruebas de hipótesis es fundamental y requiere de un conocimiento del problema bajo análisis, siendo ahí donde el trabajo interdisciplinario entre investigadores de la estadística y la economía resulta fructífero.

Capítulo 2

Conceptos Básicos

En este capítulo se presentan algunos conceptos básicos de series de tiempo. Se exploran ciertas técnicas descriptivas y gráficas que permiten identificar ciertos patrones en las series, los cuales permiten identificar características propias de algunas series que se presentarán en el Capítulo 3.

2.1. Series de tiempo

Las series de tiempo son un conjunto de datos u observaciones realizadas en forma secuencial, durante un período de tiempo. La medición entre tiempos específicos se considera por lo general igualmente espaciada y estos tiempos pueden ser medidos en minutos, horas, semanas, hasta en años. Por lo general su medición es discreta aunque la variable en estudio sea continua, por ello, este tipo de serie de tiempo se denomina *serie de tiempo discreta*, que serán las aquí tratadas. Algunos ejemplos de este tipo de series de tiempo son: la venta semanal de autos, la tasa mensual de inflación de los precios al consumidor, producción semanal de lácteos, el crecimiento anual de la población de cierta especie, etcétera; conjuntos pues, que modelan la evolución de cierta(s) variable(s) a lo largo del tiempo.

Las series de tiempo se aplican muy comúnmente en áreas como climatología, administración de riesgos, economía, control de procesos, demografía, entre otras. Por ejemplo, para determinar las características de las estaciones del año fue necesario estudiar el comportamiento del clima, pues las estaciones son períodos donde ciertas condiciones climatológicas se mantienen; de igual manera es necesario el conocimiento del clima de días anteriores para poder pronosticar el comportamiento climático de días futuros, o prever fenómenos naturales que puedan aquejar una ciudad, como tormentas, maremotos, etcétera. Las series de tiempo también permiten optimizar ciertos procesos que influyen en la mejora de algún producto, ya sea perfeccionando el producto en sí, aumentando las ganancias, logrando un ahorro de energía, mejorando el tiempo de producción, etcétera. A través de series de tiempo también es posible comparar dos sucesos en el tiempo, por ejemplo, la tasa de mortalidad en cierto año y algún evento catastrófico (guerra, hambre, sequía, etcétera). Así pues, como éstos, hay cientos de ejemplos que ilustran el uso de las series de tiempo.

Las series de tiempo pueden ser clasificadas como *deterministas*, cuando los valores de la serie pueden ser calculados exactamente con alguna función matemática, esto es, futuros valores de la serie puede predecirse de valores anteriores; pero, si es necesaria una distribución de probabilidad para obtener dichos resultados, entonces

la serie se dice *no-determinista* o aleatoria. La mayoría de las series son aleatorias, esto es, los valores futuros en la serie son parcialmente determinados por los valores anteriores.

El análisis de datos de series de tiempo involucra el uso de elementos descriptivos como gráficas y medidas descriptivas y elementos de inferencia que permiten responder preguntas acerca de la población en estudio, que podría considerarse como el conjunto de series de tiempo formadas por la parte determinista o semideterminista, combinada con todas las posibles realizaciones imaginables de la parte estocástica (Guerrero, 1990). Dentro del aspecto gráfico se analizará la metodología de descomposición de las series de tiempo, la cual permite identificar las partes determinista y no-determinista de una serie.

A través de los años las gráficas han sido una de las herramientas más utilizadas para representar información y así ha ocurrido con las series de tiempo, ya que por medio de éstas es posible visualizar la variabilidad de una manera más sencilla. La Figura 2.1 muestra una ilustración que data aproximadamente del siglo X u XI; en el fondo de ésta se observa una cuadrícula que permite ubicar las curvas que representan la inclinación de las órbitas planetarias, en función del tiempo. Esta gráfica aparece en un manuscrito descubierto en 1877 relacionado con trabajos de la física y la astronomía de aquella época. Cabe mencionar que esta ilustración es al parecer el ejemplo más antiguo que se conoce de una representación gráfica del cambio de un fenómeno en el tiempo y es una misteriosa y aislada maravilla en la historia de los gráficos, ya que la siguiente gráfica de series de tiempo conocida surge alrededor de 800 años después (Tufte, 1983). Es hasta finales de 1700 cuando los gráficos de

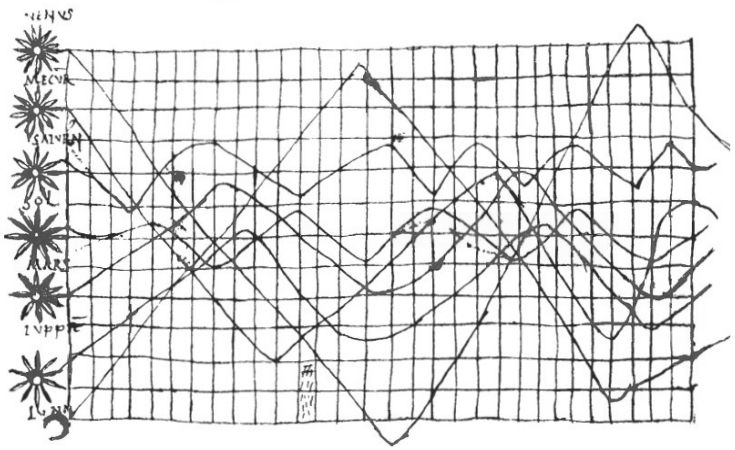


Figura 2.1: Representación gráfica de una serie de tiempo: astrónomo desconocido, siglo X u XI.

series de tiempo empiezan a aparecer en publicaciones científicas. En la Figura 2.2 se observa una gráfica desarrollada por Johann Heinrich Lambert y corresponde a una de las series más grandes de esos tiempos. Ésta muestra la variación periódica en la temperatura del suelo y su relación con la profundidad bajo la superficie. Entre

más profundidad mayor es el rezago de tiempo en la respuesta de la temperatura. Los gráficos actuales de periodicidad de series de tiempo difieren sólo un poco de aquellos de Lambert, aunque hoy en día las bases de datos son todavía más grandes. Los diseños de gráficas de series de tiempo se basan principalmente en las aportacio-

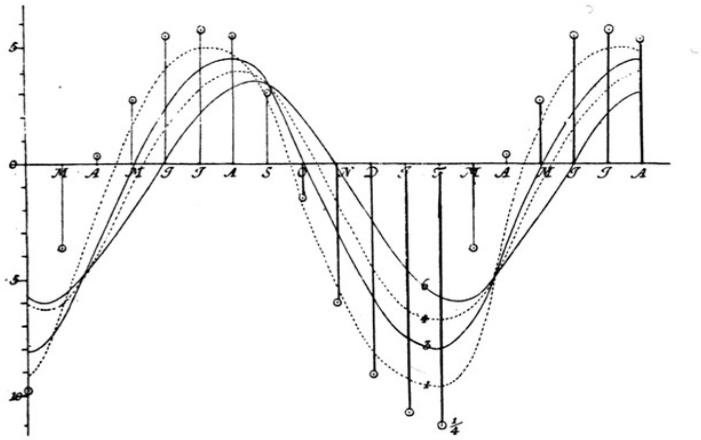


Figura 2.2: Variación periódica en la temperatura del suelo y su relación con la profundidad bajo la superficie.

nes del matemático suizo-alemán J.H. Lambert (1728-1777) y el economista sueco William Playfair (1759-1823). Al parecer, la primera serie de tiempo usando datos económicos fue publicada en el libro de Playfair "The Commercial and Political Atlas" (Londres, 1786) y se presenta en la Figura 2.3. Para Playfair los gráficos eran preferibles a las tablas, ya que éstos muestran la estructura de los datos en una perspectiva que puede ser comparable. En la primera edición de su libro, Playfair incluye 44 gráficas y lo que puede ser el primer gráfico de barras, al cual llega por la necesidad de representar datos anuales que necesitaban ser visualizados fácilmente; sin embargo, de acuerdo con Tufte (1983), Playfair se encontraba un poco escéptico sobre su creación.

Existen muchas razones por las cuales analizar series de tiempo, pero los objetivos principales, según Chatfield (2000), por los cuales se estudia este tipo de datos son:

- *Descripción.* Describir los datos usando un resúmen estadístico y/o métodos gráficos. Una gráfica de tiempo de los datos es particularmente importante.

- *Modelado.* Encontrar un modelo estadístico adecuado que describa el proceso de generación de datos. Un modelo *univariado* para una variable dada está basado solamente en datos anteriores; en cambio, un modelo *multivariado* para una variable dada puede estar fundado no sólo en los valores pasados de tal variable, si no también en el presente y el pasado de valores de otras variables predictoras, esto es, la variación en una serie puede ayudar a explicar la variación en otra serie. Por supuesto, siempre debemos recordar que todos los modelos son aproximaciones y que *la construcción de modelos es tanto un arte como una ciencia.*

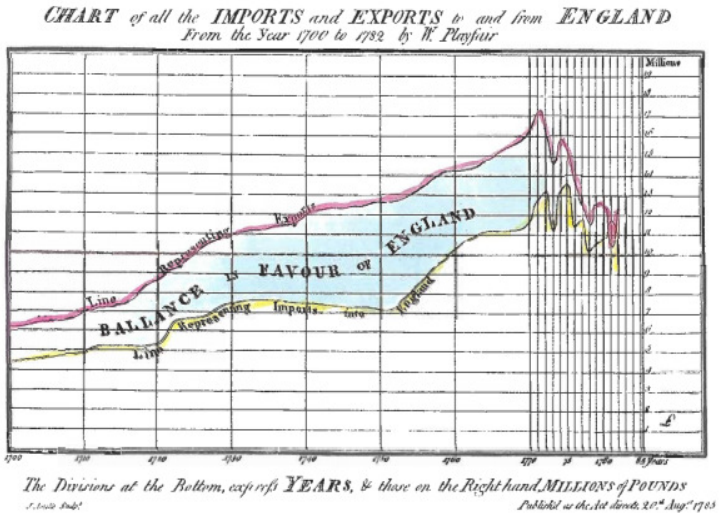


Figura 2.3: Serie de tiempo usando datos económicos: *The Commercial and Political Atlas (1786)*

- *Predicción.* Estimar los valores futuros de la serie.
- *Control.* Buenas predicciones permiten al analista tomar medidas a fin de controlar un proceso dado, ya sea industrial, económico o de alguna otra índole.

El observar y analizar gráficas de series de tiempo permite encontrar ciertos patrones, los cuales ayudarán en el proceso de realizar inferencias sobre el comportamiento de la serie. Los principales patrones que pueden observarse son los siguientes (Chatfield, 2000, p. 22, 23):

- *Tendencia.* Se puede definir como una cambio a largo plazo en la media, sin embargo, no hay una definición matemática establecida. Esta tendencia puede ser positiva o negativa, y a su vez lineal, cuadrática o de orden mayor.
- *Estacionalidad.* Se refiere a la variación periódica de cierto evento que sucede en un plazo de igual o menor a un año.
- *Cíclicos.* A diferencia del componente estacional, la variación cíclica se da en plazos mayores a un año. Estas variaciones pueden o no ser periódicas. Es común encontrar este tipo de ejemplos en actividades económicas. Sin embargo, estos patrones también se pueden encontrar en comportamientos biológicos de criaturas vivas y pueden presentarse en períodos de un día, a lo que se conoce como variaciones diurnas.

Mittal afirma que los efectos estacionales son debido a razones o periodicidades conocidas, como las estaciones del año, y los patrones cíclicos son debido a razones desconocidas (Mittal, 2007, p. 82).

Así pues, una serie de tiempo puede ser analizada descomponiéndola según los patrones ya mencionados, además de un componente que se considera como aleatorio. Por lo general una serie se puede escribir en su forma aditiva de la siguiente manera:

$$X_t = T_t + S_t + I_t + \epsilon_t, \tag{2.1.1}$$

donde denotamos por:

- T_t: valor del componente de tendencia en el período t
- S_t: valor del componente estacional en el período t
- I_t: valor del componente irregular en el período t

También se puede dar el caso de descomponer la serie en su forma multiplicativa, todo depende del comportamiento de los datos (Cowpertwait, 2009, p. 19). Los componentes de la serie 2.1.1 son considerados deterministas y sólo ϵ_t, es considerado la parte no determinista ó estocástica de la serie de tiempo. Sin embargo, en una visión más moderna, se considera que existen componentes estocásticos en la tendencia, la estacionalidad y el componente irregular (Enders, 2015, p. 3).

En la siguiente sección se presentan algunos ejemplos de series de tiempo en los que pueden identificarse, fácilmente, ciertos componentes en particular, aunque existen herramientas muy sencillas para distinguir éstos, como la descomposición que se realiza con la serie de tiempo correspondiente a los datos mensuales de pasajeros internacionales de una aerolínea entre 1949-1960, presentado por Box *et al.* (2008), y cuya descomposición aditiva puede observarse en la Figura 2.4, construida con el uso del software R, y donde se separan los componentes de tendencia, estacionalidad y el correspondiente al componente aleatorio.

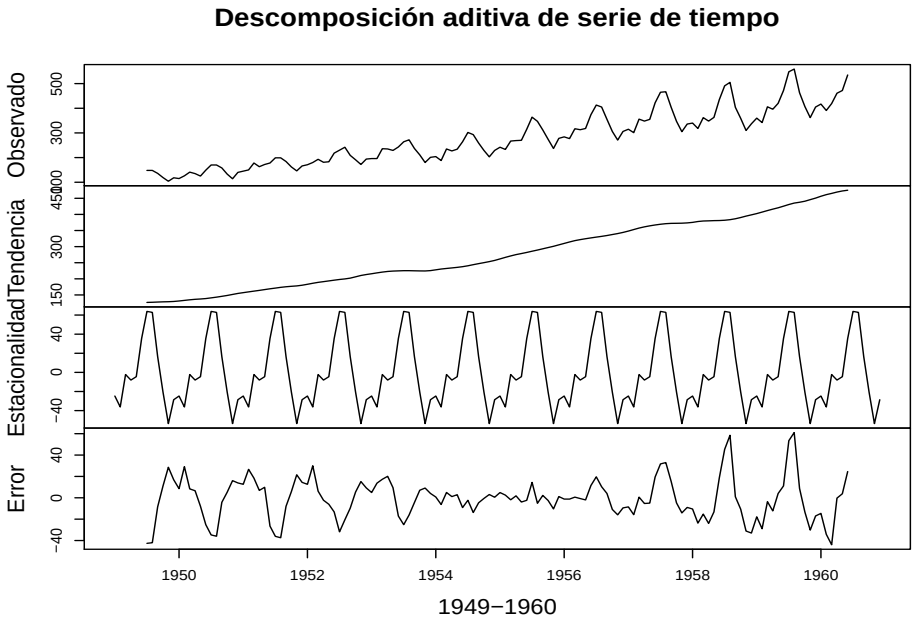


Figura 2.4: Pasajeros de una aerolínea, entre 1949 y 1960.

2.1.1. Algunos ejemplos de series de tiempo

A través de las siguientes series de tiempo se mostrarán los diferentes patrones mencionados en la Sección 2.1. Por lo general, la mayoría de las series que se presentan en la práctica combinan uno o varios de estos patrones, aunque en ocasiones también será posible observar la presencia de datos atípicos, que corresponden generalmente a eventos extraños con una probabilidad de ocurrencia pequeña, que en la mayoría de los casos será casi imposible predecir y por ello no formarán parte del modelo. Por eso, es importante notar que una serie de tiempo no explica en su totalidad el comportamiento del fenómeno, sino una aproximación a éste; cuando esta aproximación no es buena, las inferencias realizadas pueden ser totalmente erróneas.

En la Figura 2.5 se puede apreciar la serie de tiempo correspondiente a los datos del producto interno bruto en Sonora, desde 2003 hasta 2014, cifras que se representan en millones de pesos. En esta serie puede observarse una tendencia positiva, aunque en los años 2008-2009 parece que no hubo cambios. Por otra parte, en la Figura 2.6 se observan los porcentajes de población ocupada en México, mayor de 15 años y con primaria incompleta; esta serie presenta una tendencia negativa. Además de la tendencia, se pueden apreciar ciertos picos, que aunque no muy pronunciados pudieran ser el componente estacional de la serie. En la Figura 2.7 se construyó una

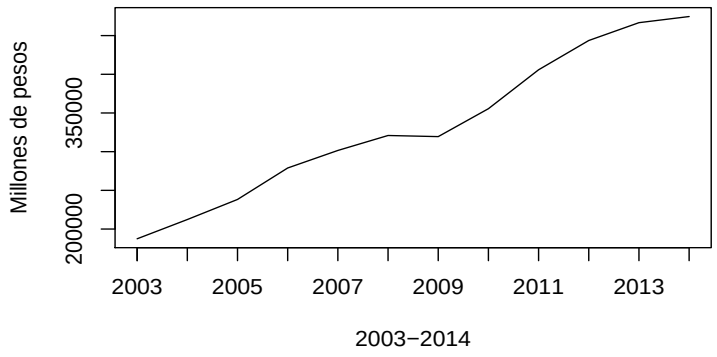


Figura 2.5: Producto Interno Bruto en Sonora.

serie de tiempo con datos referentes a los ingresos del sector público en México, que no provienen de recursos petroleros. Esta serie presenta un componente de estacionalidad, tiene una tendencia positiva en sus valores y puede observarse que año con año hay mayor variabilidad en el comportamiento de estos datos.

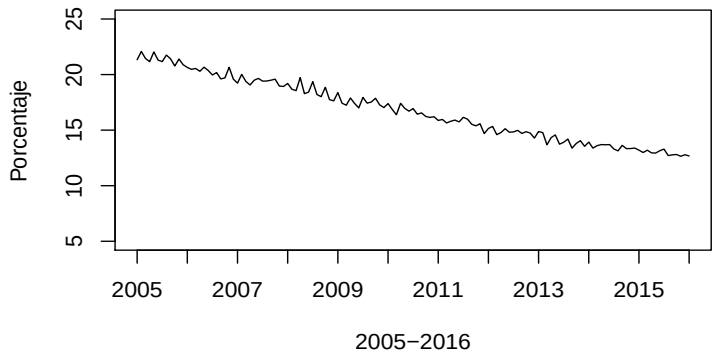


Figura 2.6: Población ocupada mayor de 15 años y con primaria incompleta.

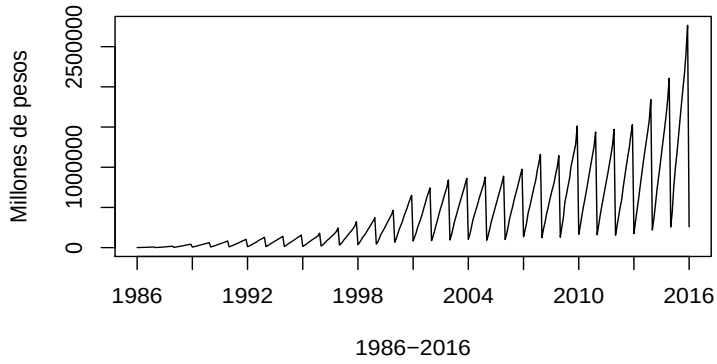


Figura 2.7: Ingresos al sector público que no provienen del petróleo.

Por otra parte, en la Figura 2.8 se muestra una serie de tiempo construida con los índices de precios de la canasta básica en México, desde Enero del 2000 hasta Mayo del 2016. En esta serie se observa una tendencia positiva, además de tener componentes estacionales.

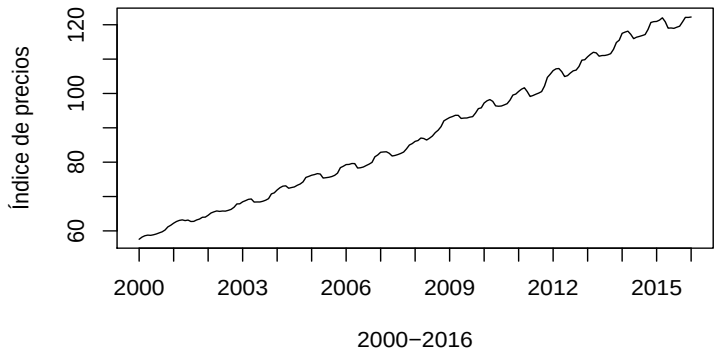


Figura 2.8: Índice de precios de la canasta básica.

Por último, en la Figura 2.9 se muestra la serie de tiempo del indicador global de la actividad económica en México desde 1980 hasta Marzo del 2016. En esta serie puede notarse un componente cíclico, el cual recordemos se observa en plazos mayores a un año.

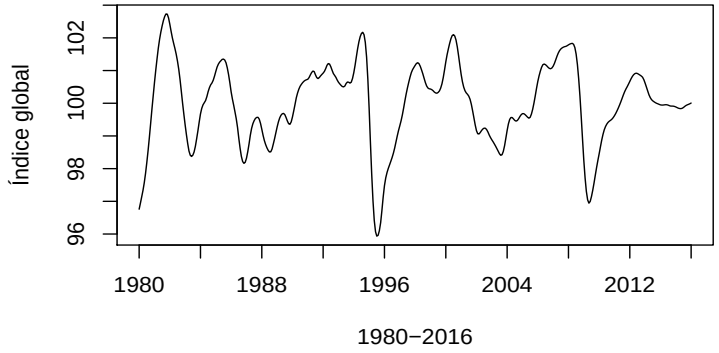


Figura 2.9: Indicador global de la actividad económica en México.

2.2. Series de tiempo y procesos estocásticos

Cuando se analiza una serie de tiempo, de manera formal, ésta puede concebirse como una sucesión de observaciones generadas por un proceso estocástico, cuyo índice se toma con respecto al tiempo. Esta realización generalmente se denota como $\{x_t\}_1^T$, mientras que el proceso estocástico será la familia de variables aleatorias $\{X_t\}_{-\infty}^{\infty}$, definida en un espacio de probabilidad apropiado. Cada una de estas sucesiones es una realización o trayectoria producida por un mecanismo probabilístico

subyacente al sistema en estudio. Así pues, una trayectoria representa una observación del proceso, o bien, una serie de tiempo. Por lo tanto, existe una relación entre series de tiempo y procesos estocásticos, por ello consideraremos las siguientes definiciones.

Definición 2.1 *Un **proceso estocástico**, $\{X_t\}$, es una familia de variables aleatorias indexadas por $t \in T \subset \Re$, y definida en un espacio de probabilidad.*

Por lo general T denota un conjunto de índices ordenados, típicamente identificado con el tiempo. En la literatura podemos encontrar conjuntos de índices como los siguientes:

- tiempo discreto $T = 1, 2, ...$
- tiempo discreto $T = ..., -2, -1, 0, 1, 2, ...$
- tiempo continuo $T = [0, \infty)$ o $T = (-\infty, \infty)$

aunque usualmente T se asocia al conjunto de los números enteros.

Así pues, una realización de un proceso estocástico con T observaciones, es una sucesión de datos observados que se utilizarán para estudiar el proceso

$$\{X_1 = x_1, X_2 = x_2, \ldots, X_T = x_T\} = \{x_t\}_{t=1}^{T}.$$

El objetivo de la modelización con series de tiempo es describir el comportamiento probabilístico del proceso estocástico que subyace y que se supone ha generado las observaciones. Para caracterizar bien este proceso sería necesario conocer la función de densidad conjunta de las variables involucradas en éste, aspecto poco realista. Sin embargo, considerando que los primeros momentos de variables aleatorias describen en gran parte su distribución, en un proceso estocástico es posible calcular ciertas características que permitan describir su comportamiento, como: medias, varianzas y covarianzas. Estas características pueden variar a lo largo del tiempo, por lo que serán funciones de éste.

Definición 2.2 ***Media de un Proceso Estocástico.*** *Sea $\{X_t\}$ un proceso estocástico con $V(X_t) < \infty$ para todo $t \in \mathbb{Z}$, la función $\mu_X : \mathbb{Z} \to \mathbb{R}$, definida por:*

$$\mu_X(t) = E(X_t),$$

es llamada la media del proceso estocástico.

Definición 2.3 ***Función de autocovarianza.*** *Sea $\{X_t\}$ un proceso estocástico con $V(X_t) < \infty$, para todo $t \in \mathbb{Z}$, se denomina función de autocovarianza a la función que asigna a cualesquiera dos períodos de tiempo t y s:*

$$\gamma_X(t, s) = Cov(X_t, X_s), \qquad \text{donde } t,s = 0 \pm 1, \pm 2, \ldots \tag{2.2.1}$$

En el análisis de series de tiempo es muy importante el concepto de estacionariedad, por ello a continuación se enuncia cuándo un proceso estocástico es estacionario.

Definición 2.4 ***Proceso estacionario.*** *Un proceso estocástico $\{X_t\}$ es estacionario si y sólo si para todo entero h, s y t, se cumple lo siguiente:*

1. $E(X_t) = \mu$,
2. $V(X_t) = \sigma^2$,
3. $\gamma_X(t, s) = \gamma_X(t + h, s + h)$.

La idea básica de estos procesos es el *equilibrio estadístico* que presentan, el cual se refiere a que las propiedades mencionadas anteriormente no cambian con el tiempo. Un proceso con las propiedades anteriores se conoce también como débilmente estacionario, covarianza-estacionario o estacionario de segundo orden. En la mayoría de las series de tiempo se cuenta con una observación para cada tiempo fijo t, esto es, la media μ_t no es la misma en cada período y hay sólo una observación para estimarla. Sin embargo, si se cumple el supuesto de que todas las observaciones comparten el mismo parámetro, se tienen todas ellas para estimarlo.

Es muy usual que el modelo de algunas series de tiempo económicas sean no-estacionarias, por ejemplo, índices como el PIB, INPC o INB, el crecimiento económico de cierto país, el ahorro nacional bruto, etcétera. Cuando las series son no-estacionarias, no es recomendable utilizar ciertos procesos que se presentarán en el Capítulo 3, aunque existen métodos para convertirlos a estacionarios, como diferenciación, aplicar logaritmos, etcétera.

Si $t = s$, tendremos que $\gamma_X(t, s) = \gamma_X(t, t) = V(X_t)$. Por lo tanto si X_t es estacionario, $\gamma_X(t, t) = V(X_t) = cte$. Por otra parte, la covarianza $\gamma_X(t, s)$, no depende de los puntos t y s, si no del número de períodos que están entre ellos, $t - s$. De esta manera, para procesos estacionarios vemos a la función de autocovarianza como función de un argumento, por lo tanto, se tiene lo siguiente,

Definición 2.5 *Sea $\{X_t\}$ un proceso estocástico estacionario con $V(X_t) < \infty$ para todo $t \in \mathbb{Z}$. La función de autocovarianza entre X_t y X_s, donde $h = t - s$, se denota por $\gamma_X(h)$, y se define como:*

$$\gamma_X(h) = Cov(X_t, X_{t-h}), \qquad \text{para todo } h \in \mathbb{N}. \tag{2.2.2}$$

Como la covarianza es simétrica en t y s, $\gamma_X(t, s) = \gamma_X(s, t)$, entonces

$$\gamma_X(h) = \gamma_X(-h), \qquad \forall h \in Z.$$

La autocovarianza mide la dirección de la dependencia lineal entre X_t y X_s y depende de las unidades de medida, por lo cual es difícil evaluar esta dependencia; una medida más útil es la función de autocorrelación, con la cual puede cuantificarse mejor dicha dependencia.

Definición 2.6 ***Función de autocorrelación.*** *La función de autocorrelación $\rho_X(h)$ de un proceso estacionario $\{X_t\}$, se define como:*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t-h}), \tag{2.2.3}$$

para todos los enteros h.

Es común denotar $\rho_X(h)$, como ρ_h y ambas notaciones se usarán indistintamente. Como puede observarse, esta definición es equivalente a:

$$\rho_X(h) = \frac{Cov(X_t, X_{t-h})}{\sqrt{V(X_t)V(X_{t-h})}}, \tag{2.2.4}$$

ya que la estacionariedad implica que $V(X_t) = V(X_{t-h})$ con lo cual

$$\sqrt{V(X_t)V(X_{t-h})} = V(X_t) = \gamma_X(0).$$

La autocorrelación mide tanto la dirección como la fuerza de la dependencia lineal entre X_t y X_{t-h}, por sus siglas en inglés también se le abrevia com *"ACF"*. La función de autocorrelación es independiente de la escala de medición de la serie de tiempo y es simétrica alrededor del cero.

Antes de presentar el concepto de autocorrelación parcial, primero veámos el siguiente ejemplo que presenta Montgomery, (2008, p. 248); considérese tres variables aleatorias, X, Y, Z. Al calcular la regresión lineal de X en Z y de Y en Z, se tiene que:

$$\hat{X} = a_1 + b_1 Z, \quad \text{donde } b_1 = \frac{Cov(Z, X)}{V(Z)},$$
$$\hat{Y} = a_2 + b_2 Z, \quad \text{donde } b_2 = \frac{Cov(Z, Y)}{V(Z)}.$$

Los errores pueden ser obtenidos como:

$$X^* = X - \hat{X} = X - a_1 + b_1 Z,$$
$$Y^* = Y - \hat{Y} = Y - a_2 + b_2 Z.$$

La **correlación parcial** entre X y Y después de ajustar para Z se define como la correlación entre X^* y Y^*. Esto es, la correlación parcial se puede ver como la correlación entre dos variables después de ser ajustadas por un factor común que las esté afectando. Ahora bien, intuitivamente, si se tienen dos series de tiempo, X_t y X_{t-h}, la función de autocorrelación parcial de X_t y X_{t-h}, es la correlación entre estas dos variables después de ajustar para $X_{t-1}, X_{t-2}, \cdots X_{t-h+1}$. Formalmente, se define de la siguiente manera:

Definición 2.7 ***Función de autocorrelación parcial.*** *Sea $\{X_t\}$, con $t \in \mathbb{Z}$, un proceso estacionario. La función de autocorrelación parcial, en el retraso h, para $h \geq 2$ se define como la correlación directa entre X_t y X_{t-h}, una vez removida la dependencia lineal de las variables intermedias X_s, con $t - h < s < t$.*

Básicamente lo que esta función realiza es determinar la correlación entre X_t y X_{t-h}, después de eliminar el efecto de las variables: $X_{t-1}, X_{t-2}, \ldots, X_{t-h+1}$. Generalmente a esta función se le puede denotar por ϕ_{hh}, $h \geq 1$, y por conveniencia se define $\phi_{00} = 1$ (Mittal, 2007).

En la mayoría de las situaciones basta con observar los primeros dos momentos, sin embargo, puede haber ocasiones en que sea necesario observar la distribución completa, para ello se utiliza el concepto que se presenta a continuación.

Definición 2.8 ***Proceso estrictamente estacionario.*** *Un proceso estocástico* $\{X_t\}$ *es estrictamente estacionario si la distribución conjunta de* $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ *y* $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})$*, es la misma para todo* $h \in \mathbb{Z}$*.*

Esto es,

$$P\{X_{t_1} \leq c_1, \cdots, X_{t_n} \leq c_n\} = P\{X_{t_1+h} \leq c_1, \cdots, X_{t_n+h} \leq c_n\}. \tag{2.2.5}$$

Esta definición implica que todas las funciones de distribución conjunta de un subconjunto de variables deben ser las mismas a las del subconjunto de variables separadas por una distancia h, por ejemplo, si $n = 1$ en (2.2.5), entonces

$$P\{X_{t_1} \leq c_1\} = P\{X_{t_1+h} \leq c_1\}. \tag{2.2.6}$$

Por otra parte, si la media existe, entonces lo anterior implica que $\mu_{t_1} = \mu_{t_1+h}$, por lo tanto μ_{t_1+h} debe ser constante. Esta condición es fuerte para la mayoría de las series de tiempo reales, ya que es difícil asegurar la estacionariedad estricta en un conjunto de datos; es por ello que generalmente se utiliza la Definición 2.4.

Existe un proceso que se conoce como ruido blanco y es muy útil para construir procesos estocásticos más complejos, como caminatas aleatorias, proceso de promedios móviles, procesos autorregresivos, etcétera, y su definición se presenta enseguida.

Definición 2.9 ***Ruido blanco.*** $\{Z_t\}$ *es proceso de ruido blanco, si* $\{Z_t\}$ *es un proceso estacionario y cumple que:*

- $E(Z_t) = 0$,
- $\gamma_Z(h) = \begin{cases} \sigma^2 & h = 0; \\ 0 & h \neq 0. \end{cases}$

Lo anterior se denota por $Z_t \sim WN(0, \sigma^2)$*.*

La función de autocorrelación de este proceso es igual a 0 si $h \neq 0$, pero si $h = 0$, entonces el resultado de ACF es 1. Cuando se cumple la condición de que las Z_i son independientes a través del tiempo, esto es Z_i y Z_j son independientes para $i \neq j$, se tendrá un proceso independiente de ruido blanco y ello implica la no-correlación, sin embargo el recíproco no siempre es cierto.

Definición 2.10 ***Caminata aleatoria.*** *Sea* $Z_t \sim WN(0, \sigma^2)$ *un proceso de ruido blanco, entonces el nuevo proceso*

$$X_t = X_{t-1} + Z_t = X_0 + \sum_{j=1}^{t} Z_j, \tag{2.2.7}$$

con $t > 0$*, se denomina caminata aleatoria.*

A continuación puede verse que este proceso es no-estacionario.

Dada una variable aleatoria inicial X_0, para $t > 0$ se tiene lo siguiente:

$$X_t = X_0 + Z_1 + Z_2 + \cdots + Z_t = X_0 + \sum_{j=1}^{t} Z_j.$$

Si la media de X_0 es finita, entonces X_t cumple con la primera condición de estacionariedad. Luego se calcula la varianza de $X_t - X_0$

$$V(X_t - X_0) = V\left(\sum_{j=1}^{t} Z_j\right) = \sum_{j=1}^{t} V(Z_j) = t\sigma^2. \tag{2.2.8}$$

Como puede observarse en (2.2.8), la varianza aumenta con el tiempo, por ello una caminata aleatoria es un proceso no estacionario, aunque su media sea constante en el tiempo.

2.3. Técnicas descriptivas gráficas

Como en todo análisis estadístico es fundamental contar con gráficos apropiados que sean informativos. Aparte de graficar una serie de tiempo, lo cual permite evaluar, en el tiempo, los patrones y comportamiento de la serie, se puede realizar otro tipo de gráfico de gran utilidad, obtenido a partir de la funciones de autocorrelación y autocorrelación parcial.

El **correlograma** es una herramienta muy importante en el análisis de las series de tiempo; en el eje y se grafica la función de autocorrelación, o bien, la función de autocorrelación parcial, y en el eje x se consideran unidades de tiempo t. Según el proceso y la condición de estacionariedad que se esté trabajando es la información que el correlograma va a proporcionar. Guerrero (1990, pp. 113) describe minuciosamente los comportamientos típicos de los correlogramas obtenidos a partir de las funciones de autocorrelación y de autocorrelación parcial, para el caso de procesos autorregresivos de orden p, $AR(p)$; procesos de promedios móviles de orden q, $MA(q)$ y los autorregresivos de promedios móviles $ARMA(p, q)$. Estos procesos se presentarán formalmente en el Capítulo 3.

En general, el análisis del correlograma es el siguiente. En el caso de un modelo $MA(q)$, el correlograma obtenido a partir de la ACF tendrá tantos valores significativos distintos de cero como orden del proceso de promedios móviles, q, y en el correlograma de la PACF se verá que la función decrece hacia cero, ya sea regular, sinusoidal o alternando valores de positivo a negativo o viceversa; por el contrario si se tiene un modelo $AR(p)$, entonces, se podría esperar que el correlograma de la ACF decrezca hacia cero (regular, sinusoidal o alternando valores de positivo a negativo o viceversa), y en el correlograma de la PACF, tendrá tantos valores significativos distintos de cero como orden del proceso autorregresivo, p. Lo anterior es para modelos sin estacionalidad. El caso del modelo $ARMA(p, q)$, no es tan sencillo de visualizar, pero tiene asociada una PACF que no desaparece después de un número finito de

retrasos. Es por ello necesario contar con la experiencia del investigador para dar una apropiada interpretación a los correlogramas. En el caso en el que el proceso sea no-estacionario, el correlograma sólo sirve para verificar ésto.

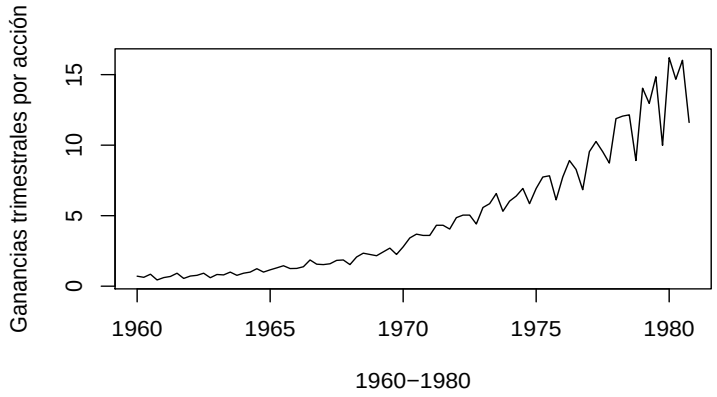


Figura 2.10: Serie de tiempo no-estacionaria: Johnson & Johnson

Ahora bien, consideremos la Figura 2.10, la cual muestra la serie de tiempo correspondiente a 84 datos que se encuentran en la base de datos JohnsonJohnson del Software R y que representan las ganancias trimestrales de la compañía Johnson & Johnson, entre 1960 y 1980. Claramente ésta es una serie de tiempo no-estacionaria, cuya media aumenta en el tiempo y el incremento en la variabilidad alrededor de esta tendencia, provoca cambios en la función de autocovarianza. Se observa también un componente cíclico, por año.
Generalmente se hacen transformaciones a la serie original aplicando el logaritmo a los datos, ésto con el fin de convertir una tendencia exponencial a una lineal. También suaviza la tendencia, si la serie muestra una variación creciente sobre el tiempo (Franses, 2014, p. 8). De acuerdo a lo anterior, se aplica a los datos la función logaritmo en base 10, su representación puede observarse en la Figura 2.11; efectivamente con esta transformación se logra visualizar que la tendencia de la serie es lineal y se logra estabilizar la varianza de la serie. Puede apreciarse que en los correlogramas siguientes se incluyen dos lineas horizontales punteadas en color azul. Éstas son muy informativas pues en el caso de ruido blanco y n grande, la función de autocorrelación muestral, $\hat{\rho}_X(h)$ con $h = 1, 2, \ldots, H$ y H fijo y arbitrario, tiene una distribución aproximadamente normal con media cero y desviación estándar dada por

$$\sigma_{\hat{\rho}_x(h)} = \frac{1}{\sqrt{n}}. \tag{2.3.1}$$

Entonces, un método aproximado para evaluar si los picos en $\hat{\rho}_X(h)$ son significativos, es determinando si éstos están fuera del intervalo $\pm 2/\sqrt{n}$. Para una sucesión de ruido blanco, aproximadamente el 95 % de la muestra de funciones de autocorrelación debieran estar dentro de estos límites (Shumway & Stoffer, 2011). En el caso de la serie de Johnson & Johnson, que cuenta con 84 datos, el rango del correlograma se

obtiene de la siguiente manera:

$$\pm\frac{2}{\sqrt{n}} = \pm\frac{2}{\sqrt{84}} = \pm 0.218,$$

lo cual explica la posición en la que se grafican estos límites.

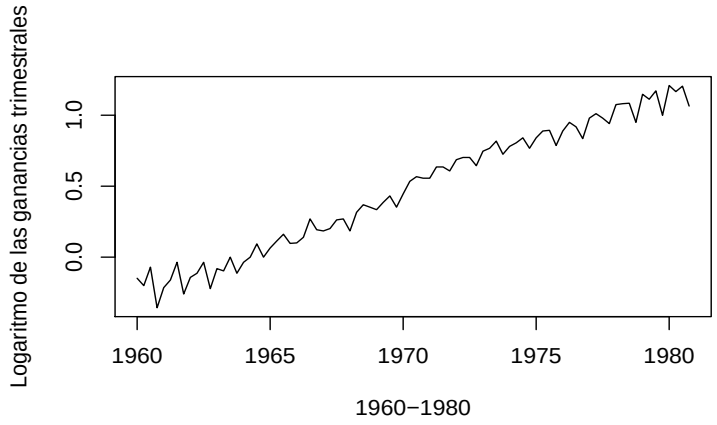


Figura 2.11: Serie de tiempo no-estacionaria, Johnson & Johnson con log(10)

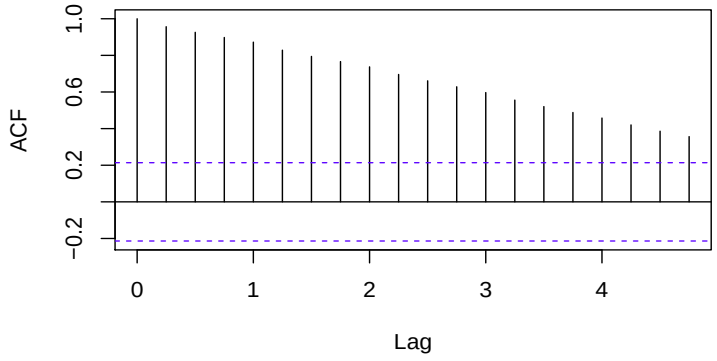


Figura 2.12: Función de autocorrelación: datos Johnson & Johnson con log(10)

En la Figura 2.12 se muestra el correlograma de la función de autocorrelación de la serie de los datos aplicando el logaritmo en base 10, y en la Figura 2.13 de la función de autocorrelación parcial de la misma serie; nótese que en la primera, los datos decaen lentamente, se puede ver que alcanzará el valor de cero para h muy grande, lo cual sugiere que nos encontramos ante una serie no-estacionaria, y en el siguiente correlograma se observa que sólo un dato es significativamente diferente de cero. De acuerdo al análisis previo que se hizo de la PACF, esto sugiere que el modelo subyacente proviene de uno $AR(1)$, sin embargo, es preciso que se haga un

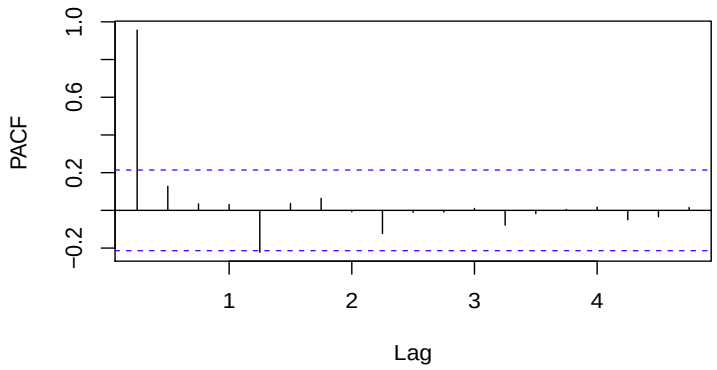


Figura 2.13: Función de autocorrelación parcial: datos Johnson & Johnson con log(10)

mayor análisis para tener certeza de dicha afirmación. Para estimar el coeficiente de correlación ρ_1 entre X_t y X_{t-1} podemos utilizar el coeficiente de correlación muestral:

$$r_1 = \frac{\sum_{t=2}^{N}(x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}, \tag{2.3.2}$$

donde

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x,$$

es la media de las primeras N observaciones.

Para calcular el coeficiente de autocorrelación entre dos observaciones con una distancia h entre ellos, se tiene la siguiente fórmula:

$$r_h = \frac{\sum_{t=h+1}^{N}(x_t - \bar{x})(x_{t-h} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}. \tag{2.3.3}$$

con $0 < h < N - 1$.

En el análisis de series de tiempo es muy importante realizar un análisis descriptivo, que combine la información gráfica y la numérica. Debe considerarse también la posibilidad de que sean necesarias transformaciones para lograr la estacionariedad en las series, aspecto importante en el análisis de éstas, ya que es un supuesto en muchas de las inferencias a realizar. Dado que en el ámbito económico abundan las series no-estacionarias, en el capítulo 4 se presentan problemas que surgen a partir de violar este supuesto de estacionariedad.

Capítulo 3

Algunos Modelos para Series de Tiempo Univariadas

En el Capítulo 2 se presentaron algunos procesos, como el de ruido blanco y de caminata aleatoria, con los cuales es posible crear procesos más complejos. En este capítulo se estudiarán los procesos autorregresivos y de promedios móviles; se presentará el proceso autorregresivo de promedios móviles (ARMA) y el autorregresivo integrado de promedios móviles (ARIMA). Todos estos procesos estacionarios son muy útiles para representar la dependencia que tienen los valores de una serie de tiempo, con sus valores pasados.

Los modelos más simples son los llamados autorregresivos; éstos generalizan la idea de regresión para representar la dependencia lineal entre una variable dependiente y una variable explicativa. En estos procesos la función de autocorrelación decae lentamente; se dice tienen una memoria larga pues el valor actual puede estar correlacionado con todos los valores previos a éste, aunque con coeficientes que van decreciendo. Estos procesos no pueden representar series de tiempo con memoria corta, como pueden hacerlo los procesos de promedios móviles, que son funciones de un número finito, generalmente pequeño, de mediciones pasadas. Combinando las propiedades de los procesos autorregresivos y los de promedios móviles, surgen los llamados procesos ARMA (autorregresivos de promedios móviles), que abarcan una amplia familia de procesos estacionarios muy útiles para representar series de tiempo.

3.1. Procesos autorregresivos

La idea principal en un proceso autorregresivo es explicar el valor actual de la serie en función de valores pasados, añadiéndole un término de error. La notación de estos modelos es generalmente $AR(p)$, donde p indica el orden del proceso, determinando el número de saltos, en el pasado, que se necesitan para predecir el valor actual.

Definición 3.1 ***Proceso autorregresivo de primer orden.*** *El proceso autorregresivo de primer orden, por sus siglas en inglés se denota como AR(1), y se define de la siguiente manera:*

$$X_t = \phi X_{t-1} + Z_t \qquad \text{donde } Z_t \sim WN(0, \sigma^2)$$

y $\phi \in \Re$.

Más delante veremos que este proceso es estacionario cuando $|\phi|$ es menor que 1. Nótese que si $\phi = 1$ se tiene un proceso de caminata aleatoria.

Para calcular la función de autocorrelación de este proceso, primeramente se calculará la covarianza entre X_t y X_{t-1}, donde

$$X_t = \phi X_{t-1} + Z_t,$$

y como $E(X_t) = 0\ \forall t$, sólo basta multiplicar la ecuación anterior por X_{t-1}:

$$X_t X_{t-1} = \phi X_{t-1}^2 + X_{t-1} Z_t,$$

para la cual calculamos su esperanza, obteniendo:

$$\begin{aligned} E(X_t X_{t-1}) &= \phi E(X_{t-1}^2) + E(X_{t-1})E(Z_t) \\ &= \phi E(X_{t-1}^2) \\ &= \phi\sigma^2. \end{aligned}$$

Finalmente la función de autocorrelación está dada por:

$$\begin{aligned} Corr(X_t, X_{t-1}) &= \frac{Cov(X_t, X_{t-1})}{\sqrt{Var(X_t)Var(X_{t-1})}} \\ &= \frac{\phi\sigma^2}{\sqrt{\sigma^2\sigma^2}} \\ &= \frac{\phi\sigma^2}{\sigma^2} \\ &= \phi. \end{aligned}$$

En general, la autocorrelación entre los procesos X_t y X_{t-h}, donde $h \geq 1$, puede expresarse (Chatfield, 2000, p. 44) como:

$$Corr(X_t, X_{t-h}) = \frac{Cov(X_t, X_{t-h})}{\sigma^2} = \phi^h. \tag{3.1.1}$$

Por otra parte, puede deducirse que la función de autocorrelación tendrá un decaimiento del tipo exponencial cuando $0 < \phi < 1$ y con signos alternados cuando $-1 < \phi < 0$.

La dependencia entre el valor presente y el pasado que un proceso $AR(1)$ establece, puede generalizarse permitiendo que X_t pueda depender no sólo de X_{t-1} sino de varios valores pasados, como puede verse en la siguiente definición de un proceso autorregresivo de orden p, $AR(p)$, el cual establece que una realización a un tiempo t es una combinación lineal de los p valores previos, más un término de ruido.

Definición 3.2 ***Proceso autorregresivo de orden*** *p. El proceso autorregresivo de orden p se denota por AR(p), y se define de la siguiente manera:*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t \qquad \text{con } Z_t \sim WN(0, \sigma^2). \tag{3.1.2}$$

Si el grado del proceso es de orden uno, entonces:

$$X_t = \phi X_{t-1} + Z_t.$$

Como se dijo anteriormente, si $\phi = 1$ se tiene una caminata aleatoria; cuando en ésta se toma la primera diferencia, se obtendrá un proceso estacionario, como puede verse a continuación:

$$\begin{aligned} X_t &= X_{t-1} + Z_t, \\ X_t - X_{t-1} &= Z_t, \\ \Delta X_t &= Z_t \quad Z_t \sim WN(0, \sigma^2). \end{aligned}$$

Ahora, si $|\phi| > 1$ la serie se considera explosiva y es no-estacionaria.

Para verificar que un proceso autorregresivo es estacionario cuando $|\phi| < 1$, haremos uso del operador de retraso B, y se probarán las tres condiciones de estacionariedad. Además de facilitar el cálculo, este operador permite hacer más compactas las ecuaciones que se manejan en el análisis de series tiempo. Básicamente, la función de este operador B, es la de recorrer el subíndice t un tiempo atrás:

$$BX_t = X_{t-1},$$

por ello se le llama operador de retraso; algunas de sus propiedades se muestran en el Apéndice A.1. Utilizando este operador tenemos que:

$$\begin{aligned} X_t &= \phi B X_t + Z_t, \\ (1 - \phi B) X_t &= Z_t. \end{aligned}$$

Luego, para $|\phi| < 1$

$$\begin{aligned} X_t &= (1 + \phi B + \phi^2 B^2 + \ldots) Z_t \\ E(X_t) &= E(Z_t) + \phi E(Z_{t-1}) + \phi^2 E(Z_{t-2}) + \ldots = 0 \\ V(X_t) &= \sigma^2 (1 + \phi^2 + \phi^4 + \ldots) \\ &= \frac{\sigma^2}{(1 - \phi^2)}. \end{aligned}$$

$$\begin{aligned} Cov(X_t, X_{t-h}) &= E[(X_t - E(X_t))(X_{t-h} - E(X_{t-h}))] \\ &= E(X_t X_{t-h}) \\ &= E[(1 + \phi B + \phi^2 B^2 + \ldots)(Z_t)(1 + \phi B + \phi^2 B^2 + \ldots)(Z_{t-h})] \\ &= 0. \end{aligned}$$

Esto es, la media, la varianza y la covarianza de la serie son constantes, si $|\phi| < 1$ y las autocorrelaciones, como se vió en (3.1.1) son de la forma ϕ^h, con $h \geq 1$, la cual tenderá a cero conforme h crece.

Ahora, el proceso $AR(p)$,

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t,$$

puede escribirse en términos del operador de retraso, como:

$$\phi_p(B)X_t = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)X_t = Z_t, \tag{3.1.3}$$

donde la ecuación $\phi_p(B) = 0$, en la que B se considera un número, real o complejo, es llamada la ecuación característica. En particular, para un proceso autorregresivo de orden 2, $AR(2)$, el cual puede expresarse como:

$$(1 - \phi_1 B - \phi_2 B^2)X_t = Z_t,$$

la estacionariedad requiere que $|\phi_2| < 1$, $\phi_2 + \phi_1 < 1$ y $\phi_2 - \phi_1 < 1$. En el Apéndice A.1 se muestran tres ejemplos que muestran cómo calcular estas raíces. Por otra parte, Guerrero (1991, p. 72) muestra que un proceso $AR(p)$ será estacionario si y sólo si las raíces de la ecuación característica se encuentran fuera del círculo unitario.

3.2. Procesos de promedios móviles

Los procesos conocidos como de promedios móviles son aquellos que explican el valor de una determinada variable en un período t, en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, que se componen de una suma de ruidos blancos, como podrá verse en las definiciones que aquí se presentan.

Definición 3.3 ***Proceso de promedios móviles de primer orden****. El proceso de promedios móviles de primer orden se denota por MA(1), y se define como:*

$$X_t = Z_t + \theta Z_{t-1} \text{ con } Z_t \sim WN(0, \sigma^2).$$

Para este proceso la media es constante e igual a cero y la función de autocorrelación puede obtenerse considerando que los términos Z_t son procesos independientes y por lo tanto:

$$E(Z_{t-h}Z_t) = 0, \qquad \forall h,$$

luego

$$\begin{aligned} Var(X_t) &= E[X_t - E(X_t)]^2 = E(X_t^2) \\ &= E(Z_t + \theta Z_{t-1})^2 \\ &= E(Z_t^2 + 2\theta Z_t Z_{t-1} + \theta^2 Z_{t-1}^2) \\ &= E(Z_t^2) + 2\theta E(Z_t Z_{t-1}) + \theta^2 E(Z_{t-1}^2) \\ &= \sigma^2 + \theta^2\sigma^2 \\ &= \sigma^2(1 + \theta^2), \end{aligned} \tag{3.2.1}$$

$$\begin{aligned} Cov(X_{t-1}, X_t) &= E[(X_{t-1} - E(X_{t-1}))(X_t - E(X_t))] \\ &= E[(X_{t-1})(X_t)] \\ &= E[(Z_{t-1} + \theta Z_{t-2})(Z_t + \theta Z_{t-1})] \\ &= E(Z_{t-1}Z_t) + \theta E(Z_{t-1}^2) + \theta E(Z_t Z_{t-2}) + \theta^2 E(Z_{t-2}Z_{t-1}) \\ &= \theta\sigma^2. \end{aligned} \tag{3.2.2}$$

De (3.2.2) y (3.2.1), se tiene que:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{Cov(X_{t-1}, X_t)}{V(X_t)} = \frac{\theta\sigma^2}{\sigma^2(1+\theta^2)} = \frac{\theta}{1+\theta^2}. \tag{3.2.3}$$

Nótese que en (3.1.1), la correlación de dos procesos con distancia $h \geq 1$, ϕ^h, tiende a cero, si $\phi < 1$ y h tiende a infinito; pero en (3.2.3), la correlación es constante, sin tomar en cuenta que tan separadas estén las variables de dicho proceso. A continuación se define el proceso de promedios móviles de orden q. Los procesos de promedios móviles son muy populares pues suavizan los datos de una serie y permiten identificar tendencias; esto es muy útil en mercados volátiles.

Definición 3.4 ***Proceso de promedios móviles de orden*** q. *El proceso de promedios móviles de grado q, denotado por MA(q), y se define como:*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad \textit{donde } Z_t \sim WN(0, \sigma^2). \tag{3.2.4}$$

Al igual que se hizo en los procesos autorregresivos, puede utilizarse el polinomio de retraso para escribir la ecuación anterior como:

$$X_t = \theta(B) Z_t, \tag{3.2.5}$$

donde $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ es el polinomio en B de orden q.

Por otra parte, es posible escribir cualquier proceso estacionario $AR(p)$ como un $MA(\infty)$. Por ejemplo considere:

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t, \\ X_{t-1} &= \phi X_{t-2} + Z_{t-1}, \end{aligned}$$

luego,

$$\begin{aligned} X_t &= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t \\ &= \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\ &= \phi^3 X_{t-3} + \phi^2 Z_{t-2} + \phi Z_{t-1} + Z_t. \end{aligned}$$

Considerando que $|\phi| < 1$, el valor ϕ^i será pequeño cuando i sea grande y al final obtendremos:

$$X_t = Z_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots$$

que resulta un proceso $MA(\infty)$. En el caso de procesos de promedios móviles, éstos pueden convertirse en procesos autorregresivos si satisfacen el que sean invertibles, lo cual se define a continuación.

Definición 3.5 ***Proceso*** $MA(p)$ ***invertible***. *Un proceso* $MA(p)$ *es invertible si puede escribirse de la forma siguiente:*

$$X_t = \sum_{1}^{\infty} \phi_j X_{t-j} + Z_t \quad \textit{donde } Z_t \sim WN(0, \sigma^2).$$

Por ejemplo, consideremos un proceso de promedios móviles de primer orden, $MA(1)$

$$X_t = Z_t - \theta Z_{t-1}, \tag{3.2.6}$$

el cual puede reescribirse como:

$$Z_t = X_t + \theta Z_{t-1}. \tag{3.2.7}$$

Ahora, de la ecuación anterior, se intercambia t por *t-1*:

$$Z_{t-1} = X_{t-1} + \theta Z_{t-2}, \tag{3.2.8}$$

y sustituyendo (3.2.8) en la ecuación (3.2.7) se tiene:

$$Z_t = X_t + \theta X_{t-1} + \theta^2 Z_{t-2}.$$

Si $|\theta| < 1$, continuamos el proceso de diferenciación anterior infinitas veces y obtenemos lo siguiente:

$$Z_t = X_t + \theta X_{t-1} + \theta^2 Z_{t-2} + \theta^3 Z_{t-3} + \cdots$$

o bien,

$$X_t = Z_t - \theta X_{t-1} - \theta^2 Z_{t-2} - \theta^3 Z_{t-3} - \cdots$$

Y así, el modelo *MA(1)* ha sido convertido en un modelo $AR(\infty)$. Se dice que el modelo *MA(1)* es invertible si y sólo si $|\theta| < 1$. Entonces, dado un modelo de probabilidad de una serie de tiempo, podemos encontrar diferentes maneras de representarlo y la selección dependerá del problema a trabajar. La importancia de la condición de invertibilidad es que todo proceso invertible está determinado, de una manera única, por su función de autocorrelación, aspecto que no se cumple en los procesos que no son invertibles. La importancia de esta condición se enfatiza para los modelos de promedios móviles, a continuación se presenta un ejemplo que muestra cómo dos procesos de MA(1) tienen la misma función de autocorrelación, es decir, para estos procesos la ACF no es suficiente si se pretende especificar el modelo subyacente, ya que se puede prestar fácilmente a confusiones.

Ejemplo 3.1 *Considérese los siguientes procesos MA(1):*

$$X_t = Z_t + \theta Z_{t-1},$$
$$X'_t = Z'_t + \frac{1}{\theta} Z'_{t-1},$$

donde $Z_t \sim WN(0, \sigma^2)$, $Z'_t \sim WN(0, \sigma'^2)$ *y* $\theta \in (-1, 1)$.

Para calcular la función de autocorrelación de cada proceso es necesario calcular su varianza y covarianza. Suponiendo que Z_t es un proceso independiente, entonces

$$E[(Z_{t-h})(Z_t)] = 0, \qquad \forall h.$$

En el caso del proceso $X_t = Z_t + \theta Z_{t-1}$, en (3.2.3) se mostró que su función de autocorrelación resulta $\rho_1 = \frac{\theta}{1+\theta^2}$.

Ahora, para el proceso $X'_t = Z'_t + \frac{1}{\theta} Z'_{t-1}$ calculemos

$$\begin{aligned} V(X'_t) &= E[X'_t - E(X'_t)]^2 = E(X_t^{'2}) \\ &= E(Z'_t + \frac{1}{\theta} Z'_{t-1})^2 \\ &= E(Z_t^{'2} + 2\frac{1}{\theta} Z'_t Z'_{t-1} + \frac{1}{\theta^2} Z_{t-1}^{'2}) \\ &= E(Z_t^{'2}) + 2\frac{1}{\theta} E(Z'_t Z'_{t-1}) + \frac{1}{\theta^2} E(Z_{t-1}^{'2}) \\ &= \sigma^2 + \frac{1}{\theta^2}\sigma^2 \\ &= \sigma^2(1 + \frac{1}{\theta^2}), \end{aligned} \tag{3.2.9}$$

$$\begin{aligned} Cov(X'_{t-1}, X'_t) &= E[(X'_{t-1} - E(X'_{t-1}))(X'_t - E(X'_t))] \\ &= E[(X'_{t-1})(X'_t)] \\ &= E[(Z'_{t-1} + \frac{1}{\theta} Z'_{t-2})(Z_t + \frac{1}{\theta} Z'_{t-1})] \\ &= E(Z'_{t-1} Z'_t) + \frac{1}{\theta} E(Z_{t-1}^{'2}) + \frac{1}{\theta} E(Z'_t Z'_{t-2}) + \frac{1}{\theta^2} E(Z'_{t-2} Z'_{t-1}) \\ &= \frac{1}{\theta}\sigma^2. \end{aligned} \tag{3.2.10}$$

A partir de (3.2.9) y (3.2.10), tenemos que:

$$\rho_1^{'} = \frac{Cov(X'_{t-1}, X'_t)}{V(X'_t)} = \frac{\frac{1}{\theta}\sigma^2}{\sigma^2(1 + \frac{1}{\theta^2})} = \frac{\theta}{1+\theta^2}. \tag{3.2.11}$$

Y así, los procesos mostrados en este ejemplo tienen la misma función de autocorrelación. Por ello, dada una función de autocorrelación, no es posible estimar un único proceso de promedios móviles, sin establecer alguna otra restricción, como las expuestas por Guerrero (1991, p.78).

Para un proceso $MA(q)$, la función de autocorrelación puede escribirse como (Chatfield, 2000, p. 46):

$$\rho_h = \begin{cases} 1 & \text{si } h = 0, \\ \frac{\sum_{i=0}^{q-h} \theta_i \theta_{i+h}}{\sum_{i=0}^{q} \theta_i^2} & \text{si } h = 1, 2, \ldots, q, \\ 0 & \text{si } h > q. \end{cases} \tag{3.2.12}$$

Donde $\theta_0 = 1$. Nótese que si el valor de h es mayor a q, la función de autocorrelación se anula; esto quiere decir que la correlación entre dos procesos, X_t y X_s, es cero si la distancia entre t y s es mayor a q. Entonces éste es un indicador que nos ayuda a calcular el orden del proceso de promedios móviles.

3.3. Generación de algunos procesos

Para ilustrar el comportamiento de los procesos que se explicaron anteriormente, consideremos el siguiente ejemplo donde se genera una muestra aleatoria de 100 datos de un proceso de ruido blanco Z_t $N(0,1)$, los cuales servirán para crear un proceso de promedios móviles, un proceso autorregresivo y una caminata aleatoria. En la Tabla 3.1 pueden observarse las primeras ocho realizaciones y en el Apéndice C.1 se muestran todos los datos generados.

Tabla 3.1: Construcción de un proceso estocástico

Tiempo	Ruido blanco	Promedios móviles	Autorregresivo	Caminata aleatoria
1	-1.6331	NA	-1.6331	-1.6331
2	-0.3666	-1.8363	-1.8364	-1.9997
3	-0.8728	-1.2027	-2.5255	-2.8725
4	-0.2469	-1.0324	-2.5198	-3.1194
5	0.1266	-0.0956	-2.1413	-2.9928
6	-0.2462	-0.1322	-2.1733	-3.2390
7	0.8541	0.6325	-1.1019	-2.3849
8	1.5789	2.3475	0.5871	-0.8060
⋮	⋮	⋮	⋮	⋮

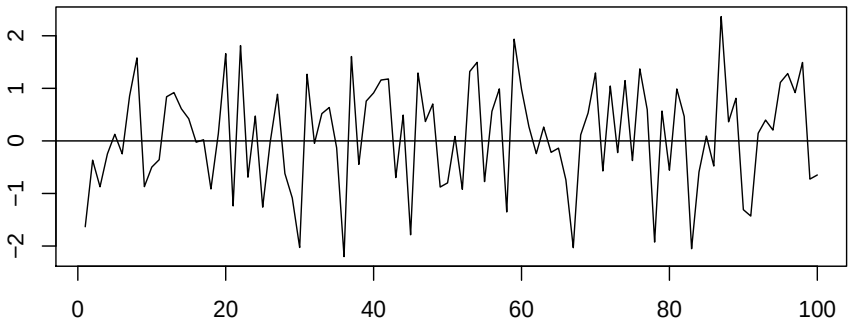


Figura 3.1: Ruido Blanco

En la Figura 3.1 se muestra la serie de tiempo que se obtiene con los datos de ruido blanco, mientras que en las Figura 3.2 se observa el correlograma de la función de autocorrelación de estos datos. Recordemos que este proceso es estacionario, por lo tanto, se espera que los valores que toma la función de autocorrelación se mantengan

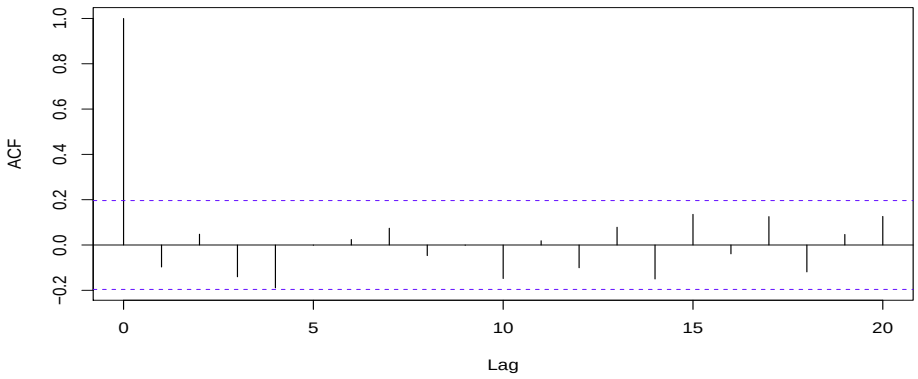


Figura 3.2: Función de autocorrelación de ruido blanco

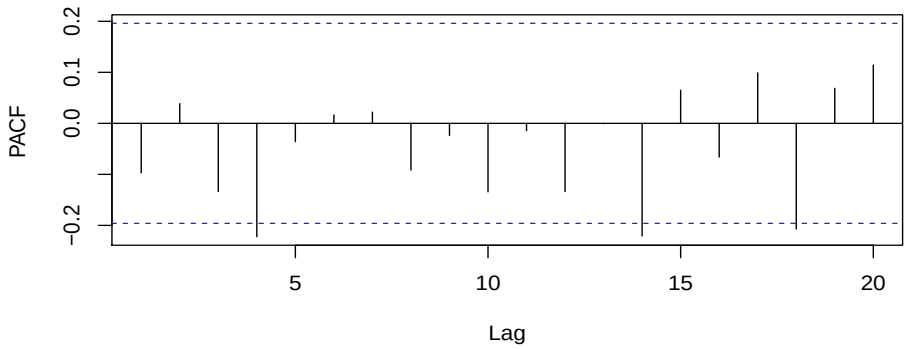


Figura 3.3: Función de autocorrelación parcial de ruido blanco

dentro del rango 0.2, es decir, que no sean significativamente diferentes de cero. En realidad se espera, debido a la variación muestral, que aproximadamente un 5 % de los valores de esta función sean significativamente diferentes de cero (Cowpertwait & Metcalfe, 2009, p. 70). La Figura 3.3 muestra la función de autocorrelación parcial que se obtiene para esta serie.

Ahora, con los datos de ruido blanco se contruye el proceso de promedios móviles de orden 1. Para ello consideramos $\theta = 0.9$, y mediante el cálculo $X_t = \theta Z_{t-1} + Z_t$, $t = 2, 3, \ldots, 100$, obtenemos la serie de tiempo mostrada en la Figura 3.4, donde puede apreciarse cómo cambia el comportamiento de este proceso; en la segunda columna de la Tabla 3.1 se pueden observar los primeros siete datos de esta serie. Por otra parte, en la Figura 3.5 se muestra el correlograma de la ACF, y según el análisis de los correlogramas que se realizó en la sección 2.3, ésto nos da un indicio del orden del modelo subyacente, en este caso, se sabe que el orden es 1; esto puede observarse en el correlograma ya que el primer gráfico de línea corresponde a

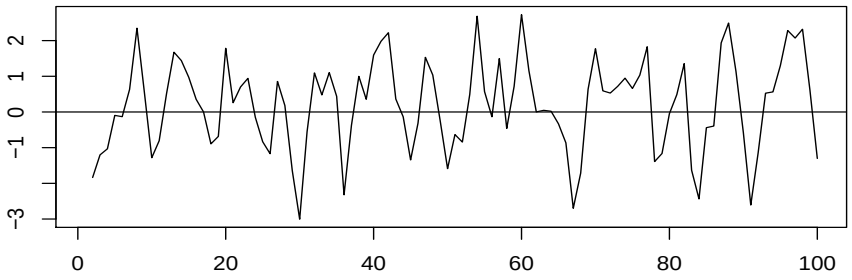


Figura 3.4: Promedios Móviles

$k = 0$, y alcanza una altura de 1, como era de esperarse a partir de (3.2.12). Ahora, teóricamente esperaríamos que para $k = 1$, $\rho_1 = \frac{0.9}{1+9.9^2} = 0.4972$, y a partir de la Figura 3.5 podría decirse que ρ_1 es significativamente diferente de cero; ésto no puede afirmarse para ρ_2, con lo cual deducimos que se trata de un proceso de promedios móviles de orden uno. Por último, en la Figura 3.6 se muestra el correlograma de la PACF del modelo con $\theta = 0.9$; en esta gráfica se observa que el comportamiento de la gráfica es alternando valores de positivo a negativo.

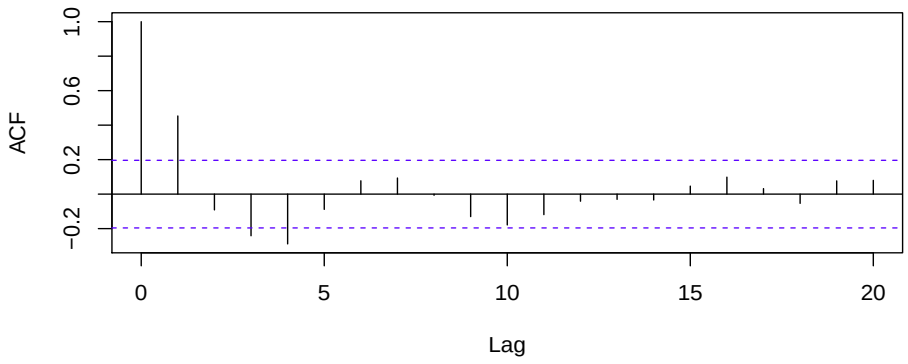


Figura 3.5: Función de autocorrelación de promedios móviles con $\theta = 0.9$

En la Tabla 3.1 también se presentan las primeras realizaciones del proceso autorregresivo de orden 1, $X_t = \phi X_{t-1} + Z_t$ con $t = 2, 3, \ldots, 100$, mostrado en la Figura 3.7, en la cual se considera $\phi = 0.9$. La gráfica de la función de autocorrelación puede observarse en la Figura 3.8, en este caso el comportamiento es sinusoidal, por lo cual es importante señalar que la función de autocorrelación muestral raramente se ajustará de manera perfecta al patrón teórico esperado. En la Figura 3.9 se observa que en el correlograma de la PACF sólo un valor es significativo, lo cual sugiere que

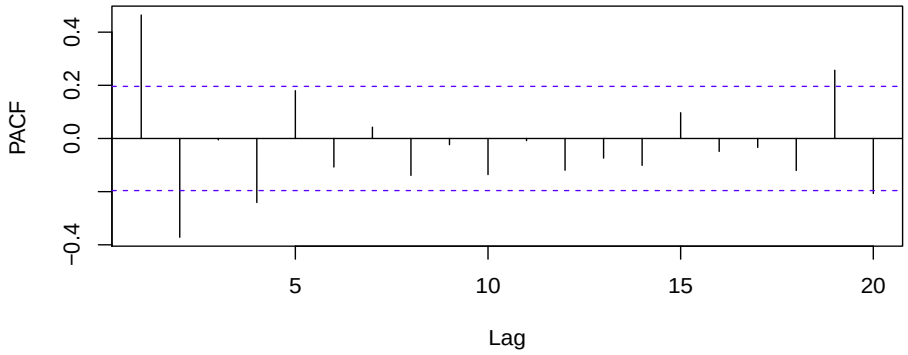


Figura 3.6: Función de autocorrelación parcial de promedios móviles con $\theta = 0.9$

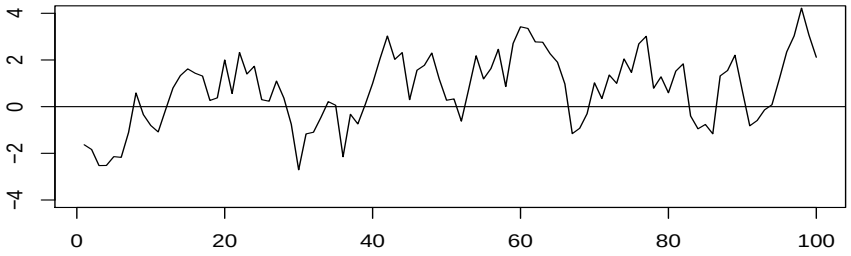


Figura 3.7: Proceso autorregresivo con $\phi = 0.9$

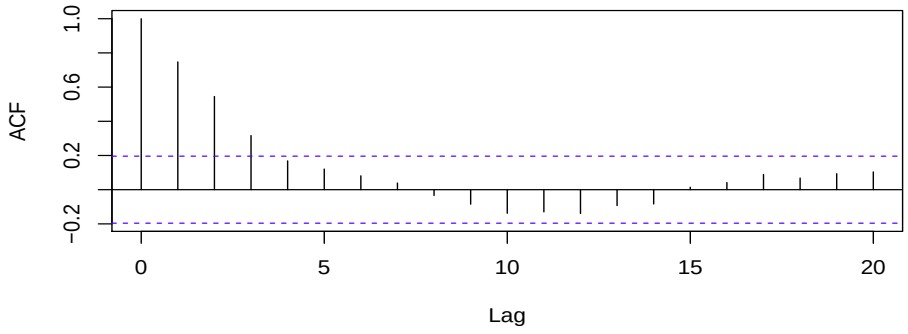


Figura 3.8: Función de autocorrelación de un proceso autorregresivo con $\phi = 0.9$

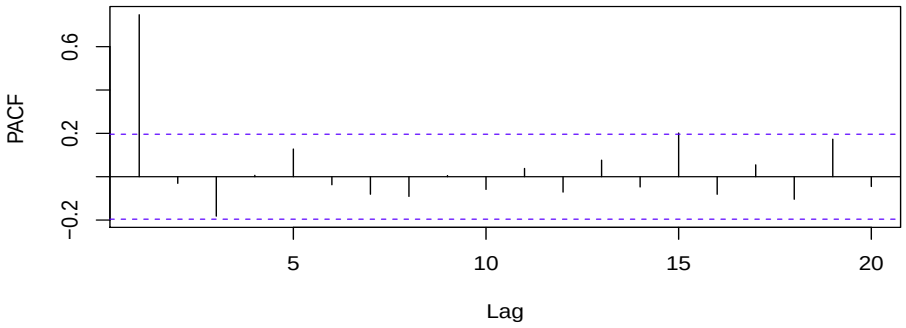


Figura 3.9: Función de autocorrelación parcial de un proceso autorregresivo con $\phi = 0.9$

el orden del modelo subyacente es 1.

Finalmente, para construir el proceso de caminata aleatoria se tomaron las sumas acumuladas de ruido blanco, es decir, $\sum_{t=1}^{100} Z_t$, o bien, también puede generarse por medio de un proceso autorregresivo con $\phi = 1$, $X_t = X_{t-1} + Z_t$; el comportamiento de este proceso se puede observar en la Figura 3.10.

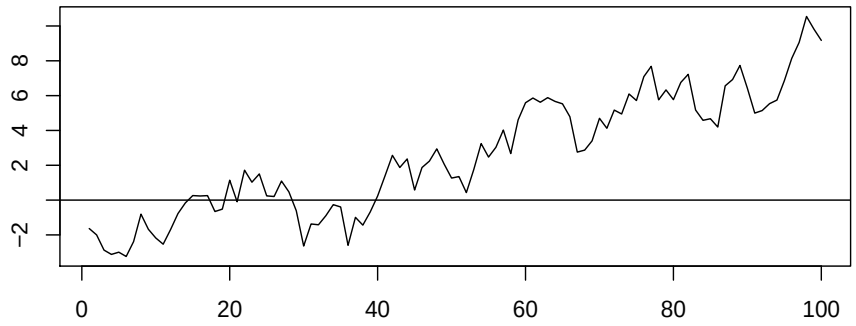


Figura 3.10: Caminata aleatoria

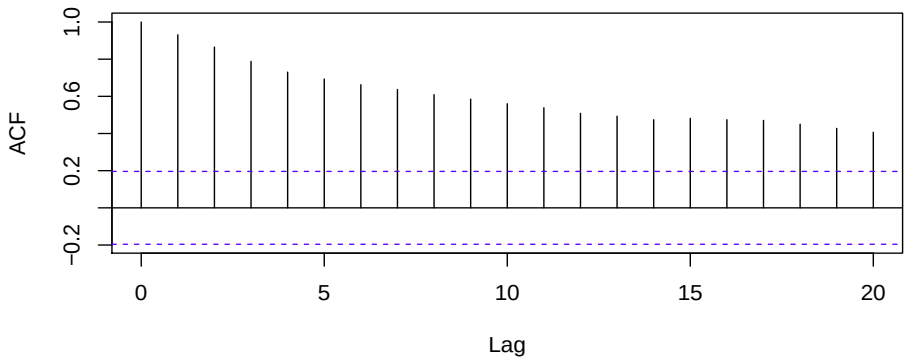


Figura 3.11: Función de autocorrelación de caminata aleatoria

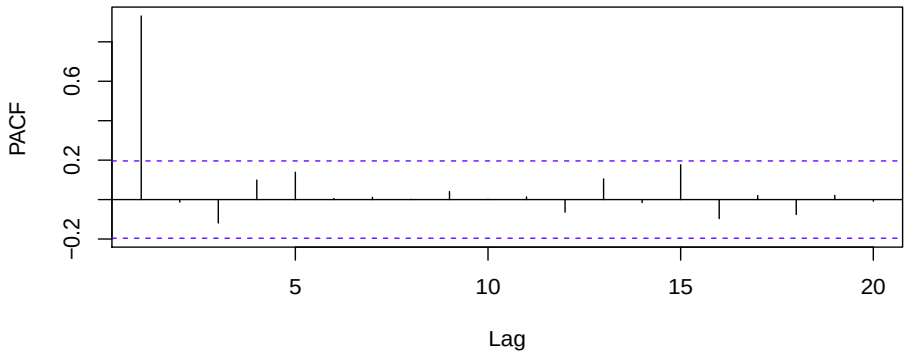


Figura 3.12: Función de autocorrelación parcial de caminata aleatoria

Su correspondiente función de autocorrelación puede observarse en la Figura 3.11, donde evidentemente se deduce que se trata de una serie no-estacionaria, pues los valores que toma la función están fuera del rango 0.2, que teóricamente define los límites para $n = 100$.

3.3.1. Ejemplos de correlogramas para algunos procesos

La Tabla 3.2 es un resúmen del comportamiento en el correlograma de los procesos expuestos en la sección anterior (Shumway, 2011, p. 75).

A continuación, se mostrarán ejemplos de correlogramas de la función de autocorrelación y de la función de autocorrelación parcial para procesos de promedios móviles de orden dos y de orden tres; y también para los procesos autorregresivos de orden dos y de orden tres. Éstos fueron creados con el Software R, con una $N(0, 1)$.

Tabla 3.2: Comportamiento de modelos $AR(p)$, $MA(q)$ y $ARMA(p,q)$

	$MA(q)$	$AR(p)$	$ARMA(p,q)$
ACF	Se corta antes del retraso q	Decrece	Decrece
PACF	Decrece	Se corta antes del retraso p	Decrece

En la Figura 3.13, se aprecian dos correlogramas para un proceso MA(2); el que se

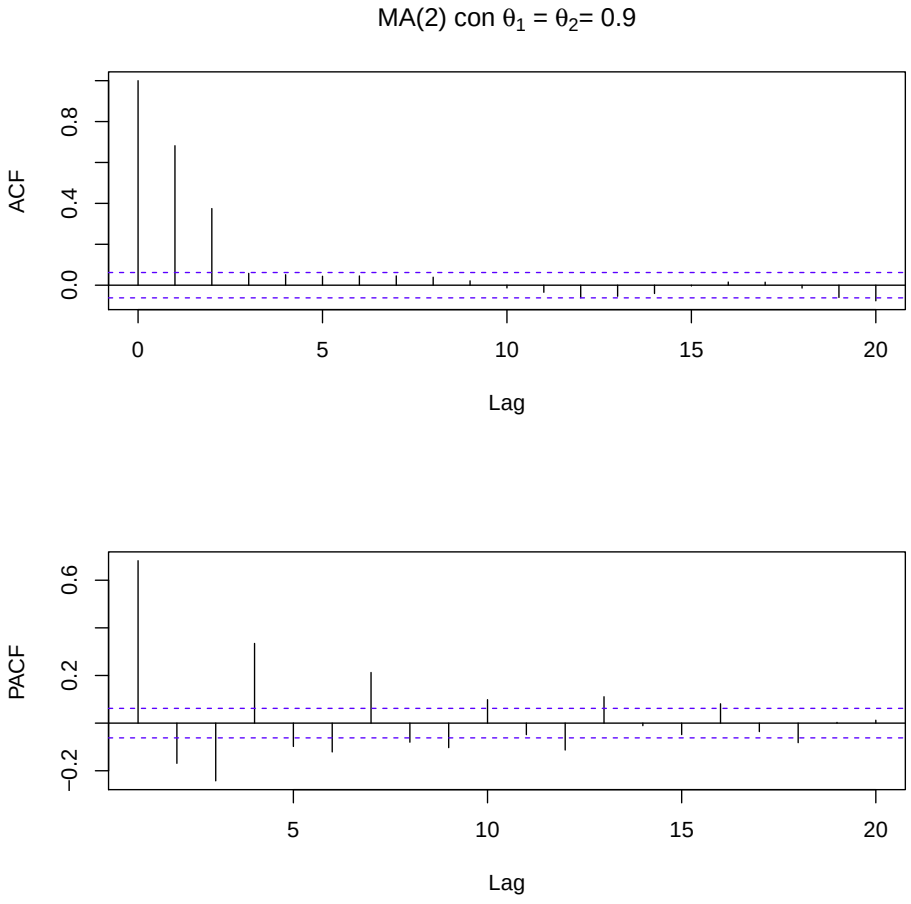


Figura 3.13: Correlograma de un proceso $MA(2)$

tomará en cuenta para visualizar el orden del modelo subyacente será el de la *ACF*, en éste se aprecia que el primer valor es uno, y después vienen los valores significativos, los cuales son dos, por lo tanto, se puede intuir que el orden del proceso es

dos, y efectivamente así es, ya que el modelo del cual se originó es de tal orden. De la misma forma, se concluye para Figura 3.14, en este caso, los valores significativos son tres, para un modelo de orden tres. Además se puede decir que ambos procesos son estacionarios. Por otra parte, se puede observar que en el correlograma para la PACF, los valores se van alternando, además se puede notar para estos casos en particular, que depende el orden, es el número de valores negativos que se toman antes de tornar uno positivo.

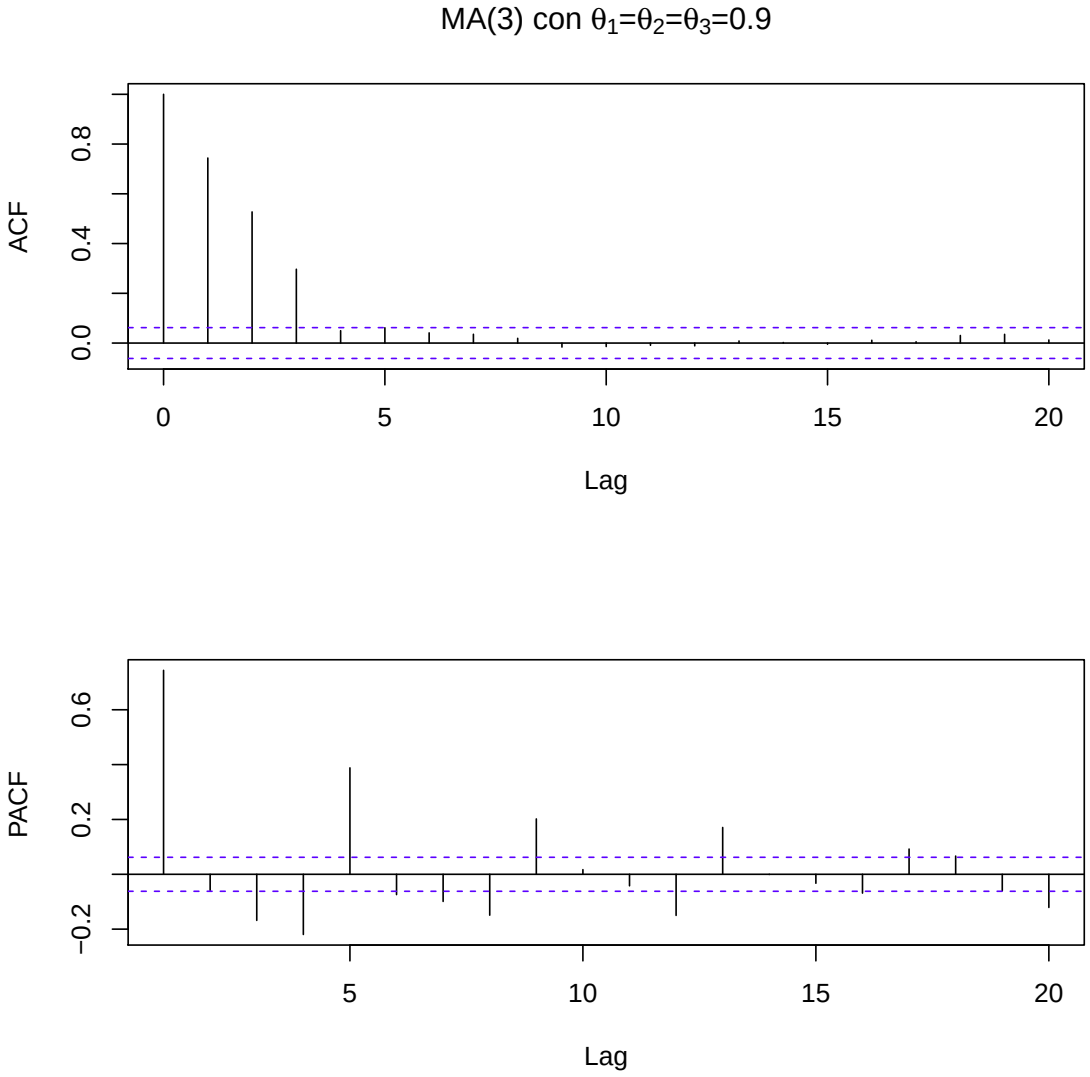


Figura 3.14: Correlograma de un proceso $MA(3)$

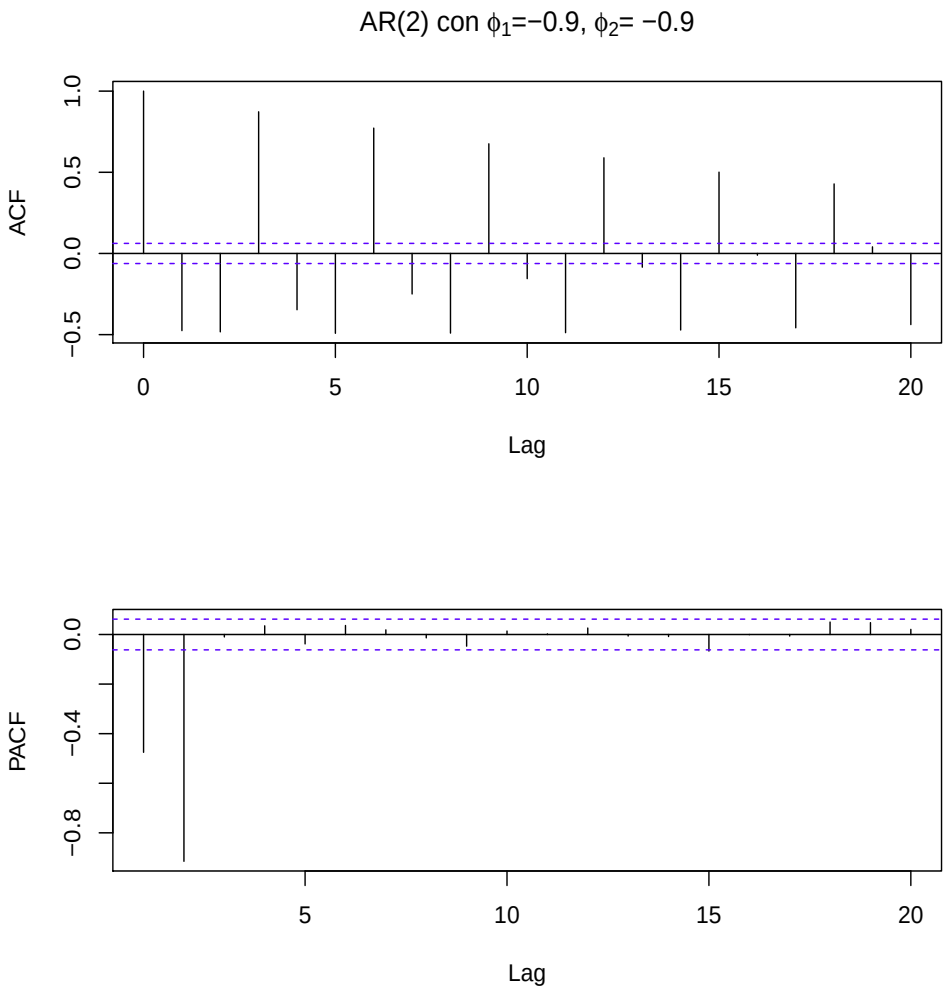


Figura 3.15: Correlograma de un proceso $AR(2)$

Luego, en la Figura 3.15 y 3.16, se aprecian los correlogramas para un procesos AR(2) y AR(3); en este caso, el correlograma que ayudará a visualizar el orden del modelo subyacente será el de la PACF; por ejemplo en la Figura 3.15 se tiene que son dos los valores significativos los cuales son valores negativos, ésto pudiera dar una idea del signo del coeficiente del modelo subyacente. De igual manera, en la Figura 3.16, los valores significativos son tres, y éstos se alternan de acuerdo al signo que corresponde cada coeficiente del modelo que fueron creados. Por otra parte, si se visualiza el correlograma para la ACF, éstos parecen tener relación con el signo de cada coeficiente, alternando valores.

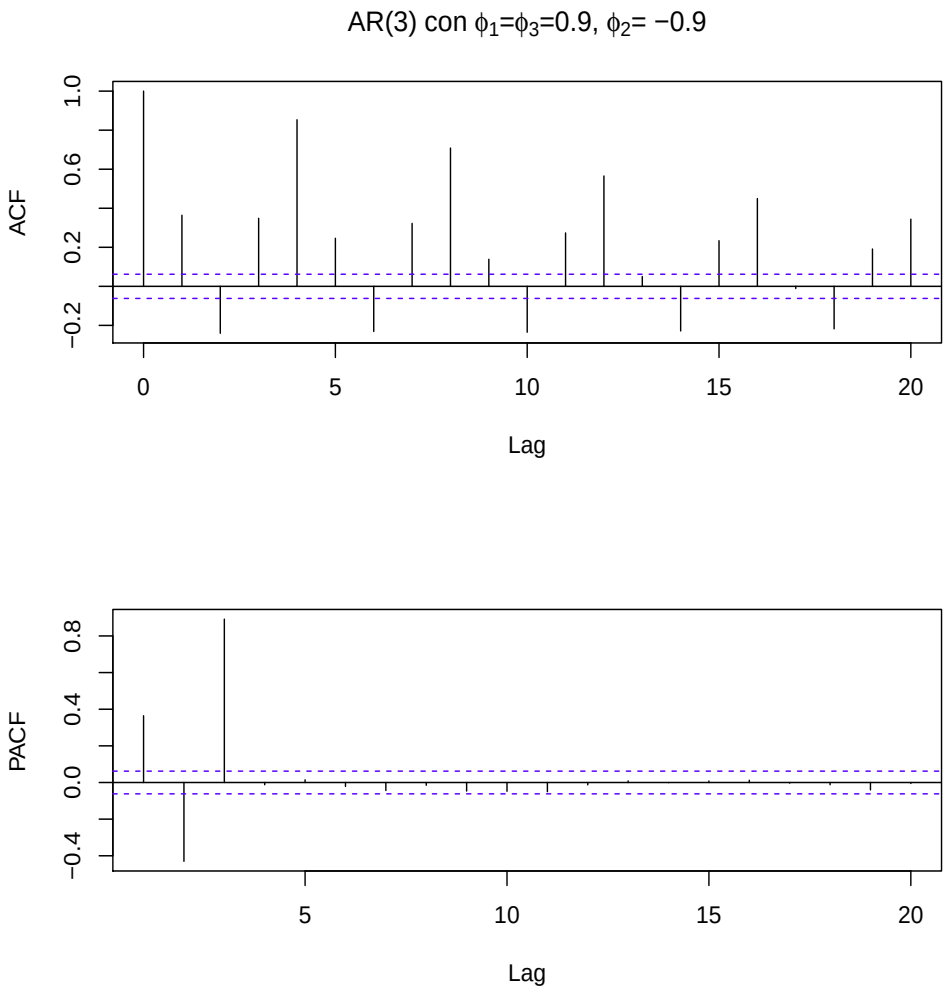


Figura 3.16: Correlograma de un proceso $AR(3)$

3.4. Otros procesos importantes

Al igual que un proceso de ruido blanco permite generar procesos más complejos, como el autorregresivo y el de promedios móviles, a su vez, a partir de éstos también pueden generarse otro tipo de procesos, tal es el caso de los *ARMA(p,q)*. La importancia de estos procesos es que muchos datos reales pueden ser aproximados, utilizando menos parámetros, cuando se mezclan procesos autorregresivos con procesos de promedios móviles. Por lo general, cuando se habla de procesos de Box & Jenkins, se tiene una mezcla de procesos autorregresivos de orden p con procesos de promedios móviles de orden q. Por otra parte, cuando un proceso tiene que ser diferenciado d veces antes de ajustar un proceso *ARMA(p,q)*, entonces, el modelo original para la serie no diferenciada será un modelo *ARIMA(p,d,q)*. A continuación se proporciona la definición de estos dos procesos.

Definición 3.6 ***Proceso autorregresivo de promedios móviles de orden (p,q).*** *Al proceso estocástico X_t, $t \in \mathbb{Z}$ se le llama proceso autorregresivo de promedios móviles de orden (p, q), y se denota por ARMA (p, q), si el proceso es estacionario y satisface la ecuación de diferencia estocástica lineal de la forma:*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{3.4.1}$$

con $Z_t \sim WN(0, \sigma^2)$, *con* $\phi, \theta \in \Re$ *y* $\phi_p \theta_q \neq 0$.

Usando el operador B, la ecuación (3.4.1) se puede reescribir como:

$$\phi(B) X_t = \theta(B) Z_t. \tag{3.4.2}$$

Ahora, X_t es llamado un proceso *ARMA(p,q)* con media μ, si $X_t - \mu$ es un proceso *ARMA (p,q).*

Como se ha mencionado a través de esta tesis, muchas series de tiempo son no-estacionarias y por lo tanto no es correcto ajustarles un modelo $AR(p)$, $MA(q)$, o $ARMA(p, q)$ directamente; por ello se realizan ciertas transformaciones que logran la estacionariedad de estas series de tiempo. Una de estas transformaciones es la diferenciación, la cual se utiliza cuantas veces sea necesario para remover la no-estacionariedad de la serie de tiempo. Por ejemplo, la primera diferencia sería $X_t - X_{t-1} = (1 - B)X_t$, la d-ésima diferencia sería $(1 - B)^d X_t$. Es claro que después de estas transformaciones cierta información puede perderse, por ello es preciso preguntarnos si lo que se va a eliminar es importante o no para las inferencias del proceso; si ese es el caso, habrá que buscar otros métodos para quitar la no-estacionariedad.

Definición 3.7 ***Proceso autorregresivo integrado de promedios móviles de orden (p, d, q).*** *Se dice que un proceso X_t es un proceso ARIMA(p,d,q), el cual puede expresarse como:*

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t,$$

si la serie de tiempo se diferencia d-veces antes de ajustar un modelo ARMA (p,q).

La letra I del acrónimo ARIMA, denota la palabra *integrado.*

En este capítulo hemos visto sólo algunos de entre varios procesos que permiten modelizar series de tiempo. Se ha retomado también el concepto de estacionariedad, aspecto importante que debe cumplirse y para el cual deben tenerse herramientas específicas que permitan identificar cuándo una serie es o no-estacionaria. Cuando se realizan inferencias con una serie no-estacionaria, se corre el riesgo de obtener resultados erróneos, los cuales se identifican como resultados espurios. Algunas de las pruebas más utilizadas para verificar la estacionariedad de una serie, se presentan en el Capítulo 4. Estas pruebas son muy utilizadas en investigaciones económicas, donde abundan las series de tiempo no-estacionarias.

Capítulo 4

El Problema de Regresión Espuria

Uno de los supuestos de los modelos clásicos de regresión es que las series de datos que se estén considerando sean estacionarias y que los errores del modelo tengan media cero y una varianza finita. En la práctica esto raramente se cumple; por ejemplo, en el ámbito económico las series de tiempo por lo general muestran una tendencia positiva o negativa, a veces con una varianza no constante, y ello origina que éstas sean no-estacionarias. El concepto de estacionariedad es importante pues sin ella, estadísticos como medias, varianzas y correlaciones no podrían describir adecuadamente a los datos, para cada tiempo o puntos de interés. Por ello, este capítulo se centra en presentar y utilizar dos pruebas de raíz unitaria, las cuales son las más recurridas en el área de economía, estas son: la *prueba de Dickey-Fuller* y la *prueba de Phillips-Perron* para verificar la estacionariedad en series de tiempo.

4.1. Regresión espuria

Cuando dos series de tiempo independientes muestran una fuerte correlación, nos encontramos ante un fenómeno denominado *correlación espuria*. Al parecer Yule (1926) fue el primero en mostrar este fenómeno, en un artículo donde estudia la proporción de casamientos ocurridos en los templos de Inglaterra y el índice de mortalidad, esto durante el período de 1866 a 1911 y donde obtiene una correlación de 0.95, valor a partir del cual erróneamente podría concluirse que hay una relación causa-efecto entre contraer matrimonio y el índice de mortalidad, lo cual no tiene sentido y es similar a muchos otros ejemplos, como el que años más tarde Hofer *et al.* (2004) retoman, con datos obtenidos en Berlín y donde observan una fuerte correlación entre el aumento de la población de cigüeñas y número de nacimientos que no son atendidos en hospitales, ejemplo con el cual podría pensarse que es cierto el cuento tradicional de la cigüeña.

Un pasatiempo interesante es tomar dos series de tiempo cuya relación no tenga sentido, esto es, que no exista causa-efecto, pero tales que la tendencia de ambas series sea similar, o bien, una contraria a la otra, y tratar de predecir el efecto que una variable tiene sobre la otra por medio de una regresión, que puede resultar significativa pero que en realidad será una *regresión espuria*. En 2014 se volvió viral el sitio de internet http://tylervigen.com/spurious-correlations, que muestra la correlación que resulta al analizar diferentes series de datos. Como ejemplo, al correlacionar lo que en Estados Unidos se gasta en ciencia, espacio y tecnología, con lo que se gasta en mascotas, resulta una correlación de 0.979564; y así como ésta, el sitio muestra

una gran variedad de correlaciones espurias que bien podrían confundir a más de uno y que resultan de no considerar que las series bajo análisis son no-estacionarias.

Granger & Newbold (1974) efectuaron un análisis detallado de las consecuencias de violar, en los modelos de regresión, la suposición de estacionariedad; mostraron, a través de simulaciones Monte Carlo, que dadas dos caminatas aleatorias es muy probable que la regresión de una en la otra produzca un coeficiente para la pendiente que resulte significativo, de acuerdo con la prueba t usual, mostrada en el Apéndice B.3. Para analizar esta situación, Granger & Newbold (1974) básicamente simularon dos caminatas aleatorias:

$$Y_t = Y_{t-1} + Z_t \qquad t = 1, 2, \ldots, n, \tag{4.1.1}$$

$$X_t = X_{t-1} + Z'_t \qquad t = 1, 2, \ldots, n, \tag{4.1.2}$$

en las que Z_t y Z'_t son procesos de ruido blanco, independientes uno del otro y calcularon la regresión:

$$Y_t = a_0 + a_1 X_t + \epsilon_t, \tag{4.1.3}$$

obteniendo, al nivel de significancia del 5 %, que aproximadamente rechazaban la hipótesis nula $H_0 : a_1 = 0$, el 75 % de las veces, cuando una prueba adecuada conduciría a un rechazo en aproximadamente un 5 % de las veces (Apéndice A.4). Por otra parte, observaron valores muy altos para el coeficiente de determinación R^2 (Apéndice B.5), observando residuales correlacionados.

Cuando se realiza un análisis de regresión en series no-estacionarias, generalmente (Granger & Newbold, 1974):

- Los estimadores de mínimos cuadrados no son consistentes (Apéndice A.2).
- Las predicciones basadas en las ecuaciones de regresión son sub-óptimas.
- Las pruebas comunes para realizar inferencia estadística no son válidas.

En realidad, cuando las series de datos en una regresión son no-estacionarios, el cálculo del coeficiente de determinación R^2 y los valores que toman los estadísticos de prueba del modelo de regresión, no siguen las distribuciones usuales.

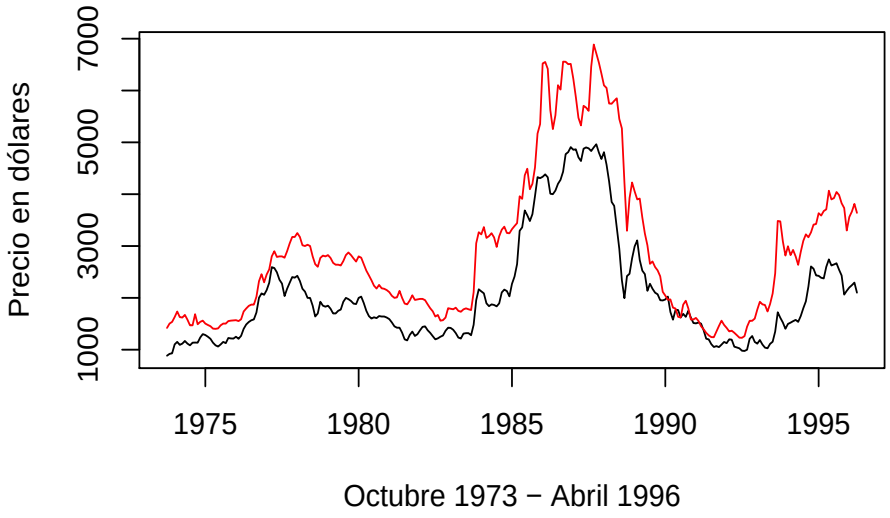


Figura 4.1: Precios por tonelada de pimienta en Estados Unidos

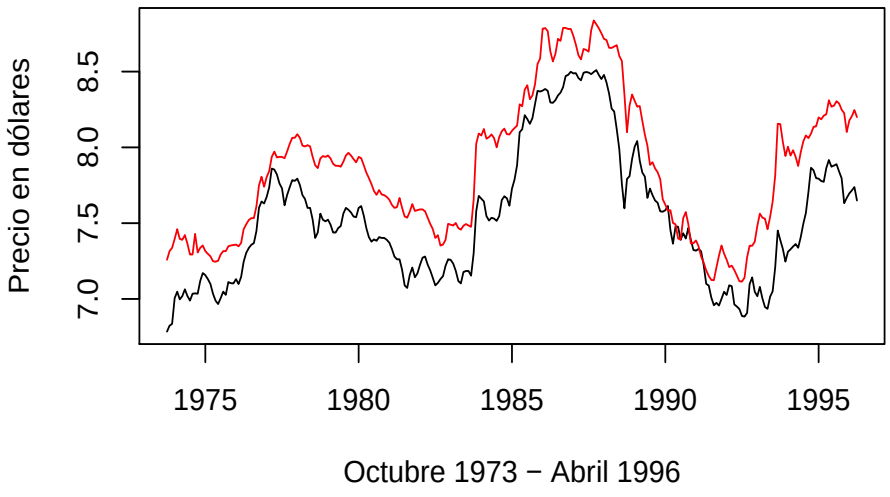


Figura 4.2: Logaritmo de los precios por tonelada de pimienta, Estados Unidos

Consideremos las series que se muestran en la Figura 4.1, donde se aprecian los datos de precios promedio de pimienta blanca (serie en color rojo) y los datos de precios promedio de pimienta negra (serie en color negro), ambos de Estados Unidos;

los datos están dados por precio por tonelada en dólares y son períodos mensuales de octubre de 1973 a abril de 1996, con un total de 271 observaciones. Como estos productos pueden ser sustitutos, es razonable esperar que los precios de ambos se parezcan a través del tiempo.
En la Figura 4.1 puede observarse que ambas series tienen un comportamiento similar. Consideremos el siguiente modelo de regresión lineal sin intercepto:

$$Y_t = \beta X_t + \epsilon_t,$$

en donde se tiene como variable dependiente, Y_t, el precio de pimienta blanca y como variable independiente, X_t, el precio de pimienta negra, y ϵ_t el error residual. Si se realiza la prueba de hipótesis siguiente:

$$H_0 : \beta = 0,$$
$$H_1 : \beta \neq 0,$$

se obtienen resultados que se muestran en la Tabla 4.1, y se observa que el coeficiente asociado a la pendiente del modelo es significativo (*p-valor*<2e-16), es decir, hay relación entre las variables; además el coeficiente de determinación resulta 0.9809, es decir, el modelo explica aproximadamente un 98 % de la variabilidad en los precios de pimienta blanca. Por otra parte, cuando se tiene el estadístico de Durbin-Watson, se tiene la siguiente prueba de hipótesis:

$$H_0 : \rho = 0,$$
$$H_1 : \rho > 0,$$

toma un valor de cercano a cero, resultando significativo (*p-valor* ≈ 0), es decir, existe correlación serial positiva en los errores (ver Apéndice B.7). De acuerdo con Granger y Newbold (1986), cuando $R^2 > d$ se tiene un buen indicador de que nos encontramos ante una regresión espuria. Sin embargo, esta aseveración no se puede tomar a la ligera, por ello estos autores exponen una discusión más amplia de la que se proporciona en el presente trabajo. Por otra parte, analizando el logaritmo de los datos se logra, de cierta manera, estabilizar un poco la varianza, como se muestra en la Figura 4.2, pero aún así, la regresión entre ellas resulta también significativa, con errores correlacionados.

Tabla 4.1: Regresión lineal: datos de pimienta blanca y negra.

Modelo de regresión: $Y_t = \beta X_t + \epsilon_t$				
	Estimación	Error est.	Estadístico t	*p-valor*
Coeficiente	$\hat{\beta} = 1.3660$	0.0116	117.7	<2e-16
	R^2=0.9809	d=0.2400 *p-valor* ≈ 0		

Sin embargo, al tomar las diferencias en los datos con logaritmo, se observa cómo la variabilidad disminuye considerablemente, como puede observarse en la Figura 4.3.

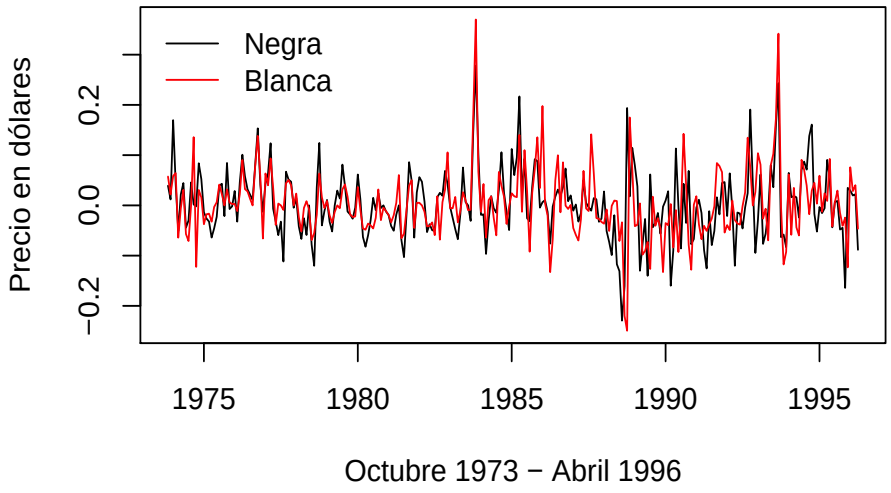


Figura 4.3: Diferencias del logaritmo de los precios por tonelada de pimienta en Estados Unidos

En este caso, el modelo de regresión lineal que se considera es el siguiente,

$$\Delta Y_t = \beta \Delta X_t + \epsilon_t,$$

en donde, $\Delta X_t = log(X_t) - log(X_{t-1})$ y $\Delta Y_t = log(Y_t) - log(Y_{t-1})$. Los resultados se muestran en la Tabla 4.2; se observa que el coeficiente de determinación es 0.3058, es decir, el modelo explica aproximadamente un 30.58 % de la variabilidad en los precios de pimienta blanca. Por otro lado, el estadístico de Durbin-Watson toma un *p-valor*=0.288, el cual resulta no significativo y se concluye que no existe correlación serial entre los errores. Sin embargo, si se observa el coeficiente asociado a la pendiente del modelo, se ve que es significativo (*p-valor*<2e-16), lo cual, a simple vista, pone en duda los resultados de la regresión; este ejemplo es peculiar, ya que a pesar de que se logra la estacionariedad en los datos como se verá más adelante, el análisis de regresión es confuso, la recomendación es aplicar las pruebas de raíz unitaria a los residuales de la regresión lineal, (Kleiber & Zeileis, 2008, p.167), sin duda es preciso realizar un estudio más detallado para tener resultados concluyentes.

Tabla 4.2: Regresión lineal de la diferencia del logaritmo de los datos de Pimienta

Modelo de regresión: $\Delta Y_t = \beta \Delta X_t + \epsilon_t$				
	Modelo de regresión	$\Delta Y_t = \beta \Delta X_t + \epsilon_t$		
	Estimación	Error est.	Estadístico t	*p-valor*
Coeficiente	$\hat{\beta} = 0.54920$	0.05045	10.89	<2e-16
	R^2= 0.3058	d=1.8585 *p-valor* = 0.288		

Como éstas, existen muchas situaciones que muestran fuerte correlación entre las variables involucradas cuando en realidad no existe tal; ello debido a que ambas pueden presentar la misma tendencia, se cruzan, o bien, posiblemente una variable oculta es la que está causando su correlación.

Otro caso que se puede estudiar es el de analizar el efecto que una serie independiente integradas de orden uno tiene sobre otra del mismo tipo. A continuación se presentan dos casos; en el primero se analizan dos series de tiempo independientes integradas de primer orden y después se les aplica una diferencia a los datos, y se compara el efecto que tiene esta transformación en los resultados.

1. **Primer caso: usando datos simulados**

   Primero considérese las siguientes observaciones que se generaron a través del software R, utilizando las semillas `set.seed(15)` y `set.seed(70)` para generar valores de X e Y, utilizando la función `rnorm()`, con la cual se generaron 100 números aleatorios provenientes de una distribución normal. Después de ello, se utilizó la función `cumsum()` para formar series de caminata aleatoria dadas por:

   $$Y_t = Y_{t-1} + Z_t \qquad t = 1, 2, \ldots, 100, \tag{4.1.4}$$
   $$X_t = X_{t-1} + Z'_t \qquad t = 1, 2, \ldots, 100, \tag{4.1.5}$$

   donde Z_t y Z'_t son observaciones de ruido blanco.

   Nótese que ambos modelos son procesos no-estacionarios. Para este caso se considera el modelo de regresión lineal con intercepto: $Y_t = \alpha + \beta X_t + \epsilon_t$ y se utiliza la misma prueba de hipótesis que el ejemplo anterior,

   $$H_0 : \beta = 0,$$
   $$H_1 : \beta \neq 0.$$

   Como Z_t y Z'_t son independientes entre sí, esperaríamos que el coeficiente de determinación R^2 tome un valor cercano a cero, es decir, que no se observe correlación entre las series. Sin embargo, en la Tabla 4.3 se observa que R^2 es 0.5386, y que la pendiente resulta significativa (p-*valor*<2e-16), con lo cual erróneamente se concluiría que existe una relación entre ellas, cuando realmente no la hay, porque así se construyeron dichas variables. Por otra parte, el estadístico de Durbin-Watson toma el valor de 0.27949 y resulta significativo (p-*valor*$\approx$ 0), ésto indica que hay correlación serial en los residuos. En este caso se puede observar la característica de que $R^2 > d$, lo cual según Granger y Newbold sugiere que nos encontramos ante una regresión espuria.

   En la Figura 4.4 se presentan las gráficas de cada una de estas caminatas aleatorias, asi como el gráfico de dispersión y la regresión lineal de Y_t sobre X_t. Obsérvese los cambios abruptos en cada serie de tiempo y la tendencia que presentan cada una, a simple vista podría concluirse que efectivamente no se presenta estacionariedad en éstas.

Tabla 4.3: Regresión lineal de Y_t en X_t

Modelo de regresión: $Y_t = \alpha + \beta X_t + \epsilon_t$				
	Modelo de regresión	$Y_t = \alpha + \beta X_t + \epsilon_t$		
	Estimación	Error est.	Estadístico t	*p-valor*
Intercepto	$\hat{\alpha}$ = -0.27491	0.39113	-0.703	0.484
Coeficiente	$\hat{\beta}$ = -0.42469	0.03971	-10.696	<2e-16
	R^2=0.5386	d=0.27949 *p-valor*≈0		

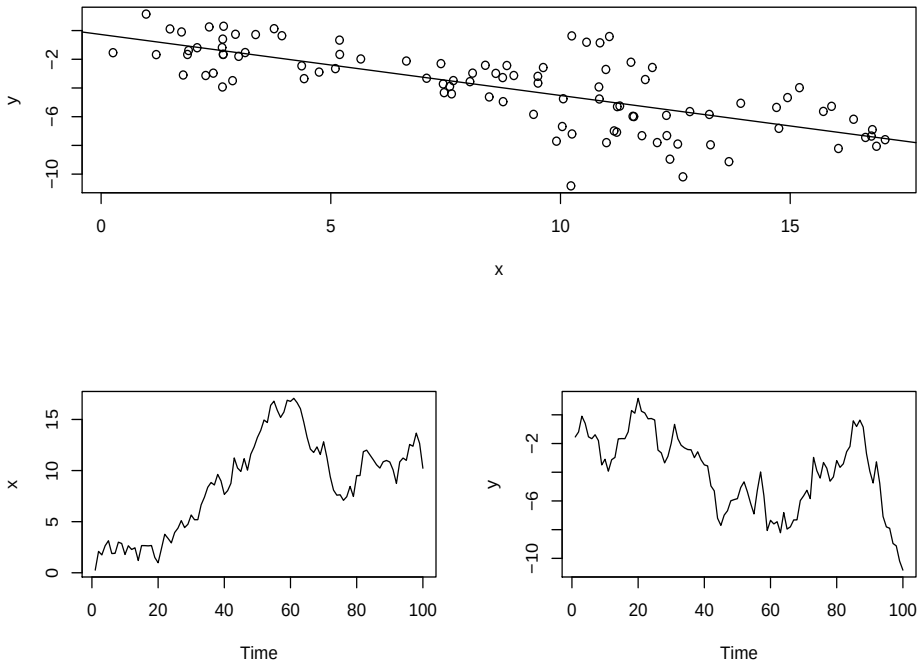


Figura 4.4: Regresión lineal de Y_t sobre X_t, datos sin diferenciar.

2. **Segundo caso: usando datos con las primeras diferencias**

   Para el segundo caso, se toma el modelo de regresión siguiente:

$$\Delta Y_t = \alpha + \beta \Delta X_t + \epsilon_t,$$

   donde

$$\Delta X_t = X_t - X_{t-1},$$
$$\Delta Y_t = Y_t - Y_{t-1},$$

y se prueba la hipótesis sobre el coeficiente de ΔX_t,

$$H_0 : \beta = 0,$$
$$H_1 : \beta \neq 0.$$

Cuando se realiza una regresión sobre las primeras diferencias, mostrada en la Tabla 4.4, se nota que el valor de R^2 resulta aproximadamente 0.02769, y el valor del estadístico de Durbin-Watson es de 1.9609, muy cercano a 2, la prueba resulta no significativa y podría decirse que la correlación desaparece. En cuanto $p-valor$ asociado al coeficiente de la pendiente de la regresión, éste toma el valor de 0.0997, por lo cual resulta no significativo. En la Figura 4.5

Tabla 4.4: Regresión lineal con datos diferenciados

Modelo de regresión: $\Delta Y_t = \alpha + \beta \Delta X_t + \epsilon_t$				
	Modelo de regresión	$\Delta Y_t = \alpha + \beta \Delta X_t + \epsilon_t$		
	Estimación	Error est.	Estad. t	*p-valor*
Intercepto	$\hat{\alpha}$ = 1.4231	0.0957	14.87	< 2e-16
Coeficiente	$\hat{\beta}$ = -0.1581	0.0951	-1.662	0.0997
	R^2=0.02769	d= 1.9609 *p-valor*= 0.854		

se muestra la gráfica del mismo modelo pero con las primeras diferencias, y la regresión asociada a ésta. Al parecer, estas series de tiempo no-estacionarios, han pasado a ser estacionarias.

Granger y Newbold (1986) afirman que "cuando una caminata aleatoria o un proceso integrado de promedios móviles está involucrado, las posibilidades de obtener una relación espuria usando los procedimientos de prueba convenciales son muy altos, aumentando con el número de variables independientes incluídas en la regresión".

El problema de regresión espuria es de gran interés en el estudio de series de tiempo económicas, pues la mayoría de ellas son no-estacionarias y si no se verifica este supuesto puede ocasionar consecuencias graves. Es por ello la importancia de contar con herramientas que permitan verificar si una serie es o no estacionaria; aspecto que se estudiará en las siguientes secciones. Sin embargo, no hay que olvidar el análisis descriptivo que debe acompañar a todo estudio es fundamental, pues permite muchas veces visualizar lo que después es corroborado con pruebas estadísticas.

4.2. Pruebas de estacionariedad

Cuando una serie de tiempo es estacionaria, su media, varianza y autocovarianza permanecen constantes sin importar el momento en el cual se miden, esto es, son invariantes con respecto al tiempo. Si para un conjunto de datos se observa que la media o la varianza cambian en el tiempo, seguramente se tendrá una serie de tiempo no-estacionaria; en éstas su comportamiento sólo puede ser analizado por períodos

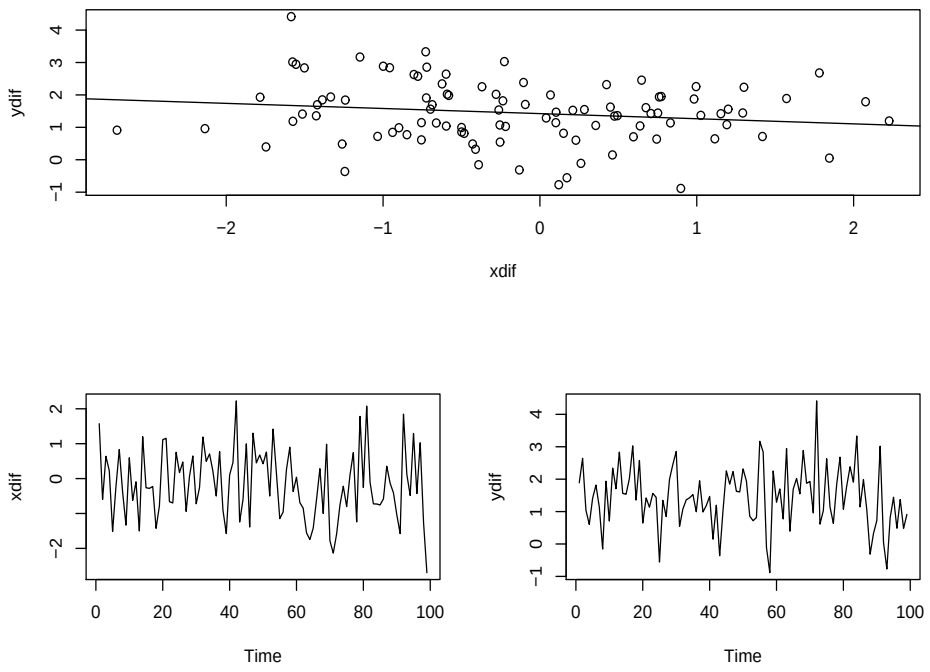


Figura 4.5: Regresión lineal de las diferencias de Y_t sobre las diferencias en X_t

y lo observado en un período no puede generalizarse a otros. El aplicar técnicas estadísticas que suponen estacionariedad cuando no la hay, resulta en un sesgo no deseable. Es necesario pues, contar con herramientas que permitan identificar cuándo una serie de tiempo es no-estacionaria.

En el ámbito económico, en series asociadas a precios de valores como son acciones o tasas de cambio, es usual observar un comportamiento similar al de un modelo de caminata aleatoria, esto es, son no-estacionarias y muchas veces puede distinguirse aquellas que son caminatas aleatorias con o sin un término constante o intersección.

Para verificar si una serie es estacionaria o no, es común utilizar las pruebas de raíz unitaria. Cuando la serie resulta no-estacionaria, en ocasiones es posible conseguir la estacionariedad a través del cálculo de diferencias. Por ejemplo, considérese el siguiente modelo:

$$Y_t = \rho Y_{t-1} + Z_t, \qquad t = 1, 2, \cdots, \cdots \tag{4.2.1}$$

donde $Z_t \sim N(0, \sigma^2)$. Si en la ecuación (4.2.1) $\rho = 1$, se tendrá un modelo de caminata aleatoria, que se sabe es no-estacionario. Sin embargo al tomar la primera diferencia se obtiene un modelo estacionario, puesto que Z_t es ruido blanco:

$$\begin{aligned} Y_t - Y_{t-1} &= Z_t, \\ \Delta Y_t &= Z_t, \end{aligned} \tag{4.2.2}$$

De acuerdo al orden del modelo serán las veces que sea necesario diferenciar; en este caso se diferenció una vez, por lo tanto, se le llama modelo integrado de orden 1, esto es, $I(1)$.

En general, en un modelo

$$Y_t = \rho Y_{t-1} + Z_t, \tag{4.2.3}$$

el cual puede escribirse como:

$$\begin{aligned} Y_t - Y_{t-1} &= (\rho - 1)Y_{t-1} + Z_t, & (4.2.4)\\ \Delta Y_t &= \delta Y_{t-1} + Z_t, & (4.2.5) \end{aligned}$$

con $\rho - 1 = \delta$, se observa que cuando $\delta < 0$, la serie original es estacionaria, en el caso que $\delta > 0$, equivalente a $\rho > 1$ corresponde a lo que se conoce como series explosivas y cuando $\rho = 1$, se tendrán series no-estacionarias. Éste es el enfoque que generalmente se maneja para las pruebas de raíces unitarias.

Existen varias pruebas de raíz unitaria; entre las más utilizadas está la prueba de Dickey-Fuller, sobre la cual Baumöhl & Lyócsa (2009) mencionan posee baja potencia (Apéndice A.3), por lo cual se han desarrollado mejores pruebas. Otras pruebas muy utilizadas pueden encontrarse en Phillips (1987), Zivot & Andrews (1992), Kwiatkowski *et al.* (1992), Elliott *et al.* (1996), etcétera. Tanto la potencia, el tamaño (Apéndice A.4) como los supuestos a cumplir, varían de una prueba a otra; inclusive hay cambios en el planteamiento de la hipótesis nula, pues cuando la mayoría de ellas establece en su hipótesis nula que la serie en estudio es no-estacionaria, la prueba de KPSS plantea que la serie es estacionaria frente a una hipótesis alternativa de que no lo es. Así pues, en la mayoría de las pruebas, el rechazo de la hipótesis nula llevaría a la conclusión que la serie es estacionaria, para un nivel de significancia dado.

En el presente trabajo solamente se estudian las pruebas más utilizadas en el ámbito económico, que son las pruebas de Dickey-Fuller y de Phillips-Perron. Estas pruebas difieren, principalmente, en la forma en que tratan la correlación serial en las pruebas de regresión.

4.2.1. Prueba de Dickey-Fuller, 1979

La prueba de raíz unitaria de Dickey-Fuller considera la hipótesis nula $\delta = 0$, donde el valor estimado del coeficiente Y_{t-1} en (4.2.2) sigue la distribución del estadístico τ o de "Dickey-Fuller"; sin embargo, cuando la hipótesis se rechaza, es decir, si es posible concluir que la serie es estacionaria, se podría utilizar el estadístico t usual, que sigue una distribución t de *Student*.

Lo que Dickey y Fuller hicieron para determinar los valores de la tabla de valores críticos fue básicamente generar cientos de sucesiones provenientes de modelos de caminatas aleatorias, en donde a cada modelo se le estimó el valor del coeficiente que acompaña al término independiente; la mayoría de estos estimadores será cercano a la unidad, o bien mayores a éste. Al realizar dicho experimento encontraron, por ejemplo, que al tomar en consideración un modelo con intersección, ocurría lo siguiente (Enders, 2015):

- 90 % de los valores estimados del coeficiente son menores que 2.58 errores estándar de la unidad.
- 95 % de los valores estimados del coeficiente son menores que 2.89 errores estándar de la unidad.
- 99 % de los valores estimados del coeficiente son menores que 3.51 errores estándar de la unidad.

Ahora bien, para evaluar la presencia de raíz unitaria Dickey y Fuller consideraron la posibilidad de los siguientes modelos de regresión en su prueba:

1. Y_t Caminata aleatoria
$$\Delta Y_t = \delta Y_{t-1} + Z_t,$$
2. Y_t Caminata aleatoria con intersección
$$\Delta Y_t = \alpha + \delta Y_{t-1} + Z_t,$$
3. Y_t Caminata aleatoria con intersección y tendencia
$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + Z_t.$$

En las ecuaciones anteriores α y β son elementos deterministas, t es el tiempo o la variable de tendencia y $\delta = \rho - 1$. En cada uno de los casos, se prueba la hipótesis nula de no-estacionariedad, frente a la alternativa de que la serie es estacionaria,

$$H_0 : \delta = 0,$$
$$H_a : \delta < 0,$$

que es equivalentemente a:

$$H_0 : \rho = 1,$$
$$H_a : \rho < 1.$$

Denotemos por τ, τ_μ y τ_τ, los estadísticos de prueba de los modelos 1, 2, 3 respectivamente, mencionados anteriormente; éstos se obtienen al tomar el cociente del parámetro estimado $\hat{\rho}$ y el error estándar de dicho estadístico.

$$\tau = \frac{\hat{\rho} - 1}{MSE(\hat{\rho})}. \tag{4.2.6}$$

Donde el estimador de ρ, según Neusser (2015), se calcula como:

$$\hat{\rho} = \frac{\frac{1}{\sigma^2 n}\sum_{t=1}^{n} Y_{t-1} Z_t}{\frac{1}{\sigma^2 n^2}\sum_{t=1}^{n} Y_{t-1}^2}. \tag{4.2.7}$$

Y el error estándar del estimador, $MSE(\hat{\rho})$, de acuerdo a Box *et al.* (2008), es igual a:

$$MSE(\hat{\rho}) = S_a \left[\sum_{t=2}^{n} (Y_{t-1} - Y_t)^2\right]^{-1/2}, \tag{4.2.8}$$

donde S_a es igual a $\sqrt{\frac{\sum_{t=2}^{n} Y_t^2 - \hat{\rho}\sum_{t=2}^{n} Y_{t-1}Y_t}{n-2}}$.

Al efectuar una prueba de hipótesis de raíz unitaria, utilizando Dickey-Fuller, se compara el valor que toma el estadístico τ, con los valores críticos de Dickey-Fuller de la Tabla C.3, mostrada en el Apéndice C; cuando el valor del estadístico resulta menor que el de la tabla, para un nivel de significancia dado, entonces se rechaza la hipótesis nula. Por otro lado, si el valor de τ es mayor que el valor crítico de tablas, entonces no se rechaza la hipótesis nula y se dice que la serie es no-estacionaria.

Cuando se rechaza H_0 en el caso 1, se concluye que la serie de tiempo es estacionaria con media cero. Si se rechaza H_0 en el caso 2, entonces se dice que la serie es estacionaria con media distinta de cero. Para el caso 3, el rechazo de H_0 implica que hay evidencia de que la serie es estacionaria alrededor de una tendencia determinista.

A continuación se retoman los datos de los precios mensuales promedio de pimienta negra y blanca, que se presentaron al inicio de este capítulo, y se efectúa una prueba de estacionariedad de Dickey-Fuller, considerando modelos sin intersección, con intersección y también con intersección y tendencia. Para esta prueba se utilizó la librería URCA del software R, la cual modela

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta Y_{t-j} + Z_t. \tag{4.2.9}$$

Ya que básicamente interesa probar la hipótesis $H_0 : \delta = 0$, a continuación se presentan las estimaciones de δ y el error estándar de su estimador. Es necesario

Modelo	Estimación	Error Estándar
$\Delta Y_t = \delta Y_{t-1} + Z_t$	0.0002246	0.0005016
$\Delta Y_t = \alpha + \delta Y_{t-1} + Z_t$	-0.01540	0.008629
$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + Z_t$	-0.01591	0.00892

escoger uno de los tres modelos anteriores. Sin embargo, antes de eso, es necesario analizar el valor de δ, por ejemplo, en el primero caso, $\delta > 0$, por lo tanto $\rho > 1$, es decir, se tiene una serie explosiva, por lo tanto este caso se descarta, y sólo se analizan los modelos restantes. Y así, utilizando la información vertida en la tabla anterior tenemos que:

$$\tau_\mu = \frac{-0.01540}{0.008629} = -1.785, \qquad \textit{p-valor} = 0.0753,$$

$$\tau_\tau = \frac{-0.01591}{0.008926} = -1.782, \qquad \textit{p-valor} = 0.0759.$$

Como el valor de τ_μ y τ_τ son mayores que los valores críticos de la tabla de Dickey-Fuller (Tabla C.3), entonces, no se rechaza la hipótesis nula. Esta conclusión se corrobora al observar el *p-valor* asociado a los valores que tomaron estos estadísticos,

en todos los casos no resultó significativo, al nivel significancia de 0.05, por lo que los datos no proporcionan evidencia para rechazar H_0 y entonces concluimos que no existe estacionariedad. Es importante observar que para cada estadístico se tiene una tabla de valores críticos, que considera además el tamaño de la muestra.

Ahora bien, si se diferencian los datos de estas series, se esperaría resulten estacionarias.

Para este caso, las estimaciones de δ y el error estándar de su estimador son:

Modelo	Estimación	Error Estándar
Sin intersección ni pendiente	0.7165	0.0586
Con intersección	0.7174	0.0587
Con intersección y pendiente	0.7176	0.0588

Utilizando la información de la tabla anterior, se tiene que:

$$\tau_\mu = \frac{-0.7174}{0.0587} = -12.2214, \qquad p\text{-}valor < 2e-16,$$

$$\tau_\tau = \frac{-0.7176}{0.0588} = -12.2040, \qquad p\text{-}valor < 2e-16.$$

De igual manera, se compara el valor de cada estadístico con los de la tabla de Dickey-Fuller; en este caso se encuentra que los estadísticos calculados son menores que los de la tabla, por lo tanto, en todos los casos se rechaza la hipótesis nula de no-estacionariedad y se concluye que las series son estacionarias. Este hecho se reafirma al observar el *p-valor* asociado a los valores de cada estadístico; en los tres casos el *p-valor* es menor al nivel de signficancia de 0.05.

La prueba de raíz unitaria de Dickey-Fuller, que hemos utilizado, es adecuada si la serie de tiempo Y_t puede caracterizarse por un modelo autorregresivo de orden uno, donde el error se considera ruido blanco. Es claro que muchas series de tiempo económicas tienen un comportamiento más complicado, por lo cual Said & Dickey (1984) generalizaron esta prueba para que pueda utilizarse en modelos generales $ARMA(p, q)$; a esta prueba se le conoce como **prueba de Dickey-Fuller Aumentada (ADF)**. Las tablas de valores críticos para los estadísticos permanecen invariantes, ya que la distribución asintótica del estadístico que se utiliza es la misma. Los modelos de regresión que se plantean en la prueba ADF son los siguientes:

- sin intersección y sin tendencia

$$\Delta Y_t = \delta Y_{t-1} + \lambda_i \sum_{i=1}^{m} \Delta Y_{t-i} + Z_t \tag{4.2.10}$$

- con intersección y sin tendencia

$$\Delta Y_t = \alpha + \delta Y_{t-1} + \lambda_i \sum_{i=1}^{m} \Delta Y_{t-i} + Z_t \tag{4.2.11}$$

- con intersección y tendencia

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \lambda_i \sum_{i=1}^{m} \Delta Y_{t-i} + Z_t \tag{4.2.12}$$

Básicamente, lo que se hace es aumentar los términos de rezago en la variable dependiente ΔY_{t-i}. La idea es incluir los términos suficientes para que el término de error en cualquiera de los tres modelos anteriores no esté serialmente correlacionado. Un criterio para ello, que se utiliza mucho en el software de series de tiempo, es el *criterio de Akaike* (Box *et al.*, 2008, pp. 59).

4.2.2. Prueba de Phillips-Perron, 1988

La prueba de Phillips-Perron (PP), es una alternativa a la de Dickey-Fuller, en la cual se utiliza el siguiente modelo:

$$Y_t = \alpha + \rho Y_{t-1} + \epsilon_t. \tag{4.2.13}$$

En esto caso, ϵ_t no necesariamente es un proceso de ruido blanco; pudiera ser cualquier proceso estacionario de media cero (Neusser, 2015), que cumple con las propiedades expuestas en el artículo de Phillips (1987).

De acuerdo con G. S. Maddala (1998), los estadísticos para estimar el parámetro ρ de la regresión (4.2.13), son los siguientes:

$$Z_\rho = T(\hat{\rho} - 1) - \frac{1}{2} \frac{(s^2 - s_e^2)}{T^{-2} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2}, \tag{4.2.14}$$

$$Z_t = \frac{s_e}{s} t_{\hat{\rho}} - \frac{1}{2} \frac{(s^2 - s_e^2)}{s[T^{-2} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2]^{(1/2)}}. \tag{4.2.15}$$

En donde $\bar{y}_{-1} = \sum_{t=1}^{T} y_t/(T-1)$. Luego, un estimador consistente de σ_e^2 está dado por $s_e^2 = T^{-1} \sum_{t=1}^{T} e_t^2$. Además, el estimador consistente de σ^2 es, s_{Tl}^2 el cual, por simplicidad, se denota por s^2. Véase Maddala, 1998, p. 79.

En las pruebas de PP, Z_ρ es la transformación del estimador $T(\hat{\rho} - 1)$ y Z_t es la transformación del estadístico t. La ventaja principal de esta prueba sobre las de Dickey y Fuller, es que es una prueba no paramétrica, lo cual amplía las situaciones en que puede aplicarse. Se ha demostrado que la prueba de PP es más potente que la de DF, es decir, rechaza la hipótesis nula más a menudo cuando ésta es falsa (Neusser (2015)).

Los estadísticos de prueba de PP tienen la misma distribución que los utilizados en la ADF, es decir, se pueden utilizar las mismas tablas para los valores críticos, con la única diferencia que ahora se consideran dos casos:

1. Y_t Caminata aleatoria con intersección

$$Y_t = \alpha + \rho Y_{t-1} + \epsilon_t, \tag{4.2.16}$$

2. Y_t Caminata aleatoria con intersección y tendencia

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \epsilon_t. \tag{4.2.17}$$

Considerando nuevamente los datos de los precios mensuales promedio de pimienta negra y blanca, a continuación se aplica la prueba de estacionariedad de Phillips-Perron, obteniendo lo siguiente:

$$\begin{aligned} Y_t &= 48.3093 + 0.9860 Y_{t-1} + \epsilon_t, \\ &\quad (30.9494)(0.0096) \end{aligned} \tag{4.2.18}$$

$$\begin{aligned} Y_t &= 48.6838 + 0.0094t + 0.9859 Y_{t-1} + \epsilon_t. \\ &\quad (31.8562)(0.1835)(0.0099) \end{aligned} \tag{4.2.19}$$

Nótese que debajo de cada coeficiente está su respectivo error estándar, además en estos modelos, el coeficiente que acompaña a Y_{t-1} es $\hat{\rho}$, por lo tanto, será necesario restar 1 a dicho valor, de tal forma obtendremos el valor de $\hat{\delta}$, y así se tiene lo siguiente:

$$\tau_\mu = \frac{0.9860 - 1}{0.0096} = -1.4583, \quad p\text{-}valor = 0.149,$$

$$\tau_\tau = \frac{0.9859 - 1}{0.0099} = -1.4242, \quad p\text{-}valor = 0.160.$$

Como el valor de τ_μ es mayor que el de la tabla de Dickey-Fuller (Tabla C.3), entonces no se rechaza la hipótesis nula, lo cual se corrobora al observar el *p-valor*, ya que es mayor al que se había fijado desde un principio, 0.05; se obtiene una conclusión análoga con el valor de τ_τ. A continuación se muestran los modelos lineales con los datos ya diferenciados.

$$\begin{aligned} Y_t &= 5.4287 + 0.2825 Y_{t-1} + \epsilon_t, \\ &\quad (13.3446)(0.0587) \end{aligned} \tag{4.2.20}$$

$$\begin{aligned} Y_t &= 5.4501 - 0.0393t - 0.2823 Y_{t-1} + \epsilon_t. \\ &\quad (13.3686)(0.1720)(0.0588) \end{aligned} \tag{4.2.21}$$

De igual manera, al valor de $\hat{\rho}$ se le resta 1, luego se calcula el estadístico de τ:

$$\tau_\mu = \frac{0.2825 - 1}{0.0587} = -12.2231, \quad p\text{-}valor < 2e - 16,$$

$$\tau_\tau = \frac{-0.2823 - 1}{0.0588} = -12.2057. \quad p\text{-}valor < 2e - 16.$$

Una vez que se han diferenciado los datos, se obtiene que los estadísticos τ_μ y τ_τ, son menores que los correspondientes valores críticos de la tabla de Dickey-Fuller, por lo tanto, se concluye que existe estacionariedad, aspecto que también puede concluirse observando el *p-valor* asociado a los valores que tomaron estos estadísticos, los cuales son menores que el nivel de significancia establecido.

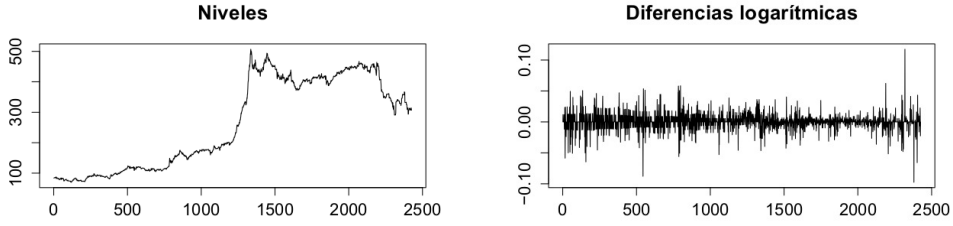
4.3. Ejemplos

A continuación se presenta un par de ejemplos. El primero de ellos tiene por objetivo mostrar los resultados espurios que pueden surgir cuando se trabaja con series no-estacionarias y la necesidad de transformaciones adecuadas para lograr la estacionariedad. La intención del segundo ejemplo es la de establecer la importancia de verificar los supuestos, antes de utilizar cualquier prueba de estacionariedad. En ambos ejemplos se utilizan tanto la prueba de Dickey-Fuller como la de Phillips-Perron.

4.3.1. Índices bursátiles: SAX y PX

En el ejemplo siguiente se utiliza el índice SAX Eslovaquia, índice bursátil que rastrea el desempeño de las grandes empresas que cotizan en la bolsa de Bratislava, y el índice PX de la República Checa, el más importante de Praga. Los datos son los precios diarios al cierre del 1 de septiembre de 1999 al 1 de septiembre del 2009, presentados por Baumöhl & Lyócsa (2009). Cuando se analiza la relación entre el cierre de los índices bursátiles de estas bolsas de valores, se obtiene que todos los parámetros de la regresión resultan significativos, en el Cuadro 4.7 se observa un valor alto para el coeficiente de determinación (0.8297) y un valor pequeño para el estadístico de Durbin-Watson, con (0.0082) si la variable dependiente es PX o (0.0077) si es de SAX. Si el análisis terminara ahí, podríamos concluir erróneamente que hay una relación causa-efecto entre estas series, que en verdad se están correlacionando por ser no-estacionarias, como se muestra a continuación.

En la Figura 4.6 (a) y 4.7 (a) se presentan la series de tiempo de cada índice bursátil; a juzgar por estas gráficas se podría decir que las series son no-estacionarias. En las Figuras 4.6 (b) y 4.7 (b) se muestra las diferencias de los logaritmos. Si se comprueba que ambas series son no-estacionarias, entonces los resultados de la regresión serán engañosos, ya que la relación espuria es causada por la tendencia que presentan ambos índices, no por la posible relación entre los precios al cierre.

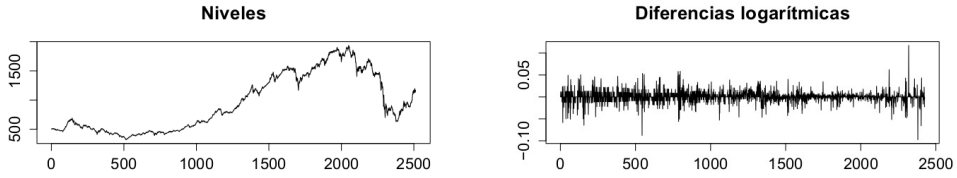


(a) Datos a niveles

(b) Diferencias logarítmicas

Figura 4.6: Índice de valores de Eslovaquia en niveles y en diferencias logarítmicas

Después de aplicar las pruebas de Dickey-Fuller y Phillips-Perron a cada conjunto de datos, se obtiene lo mostrado en las Tablas 4.5 y 4.6, en las que se observan los



(a) Datos a niveles

(b) Diferencias logarítmicas

Figura 4.7: Índice de valores de República Checa en niveles y en diferencias logarítmicas

resultados de τ_μ, τ_τ y los *p-valores* que surgen de aplicar estas pruebas de estacionariedad a los datos originales y a los datos diferenciados, tanto para SAX como para PX; en estas tablas "i" y "t" denotan intersección y tendencia, respectivamente.

Recordemos que en la práctica, un *p-valor* pequeño implica un rechazo a la hipótesis nula, concluyendo que la serie de tiempo es estacionaria. Si se establece un nivel de significancia $\alpha = 0.05$ vemos que el *p-valor* de los diferentes análisis presentados en la Tabla 4.5 resulta mayor que α, por lo que se concluye que las series son no-estacionarias, esto es, presentan una raíz unitaria y no hay evidencia para rechazar la hipótesis nula en ambos casos.

Por otra parte, en la Tabla 4.6, que corresponde a los resultados de las pruebas de estacionariedad de las diferencias del logaritmo de los datos, todos los *p-valores* son menores que 0.05, por lo tanto se rechaza la hipótesis nula y se concluye que las series son estacionarias.

Tabla 4.5: Pruebas de estacionariedad con datos originales

	DF		PP	
Índice	i	it	i	it
SAX	-1.1738	0.0854	-0.3185	-0.1245
p-valor	0.2186	0.3091	0.99	0.99
PX	-1.0534	-0.2271	-0.3364	-0.1287
p-valor	0.0294	0.0626	0.99	0.99

Tabla 4.6: Pruebas de estacionariedad con datos diferenciados

	DF		PP	
Índice	i	it	i	it
SAX	-34.3620	-34.4110	-2516.947	-50.055
p-valor	0.0000	0.0000	0.01	0.01
PX	-35.9380	-35.9510	-2189.144	-46.1824
p-valor	0.0000	0.0000	0.01	0.01

Los resultados mostrados en las Tablas 4.5 y 4.6 también se pueden analizar de la siguiente manera. Consideremos los valores que toman los estadísticos de prueba

y comparemos éstos con los valores críticos de la Tabla de Dickey-Fuller (Apéndice C); puede verse que en todos los casos el valor del estadístico es mayor que el valor crítico de la Tabla de DF, para un nivel de significancia de 5 %; concluyendo con ésto, como anteriormente se hizo, que la hipótesis no se rechaza y que las series son no-estacionarias. Luego, haciendo un análisis similar en la Tabla 4.6, puede observarse que los valores que asumen los estadísticos τ_μ y τ_τ son menores a los de la Tabla de Dickey-Fuller y en todos los casos se rechaza la hipótesis nula, esto es, no existe raíz unitaria como ya se había encontrado.

A continuación se presenta la regresión lineal que se efectuó con los datos de SAX y de PX, para los datos originales y los datos transformados; en éstas se consideró, alternadamente, cada índice bursátil como variable tanto dependiente como independiente. Los estadísticos que se tomaron en cuenta para el análisis son el estadístico t usual, el coeficiente de determinación R^2 y el estadístico de Durbin-Watson, d. Las sospechas de no-estacionariedad se verifican al calcular las diferencias logarítmicas. En la Tabla 4.8 se observa que R^2 pasa de ser mayor a 0.8, a tomar un valor cercano al 0; por otra parte d, que era casi cero en la Tabla 4.7 , resulta $d \approx 2$ en la Tabla 4.7 . Nos encontramos pues ante una regresión espuria, no importa que conjunto de datos se tome por variable dependiente. Por otra parte, es conveniente notar en la

Tabla 4.7: Resultados de la regresión con datos a niveles

		Variable dependiente	
		SAX	PX
SAX	t-test		108.69
	R^2		0.8297
	d		0.0082
PX	t-test	108.69	
	R^2	0.8297	
	d	0.0077	

Tabla 4.8: Resultados de la regesión con datos diferenciados logaritmicamente

		Variable dependiente	
		SAX	PX
SAX	t-test		0.026
	R^2		0.0000
	d		1.8761
PX	t-test	0.026	
	R^2	0.0000	
	d	2.0323	

Tabla 4.7 que $R^2 > d$, y d es cercano a cero, y como se dijo al inicio del capítulo, ésto es un buen indicador de la presencia de regresión espuria.

4.3.2. Inconsistencia de resultados

Se analizaron datos presentados en el sitio: `http://stats.stackexchange.com`, en donde uno de los participantes plantea la siguiente pregunta, que se transcribe íntegramente: *"Tengo dificultades en resolver pruebas de raíz unitaria usando las pruebas de R, adf.test y pp.test; el resultado del p-valor usando adf.test es 0.2677 (lo cual significa que existe raíz unitaria), y el resultado del $p-valor$ utilizando pp.test es 0.01696 (lo cual significa que no existe raíz unitaria). ¿Qué significa esto? Me pregunto si tengo que usar funciones diferentes para remover raíz unitaria o no"*. El participante en el blog proporcionó los datos que utilizó, mismos que se encuentran en la Tabla C.2 del Apéndice C.

Al trabajar los datos en referencia, se obtienen efectivamente los *p-valores* que se mencionan en el blog. Sin embargo, la problemática principal no radica en cuál de las dos pruebas de estacionariedad confiar, sino en verificar primero que se cumplan los supuestos que establecen algunas de las pruebas de estacionariedad. Particularmente la prueba de Dickey-Fuller supone que los errores no están correlacionados y la prueba de Phillips-Perron puede manejar situaciones donde los errores estén correlacionados e inclusive exista heterocedasticidad. Los datos en referencia no satisfacen los supuestos de la prueba de Dickey-Fuller y por ende, los resultados de ambas pruebas son muy diferentes.

4.4. Algunas observaciones

Es importante tomar en cuenta los supuestos que se consideran al hacer uso de las pruebas de raíz unitaria, pues esto dará luz sobre qué pruebas pueden ser las más adecuadas para el conjunto de datos bajo análisis.

Con respecto a la potencia de una prueba, que es la probabilidad de rechazar la hipótesis nula cuando ésta es falsa, ya se ha mencionado que las pruebas de raíz unitaria no son muy potentes, aspecto que se ha mostrado a través de simulaciones Monte Carlo (Enders, 2015). En general, las pruebas de raíz unitaria no tienen la potencia suficiente para distinguir entre modelos que tienen una raíz unitaria y aquellos cuyo coeficiente está muy cercano a considerarse una raíz unitaria. Para estudiar este aspecto más a fondo y comparar herramientas estadísticas que proporciona el software R, en el siguiente capítulo se presentan diversas simulaciones que se realizaron, bajo diferentes consideraciones, y en las que se utilizaron las pruebas de Dickey-Fuller o de Phillips-Perron, incluidas en una de las tantas librerías de R, con las que se puede trabajar series de tiempo.

Capítulo 5

Simulaciones

El propósito de este capítulo es efectuar una comparación entre las pruebas de estacionariedad más utilizadas en la práctica, las cuales son la prueba de Dickey-Fuller y la prueba de Phillips-Perron. Esto se puede realizar de diversas maneras; aquí sólo se analizará el efecto que tiene el valor de ρ y el tamaño de muestra en la decisión de rechazar o no la hipótesis nula que se plantea en estas pruebas de raíz unitaria. Esto es importante pues proporciona una idea de qué tan posible es llegar a conclusiones erróneas cuando se analiza de manera inadecuada una serie de tiempo.

La idea de efectuar estas simulaciones surge del trabajo de Baumöhl & Lyócsa (2009), titulado "Stationary of time series and the problem of spurious regression", donde se realizó la simulación siguiente: se generaron series de tiempo no-estacionarias e independientes entre sí, a partir de procesos de caminata aleatoria, tal como se presentó en el Capítulo 2, donde se mostró que estas series son no-estacionarias. Estos autores consideraron también procesos de caminata aleatoria con intersección y estudiaron la potencia de varias pruebas, entre las que se incluyen la prueba de Dickey-Fuller y la de Phillips-Perron.

5.1. Escenarios de Simulación

En el presente trabajo se realizan varias simulaciones en las que se utiliza tanto un proceso autorregresivo de primer orden con intersección como un autorregresivo de primer orden con intersección y tendencia:

- $AR(1)$ con intersección
$$Y_t = \alpha + \rho Y_{t-1} + Z_t,$$
- $AR(1)$ con intersección y tendencia,
$$Y_t = \alpha + \beta t + \rho Y_{t-1} + Z_t.$$

Los cuales se utilizaron para generar series de datos de tamaño 50, 100 y 200. Los términos Z_t de cada modelo son generados a partir de una distribución normal de media 0 y varianza σ^2, además se fija el valor inicial como 0, es decir $Y_0 = 0$; α es el coeficiente de intersección, y β el coeficiente de tendencia, que se fijaron en 0.5 y 0.1, respectivamente. Se consideraron diferentes valores para ρ: 0.8, 0.9, 0.95, 0.99, 1, 1.02 y 1.05; ello con el fin de analizar el efecto que tiene este coeficiente en la

prueba de estacionariedad. Por ejemplo, un valor de $\rho = 0.99$, es decir, un modelo teóricamente estacionario pudiera generar resultados que indiquen que se trata de un modelo no-estacionario, ya que es un valor muy cercano a 1, y podríamos concluir que tenemos un modelo de caminata aleatoria. Por otro lado, se probó con valores de ρ mayores a 1, pero muy cercanos a éste, con el fin de comparar los resultados de estas pruebas en series de tiempo explosivas.

Con base en las consideraciones anteriores, en cada una de las simulaciones efectuadas se generaron 10000 muestras, y a cada una de ellas se les aplicó las pruebas de estacionariedad utilizando la prueba de Dickey-Fuller (4.2.1) y la prueba de Phillips-Perron (4.2.2). Se analizó el efecto que tiene el tamaño de muestra y el valor de ρ en el rechazo o no de la hipótesis nula, contabilizando las veces en que el *p-valor* era menor que cierto nivel de significancia, α, el cual se fijó 0.05.

Recordemos que el planteamiento de la hipótesis nula es que la serie es no-estacionaria, frente a la alternativa de que sí lo es, es decir,

$$H_0 : \delta = 0,$$
$$H_1 : \delta < 0,$$

lo cual es equivalente a plantear:

$$H_0 : \rho = 1,$$
$$H_1 : \rho < 1.$$

En el caso de analizar datos que provienen de modelos donde se consideraron valores de $\rho > 1$, las hipótesis se plantean como:

$$H_0 : \rho = 1,$$
$$H_1 : \rho > 1,$$

esto es, la hipótesis nula establece que la serie es no-estacionaria frente a la alternativa de que es explosiva. En el software de R la función que se utiliza para probar la estacionariedad de la serie debe especificar la hipótesis alternativa como explosiva, de no ser así, la hipótesis nula se rechazaría y se concluiría que la serie es estacionaria, cuando en realidad no lo es. El definir adecuadamente la hipótesis alternativa en la prueba de DF y PP es un aspecto importante pues de lo contrario se puede llegar a una conclusión errónea, como puede observarse en el artículo de Dickey & Fuller (1981), Tabla VII, donde se analizaron datos generados con valores de $\rho = 1.02$ y $\rho = 1.05$, para los cuales se rechaza la hipótesis nula de no-estacionariedad y se concluye la estacionariedad de la serie. Una corrección a este artículo la presenta Chandra (1981), donde prueba la invalidez de ese procedimiento y corrige la prueba de Dickey-Fuller, de esta manera es posible probar la no-estacionariedad de modelos explosivos.

5.2. Simulaciones

A continuación se presenta una serie de tablas donde se contabiliza el número de veces que se rechaza la hipótesis nula de no-estacionariedad, utilizando la prueba de

Dickey-Fuller y la prueba de Phillips-Perron; todo esto bajo el criterio de comparar si el *p-valor* que se genera es menor que el nivel de significancia de la prueba, el cual se estableció en $\alpha = 0.05$.

Para las simulaciones se tomaron en cuenta procesos estacionarios $AR(1)$ con intersección, cuyos resultados se presentan a partir de la Tabla 5.1 hasta la Tabla 5.4. También se consideró el modelo $AR(1)$ con intersección y tendencia; sus resultados pueden observarse de la Tabla 5.8 a la 5.11. Por otra parte, se modelaron también procesos de caminata aleatoria con intersección, y con intersección y tendencia; los resultados aparecen de la Tabla 5.5 a la Tabla 5.12. El análisis obtenido a partir de los procesos explosivos con intersección fueron registrados en las Tablas 5.6 y 5.7, y aquellos explosivos con intersección y tendencia, pueden verse en las Tablas 5.13 y 5.14.

5.2.1. Modelos autorregresivos de orden uno con intersección

Al comparar las Tablas 5.1 a 5.4, se puede observar que el tamaño de muestra influye considerablemente en los resultados. Por ejemplo, en la Tabla 5.1, donde se simularon series considerando $\rho = 0.8$, se observa que cuando se aumenta la muestra a $n = 200$ se tiene mayor información que cuando $n = 50$, aspecto que era de esperarse pero que es importante analizar su comportamiento a través de los diferentes valores de ρ. Usando la prueba de DF, en 1126 de las 10000 muestras de tamaño 50 el *p-valor* fue menor que 0.05, es decir el 11.26 % de las veces se rechazó la hipótesis nula de no-estacionariedad, por lo cual se concluye que ese porcentaje de series provienen de un modelo estacionario, como en realidad fueron simuladas. Es con muestras de tamaño $n = 200$ donde DF detecta en un 83.98 % que las series son estacionarias. Por otro lado, la prueba de PP, para $n = 50$, muestra que el 21.47 % de las veces rechaza la hipótesis nula, y con $n = 200$, este porcentaje aumenta a un 99.91 %.

Tabla 5.1: Resultados

$\rho = 0.8$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.1126	0.3087	0.8398
Phillips-Perron	0.2147	0.7085	0.9991

En la Tabla 5.2 se tienen modelos AR(1) en donde el valor ρ es igual a 0.9. Nótese que con ambas pruebas el número de muestras que presentan un *p-valor* inferior a 0.05 es menor en los tres casos respecto a la Tabla 5.1, esto es debido a que el valor de ρ se aproxima a 1, es decir, al de un proceso de caminata aleatoria con intersección, el cual es no-estacionario. A comparación de la prueba de DF, la prueba

de PP detecta más veces la estacionariedad sin importar el tamaño de muestra, ya sea $n = 50$, $n = 100$ o $n = 200$.

Tabla 5.2: Resultados

$\rho = 0.90$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0687	0.1349	0.4242
Phillips-Perron	0.0929	0.2418	0.6938

Los procesos que se simularon con $\rho = 0.95$ están registrados en la Tabla 5.3; al igual que en los casos anteriores se puede observar que la prueba de PP concluyó la estacionariedad más veces que la prueba de DF.

Tabla 5.3: Resultados

$\rho = 0.95$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0558	0.0728	0.1769
Phillips-Perron	0.0649	0.0997	0.2555

Por otra parte, en la Tabla 5.4, el valor de $\rho = 0.99$ es muy cercano a la unidad; en este caso se observa que el tamaño de muestra no es tan determinante en la contabilidad del porcentaje de veces que se detecta estacionariedad. De igual forma, la prueba de PP sigue detectando ligeramente más veces la estacionariedad de las series.

Tabla 5.4: Resultados

$\rho = 0.99$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0465	0.0464	0.0486
Phillips-Perron	0.0530	0.0563	0.0519

Cuando el valor de ρ es la unidad, se generan procesos de caminata aleatoria con

intersección y los resultados se registran en la Tabla 5.5. Nótese que los resultados son muy parecidos cuando se toma $\rho = 0.99$. Si el tamaño de muesta es $n = 50$ usando la pueba de DF, el 4.68 % de las veces resultaron p-valores menores a 0.05, y con la prueba de PP el porcentaje fue de 5.38 %, pero si $n = 200$ con DF se obtuvo un porcentaje de 4.9 % y con PP de 6.21 %.

Tabla 5.5: Resultados

$\rho = 1$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0468	0.0458	0.0490
Phillips-Perron	0.0538	0.0590	0.0621

Por otra parte, en la Tabla 5.6 y 5.7 se presentan los resultados a partir de los modelos explosivos, esto es, aquellos en donde el valor de $\rho > 1$. Recordemos que en este caso se mantiene la hipótesis nula de $\rho = 1$, es decir, de no-estacionariedad frente a la alternativa de $\rho > 1$.

Nótese que en ambas tablas, cuando $n = 200$, las dos pruebas de estacionariedad identifican en su totalidad que las series son explosivas; el resultado en cada Tabla no varía mucho, excepto cuando $\rho = 1.02$ con $n = 50$.

Tabla 5.6: Resultados

$\rho = 1.02$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.2791	0.9875	1
Phillips-Perron	0.2608	0.9905	1

Tabla 5.7: Resultados

$\rho = 1.05$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.9906	0.9998	1
Phillips-Perron	0.9913	0.9998	1

Para el análisis de estas series se utilizó el software libre *R*, versión 3.3.1. En el caso de las pruebas de estacionariedad, se utilizó el paquete de *tseries*, el cual tiene la función adecuada para probar series de tiempo explosivas.

5.2.2. Modelos autorregresivos de orden uno con intersección y tendencia

Para las siguientes pruebas de estacionariedad donde se utilizan modelos $AR(1)$ con intersección y tendencia, se fijó 0.5 para la intersección y 0.1 para la tendencia. En la Tabla 5.8, se muestran los resultados cuando se trabajó con $\rho = 0.8$; de igual manera se observa como el tamaño de muestra influye en los resultados, al tener $n = 50$, y usar la prueba de DF, el 10.85 % de las veces se rechazó la hipótesis nula, en cambio con $n = 200$ la prueba de DF detecta que el 83.59 % de las series provienen de un modelo estacionario. También se observa que la prueba de PP detecta más veces que DF que los modelos son estacionarios, independientemente del tamaño de muestra. Si ésto se compara con los resultados de la Tabla 5.1, se verá que son muy similares entre sí, sin embargo, cuando no hay tendencia, la prueba de PP detecta, ligeramente, más veces que las series son estacionarias.

Tabla 5.8: Resultados

$\rho = 0.8$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.1085	0.3022	0.8359
Phillips-Perron	0.2137	0.6995	0.9988

En la Tabla 5.9, se registran los resultados con $\rho = 0.90$ y puede observarse como la proporción de veces que se rechaza la hipótesis nula, disminuye considerablemente en ambas pruebas.

Tabla 5.9: Resultados

$\rho = 0.90$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0663	0.1313	4215
Phillips-Perron	0.0920	0.2378	6955

Los procesos estacionarios que se modelaron con $\rho = 0.95$ están registrados en la Tabla 5.10. En este caso, a diferencia de los presentados en la Tabla 5.3, el incluir

una tendencia parece influir positivamente en los resultados, ya que las pruebas de DF y PP detectan más veces que las series son estacionarias.

Tabla 5.10: Resultados

$\rho = 0.95$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0504	0.1453	0.5201
Phillips-Perron	0.0660	0.2339	0.7829

En el caso de la Tabla 5.11, también se observa cierta diferencia en los resultados respecto a la Tabla 5.4, ya que la tendencia hace que ambas pruebas detecten la estacionariedad más veces. Además se puede ver que DF y PP detectan casi en su totalidad que los modelos provienen de uno estacionario cuando el tamaño de muestra es $n = 200$.

Tabla 5.11: Resultados

$\rho = 0.99$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0106	0.0930	0.9934
Phillips-Perron	0.0204	0.2094	1

Por otra parte, en la Tabla 5.12 se tienen los resultados con $\rho = 1$, es decir, usando caminata aleatoria con intersección y tendencia. Particularmente en este caso, puede notarse que el porcentaje de veces que detecta la estacionariedad es menor que 1 % para todos los casos. Lo cual tiene sentido puesto que los modelos simulados provienen de uno no-estacionario.

Tabla 5.12: Resultados

$\rho = 1$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.0028	0.0019	0.0010
Phillips-Perron	0.0047	0.0023	0.0008

En las Tablas 5.13 y 5.14 se registran los resultados usando modelos explosivos, con $\rho = 1.02$ y $\rho = 1.05$. En la Tabla 5.13 se rechaza la hipótesis nula de estacionariedad el 99.81 % de las veces con $n = 50$; por otra parte, en la Tabla 5.14 puede verse que la hipótesis nula se rechaza el 100 % de las veces, para cualquier tamaño de muestra.

Tabla 5.13: Resultados

$\rho = 1.02$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	0.9981	1	1
Phillips-Perron	0.9990	1	1

Tabla 5.14: Resultados

$\rho = 1.05$			
	n		
	50	100	200
Prueba de estacionariedad			
Dickey-Fuller	1	1	1
Phillips-Perron	1	1	1

Es un hecho evidente que al aumentar el tamaño de muestra, se tiene mayor información, es por eso que en la mayoría de las tablas presentadas se obtienen mejores resultados cuando $n = 200$.

La importancia de las diferentes simulaciones que se han realizado y que se presentan en este capítulo, reside en proporcionar información práctica sobre el porcentaje de error en las pruebas de Dickey-Fuller y Phillips-Perron, bajo diferentes circunstancias como son diferentes valores de ρ y distintos tamaño de muestra. Es importante pues, saber que podemos tener series estacionarias, en las que las pruebas de raíces unitarias detectarán lo contrario, como es el caso de las simulaciones efectuadas con valores de ρ cercanos a la unidad y con tamaños de muestra pequeños.

5.3. Conclusiones

En esta tesis no se pretende presentar diferentes formas de modelar una serie de tiempo, tema para el cual existen diversos y muy buenos libros. Tampoco se pretende, en los estudios de simulación efectuados, abarcar la gran cantidad de pruebas de raíces unitarias que existen hoy en día. El objetivo principal es concientizar en el

hecho de que el análisis de una serie de tiempo conlleva diferentes etapas, requiere diferentes herramientas, que deben ser adecuadas al problema en estudio.

En todo análisis estadístico es de vital importancia el aspecto descriptivo y en el análisis de series de tiempo, las herramientas presentadas en el Capítulo 2 son fundamentales, pues a partir de una simple gráfica pueden detectarse problemas como la no-estacionariedad de una serie, problema detallado en el Capítulo 4 y tema central de esta tesis. La no-estacionariedad en las series de tiempo pareciera un aspecto sencillo, pero en realidad éstas abundan en diferentes áreas, como el ámbito económico en el cual se centró la mayoría de los ejemplos incluidos. La no-estacionariedad en una serie y sus repercusiones en resultados espurios, es una razón de peso que justifica el tema bajo análisis, pues en la profunda revisión bibliográfica que se llevó a cabo, se pudo constatar que a pesar de ser éste un problema ya detectado desde hace tiempo, siguen apareciendo publicaciones con resultados espurios.

Las simulaciones aquí presentadas pueden extenderse fácilmente a otras pruebas de raíces unitarias, para las cuales es recomendable tener en cuenta la manera en que se plantean las hipótesis, ya que como se mencionó anteriormente, hay pruebas como KPSS, en donde el planteamiento de la hipótesis nula es la estacionariedad.

Es importante notar que es una práctica común y recomendable, el utilizar varias pruebas de estacionariedad en el mismo conjunto de datos, con el fin de asegurar que se tiene una idea clara de los resultados. Esta práctica, aunada siempre a la experiencia del investigador, permitirá un mejor análisis de los datos.

Apéndice A

A.1. Operador de retraso

El operador lineal de retraso B ("backshift" en inglés) permite compactar las ecuaciones estocásticas que se utilizan en las series de tiempo, además permite analizar la estructura interna de los procesos ARMA. Este operador retrasa una posición en el tiempo el subíndice t,

$$BX_t = X_{t-1}. \tag{A.1.1}$$

Las propiedades del operador B son las siguientes:

1. Si $X_t = c$, donde $c \in \mathbb{R}$

$$Bc = c.$$

2. Si B se multiplica n-veces

$$B \underbrace{\cdots}_{n\text{-veces}} BX_t = B^n X_t = X_{t-n.}$$

3. El inverso es un operador que adelanta una posición en el tiempo

$$B^{-1}X_t = X_{t+1}.$$

4. Como $B^{-1}BX_t = X_t$, tenemos que

$$B^0 = 1.$$

5. Sean $m, n \in \mathbb{Z}$, tenemos que

$$B^m B^n X_t = B^{m+n} X_t = X_{t-m-n}.$$

6. Sean $a, b \in \mathbb{R}$, $m, n \in \mathbb{N}$ y X_t, Y_t procesos estocásticos

$$(aB^m + bB^n)(X_t + Y_t) = aX_{t-m} + bX_{t-n} + aY_{t-m} + bY_{t-n}.$$

A continuación se muestran tres ejemplos de procesos autorregresivos que ilustran el procedimiento para determinar si el proceso es o no estacionario. Estos ejemplos fueron tomados de Cowpertwait & Metcalfe (2009).

1. El modelo AR(1), $x_t = \frac{1}{2}x_{t-1} + z_t$ es estacionario, ya que la raíz de $1 - \frac{1}{2}B = 0$ es $B = 2$, el cual es mayor que 1.

2. El modelo AR(2), $x_t = -\frac{1}{4}x_{t-2} + z_t$ es estacionario, ya que las raíces de $1 + \frac{1}{4}B^2 = 0$ son $B = \pm 2i$, los cuales son números complejos con $i = \sqrt{-1}$, cada uno con un valor absoluto igual a dos.

3. El modelo $x_t = \frac{1}{2}x_{t-1} + \frac{1}{2}x_{t-2} + z_t$ es no-estacionario ya que una de sus raíces es la unidad. Para probar esto, primeramente se expresa el modelo en términos del operador de retraso, $-\frac{1}{2}(B^2 + B - 2)x_t = z_t$; i.e., $-\frac{1}{2}(B-1)(B+2)x_t = z_t$. El polinomio $\Phi(B) = -\frac{1}{2}(B-1)(B+2)$ tiene unidades $B = 1, -2$. Ya que se tiene una raíz unitaria $(B = 1)$, el modelo es no-estacionario. Nótese que la otra raíz $(B = -2)$ excede a la unidad en valor absoluto. Es suficiente que una de las raíces sea unitaria para que el proceso sea no-estacionario.

A.2. Algunas propiedades de los estimadores

- Decimos que $\widehat{\theta}$ es un estimador **insesgado** de θ, si $E(\widehat{\theta}) = \theta$.

  Si $E(\widehat{\theta}) \neq \theta$, entonces se dice que $\widehat{\theta}$ es **sesgado**; el sesgo de un estimador puntual $\widehat{\theta}$ está dado por $B(\widehat{\theta}) = E(\widehat{\theta}) - \theta$. En el caso que el $lim_{n\to\infty}B(\widehat{\theta}) = 0$ entonces el estimador es **asintóticamente insesgado**.

- Dados dos estimadores insesgados $\widehat{\theta}_1$ y $\widehat{\theta}_2$ de un parámetro θ, con varianzas $V(\widehat{\theta}_1)$ y $V(\widehat{\theta}_2)$, respectivamente, entonces la **eficiencia** de $\widehat{\theta}_1$ con respecto a $\widehat{\theta}_2$, denotada $eff(\widehat{\theta}_1, \widehat{\theta}_2)$ se define como la razón:

$$eff(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{V(\widehat{\theta}_1)}{V(\widehat{\theta}_2)}.$$

  Si $eff(\theta_1, \theta_2) > 1$ se tiene que $\widehat{\theta}_1$ es un estimador más eficiente que $\widehat{\theta}_2$

- Se dice que el estimador $\widehat{\theta}_n$ es un estimador **consistente** de θ si, para cualquier número positivo ϵ:

$$lim_{n\to\infty}P(|\widehat{\theta}_n - \theta| \leq \epsilon) = 1,$$

  o bien, de forma equivalente:

$$lim_{n\to\infty}P(|\widehat{\theta}_n - \theta| > \epsilon) = 0.$$

  Usualmente los MLE son consistentes e insesgados.

- El error cuadrático medio de un estimador puntual $\widehat{\theta}$ de un parámetro θ es:

$$MSE(\widehat{\theta}) = E[(\widehat{\theta} - \theta)^2].$$

  Si $B(\widehat{\theta})$ representa el sesgo del estimador $\widehat{\theta}$, se puede demostrar que:

$$MSE(\widehat{\theta}) = V(\widehat{\theta}) + [B(\widehat{\theta})]^2.$$

A.3. Función Potencia

Sea Γ una prueba para una hipótesis nula H_0. *La función potencia* de la prueba Γ, denotada por $\pi_\Gamma(\theta)$, se define como la probabilidad de que H_0 sea rechazada cuando la distribución de la que se obtuvo la muestra fue parametrizada por θ (Mood, 1974, pp. 406).

A.4. Tamaño de una prueba

Sea Γ una prueba para una hipótesis $H_0 : \theta \in \Theta_0$, donde $\Theta_0 \in \Theta$; esto es Θ_0 es un subconjunto del espacio parametral Θ. El tamaño de la prueba Γ de H_0 se define como $\sup_{\theta\in\Theta_0}[\pi_\Gamma(\theta)]$ (Mood, 1974, pp. 407).

A.5. Covarianza

La covarianza entre dos variables aleatorias X e Y, se define como el valor esperado del producto de las desviaciones de éstas de de sus respectivas medias, μ_x y μ_y, esto es:

$$Cov(X, Y) = E(X - \mu_x)(Y - \mu_y)$$

y el correspondiente estadístico muestral es:

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

donde (x_i, y_i), $i = 1, 2, \ldots, n$ son los valores muestrales de las dos variables aleatorias X e Y, y sus respectivas medias muestrales se denotan por $\bar{x}$ y $\bar{y}$ respectivamente.

Ahora, si $Y_1, Y_2, ..., Y_n$ y $X_1, X_2, ..., X_m$ son variables aleatorias y $a_1, a_2, ..., a_n, b_1, b_2, ..., b_m$ constantes, dadas las funciónes lineales

$$U_1 = a_1Y_1, a_2Y_2, ..., a_nY_n = \sum a_iY_i$$

y

$$U_2 = b_1X_1, b_2X_2, ..., b_nX_n = \sum b_jY_j,$$

donde $E(Y_i) = \mu_i$ y $E(X_j) = \xi_j$ y

$$Cov(Y_i, X_j) = E[(Y_i - \mu_i)(X_j - \xi_j)],$$

entonces, se cumple que:

$$Cov(U_1, U_2) = \sum\sum a_ib_jCov(Y_i, X_j).$$

Un valor cero de covarianza indica que las funciones lineales no están correlacionadas y que no hay dependencia lineal entre U_1 y U_2. Entre más grande es el valor absoluto de la covarianza entre U_1 y U_2, mayor se espera sea la dependencia lineal entre ellas. Sin embargo, esta medida depende de la escala de medición por lo que resulta difícil determinar a simple vista si alguna covarianza en particular es grande o pequeña.

Apéndice B

B.1. Modelo de regresión lineal

Un modelo estadístico que relaciona una respuesta aleatoria Y con un conjunto de variables independientes $x_1, x_2, ..., x_k$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon, \tag{B.1.1}$$

donde $\beta_0, \beta_1, ..., \beta_k$ son parámetros desconocidos, ϵ es una variable aleatoria y las variables $x_1, x_2, ..., x_k$ toman valores conocidos. Supondremos que $E(\epsilon) = 0$ y por lo tanto que:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 ... + \beta_k x_k. \tag{B.1.2}$$

Para manejar los resultados y deducciones sobres los modelos de regresión lineal múltiple se hace uso del álgebra de matrices; para esto suponemos el modelo lineal (B.1.1) y hacemos n observaciones independientes, $y_1, y_2, ..., y_n$ en Y.
Podemos escribir la observación y_i como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i, \tag{B.1.3}$$

donde x_{ij} es el ajuste de la i-ésima observación para la j-ésima variable independiente, con $i = 1, 2, ..., n$. Ahora definimos las siguientes matrices con $x_0 = 1$:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} x_0 & x_{11} & x_{12} & \dots & x_{1k} \\ x_0 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_0 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Así las n ecuaciones que representan y_i como función de las x, las β y las ϵ se pueden escribir sumultáneamente como:

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\epsilon}. \tag{B.1.4}$$

B.2. Estimación por mínimos cuadrados

Para estimar los parámetros de cualquier modelo lineal hacemos uso de las ecuaciones de mínimos cuadrados.
Para n observaciones de un modelo lineal simple de la forma:

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{x} + \boldsymbol{\epsilon}.$$

tenemos

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Las ecuaciones de los mínimos cuadrados para β_0 y β_1 son:

$$n\widehat{\beta}_0 + \widehat{\beta}_1 \sum x_i = \sum y_i,$$

$$\widehat{\beta}_0 \sum x_i + \widehat{\beta}_1 \sum {x_i}^2 = \sum x_i y_i.$$

Dado que $\boldsymbol{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum {x_i}^2 \end{bmatrix}$,
y $\boldsymbol{X'Y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$, vemos que las ecuaciones de mínimos cuadrados están dados por:

$$\boldsymbol{X'X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X'Y},$$

en donde:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}, \tag{B.2.1}$$

de ahí que:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}.$$

B.2.1. Propiedades de los estimadores de regresión lineal

1. $E(\widehat{\beta}_i) = \beta_i, i = 0, 1, ..., k.$
2. $V(\widehat{\beta}_i) = c_{ii}\sigma^2$, donde c_{ii} es el elemento en la fila i e columna j de $(\boldsymbol{X'X}^{-1})$. (Recuerde que esta matriz tiene una fila y una columna enumerados con 0).
3. $Cov(\widehat{\beta}_i, \widehat{\beta}_j) = c_{ij}\sigma^2$, donde c_{ij} es el elemento en la fila i y columna j de $(\boldsymbol{X'X}^{-1})$.

4. Un estimador insesgado de σ^2 es $S^2 = SSE/[n-(k+1)]$, donde $SSE = \boldsymbol{Y'Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X'Y}$. (Obśervese que hay $k+1$ valores β_i desconocidas en el modelo).

   Si además, las ϵ_i para $i = 1, 2, ..., n$ están distribuidos normalmente:

5. Cada $\widehat{\beta}_i$ está distribuida normalmente.

6. La variable aleatoria
$$\frac{n-(k+1)S^2}{\sigma^2},$$
   tiene una distribución χ^2 con $n-(k+1)$ grados de libertad.

7. Los estadísticos $S^2\widehat{\beta}_i$ son independientes para cada $i = 0, 1, 2, ..., k$.

B.3. Prueba de hipótesis e intervalo de confianza para β_i

Sea $\boldsymbol{a}'\beta$ una combinación lineal de los parámetros y $\boldsymbol{a}'\widehat{\beta}_i$ el estimador insesgado con la misma combinación lineal de los estimadores paramétricos con varianza:

$$V(\boldsymbol{a}'\beta) = V[\boldsymbol{a}'(\boldsymbol{X'X})^{-1}\boldsymbol{a}]\sigma^2,$$

donde $\boldsymbol{X}$ es una matriz de tamaño $n \times 2$ de n variables aleatorias de un modelo lineal simple.
Una prueba para $\boldsymbol{a}'\beta$ $H_0 : \boldsymbol{a}'\beta = (\boldsymbol{a}'\beta)_0$

$$H_1 : \begin{cases} \boldsymbol{a}'\beta & > & (\boldsymbol{a}'\beta)_0 & \text{(región de rechazo de cola superior)} \\ \boldsymbol{a}'\beta & < & (\boldsymbol{a}'\beta)_0 & \text{(región de rechazo de cola inferior)} \\ \boldsymbol{a}'\beta & \neq & (\boldsymbol{a}'\beta)_0 & \text{(región de rechazo de dos colas)} \end{cases}$$

Estadístico de prueba: $T = \frac{\boldsymbol{a}'\widehat{\beta} - (\boldsymbol{a}'\beta)}{S\sqrt{\boldsymbol{a}'(\boldsymbol{X'X})^{-1}\boldsymbol{a}}}$, donde $S = \sqrt{SC_{error}/(n-2)}$ es el estimador de σ.

$$\text{Región de rechazo} : \begin{cases} t & > & t_\alpha & \text{(alternativa de cola superior)} \\ t & < & -t_\alpha & \text{(alternativa de cola inferior)} \\ |t| & \neq & t_{\alpha/2} & \text{(alternativa de dos colas)} \end{cases}$$

Observe que t_α está basada en $[n-(k+1)]$ grados de libertad.

Con base en el estadístico t dado y con coeficientes de confianza $1-\alpha$ se tiene que el intervalo de confianza $100(1-\alpha)\,\%$ para $\boldsymbol{a}'\beta$,

$$\boldsymbol{a}'\widehat{\beta} \pm t_{\alpha/2}S\sqrt{(\boldsymbol{a}'(\boldsymbol{X'X})^{-1}\boldsymbol{a})}.$$

B.4. Coeficiente de correlación

Para corregir el problema de escalas de medición en el cálculo de la covarianza, se utiliza el coeficiente de correlación ρ.
Si denotamos por $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ una muestra aleatoria de observaciones, un estimador de ρ está dado por el coeficiente de correlación muestral:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{B.4.1}$$

Con la finalidad de recordarlo fácilmente, es usual denotar $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ y $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ y así expresar r en estos términos

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}. \tag{B.4.2}$$

En el caso de un modelo de regresión lineal simple, es fácil ver que $r = \widehat{\beta}_1\sqrt{\frac{S_{xx}}{S_{yy}}}$ y a partir de una tabla de análisis de varianza, el coeficiente de correlación puede calcularse directamente de la siguiente manera: $r = \sqrt{\frac{SC_{reg}}{SC_{total}}}$

B.5. Coeficiente de determinación

Por lo general se denota con R^2 al coeficiente de determinación, que mide la proporción de la variación en los valores observados de la variable Y, explicada por el conjunto de variables independientes $X_1, X_2, ..., X_k$. Se puede calcular utilizando los elementos de la tabla de análisis de varianza:

$$R^2 = 1 - \frac{SC_{error}}{SC_{tot}}. \tag{B.5.1}$$

Si se tiene que $R^2 = 1$ entonces significa que la regresión lineal explica 100 % de la variabilidad en Y. Por otro lado, si $R^2 = 0$ entonces el modelo no explica la variabilidad en Y. Un modelo se considera bueno si tiene un coeficiente de determinación aceptable, de acuerdo al investigador.

B.6. Coeficiente de determinación ajustado

Cuando se comparan dos modelos con base en su coeficiente de determinación R^2, pero tienen diferente número de variables independientes, es conveniente utilizar un coeficiente de determinación alternativo, que es el coeficiente de determinación ajustado, que se calcula como:

$$R^2_{adj} = 1 - \frac{SC^2_{error}/(n-k)}{SC^2_{tot}/(n-1)}, \tag{B.6.1}$$

donde k es el número de parámetros en el modelo, incluyendo el término de intersección (Gujarati, 2004).

B.7. Prueba de Durbin-Watson

Los residuales se definen como la diferencia $e_i = y_i - \hat{y}_i, i = 1, 2, ..., n$, donde y_i representa el dato observado y $\hat{y}_i$ es el valor obtenido del modelo de regresión ajustado. Al realizar el análisis de regresión se realizan varios supuestos sobre estos errores, como son su independencia y el que siguen una distribución normal con media cero y varianza constante σ^2 (homoscedasticidad).

La prueba de Durbin-Watson ayuda a identificar la presencia de autocorrelación entre los residuales obtenidos del modelo de regresión. Cuando se trata de dos series de tiempo, se puede pensar en una hipótesis nula que plantea que los residuales no están correlacionados frente a una hipótesis alternativa que establece que las series siguen un modelo autoregresivo de primer orden, AR(1), esto es:

$$\begin{aligned} H_0 &: \rho = 0, \\ H_1 &: \rho > 0. \end{aligned} \tag{B.7.1}$$

El estadístico de prueba se calcula como:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}, \tag{B.7.2}$$

donde n es el tamaño de la muestra, $e_i = y_i - \hat{y}_i$, es la diferencia entre el valor observado, y_i, y el valor que resulta de la regresión, $\hat{y}_i$.

El estadístico de Durbin-Watson esta relacionado con el coeficiente de correlación de la siguiente manera:

$$d \approx 2(1 - r),$$

donde $d \in (0, 4)$ y $r \in (-1, 1)$. Si $d \approx 2$ entonces $r \approx 0$, lo cual indica que no existe correlación; pero si $d \approx 0$ entonces $r \approx 1$, es decir la correlación es positiva; por el contrario, si $d \approx 4$ entonces $r \approx -1$ y se tendrá correlación negativa.
Para este estadístico se tienen tablas que toman en cuenta el tamaño de la muestra y el numńero de regresores, las cuales arrojan dos datos, cota inferior (dL) y cota superior (dU).

- Si $d < dL$, entonces se rechaza $H_0 : \rho = 0$.
- Si $d > dU$, entonces no se rechaza $H_0 : \rho = 0$.
- Si $dl < d < dU$, entonces se tienen resultados inconclusos.

Apéndice C

C.1. Valores de ruido blanco

Tiempo	Ruido blanco	Tiempo	Ruido blanco
1	-1.6331	51	0.0861
2	-0.3666	52	-0.9212
3	-0.8728	53	1.3203
4	-0.2469	54	1.4953
5	0.1266	55	-0.7741
6	-0.2462	56	0.5621
7	0.8541	57	0.9901
8	1.5789	58	-1.3516
9	-0.8701	59	1.9353
10	-0.4990	60	0.9839
11	-0.3573	61	0.2706
12	0.8395	62	-0.2419
13	0.9196	63	0.2640
14	0.6115	64	-0.2171
15	0.4214	65	-0.1377
16	-0.0239	66	-0.7410
17	0.0215	67	-2.0317
18	-0.9138	68	0.1148
19	0.1371	69	0.5314
20	1.6610	70	1.2934
21	-1.2366	71	-0.5716
22	1.8132	72	1.0440
23	-0.6895	73	-0.2226
24	0.4729	74	1.1477
25	-1.2603	75	-0.3741
26	-0.0357	76	1.3701
27	0.8874	77	0.5940
28	-0.6194	78	-1.9247
29	-1.0852	79	0.5682
30	-2.0275	80	-0.5572
31	1.2682	81	0.9881
32	-0.0463	82	0.4649
33	0.5193	83	-2.0508

34	0.6355	84	-0.5911
35	-0.1332	85	0.0923
36	-2.2015	86	-0.4776
37	1.6054	87	2.3655
38	-0.4461	88	0.3631
39	0.7574	89	0.8107
40	0.9140	90	-1.3084
41	1.1581	91	-1.4298
42	1.1783	92	0.1450
43	-0.6974	93	0.3964
44	0.4911	94	0.2056
45	-1.7856	95	1.1121
46	1.2921	96	1.2811
47	0.3684	97	0.9187
48	0.7027	98	1.4922
49	-0.8775	99	-0.7264
50	-0.7986	100	-0.6485

Tabla C.1: Valores de ruido blanco con $\mu = 0$ y $\sigma^2 < \infty$

C.2. Valores de la sección 4.3.2

Tiempo	Valor	Tiempo	Valor
1	2935833	37	1474727
2	2622529	38	1523331
3	2719635	39	2122667
4	2625179	40	2571006
5	2311187	41	2252161
6	2101758	42	2422347
7	2552638	43	2155973
8	2883423	44	2294976
9	3128904	45	2809652
10	2959348	46	2436293
11	2759000	47	2561852
12	2233755	48	2199544
13	2560858	49	2674423
14	2548821	50	2551363
15	2625675	51	3110508
16	2326076	52	3177925
17	1662956	53	3046952
18	1772409	54	2850904

19	1797275	55	3002830
20	2639852	56	2910913
21	2799990	57	2809172
22	3133285	58	3136842
23	2438296	59	3355368
24	2583766	60	3604565
25	2610157	61	3013310
26	2493415	62	3125751
27	2094163	63	2548605
28	2174301	64	2646575
29	2283420	65	2231458
30	2505128	66	1962095
31	2873785	67	1958019
32	2339727	68	2143073
33	2985829	69	2305966
34	3037351	70	2620302
35	1828265	71	2356447
36	1038562	72	2427571

Tabla C.2: Valores dela sección 4.3.2

C.3. Valores críticos de Dickey-Fuller

	τ		τ_μ		τ_τ	
	1 %	5 %	1 %	5 %	1 %	5 %
25	-2.66	-1.95	-3.75	-3.00	-4.38	-3.60
50	-2.62	-1.95	-3.58	-2.93	-4.15	-3.50
100	-2.60	-1.95	-3.51	-2.89	-4.04	-3.45
250	-2.58	-1.95	-3.46	-2.88	-3.99	-3.43
500	-2.58	-1.95	-3.44	-2.87	-3.98	-3.42
∞	-2.58	-1.95	-3.43	-2.86	-3.96	-3.41

Tabla C.3: Valores críticos de Dickey-Fuller

Bibliografía

Baumöhl, Eduard, & Lyócsa, Stefan. 2009. Stationary of time series and the problem of spurious regression. *Munich PersonalRePEc Archive*, Septiembre.

Box, George E. P., Jenkins, Gwilym M., & Reinsel, Gregory C. 2008. *Time Series Analysis, Forecasting and Control*. Fourth edn. New Jersey: John Wiley and Sons.

Chandra, K. Suresh. 1981. *Unit Root Test for Time Series in the Presence of an Explosive*. Tech. rept. Madras School of Economics.

Chatfield, Chris. 2000. *Time-Series Forecasting*. First edn. United Kingdom: The University of Bath.

Cowpertwait, Paul S.P., & Metcalfe, Andrew V. 2009. *Introductory Time Series with R*. United States: Springer.

Dickey, David A., & Fuller, Wayne A. 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, **49**, 1057–1072.

Elliott, Graham, Rothenberg, Thomas J., & Stock, James H. 1996. Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813–836.

Enders, Walter. 2015. *Applied Econometric Time Series*. United States: Willey - interscience.

G. S. Maddala, In-Moo Kim. 1998. *Unit Roots, Cointegration, and Structural Change*. First edn. United Kingdom: Cambridge University Press.

Granger, C.W.J., & Newbold, P. 1974. Spurious regression in econometric. *Journal of Econometics*, **2**, 111–120.

Granger, C.W.J., & Newbold, P. 1986. *Forecasting Economic Time Series*. Second edn. United Kingdom: Academic Press, Inc.

Guerrero, Victor M. 1990. Desestacionalización de series de tiempo económicas: introducción a la metodología. *Comercio exterior*, **40**(11), 1035–1046.

Guerrero, Victor M. 1991. *Análisis Estadístico de Series de Tiempo Económicas*. First edn. United States: Universidad Autónoma Metropolitana.

Gujarati, Damodar N. 2004. *Econometría*. Fourth edn. United States: The McGraw-Hill.

Hofer, Thomas, Przyrembel, Hildergard, & Verleger, Silvia. 2004. New evidence for the theory of the stork. *Paediatric and perinatal epidemiology*, **18**, 88–92.

Kleiber, Christian, & Zeileis, Achim. 2008. *Applied Econometrics with R*. United States: Springer.

Kwiatkowski, Denis, Phillips, Peter C. B., Schmidt, Peter, & Shin, Yongcheol. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **52**, 159–178.

Mittal, Yashaswini. 2007. *Math 562*. Tech. rept. Universidad de Arizona.

Montgomery, Douglas C., Jennings, Cheryl L., & Kulahci, Murat. 2008. *Introduction to Time Series Analysis and Forecasting*. United States: Willey - Interscience.

Mood, Alexander McFarlane. 1974. *Introduction to the theory of statistics*. United States: Mc Graw Hill.

Neusser, Klauss. 2015. *Time Series Analysis in Economics*. Tech. rept. Universidad de Berna, Sz.

Philip Hans Franses, Dick van Dijk, & Opschoor, Anne. 2014. *Time Series Models for Business and Economic Forecasting*. Second edn. United Kingdom: Cambridge University Press.

Phillips, Peter C. B. 1987. Time series regression with a unit root. *Econometrica*, **55**(2), 277–301.

Said, Said E., & Dickey, David A. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, **71**(3), 599–607.

Shumway, Robert H., & Stoffer, David S. 2011. *Time Series Analysis and Its Applications*. Third edn. New York: Springer.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Second edn. Connecticut: Graphics Press.

Yule, G. Udny. 1926. Why do we sometimes get nonsense-correlation between time series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society*, **89**(1), 1–63.

Zivot, Eric, & Andrews, Donald W.K. 1992. Further Evidence on the great crash, the oil-price schock and the unit-root hypothesis. *Journal of Business and Economic Statistics*, **10**(3), 251–270.