



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Análisis estadístico del consumo de electricidad en México

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I A

P R E S E N T A:

MARÍA DEL MAR BAZÚA LOBATO

TUTORA:

DRA. RUTH SELENE FUENTES GARCÍA

CIUDAD UNIVERSITARIA, CD. MX,

2016





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Agradecimientos

A la memoria de Víctor, mi padre, por forjarme un pensamiento crítico.

A Perla, mi madre, por su apoyo y enseñanzas.

A Rodrigo, mi hermano, por impulsarme a titularme.

A mis tíos Jorge y Luisfe por la idea de utilizar los datos de ventas internas de electricidad.

A mis abuelos Coty y Rodolfo por su ejemplo y sus consejos.

A mi tío Rich por sus recomendaciones.

A mi abuela Isabel por su ejemplo de mujer trabajadora.

A Néstor por comprenderme e impulsarme.

Finalmente a Ruth, mi tutora, por su guía y supervisión a lo largo de este proceso.

Índice general

Introducción	5
1. La electricidad en México	6
1.1. La energía eléctrica y la industrialización	6
1.2. El sector eléctrico en la actualidad	7
1.3. Las ventas de electricidad en el mercado interno y su análisis	8
2. Características de las series de tiempo	9
2.1. Procesos estocásticos	10
2.2. Medidas de dependencia	11
2.3. Procesos Estacionarios	11
2.3.1. Estacionariedad	12
2.3.2. Funciones de autocovarianza y de autocorrelación de un proceso débilmente estacionario	13
2.3.3. Estimación de la media y de la función de autocovarianza	14
2.4. Transformaciones	15
2.4.1. Transformaciones de Box-Cox	15
2.4.2. Suavizamiento	16
2.4.3. Diferencias	18
3. Modelos autoregresivos y de promedios móviles	20
3.1. Modelos de promedios móviles	20
3.2. Modelos autoregresivos	23
3.3. Modelos autoregresivos de promedios móviles	29

3.4. Modelos autoregresivos integrados de promedios móviles	30
3.5. Estimación de los parámetros	30
3.5.1. Estimadores de momentos	31
3.5.2. Estimadores por mínimos cuadrados	32
3.5.3. Estimadores máximo verosímiles	33
3.6. Selección del orden del modelo	34
3.7. Verificación del modelo	35
3.8. Construcción de modelos ARMA	36
3.9. Pronósticos	37
4. Análisis de las series de ventas de electricidad	39
4.1. Servicios	39
4.1.1. Primera propuesta	41
4.1.2. Segunda propuesta	45
4.1.3. Comparación de los modelos	47
4.1.4. Tercera propuesta	49
4.1.5. Propuesta final	53
4.2. Doméstico	55
4.2.1. Primera propuesta	55
4.2.2. Segunda propuesta	56
4.2.3. Tercera propuesta	61
4.2.4. Propuesta final	64
Conclusiones	65
Bibliografía	69
Índice de cuadros	70
Índice de figuras	72

Introducción

La presente tesis tiene como objetivo utilizar los modelos Autoregresivos de Promedios Móviles (ARMA) para modelar las series de ventas internas de electricidad a los sectores Doméstico y Servicios en México. Mi principal intención al realizar la tesis fue la de modelar datos reales aplicando alguna de las técnicas aprendidas en los cursos de estadística. Dado que las series de ventas internas de electricidad son series de tiempo, decidí analizarlas utilizando el enfoque aprendido en Estadística 3 buscando ajustar modelos ARMA a dichas series y utilizando el software R.

La tesis está constituida por 4 capítulos.

En el capítulo 1 se brinda un panorama general del sector eléctrico de México. Se habla brevemente de la historia de la electricidad en México y se presentan tanto los sectores que conforman al sector eléctrico del país como las series de tiempo que serán analizadas.

En el capítulo 2 se presentan las características de las series de tiempo, esto es: la definición de serie de tiempo como realización de un proceso estocástico y a partir de ahí la caracterización de dicho proceso a partir de las funciones de autocovarianza y de autocorrelación. En este capítulo se presenta también a los procesos estacionarios y débilmente estacionarios junto con sus principales características. Al final del capítulo se presentan las principales transformaciones que se aplican a las series de tiempo para poder verlas como realizaciones de un proceso débilmente estacionario.

En el capítulo 3 se presenta a los modelos ARMA y al proceso para ajustar dichos modelos a una serie de tiempo. Se presentan además los principales métodos para estimar los parámetros del modelo a ajustar y cómo evaluar el ajuste del modelo.

En el capítulo 4 se encuentra el análisis de las series de ventas internas de electricidad para los sectores Servicios y Doméstico.

Finalmente, se presentan las conclusiones.

Capítulo 1

La electricidad en México

1.1. La energía eléctrica y la industrialización

Durante el siglo XIX, México tuvo muchas dificultades para incorporarse a la industrialización basada en el vapor y para aprovechar la fuerza hidráulica, pues el país carecía de carbón y en los meses de sequía los caudales de los ríos disminuían. Sin embargo, el terreno montañoso del país era ideal para la instalación de centrales eléctricas, por lo que la generación de energía eléctrica se convirtió en un factor clave para el desarrollo industrial de México.

A finales del siglo XIX, las principales industrias de México, la minera y la textil, fueron pioneras en la instalación de plantas eléctricas. En 1879, con la instalación de la primera planta termoeléctrica del país, la cual era utilizada por la fábrica textil *La Americana*, comenzó una rápida expansión de la generación de electricidad.

En la década de los cincuenta, el país experimentó un gran crecimiento económico basado principalmente en el mercado interno. El sector eléctrico fue aquel con mayor crecimiento; de 1952 a 1958, la generación de energía eléctrica tuvo un aumento del 69.7% al pasar de 5,337 MegaWattshora (MWh) a 9,057 MWh. Este crecimiento se logró gracias a una política que combinaba un intenso aprovechamiento de la capacidad instalada con la creación de nuevas plantas generadoras. La capacidad de generación de energía eléctrica ha seguido creciendo a lo largo de los años, a principios de esta década era de 53,114 MWh.

La generación de energía eléctrica representó también una gran mejora en el nivel de vida de la población del país; se comenzó con la instalación de alumbrado público y el bombeo de agua potable y posteriormente las viviendas empezaron a contar con electricidad. Según información del Censo de Población y Vivienda de 2010, el 97.78% de las viviendas del país cuenta con electricidad.

1.2. El sector eléctrico en la actualidad

Si bien en sus inicios el sector eléctrico nacional se encontraba en manos privadas, hoy en día la Comisión Federal de Electricidad (CFE) es la encargada de suministrar energía eléctrica en todo el país. Alrededor del 99.5 % de las ventas totales de energía eléctrica lo constituyen las ventas en el mercado interno y el 0.5 % restante exportaciones. A mayo del 2013, el mercado interno estaba conformado por 36.9 millones de usuarios, divididos en los sectores Doméstico, Comercial, Agrícola, Servicios e Industrial. A continuación se presenta una breve descripción de estos sectores:

- Doméstico: está constituido por los hogares, edificios, etc. Estos se encuentran conectados en baja tensión¹.
- Comercial: ventas realizadas a los comercios conectados en baja tensión, como las “tienditas de la esquina”.
- Agrícola: la electricidad utilizada para el bombeo de agua para riego agrícola; ésta puede ser distribuida en baja tensión o en media tensión².
- Servicios: energía eléctrica utilizada para el alumbrado público y para el bombeo tanto de agua potable como de aguas negras. Este sector también se encuentra conectado en baja tensión.
- Industrial: se encuentra dividido en Empresa mediana y Gran industria. El sector Empresa mediana está conformado por las industrias y empresas conectadas en media tensión mientras que el sector Gran industria lo está por aquellas conectadas en alta tensión³.

Cuadro 1.1: Número de usuarios de energía eléctrica por sector, a mayo del 2013.

Sector	Usuarios	Porcentaje
Doméstico	32,617,470	88.46 %
Agrícola	125,841	0.34 %
Comercial	366,1001	9.93 %
Servicios	191,651	0.52 %
Industrial	276,162	0.75 %

Elaboración propia con datos de la Secretaría de Energía.

El sector Doméstico es el más numeroso; el 88.46 % de los usuarios pertenecen a dicho sector. Sin embargo, sus ventas representan solamente el 22.92 % de las ventas en el mercado interno. Mientras que el sector Industrial con menos del 1 % de los usuarios representa más de la mitad de las ventas.

El sector Industrial se divide en Mediana empresa y Gran industria. De los 276,162 usuarios del sector Industrial, tan sólo el 0.33 % pertenece a la Gran industria, sin embargo, este 0.33 % de usua-

¹Baja tensión es el servicio que se suministra en niveles de tensión menores o iguales a 1,000 volts.

²Media tensión es el servicio que se suministra en niveles de tensión mayores a un kilovolt, pero menores a 35 kilovolts.

³Alta tensión a nivel subtransmisión es el servicio suministrado en niveles de tensión mayores a 35 kilovolts, pero menores a 220 kilovolts.

Cuadro 1.2: Ventas de electricidad en el mercado interno, de enero de 2013 a mayo del 2013 en MWh.

Sector	Ventas	Porcentaje
Doméstico	18,554,323.54	22.92 %
Agrícola	4,900,853.09	6.05 %
Comercial	5,364,544.24	6.63 %
Servicios	3,600,651.37	4.45 %
Industrial	48,547,419.34	59.96 %

Elaboración propia con datos de la Secretaría de Energía.

rios representa el 37.76 % de las ventas al sector Industrial.

Cuadro 1.3: Estructura del sector Industrial a mayo del 2013.

Sector	Usuarios	Ventas (MW-hora)
Empresa mediana	275,254	6,721,603.56
Gran industria	908	3,758,032.87
Industrial	276,162	10,479,636.43

Elaboración propia con datos de la Secretaría de Energía.

1.3. Las ventas de electricidad en el mercado interno y su análisis

Para el desarrollo del país, es fundamental que la CFE garantice un suministro eléctrico suficiente, oportuno y de calidad, además de contar con la solidez económica suficiente para afrontar sus gastos de operación y expansión. Las ventas de electricidad en el mercado interno reflejan el consumo de energía eléctrica en el país, el entender y pronosticar su comportamiento es fundamental en los procesos de: planeación de inversiones, políticas de precios y desarrollo del sector eléctrico nacional. De ahí la importancia de analizar la información referente a las ventas de electricidad en el mercado interno.

En la presente tesis se ejemplificará el uso de la teoría de series de tiempo y de los modelos Auto-regresivos de Promedios Móviles (ARMA). Dicho ejercicio se realizará para los sectores Doméstico y Servicios ya que ambos sectores impactan directamente en nuestra vida cotidiana y sus series de ventas permiten ilustrar la aplicación de la teoría de series de tiempo y de los modelos ARMA. Se trabajará con las series de ventas mensuales de ambos sectores de enero de 2002 a mayo de 2013, en Gigawatts-hora.

La teoría de series de tiempo y de los modelos ARMA se encuentra en los capítulos 2 y 3, respectivamente. Los análisis de las series de ventas a los sectores Doméstico y Servicios se encuentran en el capítulo 4.

Capítulo 2

Características de las series de tiempo

Una serie de tiempo es un conjunto de observaciones que han sido recolectadas de forma secuencial a lo largo del tiempo. Las series de ventas internas de electricidad que se presentaron en el primer capítulo de esta tesis son series de tiempo. Otros ejemplos de series de tiempo son: las tasas de desempleo y de natalidad, las ventas semanales de un producto, los índices de precios, la producción mensual de una fábrica, el precio de una acción en el mercado bursátil, el nivel de un río, la temperatura máxima diaria en una ciudad, etc.

El análisis de series de tiempo puede realizarse desde dos enfoques distintos: el del tiempo y el de las frecuencias. En el enfoque del tiempo, se estudian las relaciones entre el valor de una variable en un punto del tiempo y los valores de la misma en puntos anteriores del tiempo. Cuando se utiliza este enfoque se busca modelar a la serie como una función paramétrica de sus valores presente y pasados. En el enfoque de las frecuencias se estudian las oscilaciones periódicas que se encuentran de forma natural en la mayoría en los datos. En la presente tesis se utilizará el enfoque del tiempo.

Una serie de tiempo, desde el punto de vista del dominio del tiempo, puede ser continua o discreta. Como sus nombres lo indican, en una serie de tiempo continua las observaciones son recolectadas de manera continua sobre algún intervalo de tiempo, mientras que, en una serie de tiempo discreta, las observaciones son tomadas en tiempos discretos, generalmente separados por la misma distancia, por ejemplo: días, semanas, meses, años, etc.

Existen diversos propósitos por los cuales analizar una serie de tiempo. Algunos de los más comunes son: dar una descripción compacta de los datos, hacer predicciones, analizar o describir la estructura de una serie, medir el efecto de una intervención o bien medir el grado de correlación ente dos o más series de tiempo.

Para hacer inferencias acerca de una serie de tiempo es necesario encontrar un modelo matemático que permita representar a los datos, y para ello, primero hay que definir matemáticamente qué es una serie de tiempo. En la siguiente sección se presenta dicha definición.

2.1. Procesos estocásticos

En el análisis de series de tiempo, como en cualquier trabajo estadístico, es conveniente ver a los datos observados como uno de los muchos posibles conjuntos de datos que pudieron haberse presentado. Esto se logra asociando a cada punto del tiempo t una variable aleatoria $X(t) \in (-\infty, \infty)$. De esta forma, una serie de tiempo puede ser asociada a un conjunto ordenado de variables aleatorias $\{X(t), -\infty \leq t \leq \infty\}$ en el caso continuo y $\{X_t, t \in \mathbb{Z}\}$ en el caso discreto.

Una vez que se tiene esta asociación, las definiciones de proceso estocástico y de realización de un proceso estocástico son necesarias para poder definir matemáticamente a una serie de tiempo.

Definición 2.1.1. Proceso estocástico.

Un proceso estocástico es una familia de variables aleatorias $\{X_t, t \in T\}$ definida en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})^1$, donde T es un conjunto de índices tal que $\#T = \infty$.

Definición 2.1.2. Realizaciones de un proceso estocástico.

Las funciones $\{X_t(\omega), \omega \in \Omega\}$ en T son conocidas como realizaciones o trayectorias del proceso $\{X_t, t \in T\}$.

Dadas las dos definiciones anteriores, una serie de tiempo puede ser vista como una realización $\{x_t, t \in T_0\}$, o parte de una realización, de un proceso estocástico $\{X_t, t \in T\}$. Donde $T_0 \subseteq T$ y T representa a los puntos del tiempo. Generalmente T es de la forma $\{0, \pm 1, \pm 2, \pm 3, \dots\}$, $\{1, 2, 3, \dots\}$, $[0, \infty)$ ó $(-\infty, \infty)$. En las series de tiempo de ventas internas de electricidad que se analizarán en la presente tesis, $T_0 = \{1, 2, \dots, 137\}$.

Nótese que para cada $t \in T$ fija, $X_t(\omega)$ es una función de Ω , y que para cada $\omega \in \Omega$ fija, $X_t(\omega)$ es una función de T .

Una vez que se tiene la definición de una serie de tiempo como una realización de un proceso estocástico, es necesario buscar una forma de caracterizar a dicho proceso. Dado que una variable aleatoria está caracterizada por su función de distribución, resulta natural fijarse en las funciones de distribución conjunta de las variables aleatorias que constituyen al proceso.

Definición 2.1.3. Funciones de distribución de un proceso estocástico $\{X_t, t \in T \subset \mathbb{R}\}$.

Sea \mathcal{T} el conjunto de todos los vectores $\{\mathbf{t} = (t_1, \dots, t_n)' \in T^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}$. Entonces, las distribuciones finito-dimensionales de $\{X_t\}$ son las funciones $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$. Definidas para $\mathbf{t} = (t_1, \dots, t_n)' \in \mathcal{T}$ por:

$$F_{\mathbf{t}}(\mathbf{x}) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n), \quad \mathbf{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$$

La caracterización completa de un proceso estocástico $\{X_t, t \in T\}$ está dada por sus funciones de distribución finito-dimensionales $F_{\mathbf{t}}(\mathbf{x})$. Sin embargo, en la mayoría de los casos es muy difícil tanto encontrarlas como trabajar con ellas, por lo que es necesario buscar una forma más simple de

¹Un espacio muestral Ω , una σ -álgebra \mathcal{F} de subconjuntos de Ω , una función de probabilidad \mathbb{P} .

caracterizar al proceso $\{X_t, t \in T\}$.

Dado que en las series de tiempo que se analizarán en la presente tesis se tiene que $T \subseteq \mathbb{Z}$, a partir de este momento y a menos que se especifique lo contrario, se trabajará con procesos de la forma $\{X_t, t \in T\}$, donde $T \subseteq \mathbb{Z}$ y se utilizará la notación $\{X_t\}$ para denotar al proceso $\{X_t, t \in T\}$.

2.2. Medidas de dependencia

Al trabajar con un número finito de variables aleatorias, comúnmente se calcula la matriz de covarianza para conocer la estructura de dependencia que existe entre las variables. Para el análisis de series de tiempo es necesario extender el concepto de matriz de covarianza.

Definición 2.2.1. Función de autocovarianza.

Si $\{X_t\}$ es un proceso tal que $Var(X_t) < \infty$ para cada $t \in T$, entonces la función de autocovarianza $\gamma(\cdot, \cdot)$ de $\{X_t\}$ está definida por

$$\gamma(r, s) = Cov(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \quad r, s \in T. \quad (2.1)$$

Observación 2.2.1. Nótese que si X_s y X_r siguen una distribución normal, el hecho de que $\gamma(r, s) = 0$ asegura su independencia. Sin embargo, en general el que $\gamma(r, s)$ sea igual a 0 garantiza que X_r y X_s no estén relacionadas linealmente más no garantiza el que sean independientes.

Una vez que se tiene a la función de autocovarianza γ , es posible definir a la función de autocorrelación (ACF). Dicha función mide el grado de correlación entre dos puntos distintos X_r y X_s de un proceso.

Definición 2.2.2. Función de autocorrelación (ACF).

La función de autocorrelación de un proceso $\{X_t\}$ tal que $Var(X_t) < \infty \quad \forall t \in T$, está definida como:

$$\rho(r, s) = \frac{Cov(X_r, X_s)}{\sqrt{Var(X_r)}\sqrt{Var(X_s)}} = \frac{\gamma(r, s)}{\sqrt{\gamma(r, r)}\sqrt{\gamma(s, s)}}, \quad r, s \in T. \quad (2.2)$$

2.3. Procesos Estacionarios

Es deseable trabajar con procesos que posean características, tales como la media y la autocovarianza, estables a lo largo del tiempo. A este tipo de procesos se les llama estacionarios y tienen un papel fundamental en el análisis de series de tiempo.

2.3.1. Estacionariedad

A continuación se presentan las condiciones que debe cumplir un proceso $\{X_t, t \in \mathbb{Z}\}$ para ser estacionario.

Definición 2.3.1. Estacionariedad estricta.

Se dice que un proceso $\{X_t, t \in \mathbb{Z}\}$ es estrictamente estacionario si para cualquier subconjunto de variables X_{t_1}, \dots, X_{t_k} , $k \in \mathbb{Z}$ y para cualquier $h \in \mathbb{Z}$ se cumple que: $(X_{t_1}, \dots, X_{t_k})$ y $(X_{t_1+h}, \dots, X_{t_k+h})$ tienen la misma distribución.

La condición de estacionariedad estricta es demasiado fuerte para la mayoría de las aplicaciones ya que impone condiciones sobre todas las posibles distribuciones finito-dimensionales de un proceso. Es por ello que normalmente se utiliza una versión de estacionariedad más débil que solamente impone condiciones sobre los primeros dos momentos de los procesos.

Definición 2.3.2. Estacionariedad débil.

Un proceso $\{X_t, t \in \mathbb{Z}\}$, es débilmente estacionario (o estacionario de segundo orden) si satisface las siguientes condiciones:

- (I) $E[|X_t|^2] < \infty$ para todo $t \in \mathbb{Z}$.
- (II) $E[X_t] = m$ para todo $t \in \mathbb{Z}$.
- (III) $\gamma(r, s) = \gamma(r + t, s + t)$ para todo $r, s, t \in \mathbb{Z}$.

Es decir, un proceso es débilmente estacionario si posee segundo momento finito, si su esperanza no depende del tiempo y si su función de autocovarianza solamente depende del tiempo a través del rezago h , donde $h = r - s$.

Observación 2.3.1. Estacionariedad estricta implica estacionariedad débil. Sin embargo, en general, estacionariedad débil no implica estacionariedad estricta

Definición 2.3.3. Procesos de ruido blanco.

Un proceso de ruido blanco es una colección $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \dots\}$ de variables aleatorias idénticamente distribuidas tales que:

- $E[\varepsilon_i] = 0 \quad \forall i \in \mathbb{Z}$.
- $Cov(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 & \text{si } i \neq j, \\ \sigma^2 & \text{si } i = j. \end{cases} \quad \forall i, j \in \mathbb{Z}$

A lo largo de esta tesis se utilizará la notación WN para referirse a un proceso de ruido blanco.

Definición 2.3.4. Proceso Gaussiano.

El proceso $\{\varepsilon_t\}$ es Gaussiano si y sólo si todas las funciones de distribución finito dimensionales de $\{\varepsilon_t\}$ son normales multivariadas.

Los Procesos Guassinanos son un caso particular de los procesos de ruido blanco en los que estacionariedad débil implica estacionariedad estricta.

2.3.2. Funciones de autocovarianza y de autocorrelación de un proceso débilmente estacionario

Si $\{X_t, t \in \mathbb{Z}\}$ es débilmente estacionario, se tiene que $\gamma(r, s) = \gamma(r - s, 0)$ para todo $r, s \in \mathbb{Z}$. Por lo que es posible redefinir a la función de autocovarianza $\gamma(\cdot, \cdot)$ de un proceso débilmente estacionario como una función de una sola variable, a saber:

$$\gamma(h) = \gamma(h, 0) = Cov(X_{t+h}, X_t) \quad \forall t, h \in \mathbb{Z}. \quad (2.3)$$

A la función $\gamma(\cdot)$ se le conoce como función de autocovarianza de $\{X_t\}$ y a $\gamma(h)$ como su valor en el rezago h .

Observación 2.3.2. Si $\{X_t, t \in \mathbb{Z}\}$ es un proceso débilmente estacionario entonces su varianza es constante.

Observación 2.3.3. Nótese que si $\{X_t, t \in \mathbb{Z}\}$ es un proceso gaussiano y $\gamma(h) = 0$ para alguna $h \neq 0 \in \mathbb{Z}$, entonces, $\forall t \in \mathbb{Z}$ se tiene que X_t y X_{t+h} son independientes.

Proposición 2.3.1. Propiedades elementales de la función de autocovarianza.

Si $\gamma(\cdot)$ es la función de autocovarianza de un proceso débilmente estacionario $\{X_t, t \in \mathbb{Z}\}$, entonces

- (I) $\gamma(0) \geq 0$,
- (II) $|\gamma(h)| \leq \gamma(0)$ para toda $h \in \mathbb{Z}$,
- (III) $\gamma(h) = \gamma(-h)$ para toda $h \in \mathbb{Z}$.

Demostración.

- (I) Como $Var(X_t) \geq 0$ se tiene que,

$$\gamma(0) = Cov(X_{t+0}, X_t) = Var(X_t) \geq 0.$$

- (II) La segunda propiedad es una consecuencia inmediata de la desigualdad de Cauchy-Schwarz,

$$\begin{aligned} |\gamma(h)| &= |Cov(X_{t+h}, X_t)| \\ &\leq \sqrt{Var(X_{t+h})} \sqrt{Var(X_t)} \\ &= \sqrt{Var(X_t)} \sqrt{Var(X_t)} \quad \text{por la observación (2.3.2)} \\ &= Var(X_t) \\ &= \gamma(0). \end{aligned}$$

- (III) La tercera propiedad se tiene al observar que

$$\gamma(h) = Cov(X_t, X_{t+h}) = Cov(X_{t-h}, X_t) = \gamma(-h).$$

□

Si $\{X_t, t \in \mathbb{Z}\}$ es débilmente estacionario, también es posible redefinir a la función de autocorrelación parcial como una función de una sola variable, a saber:

$$\rho(h) = \text{Corr}(X_{t+h}, X_t) = \frac{\text{Cov}(X_{t+h}, X_t)}{\sqrt{\text{Var}(X_{t+h})}\sqrt{\text{Var}(X_t)}} = \frac{\gamma(h)}{\gamma(0)}, \quad \forall t, h \in \mathbb{Z}. \quad (2.4)$$

Observación 2.3.1. La función de autocorrelación $\rho(\cdot)$ cumple todas las propiedades de la función de autocovarianza y además satisface que $\rho(0) = 1$.

Observación 2.3.4. Los conceptos de estacionariedad estricta y débil fueron definidos en el caso en el que $T = \mathbb{Z}$. Es posible definir los conceptos anteriores en el caso más general en el que $T \subset \mathbb{Z}$, pero, para los propósitos de la tesis no es necesario, ya que, aunque en las aplicaciones se trabaja con observaciones de la forma $\{x_t, t \in T_0 \subset \mathbb{Z}\}$, estas pueden ser modeladas como parte de una realización de un proceso estacionario o débilmente estacionario $\{X_t, t \in \mathbb{Z}\}$.

2.3.3. Estimación de la media y de la función de autocovarianza

Un proceso débilmente estacionario $\{X_t\}$ está caracterizado por su media μ y su función de autocovarianza γ . Cuando únicamente se tienen las observaciones $\{x_1, \dots, x_n\}$ es necesario estimar a μ y γ a partir de dichas observaciones.

El estimador insesgado natural de la media de un proceso estacionario $\{X_t\}$ es la media muestral (\bar{x}) de las observaciones $\{x_1, \dots, x_n\}$. Esto es:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

El estimador $\hat{\gamma}(h)$ de la función de autocovarianza $\gamma(h)$ está dado por:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}) \quad h = 0, 1, \dots, n-1.$$

y $\hat{\gamma}(h) = \hat{\gamma}(-h)$.

La función de autocorrelación muestral está definida en términos de la función de autocovarianza muestral de la siguiente manera:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}, \quad |h| < n.$$

Los estimadores de las funciones de autocovarianza y de autocorrelación son sesgados cuando se utiliza el factor $1/n$ e incluso cuando se utiliza el factor $1/(n-h)$. Sin embargo, cuando se tienen muestras grandes, bajo algunos supuestos generales, ambos estimadores son asintóticamente insesgados. Generalmente se utiliza el factor $1/n$ ya que éste asegura que las matrices

$$\hat{\Gamma}_n(h) := [\hat{\gamma}(i-j)]_{i,j=1}^n \quad y \quad \hat{R}_n(h) := [\hat{\rho}(i-j)]_{i,j=1}^n$$

sean no-negativas definidas.

Observación 2.3.5. Los estimadores de las funciones de autocovarianza y de autocorrelación pueden ser calculados para cualquier conjunto de datos $\{x_1, \dots, x_n\}$ y no se encuentran restringidos a realizaciones de un proceso débilmente estacionario.

2.4. Transformaciones

Muchas de las series de tiempo que se observan en la vida real no son débilmente estacionarias pues en ellas puede observarse que la media no es constante a lo largo del tiempo, que la variabilidad de los datos aumenta conforme avanza el tiempo o la presencia de variaciones o de ciclos estacionales². Afortunadamente, existen diferentes técnicas que permiten transformar a las series de tiempo para que sea razonable verlas como realizaciones de algún proceso débilmente estacionario.

Las técnicas más utilizadas para convertir a una serie de tiempo en débilmente estacionaria, son las transformaciones de Box-Cox, la diferenciación y el suavizamiento utilizando promedios móviles o ajustando polinomios. A continuación se presentan dichas técnicas.

2.4.1. Transformaciones de Box-Cox

Cuando la variabilidad de una serie de tiempo $\{x_1, \dots, x_n\}$ aumenta a lo largo del tiempo, las transformaciones de Box-Cox pueden ser utilizadas para estabilizar a la misma. Éstas son de la forma:

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \text{si } \lambda \in (-\infty, \infty) - \{0\}, \\ \ln(x_t) & \text{si } \lambda = 0. \end{cases}$$

La más utilizada es el caso particular en el que $\lambda = 0$.

Si además de una variabilidad que aumenta a lo largo del tiempo, una serie presenta una media que no es constante y variaciones o ciclos estacionales, primero es necesario aplicar una transformación para estabilizar a la varianza y posteriormente aplicar alguna de las transformaciones que se presentan en las siguientes secciones para llevar la serie transformada a la estacionariedad débil.

²Las variaciones estacionales son cambios regulares en los datos que ocurren cada año calendario. Los ciclos pueden ocurrir en periodos más largos o más cortos que un año calendario.

2.4.2. Suavizamiento

Una serie de tiempo $\{x_1, \dots, x_n\}$ puede ser vista como la realización de un proceso

$$X_t = m_t + s_t + Y_t, \quad (2.5)$$

donde:

- m_t es una función que cambia lentamente y que es conocida como *tendencia*,
- s_t es una función con periodo conocido d a la que se le conoce como *estacionalidad* en la que se cumple que $s_{t+d} = s_t$ y $\sum_{i=1}^d s_i = 0$,
- Y_t es un proceso débilmente estacionario tal que $E[Y_t] = 0$.

Polinomios ajustados

En el método de suavizamiento con polinomios ajustados, el objetivo es estimar y extraer a m_t y a s_t del modelo $X_t = m_t + s_t + Y_t$, esperando que el componente restante, Y_t , sea un proceso débilmente estacionario.

Si $s_t = 0$, se tiene el caso de no-estacionariedad más simple, en él, el proceso presenta un comportamiento estacionario alrededor de una tendencia. Para estimar m_t se supone que ésta es de la forma

$$m_t = a_0 + a_1 t + \dots + a_p t^p.$$

Obsérvese que si $p = 1$ la tendencia es lineal y que si $p = 2$ la tendencia es cuadrática. Los parámetros $\{a_i\}$ se estiman mediante mínimos cuadrados ordinarios, es decir, se buscan las $\{a_i\}$ que minimicen $\sum_{i=1}^n (x_i - m_i)^2$.

En la presencia de tendencia y estacionalidad, pueden utilizarse funciones de senos y cosenos, combinadas con un polinomio de la forma $a_0 + a_1 t + \dots + a_p t^p$ para remover la tendencia. Por ejemplo, si la tendencia es cuadrática y se tiene un periodo $d = 12$, se puede utilizar una función de la forma:

$$f_t = a_0 + a_1 t + a_2 t^2 + \beta_1 \cos(2\pi t/12) + \beta_1 \text{sen}(2\pi t/12)$$

donde los estimadores $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p, \hat{\beta}_1, \hat{\beta}_2$ se obtienen mediante mínimos cuadrados ordinarios.

Promedios móviles

Otra forma de suavizar a la serie de tiempo $\{x_1, \dots, x_n\}$ consiste en utilizar promedios móviles de uno o de dos lados. Existen diferentes maneras de hacerlo, a continuación se presentan las más comunes.

Una primera forma de suavizamiento consiste en utilizar recursivamente promedios móviles de un lado, a saber:

$$\hat{m}_t = ax_t + (1 - a)\hat{m}_{t-1}, \quad t = 2, \dots, n,$$

y

$$\hat{m}_1 = x_1,$$

donde $a \in [0, 1]$ fija. A este tipo de suavizamiento se le conoce como suavizamiento exponencial. El suavizamiento exponencial puede utilizarse para hacer pronósticos.

Otra forma de suavizamiento se tiene al utilizar un promedio móvil de dos lados y simétrico:

$$m_t = \sum_{j=-k}^k a_j x_{t-j}, \quad k \in \mathbb{Z}, a \in \mathbb{R}$$

donde $a_j = a_{-j}$ y $\sum_{j=-k}^k a_j = 1$.

Si los datos $\{x_1, \dots, x_n\}$ presentan tanto tendencia como estacionalidad, generalmente en la literatura se propone hacer una primera estimación de la tendencia m_t utilizando un promedio móvil simétrico elegido especialmente para estimar y remover al componente estacional s_t y posteriormente re-estimar a la tendencia restante.

Si el periodo d es impar, $d = 2q + 1$, se estima a m_t de la siguiente forma:

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^q (X_{t+j}).$$

Si el periodo d es par, $d = 2q$, entonces se utiliza:

$$\hat{m}_t = \frac{1}{2d} x_{t-q} + \frac{1}{d} x_{t-q+1} + \dots + \frac{1}{d} x_{t+q-1} + \frac{1}{2d} x_{t+q}, \quad q < t \leq n - q.$$

Para estimar al componente estacional, para cada $k = 1, \dots, d$ primero se calcula el promedio w_k de las desviaciones $\{(x_{k+jd} - \hat{m}_{k+jd}) : q < k + jd \leq n - q\}$. Como las w_k no necesariamente suman cero, se estima al componente estacional s_k de la siguiente manera:

$$\hat{s}_k = w_k - d^{-1} \sum_{i=1}^d w_i, \quad k = 1, \dots, d,$$

donde $\hat{s}_k = \hat{s}_{k-d}$ y $k > d$.

Los datos desestacionalizados (d_t) se definen entonces de la siguiente forma:

$$d_t = x_t - \hat{s}_t, \quad t = 1, \dots, n.$$

Finalmente, la tendencia es re-estimada a partir de la serie desestacionalizada $\{d_t\}$ aplicando un promedio móvil, como el descrito anteriormente, o ajustando un polinomio a la serie $\{d_t\}$ y se trabaja con la serie $\hat{y}_t = x_t - \hat{m}_t - \hat{s}_t$.

2.4.3. Diferencias

Este método, el cual fue desarrollado ampliamente por George Box y Gwilym Jenkins durante los años sesenta, consiste en aplicar diferencias a la serie de tiempo $\{x_1, \dots, x_n\}$ hasta que las observaciones diferenciadas se parezcan a las realizaciones de algún proceso débilmente estacionario $\{Y_t\}$.

Definición 2.4.1. Operador diferencia (∇).

El operador diferencia está definido por

$$\nabla(X_t) = X_t - X_{t-1} = (1 - B)X_t \quad (2.6)$$

donde B es el operador de rezago o retraso

$$B(X_t) = X_{t-1} \quad (2.7)$$

El operador diferencia al rezago d (∇_d) está dado por:

$$\nabla_d(X_t) = (1 - B^d)X_t = X_t - X_{t-d} \quad (2.8)$$

Observación 2.4.1.

- (I) $B^j(X_t) = X_{t-j}, \quad j \geq 1$
- (II) $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t)), \quad j \geq 1$
- (III) $\nabla^j(X_t) = (1 - B)^j(X_t) \neq (1 - B^j)(X_t) = \nabla_j(X_t)$
- (IV) Los polinomios en B y ∇ son manipulados como funciones de polinomios de variable real. Por ejemplo:

$$\begin{aligned} \nabla^2(X_t) = \nabla(\nabla(X_t)) &= (1 - B)(1 - B)(X_t) \\ &= (1 - 2B + B^2)(X_t) \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned}$$

Para eliminar la tendencia y la estacionalidad de la serie de tiempo, primero se aplica el operador ∇_d al modelo $X_t = m_t + s_t + Y_t$, donde $\{s_t\}$ tiene periodo d , y se obtiene:

$$\nabla_d(X_t) = (m_t - m_{t-d}) + (Y_t - Y_{t-d}), \quad (2.9)$$

la cual da una descomposición de la diferencia ∇_d en un término de tendencia y en un término de ruido. La tendencia ($m_t - m_{t-d}$) es eliminada entonces aplicando una potencia del operador ∇ .

En las secciones 2.4.1 a 2.4.3 se presentaron los métodos más utilizados para llevar a una serie de tiempo a la estacionariedad débil. La elección del método a utilizar depende de diversos factores, entre ellos, si los estimadores de la tendencia y la estacionalidad son requeridos o si el componente estacional varía a lo largo del tiempo. Una ventaja de la diferenciación sobre los métodos de suavizamiento, consiste en que no es necesario estimar ningún parámetro. Sin embargo, como se observa en la ecuación (2.9), al diferenciar no se obtiene un estimador para el proceso Y_t . Por lo tanto, si el objetivo es estimar al proceso débilmente estacionario Y_t , a m_t o a s_t , es necesario suavizar la serie mediante polinomios o promedios móviles. Por otro lado, si el objetivo solamente es llevar a la serie de tiempo observada a la estacionariedad débil, es conveniente aplicarles el operador diferencia. Para los propósitos de esta tesis no es necesario obtener estimadores para Y_t , s_t y m_t , así que se utilizará al operador diferencia para transformar a las series de ventas de electricidad en débilmente estacionarias y solamente en los casos en los que diferenciar no sea suficiente, se utilizarán los métodos de suavizamiento.

Una vez que la tendencia y la estacionalidad han sido removidas el siguiente paso consiste en modelar a la serie débilmente estacionaria resultante $\{\hat{y}_t\}$. De no existir una estructura de dependencia entre ellos, estos pueden verse como observaciones de variables aleatorias independientes e idénticamente distribuidas, en cuyo caso no queda mucho que hacer excepto estimar su media y su varianza. Sin embargo, de existir alguna estructura de dependencia significativa entre ellos, podemos utilizar un modelo estacionario más complejo para dar cuenta de dicha estructura de dependencia.

Capítulo 3

Modelos autoregresivos y de promedios móviles

Los modelos autoregresivos y de promedios móviles (ARMA) son modelos lineales y causales utilizados para describir procesos débilmente estacionarios a través de valores pasados del mismo y de otras variables independientes.

Definición 3.0.1. Proceso lineal.

Se dice que un proceso $\{X_t\}$ es lineal si puede escribirse de la siguiente forma:

$$X_t = \mu + \sum_{-\infty}^{\infty} C_r \epsilon_{t-r},$$

donde μ es la media común, $\{C_r\}$ es una secuencia de constantes fijas y $\{\epsilon_t\}$ es un proceso de ruido blanco.

Definición 3.0.2. Proceso causal.

Se dice que un proceso lineal $\{X_t\}$ es causal si $C_r = 0 \forall r < 0$. Esto es:

$$X_t = \mu + \sum_{r=0}^{\infty} C_r \epsilon_{t-r}$$

Por lo tanto, los procesos causales no dependen de valores futuros o no observados.

3.1. Modelos de promedios móviles

En los modelos de promedios móviles, el valor actual de un proceso $\{X_t\}$ es visto como una combinación lineal del valor actual y de los q valores pasados de un proceso de ruido blanco ϵ_t .

Definición 3.1.1. Proceso de promedios móviles.

Se dice que un proceso $\{X_t\}$ es un proceso de promedios móviles de orden q , $MA(q)$, si satisface:

$$X_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.1)$$

donde $\{\varepsilon_t\}$ es un proceso de ruido blanco.

Como se mencionó en el capítulo 2, la media y la función de autocovarianza nos permiten caracterizar a un proceso débilmente estacionario y, afortunadamente, la estructura de la media y de la función de autocovarianza de los procesos MA es muy simple.

Proposición 3.1.1. Características de los modelos MA.

Sea $\{X_t\}$ un proceso que satisface la ecuación (3.1), entonces $\{X_t\}$ cumple:

(I) $E[X_t] = 0$.

(II) $var[X_t] = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$.

(III)

$$Cov(X_t, X_{t+k}) = \begin{cases} 0, & |k| > q \\ \sigma^2 \sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|}, & |k| \leq q. \end{cases}$$

Demostración.

(I) Como $\{\varepsilon_t\}$ es un proceso de ruido blanco, se tiene que:

$$\begin{aligned} E[X_t] &= E[\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t] \\ &= \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}] + \cdots + \theta_q E[\varepsilon_{t-q}] + E[\varepsilon_t] \\ &= 0. \end{aligned}$$

(II) Usando las propiedades de la varianza y que $\{\varepsilon_t\}$ es un proceso de ruido blanco:

$$\begin{aligned} Var[X_t] &= Var[\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t] \\ &= \theta_1^2 Var[\varepsilon_{t-1}] + \theta_2^2 Var[\varepsilon_{t-2}] + \cdots + \theta_q^2 Var[\varepsilon_{t-q}] + Var[\varepsilon_t] \\ &= \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2 + \cdots + \theta_q^2 \sigma^2 + \sigma^2 \\ &= (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma^2. \end{aligned}$$

(III) Ya que $E[X_t] = 0$ y $\{\varepsilon_t\}$ es un proceso de ruido blanco:

$$\begin{aligned} Cov(X_t, X_{t+k}) &= E[X_t X_{t+k}] E[X_t X_{t+k}] \\ &= E[X_t X_{t+k}] E[X_t] E[X_{t+k}] \\ &= E[\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t] E[\theta_1 \varepsilon_{t+k-1} + \cdots + \theta_q \varepsilon_{t+k-q} + \varepsilon_{t+k}] \\ &= \sigma^2 \sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|}. \end{aligned}$$

□

Observación 3.1.1. Sea $\{X_t\}$ un proceso que satisface la ecuación (3.1), entonces, por la proposición (3.1.1) se tiene:

- (I) $\{X_t\}$ es débilmente estacionario.
 (II) La función de autocorrelación $\rho(k)$ de $\{X_t\}$ es:

$$\rho(k) = \begin{cases} \sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|} / \sum_{i=0}^q \theta_i^2, & |k| \leq q, \quad k \neq 0 \\ 1, & k = 0 \\ 0, & \text{en otro caso.} \end{cases}$$

- (III) La función de autocorrelación de $\{X_t\}$ se vuelve cero después del rezago q . Gracias a ello, podemos identificar a un proceso MA(q) mediante su función de autocorrelación $\rho(k)$.

Es necesario que un proceso MA(q) cumpla con ciertas condiciones para tener asociada una única función de autocorrelación $\rho(k)$. Para clarificar ideas, consideremos el siguiente ejemplo propuesto en [4] :

Sean Y_t y Z_t dos procesos MA(1) a saber:

$$Y_t = -\theta_1 \varepsilon_{t-1} + \varepsilon_t \quad \text{y} \quad Z_t = -\frac{1}{\theta_1} \varepsilon_{t-1} + \varepsilon_t$$

Por la proposición (3.1.1) sabemos que la función de autocorrelación $\rho(k)$ de un proceso MA(1) tal que $X_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t$ es:

$$\rho(k) = \begin{cases} \theta_1 / (1 + \theta_1^2), & k = 1 \\ 1, & k = 0 \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto, las funciones de autocorrelación $\rho_Y(t)$ y $\rho_Z(t)$ de $\{Y_t\}$ y $\{Z_t\}$, respectivamente, son iguales:

$$\begin{aligned} \rho_Y(k) &= \begin{cases} -\theta_1 / (1 + \theta_1^2), & k = 1 \\ 1, & k = 0 \\ 0, & \text{eoc.} \end{cases} \quad \begin{array}{l} \text{Dividiendo el numerador y} \\ \text{el denominador entre } \theta_1^2 \end{array} \\ &= \begin{cases} -\frac{1}{\theta_1} / \left(1 + \frac{1}{\theta_1^2}\right), & k = 1 \\ 1, & k = 0 \\ 0, & \text{eoc.} \end{cases} \\ &= \rho_Z(k) \end{aligned}$$

La pregunta ahora es, si utilizamos $\rho(k)$ para identificar a un modelo MA(1) pero tenemos dos modelos MA(1) con la misma función de autocorrelación ¿cómo podemos saber con qué modelo MA(1) es preferible trabajar? Para contestar esta pregunta, observemos que podemos escribir al término de ruido blanco $\{\varepsilon_t\}$ en términos de $\{Y_t\}$ y de $\{Z_t\}$ de la siguiente forma:

$$\begin{aligned}\varepsilon_t &= Y_t + \theta_1 \varepsilon_{t-1} \\ &= Y_t + \theta_1(Y_{t-1} + \theta_1 \varepsilon_{t-2}) \\ &= Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \dots\end{aligned}\tag{3.2}$$

y

$$\begin{aligned}\varepsilon_t &= Z_t + \frac{1}{\theta_1} \varepsilon_{t-1} \\ &= Z_t + \frac{1}{\theta_1} (Z_{t-1} + \frac{1}{\theta_1} \varepsilon_{t-2}) \\ &= Z_t + \frac{1}{\theta_1} Z_{t-1} + \frac{1}{\theta_1^2} Z_{t-2} + \dots\end{aligned}\tag{3.3}$$

Si $|\theta_1| < 1$, la ecuación (3.2) converge y la ecuación (3.3) diverge. Como es preferible trabajar con una serie convergente, en este caso se prefiere a $\{Y_t\}$ sobre $\{Z_t\}$ y se dice que $\{Y_t\}$ es invertible y $\{Z_t\}$ no.

En general, las condiciones de identificabilidad de un proceso MA(q) se encuentran expresadas en función del polinomio asociado al mismo.

Definición 3.1.2. Polinomio asociado a un proceso MA(q)

El polinomio asociado a un proceso de promedios móviles es:

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

Obsérvese que el polinomio asociado a un proceso MA(q) nos permite escribir a este tipo de procesos de una forma más compacta:

$$X_t = \theta(B)\varepsilon_t \quad \text{donde } B\varepsilon_t = \varepsilon_{t-1}$$

Teorema 3.1.1. Sean X_t un proceso MA(q) y $\theta(z)$ su polinomio asociado. Entonces X_t tiene asociado de forma única una función de autocorrelación si las raíces de $\theta(z)$ se encuentran fuera del círculo unitario.

Demostración. El caso MA(1) ilustra la idea. □

3.2. Modelos autoregresivos

Los modelos autoregresivos (AR) están basados en la idea de que el valor actual de un proceso X_t puede ser visto como una función lineal de sus p valores anteriores $X_{t-1}, X_{t-2}, \dots, X_{t-p}$.

Definición 3.2.1. Procesos autoregresivos.

Se dice que un proceso $\{X_t\}$ es autoregresivo si $\{X_t\}$ es débilmente estacionario y satisface:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \quad \text{donde } \varepsilon_t \sim WN(0, \sigma^2) \quad (3.4)$$

Definición 3.2.2. Procesos autoregresivos con media μ .

Se dice que $\{X_t\}$ es un proceso AR(p) con media μ si $\{X_t - \mu\}$ es un proceso AR(p). Esto es:

$$\begin{aligned} X_t - \mu &= \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \varepsilon_t \\ &\quad \text{ó} \\ X_t &= \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \end{aligned}$$

donde $\varepsilon_t \sim WN(0, \sigma^2)$ y $\phi_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$,

A diferencia de los procesos de promedios móviles, un proceso autoregresivo no siempre es débilmente estacionario por lo que es pertinente preguntarse por las condiciones con las que dicho proceso debe cumplir para ser débilmente estacionario. Para fijar ideas se puede considerar el siguiente proceso AR(1) de acuerdo con [4]:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t. \quad (3.5)$$

Reemplazando recursivamente el valor de X_{t-1} en (3.5) podemos reescribir al proceso de la siguiente manera:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 X_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\quad \vdots \\ &= \phi_1^k X_{t-k} + \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{t-j} \quad \text{donde } \alpha_0 = 1. \end{aligned}$$

Si continuamos reemplazando recursivamente los valores de X_{t-k} y si se cumple que $\{X_t\}$ es débilmente estacionario y que $|\phi_1| < 1$, entonces, podemos reescribir a un proceso AR(1) de la siguiente forma:

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \quad \text{donde } \phi_0 = 1. \quad (3.6)$$

Entonces, la media y la función de autocovarianza de un proceso AR(1) que cumpla con la ecuación (3.6) son de la siguiente forma:

$$(I) \quad E[X_t] = E \left[\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \right] = \sum_{j=0}^{\infty} \phi_1^j E[\varepsilon_{t-j}] = 0.$$

(II)

$$\begin{aligned}
\gamma(h) &= \text{cov}(X_t, X_{t+h}) \\
&= E[X_t X_{t+h}] \\
&= E \left[\sum_{k=0}^{\infty} \phi^k \varepsilon_{t-k} \sum_{j=0}^{\infty} \phi^j \varepsilon_{t+h-j} \right] \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi^{j+h} \phi^j \quad \text{ya que } \text{cov}(\varepsilon_j, \varepsilon_k) = 0 \quad \forall k \neq j \\
&= \sigma_\varepsilon^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} \quad \text{como } |\phi| < 1 \text{ se tiene que} \\
&= \frac{\sigma_\varepsilon^2 \phi^h}{1 - \phi^2} \quad h \geq 0.
\end{aligned}$$

Resulta natural preguntarse qué pasa con un proceso AR(1) tal que $|\phi| > 1$. Volvamos a considerar a $\{X_t\}$, el proceso AR(1) que cumple con la ecuación (3.5), y supongamos ahora que $|\phi| > 1$. Claramente no podemos reescribir a este proceso de la forma presentada en la ecuación (3.6) pero sí podemos reescribir a (3.5) en términos de X_{t+1} de la siguiente forma:

$$\begin{aligned}
X_{t+1} &= \phi X_t + \varepsilon_{t+1} \\
\Rightarrow X_t &= -\frac{1}{\phi} \varepsilon_{t+1} + \frac{1}{\phi} X_{t+1}
\end{aligned} \tag{3.7}$$

$$\Rightarrow X_{t+1} = -\frac{1}{\phi} \varepsilon_{t+2} + \frac{1}{\phi} X_{t+2} \tag{3.8}$$

Sustituyendo (3.8) en (3.7) e iterando sobre t se tiene:

$$\begin{aligned}
X_t &= -\frac{1}{\phi} \varepsilon_{t+1} + \frac{1}{\phi} \left(-\frac{1}{\phi} \varepsilon_{t+2} + \frac{1}{\phi} X_{t+2} \right) \\
&= -\frac{1}{\phi} \varepsilon_{t+1} - \frac{1}{\phi^2} \varepsilon_{t+2} + \frac{1}{\phi^2} X_{t+2} \\
&= -\frac{1}{\phi} \varepsilon_{t+1} - \frac{1}{\phi^2} \varepsilon_{t+2} + \frac{1}{\phi^2} \left(-\frac{1}{\phi} \varepsilon_{t+3} + \frac{1}{\phi} X_{t+3} \right) \\
&= \\
&\vdots \\
&= -\frac{1}{\phi} \varepsilon_{t+1} - \frac{1}{\phi^2} \varepsilon_{t+2} - \frac{1}{\phi^3} \varepsilon_{t+3} - \cdots - \frac{1}{\phi^{k+1}} \varepsilon_{t+k+1} + \frac{1}{\phi^{k+1}} X_{t+k+1}
\end{aligned}$$

Como $|\phi| > 1$ se tiene que:

$$X_t = - \sum_{j=1}^{\infty} \frac{1}{\phi^j} \varepsilon_{t+j} \quad (3.9)$$

La ecuación (3.9) es la solución estacionaria de (3.5) cuando $|\phi| > 1$. Sin embargo ésta depende de valores futuros de la serie y por lo tanto no es causal.

El ejemplo anterior ilustra la necesidad de imponer restricciones sobre los parámetros de los procesos AR(p) de tal forma que su solución estacionaria solamente dependa de valores pasados. Para ello es necesario definir al polinomio asociado a un proceso AR(p).

Definición 3.2.3. Polinomio asociado a un proceso AR(p). El polinomio asociado a un proceso autoregresivo de orden p es:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

Obsérvese que el polinomio asociado a un proceso autoregresivo nos permite escribir a los procesos AR(p) de una forma más compacta, a saber:

$$\phi(B)X_t = \varepsilon_t \quad \text{donde } BX_t = X_{t-1} \quad (3.10)$$

Teorema 3.2.1. Sean X_t un proceso AR(p) y $\phi(z)$ su polinomio asociado. Entonces X_t es un proceso causal si las raíces de $\phi(z)$ se encuentran fuera del círculo unitario.

Demostración. La demostración de dicho teorema puede consultarse en [4]. □

Una vez que tenemos las restricciones sobre los parámetros de los procesos AR(p), podemos preguntarnos por la estructura de sus funciones de autocovarianza y de autocorrelación. Desgraciadamente dicha estructura no es tan simple como la de los procesos MA(q) y obtenerlas directamente es muy complejo, sin embargo, es posible obtenerlas resolviendo un sistema de ecuaciones.

Proposición 3.2.1. Características de los modelos AR.

Sea X_t un proceso AR(p) estacionario y causal, entonces, es posible calcular su función de autocovarianza realizando los siguientes pasos:

Como X_t es un proceso AR(p):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

Entonces, multiplicando por X_{t-k} se tiene que

$$X_t X_{t-k} = \phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \cdots + \phi_p X_{t-p} X_{t-k} + \varepsilon_t X_{t-k}$$

Calculando la esperanza se tiene una expresión recursiva para la función de autocovarianza al retraso k

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \cdots + \phi_p \gamma(k-p). \quad (3.11)$$

Para obtener a la función de autocorrelación basta dividir a la ecuación (3.11) entre $\gamma(0)$

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p), \quad k > 0. \quad (3.12)$$

Al conjunto de ecuaciones dadas por (3.12) se les conoce como ecuaciones de Yule Walker, que en forma matricial pueden escribirse como:

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} \quad (3.13)$$

En la observación 3.1.1 se mencionó que el ACF de un proceso MA(q) es cero después del retraso q , lo cual permite identificar el orden de este tipo de procesos. Desgraciadamente, cuando se trata de procesos AR(p) el ACF no brinda mucha información acerca del orden de dependencia. Por lo tanto es necesario encontrar alguna función que se comporte con los procesos AR de la misma forma en la que el ACF se comporta con los procesos MA; dicha función es la función de autocorrelación parcial (PACF).

Definición 3.2.4. Función de autocorrelación parcial.

La función de autocorrelación parcial puede ser vista como la correlación entre los errores de las ecuaciones de predicción, esto es:

$$\phi_{kk} = \text{corr}(X_k - P(X_k|X_1, \dots, X_{k-1}), X_0 - P(X_0|X_1, \dots, X_{k-1}))$$

Donde $P(X_k|X_1, \dots, X_{k-1})$ es la mejor proyección lineal, en términos del error cuadrático medio, de X_k en $\{X_1, \dots, X_{k-1}\}$.

La función de autocorrelación parcial ϕ_{kk} está determinada de forma única por el sistema siguiente sistema de ecuaciones:

Sea X_t un proceso débilmente estacionario con función de autocovarianza γ_k tal que $\gamma \rightarrow 0$ cuando $k \rightarrow \infty$, y suponiendo que ϕ_{kj} $j = 1, \dots, k$, $k \geq 1$, son los coeficientes del siguiente sistema de ecuaciones:

$$\begin{pmatrix} \rho(0) & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & \rho(0) & \rho(1) & \cdots & \rho(k-2) \\ \rho(2) & \rho(1) & \rho(0) & \cdots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{pmatrix} \quad k \geq 1 \quad (3.14)$$

A modo de ejemplo, calculemos la autocorrelación y la autocorrelación parcial de un proceso AR(1).

Sea X_t un proceso AR(1) débilmente estacionario y causal, $X_t = \phi_1 X_{t-1} + \varepsilon_t$. Entonces, la función de autocorrelación de X_t es:

$$\begin{aligned} \rho(h) &= \frac{\gamma(h)}{\gamma(0)} \\ &= \frac{\sigma_\varepsilon^2 \phi_1^h}{1 - \phi_1^2} / \frac{\sigma_\varepsilon^2 \phi_1^0}{1 - \phi_1^2} \\ &= \phi_1^h \end{aligned} \quad (3.15)$$

Y la función de autocorrelación parcial de X_t está dada por:

$$\begin{aligned} \rho(0)\phi_{11} &= \rho(1) \quad \text{por la ecuación (3.14)} \\ \Rightarrow \phi_{11} &= \rho(1) \end{aligned}$$

Si bien la función de autocorrelación nos permite medir la relación lineal entre dos variables X_t y X_{t+h} , la función de autocorrelación parcial mide la correlación entre X_t y X_{t+h} quitando la influencia de las variables aleatorias $X_{t+1}, \dots, X_{t+h-1}$.

Si $\{X_t\}$ es un proceso AR(p) causal con media cero, entonces, la mejor proyección lineal de X_{n+1} en $\{X_1, \dots, X_n\}$, para $n \geq p$ es:

$$P(X_{n+1}|X_1, \dots, X_n) = \phi_1 X_n + \phi_2 X_{n-1} + \cdots + \phi_p X_{n-p+1}$$

Por lo tanto, si $k > p$ entonces, $\phi_{kk} = 0$ ya que $P(X_k|X_1, \dots, X_{k-1})$ depende solamente de $\{X_{k-p}, \dots, X_{k-1}\}$. Esto nos permite usar a la función de autocorrelación parcial para encontrar el orden p de un proceso autoregresivo, de forma análoga a como se utiliza a la función de autocorrelación para identificar el orden q de un proceso de promedios móviles.

3.3. Modelos autoregresivos de promedios móviles

En las dos secciones anteriores se presentaron los modelos MA y AR. Dichos modelos son atractivos pues nos permiten describir de manera simple al proceso subyacente a una serie de tiempo observada. Sin embargo, en ocasiones se necesitaría de un modelo MA o AR de muchos parámetros para capturar la estructura de una serie de tiempo. Afortunadamente, los modelos MA y AR pueden ser combinados en los modelos ARMA permitiendo ajustar a la serie observada un modelo más parsimonioso.

Definición 3.3.1. Procesos autoregresivos de promedios móviles.

Se dice que un proceso $\{X_t\}$ es un proceso ARMA(p, q) si $\{X_t\}$ es débilmente estacionario y satisface:

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad \text{para toda } t.$$

Donde $\varepsilon_t \sim WN(0, \sigma^2)$ y $\phi(B)$ y $\theta(B)$ son los polinomios asociados a un proceso autoregresivo y a un proceso de promedios móviles respectivamente, evaluados en B .

Definición 3.3.2. Procesos autoregresivos de promedios móviles con media μ .

Se dice que un proceso $\{X_t\}$ es un proceso ARMA(p, q) con media μ si $\{X_t - \mu\}$ es un proceso ARMA(p, q).

Al igual que para los procesos AR y MA existen algunas condiciones para que los procesos ARMA estén bien definidos. Dichas condiciones están dadas en el siguiente teorema.

Teorema 3.3.1. Sea $\{X_t\}$ un proceso ARMA(p, q), entonces $\{X_t\}$ está bien definido si:

1. Las raíces de $\theta(z)$ se encuentran fuera del círculo unitario (condición de identificabilidad sobre $\theta(z)$).
2. Las raíces de $\phi(z)$ se encuentran fuera del círculo unitario (condición de causalidad sobre $\phi(z)$).
3. $\theta(z)$ y $\phi(z)$ no tienen raíces comunes (condición que asegura que el proceso está expresado en su forma reducida sobre $\theta(z)$ y $\phi(z)$).

Las condiciones 1 y 2 son análogas a aquellas de identificabilidad y causalidad de los procesos MA y AR respectivamente. Para entender más claramente la condición de que el modelo se encuentre en su forma reducida, consideremos al siguiente proceso ARMA(2, 1).

Sea X_t un proceso ARMA(2,1) tal que:

$$X_t - \frac{5}{6}X_{t-1} + \frac{1}{6}X_{t-2} = \varepsilon_t - \frac{1}{2}\varepsilon_{t-1}$$

Entonces, podemos reescribir a X_t de la siguiente forma:

$$X_t \left(1 - \frac{5}{6}B + \frac{1}{6}B^2 \right) = \varepsilon_t \left(1 - \frac{1}{2}B \right)$$

Por lo tanto, los polinomios autoregresivo y de promedios móviles de X_t , $\phi(z)$ y $\theta(z)$ respectivamente, son:

$$\begin{aligned}\phi(z) &= 1 - \frac{5}{6}z + \frac{1}{6}z^2 = \left(1 - \frac{1}{2}z\right) \left(1 - \frac{1}{3}z\right), & \phi(z) = 0 \text{ ssi } z = 2 \text{ ó } z = 3. \\ \theta(z) &= 1 - \frac{1}{2}z, & \theta(z) = 0 \text{ ssi } z = 2.\end{aligned}$$

Nótese que $\phi(z)$ y $\theta(z)$ tienen una raíz en común y que podemos reescribir al proceso $\{X_t\}$ de la siguiente forma:

$$\begin{aligned}X_t \left(1 - \frac{1}{2}B\right) \left(1 - \frac{1}{3}B\right) &= \varepsilon_t \left(1 - \frac{1}{2}B\right) \\ \Rightarrow X_t \left(1 - \frac{1}{3}B\right) &= \varepsilon_t\end{aligned}$$

$\therefore \{X_t\}$ es un proceso AR(1) aunque inicialmente había sido expresado como un proceso ARMA(2, 1).

3.4. Modelos autoregresivos integrados de promedios móviles

Los modelos Autoregresivos Integrados de Promedios Móviles (ARIMA) son una generalización de los modelos ARMA.

Definición 3.4.1. Proceso ARIMA.

Se dice que $\{X_t\}$ es un proceso ARIMA(p, d, q) si el proceso diferenciado $(1 - B)^d X_t$ es un proceso ARMA(p, q). Los procesos ARIMA pueden ser expresados de la siguiente manera:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (3.16)$$

Notemos que si $W_t = (1 - B)^d X_t$ entonces podemos reescribir a (3.16) de la siguiente forma:

$$\begin{aligned}\phi(B)W_t &= \theta(B)\varepsilon_t \\ \Rightarrow W_t &= \nabla^d X_t \text{ es un proceso ARMA}(p, q).\end{aligned}$$

3.5. Estimación de los parámetros

Para ajustar un modelo ARMA(p, q) a una serie de tiempo $\{x_1, \dots, x_n\}$ es necesario elegir los órdenes p y q del modelo y estimar los parámetros $\phi = (\phi_1, \dots, \phi_p)$, $\theta = (\theta_1, \dots, \theta_q)$, μ y σ^2 .

Dado que los órdenes p y q del modelo pueden ser (aproximadamente) determinados mediante las funciones de autocorrelación y de autocorrelación parcial, por ahora supondremos que los órdenes p y q son conocidos. Se supondrá además que la media ha sido removida de los datos, es decir, dado

que la media de un proceso Y_t puede ser estimada de la siguiente forma: $\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$, se trabajará con la serie $x_t = y_t - \bar{y}_t$ y se ajustará un modelo ARMA con media cero a la serie $\{x_t\}$.

Cuando p y q son conocidos y bajo el supuesto de que $\varepsilon_t \sim N(0, \sigma^2) \forall t$, los parámetros ϕ y θ pueden ser obtenidos utilizando el método de máxima verosimilitud. Como la función de verosimilitud de los procesos ARMA no puede ser maximizada analíticamente, ésta es maximizada numéricamente buscando al máximo dentro de la superficie de verosimilitud. Cuando se tienen valores iniciales de los parámetros a estimar, dicha búsqueda es más eficiente. En las siguientes secciones se presentan algunos métodos de estimación que permiten encontrar dichos valores iniciales y del método de máxima verosimilitud.

3.5.1. Estimadores de momentos

El método de momentos consiste en igualar los momentos poblacionales con los muestrales. Podemos encontrar los estimadores de momentos de ϕ a través de las ecuaciones de Yule-Walker sustituyendo a $\rho(h)$ por su estimador $\hat{\rho}(h)$. A los estimadores obtenidos mediante este método se les conoce como estimadores de Yule-Walker.

Procesos AR

Sea X_t un proceso AR(p). Entonces, por la ecuación (3.13) los estimadores de Yule-Walker para ϕ y σ^2 pueden calcularse de la siguiente forma:

$$\begin{aligned}\hat{\phi} &= \hat{\mathbf{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \\ \hat{\sigma}^2 &= \hat{\gamma}(0)[1 - \hat{\boldsymbol{\rho}}_p' \hat{\mathbf{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p]\end{aligned}$$

donde $\hat{\mathbf{R}}_p = [\hat{\rho}(i-j)]_{i,j=1}^p$ y $\hat{\boldsymbol{\rho}}_p = (\hat{\rho}(1), \dots, \hat{\rho}(p))'$.

Cuando la muestra es grande, los estimadores de Yule-Walker de los procesos AR(p) siguen aproximadamente una distribución normal y siguen aproximadamente la misma distribución que los estimadores máximo verosímiles.

Teorema 3.5.1. Distribución asintótica de los estimadores de Yule Walker para los modelos AR
Sea X_t un proceso AR(p) causal. Entonces, sus estimadores de Yule-Walker tienen la siguiente distribución asintótica:

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} \mathbf{N}(0, \sigma^2 \Gamma_p^{-1}) \quad \text{cuando } n \rightarrow \infty \quad (3.17)$$

donde $\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^p$

El teorema anterior nos permite encontrar intervalos de confianza asintóticos para ϕ .

Procesos MA y ARMA

Las ecuaciones de Yule-Walker de los procesos ARMA(p, q) están dadas por:

$$\hat{\gamma}(k) - \phi_1 \hat{\gamma}(k-1) - \cdots - \phi_p \hat{\gamma}(k-p) = \sigma^2 \sum_{j=k}^q \theta_j \psi_{j-k} \quad 0 \leq k \leq p+q \quad (3.18)$$

donde ψ_j está definida en términos de θ y ϕ de la siguiente forma: $\psi(z) = \theta(z)/\phi z$ y $\theta_0 := 1$.

Los estimadores de Yule-Walker para los parámetros en los modelos MA(q) y ARMA(p, q) no son lineales en los parámetros, por lo que la solución del sistema 3.18 puede no existir o no ser única. Para ilustrar lo anterior, consideremos el siguiente proceso MA(1): $Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$. Por el sistema (3.18) sabemos que:

$$\begin{aligned} \hat{\gamma}(0) &= \hat{\sigma}^2(1 + \hat{\theta}_1^2), \\ \hat{\rho}(1) &= \frac{-\hat{\theta}_1}{1 + \hat{\theta}_1^2} \\ \text{por lo tanto} \\ \hat{\sigma}^2 &= \hat{\gamma}(0)/(1 + \hat{\theta}_1^2), \\ &\text{y} \\ 0 &= \hat{\rho}(1)\hat{\theta}_1^2 + \hat{\theta}_1 + \hat{\rho}(1) \quad \text{si y sólo si} \\ \hat{\theta}_1 &= \frac{-1 \pm \sqrt{1 - 4\hat{\rho}(1)^2}}{2\hat{\rho}(1)} \end{aligned}$$

Luego, el estimador $\hat{\theta}_1$ es no lineal. En general, para los modelos MA(q) y ARMA(p, q), es muy complicado obtener estimadores de los parámetros θ_i como funciones de $\rho(i)$ analíticamente.

3.5.2. Estimadores por mínimos cuadrados

Sea X_t un proceso ARMA(p, q). Entonces, los parámetros de dicho proceso pueden ser estimados mediante la resolución numérica del siguiente problema de minimización: Como X_t es un proceso ARMA(p, q) podemos escribir al error al tiempo t de la siguiente forma:

$$\varepsilon_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

para encontrar los parámetros (ϕ, θ) que minimicen:

$$S_*(\phi, \theta) = \sum_{t=p+1}^n \varepsilon_t^2$$

Donde los valores iniciales $X_0 = X_{-1} = \cdots = X_{1-p} = \varepsilon_0 = \cdots = \varepsilon_{1-q} = 0$. Dado que los valores iniciales están condicionados a un valor, a este método se le conoce como de mínimos cuadrados condicionales.

3.5.3. Estimadores máximo verosímiles

La estimación por máxima verosimilitud tiene como objetivo encontrar los parámetros que mejor correspondan a los datos observados mediante la maximización de la función de verosimilitud.

Sea X_t un proceso ARMA(p, q), esto es:

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad \text{para toda } t, \text{ donde } \varepsilon_t \sim WN(0, \sigma^2).$$

supongamos además que $\varepsilon_t \sim N(0, \sigma^2) \forall t$. Usando este supuesto, se encuentran los estimadores máximo verosímiles $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q, \hat{\sigma}^2)$. Dichos estimadores son aquellos que maximizan a la función de verosimilitud $L(\phi, \theta, \sigma^2)$.

A continuación se mostrará el desarrollo de [4] para obtener dicha expresión cuando se trata de un proceso AR(1).

Sea $X_t, t \in \{1, \dots, n\}$, un proceso AR(1), $X_t = \phi X_{t-1} + \varepsilon_t$ donde $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d. Entonces, la función de densidad conjunta de $(\varepsilon_2, \dots, \varepsilon_n)$ está dada por:

$$f(\varepsilon_2, \dots, \varepsilon_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n-1)/2} \exp\left(\frac{-1}{2\sigma^2} \sum_{t=2}^n \varepsilon_t^2\right)$$

Como $X_2 = \phi X_1 + \varepsilon_2, \dots, X_n = \phi X_{n-1} + \varepsilon_n$, usando el método de transformación de variables y dado que el determinante del Jacobiano de dicha transformación es 1, se tiene que la función de distribución de X_2, \dots, X_n condicionada a X_1 es:

$$f(X_2, \dots, X_n | X_1) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n-1)/2} \exp\left[\frac{-1}{2\sigma^2} \sum_{t=2}^n (X_t - \phi X_{t-1})^2\right]$$

Si $\varepsilon_0, \varepsilon_{-1}, \varepsilon_{-2}, \dots$ son i.i.d, se tiene que $X_1 \sim N(0, \sigma^2/(1 - \phi^2))$, entonces,

$$f(X_1) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n-1)/2} \sqrt{1 - \phi^2} \exp\left[\frac{-1}{2\sigma^2}(1 - \phi^2)X_1^2\right]$$

Luego, la función de verosimilitud de X_1, \dots, X_n es

$$\begin{aligned} L(\phi, \sigma^2) &= f(X_1, \dots, X_n) \\ &= f(X_2, \dots, X_n | X_1) f(X_1) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \sqrt{1 - \phi^2} \exp\left\{\frac{-1}{2\sigma^2} \left[\sum_{t=2}^n (X_t - \phi X_{t-1})^2 + (1 - \phi^2)X_1^2 \right]\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \sqrt{1 - \phi^2} \exp\left[\frac{-1}{2\sigma^2} S(\phi)\right], \end{aligned}$$

donde $S(\phi) = \sum_{t=2}^n (X_t - \phi X_{t-1})^2 + (1 - \phi^2)X_1^2$. Entonces, la función de log-verosimilitud $\lambda(\phi, \sigma^2)$ de un proceso AR(1) es:

$$\begin{aligned}\lambda(\phi, \sigma^2) &= \log(L(\phi, \sigma^2)) \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 + \frac{1}{2}\log(1 - \phi^2) - \frac{1}{2\sigma^2}S(\phi)\end{aligned}$$

Obtener la expresión de la función de verosimilitud de los procesos ARMA(p, q) es complejo, en [2] puede consultarse cómo llegar a dicha expresión.

Como se mencionó anteriormente, la función de verosimilitud de los procesos ARMA(p, q) no puede ser maximizada analíticamente por lo que se maximiza numéricamente. Si se parte de algún valor inicial de los parámetros a estimar, dicha maximización es más eficiente. En el software estadístico R, se puede estimar modelos ARMA mediante mínimos cuadrados condicionales, máxima verosimilitud o una combinación de ambos (encontrando valores iniciales para el método de máxima verosimilitud a través de los estimadores por mínimos cuadrados). En la presente tesis se utilizó la última opción.

3.6. Selección del orden del modelo

En la sección anterior se presentaron distintos métodos para estimar los parámetros de un proceso ARMA(p, q) a partir de las observaciones $\{x_1, \dots, x_n\}$ y para ello se supuso que los órdenes p y q eran conocidos. En la realidad, los órdenes p y q son desconocidos por lo que es necesario encontrar alguna manera de determinarlos, al menos tentativamente.

Sea $\{x_1, \dots, x_n\}$ una realización de un proceso ARMA(p, q). Una primera forma de determinar los órdenes p y q se tiene mediante la inspección del ACF y del PACF estimados a partir de las observaciones $\{x_1, \dots, x_n\}$, como se observa en el siguiente cuadro:

Cuadro 3.1: Comportamiento del ACF y del PACF de un proceso ARMA(p, q) causal e invertible.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Va disminuyendo	Es cero después del rezago q	Va disminuyendo
PACF	Es cero después del rezago p	Va disminuyendo	Va disminuyendo

Fuente: [7]

Entonces, se pueden graficar el ACF y el PACF muestrales junto con sus respectivas bandas de $\pm 1.96\sqrt{n}^{-1}$ y utilizar el cuadro anterior para determinar los órdenes del modelo a estimar mediante la inspección de dichas gráficas. Con ellos pueden ajustarse distintos modelos ARMA(p, q) a los datos.

¹Donde 1.96 corresponde al cuantil $1 - \frac{\alpha}{2}$, $\alpha = 0.05$, de una distribución Normal(0, 1).

Otra forma de elegir los órdenes p y q del modelo consiste en utilizar estadísticos de bondad de ajuste. Los estadísticos que se usan más comunmente son: el Criterio de Información de Akaike (AIC) el Criterio de Información de Akaike corregido (AICc) y el Criterio de Información Bayesiana. A continuación se presentan dichos criterios.

Criterio de Información de Akaike

El AIC está definido como:

$$AIC = -2\log L(\phi, \theta, \sigma^2) + 2(p + q)$$

Entre menor es el AIC, mayor es la verosimilitud del modelo y mejor el ajuste.

Criterio de Información de Akaike corregido

Se ha encontrado que el AIC tiende a sobreestimar p . Dado que no es deseable ajustar un modelo con muchos parámetros, existe una versión del AIC que corrige el ajuste de un modelo con muchos parámetros mediante un factor de penalización. A dicha versión se le conoce como AICc y está definida de la siguiente manera:

$$AICc = -2\log L(\phi, \theta, \sigma^2) + \frac{2(p + q + 1)n}{n - p - q - 2}$$

Donde n es el tamaño de la muestra. Análogamente al AIC, entre menor es el AICc, mayor es la verosimilitud del modelo y mejor el ajuste.

Criterio de Información Bayesiano

Al igual que el AICc, el BIC busca contrarrestar la tendencia del AIC a sobreestimar p . El BIC está definido de la siguiente forma:

$$BIC = (n - p - q)\ln\left[\frac{n\hat{\sigma}^2}{(n - p - q)}\right] + n(1 + \ln\sqrt{2\pi}) + (p + q)\ln\left[\left(\sum_{i=1}^n X_i^2 - n\hat{\sigma}^2\right)/(p + q)\right]$$

Al igual que en el caso del AIC y del AICc, el mejor modelo es el que tiene el BIC mínimo.

3.7. Verificación del modelo

Una vez que se ajustó un modelo a la serie de tiempo $\{x_1, \dots, x_n\}$ se puede realizar un análisis de los residuos $\hat{y}_t = x_t - \hat{x}_t$, ($t = 1, \dots, n$) para verificar el ajuste del mismo.

En primer lugar, hay que verificar que los residuos provengan de un proceso de ruido blanco. Para ello puede utilizarse la prueba de Portmanteau:

H_0 : Los residuos constituyen un ruido blanco. vs. H_1 : Los residuos no constituyen un ruido blanco.

La estadística de prueba es:

$$Q = n \sum_{i=1}^k \hat{\rho}^2(i), \quad Q \sim \chi_{(k-p-q)}^2 \text{ donde } p + q < k \leq n.$$

Se rechaza H_0 cuando $Q > \chi_{1-\alpha, (k-p-q)}^2$.

También puede realizarse la prueba de Ljung-Box. Dicha prueba es una modificación de la prueba de Portmanteau:

$$\tilde{Q} = n(n+2) \sum_{i=1}^k \frac{\hat{\rho}^2(i)}{n-k}, \quad \tilde{Q} \sim \tilde{\chi}_{(k-p-q)}^2 \text{ donde } p + q < k \leq n.$$

Se rechaza H_0 cuando $\tilde{Q} > \chi_{1-\alpha, (k-p-q)}^2$.

Además de las pruebas antes mencionadas se pueden realizar los siguientes análisis sobre los residuos:

- Analizar la gráfica de los residuos y observar si existe alguna evidencia de que los residuos no sean observaciones de un proceso de ruido blanco y si la varianza de los mismos no cambia a lo largo del tiempo.
- Graficar la función de autocorrelación muestral de los residuos junto con sus bandas de confianza de $\pm 1.96/\sqrt{n^2}$ para evaluar si esta es significativamente distinta de cero para algún rezago.

Si se utilizó el método de máxima verosimilitud para ajustar el modelo, se debe verificar además que los residuos sigan una distribución normal. Para ello se puede hacer lo siguiente:

- Graficar y analizar el histograma de los residuos.
- Analizar la gráfica cuantil-cuantil normal de los residuos.
- Realizar pruebas de normalidad a los residuos. En la presente tesis se utilizaron las pruebas de Anderson-Darling, la de Lilliefors y la de Shapiro-Wilk.

3.8. Construcción de modelos ARMA

El proceso para construir modelos ARMA a partir de una serie de tiempo $\{x_t\}$ ($t = 1, \dots, n$) puede ser resumido en los siguientes pasos:

1. Graficar $\{x_t\}$ y analizar si existen evidencias de no estacionariedad débil.

²Donde 1.96 es el cuantil $1 - \frac{\alpha}{2}$ de una distribución normal estándar con $\alpha = 0.05$.

2. En caso de que no se pueda suponer que la serie de tiempo $\{x_t\}$ provenga de un proceso débilmente estacionario, aplicar las transformaciones presentadas en el capítulo 3 para llevar a $\{x_t\}$ a la estacionariedad débil.
3. Analizar la gráfica de las funciones de autocorrelación y de autocorrelación parcial muestrales de la serie $\{x_t\}$ transformada para determinar los órdenes preliminares p y q del modelo a ajustar.
4. Utilizando los órdenes p y q seleccionados en el paso anterior, ajustar un modelo ARMA mediante alguno de los métodos presentados en la sección 2.4.
5. Verificar el ajuste del modelo.

El proceso de construcción de modelos ARMA es iterativo. Es posible que del paso 5 tenga que regresarse el paso 1 en múltiples ocasiones y que se tengan que comparar varios modelos antes de elegir un modelo satisfactorio para una serie de tiempo $\{x_t\}$.

3.9. Pronósticos

Una vez que se ajustó un modelo ARMA a la serie $\{x_t\}$, es posible utilizar dicho modelo para pronosticar valores futuros de la misma. Los detalles de como obtener dichas predicciones pueden consultarse en [1].

Supongamos que tenemos un proceso estacionario $\{X_1, \dots, X_n\}$ con media μ y función de autocovarianza γ , entonces, queremos encontrar la combinación lineal de $1, X_1, \dots, X_{n-1}, X_n$ que pronostique $X_{n+h}, h > 0$, con el menor error cuadrático medio. Dicha combinación lineal es denotada como $P_n X_{n+h}$ y es de la forma

$$P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1$$

Donde los coeficientes a_1, \dots, a_n son aquellos que minimizan a la función:

$$S(a_0, \dots, a_n) = E(X_{n+h} - a_0 - a_1 X_n - \dots - a_n X_1)^2$$

El error cuadrático medio de $P_n X_{n+h}$ es:

$$E[(X_{n+h} - P_n X_{n+h})^2] = \gamma(0) - \mathbf{a}'_n \gamma_n(h) \quad \text{donde}$$

$$\mathbf{a}'_n = (a_1, \dots, a_n) \text{ y } \gamma_n(h) = (\gamma(h), \dots, \gamma(h+n-1))'$$

De forma más general, supongamos ahora que Y y W_1, \dots, W_n son variables aleatorias con segundo momento finito y que $\mu = E[Y]$, $\mu_i = E[W_i]$, $Cov(Y, Y)$, $Cov(Y, W_i)$ y $Cov(W_i, W_j)$ son conocidas. Entonces, la mejor predicción lineal de Y en términos de $\{1, W_1, \dots, W_n\}$ es:

$$P(Y | \mathbf{W}) = \mu_y + \mathbf{a}'(\mathbf{W} - \mu_W),$$

donde $\mathbf{W} = \{W_1, \dots, W_n\}$ y $\mathbf{a} = (a_1, \dots, a_n)'$ es una solución de $\Gamma\mathbf{a} = \gamma$. Nótese que $P_n X_{n+h}$ es un caso particular de $P(X_{n+h}|\mathbf{W})$ en el que $\mathbf{W} = (X_n, X_{n-1}, \dots, X_1)'$.

El error cuadrático medio de la predicción es:

$$E[(Y - P(Y|\mathbf{W}))^2] = \text{Var}(Y) - \mathbf{a}'\gamma$$

Cuando se trata de procesos ARMA (p, q) , la predicción puede ser encontrada recursivamente utilizando una versión simplificada del algoritmo de inovaciones. Detalles del algoritmo de inovaciones y de su simplificación para los procesos ARMA pueden encontrarse en [1].

Capítulo 4

Análisis de las series de ventas de electricidad

En este capítulo se presentan los análisis realizados a las series de ventas mensuales de los sectores Servicios y Doméstico, de enero del 2002 a mayo del 2013. Para realizar dichos análisis se aplicó la teoría presentada en los dos capítulos anteriores.

4.1. Servicios

Las ventas al sector Servicios están constituidas principalmente por ventas de electricidad para el alumbrado público. El registro del consumo de este sector es estimado. La estimación se obtiene de la siguiente forma: se realiza un censo de lámparas, se multiplica el número de lámparas por el número de horas que en promedio están encendidas y por el número de Gigawatts-hora(GWh) que consume una lámpara. Cada cierto tiempo se realizan censos de luminarias para actualizar el conteo.

Al examinar la serie de ventas a dicho sector (figura 4.1) se observa una tendencia que aumenta a lo largo del tiempo más no se observan ni estacionalidad ni gran variabilidad en los datos. En abril del 2012 se observa un pico en las ventas, esto se debe a un cobro retroactivo derivado de un censo de luminarias. La serie de ventas mensuales de electricidad al sector Servicios de enero de 2002 a mayo del 2013 será denotada como $serv_t$, $t \in \{1, \dots, 137\}$.

Claramente $serv_t$ no es una serie débilmente estacionaria. En la figura 4.2 se observa la función de autocorrelación estimada de $serv_t$. Como la autocorrelación de orden uno es cercana .9 , se aplicó una diferencia al rezago uno, es decir:

$$\nabla serv_t = serv_t - serv_{t-1} \quad \text{donde } 1 < t \leq 137.$$

La gráfica de la serie diferenciada $\nabla serv_t$, figura 4.3, sugiere que dicha serie es débilmente estacionaria. En la gráfica del ACF de $\nabla serv_t$ que se presenta en la figura 4.4 puede observarse que la mayor autocorrelación de la serie se presente al rezago uno, sin embargo, puede suponerse que dicha

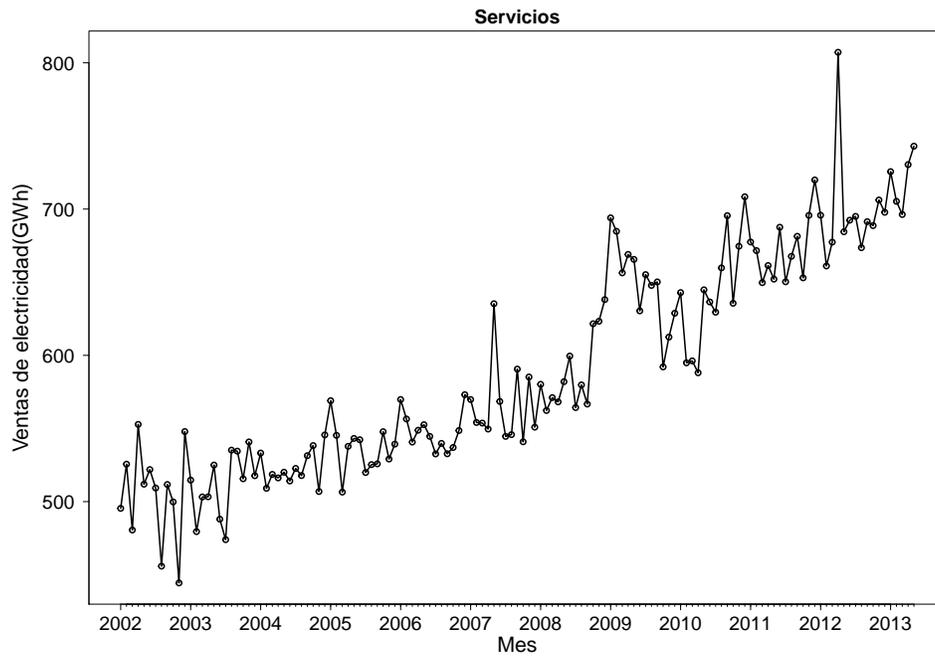


Figura 4.1: Ventas de electricidad al sector Servicios de enero del 2002 a mayo del 2013 en GWh.

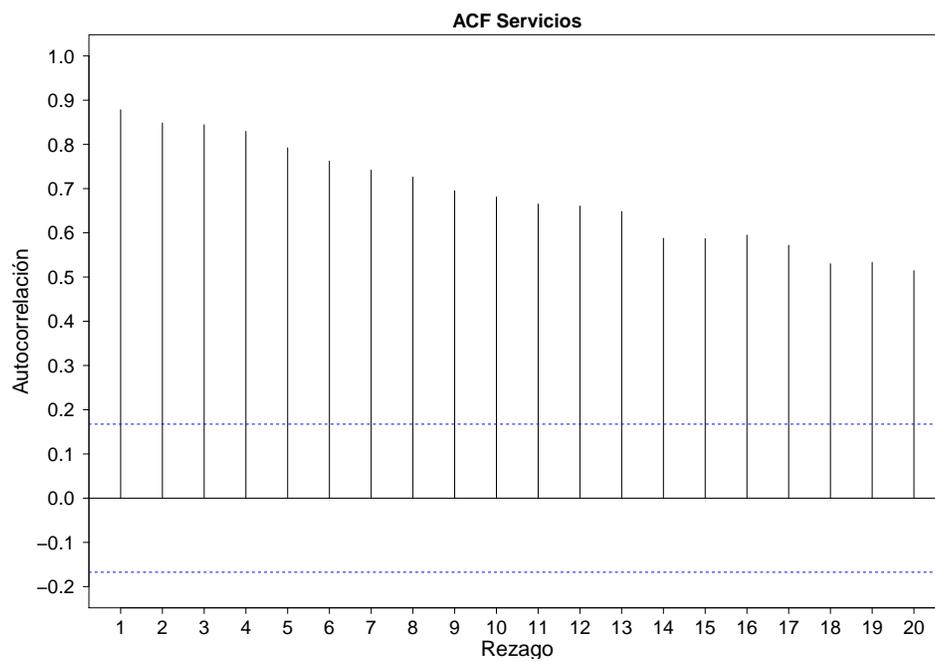


Figura 4.2: ACF de la serie de ventas de electricidad al sector Servicios.

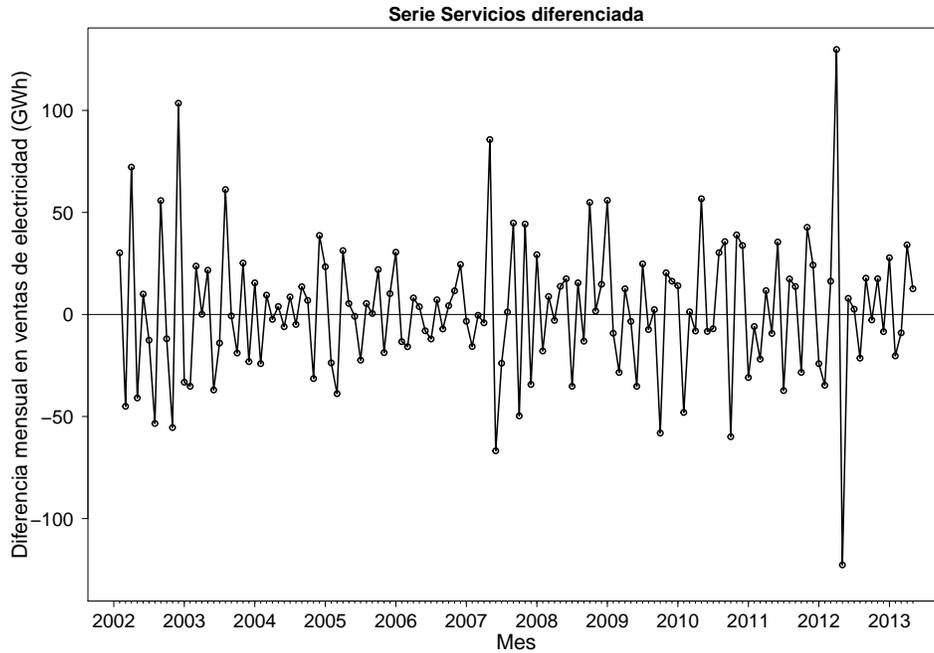


Figura 4.3: Grafica de la serie de ventas de electricidad al sector Servicios diferenciada ($\nabla serv_t$).

serie es débilmente estacionaria.

Las gráficas del ACF y del PACF de $\nabla serv_t$ sugieren ajustar, como primera propuesta, un modelo ARMA(3,1) a la serie diferenciada $\nabla serv_t$ o equivalentemente ajustar un modelo ARIMA(3,1,1) a la serie original $serv_t$.

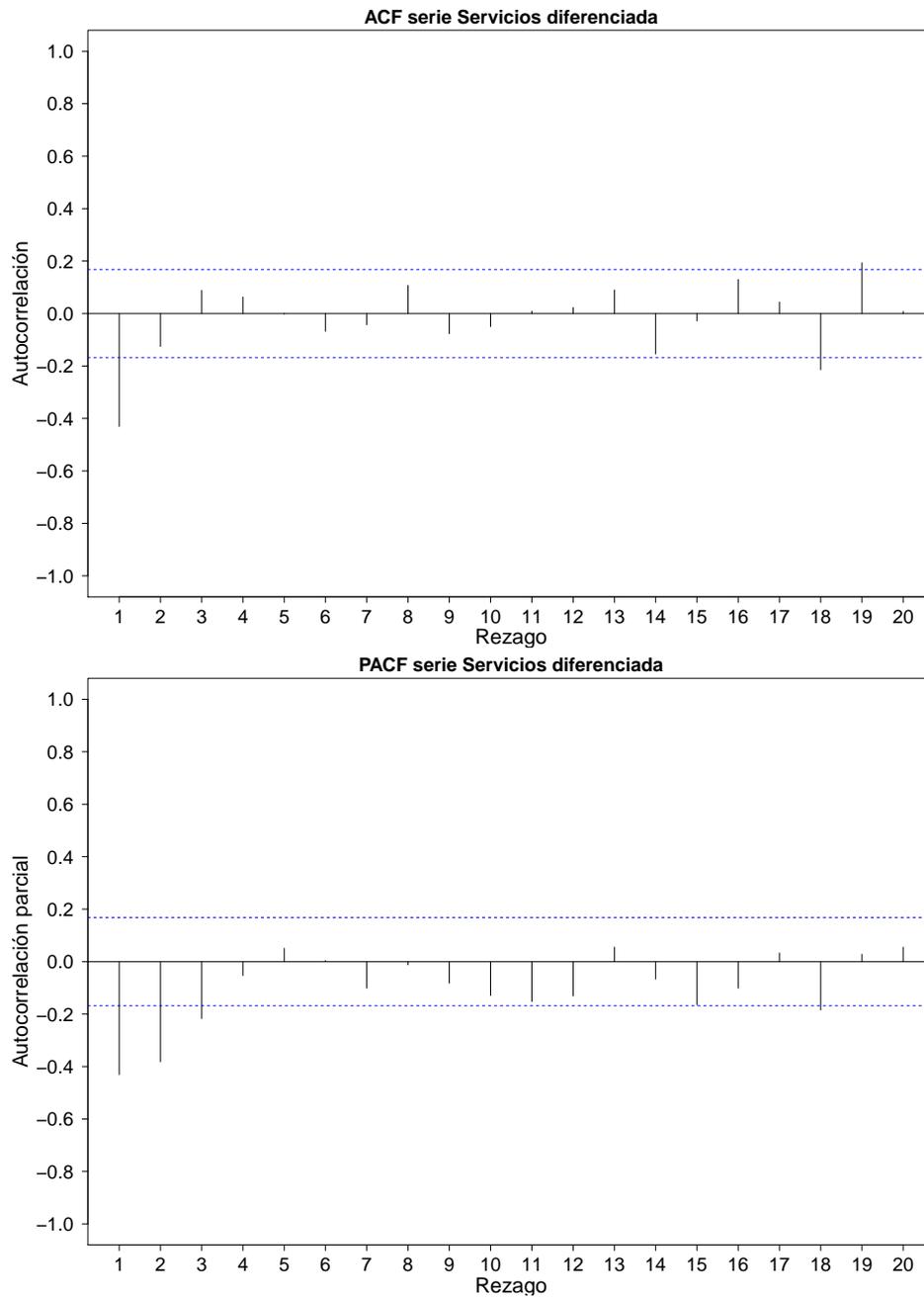
4.1.1. Primera propuesta

Sea $X_t = \nabla serv_t$. Se ajustó un modelo ARMA(3,1) a la serie X_t utilizando el método de máxima verosimilitud y utilizando como valores iniciales los estimadores por mínimos cuadrados. El modelo ajustado (Modelo 1) es :

$$\begin{aligned} X_t &= \mu(1 - \phi_1 - \phi_2 - \phi_3) + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= 3.4641 - 0.5271 X_{t-1} - 0.4243 X_{t-2} - 0.1669 X_{t-3} - 0.1671 \varepsilon_{t-1} + \varepsilon_t. \end{aligned} \quad (4.1)$$

Después de ajustar el modelo es necesario verificar que los supuestos del mismo se cumplan, esto es: verificar que los residuos del modelo constituyan un ruido blanco y que sigan una distribución normal.

En la figura 4.5 se observan los residuos del Modelo 1 y en la figura 4.6 la función de autocorrelación estimada de los mismos. En ninguna de las dos figuras se observan indicios de que los residuos estén correlacionados. Además se realizó una prueba de Ljung-Box (H_0 los residuos constituyen un

Figura 4.4: ACF y PACF de la serie $\nabla serv_t$.

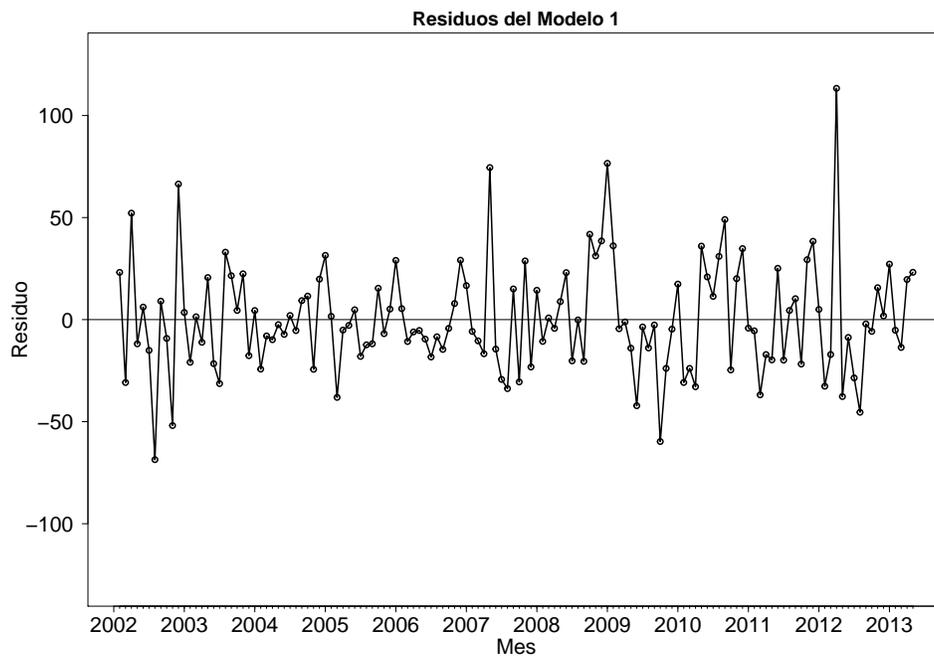


Figura 4.5: Residuos del primer modelo ajustado a la serie $\nabla serv_t$.

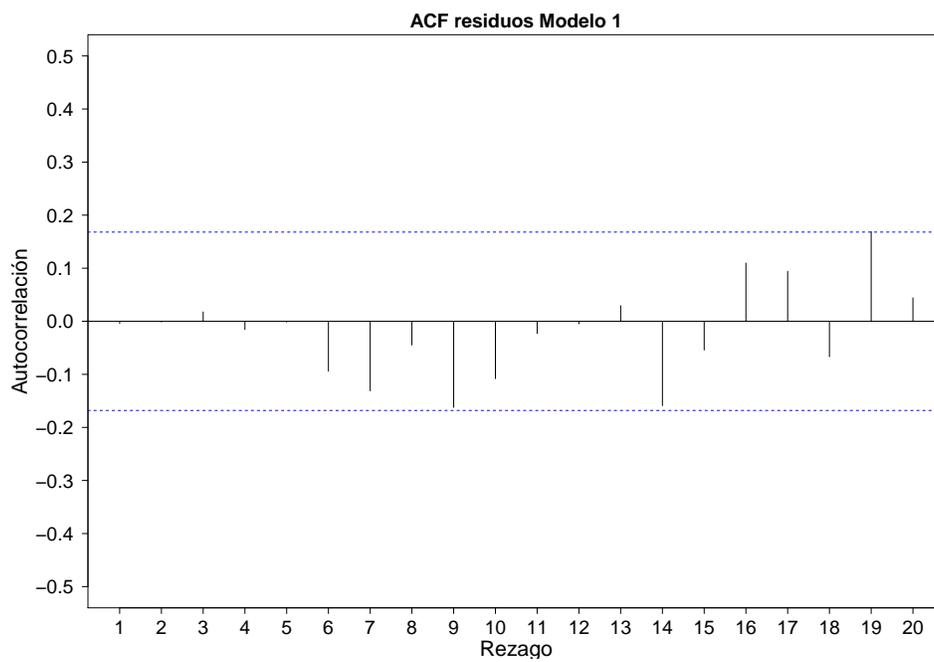


Figura 4.6: ACF de los residuos del primer modelo ajustado a la serie $\nabla serv_t$.

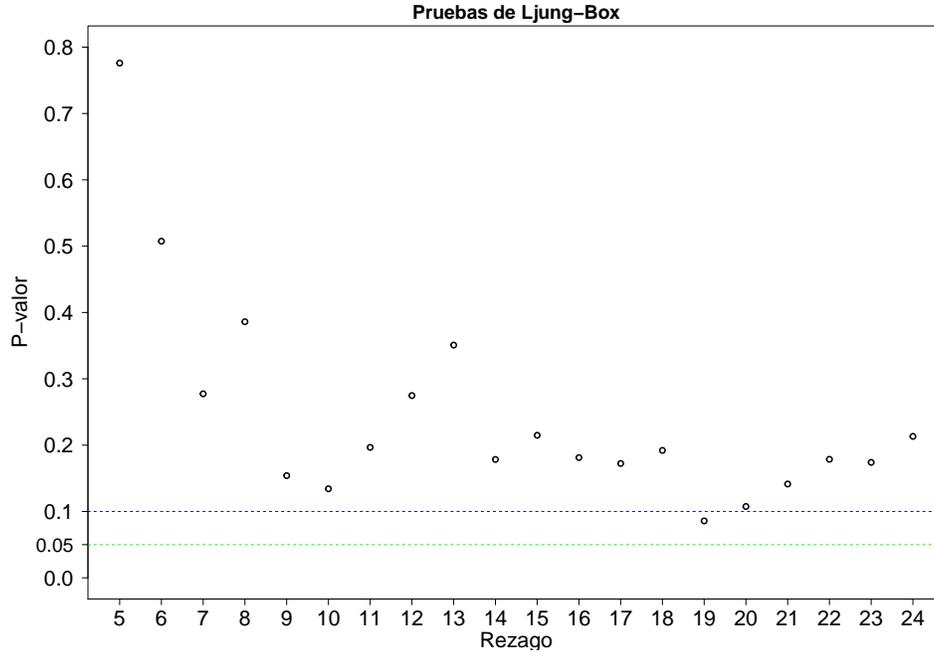


Figura 4.7: Prueba de Ljung-Box para los residuos del primer modelo ajustado a la serie $\nabla serv_t$.

ruido blanco v.s. H_1 los residuos no constituyen un ruido blanco) para los rezagos 5 a 24 en las que se obtuvieron los resultados que se muestran en la figura 4.7. En dicha figura se observa que los p-valores para todos los rezagos considerados son mayores a 0.05 y que solamente al rezago 19 el p-valor es menor a 0.1 por lo que para ningún rezago se rechazó la hipótesis nula y puede decirse que los residuos constituyen un ruido blanco.

Falta verificar que los residuos del modelo ajustado sigan una distribución normal. Para ello, primero se examinaron el histograma y la gráfica cuantil-cuantil de los residuos. En la figura 4.8 se encuentran ambas gráficas, en ellas se observa que los datos parecen seguir una distribución normal excepto por que la distribución de los residuos presenta una cola derecha pesada.

Se realizaron las tres pruebas de normalidad mencionadas en el capítulo anterior y arrojaron los resultados que se encuentran en el cuadro 4.1. En ellos se observan p-valores pequeños por lo que no se puede decir que los residuos siguen una distribución normal. Por lo tanto, se buscará ajustar otro modelo en el que el supuesto de normalidad se sostenga mejor.

Cuadro 4.1: Pruebas de normalidad a los residuos del primer modelo ajustado.

Prueba	Estadístico	P-valor
Anderson-Darling	1.1054	0.006541
Lilliefors	0.0841	0.01965
Shapiro-Wilk	0.96225	0.000819

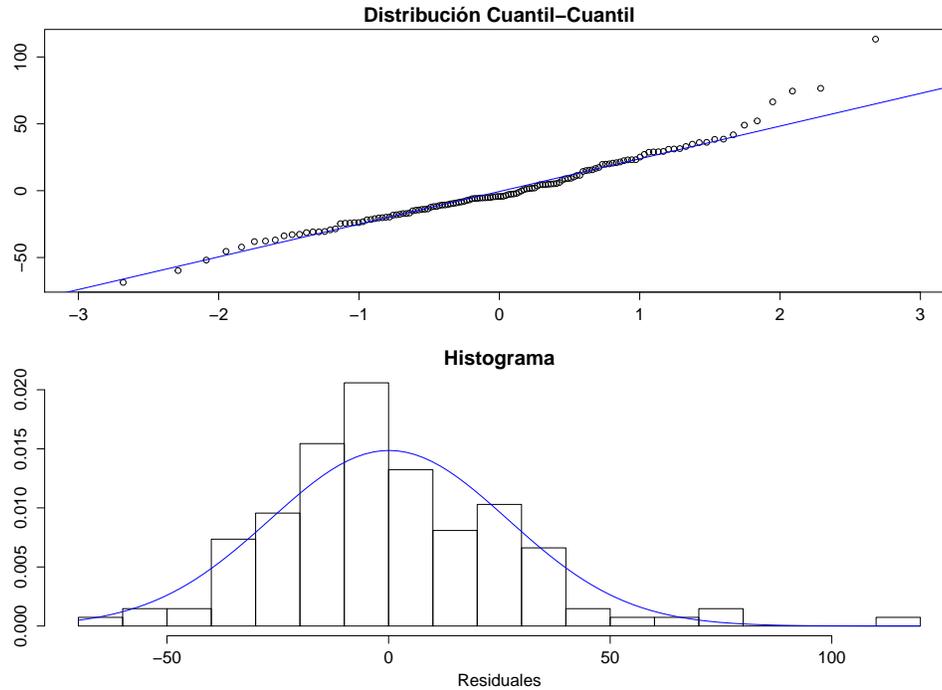


Figura 4.8: Gráfica cuantil-cuantil e histograma de los residuos del primer modelo ajustado a la serie $\nabla serv_t$.

4.1.2. Segunda propuesta

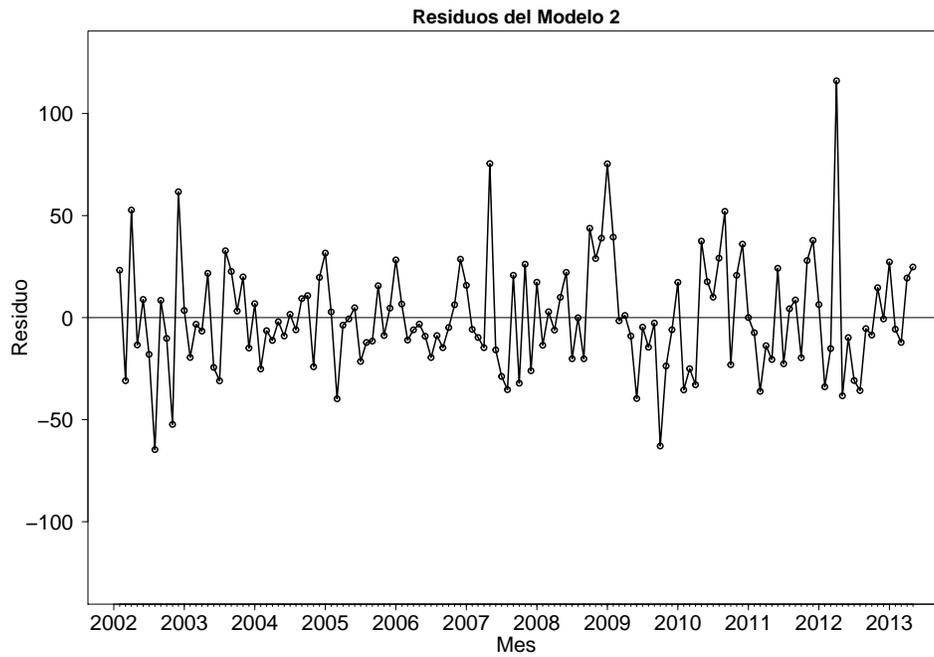
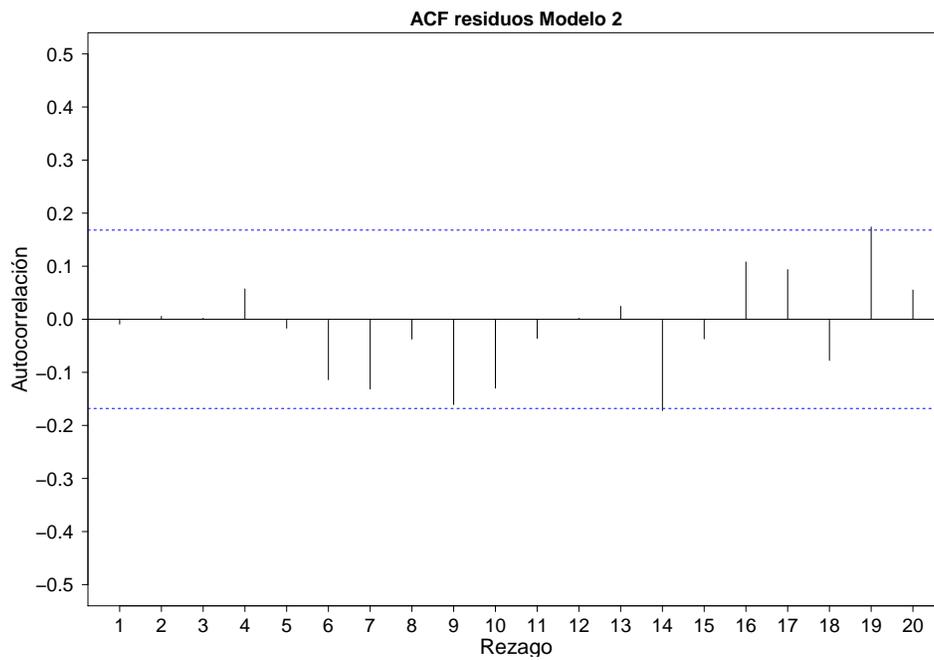
Al analizar el ACF y el PACF de la serie $\nabla serv_t$ por primera vez, figura 4.4, se propuso ajustar un modelo ARMA(3,1) a la misma. Sin embargo, al ajustar dicho modelo los supuestos de normalidad no se sostuvieron. Al realizar un segundo análisis del ACF y del PACF de la serie se observa que el valor del PACF al rezago 3 apenas y sale de las bandas de confianza, por lo que como segunda propuesta se ajustará un modelo ARMA(2,1) a la serie $\nabla serv_t$.

El modelo ajustado (Modelo 2) es:

$$\begin{aligned}
 X_t &= \mu(1 - \phi_1 - \phi_2) + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \\
 &= 2.3721 - 0.223X_{t-1} - 0.2303X_{t-2} - 0.466\varepsilon_{t-1} + \varepsilon_t.
 \end{aligned}
 \tag{4.2}$$

A continuación se verificarán los supuestos del modelo. En la figura 4.9 y en figura 4.10 se presentan las gráficas de los residuos y de la función de autocorrelación de los mismos, respectivamente. En ninguna de las dos se observan indicios de correlación entre los residuos.

Además, se realizaron pruebas de Ljung-Box para los rezagos 4 a 24 (figura 4.11) y en ellas los residuos parecen ser consistentes con la hipótesis nula por lo que, puede decirse que constituyen un ruido blanco.

Figura 4.9: Residuos del segundo modelo ajustado a la serie $\nabla serv_t$.Figura 4.10: ACF de los residuos del segundo modelo ajustado a la serie $\nabla serv_t$.

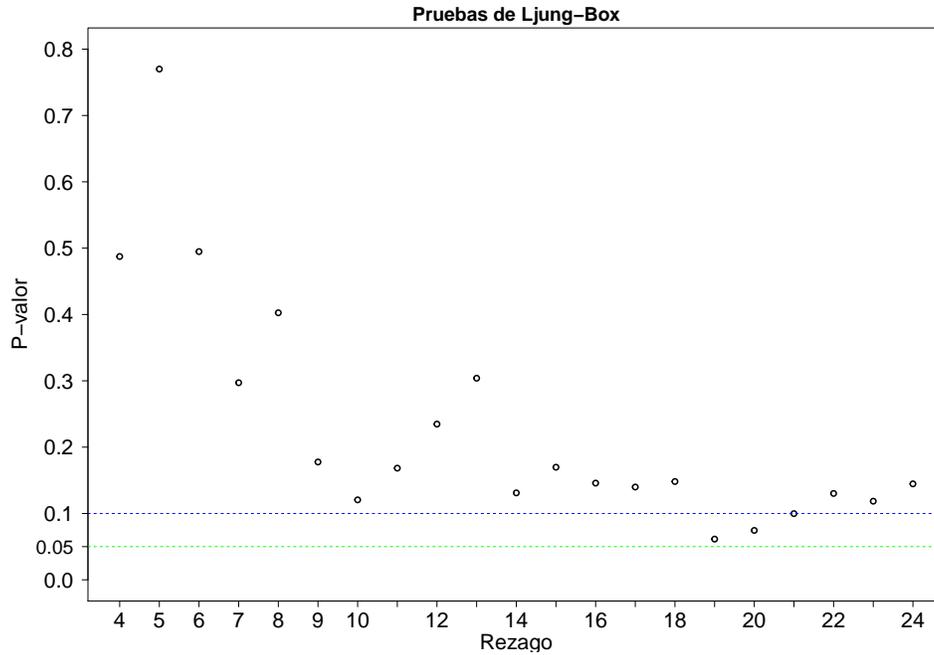


Figura 4.11: Prueba de Ljung-Box para los residuos del segundo modelo ajustado a la serie $\nabla serv_t$.

Se analizaron el histograma y la gráfica cuantil-cuantil y de los residuos del Modelo 2 que se encuentran en la figura 4.12. En dichas gráficas puede observarse que la distribución de los residuos del Modelo 2 es muy parecida a la de los residuos del Modelo 1 y que la cola derecha es pesada, sin embargo es menos pesada que la de los residuos del Modelo 1.

Al igual que para el Modelo 1, se realizaron tres pruebas de normalidad cuyos resultados se encuentran en el cuadro 4.2. Para las prueba de Anderson-Darling y de Shapiro-Wilk el supuesto de normalidad no se sostiene, sin embargo para la prueba de Lilliefors sí. Debido a que se tiene el dato atípico correspondiente a abril del 2012, se tomará por aproximado el resultado y se aceptará la hipótesis de normalidad.

Cuadro 4.2: Pruebas de normalidad a los residuos del Modelo 2.

Prueba	Estadístico	P-valor
Anderson-Darling	1.0696	0.008022
Lilliefors	0.0745	0.0618
Shapiro-Wilk	0.96037	0.0005604

4.1.3. Comparación de los modelos

Se compararon los Criterios de Información Bayesiano y de Akaike, cuadro 4.3, de ambos modelos y en todos los casos se eligió al Modelo 2. Además el supuesto de normalidad se sostiene mejor para el Modelo 2 que para el Modelo 1 y el Modelo 2 es más simple.

Por las razones anteriores se prefirió al Modelo 2 y se utilizó para realizar pronósticos a doce meses.

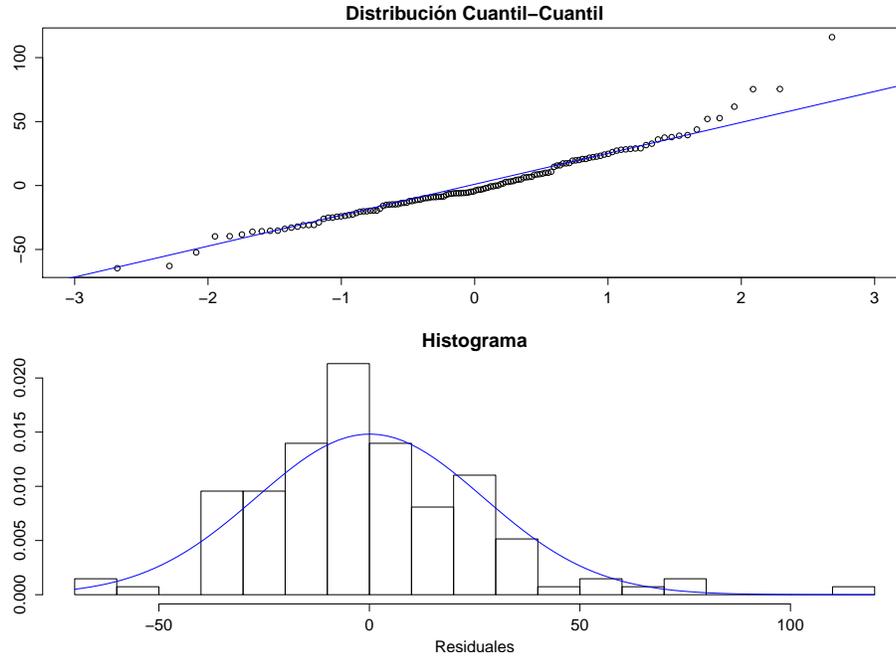


Figura 4.12: Gráfica cuantil-cuantil e histograma de los residuos del segundo modelo ajustado a la serie $\nabla serv_t$.

Cuadro 4.3: Verosimilitud de los modelos ajustados a la serie $\nabla serv_t$.

Modelo	AIC	AICc	BIC
Modelo 1	1292.53	1293.18	1310
Modelo 2	1291.28	1291.74	1305.84

Se obtuvo un pronóstico a doce meses para la serie $x_t = \nabla serv_t = serv_t - serv_{t-1}$. Dado que los valores de la serie de ventas al sector servicios de junio del 2013 a mayo del 2013 ya fueron observados, los valores pronosticados fueron invertidos para obtener pronósticos a doce meses de la serie original $serv_t$, esto es: $serv_t = \nabla serv_t + serv_{t-1}$ y compararlos con los valores observados. Se obtuvieron los siguientes resultados:

Cuadro 4.4: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.

	Jun. 2013	Jul. 2013	Ago. 2013	Sep. 2013	Oct. 2013	Nov. 2013
Pronóstico	723.2	727.1	733.1	733.2	734.2	736.3
Valor observado	1,401.1	708.5	708.8	714.0	689.3	718.5
	Dic. 2013	Jan. 2014	Feb. 2014	Mar. 2014	Abr. 2014	Mayo 2014
Pronóstico	738.0	739.5	741.2	742.8	744.4	746.1
Valor observado	720.0	716.2	728.7	696.9	730.2	735.8

En el cuadro anterior se observa que en junio del 2013 se presentó un pico debido a un ajuste que el modelo no pudo captar y predecir, sin embargo, para los meses posteriores el modelo logra captar la tendencia de la serie de ventas al sector Servicios. En la siguiente gráfica se muestran dichos pronósticos con sus intervalos de confianza al 95 %.

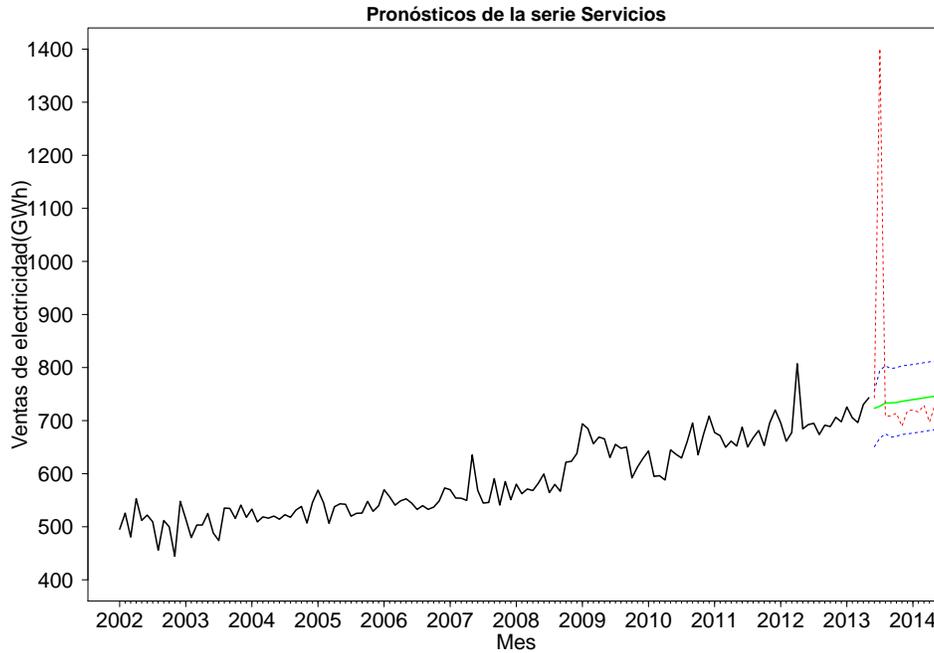


Figura 4.13: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.

Debido a que el dato de junio del 2013 es un dato atípico, el valor del mismo fue reemplazado con el promedio de los valores de mayo del 2013 y julio del 2013. La gráfica de los pronósticos con el valor de junio del 2013 corregido se presenta en la gráfica 4.14.

4.1.4. Tercera propuesta

Debido a que, al realizar las pruebas de normalidad del Modelo 2, se observaron p-valores muy pequeños para las pruebas de Anderson-Darling y de Shapiro-Wilk y como la variabilidad de la serie parece aumentar a partir del 2009, se decidió probar con el logaritmo de la serie $serv_t$ y ajustar un modelo ARMA(2,0,1), es decir un modelo con los ordenes p y q del Modelo 2, a la serie $Z_t = \nabla \log(serv_t)$.

El modelo ajustado (Modelo 3) es:

$$\begin{aligned} Z_t &= \mu(1 - \phi_1 - \phi_2) + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= 0.004 - 0.2484 Z_{t-1} - 0.2357 Z_{t-2} - 0.4727 \varepsilon_{t-1} + \varepsilon_t. \end{aligned} \quad (4.3)$$

A continuación se verificarán los supuestos del modelo. En las figuras 4.16 y 4.17 se presentan la gráfica de los residuos y la gráfica de la función de autocorrelación de los mismos. En ellas no se observan evidencias de correlación entre los residuos. Se realizaron además pruebas de Ljung-Box para los rezagos 4 a 24 (figura 4.18) y en todas ellas se aceptó la hipótesis de que los residuos constituyen un proceso de ruido blanco.

En cuanto a la normalidad, se analizaron el histograma y la gráfica cuantil-cuantil de los residuos (figura 4.19) y no se encontraron grandes desviaciones de la normalidad. Se realizaron las pruebas de

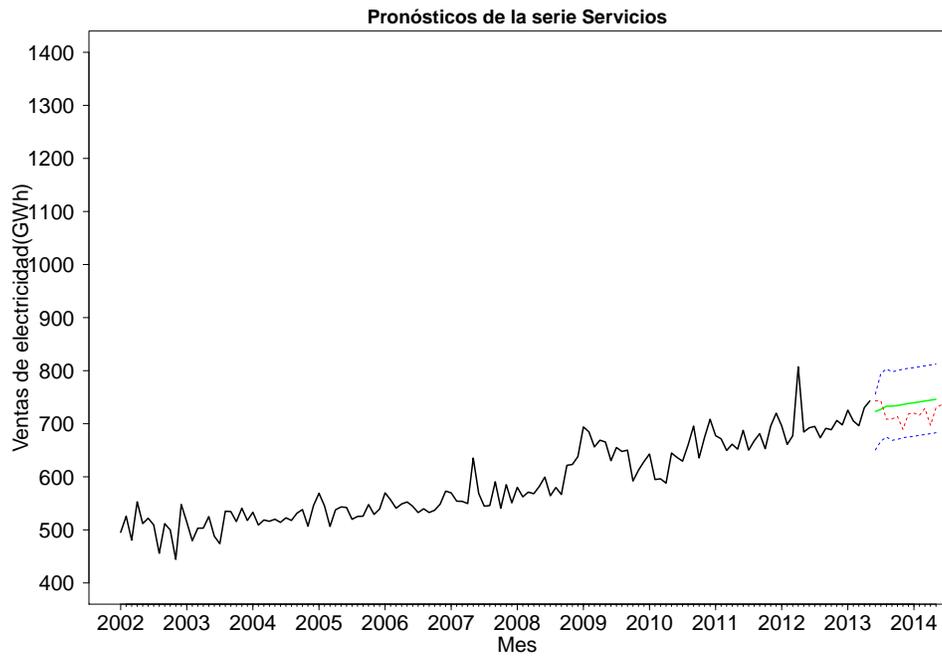


Figura 4.14: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.

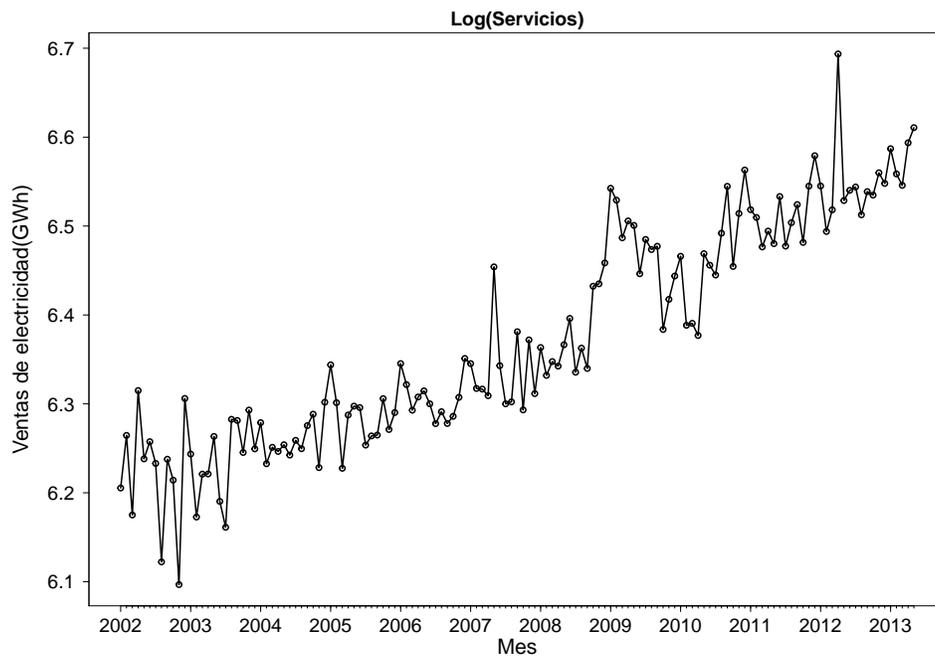


Figura 4.15: Gráfica del logaritmo de la serie de ventas de electricidad al sector Servicios.

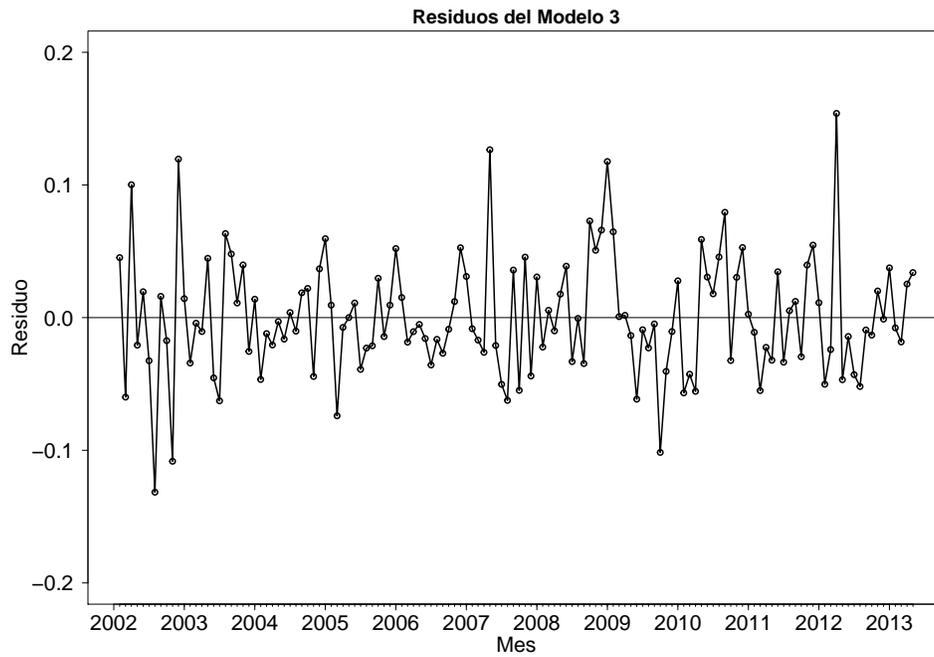


Figura 4.16: Residuos del modelo ajustado a la serie $\nabla \log(serv_t)$.

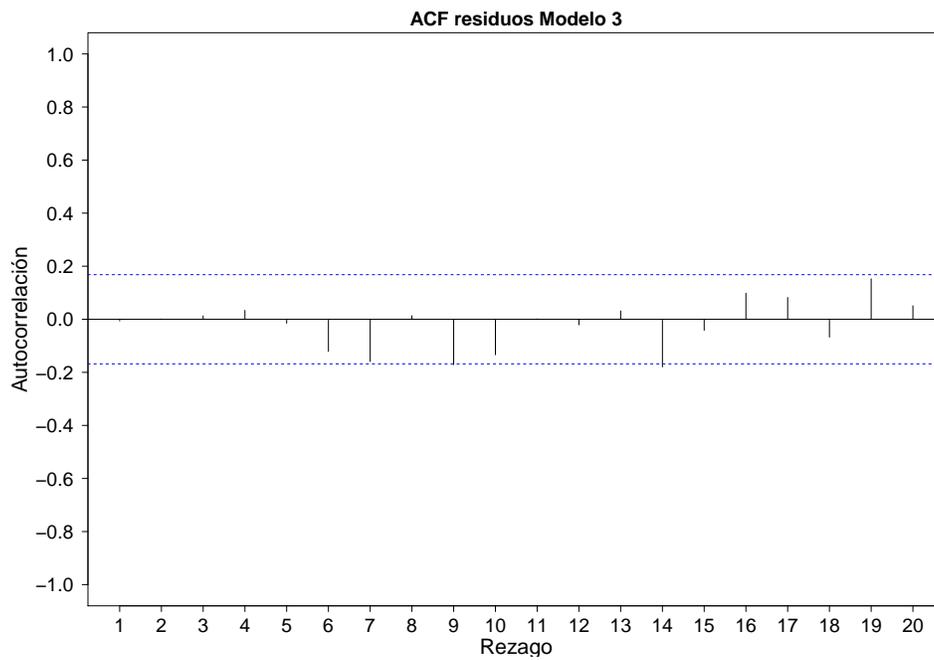


Figura 4.17: Función de autocorrelación de los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$.

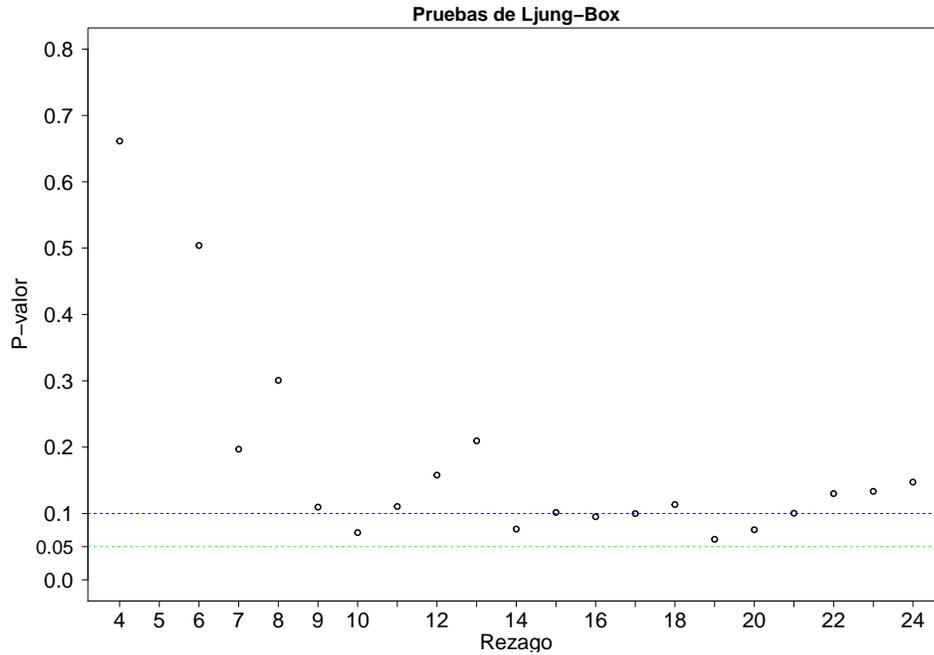


Figura 4.18: Prueba de Ljung-Box para los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$.

normalidad Anderson-Darling, Lilliefors y Shapiro-Wilk, los resultados se muestran en el cuadro 4.5; para las prueba de Anderson-Darling y Shapiro-Wilk el supuesto de normalidad sigue sin sostenerse pero se observan p-valores menos pequeños que en el Modelo 2. Para la prueba de Lilliefors el supuesto de normalidad sí se sostiene.

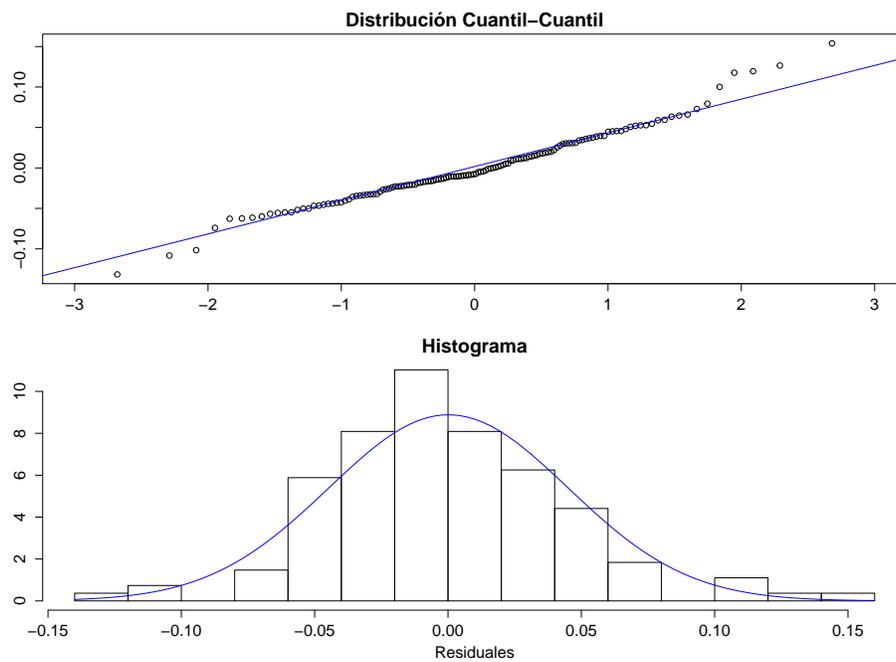


Figura 4.19: Gráfica cuantil-cuantil e histograma de los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$.

Cuadro 4.5: Pruebas de normalidad a los residuos del Modelo 3.

Prueba	Estadístico	P-valor
Anderson-Darling	0.89061	0.02227
Lilliefors	0.074031	0.06502
Shapiro-Wilk	0.97503	0.01333

4.1.5. Propuesta final

Al comparar las gráficas cuantil-cuantil, los histogramas y las pruebas de normalidad de los modelos 2 y 3, se encontró que los supuestos de normalidad se sostienen mejor para el Modelo 3. Debido a lo anterior, se decidió aplicar un logaritmo y luego una diferencia a los datos y utilizar el Modelo 3 para realizar pronósticos.

Se obtuvo un pronóstico a doce meses para la serie $x_t = \nabla \log(serv_t) = \log(serv_t) - \log(serv_{t-1})$. Los valores pronosticados fueron invertidos para obtener pronósticos a doce meses de la serie original $serv_t$ y compararlos con los valores observados. Se obtuvieron los siguientes resultados:

Cuadro 4.6: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios

	Jun. 2013	Jul. 2013	Ago. 2013	Sep. 2013	Oct. 2013	Nov. 2013
Pronóstico	722.8	727.7	734.1	734.3	735.6	738.2
Valor observado	1,401.1	708.5	708.8	714.0	689.3	718.5
	Dic. 2013	Jan. 2014	Feb. 2014	Mar. 2014	Abr. 2014	Mayo 2014
Pronóstico	740.2	742.0	744.0	746.1	748.1	750.1
Valor observado	720.0	716.2	728.7	696.9	730.2	735.8

En la siguiente gráfica se muestran dichos pronósticos con sus intervalos de confianza al 95 %.

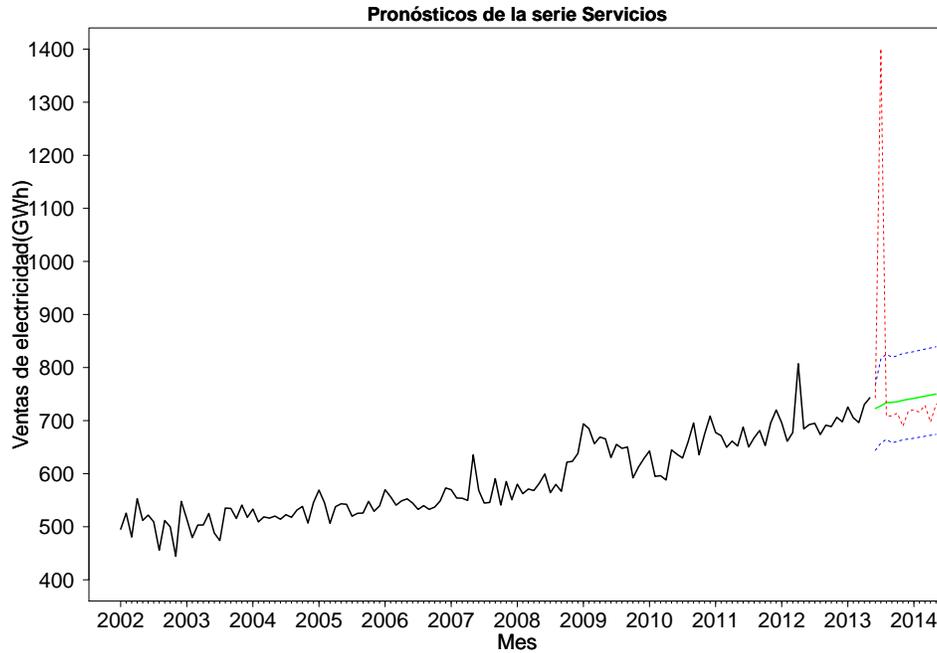


Figura 4.20: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.

Debido a que el dato de junio del 2013 es un dato atípico, al igual que para el Modelo 2, el valor del mismo fue reemplazado con el promedio de los valores de mayo del 2013 y julio del 2013. La gráfica de los pronósticos con el valor de junio del 2013 corregido se presenta a continuación:

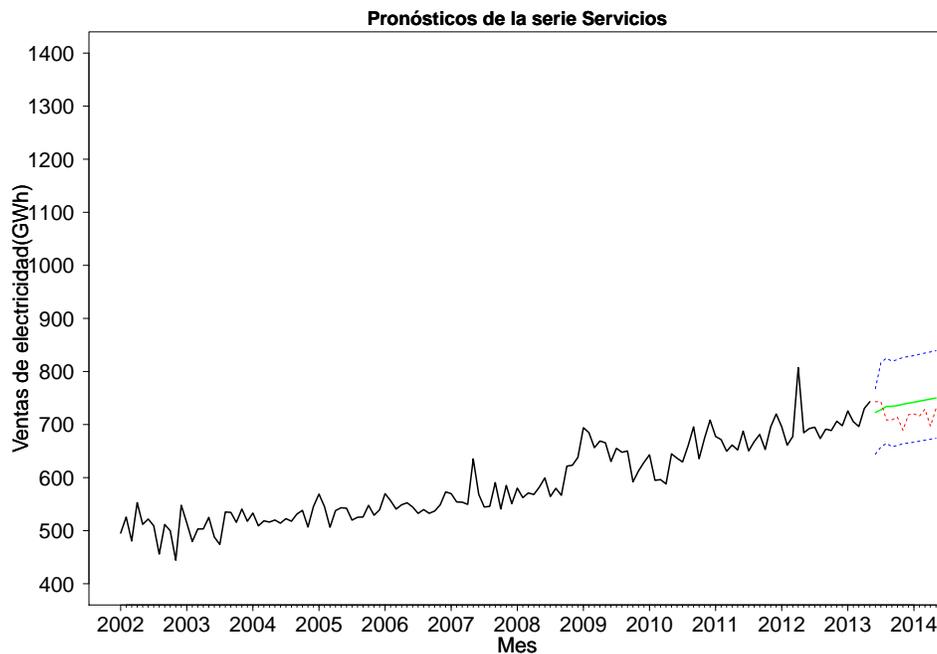


Figura 4.21: Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.

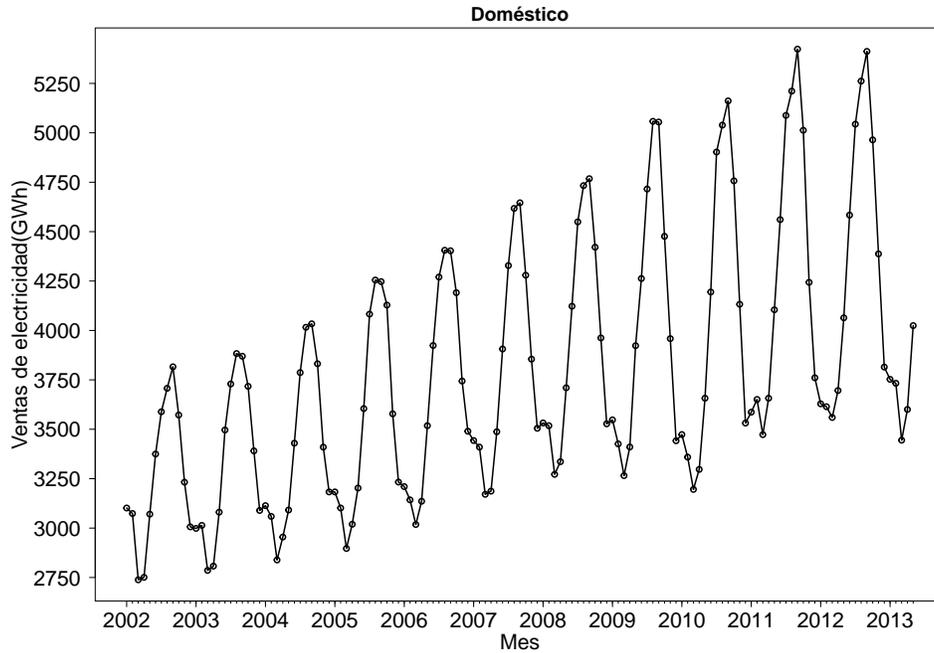


Figura 4.22: Ventas de electricidad mensuales al sector Doméstico de enero de 2002 a mayo de 2013 en GWh.

4.2. Doméstico

En la figura 4.22 se muestra la gráfica de la serie de ventas mensuales de electricidad, en Gigawhats-hora, al sector Doméstico, de enero de 2002 a mayo de 2013. Esta serie será denotada como dom_t , $t = 1, \dots, 137$. En ella puede observarse una tendencia a la alza y un componente estacional con periodo $d = 12$. Además se observa que, para cada año, los meses con mayor consumo de electricidad son julio, agosto, septiembre y octubre, esto se debe al uso de aire acondicionado durante el verano. Así mismo, puede observarse que la variabilidad de la serie aumenta a lo largo del tiempo, lo cual sugiere aplicar alguna transformación de Box-Cox a los datos. En la gráfica del logaritmo de la serie de ventas al sector Doméstico, $\log(dom_t)$, que se encuentra en la figura 4.23, se puede observar que si bien quedan algunos picos con una varianza un poco mayor en general la varianza de la serie se estabiliza. Es necesario entonces remover al componente estacional y a la tendencia de la serie $\log(dom_t)$.

4.2.1. Primera propuesta

En un primer intento para remover al componente estacional, al logaritmo de los datos les fue aplicado el operador ∇_{12} , esto es:

$$\nabla_{12}\log(dom_t) = \log(dom_t) - \log(dom_{t-12}) \quad \text{donde } 12 < t \leq 137.$$

Como se mencionó en el capítulo 2, para que un proceso sea débilmente estacionario, su media, su función de covarianza y su función de autocorrelación no deben depender del tiempo. Sin embargo, al examinar la función de autocorrelación estimada $\hat{\rho}(h)$ a partir de la serie $\nabla_{12}\log(dom_t)$ se tiene

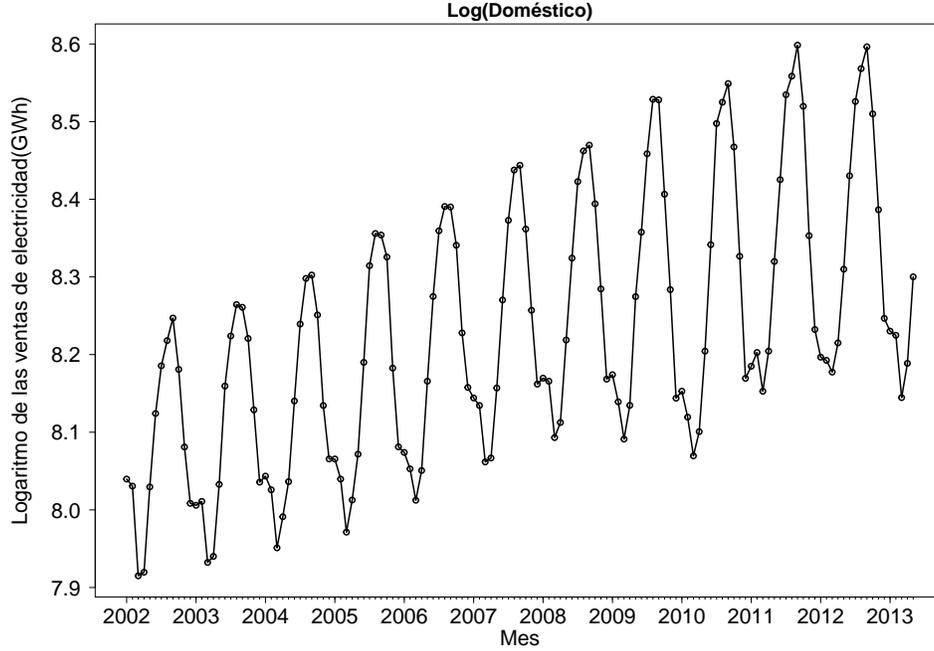


Figura 4.23: Logaritmo de la serie de ventas al sector Doméstico

que $\hat{\rho}(1) = 0.6749$, $\hat{\rho}(2) = 0.3863$, $\hat{\rho}(3) = 0.27011$, etc. por lo que no es posible suponer que dicha serie sea una realización de un proceso débilmente estacionario. Dado que la mayor autocorrelación de la serie $\nabla_{12}\log(dom_t)$ se presentó cuando $h=1$, se decidió aplicar una diferencia de orden 1 a la misma. En la figura 4.24 se muestra a la serie $\nabla\nabla_{12}\log(dom_t)$, aparentemente esta serie no presenta ni tendencia ni estacionalidad y además parece que su variabilidad no aumenta a lo largo del tiempo, por lo tanto, parecería razonable suponer que se trata de una realización de un proceso débilmente estacionario. Sin embargo, al examinar a la función $\hat{\rho}(h)$ de dicha serie, figura 4.25, se puede observar que $\hat{\rho}(2) = -0.256$, $\hat{\rho}(12) = -0.34489$, $\hat{\rho}(14) = 0.236$ y $\hat{\rho}(19) = -0.188$. Derivado de este análisis se sugeriría ajustar un modelo con al menos 12 ó 14 parámetros lo cual no es conveniente ni deseable. Debido a lo anterior se buscará utilizar otro método para llevar a la serie dom_t a la estacionariedad débil.

4.2.2. Segunda propuesta

Como alternativa a la diferenciación, la tendencia y el componente estacional fueron estimados utilizando el método de promedios móviles que se encuentra en el capítulo 2 de la presente tesis. En él, en primer lugar se busca estimar al componente estacional, para ello se hace una primera estimación de la tendencia (\hat{m}) que es removida de la serie, luego, la serie sin tendencia es utilizada para estimar al componente estacional.

La tendencia fue estimada de la siguiente forma:

$$\hat{m}_t = \frac{1}{24}dom_{t-6} + \frac{1}{12} \sum_{i=-5}^5 dom_{t+i} + \frac{1}{24}dom_{t+6}, \quad 6 < t \leq 132.$$

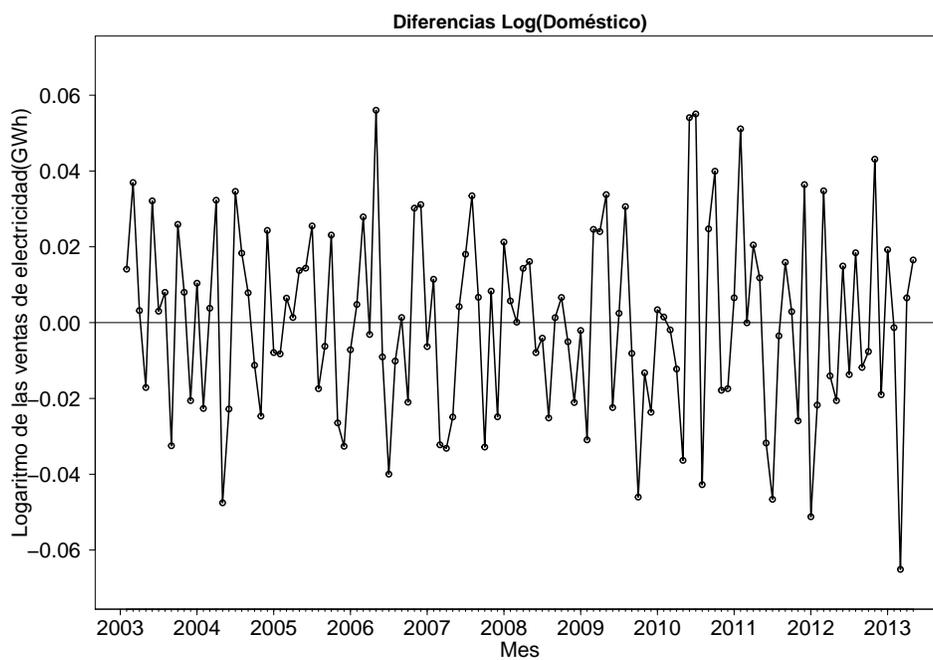


Figura 4.24: Gráfica de la serie $\nabla\nabla_{12}\log(dom)$.

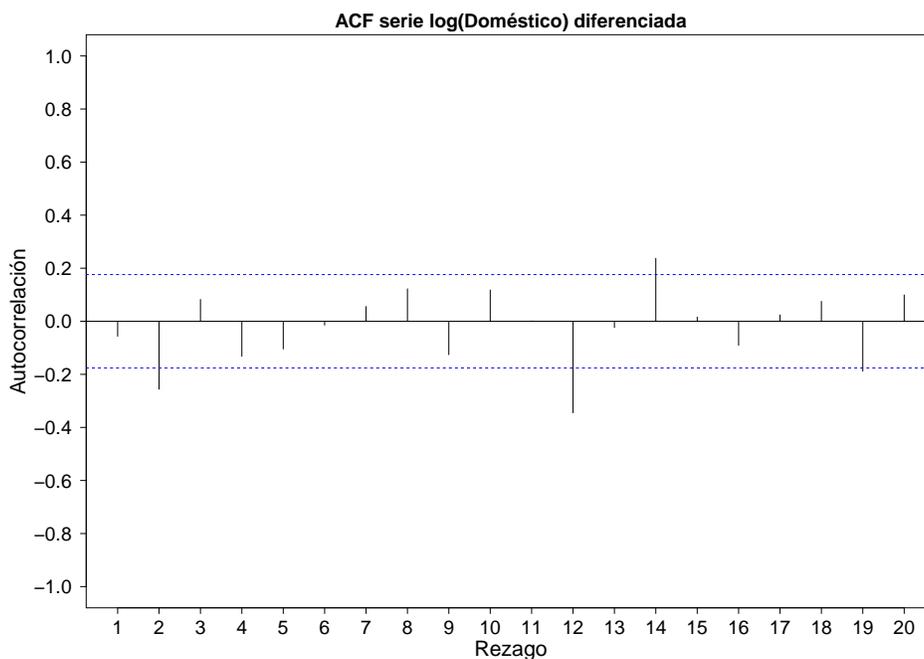


Figura 4.25: Función de autocorrelación de la serie $\nabla\nabla_{12}\log(dom)$.

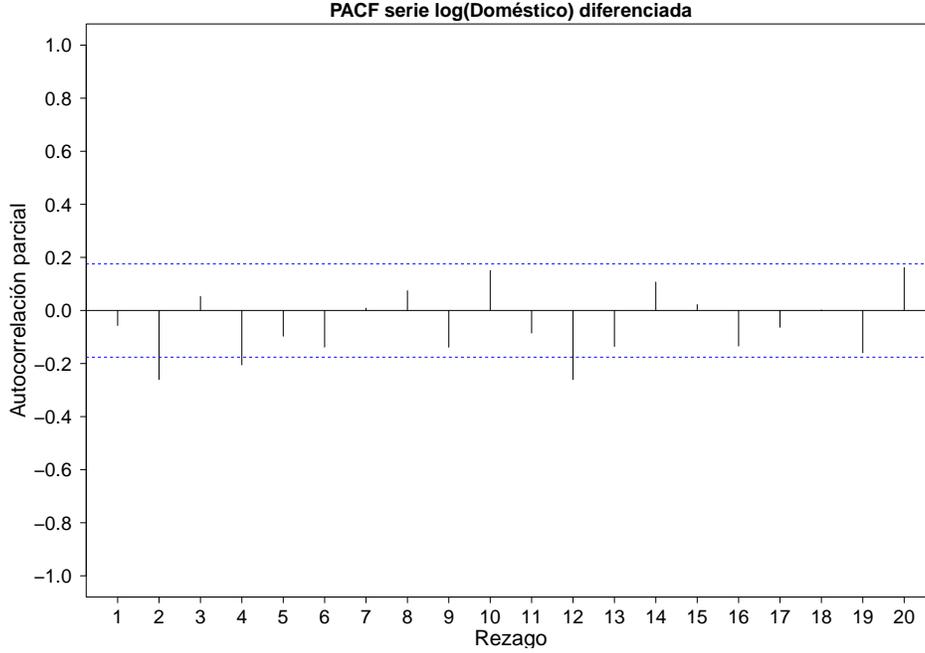


Figura 4.26: Función de autocorrelación parcial de la serie $\nabla\nabla_{12}\log(dom)$.

Los componentes estacionales \hat{w}_k y \hat{s}_k con $k \in \{1, \dots, 12\}$ fueron estimados a partir de \hat{m}_t de la siguiente forma:

Sea $x_i = \log(dom_i)$ entonces, $\forall k \in \{1, \dots, 12\}$ se tiene

$$\hat{w}_k = \text{promedio}\{x_{k+12j} - \hat{m}_{k+12j} : \frac{6-k}{12} < j \leq \frac{131-k}{12}\}$$

$$\text{y } \hat{s}_k = \hat{w}_k - \bar{w} \quad \text{donde} \quad \bar{w} = \sum_{k=1}^{12} \hat{w}_k$$

Una vez que se estimó al componente estacional \hat{s}_k , este fue removido de la serie $\log(dom)_t$ y finalmente la tendencia de esta nueva serie sin estacionalidad, $domsuav_t$, fue estimada utilizando un promedio móvil. En este caso se utilizó:

$$\hat{m}_t = \frac{1}{24}domsuav_{t-6} + \frac{1}{12} \sum_{i=-5}^5 domsuav_{t+i} + \frac{1}{24}domsuav_{t+6}, \quad 6 < t \leq 132.$$

La serie $\hat{y}_t = \log(dom_t) - \hat{s}_t - \hat{m}_t$, figura 4.30, ya no presenta ni tendencia. Sin embargo sí presenta estacionalidad, y si bien la varianza no parece aumentar a lo largo del tiempo, se siguen observando varios picos con una varianza mayor. En la figura 4.31 se observa la gráfica de su función de autocorrelación, claramente la serie \hat{y}_t no es débilmente estacionaria.

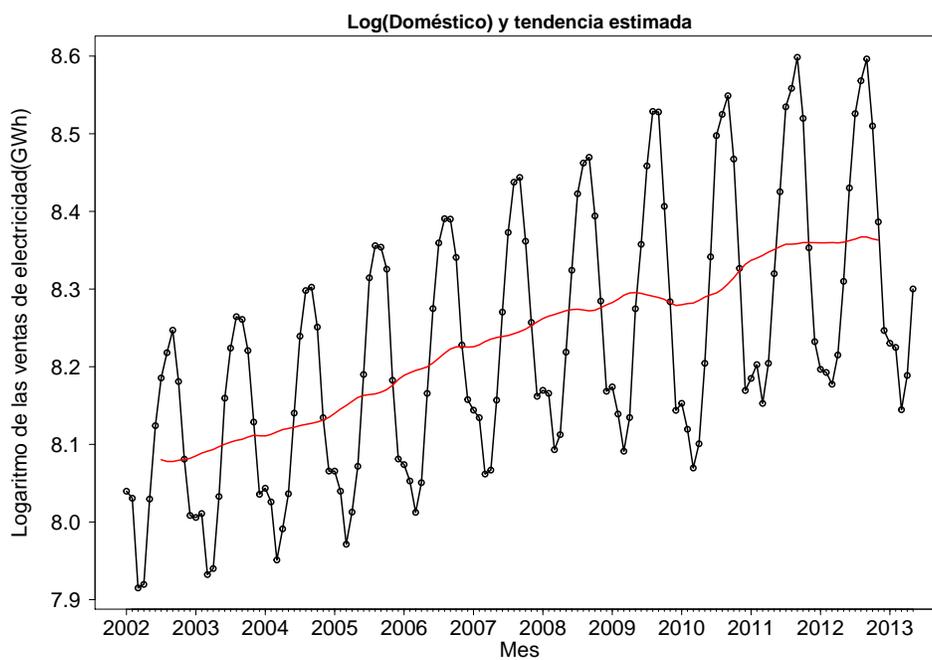


Figura 4.27: Tendencia estimada de la serie $\log(dom)$.

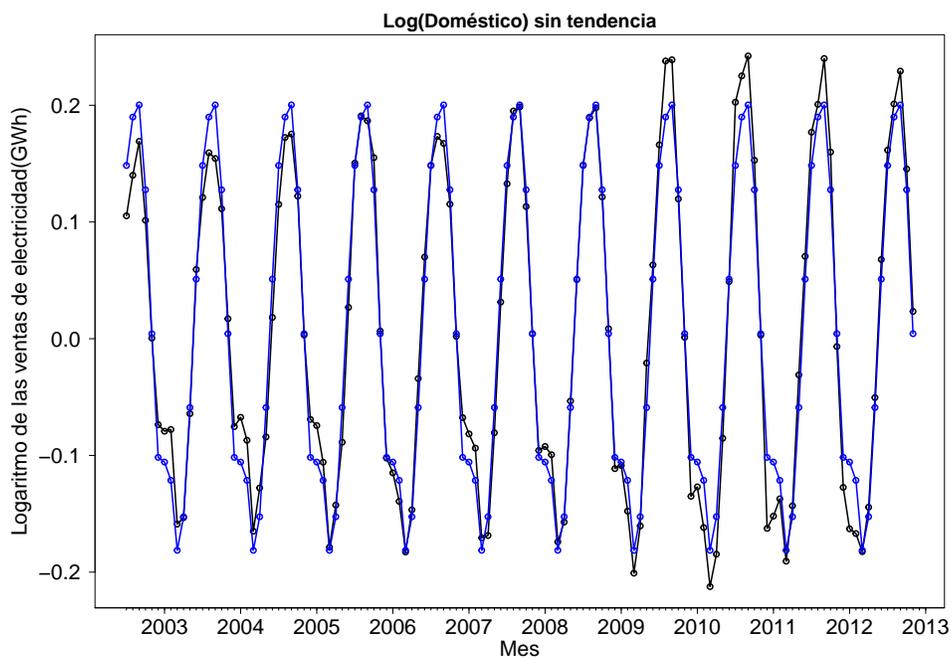


Figura 4.28: Serie sin tendencia $(\log(dom))_t - \hat{m}_t$ y componente estacional estimado \hat{s}_t .

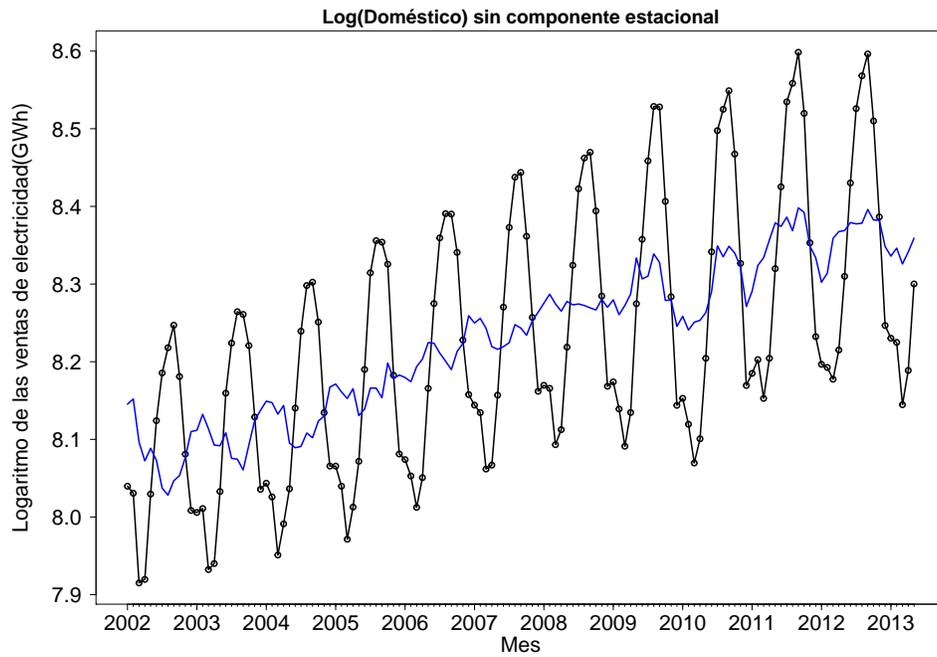


Figura 4.29: Serie $\log(dom_t)$ sin el componente estacional \hat{s}_t y $\log(dom_t)$.

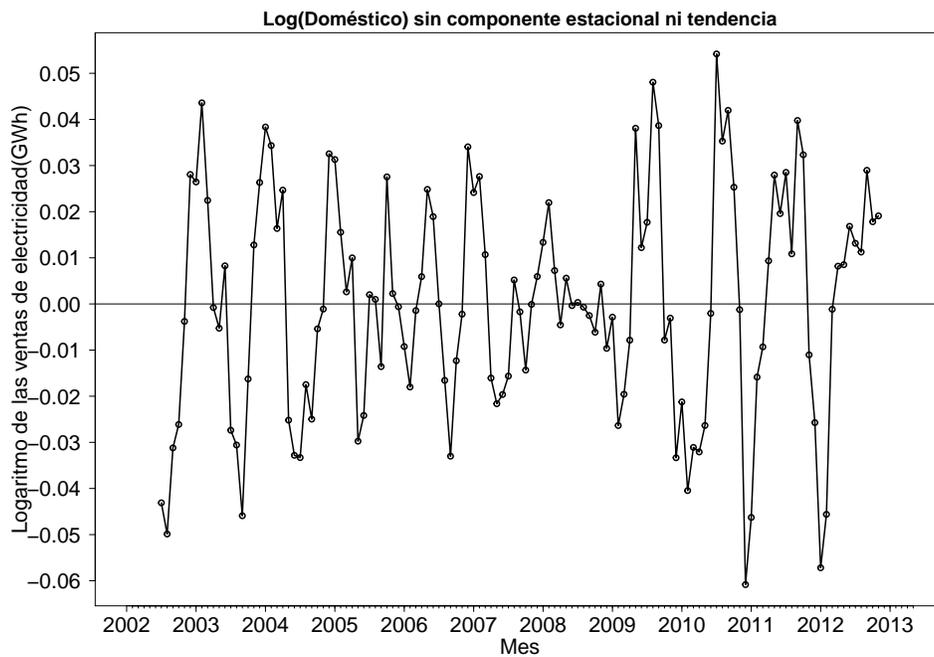


Figura 4.30: Serie $\log(dom_t) - \hat{s}_t - \hat{\mu}_t$.

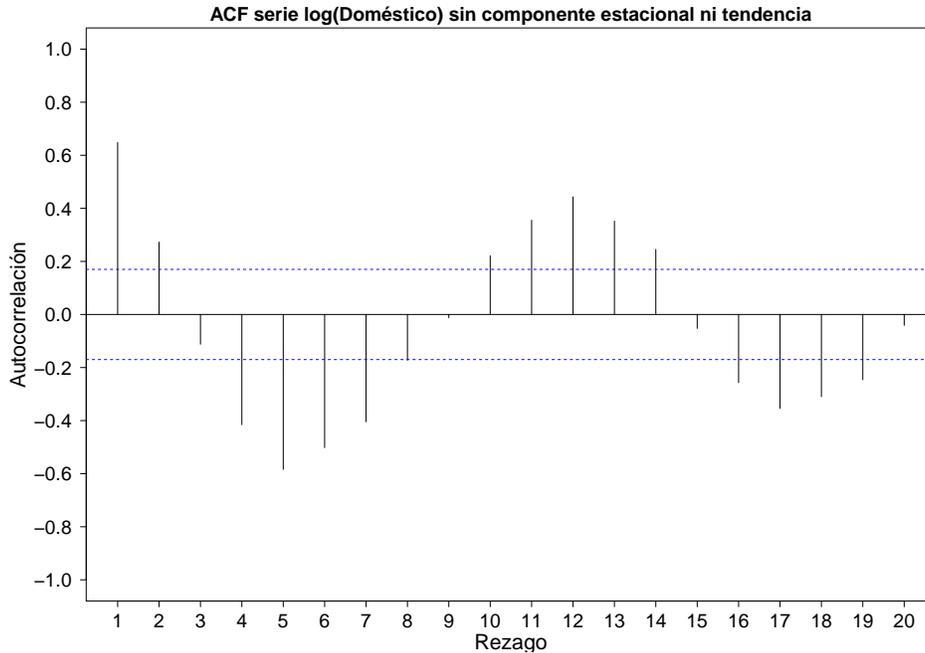


Figura 4.31: Función de autocorrelación de la serie $\log(dom_t) - \hat{s}_t - \hat{\mu}_t$.

4.2.3. Tercera propuesta

Dado que la varianza de la serie dom_t crece mucho a lo largo del tiempo, ni aplicar el logaritmo y después diferenciar ni aplicar el logaritmo y después utilizar promedios móviles fue suficiente para llevar a la serie a la estacionariedad débil, pues en el primer caso la serie resultante presenta algunos picos de mayor varianza que obligarían a ajustar un modelo ARIMA con al menos 12 parámetros y en el segundo se siguió observando estacionalidad. Debido a lo anterior se decidió probar el ajuste de un modelo ARCH(1) ya que es un modelo simple que permite que la varianza de la serie dependa del tiempo y de la observación anterior.

A diferencia de los modelos ARMA que asumen que la varianza de un proceso $\{X_t\}$ es constante a lo largo del tiempo, los modelos ARCH (de heterocedasticidad condicional autorregresivos) consideran que dicha varianza no es constante en el tiempo, es decir, $var(X_t) = \sigma_t^2$. El modelo ARCH(1) es el modelo más simple de la familia de modelos ARCH, este modelo describe la dependencia de una proceso $\{X_t\}$ mediante una función cuadrática de su valor anterior.

Definición 4.2.1. ARCH(1)

En un modelo ARCH(1) una serie de tiempo $\{X_t\}$ es descrita de la siguiente forma:

$$X_t = \sigma_t \epsilon_t \quad (4.4)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 \quad (4.5)$$

donde $\epsilon_t \sim \text{iid } N(0, 1)$ y $\alpha_0 > 0$ y $\alpha_1 \geq 0$ para garantizar que la varianza sea positiva.

Un modelo ARCH(1) puede ser reescrito de la siguiente forma:

$$\begin{aligned}
 X_t^2 &= \sigma_t^2 + X_t^2 - \sigma_t^2 \\
 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \sigma_t^2 \epsilon_t^2 - \sigma_t^2 \\
 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \sigma_t^2 (\epsilon_t^2 - 1) \\
 &= \alpha_0 + \alpha_1 X_{t-1}^2 + v_t
 \end{aligned}$$

Por lo tanto, un modelo ARCH(1) puede ser visto como un modelo AR(1) para $\{X_t^2\}$ con un nuevo ruido v_t . Dado que los modelos ARCH no son el tema de esta tesis y solamente se está ajustando un modelo ARCH a manera de prueba, no se hablará más a profundidad la teoría de este tipo de modelos.

Antes de ajustar un modelo ARCH a la serie dom_t a esta le fue aplicada primero una diferencia de orden 12 para remover la estacionalidad y después una diferencia de orden uno para remover la tendencia y se obtuvo la serie $\nabla\nabla_{12}dom_t$, figura 4.32. A dicha serie le fue ajustado el modelo ARCH(1) y se obtuvieron los siguientes resultados:

Cuadro 4.7: Coeficientes del modelo ARCH(1) ajustado.

	Estimador	Error est.	Prueba t.	P(> t)
α_0	8.715e+03	1.508e+03	5.781	7.44e-09
α_1	6.782e-02	9.792e-02	0.693	0.489

Del cuadro anterior podemos notar que si bien α_0 es estadísticamente significativo, el valor de α_1 es pequeño y no existe evidencia suficiente para afirmar que sea estadísticamente significativo. Ahora, verifiquemos si los residuos cumplen con los supuestos de ser un ruido blanco y de normalidad.

En el caso de los modelos ARCH es necesario verificar que los residuos al cuadrado no estén correlacionados. Para ello se realizó una prueba de Ljung-Box sobre ellos; el p-valor de dicha prueba fue de 0.9566 por lo que puede decirse que los residuos al cuadrado no están correlacionados.

Para verificar el supuesto de normalidad, se analizaron también el histograma y la gráfica de la función cuantil-cuantil (figura 4.34). En dichas gráficas se observa que los residuos parecen seguir una distribución normal. Además del análisis gráfico se realizaron tres pruebas de normalidad, los resultados de dichas pruebas (cuadro 4.8) permiten suponer que los residuos siguen una distribución normal.

Cuadro 4.8: Pruebas de normalidad a los residuos del modelo ARCH(1).

Prueba	Estadístico	P-valor
Anderson-Darling	0.2315	0.7984
Lilliefors	0.0462	0.7509
Shapiro-Wilk	0.99408	0.8878

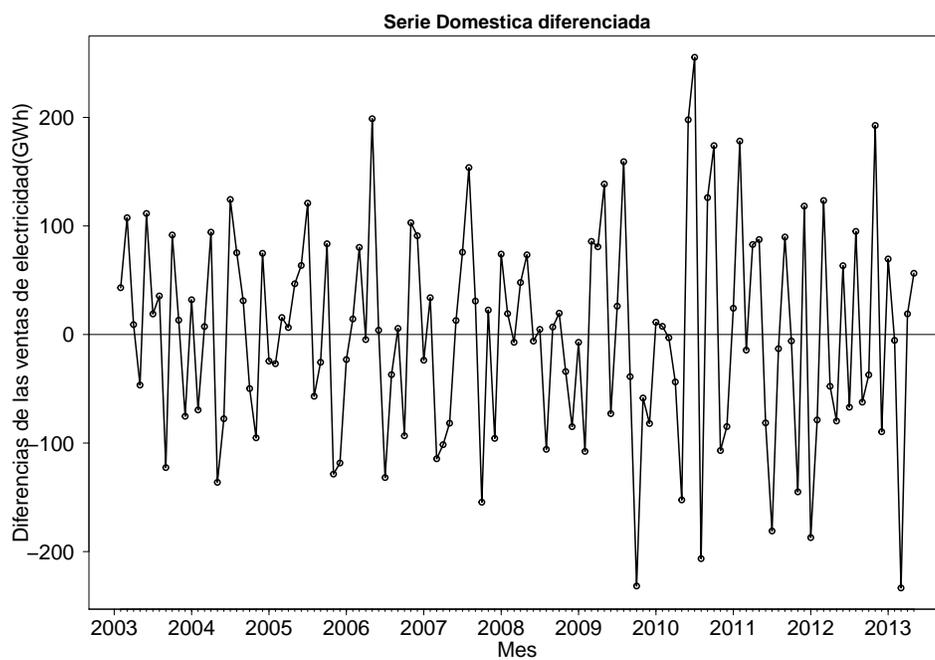
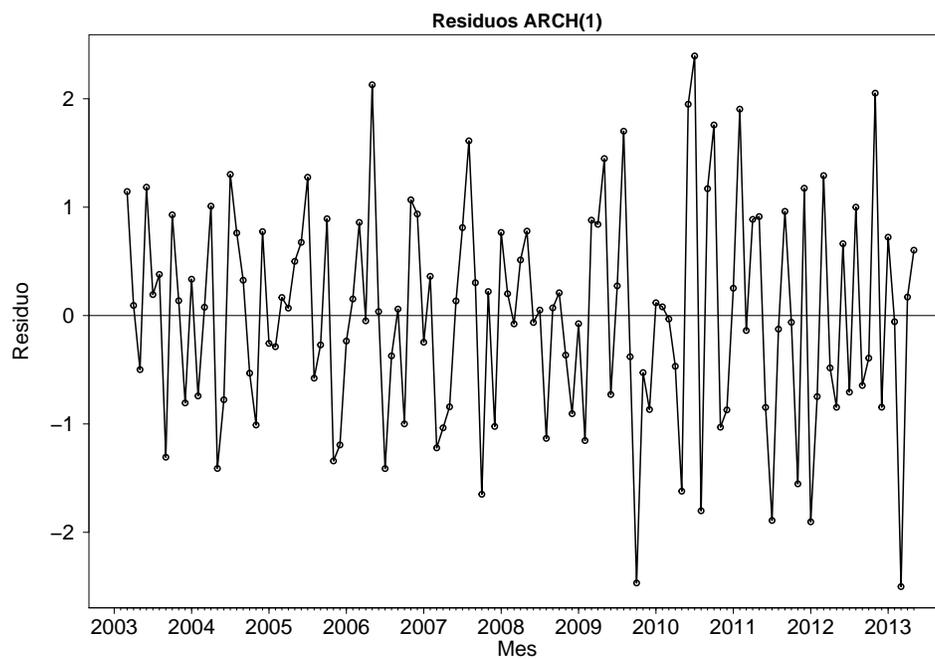
Figura 4.32: Serie $\nabla \nabla_{12} dom_t$.

Figura 4.33: Gráfica de los residuos del modelo ARCH(1).

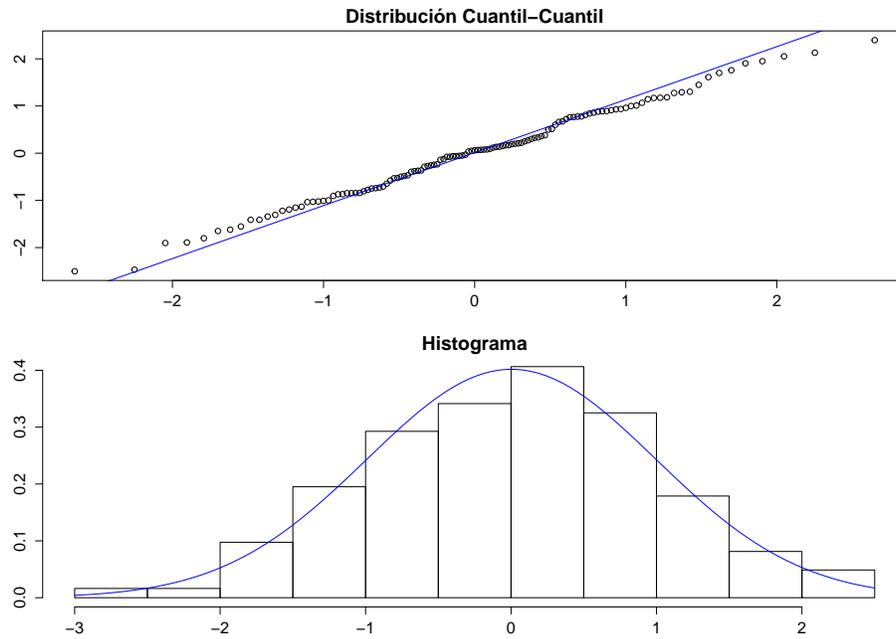


Figura 4.34: Gráfica cuantil-cuantil e histograma de los residuos del modelo ARCH(1).

4.2.4. Propuesta final

Todos los supuestos sobre los residuos se cumplen. Sin embargo, el parámetro α_1 no resultó estadísticamente significativo por lo que, si bien el modelo ARCH(1) que se encuentra en la cuadro 4.7 no es una propuesta del todo descabellada para modelar a la serie doméstica diferenciada, puede ser conveniente buscar algún otro modelo más complejo para modelar a dicha serie.

Conclusiones

El objetivo principal de esta tesis ha sido utilizar los modelos Autoregresivos de Promedios Móviles (ARMA) para modelar las series de ventas internas de electricidad a los sectores Doméstico y Servicios en México.

Ambas series presentaron comportamientos muy distintos, sin embargo, la población aumenta y con ello el consumo de electricidad de los dos sectores antes mencionados.

Mi hipótesis inicial era que las ventas de ambos sectores iban a ser relativamente sencillas de modelar mediante modelos ARMA y en especial aquellas del sector Doméstico, pues presentaban una clara tendencia creciente y una marcada estacionalidad de orden 12.

El análisis de la serie de ventas al sector Servicios me permitió encontrar un modelo ARMA sin mayores complicaciones. Tras un análisis inicial de la función de autocorrelación y en dos iteraciones fue posible encontrar un modelo con un ajuste razonable a los datos diferenciados y además se trataba de un modelo simple, un ARMA(2, 1). Con dicho modelo se obtuvo una predicción a doce meses que podría servir a la CFE para garantizar un suministro oportuno de electricidad y tomar decisiones relacionadas con gastos y tarifas. En mi opinión, el valor de haber realizado el ejercicio de ajustar un modelo ARMA a dicha serie consiste en que permitió ilustrar que la metodología para construir modelos ARMA funciona y que en efecto permite construir modelos que nos ayudan a explicar el comportamiento de los datos observados.

Para el sector Doméstico la historia fue completamente diferente. Al realizar una primera inspección de la gráfica de esta serie parecía que sería una serie muy fácil de modelar con modelos ARMA sin embargo no fue así; ninguno de los métodos para llevar a la serie a la estacionariedad débil fueron efectivos ya que al examinar las funciones de autocorrelación y autocorrelación parcial se sugería ajustar modelos con muchos parámetros. Al analizar dichos resultados, se observó que la varianza de la serie crecía a un paso tal que no podía ser llevada a la estacionariedad débil y suponer una varianza constante, por lo que se decidió intentar modelar a esta serie con un modelo ARCH(1), ya que este modelo supone que la varianza depende del tiempo. Sin embargo, el modelo ARCH(1) ajustado tampoco fue del todo satisfactorio, por lo que explorar otros tipos de modelos para modelar dicha serie puede ser un ejercicio interesante en el futuro. Lo que queda muy claro, simplemente al observar la gráfica de la serie, es que en México, el mayor consumo de electricidad en los hogares se presenta de julio a octubre, por lo que la CFE podría tomar medidas para garantizar el suministro de electricidad en dichos meses.

Finalmente, el mayor aprendizaje que tuve al realizar este trabajo fue que al trabajar con modelos estadísticos debemos buscar modelos para responder a las preguntas que nos planteemos más no debemos perdernos en la búsqueda de el modelo perfecto, pues el modelo perfecto no existe. Al final del día, un modelo estadístico no es más que una herramienta coherente que nos ayuda a entender mejor el mundo en el que vivimos y a tomar decisiones.

Datos

Año	Mes	Domestico	Servicios	Año	Mes	Domestico	Servicios	Año	Mes	Domestico	Servicios
2002	Enero	3101.6	495.4	2006	Enero	3209.7	569.8	2010	Enero	3473.2	642.9
2002	Febrero	3073.6	525.6	2006	Febrero	3142.7	556.5	2010	Febrero	3359.1	594.9
2002	Marzo	2738.3	480.6	2006	Marzo	3018.2	540.7	2010	Marzo	3195.7	596.2
2002	Abril	2750.9	552.8	2006	Abril	3135.9	548.8	2010	Abril	3296.9	588.1
2002	Mayo	3070.4	511.9	2006	Mayo	3518.1	552.6	2010	Mayo	3657	644.8
2002	Junio	3374.8	521.9	2006	Junio	3923.9	544.6	2010	Junio	4194.5	636.5
2002	Julio	3588.6	509.3	2006	Julio	4270.2	532.6	2010	Julio	4902.8	629.5
2002	Agosto	3707.1	455.9	2006	Agosto	4405.6	539.8	2010	Agosto	5038.9	659.8
2002	Septiembre	3816.1	511.7	2006	Septiembre	4403.1	532.7	2010	Septiembre	5161.4	695.5
2002	Octubre	3572.1	499.8	2006	Octubre	4191.3	537	2010	Octubre	4757	635.6
2002	Noviembre	3232.3	444.4	2006	Noviembre	3743.9	548.6	2010	Noviembre	4132.6	674.6
2002	Diciembre	3006.2	547.9	2006	Diciembre	3489.9	573.1	2010	Diciembre	3531.2	708.4
2003	Enero	2998.5	514.7	2007	Enero	3442.9	569.8	2011	Enero	3586.6	677.5
2003	Febrero	3013.6	479.5	2007	Febrero	3409.8	554.1	2011	Febrero	3650.7	671.6
2003	Marzo	2785.9	503.2	2007	Marzo	3170.8	553.7	2011	Marzo	3472.9	649.7
2003	Abril	2807.6	503.3	2007	Abril	3187	549.6	2011	Abril	3657	661.4
2003	Mayo	3080.5	525	2007	Mayo	3487.5	635.3	2011	Mayo	4104.6	652.1
2003	Junio	3496.4	488	2007	Junio	3906.2	568.5	2011	Junio	4560.7	687.6
2003	Julio	3729.1	474	2007	Julio	4328.3	544.6	2011	Julio	5088	650.3
2003	Agosto	3883.1	535.2	2007	Agosto	4617.5	545.8	2011	Agosto	5211	667.7
2003	Septiembre	3869.5	534.5	2007	Septiembre	4645.8	590.6	2011	Septiembre	5423.3	681.4
2003	Octubre	3717.2	515.6	2007	Octubre	4279.5	540.9	2011	Octubre	5012.9	653
2003	Noviembre	3390.6	540.8	2007	Noviembre	3854.6	585.2	2011	Noviembre	4243.5	695.7
2003	Diciembre	3089.2	517.7	2007	Diciembre	3504.9	550.9	2011	Diciembre	3760.4	719.9
2004	Enero	3113.5	533.2	2008	Enero	3532	580.2	2012	Enero	3628.7	695.8
2004	Febrero	3059.1	509.1	2008	Febrero	3518.1	562.3	2012	Febrero	3614.1	661.1
2004	Marzo	2838.7	518.6	2008	Marzo	3271.9	571.1	2012	Marzo	3559.7	677.4
2004	Abril	2954.7	516.2	2008	Abril	3336	568.2	2012	Abril	3696.1	807.2
2004	Mayo	3091.4	520.1	2008	Mayo	3709.9	582	2012	Mayo	4064	684.5
2004	Junio	3429.7	514.1	2008	Junio	4122.5	599.5	2012	Junio	4583.5	692.4
2004	Julio	3786.8	522.7	2008	Julio	4549.3	564.3	2012	Julio	5043.8	695
2004	Agosto	4016.1	517.8	2008	Agosto	4732.7	579.8	2012	Agosto	5261.8	673.6
2004	Septiembre	4033.6	531.4	2008	Septiembre	4767.9	566.7	2012	Septiembre	5411.8	691.4
2004	Octubre	3831.5	538.3	2008	Octubre	4421.2	621.6	2012	Octubre	4964.2	688.7
2004	Noviembre	3409.7	506.9	2008	Noviembre	3962.2	623.3	2012	Noviembre	4387.4	706.2
2004	Diciembre	3183.1	545.6	2008	Diciembre	3527.6	638.1	2012	Diciembre	3814.7	697.8
2005	Enero	3182.9	569	2009	Enero	3547.5	694	2013	Enero	3752.6	725.6
2005	Febrero	3101.6	545.3	2009	Febrero	3425.9	684.8	2013	Febrero	3732.6	705.3
2005	Marzo	2896.8	506.5	2009	Marzo	3265.5	656.4	2013	Marzo	3444.7	696.3
2005	Abril	3019.2	537.8	2009	Abril	3410.4	669	2013	Abril	3600.1	730.4
2005	Mayo	3202.6	543.2	2009	Mayo	3922.9	665.6	2013	Mayo	4024.4	743
2005	Junio	3604.5	542.3	2009	Junio	4262.6	630.4				
2005	Julio	4082.6	519.9	2009	Julio	4715.5	655.2				
2005	Agosto	4255	525.3	2009	Agosto	5058.1	647.8				
2005	Septiembre	4246.9	525.8	2009	Septiembre	5054.5	650.2				
2005	Octubre	4128.4	547.8	2009	Octubre	4476.1	592.1				
2005	Noviembre	3578	529.1	2009	Noviembre	3958.6	612.5				
2005	Diciembre	3233	539.3	2009	Diciembre	3442	628.8				

Bibliografía

Libros

- [1] Peter J. Brockwell & Richard A. Davis; Introduction to Time Series and Forecasting; Springer; 2002.
- [2] Peter J. Brockwell & Richard A. Davis; Time Series: Theory and Methods; Springer; 1991.
- [3] Ciro Cardoso, Carmen Reyna et al.; México en el siglo XIX (1821-1910) Historia económica y de la estructura social; Nueva Imagen; 1994.
- [4] Ngai Hang Chan; Time Series Applications to finance with R and S-Plus; Wiley; 2010.
- [5] William H. Greene; Econometric Analysis; Prentice Hall; 2002.
- [6] Olga Pellicer de Brody y Esteban L. Mancilla; Historia de la Revolución Mexicana (Periodo 1952-1960); El Colegio de México; 1978.
- [7] Robert H. Shumway & David S. Stoffer; Time Series Analysis and Its Applications With R Examples; Springer; 2006.
- [8] Ruey S. Tsay; Analysis of Financial Time Series; Wiley; 2005.
- [9] William W. S. Wey; Time Series Analysis; Pearson; 2006.

Índice de cuadros

1.1. Número de usuarios de energía eléctrica por sector, a mayo del 2013.	7
1.2. Ventas de electricidad en el mercado interno, de enero de 2013 a mayo del 2013 en MWh.	8
1.3. Estructura del sector Industrial a mayo del 2013.	8
3.1. Comportamiento del ACF y del PACF de un proceso $ARMA(p, q)$ causal e invertible.	34
4.1. Pruebas de normalidad a los residuos del primer modelo ajustado.	44
4.2. Pruebas de normalidad a los residuos del Modelo 2.	47
4.3. Verosimilitud de los modelos ajustados a la serie $\nabla serv_t$	48
4.4. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.	48
4.5. Pruebas de normalidad a los residuos del Modelo 3.	53
4.6. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios	53
4.7. Coeficientes del modelo ARCH(1) ajustado.	62
4.8. Pruebas de normalidad a los residuos del modelo ARCH(1).	62

Índice de figuras

4.1. Ventas de electricidad al sector Servicios de enero del 2002 a mayo del 2013 en GWh.	40
4.2. ACF de la serie de ventas de electricidad al sector Servicios.	40
4.3. Gráfica de la serie de ventas de electricidad al sector Servicios diferenciada ($\nabla serv_t$).	41
4.4. ACF y PACF de la serie $\nabla serv_t$	42
4.5. Residuos del primer modelo ajustado a la serie $\nabla serv_t$	43
4.6. ACF de los residuos del primer modelo ajustado a la serie $\nabla serv_t$	43
4.7. Prueba de Ljung-Box para los residuos del primer modelo ajustado a la serie $\nabla serv_t$	44
4.8. Gráfica cuantil-cuantil e histograma de los residuos del primer modelo ajustado a la serie $\nabla serv_t$	45
4.9. Residuos del segundo modelo ajustado a la serie $\nabla serv_t$	46
4.10. ACF de los residuos del segundo modelo ajustado a la serie $\nabla serv_t$	46
4.11. Prueba de Ljung-Box para los residuos del segundo modelo ajustado a la serie $\nabla serv_t$	47
4.12. Gráfica cuantil-cuantil e histograma de los residuos del segundo modelo ajustado a la serie $\nabla serv_t$	48
4.13. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.	49
4.14. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.	50
4.15. Gráfica del logaritmo de la serie de ventas de electricidad al sector Servicios.	50
4.16. Residuos del modelo ajustado a la serie $\nabla \log(serv_t)$	51
4.17. Función de autocorrelación de los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$	51
4.18. Prueba de Ljung-Box para los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$	52
4.19. Gráfica cuantil-cuantil e histograma de los residuos del modelo ajustado a la serie $\nabla \log(serv_t)$	52
4.20. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios.	54

4.21. Pronóstico a doce meses de la serie de ventas de electricidad al sector Servicios. . . .	54
4.22. Ventas de electricidad mensuales al sector Doméstico de enero de 2002 a mayo de 2013 en GWh.	55
4.23. Logaritmo de la serie de ventas al sector Doméstico	56
4.24. Gráfica de la serie $\nabla\nabla_{12}\log(dom)$	57
4.25. Función de autocorrelación de la serie $\nabla\nabla_{12}\log(dom)$	57
4.26. Función de autocorrelación parcial de la serie $\nabla\nabla_{12}\log(dom)$	58
4.27. Tendencia estimada de la serie $\log(dom)$	59
4.28. Serie sin tendencia $(\log(dom)_t - \hat{m}_t)$ y componente estacional estimado \hat{s}_t	59
4.29. Series $\log(dom_t)$ sin el componente estacional \hat{s}_t y $\log(dom_t)$	60
4.30. Serie $\log(dom_t) - \hat{s}_t - \hat{\mu}_t$	60
4.31. Función de autocorrelación de la serie $\log(dom_t) - \hat{s}_t - \hat{\mu}_t$	61
4.32. Serie $\nabla\nabla_{12}dom_t$	63
4.33. Gráfica de los residuos del modelo ARCH(1).	63
4.34. Gráfica cuantil-cuantil e histograma de los residuos del modelo ARCH(1).	64