



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

APLICACIÓN DE ANÁLISIS ESTADÍSTICO
DE SERIES DE TIEMPO
A LAS VENTAS DE UN MICRO NEGOCIO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

LUIS HUMBERTO CARO HERNÁNDEZ



DIRECTOR DE TESIS:
DR. RICARDO RAMIREZ ALDANA

2017

CIUDAD UNIVERSITARIA, CDMX



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno

Caro

Hernández

Luis Humberto

66 47 63 26

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

409076662

2. Datos del tutor

Dr

Ricardo

Ramírez

Aldana

3. Datos del sinodal 1

Dr

Alberto

Contreras

Cristán

4. Datos del sinodal 2

Dra

Ruth Selene

Fuentes

García

5. Datos del sinodal 3

Dra

Lizabeth

Naranjo

Albarrán

6. Datos del sinodal 4

Act

Ángel Manuel

Godoy

Aguilar

7. Datos del trabajo escrito

Aplicación de análisis estadístico de series
de tiempo a las ventas de un micro negocio

106 p

2017

*Agradezco y dedico este trabajo a
Victoria, Rubén, Victor, Daniel y Karla
su esfuerzo fue esencial en la elaboración de este trabajo,
y su presencia vital al formarme como persona*

Índice general

Introducción	V
1. Modelo de regresión lineal simple	1
1.1. Conceptos básicos del modelo de regresión lineal simple	1
1.2. Estimación de parámetros	4
1.3. Coeficiente de determinación y de correlación	6
1.4. Construcción de intervalos de confianza para los parámetros	7
1.5. Análisis de varianza	10
1.6. Detección de autocorrelación	12
1.7. Aplicación	15
2. Análisis clásico de series temporales	21
2.1. Conceptos básicos del análisis clásico de series temporales	21
2.1.1. Definición de series de tiempo	21
2.1.2. Componentes de una serie temporal	23
2.2. Modelos aditivo y multiplicativo y criterios para detectarlos	26
2.3. Análisis de estacionalidad	28
2.4. Análisis de tendencia	33
2.5. Métodos de alisado	36
2.5.1. Alisado exponencial simple	37
2.5.2. Alisado exponencial doble (modelo de Holt)	39
2.5.3. Modelo Holt-Winters	40
2.6. Aplicación	43
3. Metodología Box-Jenkins	51
3.1. Conceptos básicos	51
3.1.1. Funciones de autocovarianza y autocorrelación	53
3.1.2. Función de autocorrelación parcial	55
3.2. Modelos $MA(q)$	56
3.3. Modelos $AR(p)$	61
3.4. Modelos $ARMA(p, q)$	67
3.5. Modelos $ARIMA$	70
3.6. Modelos $SARIMA$	71
3.7. Predicción	74
3.8. Diagnóstico del modelo	76

3.9. Aplicación	77
4. Modelos de regresión dinámicos (Regresión con errores ARMA)	85
4.1. Conceptos básicos de los modelos de regresión con errores ARMA	85
4.2. Tendencia en modelos de regresión dinámicos	86
4.3. Estimación de los coeficientes de regresión	87
4.4. Identificación del modelo	88
4.5. Aplicación	90
5. Conclusiones	101
Bibliografía	105

Introducción

Este trabajo pretende abordar el tema de la estimación de los ingresos futuros, un problema muy común en los negocios, y de mucha utilidad a los comerciantes pues les permite planificar sus gastos en función del dinero que estimen recibir por concepto de sus ventas. Particularmente el interés de este trabajo está en estudiar los micro negocios dedicados a la venta de artículos de tlapalería y ferretería. Los gastos a planificar en este tipo de negocios son compra de mercancía, pago de sueldos, renta o impuestos, éstos en su mayoría gastos fijos, pero también deben prever gastos de algunos siniestros como reparación del inmueble o robo. La suma de estos puede ocasionar fuertes impactos en la economía de un negocio, provocando fluctuaciones en sus ingresos y en ocasiones el cierre del establecimiento. Por tanto, la planificación de gastos y predicción de los ingresos debe situarse como un procedimiento de vital importancia.

El objetivo de esta tesis es obtener predicciones de las ventas futuras de un establecimiento, aplicando distintos métodos estadísticos. El análisis realizado se conoce como análisis de series de tiempo, pues hace uso de datos recolectados a través del tiempo. Para la elaboración de este trabajo se utilizaron los datos de un negocio familiar de este ramo, ubicado en el municipio Nezahualcóyotl, Edo. de México. Al mismo tiempo, estas proyecciones pueden ser de utilidad a los dueños quienes, con la información generada, tienen la posibilidad de mejorar el desempeño de su empresa.

La estructura del documento es de cuatro capítulos teórico-prácticos y un último capítulo de conclusiones. En cada uno de los primeros cuatro capítulos se explica el fundamento teórico-matemático de una metodología diferente de análisis de series de tiempo. El orden en que aparece cada tema va de acuerdo a la complejidad que presenta, de manera que los primeros capítulos tratan los temas más sencillos, mientras que los últimos contienen temas más complejos y recientes en el campo del análisis de series de tiempo. Al final de cada capítulo se presenta la aplicación de la metodología a los datos reales y se comentan brevemente los resultados.

En el capítulo 1 se explica el análisis de regresión lineal simple, un análisis cuyo objetivo es la predicción de una variable a partir de otra. Este análisis no fue creado para el tratamiento de datos recolectados a través del tiempo, pero la intención es introducir los conceptos básicos de un modelo estadístico, como su construcción, estimación de parámetros, la interpretación de los parámetros obtenidos, la medición del ajuste y las pruebas de hipótesis asociadas con la significancia de los parámetros y la validación de supuestos. Todos estos son conceptos que aparecen en cualquier tipo de análisis estadístico, particularmente en los análisis de series de

tiempo. El modelo de regresión lineal se retoma en posteriores capítulos como un modelo de apoyo en el modelamiento de ciertos comportamientos en la serie o incluso como parte de un modelo de predicción avanzada.

En el capítulo 2 se exponen los métodos clásicos del análisis de series de tiempo, los cuales parten de la idea que las series están integradas por cuatro componentes: tendencia, ciclo, estacionalidad y componente irregular. Para obtener estimaciones futuras de los datos, se supone un esquema en que los componentes están combinados, ya sea sumados (esquema aditivo) o multiplicados (esquema multiplicativo), y a través del esquema elegido se separa cada componente encontrando patrones de regularidad que puedan ser reproducidos.

En este capítulo también se aborda la forma correcta de iniciar el estudio de series temporales eliminando alteraciones en la serie por la misma medición del tiempo (efecto calendario), y el alza de precios (inflación) que puede dar la apariencia de haber tendencia donde no la hay. Al final del capítulo se presentan los métodos de alisado exponencial (también llamados de suavizamiento exponencial o *exponential smoothing* en inglés), tres métodos clasificados como clásicos pero con diferencias marcadas con los métodos arriba mencionados, pues se retroalimentan a medida que se generan nuevos datos, lo que permite obtener resultados satisfactorios con menor número de observaciones.

El capítulo 3 habla sobre la metodología Box-Jenkins la cual hace uso de procesos estocásticos, esto implica visualizar a la serie como un conjunto de variables aleatorias, cada una en cierto tiempo, de forma que cada observación corresponde al valor que toma la variable aleatoria en un tiempo dado. Para modelar el comportamiento de una serie real se parte del supuesto de que las variables están correlacionadas entre sí. En ocasiones cada variable puede verse como una regresión en donde los predictores son variables de la misma serie en momentos pasados (modelos autorregresivos), en otras ocasiones pueden verse como un conjunto de choques aleatorios (modelos de promedios móviles). La combinación de estos modelos se conoce como procesos ARMA y permiten modelar una amplia gama de series temporales siempre que sean estacionarias, es decir, que la media general y la correlación entre variables no dependan del tiempo. Aunque dichos modelos no puedan ser utilizados para series con tendencia o estacionalidad, algunas sencillas modificaciones permiten considerar tales comportamientos en las series. Así las series que presentan tendencia pueden ser modeladas por procesos ARIMA, las series que presentan tendencia y estacionalidad se modelan con procesos SARIMA y series que presentan estacionalidad pero no tendencia pueden ser modeladas por procesos SARMA.

Bajo esta metodología, la serie analizada puede ser generada por alguno de los procesos mencionados por lo que parte del capítulo trata sobre las técnicas de identificación de modelo, en esta etapa se observa la autocorrelación de las variables y los tipos de comportamientos presentes en la serie. En cualquiera de estos procesos, la estimación de los parámetros se hace necesariamente con ayuda de un paquete estadístico que a lo largo de todo el trabajo será el programa R. Después de estimar los parámetros de varios modelos posibles se debe elegir entre alguno de ellos, esta elección puede estar en función del ajuste del modelo pero también debe corroborarse que los residuales cumplan ciertos supuestos, por lo que se mencionan las

pruebas de hipótesis correspondientes.

El capítulo 4 es el último referente a métodos de análisis de series temporales, y trata el análisis dinámico, para algunos autores considerado como un método avanzado de predicción. Se trata de una combinación entre los modelos ARMA del capítulo 3 con el análisis de regresión visto en el capítulo 1. En otras palabras el objetivo es predecir una serie de tiempo en función de otra serie mediante una regresión lineal, modificando el supuesto de que los errores no deban estar correlacionados entre sí, sino que se modelen con alguno de los procesos ARMA, ARIMA, SARMA o SARIMA. Con ello se logra obtener el peso de cada variable explicativa sobre la explicada de forma confiable al no violar supuestos y reducir el error de predicción con respecto a los métodos anteriores.

En el texto se indica que por el hecho de incluir procesos ARMA, su estimación requiere de un paquete estadístico y varios pasos para ajustar un modelo adecuado. Sin embargo, es posible ahorrar algunos pasos en la identificación del modelo debido a que existen procedimientos que a partir de las estimaciones hechas con un modelo genérico permiten obtener un mejor modelo en pocos pasos.

Por último en el capítulo 5 se presentan las conclusiones del trabajo, en estas se resume la información que proporcionó cada método sobre la serie. Se compara la precisión de las predicciones de los modelos utilizados y se comentan las ventajas y desventajas más importantes de cada modelo. También se emiten recomendaciones a los dueños del negocio con las cuales mejorar a futuro la precisión de las estimaciones, y de la elección del modelo de estimación según las circunstancias.

Capítulo 1

Modelo de regresión lineal simple

1.1. Conceptos básicos del modelo de regresión lineal simple

El análisis de regresión estudia la dependencia de una variable (variable dependiente) respecto de una o más variables (variables explicativas) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos de la(s) segunda(s).

Algunos ejemplos comunes de este tipo son: la dependencia del consumo personal con respecto al ingreso, la demanda de un producto en relación al gasto en publicidad, o bien el rendimiento de un cultivo en relación de la temperatura, lluvia, cantidad de sol y fertilizantes. Si el análisis tiene solo una variable explicativa se le conoce como análisis de regresión lineal simple, mientras que si tiene dos o más variables explicativas se le conoce como análisis de regresión lineal múltiple. En este capítulo se abordará únicamente el modelo de regresión lineal simple.

Gujarati y Porter (2010) mencionan algunas características del análisis de regresión que lo diferencian de otra clase de análisis:

- En el análisis de regresión interesa lo que se conoce como dependencia estadística entre variables, no así la funcional o determinista, propia de la física clásica.
- Una relación estadística por sí misma no puede implicar causalidad, para pensar en causalidad se debe acudir a consideraciones a priori o teóricas. Por ejemplo, si bien se sabe que el rendimiento del cultivo depende de la lluvia (entre otras cosas), no existe una razón estadística que impida pensar que la lluvia depende del cultivo, pero la relación de dependencia e independencia que se da a las variables se debe a cuestiones no estadísticas como el sentido común.
- A diferencia de un análisis de correlación, en donde el objetivo es medir el grado de asociación lineal entre dos variables, el análisis de regresión pretende estimar el valor

promedio de una variable con base en los valores fijos de otras. En el análisis de correlación no hay diferencia entre la variable dependiente y explicativa. Por ejemplo, la correlación entre las calificaciones de exámenes de matemáticas y estadística es la misma que entre las calificaciones de los exámenes de estadística y los exámenes de matemáticas.

El objetivo del análisis de regresión simple es averiguar la forma en que varía el valor promedio de una variable dependiente Y de acuerdo al valor dado de una variable explicativa X ; en términos de probabilidad decimos que el valor esperado de la variable dependiente Y dado un valor X_i de la variable explicativa X está en función de ese valor X_i , esto es:

$$E(Y|X_i) = f(X_i), \quad i = \{1, \dots, N\}$$

donde N corresponde al número de observaciones en la población, y la función $f(X_i)$ se conoce como función de esperanza condicional o *función de regresión poblacional (FRP)*.

El modelo de regresión lineal simple tiene la siguiente FRP:

$$f(X_i) = \beta_1 + \beta_2 X_i$$

donde β_1 y β_2 son parámetros desconocidos pero fijos, que se denominan coeficientes de regresión; a β_1 se le conoce como coeficiente de intercepción, o término constante, y a β_2 como coeficiente de pendiente.

La linealidad se puede interpretar de dos formas, linealidad en las variables y linealidad en los parámetros. Linealidad en las variables indica que la FRP es una función lineal de la variable X , es decir que la variable explicativa está elevada al exponente uno en la FRP, por ejemplo, $f(X_i) = \beta_1 + \beta_2 X_i$. Por otro lado, la linealidad en los parámetros, indica que la FRP es una función lineal de los coeficientes de regresión, por tanto sólo pueden estar elevados a la potencia uno, por ejemplo:

$$f(X_i) = \beta_1 + \beta_2 X_i^2$$

$$f(X_i) = \beta_1 + \beta_2 X_i + \beta_3 X_i^2,$$

son modelos de regresión lineales en los parámetros pero no en las variables. La linealidad en los parámetros es la pertinente en el desarrollo de la teoría del modelo, por lo que en regresión cuando se habla de linealidad se refiere a linealidad en los parámetros.

El *Modelo Clásico de Regresión Lineal (MCRL)* se compone de un componente determinista, la FRP, más un componente estocástico, el término aleatorio llamado u_i , es decir:

$$Y_i = \beta_1 + \beta_2 X_i + u_i,$$

por tanto, u_i es la desviación del valor observado Y_i con respecto a su media condicional $u_i = Y_i - E(Y|X_i)$.

En la mayoría de los casos ocurre que sólo se dispone de una muestra de la población para hacer el estudio y no de toda la población, por ello usamos el concepto de función de regresión muestral (FRM) que como la FRP es la esperanza condicional de Y dado un valor de X pero

con la diferencia de que se trabaja con una muestra de la población. Esta se expresa de la siguiente manera:

$$E(\widehat{Y}|X_i) = \hat{\beta}_1 + \hat{\beta}_2 X_i,$$

donde $E(\widehat{Y}|X_i)$ es el estimador de $E(Y|X_i)$, $\hat{\beta}_1$ el estimador de β_1 y $\hat{\beta}_2$ el estimador de β_2 . El modelo de regresión lineal simple usando la FRM queda de la siguiente manera¹:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i, \quad i = \{1, \dots, n\}$$

donde n corresponde al número de observaciones en la muestra, y \hat{u}_i es el error muestral. El objetivo principal del análisis de regresión es estimar la FRP con base en la FRM. Este término de perturbación estocástica es de gran importancia en el análisis de regresión por muchas razones entre ellas las siguientes:

- Una teoría puede tener la certeza de que X afecta a Y pero ignoraría o no tendría certeza sobre otras variables, por lo que u_i es un sustituto de todas las variables excluidas en el modelo.
- Aunque se conozcan las variables que afecten a Y además de X es común que no se tenga la información disponible sobre las mismas.
- Es posible que la influencia conjunta de todas las demás variables o varias de ellas sea muy pequeña, y su efecto combinado sea mejor tratarlo como una variable aleatoria u_i .
- A pesar de introducir al modelo todas las variables pertinentes es posible seguir observando aleatoriedad intrínseca en la variable dependiente Y que no sea explicada.
- Principio de parsimonia: Conviene mantener el modelo de regresión lo más sencillo posible ya que si la teoría explica sustancialmente la relación y no tiene argumentos bastante sólidos para incluir más variables, es mejor que el término u_i represente a todas las demás variables.
- En modelos con dos variables (explicativa y dependiente) el modelo puede inferirse al observar un diagrama de dispersión, por ejemplo, podría verse si los datos se ajustan a una línea recta como en el MCRL, o bien a una parábola o haber una relación exponencial, etc., pero en un modelo de regresión múltiple no es fácil determinar la forma funcional adecuada.

Por otra parte, el MCRL plantea siete supuestos de la forma en que se generan X_i y u_i :

- 1) El modelo de regresión es lineal en los parámetros, que como se indicó anteriormente puede ser o no ser lineal en las variables.
- 2) Los valores que toma la variable explicativa X pueden considerarse fijos en muestras repetidas, o haber sido muestreados junto con la variable dependiente, en este caso se supone que X es independiente del término de error, esto es, $Cov(X_i, u_i) = 0$.

¹Gujarati, D. N. y Porter, D. C., 2010, *Econometría*, (Pilar Carril, trad.). México: Mc Graw Hill, p. 44

- 3) El valor medio del error u_i es igual a cero, es decir, $E[u_i] = 0$.
- 4) Homocedasticidad o varianza constante de u_i . La varianza del término de error, es la misma sin importar el valor de X .
- 5) No hay autocorrelación entre las perturbaciones. Dados dos valores cualesquiera de X , X_i y X_j (i diferente de j), la correlación entre dos u_i y u_j es cero. En otras palabras, estas observaciones se muestrean de manera independiente. Simbólicamente, $Cov(u_i, u_j | X_i, X_j) = 0$, o bien, $Cov(u_i, u_j) = 0$ si X no es estocástica.
- 6) El número de observaciones n debe ser mayor al número de parámetros por estimar.
- 7) No todos los valores de X deben ser iguales.

Si a los supuestos del MCRL se añade el supuesto de que u_i se distribuye como una variable aleatoria normal con media cero y varianza σ^2 , es decir, $u_i \sim N(0, \sigma^2)$, obtenemos el *Modelo Clásico de Regresión Lineal Normal (MCRLN)*. Ya que la FRM es una aproximación de la FRP, es necesario encontrar los parámetros muestrales $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\sigma}^2$ de manera en que se acerquen lo más posible a β_1 , β_2 y σ^2 .

1.2. Estimación de parámetros

Dados los n pares de observaciones (X_i, Y_i) , donde n corresponde al tamaño de la muestra, interesa determinar la FRM, o bien elegir los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$, de modo que sea lo más cercana posible a los valores de Y observados.

Un primer criterio en el que se puede pensar intuitivamente es el de minimizar la suma de los errores (o residuales) $\sum_{i=1}^n \hat{u}_i$. Sin embargo, este criterio da igual peso a los errores grandes o pequeños, dejando la posibilidad de que desviaciones muy grandes positivas, se compensen con desviaciones muy grandes negativas (al igual que con las pequeñas), es decir que la suma podría incluso ser cero pero la dispersión de los datos podría ser muy grande. El método de *Mínimos Cuadrados Ordinarios (MCO)* establece como criterio minimizar la suma de los cuadrados de los residuales (SCR):

$$SCR = \sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i)^2,$$

lo cual ocurre cuando $\frac{\partial SCR}{\partial \hat{\beta}_1} = 0$ y $\frac{\partial SCR}{\partial \hat{\beta}_2} = 0$, que respectivamente corresponden a las siguientes ecuaciones:

$$-\sum_{i=1}^n Y_i + n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n X_i = 0 \quad (1.1)$$

$$-\sum_{i=1}^n Y_i X_i + \hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \sum_{i=1}^n X_i^2 = 0 \quad (1.2)$$

Estas son conocidas como las ecuaciones normales de los coeficientes de regresión. Al dejar de un lado de las ecuaciones los términos que contienen a los coeficientes de regresión, podemos

observarlas como un sistema de ecuaciones con dos variables las cuales son $\hat{\beta}_1$ y $\hat{\beta}_2$. Si se sustituye $\sum_{i=1}^n X_i = n\bar{X}$ y se despeja a $\hat{\beta}_1$ en (1.1) se obtiene:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}, \quad (1.3)$$

que puede substituirse en la ecuación (1.2), de donde:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n Y_i X_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2}. \quad (1.4)$$

Otra forma de escribir el mismo estimador es usando expresiones alternativas para los estimadores muestrales de la covarianza y varianza. Los estimadores insesgados de tienen como denominador $n - 1$, pero si se usa n como denominador los estimadores son los siguientes $\widehat{Cov}(X, Y) = \frac{1}{n} \left(\sum_{i=1}^n X_i Y_i \right) - \bar{X}\bar{Y}$, $\widehat{Var}(X) = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 \right) - \bar{X}^2$, puede verse que si en la

ecuación (1.4) se divide entre n arriba y abajo (es decir, se multiplica por $1 = \frac{1/n}{1/n}$), entonces

$\hat{\beta}_2$ también puede verse como:

$$\hat{\beta}_2 = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)} \quad (1.5)$$

Mientras que $\hat{\beta}_1$ se obtiene a través de la ecuación (1.3).

Para obtener el estimador de σ^2 se usa el hecho de que la variable $\frac{\sum_{i=1}^n \hat{u}_i^2}{\sigma^2}$ se distribuye como una variable aleatoria χ^2 con $n - 2$ grados de libertad. Dado que la esperanza de esta variable es $n - 2$, al despejar se obtiene:

$$E \left(\sum_{i=1}^n \hat{u}_i^2 \right) = (n - 2)\sigma^2$$

por tanto, si se define:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n u_i^2}{n - 2}$$

éste resulta ser un estimador insesgado para σ^2 , y donde el término $n - 2$, como se dijo, es conocido como el número de grados de libertad (gl). El término corresponde al número total de observaciones en la muestra menos el número de parámetros que se estiman o incluyen en el modelo. En otras palabras, es la cantidad de observaciones independientes de un total de n observaciones. Así, para calcular $\sum_{i=1}^n \hat{u}_i^2$, es necesario obtener antes $\hat{\beta}_1$ y $\hat{\beta}_2$. Por consiguiente, estas dos estimaciones imponen dos restricciones a $\sum_{i=1}^n \hat{u}_i^2$, obteniendo entonces, $n - 2$ observaciones independientes.

1.3. Coeficiente de determinación y de correlación

El coeficiente de determinación es una medida de la bondad de ajuste, esto es, dice cuán bien se ajusta la línea de regresión a los datos. Para definirlo debemos tener presente las siguientes sumas:

- La Suma de Cuadrados Totales $SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$ es la suma que mide la variación total de la variable Y con respecto a su media.
- La Suma de Cuadrados Explicada $SCE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$, es la suma asociada a la variación de los valores estimados \hat{Y}_i con respecto a la media. También se le conoce como suma de cuadrados debida a la regresión.
- La Suma de Cuadrados de los Residuos $SCR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$, es la suma asociada a la variación residual, es decir, la variación de los valores estimados con respecto de sus valores observados.

La Suma de Cuadrados Totales tiene la propiedad de descomponerse en función de las dos sumas restantes:

$$SCT = SCE + SCR,$$

lo cual muestra que la variación total en los valores Y observados alrededor del valor de su media puede dividirse en dos partes, una atribuible a la línea de regresión y la otra a algún efecto aleatorio, pues no todas las observaciones de Y caen sobre la línea ajustada. El coeficiente de determinación, denotado por R^2 , mide el porcentaje de variación total en Y explicada por el modelo de regresión, matemáticamente tiene la siguiente expresión:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SCE}{SCT}.$$

Este coeficiente tiene las siguientes dos propiedades:

- 1) Es una cantidad no negativa.
- 2) Sus límites son $0 \leq R^2 \leq 1$. Entre más se acerque a 1 (entre más cercano sea la SCE a la SCT) mejor será el ajuste ya que la variación total de Y es en mayor medida atribuible a la regresión que al error.

Por otro lado, el coeficiente de correlación es una medida del grado de asociación lineal entre dos variables cuantitativas. El coeficiente de correlación muestral se calcula como:

$$R = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2][n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]}}.$$

Entre sus propiedades más importantes se encuentran:

- 1) Cae dentro de los límites $-1 \leq R \leq 1$. El signo lo determina el numerador el cual mide la covariación entre las dos variables.

- 2) Es simétrico, es decir, el coeficiente de correlación entre X y Y es el mismo que entre Y y X .
- 3) Es una medida de asociación lineal o dependencia lineal, su uso en la descripción de relaciones no lineales no tiene significado.
- 4) Una correlación igual a cero no necesariamente implica independencia.

Obsérvese que el coeficiente de correlación al cuadrado corresponde al coeficiente de determinación.

1.4. Construcción de intervalos de confianza para los parámetros

En lugar de depender de un solo estimador puntual, se puede construir un intervalo alrededor del estimador puntual, por ejemplo, dentro de dos o tres errores estándar a cada lado del estimador puntual, tal que este intervalo incluya al verdadero valor del parámetro con una alta probabilidad, por ejemplo, un 95 %. Ésta es, a grandes rasgos, la idea básica de la estimación por intervalos.

Si tomamos por ejemplo al parámetro β_2 el objetivo es encontrar dos números positivos, δ y α , de modo que la probabilidad de que el intervalo aleatorio $(\hat{\beta}_2 - \delta, \hat{\beta}_2 + \delta)$ contenga al verdadero β_2 sea $1 - \alpha$, donde a α se le conoce como nivel de significancia. Partiendo del MCRLN los estimadores por MCO tienen las siguientes propiedades:

- 1) Son insesgados, es decir, el valor esperado del estimador es el parámetro. Matemáticamente: $E(\hat{\beta}_2) = \beta_2$ y $E(\hat{\beta}_1) = \beta_1$.
- 2) Tiene varianza mínima dentro de la clase de todos los estimadores lineales insesgados.
- 3) Presentan consistencia; es decir, a medida que el tamaño de la muestra aumenta indefinidamente, los estimadores convergen hacia sus verdaderos valores poblacionales.
- 4) Al ser funciones lineales de u_i , $\hat{\beta}_1$ y $\hat{\beta}_2$ están normalmente distribuidas:

$$\hat{\beta}_1 \sim N(\beta_1, \sigma_{\beta_1}^2)$$

$$\hat{\beta}_2 \sim N(\beta_2, \sigma_{\beta_2}^2)$$

donde $\sigma_{\beta_1}^2$ y $\sigma_{\beta_2}^2$ corresponden respectivamente a $Var(\hat{\beta}_1)$ y $Var(\hat{\beta}_2)$. Estas varianzas tienen las siguientes expresiones: $\sigma_{\beta_1}^2 = \sigma^2 \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}$ y $\sigma_{\beta_2}^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$, recordando que σ^2 es la varianza del error u_i .

- 5) Como se mencionó en la sección anterior el término $(n - 2)(\hat{\sigma}^2/\sigma^2)$ está distribuido como una variable χ^2 (ji-cuadrada), con $(n - 2)$ grados de libertad.

6) $\hat{\beta}_1$ y $\hat{\beta}_2$ se distribuyen de manera independiente respecto de $\hat{\sigma}^2$.

Teorema²: Sea Z_1 una variable normal estándar y Z_2 una variable χ^2 con k grados de libertad e independiente de Z_1 , entonces la variable definida como:

$$t = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}},$$

sigue la distribución t de Student con k grados de libertad. A continuación se calculan los intervalos de confianza de los parámetros utilizando el teorema y las propiedades anteriores.

Para el intervalo de confianza de β_2 usamos $Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\beta_2}}$ y $Z_2 = (n - 2)\frac{\hat{\sigma}^2}{\sigma^2}$, ya que por la propiedad 4, al ser $\hat{\beta}_2$ una variable normalmente distribuida entonces Z_1 se distribuye como una variable normal estándar. Por la propiedad 5 la variable Z_2 tiene una distribución ji-cuadrada con $n - 2$ grados de libertad. Además utilizando la propiedad 6 se llega a que Z_1 es independiente de Z_2 , por lo tanto, la variable t queda expresada de la siguiente forma:

$$t = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}} = \frac{(\hat{\beta}_2 - \beta_2)}{\hat{\sigma}_{\beta_2}},$$

en donde $\hat{\sigma}_{\beta_2} = \frac{\hat{\sigma}}{\sqrt{\sum(X_i - \bar{X})^2}}$, es el estimador de la desviación estándar de $\hat{\beta}_2$. Utilizando los percentiles de la distribución t podemos crear un intervalo como este:

$$P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha, \quad (1.6)$$

donde $t_{\alpha/2}$ es tal que $P(t > t_{\alpha/2}) = \alpha/2$. Al sustituir t y reorganizar los términos el intervalo de confianza queda de la siguiente forma:

$$P(\hat{\beta}_2 - t_{\alpha/2}\hat{\sigma}_{\beta_2} \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\hat{\sigma}_{\beta_2}) = 1 - \alpha, \quad (1.7)$$

por lo tanto, el intervalo de confianza al $(1 - \alpha)\%$ para β_2 es:

$$(\hat{\beta}_2 - t_{\alpha/2}\hat{\sigma}_{\beta_2}, \hat{\beta}_2 + t_{\alpha/2}\hat{\sigma}_{\beta_2}). \quad (1.8)$$

La interpretación de este intervalo de confianza es: dado el coeficiente de confianza de 95 %, en 95 de cada 100 casos, los intervalos como (1.8) contendrán al verdadero valor de β_2 . Pero no se puede afirmar que la probabilidad de que el intervalo específico de la ecuación contenga al verdadero β_2 sea de 95 %, porque este intervalo es ahora fijo y no aleatorio; por consiguiente, β_2 se encontrará o no dentro de él: la probabilidad de que el intervalo fijo específico incluya al verdadero valor de β_2 es por consiguiente 1 ó 0. De manera análoga se obtiene el intervalo para β_1 , el cual es el siguiente:

$$(\hat{\beta}_1 - t_{\alpha/2}\hat{\sigma}_{\beta_1}, \hat{\beta}_1 + t_{\alpha/2}\hat{\sigma}_{\beta_1}), \quad (1.9)$$

²Puede verse en Gujarati, D. N. y Porter, D. C., 2010, *Econometría*, (Pilar Carril, trad.). México: Mc Graw Hill, p. 144.

en donde $\hat{\sigma}_{\beta_1} = \hat{\sigma} \sqrt{\frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2}}$, es el estimador de la desviación estándar de σ_{β_1} . El intervalo de confianza para σ^2 se obtiene usando los percentiles de la distribución χ^2 ya que la variable $\chi^2 = (n - 2)(\hat{\sigma}^2/\sigma^2)$ tiene una distribución χ^2 , por tanto:

$$P(\chi_{\alpha/2}^2 \leq \chi^2 \leq \chi_{1-\alpha/2}^2) = 1 - \alpha$$

$$P\left[(n - 2) \frac{\hat{\sigma}^2}{\chi_{1-\alpha/2}^2} \leq \sigma^2 \leq (n - 2) \frac{\hat{\sigma}^2}{\chi_{\alpha/2}^2}\right] = 1 - \alpha,$$

quedando el intervalo de confianza de σ^2 al $(1 - \alpha)\%$ de la siguiente manera:

$$\left((n - 2) \frac{\hat{\sigma}^2}{\chi_{1-\alpha/2}^2}, (n - 2) \frac{\hat{\sigma}^2}{\chi_{\alpha/2}^2} \right).$$

Los intervalos de confianza se usan como un método para hacer pruebas de hipótesis aunque no es el único método. Una prueba de hipótesis es un procedimiento basado en evidencia de la muestra y la teoría de probabilidad para determinar si la hipótesis es una afirmación razonable. En una prueba de hipótesis suele probarse la hipótesis nula denotada por H_0 frente a una hipótesis alternativa denotada con H_1 . Por ejemplo se pueden plantear las hipótesis:

$$H_0 : \beta_2 = 0.5 \quad vs \quad H_1 : \beta_2 \neq 0.5. \quad (1.10)$$

Se conoce como *error tipo I* aceptar H_0 dado que sea falsa. El nivel de significancia α se define como la probabilidad cometer un *error tipo I*. Para contrastar H_0 contra H_1 se usa una *regla de decisión*, bajo el enfoque del método del intervalo de confianza la regla es: construir un intervalo de confianza para β_2 a $100 \times (1 - \alpha)\%$, si 0.5 (es decir, el valor de β_2 en H_0) se encuentra dentro de este intervalo de confianza no se rechaza H_0 , pero si está fuera del intervalo se rechaza H_0 .

Dentro de las pruebas de hipótesis existe otro enfoque llamado prueba de significancia. Esta consiste en que la decisión de aceptar o rechazar H_0 se toma con base en el valor del estadístico de prueba obtenido con los datos disponibles. Por ejemplo, si se quiere probar la hipótesis en (1.10) se toma la ecuación (1.6) y se transforma en la siguiente:

$$P(\beta_2^* - t_{\alpha/2} \hat{\sigma}_{\beta_2} \leq \hat{\beta}_2 \leq \beta_2^* + t_{\alpha/2} \hat{\sigma}_{\beta_2}) = 1 - \alpha, \quad (1.11)$$

donde β_2^* es el valor de β_2 según H_0 (0.5 en este ejemplo). Nótese que (1.11) es similar a la ecuación (1.7) con la diferencia que en medio de la desigualdad está el estadístico $\hat{\beta}_2$. Con esto se obtiene un intervalo de confianza para la variable aleatoria $\hat{\beta}_2$ para un nivel de significancia α . Esta prueba puede interpretarse de la siguiente manera: Si se toma como cierta la hipótesis H_0 , entonces la probabilidad de que la variable aleatoria $\hat{\beta}_2$ esté en el intervalo $(\beta_2^* - t_{\alpha/2} \hat{\sigma}_{\beta_2}, \beta_2^* + t_{\alpha/2} \hat{\sigma}_{\beta_2})$ es de $(1 - \alpha)\%$. Si al calcular la variable $\hat{\beta}_2$, esta se encuentra dentro del intervalo, entonces no se rechaza H_0 , de no estar dentro del intervalo se rechaza tal hipótesis.

Sin embargo, no es necesario obtener un intervalo de confianza para $\hat{\beta}_2$. Para probar la misma

hipótesis en (1.10) se calcula el valor de $t = \frac{(\hat{\beta}_2 - \beta_2)}{\hat{\sigma}_{\beta_2}}$ sustituyendo el valor de β_2 dado en la hipótesis nula, si t cae entre los percentiles (también llamados valores críticos t) dados en (1.6) entonces no se rechaza la hipótesis nula, y por el contrario de no estar dentro de ellos indica que hay elementos significativos para pensar que el valor hipotético de β_2 no es compatible con los datos de la muestra. Al usar la distribución t a dicha prueba se le llama la prueba t . En el análisis de regresión es importante determinar si el valor de β_2 pudiera ser igual a cero ya que en ese caso la variable independiente no estaría afectando de ninguna manera a la variable dependiente, y el análisis no tendría sentido.

1.5. Análisis de varianza

La comparación simultánea de varias medias poblacionales se denomina análisis de varianza (ANOVA). En el caso de la regresión lineal se emplea para probar la hipótesis de que todos los coeficientes de regresión (a excepción del intercepto) son iguales a cero, en cuyo caso, como se mencionó, el análisis no tendría sentido. Por tanto, el análisis de varianza en la regresión lineal simple proporciona una forma alterna pero complementaria de probar la hipótesis nula $\beta_2 = 0$. Para realizar dicha prueba se emplea la distribución F que tiene las siguientes características:

- La distribución F no puede ser negativa: El menor valor que F puede tomar es cero.
- Es una distribución continua: Supone un número infinito de valores entre el cero y el infinito positivo.
- Es sesgada de manera positiva: La cola larga de la distribución es hacia el lado derecho. Cuando el número de grados de libertad aumenta, tanto en el numerador como en el denominador, la distribución se aproxima a ser normal.
- Es asintótica: Cuando los valores de X aumentan, la curva F se aproxima al eje X pero nunca lo toca.

Teorema: Si Z_1 y Z_2 son variables ji-cuadrada independientemente distribuidas, con k_1 y k_2 gl respectivamente, entonces la variable:

$$F = \frac{Z_1/k_1}{Z_2/k_2}, \quad (1.12)$$

tiene una distribución F con k_1 y k_2 grados de libertad. Gujarati y Porter demuestran que $Z_1 = \sum(\hat{Y}_i - \bar{Y})^2$ (la SCE) y $Z_2 = \sum(Y_i - \hat{Y}_i)^2$ (la SCR) son variables aleatorias con 1 y $(n - 2)$ gl respectivamente. Si se añade la hipótesis nula de que $\beta_2 = 0$, entonces, la variable F en (1.12) ahora sustituida como

$$F = \frac{SCE/1}{SCR/(n - 2)},$$

tiene asociada una distribución F con 1 gl en el numerador y $(n - 2)$ gl en el denominador.

La idea intuitiva del análisis de varianza es obtener dos expresiones distintas de varianzas asociadas al modelo de regresión para después obtener la razón de dichas estimaciones y si esta razón es aproximadamente 1, entonces las dos estimaciones son iguales y se concluye que las medias poblacionales son iguales. En este caso el denominador como se vio en la sección 1.1 es el estimador insesgado para σ^2 . Mientras que el numerador (la SCE) puede demostrarse³ que tiene un valor esperado de

$$\sigma^2 + \beta_2^2 \sum_{i=1}^n (X_i - \bar{X})^2$$

Bajo la hipótesis de que $\beta_2 = 0$ el valor esperado de la SCE es también σ^2 . Por tanto, en caso de que la hipótesis nula sea cierta tendríamos dos estimadores insesgados diferentes para el mismo parámetro σ^2 y la razón F debería ser cercana a 1, pero si la hipótesis nula es rechazada, entonces el cociente es mayor que 1. Es por ello que la región de rechazo de la hipótesis nula está en la cola derecha de la distribución.

La distribución F sirve como un árbitro para indicar en qué instancia la razón de las varianzas muestrales es mucho mayor que 1, el procedimiento es buscar el percentil de $F_{(1,n-2)}$ (con un gl en el numerador y $n - 2 gl$ en el denominador) con el nivel de significancia α elegido tal que:

$$P(F > F_{(1,n-2)}) = \alpha.$$

La regla de decisión es rechazar H_0 si F rebasa el cuantil asociado a la prueba $F_{(1,n-2)}$, ya que en este valor crítico la probabilidad de rechazar H_0 dado que sea cierta es el $100\alpha\%$ y si F es mayor entonces la probabilidad de cometer dicho error será aún menor.

Otra herramienta para decidir aceptar o rechazar una hipótesis nula, tanto para la prueba t como para la prueba F, es el *valor p* (o *p value*) del estadístico de prueba. Este se define como la probabilidad exacta de cometer un error tipo I dado un valor del estadístico. Por ejemplo, para un análisis de regresión con 13 observaciones y un valor $t = 10.3428$ se obtiene un *valor p* = 0.00009. Es decir que la probabilidad de rechazar $H_0 : \beta_2 = 0$ dado que es verdadera, para el valor $t = 10.3428$ de una variable aleatoria con distribución t con 11 gl es 0.00009. En la práctica suele fijarse el nivel de significancia α y rechazar la hipótesis nula si el *valor p* es menor que α , la ventaja de obtener el *valor p* es se tiene más libertad de elegir α asumiendo mayor o menor riesgo de equivocación. En este ejemplo se rechaza H_0 si se fija α en 0.05, pero si se es muy adverso al riesgo también puede rechazarse H_0 eligiendo $\alpha = 0.001$.

En el cuadro 1.1 se muestra la tabla ANOVA la cual resume la información del análisis de varianza. En esta se muestran las fuentes de variación: debida a la regresión, debida al residuo y la variación total. En la columna gl los grados de libertad correspondientes. En la columna SC las sumas de cuadrados asociadas a cada fuente de variación, como se vio en la Sección 1.2 la *SCE* se asocia con la variación explicada por la regresión, la *SCR* se asocia con la variación que explica el residuo y la *SCT* se asocia con la variación total de la variable. En la columna SCM la suma de cuadrados promedio de cada una de las fuentes de variación, que

³Véase K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, Nueva York, 1960, pp. 278-280.

Fuente de variación	gl	SC	SCM	F
Regresión	1	SCE	$\frac{SCE}{1}$	$\frac{SCE/1}{SCR/(n-2)}$
Residuo o error	$n - 2$	SCR	$\frac{SCR}{(n-2)}$	
Total	$n - 1$	SCT		

Cuadro 1.1: Tabla ANOVA

corresponde a la suma de cuadrados de la fuente entre los grados de libertad correspondientes. Por último el estadístico F que corresponde ser la división de la suma de cuadrados medios de la regresión entre la suma de cuadrados medios del residuo.

1.6. Detección de autocorrelación

Uno los supuestos del MCRL es la no autocorrelación de las perturbaciones. Su detección es importante pues en presencia de autocorrelación los estimadores de MCO, a pesar de ser lineales, insesgados y tener distribución asintóticamente normal (es decir, en muestras grandes), dejan de tener varianza mínima entre todos los estimadores lineales insesgados. En resumen, no son eficientes en relación con los demás estimadores lineales e insesgados. Dicho de otro modo, es posible que no sean los Mejores Estimadores Lineales Insesgados (MELI). Como resultado, las pruebas usuales t y F pueden no ser válidas.

Prueba d de Durbin-Watson

Una de las pruebas más conocidas para detectar correlación serial es la prueba Durbin-Watson. La hipótesis nula es siempre la no autocorrelación⁴, es decir $H_0 : \rho = 0$, donde a ρ se conoce como coeficiente de autocorrelación, y se usa el estadístico d definido como:

$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_i^2}$$

en el que puede verse que el numerador del estadístico el número de observaciones es $n - 1$ porque se pierde una observación al obtener las diferencias consecutivas. Gujarati y Porter (2010) explican que la construcción del estadístico d se basa en los siguientes supuestos:

- 1) El modelo de regresión incluye el término del intercepto. Si dicho término no está presente (a lo que se le conoce como regresión a través del origen) es esencial efectuar de nuevo la regresión con dicho término para obtener la SCR.
- 2) Las variables explicativas, X , son no estocásticas, es decir, son fijas en muestreo repetido.

⁴Anderson, Sweeney, Williams, Camm, y Cochran, *Statistics for Business & Economics*, Boston, Cengage Learning, 2015, pp.793-796

- 3) Las perturbaciones u_t se generan mediante el esquema autorregresivo de primer orden o modelo $AR(1)$:

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad (1.13)$$

donde ε_t es término de error de ruido blanco, el cual satisface: $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_{\varepsilon_t}^2$ y $Cov(\varepsilon_t, \varepsilon_{t+s}) = 0$ para $s \neq 0$. Por tanto, no se puede utilizar para detectar esquemas autorregresivos de orden superior.

- 4) Se supone que el término de error u_t está normalmente distribuido.
 5) No hay observaciones faltantes en los datos.

A diferencia de las pruebas t o F , en esta prueba no hay un valor crítico único que lleve al rechazo o a la aceptación de la hipótesis nula de que no hay correlación serial de primer orden en las perturbaciones u_i . Sin embargo, existen unos límites superior d_U e inferior d_L tales que si el estadístico d cae por fuera de estos valores críticos, puede tomarse una decisión respecto de la presencia de correlación serial positiva o negativa. Por el contrario si el estadístico cae dentro del intervalo la prueba no es concluyente y no puede tomarse una decisión. Estos límites dependen del número de observaciones n , del nivel de confianza α y del número de variables explicativas, mas no de los valores que adquieren estas variables explicativas.

Debido a que las sumas $\sum_{t=2}^n \hat{u}_t$ y $\sum_{t=2}^n \hat{u}_{t-1}$ difieren solo en un término puede expresarse d como la siguiente aproximación:

$$d \approx 2 \left(1 - \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2} \right),$$

y si se define

$$\hat{\rho} = \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2},$$

como el coeficiente de autocorrelación muestral, entonces la aproximación para d puede expresarse como:

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

Ya que ρ tiene como límites $-1 \leq \rho \leq 1$, el estadístico d toma los valores $0 \leq d \leq 4$. En donde 0 implica una autocorrelación positiva perfecta, 4 una autocorrelación negativa perfecta y 2 no hay autocorrelación. Las reglas de decisión se muestran en el cuadro 1.2⁵.

Como se observa, la prueba no solo indica si existe o no correlación serial entre las perturbaciones, sino también en caso de haber autocorrelación, si esta es positiva o negativa. La autocorrelación positiva se presenta cuando a una perturbación positiva sigue una perturbación también positiva, o a una negativa sigue una negativa. Las series de tiempo económicas por lo general muestra autocorrelación positiva, pues en ellas los datos se desplazan hacia arriba o hacia abajo en extensos periodos. La autocorrelación negativa se presenta cuando a una perturbación positiva sigue una perturbación negativa, es decir exhiben un movimiento ascendente y descendente constante.

⁵Anderson, et al., *op. cit.*, pp.795-796

Hipótesis alternativa	Rango	Decisión
$H_1 : \rho > 0$	$d < d_L$	Hay autocorrelación positiva
	$d_L \leq d \leq d_U$	La prueba no es concluyente
	$d_U < d$	No hay evidencia de autocorrelación positiva
$H_1 : \rho < 0$	$d > 4 - d_L$	Hay autocorrelación negativa
	$4 - d_U \leq d \leq 4 - d_L$	La prueba no es concluyente
	$d < 4 - d_U$	No hay evidencia de autocorrelación negativa
$H_1 : \rho \neq 0$	$d < d_L$	Hay autocorrelación positiva
	$d > 4 - d_L$	Hay autocorrelación negativa
	$d_L \leq d \leq d_U$	La prueba no es concluyente
	$4 - d_U \leq d \leq 4 - d_L$	La prueba no es concluyente
	$d_U \leq d \leq 4 - d_U$	No hay evidencia de autocorrelación

Cuadro 1.2: Reglas de decisión de la prueba d

En algunas ocasiones los residuales no están correlacionados con un retraso de orden uno como en (1.13), sino que pueden estar correlacionados a ordenes mayores, por lo que pueden usarse otras pruebas.

La prueba Breusch-Godfrey

Una prueba más general de autocorrelación es la prueba de Breusch-Godfrey en el sentido de que permite como variables explicativas los valores rezagados de la variable dependiente⁶ y considera esquemas autorregresivos de orden mayor. De hecho, en el contexto de series de tiempo su uso es preferible al de la prueba Durbin-Watson. La prueba parte de un modelo de regresión simple:

$$Y_t = \beta_1 + \beta_2 X_t + u_t, \quad (1.14)$$

en el que el término u_t tiene un comportamiento autorregresivo de orden p (AR(p)), esto es:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$$

en donde la variable ε_t se distribuye normal y no está autocorrelacionada, y la hipótesis nula de la prueba es:

$$H_0 = \rho_1 = \rho_2 = \cdots = \rho_p = 0,$$

es decir que no existe correlación de ningún orden. Para obtener el estadístico de prueba existen dos pasos previos, el primero es estimar por MCO el modelo en (1.14) y obtener los

⁶Gujarati y Porter no mencionan explícitamente como hacer dicha prueba pero sugieren los siguientes artículos: L.G. Godfrey, "Testing Against General and Moving Average Error Models When the Regressor include Lagged Dependent Variable", *Econometrica*, Vol. 46, 1978, pp. 1293-1302 y T.S. Breusch, "Testing for Autocorrelation in Dynamic Models", *Australian Economic Papers*, vol. 17, 1978, pp. 334-355.

residuos \hat{u}_t , el segundo es estimar una regresión auxiliar siguiendo este esquema:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \dots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t. \quad (1.15)$$

El estadístico de prueba es:

$$(n - p)R^2 \sim \chi_p^2,$$

donde n es el número de observaciones, p el orden autorregresivo escogido y R^2 es el coeficiente de determinación de la regresión auxiliar en (1.15). Bajo la hipótesis nula tiene una distribución χ^2 con p grados de libertad, de manera que fijando un error $\alpha = 0.05$, si el estadístico es más grande que el percentil al 95% de confianza (o bien se obtiene un *valor p* menor a 0.05) se rechaza la hipótesis nula, con lo que al menos una autocorrelación será distinta de cero.

1.7. Aplicación

Este primer ejemplo pretende ilustrar los conceptos de regresión y verificación de supuestos vistos en el capítulo. En la siguiente tabla se muestran las ventas mensuales de un negocio⁷ de Julio 2010 a Diciembre 2014.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010							297.41	336.36	301.57	325.61	299.45	409.27
2011	331.66	293.64	314.65	284.23	310.00	317.13	371.82	335.91	351.68	322.03	287.19	437.06
2012	355.62	322.89	305.56	325.37	350.08	327.10	362.95	361.14	361.54	360.99	363.74	469.71
2013	423.83	348.76	332.52	339.63	352.11	337.41	338.54	359.51	326.63	366.91	357.81	496.83
2014	404.69	337.32	316.56	340.23	346.21	356.49	354.83	377.24	326.42	359.16	338.84	465.90

Cuadro 1.3: Ventas mensuales

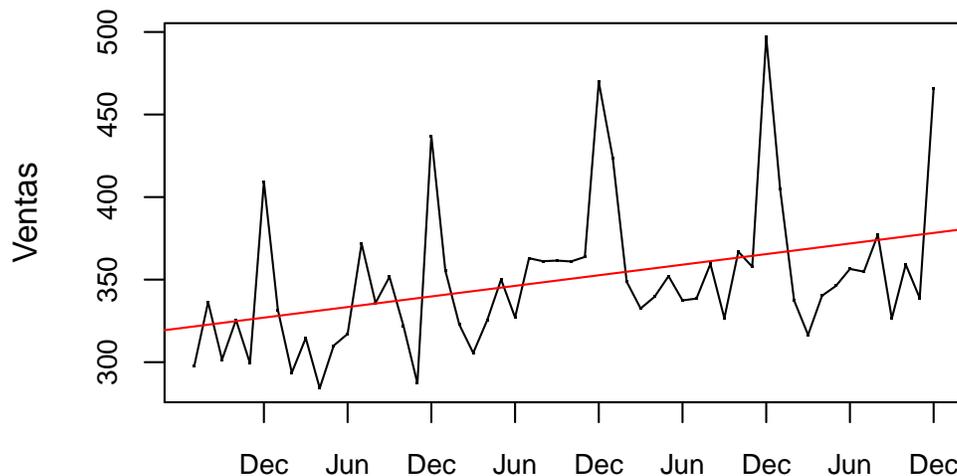


Figura 1.1: Ventas mensuales

⁷Los datos están transformados a cierta escala, sin afectar la interpretación de los resultados obtenidos.

Por tanto si se toma al tiempo como la variable explicativa de forma que $X_i = \text{Mes } i$, y las ventas como variable dependiente de forma que $Y_i = \text{Venta del mes } i$, con $i = \{1, \dots, 54\}$, entonces se puede pensar en ajustar un modelo de regresión lineal simple con el objetivo de explicar las ventas de cierto mes i en función del tiempo.

El valor esperado de las ventas de un mes i (Y_i) dado el mes i (X_i) es $f(X_i) = \beta_1 + \beta_2 X_i$, la llamada función de regresión poblacional (*FRP*). El término independiente β_1 puede interpretarse como la venta del mes cero (Junio 2010), mientras que β_2 puede verse como el incremento de la venta promedio por mes. La Figura 1.1 es la gráfica de dichas ventas con respecto al tiempo. En ella se observa que la relación entre estas dos variables no es exactamente lineal, pero no se descarta el modelo ya que conforme transcurre el tiempo las ventas aumentan, y parece haber una tendencia creciente en las mismas.

El Modelo Clásico de Regresión Lineal (*MCRL*) indica que las ventas del mes i se componen de la FRP más un componente estocástico denotado u_i :

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \quad i = \{1, \dots, 54\}. \quad (1.16)$$

Debido a que se trabaja con una muestra de los datos, el MCRL se estima con la función de regresión muestral *FRM* y un error muestral, con lo que el modelo a probar es el siguiente:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i, \quad i = \{1, \dots, 54\}. \quad (1.17)$$

Para la estimación de parámetros del análisis de regresión lineal simple se utilizó el software estadístico R. Una vez ingresado el modelo, el resumen de los resultados muestra la siguiente salida:

```
> regresion<-lm(serie~mes)
> summary(regresion)
```

Call:

```
lm(formula = serie ~ mes)
```

Residuals:

```
    Min      1Q  Median      3Q     Max
-52.10 -23.07 -15.46  12.16 131.37
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 320.5611    11.4017  28.115 < 2e-16 ***
mes          1.0690     0.3607   2.964  0.00458 **
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 41.31 on 52 degrees of freedom

Multiple R-squared: 0.1445, Adjusted R-squared: 0.128

F-statistic: 8.783 on 1 and 52 DF, p-value: 0.004579

El estimador para β_1 (la venta del mes cero) es $\hat{\beta}_1 = 320.56$, en este punto es adecuado preguntarse: ¿es posible que el intercepto sea cero y en realidad no sea necesario este parámetro en el modelo? Para responder a la pregunta se recurre a la prueba t , fijando primero un nivel de significancia del 5%. Bajo la hipótesis $H_0 : \beta_1 = 0$, el valor t es: $t=28.115$, el cual no se encuentra en el intervalo $(-t_{0.025}, t_{0.025})$ que es $(-2.006, 2.006)$, por lo que se rechaza dicha hipótesis. Por tanto el parámetro β_1 que representa la venta inicial tiene sentido en el modelo.

El estimador para β_2 (la pendiente de crecimiento mensual) es $\hat{\beta}_2 = 1.0690$. Para este parámetro es aún más importante saber si su valor puede ser cero, pues de serlo, el tiempo sería un factor que no explica el comportamiento promedio de las ventas y el modelo no tendría sentido. La probabilidad de rechazar $H_0 : \beta_2 = 0$ dado que es verdadera, para el valor $t = 2.964$ de una variable aleatoria t con 52 *gl* es 0.00458 o bien 0.458 %, dado que se fijó un error máximo del 5% se rechaza la hipótesis nula. Por tanto el modelo tiene sentido pues el tiempo “influye” sobre el comportamiento de las ventas.

Véase que se aplicó la prueba t para β_1 mediante la comparación con el valor crítico t y para β_2 mediante el valor p , pero pueden usarse indistintamente y obtener el mismo resultado en la inferencia. Por otro lado, como se vio en la sección 1.3, se puede hacer inferencia sobre los parámetros a través de sus intervalos de confianza descritos en las ecuaciones (1.8) y (1.9) los cuales son los siguientes, a un 95 % de confianza:

```
> confint(regresion, level=.95)
                2.5 %      97.5 %
(Intercept) 297.6818392 343.440327
mes          0.3451773   1.792791
```

Como en ninguno de estos intervalos está el cero, se comprueba el hecho de que se rechazan las hipótesis nulas $\beta_1 = 0$ y $\beta_2 = 0$.

El estimador para la desviación residual es $\hat{\sigma} = 41.31$. Respecto a la bondad de ajuste el resumen de R muestra que $R^2 = 0.1445$, es decir que el 14.45 % de la variación de las ventas es explicada por el modelo, en otras palabras el avance del tiempo explica aproximadamente el 14 % de la variación de las ventas, el otro 86 % de variación restante lo explica el componente aleatorio. Este porcentaje no representa un buen ajuste y es atribuible al hecho de que los residuales \hat{u}_t están diseñados para captar la variación de efectos que son aleatorios de variables que no conocemos como gasto en publicidad, compras o número de clientes diarios. Sin embargo, en los datos existen elementos claramente no aleatorios, véase que las ventas siguen un patrón anual, este efecto llamado estacionalidad será de vital importancia en el estudio de estos datos a lo largo de este trabajo. También debe considerarse eliminar de la serie la inflación y el efecto calendario, puntos a tratar en el siguiente capítulo.

Sobre el análisis de varianza el resumen muestra el estadístico F y el *valor p* del modelo, utilizando este último la interpretación es la siguiente: al fijar un error del 5 %, si se rechaza la hipótesis nula $H_0 : \beta_2 = 0$ la probabilidad de equivocarse es de 0.46 %, como es menor al error permitido puede rechazarse tal hipótesis. Como se mencionó en el caso de la regresión

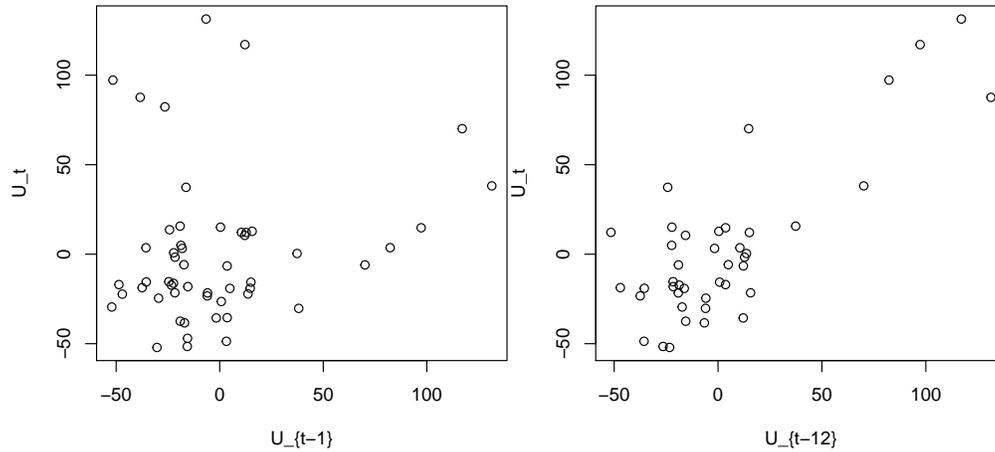


Figura 1.2: Gráfica de residuos contra residuos retrasados 1 y 12 periodos

lineal simple, el análisis de varianza probó la misma hipótesis que la prueba t para el parámetro β_2 reafirmando la decisión.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
mes	1	14989.72	14989.72	8.78	0.0046
Residuals	52	88747.87	1706.69		

Cuadro 1.4: Tabla ANOVA del modelo

En la tabla ANOVA del modelo (Cuadro 1.4) puede verse que la suma de los cuadrados de los residuos es mayor que la suma de los cuadrados de la regresión, algo que ya se había visto en el cálculo de R^2 . Pero al tomar en cuenta los grados de libertad, la variable explicativa “mes” tiene mayor suma de cuadrados media que los residuos, estos valores son estimaciones para $\sigma^2 + \beta_2^2 \sum_{i=1}^n (X_i - \bar{X})^2$ y σ^2 respectivamente. De hecho es 8.78 veces mayor, cantidad que representa el estadístico F . Para una variable con distribución F de 1 y 52 *gl*, el 95% de las observaciones se encuentran por debajo del valor crítico 4.03, pero el estadístico $F = 8.78$ indica que es improbable (con un nivel de significancia del 5%) que la muestra provenga de una distribución F , y por tanto se rechaza la hipótesis nula $\beta_2 = 0$.

La validez de las pruebas de hipótesis mencionadas depende de la validación de supuestos entre ellos la no autocorrelación entre los residuales. La figura 1.2 muestra la relación de los residuales retrasados 1 y 12 lugares, en la primera podría existir autocorrelación aunque el efecto es mucho más evidente entre los residuales desplazados 12 lugares. Para confirmar se aplican las pruebas d y Breusch-Godfrey con las funciones `dwtest` y `bgtest` del software R, en la librería `lmtest`.

```
> library(lmtest)
> dwtest(regresion, alternative="two.sided")
```

Durbin-Watson test

```

data:  regresion
DW = 1.6035, p-value = 0.1042
alternative hypothesis: true autocorrelation is not 0

```

Consultando en tablas de del estadístico para $n = 14$, una variable explicativa y $\alpha = 0.05$ los límites son $d_L = 1.528$ y $d_U = 1.601$. Para probar la hipótesis de que no hay autocorrelación de los residuos, debe ocurrir $d_U < d < 4 - d_U$, lo cual ocurre pues $1.601 < 1.6035 < 2.399$, por apenas muy poco la prueba concluye que no hay autocorrelación. Como se dijo anteriormente la prueba Durbin Watson puede identificar autocorrelación cuando se presenta como en 1.13, y bajo el supuesto de que éstos se distribuyen de manera normal. Aunque en este caso la prueba concluyó que no existe autocorrelación, en el gráfico Q-Q plot de los residuales estandarizados (figura 1.3), se observa que los puntos del gráfico no se acercan a la línea de identidad, en particular en el centro y parte derecha de la escala. Esto lleva a pensar que los residuales no son normales, y por tanto, la prueba Durbin Watson podría no reflejar resultados válidos. Más detalles sobre esta y otras pruebas de normalidad pueden consultarse en la sección (3.7).

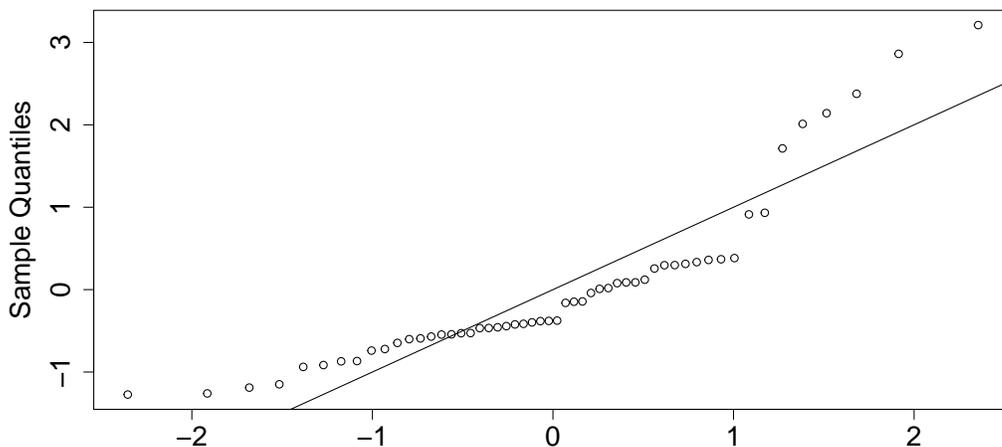


Figura 1.3: Gráfico Q-Q plot de los residuales estandarizados

```
> bgtest(regresion,order=12,type="Chisq")
```

Breusch-Godfrey test for serial correlation of order up to 12

```

data:  regresion
LM test = 32.153, df = 12, p-value = 0.00131

```

Al aplicar la prueba Breusch-Godfrey de orden $p = 12$ y $\alpha = 0.05$, los resultados indican que el estadístico 1.6 es muy grande para provenir de una distribución ji-cuadrada con 12 grados de libertad, por lo que en efecto alguna autocorrelación $\hat{\rho}_i$ con $i = \{1 \dots, 12\}$ es distinta de cero. De hecho es muy probable que sea la de orden 12, pues la misma prueba para órdenes menores no mostró autocorrelación.

Debido a que existe algún tipo de autocorrelación temporal, las pruebas t y F no proporcionan conclusiones válidas. La idea de que el incremento en las ventas por trimestre se explica por el paso del tiempo no es realista. El tiempo en realidad no explica por sí mismo de manera directa las variaciones de cualquier variable que se recoja a través del tiempo, más bien dichas variaciones son producto de otras variables como el gasto en el resurtido de mercancía, eventos específicos que ocurren en el año, por ejemplo las vacaciones, fiestas decembrinas, inicio o fin de ciclo escolar, etc.; el tiempo puede usarse para establecer patrones de regularidad en el comportamiento de cierta variable.

Basta con ver la figura 1.1 para notar que el modelo no podría predecir confiablemente los datos. Por supuesto este ejemplo fue tomado a propósito a manera de introducción de los datos y conceptos que se estarán manejando, muchos de los cuales se retomarán en capítulos posteriores, especialmente en el capítulo 4.

Los siguientes capítulos abordan de mejor manera el tratamiento de datos de tipo series de tiempo, en el siguiente capítulo incluso se observa como el análisis de regresión puede usarse cómo un método de apoyo para poder captar la tendencia en el análisis clásico de series de tiempo.

Capítulo 2

Análisis clásico de series temporales

2.1. Conceptos básicos del análisis clásico de series temporales

2.1.1. Definición de series de tiempo

Una serie de tiempo se puede definir como “una sucesión de observaciones correspondientes a una variable en distintos momentos de tiempo. En general, las diferentes observaciones se obtienen a intervalos regulares y de duración constante. Así las series pueden tener periodicidad anual, semestral, mensual, trimestral, etc., según los periodos de tiempo en que vengan recogidos los datos que la componen”¹. Ejemplos de series de tiempo son las ventas trimestrales de una empresa, el número de clientes por hora que llegan a un negocio, el precio diario del dólar o la producción anual de maíz a nivel nacional.

En general conviene que los intervalos de tiempo entre observaciones consecutivas sean de la misma longitud para evitar los llamados problemas de calendario que suceden, por ejemplo, por el hecho de que no todos los meses tienen el mismo número de días y también porque algunas festividades no caen todos los años en las mismas fechas, este y otros temas relacionados al tratamiento de los datos se tratarán más adelante en esta misma sección.

El estudio de las series temporales es una de las técnicas más habituales cuando el objetivo es la predicción de valores futuros de la variable, aunque también puede utilizarse con el único propósito de describir la historia de una determinada variable. Es necesario el análisis de los datos para obtener conclusiones más fiables ya que de ello dependen decisiones, pues si los datos se tratan directamente pueden tener una interpretación errónea, induciendo a conclusiones y decisiones equivocadas.

Morilla (2000) explica que existen análisis que tienen distintas metodologías, cada una trabaja bajo cierto conjunto de hipótesis, lo que da lugar a los siguientes enfoques:

¹Morilla, *Análisis de Series Temporales*, Madrid, La Muralla, 2000, p. 12.

Enfoque determinista: el enfoque determinista supone que la variable presenta un patrón fijo, por lo que las irregularidades de la serie son vistas como una desviación respecto a una pauta de comportamiento sistemático.

Enfoque estocástico: en el enfoque estocástico se percibe la generación de la serie a través de la teoría de los procesos estocásticos. La metodología Box-Jenkins es la de más difusión en este enfoque.

Bajo este contexto, este capítulo habla del análisis clásico de series temporales, el cual tiene un enfoque determinista pues considera a la serie integrada por cuatro componentes: tendencia, ciclo, estacionalidad y componente irregular, de las que se hablará con más detalles en la siguiente sección.

En esta metodología se asume que las observaciones de la serie están autocorrelacionados. Por ejemplo, si la ocupación hotelera ha crecido de un año a otro es posible esperar que el siguiente año también crezca, o bien si las ventas del mes de diciembre han sido históricamente las más altas, por qué no esperar que en diciembre del año actual se siga comportando igual. No tomar en cuenta la autocorrelación puede llevar a conclusiones erróneas, su tratamiento es parte del proceso de modelación.

También Morilla (2000) comenta que el análisis de series temporales abarca las siguientes etapas:

1. Tratamiento de los datos. Esta es la parte previa a la utilización de cualquier técnica, en ésta se considera la naturaleza cronológica de los datos. Entre los aspectos que se consideran están:
 - a) Los cambios en el valor del dinero, pues repercuten en aquellas series expresadas en términos monetarios ya que el valor del dinero varía a causa de la inflación. Para poder trabajar con los datos es necesario expresarlos de manera en que todos ellos sean comparables a través del tiempo con un período base, a este proceso se les denomina deflatación. Básicamente consiste en eliminar el efecto de la inflación dividiendo cada elemento entre su correspondiente índice de precios, usualmente el índice de precios al consumidor. No se considera con ello que la inflación sea constante pues el índice de precios se obtiene de manera mensual, por lo que es una serie de tiempo como también lo es la inflación.
 - b) Los efectos de calendario, los cuales se refieren al problema que surge por el hecho de que no todos los meses tienen el mismo número de días, o que no todos los días festivos coincidan el mismo día todos los años. En una serie de ventas mensual, las ventas en los meses con 31 días podrían verse más altas que aquellos con 30 días, por el simple hecho de tener un día más de venta. Lo que suele hacerse en la práctica es multiplicar por un factor de ajuste para que todos los meses parezcan tener el mismo número de días (30 en general). Así, por ejemplo, de los 31 días de Enero se toman solo treinta para el análisis, lo que se consigue multiplicando las ventas de Enero por el factor (30/31).

c) Un punto que adicional en el pre-procesamiento de datos es el de los valores atípicos (*outliers*) y los valores faltantes (*missing values*). Sobre este tema Morilla (2000) comenta que “conviene eliminar o no tomar en cuenta esos picos u observaciones anormales, con el fin de que factores esporádicos no distorsionen los resultados del análisis”. Pero se tiene el inconveniente de que la mayoría de los métodos no pueden aplicarse sobre datos incompletos, así que las opciones son trabajar con menor cantidad de información o bien reemplazar estos valores, un proceso llamado imputación. El tema es muy extenso, y aunque no es el tema principal de esta tesis se recomienda el trabajo de Steffen Moritz, Alexis Sardá, Thomas Bartz-Beielstein, Martin Zaefferer, Jörg Stork (2015) “*Comparison of different Methods for Univariate Time Series Imputation in R*”² quienes hablan sobre la forma en que se generan los valores perdidos, y comparan distintos métodos de imputación. En el caso de valores atípicos trabajos como el de Kraiser y Maravall (1999)³ presentan los tipos de valores atípicos así como su detección y corrección bajo el metodología de Box-Jenkins. Existen de hecho funciones y librerías enteras en R dedicadas a la detección y corrección de estos valores, por ejemplo `tsoutlier`, librería cuyo autor es Javier López-de-Lacalle así como varias funciones en la librería `Forecast` por Rob J. Hyndman⁴.

2. Representación gráfica. Ver gráficamente el comportamiento de la serie facilita la toma de decisiones acerca de los métodos más adecuados a aplicar en ella.
3. Modelización. En esta etapa la finalidad es buscar el modelo que mejor se ajuste a los datos de la serie, es decir que explique mejor la realidad que se desea analizar.
4. Validación del modelo. Para este punto se debe conocer si el modelo que representa la serie tiene validez para describir su historia y/o hacer predicciones.
5. Predicción. El objetivo principal del análisis de series es la predicción, que consiste en determinar el valor futuro que tomará la variable aplicando un modelo determinado.

2.1.2. Componentes de una serie temporal

Partiendo del análisis clásico de series de tiempo, una serie se concibe integrada por movimientos a corto, medio y largo plazo, de alguna forma superpuestos. Por tanto, cada uno de estos movimientos es tratado como una componente de la serie, su estudio consiste en determinar los patrones de regularidad en cada uno de ellos, con la finalidad de reproducir el comportamiento de la serie. Morilla (1999) explica las componentes de la siguiente forma:

²Puede consultarse en: <https://arxiv.org/ftp/arxiv/papers/1510/1510.03924.pdf>.

³Puede consultarse en: <http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSerriadas/DocumentosTrabajo/99/Fic/dt9915e.pdf>.

⁴Mayores detalles sobre estas librerías pueden consultarse en: <https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf> y <https://cran.r-project.org/web/packages/forecast/forecast.pdf> respectivamente.

Tendencia:

Se considera tendencia al movimiento suave y regular de la serie a largo plazo. Esta es de gran interés ya que refleja la dirección del movimiento de una determinada variable. Puede detectarse si a largo plazo la serie adopta una marcha persistente en su crecimiento, decrecimiento o estabilidad. No se restringe a ser de comportamiento lineal estable, Hyndman y Athanasopoulos (2012) explican que “No tiene que ser lineal. A veces nos referimos a una tendencia de “dirección cambiante” cuando esta podría ir de una tendencia creciente a una tendencia decreciente.”. Para observarla es necesario disponer de un horizonte temporal amplio.

Estacionalidad:

Morilla (2000) define la estacionalidad como los movimientos regulares de la serie que tienen una periodicidad menor al año. Probablemente esta definición considera series de tiempo económicas, aunque de una manera más general Hyndman y Athanasopoulos (2012) mencionan que “La estacionalidad es siempre de un periodo fijo y conocido”⁵, sin restringirla a un periodo anual. Hay muchos casos de series donde se observa claramente esta componente y la mayor parte de las veces obedece a factores institucionales o climatológicos. Por ejemplo, la serie de tiempo de las ventas mensuales de abrigos se ve afectada por el factor climatológico pues sus ventas más altas se presentan en invierno⁶. Por otra parte la estacionalidad puede considerarse estable periodo a periodo o por el contrario presentarse afectada por una cierta tendencia.

Ciclo:

Un ciclo se distingue por una serie de movimientos ascendentes y descendentes separados por puntos de inflexión que en la terminología económica corresponden a las denominadas fases de recuperación, prosperidad, recesión y depresión. Una vez más Hyndman y Athanasopoulos (2012) proporcionan una visión ligeramente distinta que da más claridad a los conceptos, pues indican: “Un patrón cíclico existe cuando los datos exhiben subidas y bajadas que no son de un periodo fijo”. Además también lo diferencian de la estacionalidad según lo siguiente: “si las fluctuaciones no son de un periodo fijo entonces son cíclicas, si el periodo no cambia y están asociadas con un aspecto del calendario, entonces el patrón es estacional. En general, la duración promedio del ciclo es más larga que la duración del patrón estacional, y la magnitud de los ciclos tiende a ser más variable que la magnitud de los patrones estacionales”⁷.

Componente irregular:

Se caracteriza porque no corresponde un comportamiento sistemático o regular y en

⁵Traducción mía. Hyndman, R.J. & Athanasopoulos G. (2012), *Forecasting: principles and practice*. Recuperado de: <https://www.otexts.org/fpp/6/1>

⁶Como aclaración, cuando en este trabajo se mencionen las “estaciones” de la serie no se refiere a las cuatro estaciones climatológicas, sino a los periodos de tiempo en que está dividido el año para su estudio ya sean semanas, quincenas, meses, trimestres, etc.

⁷Traducción mía. Hyndman, R.J. & Athanasopoulos G. (2012), *Forecasting: principles and practice*. Recuperado de: <https://www.otexts.org/fpp/6/1>

consecuencia no sería posible su predicción. En el enfoque clásico se atribuye esta irregularidad al azar; sin embargo, en los análisis habituales de las series de tiempo, algunas veces es posible encontrar la causa que provoca esta irregularidad. Por ejemplo, puede ocurrir que las ventas de un negocio se hayan caído en cierto mes debido a una huelga, o bien desastres naturales, guerras, etc. Esto lleva de nuevo al tema de valores perdidos y valores atípicos mencionado anteriormente.

En una serie no tienen por qué estar presente todas las componentes, pero la componente irregular deberá ser incluida siempre, ya que estamos trabajando con series que no son deterministas y por tanto estarán afectadas por una perturbación al menos de carácter aleatorio.

Antes de poder llevar a cabo el análisis de las componentes, se deben tomar en cuenta las siguientes hipótesis:

- 1) La componente ciclo, de existir, está asociada con la tendencia formando una sola componente, llamada ciclo-tendencia.
- 2) Debemos asumir una cierta estabilidad en la estructura del fenómeno estudiado, de tal forma que si la serie presenta cambios bruscos de tendencia, sería más adecuado dividir el periodo de observación y estudiar cada parte por separado. Por otro lado, un cambio de tendencia en los datos puede deberse a alguna causa específica, y se requiera la información de en qué medida afectó a la serie. Por ejemplo, la entrada en vigor de una nueva legislación puede afectar el precio de las acciones de alguna empresa en particular, compra de publicidad se supondría elevaría las ventas a partir de la emisión del anuncio, o bien la implementación medidas para reducir contaminación. Este análisis es llamado *intervention analysis* en inglés, y aunque en este trabajo no se entrará en detalles sobre este tema, puede consultarse mayor información en el capítulo 10 del libro *Time series analysis, univariate and multivariate methods* por William W.S. Wei.
- 3) Hay que tener en cuenta el periodo de observación, pues puede suceder que, de forma incorrecta, se efectúe una predicción decreciente de la tendencia, basándose en que las observaciones finales formaban la parte decreciente de un ciclo de actividad.
- 4) Asumiremos que dentro del año, la influencia de la estacionalidad sobre el nivel medio de la serie es nulo, es decir, a lo largo de un año los efectos positivos se supone que se compensan con los negativos resultando nulo el efecto neto (razón por la cual las series con periodicidad anual no presentan efectos de la estacionalidad).

2.2. Modelos aditivo y multiplicativo y criterios para detectarlos

El análisis clásico se identifica como Y_t al valor de la variable en el tiempo t . La idea principal es que Y_t es de la forma:

$$Y_t = f(TC_t, E_t, I_t),$$

donde

TC_t : es la componente de ciclo-tendencia en el tiempo t ,

E_t : es la componente estacional en el tiempo t ,

I_t : es la componente irregular en el tiempo t .

Es decir que la variable está en función de estas cuatro componentes (aunque ciclo y tendencia son tratadas como una sola). La mayoría de las series pueden trabajarse con alguno de los siguientes esquemas:

Esquema multiplicativo

$$Y_t = TC_t \times E_t \times I_t$$

Este esquema puede describirse de la siguiente forma “El modelo multiplicativo asume que las cuatro componentes se deben a causas diferentes pero no son necesariamente independientes y pueden afectarse entre sí. Describe la tendencia como un valor absoluto mientras que los otros componentes [...] son expresados como tasas o porcentajes.”⁸.

Esquema aditivo

$$Y_t = TC_t + E_t + I_t$$

También Jain y Jhunjhunwala explican sobre este modelo: “El modelo aditivo asume que los cuatro componentes de de la serie de tiempo son independientes uno de otro y describe a todos los componentes como valores absolutos”⁹.

Los criterios para detectar si las componentes se asocian de manera aditiva o multiplicativa según Morilla son los siguientes:

- Representación gráfica de la serie

Si el gráfico de la serie tiene sus oscilaciones más o menos de la misma amplitud en torno a la tendencia sería un indicador de aditividad en las componentes, es decir, estaríamos admitiendo una hipótesis de independencia entre la tendencia y las demás fluctuaciones. Por el contrario, cuando las variables muestran oscilaciones no constantes si no cada vez mayores o bien cada vez menores a través del tiempo, entonces el esquema multiplicativo dará mejores resultados. La figura 2.1 explica gráficamente lo descrito.

- Análisis del gráfico desviación típica media

⁸Jain & Jhunjhunwala, 2007, *Business Statistics*, New Delhi, Tata McGraw-Hill, p. 8.5

⁹Jain & Jhunjhunwala, *op. cit.*, p. 8.5

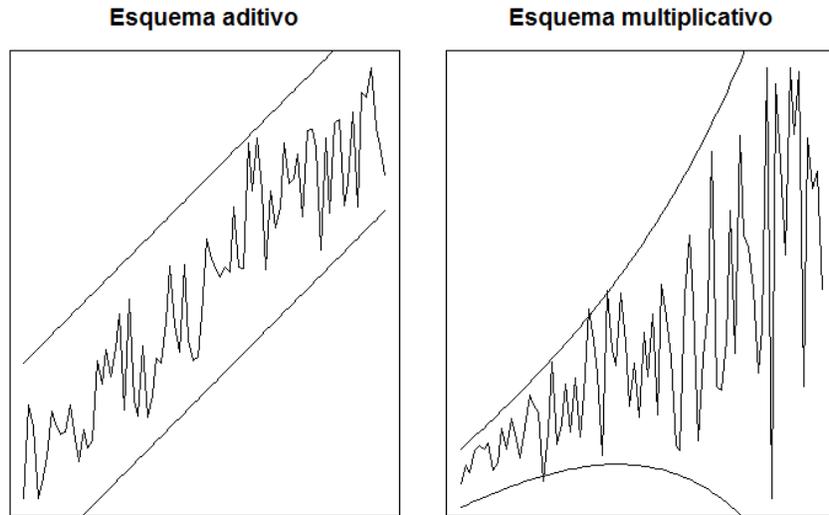


Figura 2.1: Esquema de los modelos aditivo y multiplicativo.

Este método consiste en obtener y representar gráficamente mediante un diagrama cartesiano, la media y la desviación estándar de la variable dentro de cada año, y con base en ello analizar la nube de puntos resultante.

Si el aspecto de la nube de puntos corresponde con una línea más o menos creciente elegiríamos una asociación multiplicativa entre las componentes. Por el contrario si la nube de puntos se distribuye de forma aleatoria, el esquema aditivo sería más apropiado.

- Método de las diferencias y cocientes estacionales

La diferencia estacional es la resta entre dos datos consecutivos, pero de la misma estación. Por ejemplo, una serie de estacionalidad anual se expresa como:

$$d_{j,i} = Y_{j,i} - Y_{j-1,i} \quad (2.1)$$

donde $Y_{j,i}$ es el valor de la serie en el año j con $j = \{1, \dots, e\}$ donde e es el número de años, en la estación i con $i = \{1, \dots, k\}$, existiendo k estaciones en el año. Por ejemplo, si la serie es trimestral $k = 4$ o si es mensual $k = 12$.

Mientras que el cociente estacional se obtiene por la división de dos datos consecutivos, pero de la misma estación, es decir:

$$c_{j,i} = Y_{j,i}/Y_{j-1,i} \quad (2.2)$$

Al obtener las series de diferencias y cocientes estacionales, se perderán en cada una k observaciones, correspondientes al primer año.

Como siguiente paso se obtienen los coeficientes de variación para las diferencias y cocientes estacionales, estos son respectivamente:

$$CV(d) = \frac{\text{Desviación típica } (d)}{\text{media } (d)}$$

$$CV(c) = \frac{\text{Desviación típica } (c)}{\text{media } (c)}$$

Por último se toma una decisión con la siguiente regla:

Si $CV(d)$ es menor que $CV(c)$ se elige el modelo aditivo.

Si $CV(c)$ es menor que $CV(d)$ se elige el modelo multiplicativo.

En otras palabras, los coeficientes estacionales están asociados con una relación multiplicativa entre la tendencia y la estacionalidad, mientras que las diferencias estacionales están asociadas con una relación aditiva entre la tendencia y la estacionalidad. Se elige el modelo que presente menor variación, para lo cual comparamos los coeficientes de variación, ya que es una medida que permite la comparación de varianzas en grupos de distinta escala.

Una vez adaptado el modelo de integración de componentes que mejor se adapta a la serie, el siguiente paso es la identificación de cada una de ellas. En las siguientes secciones se habla del análisis de componentes, que se refiere a encontrar el patrón de regularidad de cada componente. Específicamente, se tratan los análisis de estacionalidad y tendencia, pues la componente irregular por definición no tiene un patrón de regularidad, y la componente ciclo se trata conjuntamente con la tendencia debido a que la duración de un ciclo no es constante a lo largo de la serie.

2.3. Análisis de estacionalidad

Antes de iniciar cabe aclarar que los métodos de análisis de estacionalidad y análisis de tendencia de la siguiente sección son parte del análisis clásico que Morilla (2000) presenta en *Análisis de series temporales*.

El análisis de estacionalidad se puede realizar con dos propósitos diferentes:

El primero es eliminar las fluctuaciones de carácter periódico que se presentan en la serie a lo largo del año, con el objetivo de hacer comparables los datos correspondientes a estaciones diferentes, a este proceso se le llama desestacionalización. Por ejemplo, si se compara la variación del índice de precios al consumidor, entre los meses de noviembre y diciembre, se verá que ésta es muy alta, pues diciembre es un mes en el que los precios sufren un aumento considerable como consecuencia de los mayores gastos que acompañan a la fiesta de Navidad. Por tanto el mes de noviembre no sería comparable si no se elimina previamente la estacionalidad presente en la serie.

El segundo propósito del análisis de estacionalidad es la predicción, es decir anticipar un valor de la variable, por lo que se necesita tener conocimiento de esta componente.

Los diversos métodos que se utilizan para el análisis de la estacionalidad estiman esta componente a través de índices, expresando con ello las variaciones porcentuales de las diferentes estaciones del año respecto a la media anual.

Método de la razón a la media móvil

Como se verá en la siguiente sección, la media móvil es un método de estimación de tendencia, usado en los casos en que la tendencia no tiene un patrón bien definido, es decir, no es creciente ni decreciente durante todo el periodo de tiempo observado. En esta sección es necesario explicar el método, pues a partir de estas medias móviles se obtienen los índices con los que se estima la componente estacional de la serie.

Para un conjunto de datos $\{Y_1, Y_2, \dots, Y_T\}$ donde Y_T es el último elemento de la serie, se denomina media móvil de orden p a la serie de promedios de p observaciones. Los valores de esta nueva serie se asignan al momento central, por tanto sea $p = 2q + 1$, la media móvil de orden p del elemento t se define como:

$$Y_t^* = \frac{1}{2q + 1} \sum_{j=-q}^q Y_{t+j}$$

Por ejemplo, en un conjunto de diez datos, el primer elemento usando la media móvil $p = 5$ corresponde a: $Y_3^* = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5}$, véase que se asigna al momento $t = 3$, ya que Y_3 es el elemento central de los datos considerados. Es decir que este primer elemento está representando al elemento Y_3 . El siguiente elemento en la serie de medias móviles, que promedia de Y_2 a Y_6 , se asigna al momento $t = 4$, y así sucesivamente quedando todos los elementos de la media móvil centrados en los siguientes momentos $\{3, 4, 5, 6, 7, 8\}$. Finalmente se obtienen seis elementos ya que este método conlleva a la pérdida de observaciones, particularmente en este ejemplo se pierden cuatro observaciones de la serie original. Para que la media móvil quede totalmente centrada es necesario que el orden de la media sea un número impar.

Sin embargo, también es posible hacer series de promedios móviles de un orden p par¹⁰. Un ejemplo ayudará a ver que existe un problema y una sencilla solución, si se hace una media móvil de orden $p = 4$, los elementos estarán centrados en los momentos $\{2.5, 3.5, 4.5, \dots, 8.5\}$, pero haciendo nuevamente una media móvil de orden dos a la serie de promedios móviles obtenida, la serie estará centrada en los momentos $\{3, 4, \dots, 8\}$.

Morilla explica la intuitividad de la media móvil hablando primero de la media global pues “la media global de una serie constituye una representación de la misma siempre que su variabilidad no sea grande”, luego explica que el simple hecho de que una serie tenga tendencia y que además varíe a lo largo del tiempo hace que una media global no sea adecuada para describir la serie y concluye que por tanto se necesita “recurrir al concepto de media móvil con el fin de que el promedio se vaya adaptando a las circunstancias cambiantes que se produzcan en la historia cambiante de la serie”¹¹.

En el contexto del análisis de estacionalidad explica “el orden de la media más apropiado se hará coincidir precisamente con la periodicidad con que vengán expresados los datos dentro del año”¹², esto es, si la serie tiene periodicidad mensual se toma un promedio móvil de orden

¹⁰Morilla, *op. cit.*, p. 60.

¹¹Morilla, *op. cit.*, p. 59.

¹²Morilla, *op. cit.*, p. 61.

$p = 12$. Esto trae como consecuencia que las componentes estacional e irregular tienda a anularse en promedio y se obtiene una serie más representativa de la componente tendencia. Con el objetivo de estimar la componente estacional, el método de la razón a la media móvil, que explica Morilla (2000) sigue los siguientes pasos:

- 1) Estimar la tendencia ciclo a través de la obtención de una media móvil, cuyo orden p permita la eliminación de la componente estacional, este debe ser igual al número de estaciones en el año. En las series económicas es usual que los datos tengan una periodicidad mensual, bimestral, o trimestral, en todas ellas el número de observaciones por periodo es par: 12 meses, 6 bimestres, 4 trimestres en el año. Luego como se comentó para quedar completamente centrada se hace una nueva media móvil de orden 2.

En este primer paso, además de quedar eliminada la estacionalidad también se habrá purgado parte de las irregularidades, puesto que toda media móvil proporciona una serie más suavizada.

- 2) Obtener las razones o porcentajes con respecto a las medias móviles, es decir los cocientes de la serie original entre las medias móviles calculadas en el paso anterior. Como resultado se obtiene una primera estimación de la estacionalidad. El cálculo de las medias móviles, es una estimación de la tendencia-ciclo $TC_{j,i}$, con $j = \{1 \dots e\}$ y $i = \{1 \dots k\}$. Partiendo del modelo multiplicativo, en el que $Y_{j,i} = TC_{j,i} \times E_{j,i} \times I_{j,i}$ si se divide cada elemento de la serie original entre su correspondiente media móvil se obtiene el elemento estacional de la serie:

$$\frac{Y_{j,i}}{TC_{j,i}} = E_{j,i} \times I_{j,i} = IBE_{j,i}, \quad (2.3)$$

a estos cocientes se les nombra los Índices Brutos de Estacionalidad (IBE). Brutos en el sentido de que aún conservan el componente irregular. En el caso de tratar con el modelo aditivo la eliminación del componente ciclo-tendencia se hace por sustracción, es decir:

$$Y_{j,i} - TC_{j,i} = E_{j,i} + I_{j,i} = IBE_{j,i}.$$

- 3) Calcular la media de todos los IBE de una misma estación. Este paso se hace bajo el supuesto de que la estacionalidad es estable de un año a otro, por lo que las variaciones de los IBE de una misma estación serían poco significativas, entonces sería válido resumir los distintos índices de una misma estación en uno solo, y el cálculo de la media consigue eliminar en su mayor parte las variaciones irregulares. Se consiguen con ello k índices de estacionalidad (IE) con la siguiente ecuación:

$$IE_i = \frac{\sum_{j=1}^{e-1} IBE_{j,i}}{e-1} \quad i = 1, \dots, k$$

- 4) Normalización de los índices de estacionalidad. Los índices no deben afectar, en principio, el nivel de la serie, su suma debe ser igual a $100k$, pero nada garantiza que se

consiga ello con las operaciones realizadas hasta el momento. En este paso se obtiene primero la media de todos los IE , a la que se denota por IE (sin subíndices).

$$IE = \frac{\sum_{i=1}^k IE_i}{k} \quad (2.4)$$

Luego, los índices de estacionalidad normalizados están dados por:

$$IEN_i = \frac{IE_i}{IE}. \quad (2.5)$$

Los IEN_i son por tanto, los k valores finales que se usan como estimación de las influencias estacionales, correspondientes a las diferentes estaciones del año, y que se mantendrán constantes a lo largo del tiempo (en el supuesto de estacionalidad estable). Su interpretación es el porcentaje de desviación de cada observación con respecto a la tendencia.

Es posible que sintetizar las variaciones estacionales en k índices no sea conveniente porque la estacionalidad no sea estable, observándose algún tipo de tendencia en la misma para ello podemos:

- 1) Se representa gráficamente todos los IBE para cada estación observando su comportamiento a través de los años. En esta inspección puede intuirse la función que mejor ajuste a la nube de puntos. Es decir para cada estación tendremos una función de la forma:

$$IBE_i = f(t) + \varepsilon_t \quad i = 1, \dots, k.$$

Por ejemplo, para el caso de una serie trimestral, si se nota que los puntos se acoplan a una línea recta en cada estación, obtendremos cuatro funciones:

$$IBE_1 = \hat{a}_1 + \hat{b}_1 + \varepsilon_t$$

$$IBE_2 = \hat{a}_2 + \hat{b}_2 + \varepsilon_t$$

$$IBE_3 = \hat{a}_3 + \hat{b}_3 + \varepsilon_t$$

$$IBE_4 = \hat{a}_4 + \hat{b}_4 + \varepsilon_t.$$

- 2) Se estima la función por mínimos cuadrados ordinarios. En este proceso se elimina parte de la irregularidad.
- 3) Dando valores a t obtendremos estimaciones de estacionalidad para cada trimestre y año, incluso estas funciones se pueden utilizar para predecir la estacionalidad en años futuros.

La desestacionalización

Como se comentó al principio de la sección la estimación del componente estacional se hace con la finalidad de hacer diagnósticos más correctos de la realidad, ya que los movimientos cíclicos y de tendencia se apreciarán mucho mejor si se observa la serie desestacionalizada. Para que el proceso de desestacionalización pueda ser exitoso debe tomarse en cuenta que: movimientos cíclicos y de tendencia se apreciarán mucho mejor si se observa la serie desestacionalizada. Morilla (2000) menciona el proceso de desestacionalización debe tomar en cuenta lo siguiente:

- a) No se deben distorsionar los movimientos cíclicos básicos.¹³
- b) Las fluctuaciones irregulares deben comportarse de forma aleatoria cuando se considera que la autocorrelación tiene periodicidad alrededor de 12 (si la serie es mensual).
- c) La suma de la serie desestacionalizada debe ser igual a la suma de la serie original, ya que la desestacionalización no debe interferir en el resultado final.

Bajo el esquema multiplicativo, la serie desestacionalizada se obtiene dividiendo la serie original entre los índices de estacionalidad normalizados:

$$\frac{Y_{t,i}}{IEN_i} = \frac{TC_{t,i} \times E_{t,i} \times I_{t,i}}{IEN_i} \approx TC_{t,i} \times I_{t,i}, \quad (2.6)$$

mientras que bajo el esquema aditivo se hace por sustracción

$$Y_{t,i} - IEN_i = TC_{t,i} + E_{t,i} + I_{t,i} - IEN_i \approx TC_{t,i} + I_{t,i}.$$

Método de las diferencias estacionales

Este método consiste en crear una nueva serie la cual tendrá eliminado el efecto de la componente estacional, para ello debe suponerse que la serie original sigue el modelo aditivo y que el componente estacional es tal que ocurre lo siguiente:

$$E_t = E_{t-s},$$

además los componentes estacionales suman cero, esto es:

$$\sum_{k=1}^s E_k = 0.$$

Tomando en cuenta los aspectos anteriores vemos que la serie

$$Z_t = Y_t - Y_{t-s},$$

tiene eliminado el componente estacional, pues los elementos E_{t-s} y E_t se anulan al restarse, por lo que en la nueva serie Z_t solo están presentes los componentes tendencia y error. Bajo este método no se desestacionaliza propiamente la serie original sino a la nueva serie.

¹³Morilla no profundiza en el punto pero puede entenderse que el proceso de desestacionalización, como su nombre no indica sólo debe quitar el efecto de la estacionalidad, y no debe afectar la ciclo-tendencia de la serie.

2.4. Análisis de tendencia

Para el estudio de la tendencia se emplean dos enfoques distintos, el global y el local. En el enfoque global estimación de la tendencia se mantendría constante para todo el periodo. Y por tanto los métodos usados para estimarla, consisten en el ajuste de una función matemática dependiente del tiempo. Se emplea con la finalidad de realizar predicciones a largo plazo.

El enfoque local es el preferido cuando se trata de efectuar predicciones a corto plazo, ya que como la estimación de la tendencia va cambiando a medida que se añaden nuevas observaciones, los métodos adaptan mejor las circunstancias cambiantes. Las medias móviles así como los métodos de alisado (que se explicarán en la sección 2.5 de este capítulo) utilizan este enfoque local.

Método del ajuste analítico

Este método tiene un enfoque global, consiste en tratar de modelar la tendencia mediante funciones matemáticas dependientes del tiempo. Formalmente: $T_t = f(t)$, en donde la función f puede adoptar múltiples expresiones. El supuesto más habitual es el de tendencia lineal, aunque muchas series económicas se ajustan mejor con funciones de tipo exponencial o parabólica. El gráfico de la serie nos puede dar una idea aproximada de qué tipo de función cabe ajustar en cada caso.

Para la aplicación de este método se requiere previamente que la componente estacional, de existir sea eliminada, pues de lo contrario las oscilaciones interanuales perturbarían la identificación de la tendencia. De esta forma se considerará que la serie sólo tiene dos componentes, tendencia y residuo, con lo cual se observaría la serie en la forma: $Y_t = T_t + I_t$ con lo que se intenta extraer la componente tendencia ajustando un modelo como el modelo de regresión lineal.

En este caso, no se usa con el mismo planteamiento que en el análisis de regresión, ya que no se pretende analizar la capacidad explicativa de la variable tiempo, sino encontrar regularidades en el mismo.

Algunos de los modelos de tendencia más habituales, para el método del ajuste analítico son:

- Modelo de tendencia lineal

En este modelo, la serie sin efecto de la componente estacional, tiene el siguiente comportamiento:

$$Y_t = a + bt + \varepsilon_t.$$

En donde la tendencia a es el nivel de tendencia para $t = 0$, y b representa el incremento o decremento esperado cuando pasamos de un año a otro. Los estimadores de a y b corresponden a los estimadores β_1 y β_2 , por lo que el cálculo de estimación es el mismo. Para poder elegir un modelo de tendencia lineal, gráficamente los puntos deben estar más o menos alineados en torno a una recta.

En el análisis de regresión el modelo carecía de sentido si se llegaba a la conclusión de que el parámetro de la pendiente fuese cero, pero tratándose del análisis de tendencia,

el modelo con pendiente cero se refiere a que la tendencia es constante con el paso del tiempo.

Una desventaja importante en este e incluso en los siguientes modelos de tendencia es que los errores se suponen normalmente distribuidos y no correlacionados y de varianza constante¹⁴. Si bien suponemos que la serie no tiene la componente estacional, la autocorrelación de los errores no provienen necesariamente de esta componente, por lo que cabe la posibilidad de que los errores estén correlacionados y no se tengan estimadores confiables.

- Modelo de tendencia cuadrática o parabólica

Este es el caso en el que $f(t)$ ya no tiene una forma lineal sino parabólica, es decir:

$$Y_t = a_1 + a_2t + a_3t^2 + \varepsilon_t$$

También es un modelo de regresión lineal, pues la linealidad de la regresión no depende del exponente de la variable t , si no del exponente de los parámetros a_i , sin embargo, este se convierte en un modelo de regresión lineal múltiple.

La manera de detectar si los datos se adecuan a un modelo parabólico es la siguiente. Se define la primera diferencia de una serie como:

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

mientras que la segunda diferencia se calcula como:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}.$$

Tomando solo la parte determinista del modelo de tendencia parabólica para hacer las diferencias, puede verse que la primera diferencia tiene forma de una ecuación lineal:

$$\begin{aligned} \Delta Y_t &= a_1 + a_2t + a_3t^2 - [a_1 + a_2(t-1) + a_3(t-1)^2] \\ &= a_2 + a_3 + 2a_3t \end{aligned}$$

Entonces la serie de diferencias debe parecerse a una recta, en donde $a_2 + a_3$ es la ordenada al origen. Mientras que la serie de segundas diferencias resulta ser una constante:

$$\begin{aligned} \Delta^2 Y_t &= a_2 + a_3 + 2a_3t - [a_2 + a_3 + 2a_3(t-1)] \\ &= 2a_3 \end{aligned}$$

De donde se deduce que esta serie debe oscilar alrededor de un término constante. Entonces al obtener la primera diferencia de una serie, si resulta oscilar alrededor de una línea recta, o si la segunda diferencia parece oscilar alrededor de un término constante, son indicios para pensar que el modelo de tendencia parabólica es adecuado.

De hecho la n -ésima diferencia filtra a un polinomio de orden n , esto es, si la serie se ajusta a un polinomio de orden n , la n -ésima diferencia debería oscilar alrededor de un valor constante.

¹⁴Morilla (2000), *op. cit.*, pp. 48-49.

- Modelo tendencia exponencial

Este modelo tiene la forma:

$$Y_t = e^{a+bt}$$

En general el ajuste exponencial está indicado cuando la serie de datos varían en progresión geométrica. Una forma de saber si los datos se adecúan a este modelo es considerar la serie de crecimientos dada por:

$$t.c. = \frac{(Y_t - Y_{t-1})}{Y_{t-1}} = \frac{Y_t}{Y_{t-1}} - 1.$$

Se elige el modelo exponencial si el comportamiento de los crecimientos a lo largo del tiempo parece ser constante, ya que la función exponencial se caracteriza por tener una tasa de crecimiento constante. Este modelo puede manejarse como un modelo de regresión lineal simple tomando el logaritmo en ambos lados de la ecuación:

$$\ln Y_t = \ln(e^{a+bt}) = (a + bt)\ln(e) = a + bt.$$

Es decir que si se trabaja con la serie de logaritmos de Y_t , ésta se comporta como una línea recta, a la que se modela como una regresión lineal simple para obtener los parámetros a y b .

Método de la media móvil

El ajuste de una función matemática para representar la tendencia de una serie es apropiado cuando se observa un crecimiento o decrecimiento continuo en los valores de la serie. Sin embargo, cuando el ritmo de crecimiento cambia a lo largo del tiempo, entonces la función puede aproximarse para intervalos cortos de tiempo. Este es el caso del método de la media móvil que corresponde a un enfoque local del análisis de tendencia.

En la sección anterior se mencionó el método de la media móvil para la estimación de la estacionalidad. En esta sección el objetivo es la estimación de la tendencia, para lo cual también es importante la eliminación de la estacionalidad, pues de conservarse, no se distinguiría claramente la componente de tendencia. Igual que en la sección anterior, para llegar a este objetivo se calcula la media móvil de orden p tal que coincida con la periodicidad de la serie, de no ser así, la estacionalidad no quedará totalmente eliminada.

Una vez eliminada la estacionalidad, las medias móviles obtenidas son una representación de la evolución de la tendencia de una manera más sencilla que con las funciones analíticas, y que se adapta mejor los cambios de tendencia. Sin embargo, para fines de predicción no es la mejor alternativa ya que este método no tiene una función matemática que pueda extrapolarse en función de tiempo. Su uso es mejor alternativa para fines descriptivos. En resumen, el método de la media móvil para la estimación de tendencia tiene los siguientes inconvenientes:

- Pérdida de observaciones.
- Dificultad para hacer predicciones, pues no tiene una función que permita la extrapolación.

- No disponemos de una medida de fiabilidad como el coeficiente de correlación o el coeficiente de determinación.
- Los promedios móviles pueden generar ciclos u otros movimientos que no estaban presentes en los datos originales.
- La aplicación de medias móviles introduce autocorrelación incluso en las series aleatorias puras y, además, puede modificar la autocorrelación previa.

Método de las diferencias

Este método no estima o identifica la tendencia de la serie, sino que la elimina. El procedimiento es obtener una nueva serie Z_t en la que cada elemento es la diferencia de la forma:

$$Z_t = Y_t - Y_{t-1}.$$

Bajo el supuesto de que la serie admite una tendencia lineal, si se resta a cada valor de la serie su valor anterior, la tendencia quedará eliminada de la serie resultante. A esta operación se denomina diferenciar una serie y se utiliza habitualmente en la modelización ARIMA de la metodología Box-Jenkins, de la que se habla en el siguiente capítulo. Si la primera diferencia no consigue eliminar la tendencia, la hipótesis de linealidad puede no ser la más indicada. En este caso, si un ajuste parabólico pareciera adecuado, se somete la serie a una nueva diferenciación. De manera general, puede aplicarse la n -ésima diferencia en caso de que la serie presente una tendencia polinomial de orden n .

Morilla resume los inconvenientes de este método de la siguiente manera¹⁵:

- Por cada diferenciación de la serie se pierde una observación.
- Se requiere que la serie no presente estacionalidad o que haya sido previamente eliminada.
- Puede que con la diferenciación se elimine alguna otra componente.
- Se dificultan las predicciones pues con esta técnica no se identifica la tendencia.

Sobre el último punto, si bien la diferenciación es invertible, Morilla probablemente se refiere a que no se tiene una función matemática dependiente del tiempo que permita hacer predicciones mediante extrapolación.

2.5. Métodos de alisado

Los métodos de alisado son técnicas de análisis de series temporales que se retroalimentan a medida que se generan nuevos datos. Facilitan la predicción pues usan, para tal fin, fórmulas recurrentes (es decir que la estimación de cada término está en función de términos anteriores). Estas técnicas también son de enfoque determinista como los modelos de descomposición, pero tienen las siguientes diferencias:

¹⁵Morilla, *op. cit.*, pp. 69-70.

- 1) Los modelos de descomposición, son de validez global ya que intentan extraer las regularidades a través de funciones dependientes del tiempo. Mientras que los métodos de alisado son de validez local, pues tratan de combinar en cada momento las observaciones pasadas con el fin de descubrir la estructura del fenómeno.

Es por ello que los métodos de alisado dan mejores resultados cuando la estructura de los datos se muestra inestable.

- 2) Las técnicas de alisado son más indicadas cuando se quiere predecir sobre todo en el corto plazo, ya que los cambios estructurales se detectan antes.
- 3) Los resultados son satisfactorios incluso cuando se dispone de un menor número de observaciones.

En esta sección se usarán los siguientes términos y notaciones:

- La información de la variable desde el momento 1 hasta el momento T , se denota por: $Y_1, Y_2, \dots, Y_t, \dots, Y_{T-1}, Y_T$.
- *Predicciones intramuestrales*: son las estimaciones que se hacen de la variable dentro del periodo que abarque la muestra de observaciones. Las predicciones intramuestrales un periodo hacia adelante se representan como: $\hat{Y}_{t+1/t}$.
- *Predicciones extramuestrales*: son las realizadas para momentos futuros al periodo de observación. Las predicciones que se hacen uno, dos o h periodos adelante, condicionadas a la información con que se dispone en el momento T se denotan como: $\hat{Y}_{T+1/T}, \hat{Y}_{T+2/T}, \dots, \hat{Y}_{T+h/T}$.
- El error de predicción se denota como: $e_t = Y_t - \hat{Y}_t$.

El esquema básico de referencia en estos modelos es el de una serie que oscila en torno a su valor medio, es decir:

$$Y_t = a_t + \varepsilon_t,$$

donde a_t representa el nivel medio de la serie y ε_t designa las desviaciones de carácter aleatorio que se producen en torno a ese valor medio. Se asume que en la variable ε_t no existe autocorrelación. Además el nivel a_t depende del tiempo t porque estos métodos ajustan el nivel la serie de acuerdo a la información que se disponga de cada uno, es por ello que estas técnicas se denominan de estructura variable.

2.5.1. Alisado exponencial simple

La técnica del alisado exponencial simple se usa en series que no tienen tendencia ni estacionalidad, o bien en series en las que estas componentes ya fueron eliminadas. La idea principal de esta técnica es que supone que la información más cercana al momento de la predicción es más relevante que las observaciones más alejadas. Por ello, para hacer una predicción sobre un cierto período, se toma un promedio ponderado de todos los datos anteriores.

Un promedio o media ponderada es un promedio en el que no todas las observaciones tienen el mismo peso, por ejemplo, un curso con tres exámenes que valen 20 %, 60 % y 20 %, si las calificaciones de estos son 9, 7 y 8, entonces la calificación final será $9(0.2) + 7(0.6) + 8(0.2) = 7.6$ mientras que en un promedio normal la calificación sería 8. Nótese que los pesos suman la unidad (o bien el 100 %).

En el alisado exponencial simple se da mayor peso a los datos más recientes, y los pesos decrecen de forma exponencial. Matemáticamente el nivel a_t se estima de la siguiente forma:

$$\hat{a}_t = A_t = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots = \alpha \sum_{i=0}^{\infty} Y_{t-i} (1 - \alpha)^i, \quad (2.7)$$

donde el parámetro α es la *constante de alisamiento*, sus límites son $0 < \alpha < 1$. En este caso la suma de los pesos no es igual a uno para un número pequeño de observaciones, sino que mientras más observaciones se tengan más se acerca dicha suma a la unidad, es decir, $\sum_{i=0}^{\infty} \alpha(1 - \alpha)^i = 1$.

Las predicciones intramuestrales se calculan de la siguiente manera:

$$\hat{Y}_{t+1/t} = A_t \quad \hat{Y}_{t+2/t+1} = A_{t+1} \quad \hat{Y}_{T/T-1} = A_{T-1}$$

de manera que, la estimación para el momento $t + 1$ contando con los primeros t datos, es la alisada con los primeros t términos. Mientras que las predicciones extramuestrales son siempre A_T , pues la única información real es la que se tiene hasta el momento T :

$$\hat{Y}_{T+1/T} = A_T \quad \hat{Y}_{T+2/T} = A_T \quad \hat{Y}_{T+h/T} = A_T$$

Una forma más simple de calcular las predicciones, se obtiene al factorizar en la fórmula (2.7) el término $(1 - \alpha)$, obteniendo la siguiente fórmula recursiva:

$$\begin{aligned} A_t &= \alpha Y_t + (1 - \alpha)[\alpha Y_{t-1} + \alpha(1 - \alpha)Y_{t-2} + \dots] \\ A_t &= \alpha Y_t + (1 - \alpha)A_{t-1} \end{aligned}$$

Es recursiva en el sentido de que la estimación del periodo actual A_t requiere de la estimación del periodo anterior A_{t-1} , ésta a su vez requiere del promedio A_{t-2} , y así hasta llegar al término A_1 . Con esto las estimaciones son calculadas como:

$$\hat{Y}_{t+1/t} = \alpha Y_t + (1 - \alpha)\hat{Y}_{t/t-1}$$

De este modo las nuevas observaciones se utilizarán para actualizar las estimaciones hechas en periodos anteriores. De hecho, el uso de fórmulas recursivas para hacer promedios ponderados es en lo que se basan los métodos de alisado. Para usar esta fórmula recursiva debe asignarse un valor inicial a A_1 , para lo cual se tienen estas opciones:

- Considerar $A_1 = Y_1$
- Utilizar la media de los primeros valores de la serie.

- Elegir como valor inicial la media muestral de todos los valores de la serie, que sería adecuado para series que muestren una cierta estabilidad.

El uso de α proporciona una serie con menor o mayor suavidad. Si este parámetro se aproxima a cero querría decir que se le da mayor importancia a la parte de alisado lo que provocará una serie con menos irregularidades. Por el contrario si α está cercano a uno, los valores más recientes pesarán más, las estimaciones se ajustarán más a ellos y proporcionará una serie de predicciones con menor alisamiento. Sin embargo, el valor más adecuado para α es el que minimice el error cuadrático medio, ECM , este se define como:

$$ECM = \sum_{t=1}^T \frac{(Y_t - \hat{Y}_t)^2}{T}.$$

2.5.2. Alisado exponencial doble (modelo de Holt)

El modelo de Holt se aplica a las series que presenten una tendencia lineal pero no estacionalidad, puede entonces aplicarse sin problemas a series con periodicidad anual pues éstas no presentan estacionalidad, o bien si se trabaja con una periodicidad menor a la anual (mensual, trimestral, etc.) debe eliminarse el efecto de la estacionalidad para poder trabajar con este método. El método Holt trabaja con dos constantes de alisado (α y β) con la finalidad de que el método se adapte con mayor flexibilidad a los valores de la serie.

El modelo funciona de esta manera, se estima mediante promedios ponderados (o bien alisamiento) tanto el nivel de tendencia (a_t) como la pendiente de un periodo a otro (b_t), con el uso de fórmulas recursivas. Es por ello que también se le conoce como el alisado exponencial doble. Las ecuaciones de alisado son las siguientes:

Ecuación de nivel

$$a_t = \alpha Y_t + (1 - \alpha)(a_{t-1} + b_{t-1})$$

En ella se observa que la estimación de nivel del período t , se compone de a_{t-1} más b_{t-1} , para evitar que las predicciones queden sistemáticamente sesgadas.

Ecuación de pendiente

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \quad \text{con } 0 < \alpha < 1 \text{ y } 0 < \beta < 1.$$

Los valores de la pendiente se actualizan, según se observa una fórmula, a través de una media ponderada entre el último valor estimado, b_{t-1} y la diferencia entre las dos estimaciones más recientes del nivel, $a_t - a_{t-1}$.

Las predicciones para el periodo muestral son de la forma:

$$\hat{Y}_{t/t-1} = a_{t-1} + b_{t-1},$$

mientras que las predicciones extramuestrales se obtienen como:

$$\hat{Y}_{T+h/T} = a_T + b_T h.$$

2.5.3. Modelo Holt-Winters

El modelo de Holt-Winters una extensión del modelo de Holt que permite modelar además del nivel y la pendiente, el comportamiento estacional. Las series susceptibles de ser tratadas mediante la técnica de Holt -Winters son aquellas que contienen una componente de tendencia aproximadamente lineal y otra de carácter estacional. Este factor estacional, denotado por c_t , puede incluirse de forma aditiva o de forma multiplicativa, dando lugar a los siguientes modelos:

Asociación multiplicativa

$$Y_t = T_t \times E_t + \varepsilon_t = (a_t + b_t \times t) \times c_t + \varepsilon_t$$

Asociación aditiva

$$Y_t = (T_t + E_t) + \varepsilon_t = (a_t + b_t \times t) + c_t + \varepsilon_t$$

Observamos que en ambos casos la perturbación aleatoria se introduce aditivamente. Como en los otros métodos de alisado, el de Holt-Winters también estima la pendiente, el nivel y ahora el factor estacional de manera recursiva. Las ecuaciones de actualización para el caso de la asociación multiplicativa son las siguientes:

- Ecuación de nivel

$$a_t = \alpha \frac{Y_t}{c_{t-k}} + (1 - \alpha)(a_{t-1} + b_{t-1})$$

- Ecuación de pendiente

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

- Ecuación del factor estacional

$$c_t = \gamma \frac{Y_t}{a_t} + (1 - \gamma)c_{t-k}$$

En donde k es el número de estaciones al año, de manera que $k = 12$ si la serie es mensual, mientras que α , β y γ las constantes de alisado correspondientes, todas ellas entre cero y uno.

En el caso de la asociación aditiva las ecuaciones de actualización son:

- Ecuación de nivel

$$a_t = \alpha(Y_t - c_{t-k}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

- Ecuación de pendiente

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

- Ecuación del factor estacional

$$c_t = \gamma(Y_t - a_t) + (1 - \gamma)c_{t-k}$$

La ecuación de nivel similar a la empleada por el modelo Holt, sólo que en este caso debe quitarse la estacionalidad a la nueva observación Y_t , es por ello que aparece dividida entre componente estacional del mismo periodo pero del año anterior (c_{t-k} , donde como se dijo k es el número de estaciones en el año). De manera similar una ecuación del factor estacional, a la observación Y_t se le divide entre a_t , para observar solamente el componente estacional de la observación. Las fórmulas que se usan para predicción en el modelo multiplicativo son las siguientes:

Para el periodo muestral $\hat{Y}_{t/t-1} = (a_{t-1} + b_{t-1})c_{t-k}$

Para predicciones extramuestrales $\hat{Y}_{T+h/T} = (a_T + b_T h)c_{T+h-k}$.

Mientras que en el modelo aditivo se utilizan las siguientes:

Para el periodo muestral $\hat{Y}_{t/t-1} = a_{t-1} + b_{t-1} + c_{t-k}$

Para predicciones extramuestrales $\hat{Y}_{T+h/T} = a_T + b_T h + c_{T+h-k}$.

Para iniciar los cálculos del modelo deben conocerse los valores iniciales de los componentes nivel, pendiente y estacionalidad. Este es un tema bastante amplio en el que existe más de una respuesta. Hyndman (1998) propone los siguientes valores iniciales:

$$a_0 = \frac{\sum_{i=1}^k Y_i}{k}, \quad (2.8)$$

$$b_0 = \frac{\sum_{i=k+1}^{2k} Y_i - \sum_{i=1}^k Y_i}{k^2}, \quad (2.9)$$

en dónde a_0 es el valor del nivel inicial, puede verse que es el promedio de las observaciones del primer año; mientras que la pendiente inicial b_0 es equivalente a obtener el promedio de las pendientes entre cada estación en los primeros dos años. Por último, para los primeros k componentes estacionales, si el modelo es multiplicativo, se usa:

$$c_i = \frac{Y_i}{a_0} \quad \text{con } i = 1, \dots, k, \quad (2.10)$$

mientras que si el modelo es aditivo se usa:

$$c_i = Y_i - a_0 \quad \text{con } i = 1, \dots, k. \quad (2.11)$$

Hyndman comenta como ventajas que son estimadores fáciles de calcular y producen resultados aceptables, sin embargo, tienen la desventaja de provocar resultados poco fiables en series cortas y que las estimaciones pueden hacerse a partir de la observación $k+1$, con lo que se pierde un periodo de observaciones. La función `HoltWinters` en R utiliza las estimaciones en (2.8) y (2.9) cuando la serie no es estacional, o bien sólo la estimación en (2.8) si la serie tampoco presenta tendencia.

Morilla (2000) sugiere realizar una regresión lineal simple a los datos previamente desestacionalizados con el método de la media móvil, y utilizar los parámetros ordenada al origen y pendiente como a_0 y b_0 respectivamente, mientras que para los factores estacionales pueden usarse los índices proporcionados por el método de la media móvil antes explicado. Al respecto, Hyndman (2008) también recomienda la regresión con valores desestacionalizados y obtener los parámetros a y b como los valores iniciales, pero advierte no hacer dicha regresión con valores no desestacionalizados. También menciona que este procedimiento es utilizado en la función `HoltWinters` de la librería `stats` de R, cuando la serie es estacional y se usan los primeros dos años de los datos¹⁶, aunque, los parámetros iniciales, también pueden ser asignados por el usuario.

Por último, Hyndman (2008) habla sobre un nuevo método de estimación de los valores iniciales, en el que estos son estimados junto con los parámetros de suavizamiento α , β y γ por el método de máxima verosimilitud. Si bien este método sigue las ecuaciones de Holt-Winters, es en esencia distinto, pues en la teoría detrás de este método están los modelos llamados modelos de espacio de estado. Estos contemplan una variedad cercana a 30 modelos de los cuales los tres antes mencionados son casos particulares. Su estimación es también más compleja y no se revisará en este trabajo, pero como ventaja no solo pueden estimar valores futuros sino también intervalos de confianza para los mismos. La función `ets` de R utiliza este método de estimación.

¹⁶Hyndman, R. J. (29 de Noviembre de 2010). Initializing the Holt-Winters method [Mensaje en un blog]. Recuperado de: <http://www.r-bloggers.com/initializing-the-holt-winters-method/>

2.6. Aplicación

Se seguirán utilizando los mismos datos mensuales del negocio dedicado a la venta de artículos de ferretería y tlapalería en el municipio de Nezahualcóyotl, mostrados en la figura 1.1. Para empezar a trabajar la base de datos deben tomarse en cuenta los efectos de calendario de inflación. Para quitar el efecto calendario, a los datos de los meses Enero, Febrero, Marzo, Mayo, Julio, Agosto, Octubre y Diciembre se dividen entre sus 31 días (en el caso de las ventas de Febrero se dividen entre 28 días), y luego se multiplican por 30 para hacer parecer que tienen 30 días.

También es importante remover el efecto inflacionario, pues los datos no son comparables a través del tiempo, por ejemplo, las ventas en Noviembre 2012 no se pueden comparar con las ventas de Julio 2010, pues en ese lapso de tiempo los precios subieron un 10.22%. La inflación mensual se mide con los índices de precios al consumidor, información que se obtiene a través del portal del INEGI (Instituto Nacional de Estadística Geografía e Informática). El incremento porcentual entre un índice y otro es la inflación mensual. Luego debe elegirse un período base, en este caso se toma el primer periodo en la serie, de forma que todos los elementos de la serie podrán compararse con el mes de Julio 2010. Para obtener datos deflactados (sin inflación) véase un ejemplo, si las ventas no subieran por el fenómeno de la inflación, las ventas en Noviembre 2012 serían de $363.74/(1+10.22\%)=330.01$. De la misma forma, las ventas de los demás meses se dividen entre la inflación respecto del mes base Julio 2010.

El Cuadro 2.1 muestra los datos sin el efecto calendario, y sin efecto de la inflación. Al graficar los datos, como se muestra en la Figura 2.2 vemos que esta serie no muestra una fuerte tendencia al alza, sin embargo, sí muestra un fuerte comportamiento estacional con el más alto pico en Diciembre.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010							287.82	324.61	299.17	310.68	292.89	385.49
2011	310.87	303.59	293.26	273.76	291.10	307.74	347.50	313.44	338.27	297.75	271.45	396.52
2012	320.37	310.31	274.55	303.05	316.54	304.22	324.85	322.26	331.91	319.10	330.01	411.46
2013	369.78	335.23	286.59	302.28	304.29	301.49	292.83	310.09	290.03	313.79	313.29	418.58
2014	337.93	311.07	262.95	292.58	289.04	307.01	294.92	312.42	278.11	294.51	283.42	377.47

Cuadro 2.1: Ventas mensuales sin los efectos de calendario ni de inflación

Es momento de decidir si se adopta el esquema aditivo o el multiplicativo para el tratamiento de la serie. De manera gráfica, al observar la Figura 2.2, se observa que los picos de Diciembre crecen a cada año, a excepción de Diciembre 2014. Sin embargo, a simple vista no puede asegurarse que las oscilaciones son de la misma amplitud. Según el gráfico desviación típica media (Figura 2.3)¹⁷, la media parece crecer junto con la varianza, y proporciona elementos para pensar que es más adecuado el esquema multiplicativo. Si bien esto no es del todo cierto en el último año, se confirmará con el método de las diferencias y los cocientes estacionales cuál es el modelo al que más se ajusta la serie.

¹⁷Se tomaron en cuenta los cuatro años con información completa.

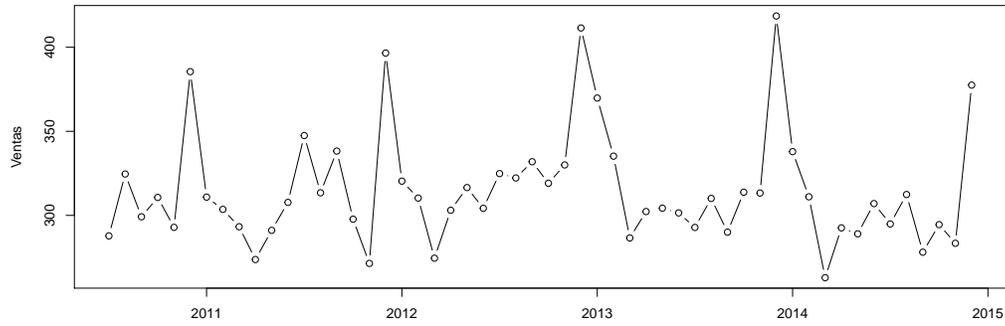


Figura 2.2: Ventas mensuales sin los efectos de calendario ni de inflación

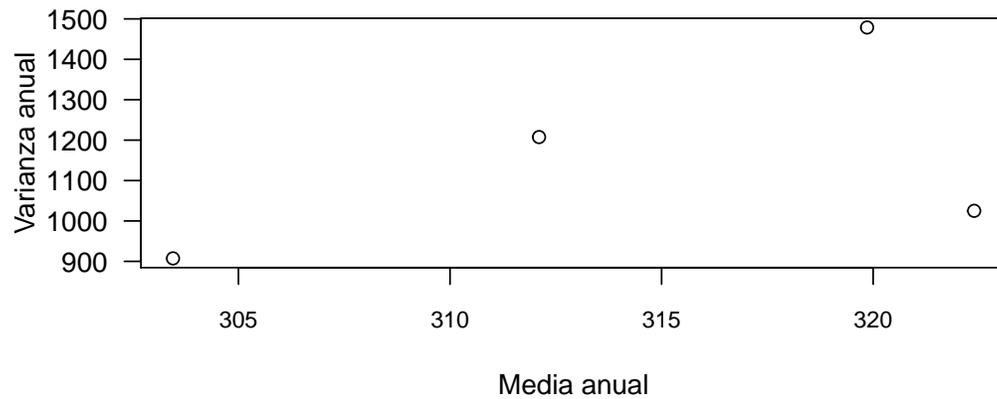


Figura 2.3: Gráfico de desviación típica media

Con ayuda del software *R* se hacen cálculos de las diferencias y cocientes estacionales usando las fórmulas (2.1) y (2.2) y se obtienen los coeficientes de variación correspondientes.

```
> cv_dif #Coeficiente de Variación de diferencias estacionales
```

```
[1] 26.20403
```

```
> cv_cof #Coeficiente de Variación de cocientes estacionales
```

```
[1] 0.0802564
```

Al ser menor la variación de los cocientes, se prefiere el esquema multiplicativo, rectificando el resultado anterior.

Análisis de la estacionalidad

Para estudiar la estacionalidad de la serie se usará el método de la media móvil que permitirá estimarla y posteriormente eliminar su efecto de los datos. Al aplicar una media móvil de orden $p = 12$ se elimina el efecto estacional, por ser el número de meses en el año. Pero esta serie no está correctamente centrada, por ejemplo, si las observaciones de la serie las denotamos como Y_1, Y_2, Y_3, \dots , entonces la serie de promedios móviles estiman a los elementos $Y_{6.5}^*, Y_{7.5}^*, Y_{8.5}^*, \dots$, sin embargo, las observaciones no se encuentran en esas posiciones y no tendría sentido describir ni hacer predicciones para puntos en el tiempo que no se observan. Para centrar los datos se aplica una media móvil de orden $p = 2$, de manera que los nuevos promedios móviles sean $Y_7^*, Y_8^*, Y_9^*, \dots$, quedan correctamente centrados. La nueva serie de medias móviles ya centradas y con la pérdida de doce observaciones puede observarse en la Figura 2.4.

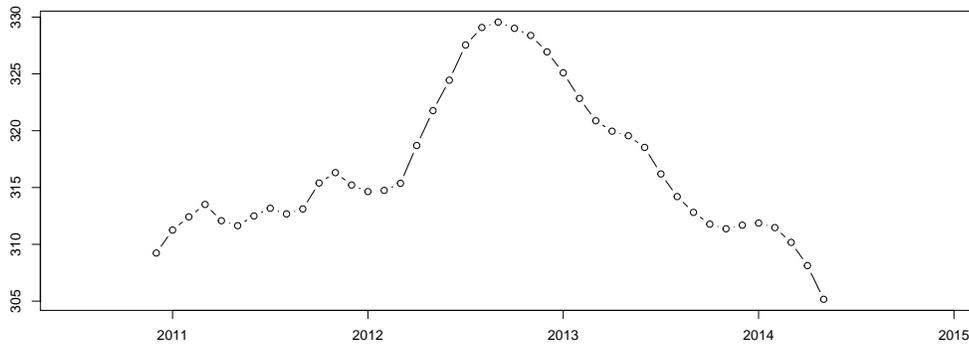


Figura 2.4: Media móvil de orden $p=2$ aplicada a la media móvil de orden $p=12$

Mediante el método de las razones a la media móvil se obtienen los índices brutos de estacionalidad, como en la fórmula (2.3) y graficados en la figura 2.5 de 2011 a 2013, pues solo en estos años se tienen índices en todos los meses. En esta se aprecia que en 2011 el comportamiento estacional fue ligeramente distinto en algunos meses con respecto a 2012 y 2013, pero en ningún caso parece observarse tendencia en algún índice, además los IBE entre 2012 y 2013 son más parecidos entre sí, por lo que podemos suponer que la componente estacional es estable.

Habiendo visto que cada mes tiene un comportamiento similar año con año, tiene sentido el cálculo de los índices estacionales normalizados (IEN), obtenidos según las ecuaciones (2.3), (2.4) y (2.5), lo que proporcionó los siguientes resultados:

```
[1] 1.0508474 0.9980819 0.8991808 0.9237358 0.9561213 0.9557411 1.0085723
[8] 0.9889644 1.0042896 0.9728935 0.9558657 1.2857061
```

A partir de estos índices se ve que el mes con mayor venta es Diciembre, seguramente relacionado con el aguinaldo que se da a los trabajadores, mientras que Marzo suele ser el mes con menores ventas, probablemente relacionado con las vacaciones de semana santa, pues incluso en la figura 2.5 puede verse que en 2011 cuando esta festividad fue a mediados de

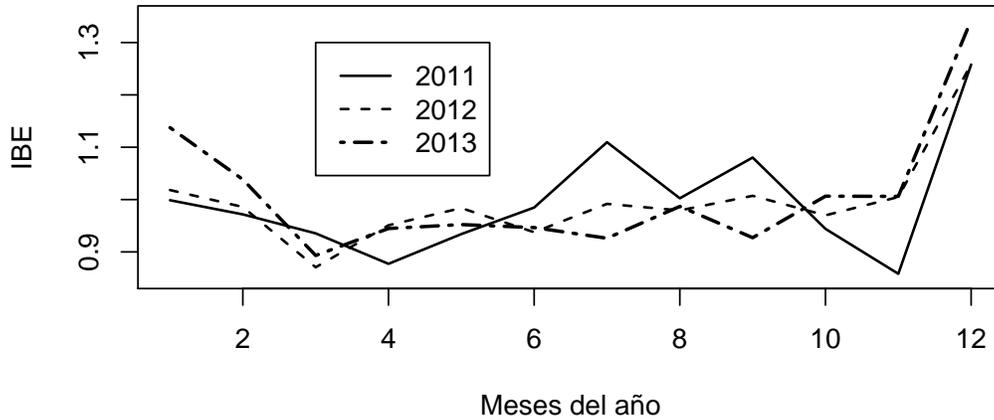


Figura 2.5: Índices Brutos de Estacionalidad de los años 2011 a 2013

abril, el efecto se recorrió al mes de abril, en 2012 y 2013 ocurrió en marzo y principios de abril respectivamente pero afectó más en el mes de marzo. Es de notar que esto no ocurre en las vacaciones de Julio pues estas son vacaciones escolares, y dado el giro del negocio, no tienen tanto impacto.

La obtención de la serie desestacionalizada es el primer propósito del análisis de estacionalidad, se obtuvo dividiendo cada observación entre su IEN correspondiente (ecuación 2.6) y se muestra en la figura 2.6. Este procedimiento permite remover la estacionalidad de la serie sin afectar la suma anual. De esta forma los meses son comparables entre sí, por ejemplo, se sabe que en Diciembre se tienen ventas más altas que en Noviembre por el efecto estacional, pero sin el efecto estacional esto solo ocurre en 2011, en los siguientes años la venta de Noviembre es siempre mayor, debido a que se está observando propiamente las componentes tendencia y error de la serie. Esta serie es la base para el análisis de tendencia.

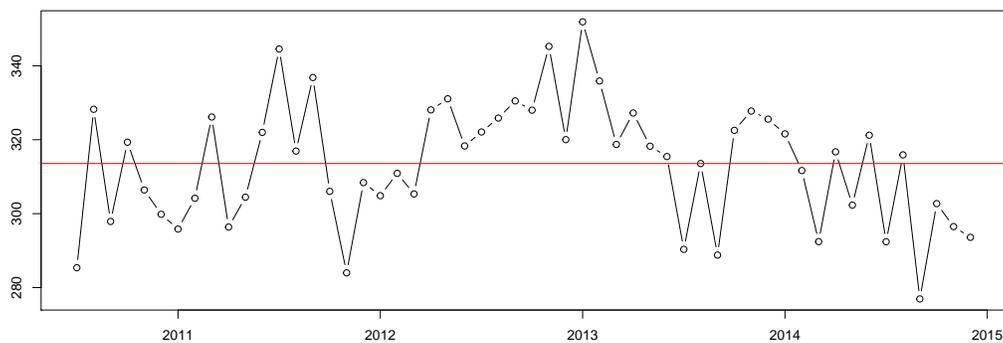


Figura 2.6: Serie de ventas sin el efecto de la componente estacional

Análisis de la tendencia

A primera vista la serie desestacionalizada parece tener una tendencia constante, pues las observaciones oscilan en general alrededor de la media. Sin embargo, en algunos periodos se agrupan más por arriba de la media, mientras que en otros se agrupan más por debajo de la

media. Lo cual hace pensar que el comportamiento de la tendencia de la serie puede ser una combinación de tendencia constante, creciente y decreciente.

Debe tenerse presente que si bien se eliminó el efecto estacional, la serie aún presenta el efecto aleatorio, por lo que parece conveniente utilizar una media móvil para suavizar la serie y así obtener una mejor visibilidad de la tendencia. Como la estacionalidad ya fue removida, puede utilizarse cualquier orden p impar para que el centrado de la serie sea automático, y que no involucre pérdida de tantas observaciones. Desde un orden $p = 3$, aunque un orden tan pequeño podría no suavizar lo suficiente la serie, probaremos con los órdenes $p = 5$, $p = 9$, $p = 13$ (figura 2.7).

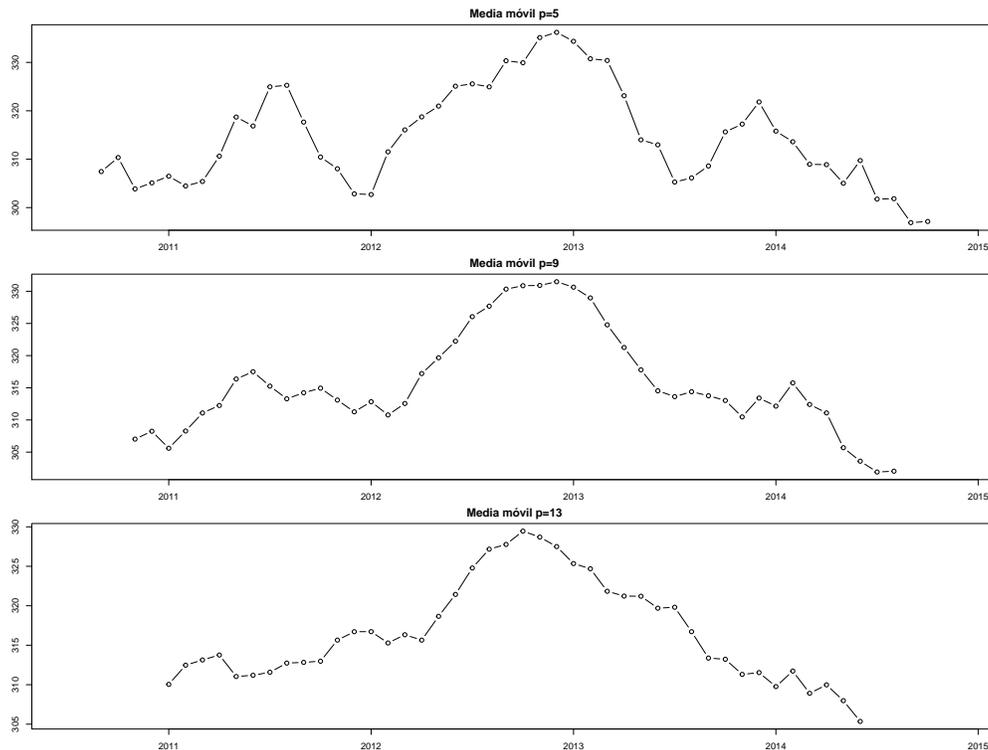


Figura 2.7: Medias móviles de la serie con órdenes $p=5$, $p=9$, $p=13$

En las medias móviles con orden $p = 5$ y $p = 9$, se alcanza a distinguir de ligeras tendencias a la alza de inicios a mediados de 2011, luego cae a la baja hasta finales de 2011, vuelve a subir durante todo 2012 y en 2013 cae la mitad del año, mientras que la otra mitad se recuperó ligeramente, sin llegar al nivel máximo de finales de 2012, que además es el año con movimientos más constantes a la alza. Para la media móvil de orden $p = 13$ el movimiento de tendencia más alto es en el año 2012, mientras que en 2011 ya no se observan picos sino más bien una tendencia constante, y para los años 2013 y 2014 se ve una clara tendencia a la baja.

El movimiento más visible es el de alza hasta 2012 y baja a partir de 2013. Es probable que el movimiento a la baja en ventas esté relacionado con dos hechos, el primero es el robo del vehículo utilizado para transportar mercancías, registrado a finales de 2012; el segundo es que, al ser un negocio familiar los hijos estaban encargados completamente de surtir mercan-

cía, y empezaron a alternar su actividad en el negocio con su actividad profesional. Debido a que la tendencia de la serie no es constante a través del tiempo sería más adecuado hacer estimaciones a través de los métodos de alisado, que con los métodos clásicos.

Predicción

Como la serie tiene estacionalidad se ocupará el método de Holt-Winters, en el que los factores estacionales iniciales serán los dados por los índices estacionales normalizados ya calculados. Para el nivel y la pendiente iniciales se calculó la ordenada al origen y la pendiente de los primeros dos años desestacionalizados.

En el alisamiento se consideraron todas las observaciones sin importar que los índices estacionales calculados no tomaron en cuenta las observaciones de 2010, simplemente se supondrá que en 2010 se tiene el mismo patrón estacional, con el propósito de aprovechar toda la información posible. La función `HoltWinters` de R, hace los cálculos del suavizamiento y encuentra los valores de α , β y γ que minimizan el ECM. Es importante mencionar que esta función empieza el suavizamiento en la observación uno del segundo año, es decir en Julio 2011, pues considera que el primer año se perdió en la estimación de los valores iniciales. Pero en este caso, los valores iniciales fueron calculados para que el alisamiento empiece en Julio 2010, por lo que para conseguir el alisamiento desde la primera observación se agregó a la serie todo un año de observaciones anteriores no reales antes del primer año, que el programa simplemente no considera en los cálculos.

La función fue llamada de la siguiente manera: `HoltWinters(serieiek,seasonal="mult",s.start=ien3, b.start=i2,l.start=g2)`, con lo que se aplica el método Holt-Winters al objeto `serieiek` que contiene a la serie con un año extra de observaciones, con el argumento `seasonal` se indica que utilice el método multiplicativo, y los parámetros iniciales de estacionalidad, tendencia y nivel se asignan utilizando los argumentos `s.start`, `b.start` y `l.start` respectivamente. Con este procedimiento se obtuvo un ECM de 243.4521, la gráfica 2.8 muestra en rojo el alisamiento y predicción de la serie bajo el procedimiento descrito.

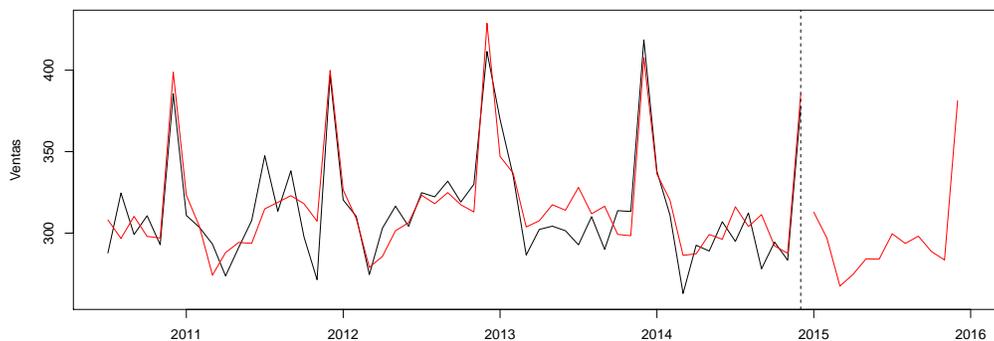


Figura 2.8: Alisamiento y predicción con Holt-Winters usando la función `HoltWinters`

Por otro lado se aplicó también el suavizamiento con la función `ets` la cual también uti-

liza el modelo Holt-Winters indicando usar el modelo “MAM”¹⁸. Como se mencionó, esta función incorpora una rutina que estima no solo los parámetros α , β y γ , sino también los valores iniciales a_0 y b_0 que minimizan el ECM. El resultado de este procedimiento llamado con la función `ets(serie2, model="MAM", damped=FALSE)`¹⁹ tuvo un ECM=229.2934 y el alisamiento puede verse en la figura 2.9.

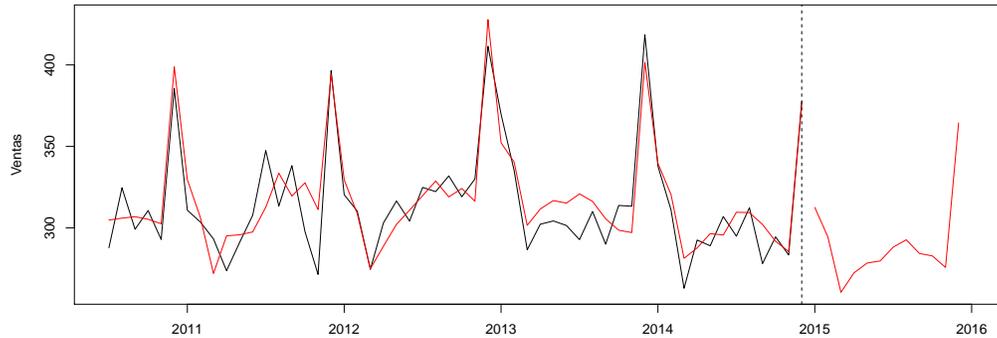


Figura 2.9: Alisamiento y predicción con Holt-Winters según la función `ets`.

¹⁸En los modelos State Space Models “MAM” indica que el Error (nivel), Tendencia y Estacionalidad (ETS) se comportan de manera multiplicativa, aditiva y multiplicativa respectivamente.

¹⁹Si el argumento `damped` se asigna en `TRUE` los cálculos del método son ligeramente diferentes a los de Holt-Winters, por lo que es necesario asignarlo en `FALSE`.

Capítulo 3

Metodología Box-Jenkins

3.1. Conceptos básicos

La metodología Box-Jenkins es la de más difusión en el enfoque estocástico del estudio de series de tiempo, se basa en teoría de procesos estocásticos. Las siguientes definiciones se usarán de manera reiterada en adelante.

Un *proceso estocástico* es una colección de variables aleatorias $\{X_t\}$ en donde el subíndice t puede ser una variable continua (adopta cualquier valor en un intervalo específico) o discreta (adopta sólo ciertos valores y existen *vacíos* entre ellos). Para el caso de este trabajo se considerará siempre que t es discreta y varía sobre los números enteros.

Una *realización* de un proceso estocástico es el conjunto los valores observados de las variables $\{X_t\}$ de un proceso estocástico. Dichos valores se denotan con letras minúsculas, así x_1 es el valor observado de la variable aleatoria X_1 , x_2 el valor observado de la variable X_2 , etc. En este sentido una serie de tiempo puede referirse al proceso estocástico como tal, o bien a una realización de este.

El propósito de este análisis es obtener inferencias de las series, a través del establecimiento de un modelo de probabilidad hipotético para representar los datos. Para los datos observados $\{x_t\}$ se entiende por *modelo de serie de tiempo* como la especificación de las distribuciones conjuntas (o posiblemente sólo las medias y covarianzas) de una secuencia de variables aleatorias $\{X_t\}$, para las cuales se da por sentado que $\{x_t\}$ es una realización.

Un modelo probabilístico completo para una serie de tiempo, especifica todas las distribuciones conjuntas $P[X_1 \leq x_1, \dots, X_n \leq x_n]$ para todos los posibles valores de x_i y para $n = \{1, 2, \dots\}$. Lo cual sólo ocurre cuando se comprende el mecanismo por el cual son generados los datos, por lo que en la práctica, difícilmente se obtienen tales distribuciones conjuntas. En vez de eso se especifican sólo el primer y segundo momentos de las distribuciones conjuntas, esto es, los valores esperados $E(X_t)$ y los productos esperados $E(X_{t+h}X_t)$, para $t = \{1, 2, \dots\}$ y $h = \{1, 2, \dots\}$.

Bajo estas consideraciones el tratamiento de una serie de tiempo según esta metodología es

el siguiente:

- 1) Graficar la serie y examinar los principales aspectos de la gráfica, particularmente si existen: tendencia, componente estacional, cambios de comportamiento y observaciones aisladas.
- 2) Homogeneizar varianza de los datos, remover la tendencia y el componente estacional para conseguir series estacionarias de segundo orden, las cuales se abordan en la siguiente sección. Una serie de tiempo es estacionaria si tiene propiedades estadísticas similares a la misma serie de tiempo desplazada h periodos de tiempo, para cualquier entero h , estas propiedades incluyen varianza y media constantes. El tema de estacionariedad se abordará con mayor detalle en la siguiente sección, pero ahora es relevante porque la teoría está diseñada para aplicarse sobre series estacionarias. Si bien en la realidad es muy raro encontrar series de este tipo, pueden aplicarse diferentes transformaciones a la serie original que permitan obtener una serie transformada estacionaria a la cual aplicar la metodología.

Un método frecuentemente usado para estabilizar la varianza de una serie es aplicar la transformación de Box-Cox:

$$f(X_t) = \begin{cases} \lambda^{-1}(X_t^\lambda - 1) & \text{para } X_t \geq 0, \lambda > 0 \\ \log(X_t) & \text{para } X_t > 0, \lambda = 0. \end{cases} \quad (3.1)$$

Algunos autores mencionan que λ se elige de tal forma que la relación $\frac{\sigma_t}{\mu_t^{1-\lambda}}$ sea constante, en donde σ_t y μ_t son la desviación estándar y la media de X_t , para $t = 1, 2, \dots, N$ con N el número de observaciones de la serie. Sin embargo, no es posible obtener una estimación de la varianza a partir de una sola observación. Lo que se hace es partir los datos en H grupos con el mismo número de observaciones, pueden excluirse n observaciones al principio o al final de la serie con el objetivo de procurar homogeneidad en los grupos (por ejemplo años incompletos), y se calcula la media y la desviación estándar muestral de estos grupos. Para el caso de esta transformación se denotan como S_h y \bar{X}_h a la desviación y media muestrales dentro del grupo h , para $h = 1, 2, \dots, H$, respectivamente. Luego se hace una tabla con los cálculos de $\frac{S_h}{\bar{X}_h^{1-\lambda}}$ para distintos valores de λ que van de -1 a 1. El valor de λ que genere una menor variación (que puede medirse con el coeficiente de variación) ente los valores $\frac{S_h}{\bar{X}_h^{1-\lambda}}$ será el valor de λ más adecuado para realizar la transformación.

Sobre remover la tendencia y la estacionalidad podría ocuparse algún método de los descritos en el capítulo anterior, pero en este capítulo se hablará de los procedimientos propios de la metodología Box-Jenkins, los cuales remueven esos componentes a través de la diferenciación, en las secciones 3.5 y 3.6 se ahonda sobre el tema.

- 3) Si se ocuparon los métodos vistos en el capítulo 2 o bien no fue necesario remover tendencia y estacionalidad, se escoge un modelo ARMA que se ajuste a la serie transformada, haciendo uso de las funciones muestrales de autocorrelación y autocorrelación parcial que se definirán en las secciones siguientes. Por otro lado si se utilizan modelos

ARIMA o SARIMA, antes de la elección de un modelo, se elige el grado de diferenciación, que en sí mismo es un parámetro de los modelos.

A lo largo de todo este capítulo se describen los modelos teóricos ARMA, ARIMA y SARIMA, así como los procedimientos que permiten encontrar el modelo con mejor ajuste a los datos.

- 4) La predicción de valores futuros de la serie se logra a través de la predicción de la serie transformada, y luego invirtiendo la transformaciones que se le hayan aplicado.

3.1.1. Funciones de autocovarianza y autocorrelación

Las funciones de autocovarianza y autocorrelación se definen para series estacionarias, a su vez las series estacionarias se definen a través de las siguientes dos funciones:

Sea $\{X_t\}$ una serie de tiempo con segundo momento finito, esto es, $E(X_t^2) < \infty$,

- La función media de $\{X_t\}$ es

$$\mu_X(t) = E(X_t) \quad (3.2)$$

- Y la función de covarianza de $\{X_t\}$ es

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))] \quad (3.3)$$

Entonces un **proceso estacionario** (débilmente o de segundo orden) $\{X_t\}$ es aquel que cumple las siguientes dos condiciones:

- (i) El valor de $\mu_X(t)$ no depende del tiempo t .
- (ii) El valor de $\gamma_X(t+h, t)$ no depende del tiempo t para cualquier valor h .

Por otro lado una serie de tiempo $\{X_t\}$ es estrictamente estacionaria si cumple que para toda $n > 0$ y para toda h en los enteros, los vectores (X_1, \dots, X_n) y $(X_{1+h}, \dots, X_{n+h})$ tienen las mismas distribuciones conjuntas. Sin embargo, cuando se hable sobre una serie de tiempo estacionaria, se referirá siempre a la estacionariedad débil.

Definido el concepto de series estacionarias, las funciones de autocorrelación y autocovarianza se definen de la siguiente forma:

Sea $\{X_t\}$ una serie de tiempo estacionaria, la **función de autocovarianza** (ACVF por sus siglas en inglés) de $\{X_t\}$ con retraso h es:

$$\gamma_X(h) = Cov(X_{t+h}, X_t). \quad (3.4)$$

En ella se puede apreciar que sólo se requiere el parámetro h , debido a que al ser una serie estacionaria la covarianza no depende de la variable tiempo t , si no solo del retraso h .

Las propiedades de γ_x son:

$$\begin{aligned}\gamma_X(0) &\geq 0, \\ |\gamma_X(h)| &\leq \gamma_X(0) \quad \text{para toda } h, \\ \gamma_X(h) &= \gamma_X(-h) \quad \text{para toda } h.\end{aligned}\tag{3.5}$$

Por la igualdad de la tercera línea de (3.5), se entiende que la función $\gamma(\cdot)$ es par, por lo cual es común que al hablar de h siempre se trate como un valor positivo.

Por otro lado la **función de autocorrelación** (ACF por sus siglas en inglés) de $\{X_t\}$ con retraso h es:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Corr}(X_{t+h}, X_t),$$

y mide la correlación entre la variable X_t y X_{t+h} . Estas funciones se pueden calcular siempre que se tenga información al menos sobre la media y varianza teóricas de todas las variables aleatorias en la serie. Por ejemplo, un proceso de *ruido blanco* es una secuencia de variables aleatorias normales no correlacionadas, cada una con media cero y varianza σ^2 (de donde se desprende que $\sigma^2 = E(X_t^2) < \infty$). Este es un proceso estacionario ya que su función media y de covarianza no dependen del tiempo t , sus funciones de autocovarianza y autocorrelación son las siguientes:

$$\begin{aligned}\gamma_X(h) &= \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases} \\ \rho_X(h) &= \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0. \end{cases}\end{aligned}$$

Cabe mencionar que es común usar la notación $\gamma_h = \gamma_X(h)$, para la función de autocovarianza y $\rho_h = \rho_X(h)$ para la función de autocorrelación.

Sin embargo, en la práctica no se tienen realmente los valores teóricos sobre la media y varianza de todas las variables X_1, \dots, X_n , sino que se cuenta solamente con las observaciones de dichas variables x_1, \dots, x_n (es decir, se cuenta con una realización del proceso). Por lo que se usan las versiones muestrales de las funciones ACVF y ACF.

Sean $\{x_t\}$ con $t = \{1, \dots, n\}$ las observaciones de una serie de tiempo, y \bar{x} la media muestral de $\{x_t\}$, con $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$, entonces la función de autocovarianza muestral, con retraso h es:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n,\tag{3.6}$$

y la función de autocorrelación muestral con retraso h es:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n.\tag{3.7}$$

La importancia de la ACF, como se verá en las siguientes secciones, está en su utilidad para identificar los parámetros relacionados con los términos de un modelo de promedios móviles.

3.1.2. Función de autocorrelación parcial

La función de autocorrelación parcial (PACF) de un proceso estacionario X_t , denotada por ϕ_{hh} , para $h = 1, 2, \dots$, es:

$$\phi_{11} = \text{Corr}(X_{t+1}, X_t) = \rho(1), \quad (3.8)$$

mientras que

$$\phi_{hh} = \text{Corr}(X_{t+h} - \widehat{X}_{t+h}, X_t - \widehat{X}_t), \quad \text{para } h \geq 2, \quad (3.9)$$

donde $\widehat{X}_{t+h} = \hat{\beta}_1 X_{t+h-1} + \hat{\beta}_2 X_{t+h-2} + \dots + \hat{\beta}_{h-1} X_{t+1}$, es la regresión de X_{t+h} sobre las variables intermedias entre X_{t+h} y X_t , por otro lado $\widehat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \dots + \beta_{h-1} X_{t+h-1}$, es la regresión de X_t sobre las variables intermedias entre X_t y X_{t+h} . Debido a que X_t es un proceso estacionario los coeficientes β_i son los mismos para ambas regresiones¹.

Como se observa, la PACF no es otra cosa que la correlación entre los residuos de la regresión de X_t sobre las variables intermedias $\{X_{t+1}, X_{t+2}, \dots, X_{t+h-1}\}$ y los residuos de la regresión de X_{t+h} sobre las mismas variables intermedias. De manera intuitiva, esta correlación entre X_t y X_{t+h} , tiene removido el efecto lineal de las variables intermedias $\{X_{t+1}, X_{t+2}, \dots, X_{t+h-1}\}$.

A partir de las ecuaciones de *Yule-Walker* en (3.36) se obtienen expresiones más sencillas, mostradas en (3.10).

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots, \quad \phi_{pp} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_2 \\ & & & \dots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & \rho_p \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_{p-2} \\ & & & \dots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (3.10)$$

La función de autocorrelación parcial muestral, sustituye los valores ρ_i por las correspondientes estimaciones muestrales $\hat{\rho}_i$.

¹Shumway, R. H. & Stoffer, D. S., *Time Series Analysis and Its Applications: With R Examples*, New York, USA: Springer, 2011, p. 165.

3.2. Modelos MA(q)

Un proceso $\{X_t\}$ es un proceso de promedios móviles MA(q) con media μ si $\{X_t\}$ puede escribirse como:

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \quad (3.11)$$

donde $\mu, \theta_1, \dots, \theta_q$ son constantes reales y $\{\varepsilon_t\}$ un proceso de ruido blanco. En realidad los pesos θ_i no necesariamente suman uno, por lo que el proceso no es un promedio de las variables del ruido blanco, más bien el modelo intenta expresar a X_t como una suma finita ponderada de choques aleatorios independientes.

Cuando $\mu = 0$ se dice que el proceso X_t está centrado en cero, esto no ocurre usualmente, pero un proceso puede ser modificado, de tal modo que una nueva serie $X_t^* = X_t - \mu$ para toda t , esté centrada en cero, lo que facilita el tratamiento de las series. Con el propósito de poder expresar el modelo de una forma más compacta, se utiliza el *operador retraso* B (del inglés Backward), que se define por las siguientes relaciones:

$$\begin{aligned} BX_t &= X_{t-1} \\ B^2X_t &= B(BX_t) = X_{t-2} \\ &\vdots \\ B^kX_t &= B(B^{k-1}X_t) = X_{t-k}. \end{aligned} \quad (3.12)$$

Un modelo con $\mu = 0$ puede escribirse como $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$ y en forma de operador:

$$X_t = \theta(B)\varepsilon_t \quad (3.13)$$

en donde $\theta(B) = (1 + \theta_1B + \theta_2B^2 + \cdots + \theta_qB^q)$ es el polinomio de promedios móviles. Cabe mencionar que B puede verse como la variable de un polinomio de grado q , en el cual sus soluciones pueden ser números reales o complejos, lo cual resulta conveniente para probar las propiedades de causalidad e invertibilidad deseables e incluso necesarias en la modelación de series de tiempo.

Los modelos MA(q) son **procesos causales**, estos son aquellos que pueden representarse como $X_t = \mu + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \dots$, y la suma del valor absoluto de los pesos converge, es decir $\sum_{i=0}^{\infty} |\psi_i| < \infty$. O bien si el proceso tiene media cero un proceso causal tiene la forma:

$$X_t = \psi(B)\varepsilon_t \quad (3.14)$$

donde $\psi(B) = (1 + \psi_1B + \psi_2B^2 + \dots)$. En el caso de los modelos MA(q), $\psi_k = 0$ para $k > q$, por lo que $\sum_{i=0}^{\infty} |\psi_i| = \sum_{i=1}^q |\theta_i| < \infty$, de ahí que todos los procesos de promedios móviles sean causales.

La explicación intuitiva del significado de causalidad se tratará en la siguiente sección ya que es una cualidad más relacionada a los procesos AR(p). Por el momento basta decir que todos los procesos causales son estacionarios, y por tanto **todos los procesos MA(q) son estacionarios**. Lo anterior también puede comprobarse al ver que su función media y de

autocovarianza no dependen del tiempo t . Para un proceso MA(q) centrado en cero se tiene que:

$$\begin{aligned} E[X_t] &= 0 \\ \gamma(0) &= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma_\varepsilon^2 \\ \gamma(h) &= \begin{cases} \sigma_\varepsilon^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h} & h = 1, \dots, q \\ 0 & h \geq q + 1 \end{cases} \end{aligned} \quad (3.15)$$

Se deduce entonces la ACF como:

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{i=0}^{q-h} \theta_i \theta_{i+h}}{\sum_{i=0}^q \theta_i^2} & h = 1, \dots, q \\ 0 & h \geq q + 1 \end{cases} \quad (3.16)$$

Otra importante propiedad de los procesos estocásticos es la **invertibilidad**. Un proceso es invertible si existen constantes $\{\pi_i\}$ tales que $\sum_{i=0}^{\infty} |\pi_i| < \infty$ de manera que el término de error ε_t pueda expresarse como una suma de los valores del proceso ponderados por los pesos $\{\pi_i\}$, es decir, $\varepsilon_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}$ con $\pi_0 = 1$, o bien, en forma de operador:

$$\pi(B)X_t = \varepsilon_t, \quad (3.17)$$

en donde $\pi(B) = (1 + \pi_1 B + \pi_2 B^2 + \dots)$.

¿Qué significa que un proceso sea invertible? Para los modelos MA(q) ocurre que varios modelos pueden tener la misma función de autocorrelación pero solo uno de ellos es invertible, o en otras palabras solo uno puede invertirse de un proceso media móvil a un proceso autorregresivo como los abordados en la siguiente sección. Un ejemplo que explican Shumway y Stoffer (2011) ilustra lo anterior, véanse los siguientes dos modelos:

$$\begin{aligned} X_t &= u_t + \frac{1}{5}u_{t-1}, & u_t &\sim \text{iid } N(0, 25) \\ Y_t &= v_t + 5v_{t-1}, & v_t &\sim \text{iid } N(0, 1) \end{aligned}$$

donde *iid* significa que los errores son variables independientes e idénticamente distribuidas, cada error u_t tiene una distribución normal con media 0 y varianza 25 mientras que cada error v_t es una variable normal estándar. A pesar de ello, usando la ecuación (3.15), ambos procesos tienen la misma función de autocovarianza:

$$\gamma(h) = \begin{cases} 26 & h = 0, \\ 5 & h = 1, \\ 0 & h > 1. \end{cases} \quad (3.18)$$

y por tanto la misma función de autocorrelación. Por otro lado,

$$\begin{aligned} u_t = X_t - \frac{1}{5}u_{t-1} &= X_t - \frac{1}{5}(X_{t-1} - \frac{1}{5}u_{t-2}) = \sum_{j=0}^{\infty} \left(-\frac{1}{5}\right)^j X_{t-j} \\ v_t = Y_t - 5v_{t-1} &= Y_t - 5(Y_{t-1} - 5v_{t-2}) = \sum_{j=0}^{\infty} (-5)^j Y_{t-j} \end{aligned} \quad (3.19)$$

De las ecuaciones (3.19) se deduce que al poner el error de cada serie en función de sus respectivos procesos, ambos pueden expresarse como una suma de las variables del proceso, pero sólo uno de los procesos es invertible, pues la condición de que la suma de los pesos sea finita ($\sum_{i=0}^{\infty} |\pi_i| < \infty$), se cumple sólo en el primer caso, en donde $\sum_{j=0}^{\infty} |(-\frac{1}{5})^j| < \infty$. En conclusión, la utilidad de la invertibilidad está en el hecho de que, si se tiene la función de autocovarianza para la cual exista más de un modelo que cumpla tener dicha función, por conveniencia se elige el modelo que sea invertible, pues éste es único.

Existen varias maneras para comprobar cuando un proceso de promedios móviles es invertible, una de las más intuitivas se apoya del siguiente resultado de series de potencias: $1 + a + a^2 + a^3 + \dots + a^n = \frac{1-a^{n+1}}{1-a}$, que cuando n tiende a infinito toma la forma:

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}, \quad (3.20)$$

lo cual ocurre siempre que la razón a sea menor que 1 en valor absoluto, en caso contrario la serie de potencias no se acerca a un valor finito. El proceso MA(1) se define por la siguiente ecuación:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad (3.21)$$

que también puede verse como $X_t = (1 + \theta_1 B)\varepsilon_t$, y suponiendo que $|\theta_1| < 1$, entonces:

$$\varepsilon_t = \frac{1}{1 + \theta_1 B} X_t = \sum_{i=0}^{\infty} (-\theta_1 B)^i X_t = \sum_{i=0}^{\infty} (-\theta_1)^i B^i X_t = \sum_{i=0}^{\infty} (-\theta_1)^i X_{t-i} \quad (3.22)$$

Es importante observar que la segunda igualdad de (3.22) es posible sólo si $|\theta_1| < 1^2$. Por otro lado cuando X_t se multiplica por la suma se ve afectada por el operador B^i , obteniendo la suma buscada en (3.17), y además la suma de los pesos $\sum_{i=0}^{\infty} \theta_1^i$ es finita por (3.20) ya que $|\theta_1| < 1$. En conclusión $|\theta_1| < 1$ es la condición necesaria para que X_t sea invertible.

Visto de otra forma, si el polinomio $\theta(z)$ correspondiente al proceso MA(1), donde $\theta(z) = 1 + \theta_1 z$, se iguala a cero, entonces la raíz $z = \frac{1}{\theta_1}$ debe ser mayor a 1 en valor absoluto para asegurar que $|\theta_1| < 1$, y con ello la invertibilidad del proceso.

Para generalizar el resultado a los modelos MA(q), debe tomarse en cuenta que todo polinomio de grado n con coeficientes reales y valor independiente unitario puede ser representado como una multiplicación de n polinomios de grado 1. Por tanto cualquier polinomio $\theta(z)$ de un proceso MA(q) puede verse como:

$$\theta(z) = (1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q) = (1 - \eta_1 z)(1 - \eta_2 z) \dots (1 - \eta_q z). \quad (3.23)$$

Según lo cual un proceso MA (q) puede verse como en la ecuación (3.24), siempre que cada raíz z de $(1 - \eta_i z) = 0$ con $i = 1, \dots, q$ sea mayor a 1 en valor absoluto, o bien si z es un

²En principio la condición a cumplirse es $|\theta_1 B| < 1$, Sabau comenta en su libro *Análisis Económico Dinámico* (2011, p. 97) que el operador B puede verse como una variable compleja con módulo unitario ($|B| = 1$), por lo que $|\theta_1 B| < 1$ se cumple siempre que $|\theta_1| < 1$.

número complejo, esté fuera del círculo unitario. Con que una sola raíz z tome valores dentro del círculo unitario, el inverso del polinomio dará lugar a una suma divergente, y se concluirá que el modelo no es invertible.

$$\varepsilon_t = \frac{1}{\theta(B)} X_t = \frac{1}{(1 - \eta_1 B)} \frac{1}{(1 - \eta_2 B)} \cdots \frac{1}{(1 - \eta_q B)} X_t \quad (3.24)$$

Por tanto, la condición para que cualquier proceso MA(q) sea invertible es que cada una de las raíces z del polinomio $\theta(z) = 0$ en (3.23), que pueden ser reales o complejas, estén fuera del círculo unitario.

Algunos autores mencionan que para que un proceso MA(q) sea invertible se requiere que la matriz de autocorrelación definida como:

$$P_N = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\ & & & \cdots & \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \cdots & 1 \end{pmatrix} \quad (3.25)$$

debe ser *definida positiva*, es decir, que el determinante de P_N y todos sus menores principales deben ser positivos, como en la ecuación (3.26).

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0, \begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0, \dots, \begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\ & & & \cdots & \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \cdots & 1 \end{vmatrix} > 0 \quad (3.26)$$

Sin embargo, es preferible el uso de los polinomios para probar invertibilidad, por su utilidad en las pruebas de raíces unitarias, importantes en la validación de supuestos y la estimación de parámetros.

Identificación del parámetro q

La manera de identificar un proceso MA(q) es a través de su ACF, ya que, como se ve en la fórmula (3.16) resulta que $\rho_h = 0$ para $h > q$, entonces se elige como q el valor de h más grande para el que $\rho_h \neq 0$. Y aunque no se tengan los valores de las θ_i , puede hacerse inferencia a partir de los valores muestrales de ρ . Sin embargo, debido a que la función $\hat{\rho}_h$ está afectada por variaciones muestrales, debe establecerse un criterio para saber desde que momento las autocorrelaciones muestrales son cero.

Una prueba de hipótesis ayuda a determinar cuando el parámetro ρ , por sus variaciones muestrales, pueda ser estadísticamente igual a cero, para lo cual debe considerarse la distribución de $\hat{\rho}_h$. Wei (2006) menciona que para n grande la distribución de $\hat{\rho}_h$ es aproximadamente normal con media ρ_h , y la varianza está dada por la fórmula de Bartlett ³. Para modelos en los que $\rho_h = 0$ a partir de $h > m$, como el modelo MA(q), se tiene que

$$\text{Var}(\hat{\rho}_h) \simeq \frac{1}{n} (1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_m^2).$$

³La fórmula de Bartlett puede consultarse en el libro de Brockwell & Davis (2002), *Introduction to time series and forecasting*, p. 61.

Pero ya que no se cuenta con los valores reales de ρ_i ($i = 1, \dots, m$) pueden usarse sus respectivas estimaciones muestrales de forma que:

$$S_{\hat{\rho}_h} \simeq \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \dots + 2\hat{\rho}_m^2)}$$

es el error estándar de $\hat{\rho}_h$.

Según el enfoque de pruebas de significancia, no se rechaza la hipótesis nula $H_0 : \rho_h = 0$ al 95 % de confianza si el estimador $\hat{\rho}_h$ está dentro del intervalo $(-1.96S_{\hat{\rho}_h}, +1.96S_{\hat{\rho}_h})$. Sin embargo es más común utilizar el intervalo de confianza de un ruido blanco. Es decir, si el proceso X_t es un ruido blanco entonces $\hat{\rho}_i \sim N(0, \frac{1}{n})$ para $i = 1, 2, \dots$, por lo que no se rechaza la hipótesis $H_0 : \rho_i = 0$ al 95 % de confianza si el estimador $\hat{\rho}_i$ está dentro del intervalo $(-1.96/\sqrt{n}, +1.96/\sqrt{n})$. Como este intervalo es más estrecho, si $\hat{\rho}_i$ está dentro de éste también está dentro de $(-1.96S_{\hat{\rho}_h}, +1.96S_{\hat{\rho}_h})$.

La figura (3.1) muestra las funciones ACF y PACF muestrales de una serie MA(1) simulada, las líneas punteadas indican los límites inferior y superior del intervalo $(-1.96/\sqrt{n}, +1.96/\sqrt{n})$ donde $n = 100$. Se comprueba que las variaciones muestrales provocan que $\hat{\rho}_i$ no sea exactamente cero para $i > 1$, pero al estar dentro del intervalo de confianza se consideran estadísticamente cero y se identifica que $q = 1$.

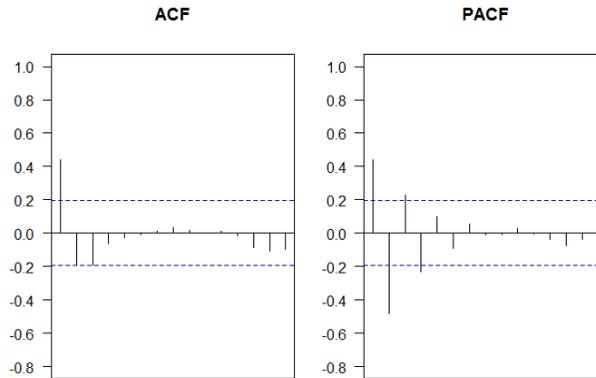


Figura 3.1: ACF y PACF muestrales de una serie MA(1) simulada.

3.3. Modelos AR(p)

Un proceso $\{X_t\}$ es un proceso autorregresivo de orden p con media μ si $\{X_t\}$ puede escribirse como:

$$X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad (3.27)$$

donde $\mu, \phi_1, \dots, \phi_p$ son números reales y $\{\varepsilon_t\}$ un proceso de ruido blanco. A este proceso se le denota AR(p). Se le conoce como autorregresivo debido a que la ecuación (3.27) se asemeja a la de una regresión lineal (múltiple si $p > 1$), pero con la diferencia de que en este caso las variables control no son variables independientes sino p variables anteriores al tiempo t del mismo proceso que sí presentan dependencia entre ellas. Por lo demás, también incluye los términos aleatorio (ε_t) y los coeficientes de regresión ($\mu, \phi_1, \dots, \phi_p$), propios del modelo de regresión.

El modelo AR(p), centrado (con $\mu = 0$), puede verse como $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t$, o también de la siguiente forma:

$$\phi(B)X_t = \varepsilon_t, \quad (3.28)$$

en donde $\phi(B)$ es el polinomio autorregresivo definido por:

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p). \quad (3.29)$$

El caso más simple es el modelo AR(1):

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \varepsilon_t \\ (1 - \phi_1 B)X_t &= \varepsilon_t, \end{aligned} \quad (3.30)$$

a partir de este pueden verse propiedades importantes que posteriormente pueden generalizarse a los modelos AR(p). Primero es destacable que el proceso es causal cuando se satisface la condición de que $|\phi_1| < 1$. Esto se comprueba de manera muy similar a como se comprueba la invertibilidad en los procesos MA(q):

$$X_t = \frac{1}{1 - \phi_1 B} \varepsilon_t = \sum_{i=0}^{\infty} (\phi_1 B)^i \varepsilon_t = \sum_{i=0}^{\infty} \phi_1^i B^i \varepsilon_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \quad (3.31)$$

Como se observa en (3.31) la condición de que $|\phi_1| < 1$ para que el proceso sea causal, se requiere desde la segunda igualdad. Cuando una serie es causal el modelo tiene sentido.

Tómese por ejemplo el caso de que $|\phi_1| > 1$ en (3.30), el cual no es causal (ni estacionario), se observaría que los valores X_t de la serie crecerían bastante en magnitud mientras el tiempo t corre. Esto se debe a que la suma $X_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$ no converge (en media cuadrada), pues $|\phi_1|^i$ crece sin límite, a esta clase de modelos se les llama explosivos. Una posible solución para convertir este modelo en estacionario es expresarlo como $X_t = \phi_1^{-1} X_{t+1} - \phi_1^{-1} \varepsilon_{t+1}$, de manera que cada elemento $X_t = -\sum_{i=0}^{\infty} \phi_1^{-i} \varepsilon_{t+i}$ converge debido a que $|\phi_1^{-1}| < 1$. Sin embargo, aunque se logró con ello que el proceso fuese estacionario, es también dependiente del futuro, y resultaría inútil querer trabajar con este modelo, pues requeriría conocer previamente los

datos futuros para poder predecir el futuro. Cuando $\phi_1 = 1$ se le llama un proceso de caminata aleatoria sin deriva, para el cual también es comprobable que no es causal ni estacionario.

Similar a lo que sucede en la ecuación (3.23), también el polinomio $\phi(z)$ puede escribirse como la multiplicación de p ecuaciones de grado uno:

$$\phi(z) = (1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p) = (1 - \zeta_1 z)(1 - \zeta_2 z) \dots (1 - \zeta_p z). \quad (3.32)$$

Y así como $|\phi_1|$ debe ser menor a uno en (3.31), las ζ_i deben de estar dentro del círculo unitario como condición para hacer el proceso causal. En conclusión:

- No todos los procesos AR(p) son causales.
- Si se trabaja con modelos que no sean causales, la dependencia del futuro llevaría a la invalidez de los resultados obtenidos en el análisis mientras que el uso de modelos causales asegura que el modelo tenga sentido.
- La causalidad de los modelos AR(p) se comprueba exactamente igual que la invertibilidad en los procesos MA(q). Así, las raíces del polinomio $\phi(z) = 0$, en (3.29), deben estar fuera del círculo unitario para que el proceso AR(p) sea causal.

Debido a que todo proceso causal es estacionario, puede deducirse que la condición para verificar que el proceso AR(1) sea estacionario es la misma para verificar que sea causal. Si el proceso está centrado en cero, entonces $E[X_t] = 0$ para toda t . Por otro lado la varianza del proceso puede obtenerse de la siguiente forma, primero nótese que $Cov(X_{t-h}, \varepsilon_t) = 0$ pues:

$$\begin{aligned} Cov(X_{t-h}, \varepsilon_t) &= E[X_{t-h}\varepsilon_t] - E[X_{t-h}]E[\varepsilon_t] \\ &= E\left[\sum_{i=1}^{\infty} \psi_i \varepsilon_{t-h-i} \varepsilon_t\right] - E[X_{t-h}]E[\varepsilon_t] \\ &\text{y por ser un proceso de ruido blanco} \\ &= 0 - E[X_{t-h}]0 = 0. \end{aligned} \quad (3.33)$$

Luego se obtiene la varianza de X_t :

$$\begin{aligned} Var(X_t) = \gamma_X(0) &= Var(\phi_1 X_{t-1} + \varepsilon_t) \\ &= Var(\phi_1 X_{t-1}) + Var(\varepsilon_t) + 2Cov(\phi_1 X_{t-1}, \varepsilon_t) \\ &= \phi_1^2 Var(X_{t-1}) + \sigma_\varepsilon^2 + 0 \\ &\text{por (3.33)} \\ &= \phi_1^2 \gamma_X(0) + \sigma_\varepsilon^2 \\ &\text{factorizando y despejando } \gamma_X(0) \text{ se obtiene} \\ (1 - \phi_1^2) \gamma_X(0) &= \sigma_\varepsilon^2 \\ \gamma_X(0) &= \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)} \end{aligned} \quad (3.34)$$

La función de autocovarianza se obtiene al multiplicar ambos lados de la igualdad (3.30) por X_{t-h} y tomando esperanza:

$$\begin{aligned}
Cov(X_{t-h}, X_t) &= \gamma_X(h) = E[X_{t-h}X_t] \\
&= E[X_{t-h}\phi_1 X_{t-1} + X_{t-h}\varepsilon_t] \\
&= Cov(X_{t-h}, \phi_1 X_{t-1}) + Cov(X_{t-h}, \varepsilon_t) \\
&= \phi_1 \gamma_X(h-1) + 0 \\
&\quad \text{repetiendo } h \text{ veces se obtiene} \\
&= \phi_1^h \gamma_X(0) \\
&\quad \text{y por (3.34) se obtiene} \\
&= \phi_1^h \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}.
\end{aligned} \tag{3.35}$$

Con ello, la función de autocorrelación es $\rho_X(h) = \phi_1^{|h|}$, la cual tiende a cero conforme las diferencias h se hacen más grandes. Véase que, en efecto, la condición de que $|\phi_1| < 1$ es necesaria para que el proceso sea estacionario, ya que evita que la varianza del proceso en (3.34) adquiera valores negativos.

Con el ejemplo del modelo AR(1), puede notarse que a diferencia de los modelos MA(q) no existe una sola ecuación con la que se obtenga la autocovarianza de cualquier proceso AR(p), pero es posible formar un sistema de ecuaciones con las cuales es sencillo calcular la autocovarianza y autocorrelación de cualquiera de estos. Se les conoce como las ecuaciones *Yule-Walker*:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \cdots + \phi_p \rho_{p-2} \\
&\quad \dots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p.
\end{aligned} \tag{3.36}$$

De esta forma, para un AR(2), según las ecuaciones (3.36), se tiene que $\rho_1 = \phi_1 + \phi_2 \rho_1$ y $\rho_2 = \phi_1 \rho_1 + \phi_2$, por tanto:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 \tag{3.37}$$

y siguiendo las mismas ecuaciones de forma iterativa, se observa que $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$ para $k \geq 3$. Estas ecuaciones pueden generalizarse para cualquier proceso AR(p) y se conocen como las ecuaciones Yule-Walker extendidas.

En cuanto a la propiedad de invertibilidad, ocurre que **todo proceso AR(p) es invertible**, ya que pueden representarse como $\pi(B)X_t = \varepsilon_t$, en este caso $\sum_{i=1}^{\infty} |\pi_i| = \sum_{i=1}^p |\phi_i|$ que por ser p sumandos, con p constante, es una suma finita. Esto quiere decir que si se tiene la ACF de un proceso autorregresivo, entonces existe un único modelo AR(p) que genera el la función de autocorrelación. Sin embargo, las funciones de autocorrelación muestrales no proporcionan

suficiente información para identificar el parámetro p .

Identificación del parámetro p

Obsérvese que la función de autocorrelación del AR(1) $\rho_X(h) = \phi_1^{|h|}$ y en general la de cualquier proceso AR(p) decae exponencialmente, y no como la ACF de cualquier modelo MA(q) que toma el valor cero en todos los valores de h mayores a q . Esto hace que, en comparación, sea más difícil identificar el parámetro p de los modelos AR(p), pues la función $\rho(h)$ no se trunca cuando h es mayor que p .

En los procesos AR(p), la función de autocorrelación parcial PACF, que se vio en la sección 3.3., se usa para identificar el orden p de un proceso autorregresivo. Véase, por ejemplo, el proceso AR(1), por definición $\phi_{11} = \rho(1) = \phi$. El cálculo de ϕ_{22} considera las variables X_t , X_{t+1} y X_{t+2} , para hacer dos regresiones lineales la primera tiene a X_{t+2} como variable dependiente y X_{t+1} como variable independiente, la segunda tiene a X_t como variable dependiente y X_{t+1} como variable independiente. Shumway y Stoffer (2011) muestran que en ambos casos el coeficiente de pendiente $\beta = \phi$, por lo que $\hat{X}_{t+2} = \phi X_{t+1}$ y $\hat{X}_t = \phi X_{t+1}$. Al calcular ϕ_{22} obtenemos:

$$\begin{aligned}\phi_{22} &= \text{Corr}(X_{t+2} - \hat{X}_{t+2}, X_t - \hat{X}_t) \\ &= \text{Corr}(X_{t+2} - \phi X_{t+1}, X_t - \phi X_{t+1}) \\ &= \text{Corr}(\varepsilon_{t+2}, X_t - \phi X_{t+1}) \\ &\text{por causalidad} \\ &= 0\end{aligned}\tag{3.38}$$

También muestran que la correlación parcial $\phi_{hh} = 0$ para $h > 1$, y de manera más general, en cualquier modelo AR(p) se tiene que $\phi_{hh} = 0$ para $h > p$.

El estadístico muestral $\hat{\phi}_{hh}$ visto en la fórmula (3.10), pierde observaciones conforme h es más grande, por lo que se recomienda no considerar un rezago h superior a $n/3$, con n el número de observaciones. Para determinar a partir de qué valor de h los coeficientes muestrales $\hat{\phi}_{hh}$ son estadísticamente cero se hace uso de su distribución. Los coeficientes $\hat{\phi}_{hh}$ se distribuyen de manera normal para una muestra grande, esto es:

$$\hat{\phi}_{\mathbf{hh}} \approx N(\phi, n^{-1}\sigma^2\Gamma_p^{-1})\tag{3.39}$$

donde $\hat{\phi}_{\mathbf{hh}}$ y $\phi_{\mathbf{hh}}$ son los vectores de autocorrelación parcial muestral y autocorrelación parcial respectivamente, y Γ_p es la matriz de varianzas y covarianzas del proceso⁵.

Así como en la identificación del parámetro q , suele usarse la banda de confianza de un ruido blanco, en el que sus coeficientes $\hat{\phi}_{hh}$ tienen una distribución asintóticamente normal con media cero y varianza $\text{Var}(\hat{\phi}_{hh}) \simeq \frac{1}{n}$.⁶ Si $\hat{\phi}_{hh}$ se encuentra en el intervalo $(-1.96/\sqrt{n}, +1.96/\sqrt{n})$

⁴Shumway & Stoffer, *op. cit.*, p. 106.

⁵Brockwell & Davis (2002), *Introduction to time series and forecasting*, p. 141.

⁶Wei W., 2006, *Time Series Analysis Univariate and Multivariate Methods*. Boston, USA: Pearson Addison Wesley, p. 22.

entonces se considera que es estadísticamente cero. Este intervalo es más estrecho que el que se obtendría con la distribución mencionada en el párrafo anterior, el modelo se identifica por el hecho de que $\phi_{hh} = 0$ cuando h es mayor que p . Nótese que es el mismo intervalo que se usa para identificar cuando $\rho_h = 0$ y el uso de la banda de confianza de un ruido blanco se justifica de manera similar. Al respecto, Brockwell y Davis (2002) justifican el uso de la banda de confianza de un ruido blanco de una forma distinta, explican que si se ajusta un modelo autorregresivo de orden m a un modelo $AR(p)$ con $m > p$, entonces $\hat{\phi}_{mmm}$ (que es el valor muestral de la autocorrelación parcial de retraso m) tiene una distribución aproximadamente normal con media cero y varianza $1/n$; por tanto si el modelo $AR(p)$ es el adecuado para los datos, entonces los valores $\hat{\phi}_{kk}$ con $k > p$ deberán ser compatibles con la distribución $N(0, 1/n)$.

La figura (3.2) muestra las funciones ACF y PACF muestrales de una serie $AR(1)$ simulada, las líneas punteadas indican los límites inferior y superior del intervalo $(-1.96/\sqrt{n}, +1.96/\sqrt{n})$ donde $n = 100$. Se comprueba que las variaciones muestrales provocan que $\hat{\phi}_{hh}$ no sea exactamente cero para $h > 1$, pero al estar dentro del intervalo de confianza se consideran estadísticamente cero y se identifica que $p = 1$.

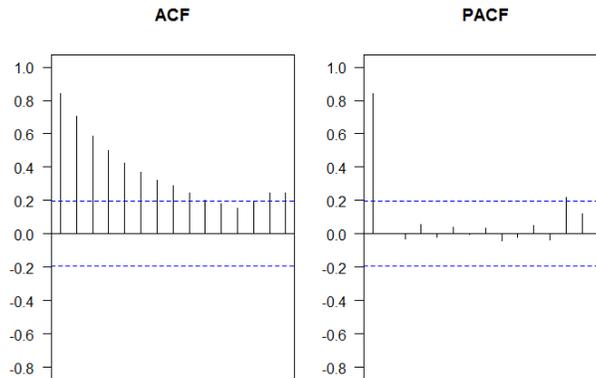


Figura 3.2: ACF y PACF muestrales de una serie $AR(1)$ simulada.

Pruebas de raíces unitarias

En la práctica, para determinar si la serie es estacionaria, no se cuenta con el polinomio sino con los valores de la serie. Es por ello que las pruebas de raíces unitarias determinan si los datos indican la presencia estadística de alguna raíz ζ_i fuera del círculo unitario. Retomando el caso del proceso $AR(1)$, las hipótesis a contrastar son:

$$H_0 : |\phi_1| \geq 1 \quad vs \quad H_1 : |\phi_1| < 1, \quad (3.40)$$

en donde H_0 indica que el proceso no es estacionario, mientras que H_1 indicaría que si lo es. Sin embargo, lo que en realidad importa es comprobar la hipótesis alternativa, y cómo es relativamente más sencillo probar una hipótesis con una distribución específica, entonces la prueba se realiza con las siguientes hipótesis:

$$H_0 : \phi_1 = 1 \quad vs \quad H_1 : |\phi_1| < 1 \quad (3.41)$$

Motivo por el cual a estas pruebas se les conoce como “pruebas de raíces unitarias”. Una de las más famosas es la *prueba de Dickey-Fuller* (DF). En esta se visualiza la diferencia del proceso como una regresión lineal, es decir:

$$\begin{aligned}
 \nabla X_t &= X_t - X_{t-1} \\
 &= \phi_1 X_{t-1} + \varepsilon_t - X_{t-1} \\
 &= (\phi_1 - 1)X_{t-1} + \varepsilon_t \\
 &= \varphi_1 X_{t-1} + \varepsilon_t.
 \end{aligned}
 \tag{3.42}$$

Según esta ecuación, puede obtenerse un estimador para $\varphi_1 = \phi_1 - 1$ mediante una regresión lineal, en la misma forma en que se obtuvo el parámetro de la pendiente en el Capítulo 1, y podrán contrastarse las hipótesis:

$$H_0 : \varphi_1 = 0 \quad vs \quad H_1 : \varphi_1 \neq 0, \tag{3.43}$$

que son equivalentes a las de (3.41) (de aceptar $H_0 : \varphi_1 = 0$ querrá decir que ϕ_1 es una raíz fuera del círculo unitario). De hecho, el estadístico t que se usó en el test de significancia de la pendiente en la regresión lineal, es el mismo que se usa en esta prueba para contrastar las hipótesis. Sin embargo, no se compara contra los valores críticos de una distribución t , pues estos no son adecuados en presencia de inestabilidad bajo la hipótesis nula. Los valores críticos propuestos por Dickey-Fuller para esta prueba, al 10 %, 5 % y 1 % de significancia son: -2.5808, -2.8877 y -3.4902; de manera que se rechaza la hipótesis nula al $(1 - \alpha) \%$ de confianza si el estadístico t es menor que el valor crítico al $\alpha \%$ de significancia.

La generalización este test a los modelos AR(p) se le conoce como el test Dickey-Fuller aumentado (DFA), siguiendo básicamente el mismo proceso de visualizar el proceso diferenciado como una regresión lineal, el parámetro pendiente φ ahora es:

$$\varphi = \phi_1 + \phi_2 + \dots + \phi_p - 1 \tag{3.44}$$

se deduce de (3.32), que $\varphi = -\phi(1)$, y que si se acepta que $H_0 : \varphi = -\phi(1) = 0$ indicará que al menos una de las raíces ζ_i es unitaria. Sin embargo, esta prueba tiene las desventajas de que no se sabe a priori qué valor de p usar, y para que la prueba sea válida debe comprobarse que los errores ε_t se comporten como un ruido blanco.

Existen otras pruebas que intentan atacar estos problemas, como la Phillips-Perron (el llamado *PP test*) el cual no tiene una distribución tan restrictiva a los errores, sin embargo, se sabe que en general las pruebas de raíces unitarias tienden a encontrar alguna raíz unitaria donde no la hay. Una de las recomendaciones frecuentes es corroborar por otros medios si se está tratando con una serie no estacionaria, como la inspección visual. Por ejemplo las series en donde el ACF no decrece rápidamente tienen alguna tendencia lo que las hace no estacionarias.

3.4. Modelos ARMA(p, q)

Un proceso $\{X_t\}$ es un proceso autorregresivo de media móvil de ordenes p y q , con media $\mu = 0$ si $\{X_t\}$ puede expresarse como:

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

o bien en función de los polinomios $\theta(B)$ y $\phi(B)$ se tiene que $\phi(B)X_t = \theta(B)\varepsilon_t$, aunque también es usual encontrar X_t despejado en la parte izquierda de la ecuación:

$$X_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t \quad (3.45)$$

donde $\{\varepsilon_t\}$ es un proceso de ruido blanco y los polinomios $\theta(B)$ y $\phi(B)$ no tienen raíces en común. A este proceso se le denota ARMA(p, q). La existencia de una raíz común implicaría que el modelo no tuviese órdenes (p, q) sino $(p - 1, q - 1)$.

De manera intuitiva el modelo puede verse como la mezcla de los dos procesos antes descritos. Al respecto de ellos puede observarse la existencia de una dualidad entre los procesos MA(q) y AR(p): los primeros son siempre causales y su invertibilidad se comprueba al estar todas las raíces z del polinomio $\theta(z) = 0$ fuera del círculo unitario, mientras que los segundos son siempre invertibles y son causales cuando todas las raíces z del polinomio $\phi(z) = 0$ están fuera del círculo unitario. Generalmente dichos procesos no se encuentran por separado de manera natural, sino que suelen estar mezclados como en el modelo ARMA(p, q). Este es un proceso tan general, que fue establecido por Wold (1938) que todo proceso estacionario puede ser representado como un proceso ARMA(p, q).

Para que un proceso ARMA sea causal (y por ende estacionario) se requiere que las raíces de $\phi(z) = 0$ estén fuera del círculo unitario. Del mismo modo será invertible si las raíces de $\theta(z) = 0$ también están fuera del círculo unitario.

Identificación de los parámetros p y q

Cuadro 3.1: Comportamiento de la ACF y la PACF para los modelos AR, MA y ARMA

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Disminuye lentamente	Se trunca después del retraso q	Disminuye lentamente después del retraso $(q - p)$
PACF	Se trunca después del retraso p	Disminuye lentamente	Disminuye lentamente después del retraso $(p - q)$

Las funciones muestrales de autocorrelación y autocorrelación parcial son suficientes para identificar si las observaciones de una serie de tiempo tienen el comportamiento de un modelo AR(p) o un MA(q), e incluso permiten encontrar el valor de p y q más adecuados. En los modelos ARMA(p, q) dichas funciones sólo permiten saber si las observaciones se adaptan al

modelo ARMA, pero es menos factible usarlas para identificar los valores de p y q .

Como se muestra en el cuadro 3.1 las funciones ACF y PACF para un proceso ARMA no se truncan sino que van disminuyendo. Esto se debe a que se supone al proceso como causal e invertible, entonces la parte autorregresiva tiene una representación media móvil infinita por lo que la función de autocorrelación no se trunca, del mismo modo la parte media móvil tiene una representación autorregresiva infinita y por tanto la función de autocorrelación parcial no se trunca.

Tsay(2005) presenta una función que permite identificar los órdenes p y q de un modelo ARMA, llamada función de autocorrelación extendida (EACF por sus siglas en inglés). Un ejemplo ayudará a entender el funcionamiento de dicha función, supóngase que se tiene un modelo ARMA(1,1), es decir, $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$, por tanto los residuales $W_{1,t} = X_t - \hat{\phi} X_{t-1}$ se comportan como un proceso MA(q). A partir de W_t se obtienen estimaciones de $\rho_{(1,j)}$ con $j \geq 1$ donde el subíndice $(1, j)$ quiere decir que se estiman las correlaciones extendidas de retraso j de primer orden, ya el modelo autorregresivo de la serie es de orden 1.

En una serie de datos no se sabe de antemano los valores de p y q , así que se calcula la EACF de ordenes autorregresivos mayores que 1, es decir para el orden 2, se calculan los residuales $W_{2,t} = X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2}$ y las autocorrelaciones muestrales extendidas $\hat{\rho}_{2,j}$, con $j \geq 1$. Los estimadores de ϕ_1 y ϕ_2 pueden obtenerse a partir de las ecuaciones Yule-Walker generalizadas, de manera análoga para órdenes autorregresivos mayores. La información de los valores $\hat{\rho}_{(i,j)}$ se presenta en forma tabulada como en el cuadro 3.2.

AR	MA						
	1	2	3	4	5	6	7
0	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_6$	$\hat{\rho}_7$
1	$\hat{\rho}_{(1,1)}$	$\hat{\rho}_{(1,2)}$	$\hat{\rho}_{(1,3)}$	$\hat{\rho}_{(1,4)}$	$\hat{\rho}_{(1,5)}$	$\hat{\rho}_{(1,6)}$	$\hat{\rho}_{(1,7)}$
2	$\hat{\rho}_{(2,1)}$	$\hat{\rho}_{(2,2)}$	$\hat{\rho}_{(2,3)}$	$\hat{\rho}_{(2,4)}$	$\hat{\rho}_{(2,5)}$	$\hat{\rho}_{(2,6)}$	$\hat{\rho}_{(2,7)}$
3	$\hat{\rho}_{(3,1)}$	$\hat{\rho}_{(3,2)}$	$\hat{\rho}_{(3,3)}$	$\hat{\rho}_{(3,4)}$	$\hat{\rho}_{(3,5)}$	$\hat{\rho}_{(3,6)}$	$\hat{\rho}_{(3,7)}$
4	$\hat{\rho}_{(4,1)}$	$\hat{\rho}_{(4,2)}$	$\hat{\rho}_{(4,3)}$	$\hat{\rho}_{(4,4)}$	$\hat{\rho}_{(4,5)}$	$\hat{\rho}_{(4,6)}$	$\hat{\rho}_{(4,7)}$
5	$\hat{\rho}_{(5,1)}$	$\hat{\rho}_{(5,2)}$	$\hat{\rho}_{(5,3)}$	$\hat{\rho}_{(5,4)}$	$\hat{\rho}_{(5,5)}$	$\hat{\rho}_{(5,6)}$	$\hat{\rho}_{(5,7)}$

Cuadro 3.2: Función EACF muestral, en donde cada $\hat{\rho}_{(i,j)}$ es el valor de la autocorrelación muestral de retraso j y orden autorregresivo i .

Para un modelo ARMA(p, q), la tabla 3.2 tendrá un patrón “triangular” en donde los valores $\hat{\rho}_{(i,j)}$ dentro de este triángulo, son estadísticamente cero y la posición (i, j) del pico superior izquierdo de este triángulo corresponderá con los órdenes (p, q) del modelo ARMA. Por ejemplo, en un modelo ARMA(1,1) todas las autocorrelaciones $\hat{\rho}_i$ serán significativas, las autocorrelaciones $\hat{\rho}_{(1,j)}$ serán no significativas para $j \geq 2$, las autocorrelaciones $\hat{\rho}_{(2,j)}$ serán no significativas para $j \geq 3$, etc. En su forma simplificada la tabla sólo muestra O's y X's denotando valores que son cero y distintos de cero respectivamente.

Otra opción de selección de los órdenes para estos modelos, a través de una opción sistemática es el *Criterio de Información de Akaike Corregido* (AICC) mostrado en (3.46). El AICC se usa como un método de comparación de modelos basado en la teoría de la información y utiliza para su cálculo la estimación de los parámetros ϕ y θ por *Máxima Verosimilitud* (MV). Este último es un método de estimación de parámetros que funciona bajo la idea de tomar como estimación de los parámetros aquellos que maximicen la probabilidad de obtener la muestra observada, dicha probabilidad es la llamada función de verosimilitud $L(\bullet)$. El índice AICC se calcula como:

$$AICC = -2\ln[L(\hat{\phi}_p, \hat{\theta}_q, \hat{\sigma}_\varepsilon^2)] + \frac{2(p+q+1)n}{(n-p-q-2)} \quad (3.46)$$

En la ecuación (3.46), n es el número de observaciones, p y q los órdenes del modelo, y $\hat{\phi}_p$, $\hat{\theta}_q$, y $\hat{\sigma}_\varepsilon^2$ son los vectores de los parámetros estimados de los por MV, por lo que $L(\hat{\phi}_p, \hat{\theta}_q, \hat{\sigma}_\varepsilon^2)$ denota el máximo valor que alcanza la función de verosimilitud.

La estrategia para la identificación de p y q por el criterio de AICC, podría sintetizarse de la siguiente manera: establecer un rango de valores posibles para p y q , posteriormente estimar a través de MV los parámetros ϕ_i ($i = 1, \dots, p$), θ_j ($j = 1, \dots, q$) y σ_ε^2 para los posibles modelos ARMA(p, q). Por último elegir el modelo con la combinación de órdenes p y q que minimice el valor del estadístico AICC. Por su construcción, el término negativo recompensa la bondad de ajuste y el término positivo penaliza el sobreajuste, razón por la que se escoge el modelo con el valor más pequeño.

El valor de este estadístico no tiene unidades y carece de sentido cuando se tiene el de un solo modelo, pero cuando se tienen varios modelos, permite compararlos e indica que el que tenga el valor más bajo es el que explica mejor los datos con el menor número de parámetros, siendo el más cercano a la realidad. Esta indefinición de la escala, puede resultar en que los valores sean incluso negativos, en tal caso se sigue el mismo criterio de elegir el modelo con el menor AICC.

Raíces unitarias en los modelos ARMA

La existencia de raíces unitarias en cualquiera de los dos lados de la ecuación tendría implicaciones importantes. Debido a que los procesos ARMA(p, q) modelan procesos estacionarios, la presencia de una raíz unitaria en el polinomio autorregresivo invalidaría los resultados de la estimación. En la ecuación (3.47) se muestra un modelo ARMA en el que una de las p raíces de la parte autorregresiva es unitaria, modelar este tipo de series no estacionarias es un tema del siguiente sección.

$$\phi_{p-1}(B)(1-B)X_t = \theta_q(B)\varepsilon_t \quad (3.47)$$

Por otro lado la presencia de una raíz unitaria en el polinomio de promedios móviles genera no invertibilidad. Esto significa que el proceso no tiene la representación única, y la causa de esta es la sobre diferenciación. Como se aprecia en la ecuación (3.48), aplicar una diferencia extra al proceso ARMA(p, q) se está agregando una raíz unitaria al polinomio del proceso MA.

$$\phi_p(B)\nabla X_t = \theta_q(B)(1-B)\varepsilon_t \quad (3.48)$$

3.5. Modelos ARIMA

A partir de esta sección se describen modelos para procesos no estacionarios. Al tratar con estos, como primera instancia suelen aplicarse transformaciones a los datos para obtener series estacionarias y con decrecimiento rápido de la función de autocorrelación. El primer modelo por analizar es el proceso ARIMA.

Un proceso $\{X_t\}$ es un proceso autorregresivo integrado de promedios móviles ARIMA(p, d, q) si $\{X_t\}$ puede expresarse como:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad \text{para } d \geq 1 \quad (3.49)$$

donde ε_t es un proceso de ruido blanco. El modelo indica que la serie diferenciada d veces puede modelarse como un proceso ARMA(p, q). Es usual escribir el modelo usando el operador $\nabla = (1 - B)$, por lo que (3.49) es equivalente a $\phi(B)\nabla^d(X_t) = \theta(B)\varepsilon_t$, $d \geq 1$. Véase que $\nabla X_t = X_t - X_{t-1}$.

Nótese que el modelo ARIMA con $d = 1$ realiza estimaciones a partir de la serie $(1 - B)X_t$ mientras que el modelo ARMA en (3.47) hace estimaciones sobre X_t , es decir, aunque la expresión matemática es la misma, el proceso ARIMA está efectivamente diferenciado.

Para mirar el proceso desde una perspectiva intuitiva deben recordarse los siguientes aspectos:

- Comúnmente las series no son estacionarias por alguno(s) de los siguientes tres motivos: 1) exhiben algún tipo de tendencia, 2) su varianza no es constante o 3) presentan de algún factor tipo semi determinista como puede ser la estacionalidad.
- Cuando una serie presenta una tendencia de grado d , esta puede removerse de la serie tomando las diferencias de orden d (como se vio en el capítulo anterior).

Por tanto, intuitivamente $\{X_t\}$ es un proceso ARIMA si $\nabla^d X_t$, el proceso con tendencia removida, puede modelarse como un proceso ARMA(p, q) causal. Los modelos ARIMA permiten modelar series que originalmente no son estacionarias, cuando la razón de la no estacionariedad es la presencia de una tendencia lineal o polinomial.

Además el parámetro μ como el de la ecuación (3.27), en estos modelos no tiene el mismo papel que en los modelos anteriores, ya que μ no está relacionado con la media del proceso sino que representa el término de pendiente determinística. Tómese como ejemplo el siguiente proceso con $d = 1$, $p = 0$ y $q = 0$ (también conocido como caminata aleatoria con deriva):

$$\begin{aligned} (1 - B)X_t &= \mu + \varepsilon_t \\ \text{entonces} \\ X_t &= X_{t-1} + \mu + \varepsilon_t \\ X_t &= X_{t-2} + 2\mu + \varepsilon_t + \varepsilon_{t-1} \\ &\vdots \\ &= X_k + (t - k)\mu + \sum_{j=k+1}^t \varepsilon_j \quad \text{para } t > k. \end{aligned} \quad (3.50)$$

En tal modelo se observa como el parámetro μ es ahora la pendiente de una tendencia determinística lo cual ocurre para cualquier modelo ARIMA cuando $d > 0$. Wei (2006) explica que para t grande este término se hace muy dominante en el comportamiento de la serie, forzándola a seguir un patrón determinístico, por lo que en general cuando $d > 0$ se asume que $\mu = 0$, a menos que sea claro en los datos o por la naturaleza del problema, que un componente determinístico sea realmente necesario.

Identificación del parámetro d

Sabau (2011) explica el siguiente algoritmo para determinar el orden de integración se logra mediante una aplicación iterativa de la prueba DFA de la siguiente manera⁷:

- 1) Fijar $d = 0$.
- 2) Aplicar la prueba DFA a la serie $\nabla^d X_t$.
- 3) En caso de no rechazar la hipótesis nula $H_0 : \varphi_1 = 0$, aumentar el grado de diferenciación d en una unidad, y regresar al paso dos.
- 4) En caso de aceptar la hipótesis alternativa $H_1 : \varphi_1 < 0$, tomar el último valor de d como el orden de integración.

Si no se diferencia lo suficiente no se alcanzará a remover la tendencia y seguirá siendo un proceso no estacionario, con presencia de raíces unitarias, ante la duda de la existencia de estas hay que volver a diferenciar. Por otro lado diferenciar de más el proceso provoca no invertibilidad como se aprecia en la ecuación (3.48), ante la duda de una raíz unitaria en el polinomio MA hay que diferenciar menos. Si existe la duda legítima mejor diferenciar de más que de menos ya que la no invertibilidad es menos grave que la no estacionariedad.

Wei (2006) menciona otra forma de identificar el valor del parámetro d de esta forma: si la ACF muestral decae lentamente y la PACF muestral se corta después del retraso 1, indica que la diferenciación es necesaria⁸.

Como dato adicional, Shumway y Stoffer (2011) explican que si se ha aplicado el logaritmo a la serie y luego se aplica una diferencia, la serie transformada se interpreta (por propiedades de los logaritmos) como la serie de las tasas de crecimientos.

3.6. Modelos SARIMA

Un proceso $\{X_t\}$ es un modelo estacional autorregresivo integrado de promedios móviles (SARIMA) $(p,d,q) \times (P,D,Q)$, con periodo s si la serie diferenciada $Y_t = (1 - B)^d(1 - B^s)^D X_t$ es un proceso ARMA definido por:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)\varepsilon_t \quad (3.51)$$

⁷Sabau H., *Análisis Económico Dinámico*, México: Universidad Iberoamericana, 2011, p. 125.

⁸Wei W., 2006, *Time Series Analysis Univariate and Multivariate Methods*. Boston, USA: Pearson Addison Wesley, p.109

con $\phi(B)$ y $\theta(B)$ definidos como antes y adicionalmente:

$$\begin{aligned}\Phi(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \\ \Theta(B^s) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs},\end{aligned}\tag{3.52}$$

donde s representa el número de observaciones por periodo, por ejemplo, si la serie fuese recogida mensualmente $s = 12$, si fuese recogida trimestralmente $s = 4$, etc. Sea X_t una serie con comportamiento estacional a la que se le ajusta un modelo ARIMA como en la ecuación (3.53), ocurre que el término de error b_t no se comporta como un ruido blanco. Esto es, los polinomios $\phi(z)$ y $\theta(z)$ modelan el comportamiento no estacional de la serie, es decir, el comportamiento entre observaciones consecutivas, por ejemplo, el comportamiento entre meses del mismo año si la serie fuese mensual, pero no hay en el modelo elementos que expliquen el comportamiento estacional. En el modelo SARIMA se intenta solucionar dicho problema tratando a b_t como otra serie ARIMA la cual modele el comportamiento estacional de X_t , en el ejemplo de la serie mensual modelaría el comportamiento de los meses entre años.

$$\phi(B)(1 - B)^d X_t = \theta(B)b_t\tag{3.53}$$

$$\Phi(B^s)(1 - B^s)^D b_t = \Theta(B^s)\varepsilon_t\tag{3.54}$$

En (3.54) los parámetros P , D y Q tienen efecto sobre la serie b_t y ε_t corresponde a un proceso de ruido blanco. De lo anterior puede inferirse que el término de integración D tiene la función de eliminar, en caso de existir, alguna tendencia en el comportamiento estacional; sin embargo, rara vez es mayor que uno, pues el comportamiento estacional suele ser muy similar año con año. Por otro lado los polinomios $\Phi(B^s)$ y $\Theta(B^s)$ definidos como en (3.52), tienen la función de modelar la parte autorregresiva y media móvil del comportamiento estacional de la serie, por ejemplo en el caso de una serie mensual, estos modelarán las relaciones de un mes con el mismo mes de años anteriores pues en los polinomios solo hay elementos no nulos cada 12 diferencias. Brockwell y Davis (2002) mencionan que en raras ocasiones D es mayor que 1 mientras que P y Q no suelen ser mayores a 3.

Si se sustituye (3.54) en (3.53), y se define $Y_t = (1 - B)^d(1 - B^s)^D X_t$, se obtendrá el modelo SARIMA presentado en (3.51), en el que se observa que los elementos estacionales están multiplicados por los respectivos elementos no estacionales, razón por la que es conocido como el modelo ARIMA estacional multiplicativo.

En cuanto a las propiedades de causalidad e invertibilidad, el proceso Y_t es causal si y sólo si las raíces de los polinomios $\phi(z)$ y $\Phi(z)$ están fuera del círculo unitario. La misma condición se pide en los polinomios $\theta(z)$ y $\Theta(z)$ para que el proceso sea invertible.

Puede ocurrir que una serie no presente tendencia de manera estacional ni de forma no estacional, es decir, que el proceso tenga media, varianza y comportamiento estacional constantes. Este comportamiento se modela mediante un modelo tipo $SARMA(p, q) \times (P, Q)$, su representación matemática corresponde a:

$$\phi(B)\Phi(B^s)X_t = \theta(B)\Theta(B^s)\varepsilon_t.\tag{3.55}$$

En (3.55) puede verse que el ajuste es directamente sobre la variable X_t y no sobre una variable diferenciada, pues al no presentar tendencias los parámetros d y D son cero. Este modelo puede verse como un caso particular del modelo ARMA con parámetros $p + sP$ y $q + sQ$, con una gran cantidad de coeficientes intermedios nulos.

	$AR(P)_s$	$MA(Q)_s$	$ARMA(P, Q)_s$
ACF*	Disminuye lentamente para retrasos $h = ks$ con $k = 1, 2, \dots$,	Se trunca después del retraso Qs	Disminuye lentamente para retrasos $h = ks$ con $k = 1, 2, \dots$,
PACF*	Se trunca después del retraso Ps	Disminuye lentamente para retrasos $h = ks$ con $k = 1, 2, \dots$,	Disminuye lentamente para retrasos $h = ks$ con $k = 1, 2, \dots$,

* Los valores en retrasos $h \neq ks$ para $k = 1, 2, \dots$, son cero.

Cuadro 3.3: Comportamiento de las funciones ACF y PACF para modelos puramente estacionales

Identificación de los parámetros P, D y Q

El cuadro 3.3 sirve como guía en la identificación de los parámetros cuando se trata de procesos puramente estacionales, en los que únicamente P y Q son distintos de cero⁹. Como se observa, la identificación de los órdenes P y Q de los procesos SARMA, es muy similar a como se identifican en los procesos ARMA, la diferencia es que en los modelos SARMA se pone atención a las funciones ACF y PACF en las diferencias múltiplo de la estacionalidad.

Puede haber más de una opción que parezca admisible en la modelación del proceso, por lo cual una posibilidad es elegir aquella combinación de los parámetros que proporcione el estadístico AICC más bajo.

Sobre la identificación de los parámetros de un modelo SARIMA Shumway y Stoffer (2011) comentan que es conveniente identificar primero los órdenes de integración (d y D). En un ejemplo identifican primero el orden no estacional, y si con este primer orden de diferenciación basta para que el proceso tengan media y varianza constante entonces se asume que $D = 0$, y se procede a identificar los órdenes P y Q como en el modelo SARMA (con la guía del cuadro 3.3). De hecho plantean varias opciones sobre los valores que pueden tomar P , Q así como de p y q , para posteriormente comparar los modelos posibles con el estadístico AICC¹⁰.

⁹Shumway & Stoffer, *op. cit.*, p. 155.

¹⁰Shumway & Stoffer, *op. cit.*, pp. 158-161.

3.7. Predicción

Como explican Brockwell y Davis (2002), el problema de la predicción consiste en estimar el valor de X_{n+h} , con $h > 0$, de una serie de tiempo estacionaria en función de los valores X_n, \dots, X_1 , con media μ y función de autocovarianza γ conocidas. Explican que la estimación de X_{n+h} (denotada por $P_n X_{n+h}$ o \hat{X}_{n+h}) es de la forma:

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

y la meta es encontrar los valores a_0, \dots, a_n de tal que se minimice el error cuadrático medio, que es:

$$S(a_0, \dots, a_n) = E(X_{n+h} - \hat{X}_{n+h})^2. \quad (3.56)$$

Un primer camino a tomar en el proceso de predicción es tratar de minimizar directamente (3.56). Después de aplicar $n+1$ derivadas parciales se llega a que $a_0 = \mu(1 - \sum_{i=1}^n a_i)$, mientras que los valores de a_1, \dots, a_n se obtienen a partir del siguiente sistema de ecuaciones:

$$\Gamma_n(a_1, \dots, a_n)' = \gamma_n(h),$$

donde Γ_n es la matriz de autocovarianzas de (X_1, \dots, X_n) y $\gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))'$. En el caso particular en que $n = p$ los valores a_i ($i = 1, \dots, p$) proporcionan estimaciones para los valores de ϕ_p de un modelo $AR(p)$ ¹¹ y es justo la razón de que los estimadores de ϕ_i en (3.10) obtenidos con las ecuaciones Yule-Walker estén expresados como solución de este sistema de ecuaciones. De hecho, los coeficientes a_i se ven usualmente renombrados en la literatura como $a_i = \phi_{ni}$ ($i = 1, \dots, n$).

Aplicar esta idea para hacer estimaciones tiene el inconveniente de gastar muchos recursos computacionales, por lo que se desarrollaron procedimientos que permiten realizar los cálculos de manera recursiva como es el algoritmo *Durbin-Levinson*¹². El procedimiento de este algoritmo consiste en obtener \hat{X}_{n+1} basado en las primeras n observaciones y utilizar este resultado para estimar \hat{X}_{n+2} basado en las primeras $n+1$ observaciones, y así de manera sucesiva. A través del algoritmo Durbin Levinson, los predictores de un paso se obtienen según la expresión:

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}, \quad j = 1, \dots, n \quad (3.57)$$

para lo cual se obtienen, también de manera sucesiva, los valores ϕ_{nj} ($j = 1, \dots, n$). Sin embargo, (3.57) no es la única representación de \hat{X}_{n+1} que permite llevar a cabo el cálculo de las predicciones, Brockwell y Davis (2002) mencionan que a través de

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) \quad (3.58)$$

también es posible realizar los cálculos de predicción. En (3.58), a los términos $X_n - \hat{X}_n$ se les llama errores de predicción de un paso (o *innovaciones*), de ahí que el algoritmo que

¹¹Shumway & Stoffer, *op. cit.*, p. 113.

¹²Brockwell & Davis, 2002, *op. cit.*, p. 69.

permite estimar los valores de θ_{nj} , predicciones y error cuadrático medio se le llame algoritmo *innovaciones*¹³. Bajo esta representación, la estimación de h pasos \hat{X}_{n+h} es de la forma:

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), \quad (3.59)$$

su cálculo es posible ya que el algoritmo permite estimar no sólo \hat{X}_{n+1} sino también los valores θ_{nj} .

De manera análoga al algoritmo Durbin-Levinson, el algoritmo innovaciones permite estimar los parámetros $(\theta_1, \dots, \theta_q)$ de un modelo MA(q), ya que $\theta_{nj} \rightarrow \theta_j$ para $n \rightarrow \infty$ ¹⁴. En este sentido los estimadores obtenidos se consideran preliminares, pues se utilizan como valores iniciales, pues conforman la información de partida del método de máxima verosimilitud¹⁵.

El algoritmo innovaciones funciona para cualquier serie con segundo momento finito, más específicamente, para un modelo ARMA(p, q) las ecuaciones de predicción son de la forma:

$$\hat{X}_{n+h} = \sum_{i=1}^p P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}),$$

siempre que $n > m = \max(p, q)$ y $h \geq 1$ ¹⁶. En el caso de los modelos ARMA, los estimadores de ϕ_1, \dots, ϕ_p pueden obtenerse con el algoritmo bajo el supuesto de causalidad¹⁷. Para modelos ARIMA y SARIMA los mismos métodos pueden adaptarse para obtener ecuaciones de predicción, estas pueden consultarse en Brockwell y Davis, 2002 páginas 198-210.

Por último debe hacerse notar que las predicciones realizadas se hicieron tomando en cuenta la serie afectada por la transformación de Box-Cox, el efecto calendario y la inflación. Estas transformaciones deben realizarse de manera inversa para que la serie y las predicciones puedan visualizarse en escala de pesos, y cumplir con el fin principal del análisis.

¹³Las ecuaciones del algoritmo innovaciones pueden consultarse en Shumway y Stoffer, *Time series analysis and its applications*, p. 114.

¹⁴Brockwell & Davis, 2002, *op. cit.*, p. 101.

¹⁵Los detalles sobre la estimación por máxima verosimilitud pueden consultarse en Brockwell y Davis, 2002, *op. cit.*, pags. 158-161.

¹⁶Brockwell y Davis, 2002, *op. cit.*, p. 104.

¹⁷Véase Brockwell y Davis, 2002, *op. cit.*, pags. 154-155.

3.8. Diagnóstico del modelo

El siguiente paso en el ajuste de un modelo cualquiera de los tratados en este capítulo es el diagnóstico del modelo. El objetivo en este punto es revisar que los supuestos que se hicieron al respecto de los residuales se hayan cumplido, o bien saber en qué grado se cumplen, como son el comportamiento normal y la no correlación entre ellos. Al realizar el diagnóstico del modelo, los residuales se definen como:

$$\hat{W}_t = (X_t - \hat{X}_t) / \sqrt{r_{t-1}}, \quad \text{para } t = 1, \dots, n$$

donde \hat{X}_t es el mejor predictor lineal de X_t basado en las observaciones al tiempo $t - 1$, $r_{t-1} = E(X_t - \hat{X}_t)^2 / \sigma^2$ y σ^2 es la varianza del ruido blanco del modelo ajustado. Para muestras grandes \hat{W}_t tienen propiedades similares al proceso de ruido blanco ε_t . Lo más común es trabajar con los residuales rescalados que se definen como:

$$\hat{R}_t = \hat{W}_t / \hat{\sigma}, \quad t = 1, \dots, n \quad (3.60)$$

donde $\hat{\sigma} = \sqrt{(\sum_{t=1}^n W_t^2) / n}$. Para revisar la normalidad en los residuales rescalados se realizan comúnmente alguno de los siguientes análisis:

- Un histograma de los residuales debe parecerse al histograma de una distribución normal estándar.
- La gráfica conocida como Q-Q plot muestra en el eje vertical, los cuantiles del conjunto de datos, mientras que en el eje horizontal muestra los cuantiles teóricos de la distribución normal (0,1) dado que se trabaja con residuos estandarizados. Es decir, cada punto en la gráfica representa el cuantil teórico contra el cuantil empírico, por lo que se esperaría que ambos valores sean muy parecidos. De esta forma mientras más se asemejan los puntos a la línea recta de la identidad más se acercan los datos empíricos a una distribución normal.
- La prueba de normalidad de Anderson-Darling prueba si los datos ordenados $\{Y_1 < \dots < Y_N\}$ provienen de una distribución normal, el estadístico de prueba es:

$$A^2 = -N - \sum_{k=1}^N \frac{2k-1}{N} [\ln F(Y_k) + \ln(1 - F(Y_{N+1-k}))]. \quad (3.61)$$

Con N el número de datos y $F()$ la función de distribución normal estándar. En este caso la hipótesis nula es H_0 : La muestra aleatoria proviene de una distribución normal y se rechaza con un nivel de significancia $\alpha = 0.05$ si el estadístico A^2 es mayor que 0.752, o bien un *valor p* menor a 0.05. El test asume que los errores no están correlacionados, por lo que es conveniente utilizar alguna de las pruebas siguientes para corroborar este supuesto.

Al respecto de la aleatoriedad y autocorrelación algunas pruebas adecuadas son:

- La prueba de corrida de rachas debe mostrar que los residuales son aleatorios. Las hipótesis a contrastar en esta prueba son H_0 : Los residuales son aleatorios contra H_a : Los residuales no son aleatorios. Por lo que, con un valor de significancia $\alpha = 0.05$, buscamos que el *valor p* de la prueba sea más grande que α para aceptar la hipótesis nula.
- Las autocorrelaciones muestrales $\hat{\rho}_e(h)$ deben ser estadísticamente cero para todos los valores de h .
- El estadístico de Ljung-Box-Pierce, mostrado en (3.62), debe mostrar que el efecto conjunto de las autocorrelaciones, hasta la diferencia H , no sea significativo.

$$Q = n(n + 2) \sum_{h=1}^H \frac{\hat{\rho}_e^2(h)}{n - h} \quad (3.62)$$

donde $\hat{\rho}_e$ es el coeficiente de autocorrelación de los residuos y n el número de observaciones. Considerando las primeras H autocorrelaciones, las hipótesis a contrastar en esta prueba son H_0 : No hay autocorrelación contra H_a : Por lo menos una autocorrelación es diferente de cero. Bajo H_0 el estadístico tiene una distribución ji-cuadrada con $H - p - q$ grados de libertad (por lo que la prueba puede realizarse para H mayor a $p + q$). Para un valor de significancia $\alpha = 0.05$, buscamos que el *valor p* de la prueba sea más grande que α y así aceptar la hipótesis nula.

3.9. Aplicación

Continuando con el ejemplo, sobre la serie de ventas se hará el análisis a través de la metodología de Box-Jenkins. Los datos son los mismos que se presentaron en el cuadro 2.1, también se retomaron los cambios hechos a la base en el ejemplo anterior pues también deben considerarse los datos sin los efectos de la inflación y calendario. A continuación se presenta la selección del modelo siguiendo los pasos sugeridos en la literatura.

Paso 1: Graficar la serie y escoger las transformaciones apropiadas

La gráfica de esta serie se encuentra en la figura 2.2. Las primeras transformaciones que deben aplicarse son las de reducción de varianza antes que las de diferenciación. En este caso se aplicó el logaritmo a la serie, pues de las transformaciones Box-Cox, con distintos valores de $0 \leq \lambda \leq 1.5$, $\lambda = 0$ (el logaritmo) fue la que mostró mayor reducción de varianza. A esta serie se le llamará Y_t , puede verse en la figura 3.3.

En ella también se observa una línea que marca el promedio general de la serie, y guiándonos inicialmente por el hecho de que no hay claras desviaciones de esa línea, y que la varianza pareciera constante a través del tiempo, se trabajará bajo la hipótesis de que no se necesitan diferencias para que la serie sea estacionaria.

Paso 2: Determinar si es necesario aplicar algún grado de diferenciación

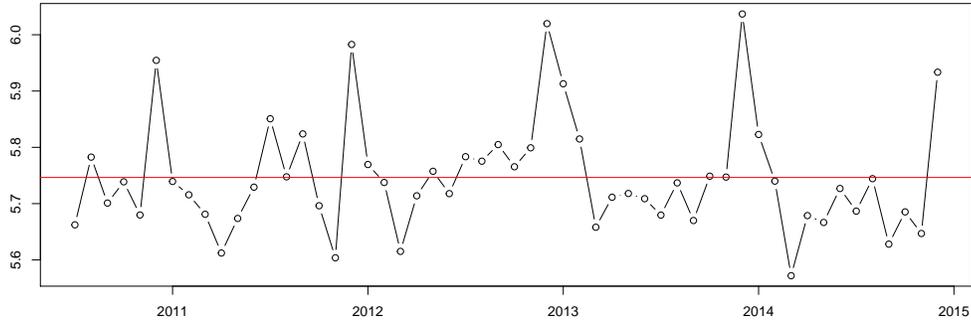


Figura 3.3: Serie de datos una vez aplicada la transformación Box-Cox $Y_t = \ln(X_t)$

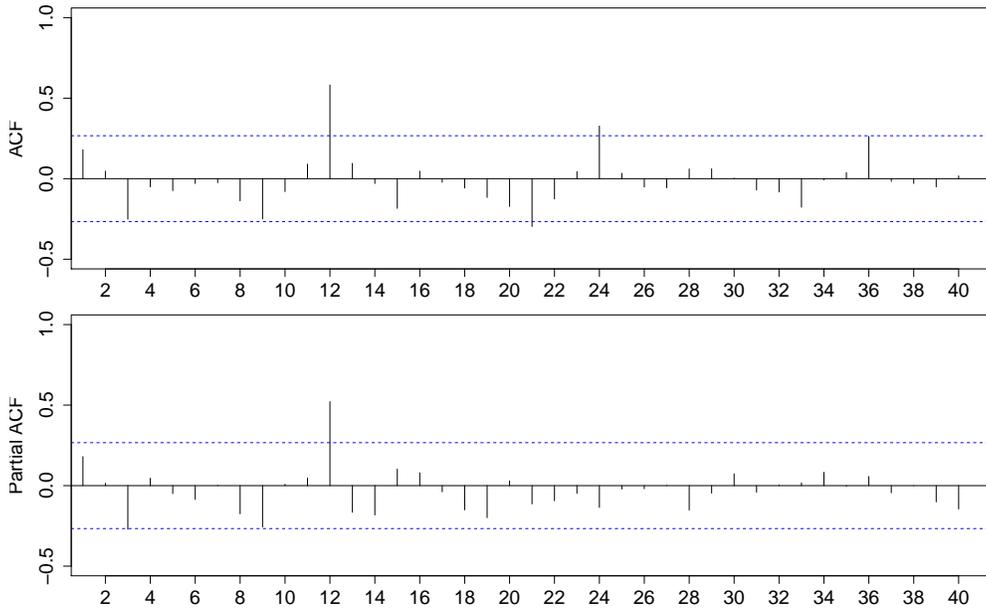


Figura 3.4: Funciones ACF y PACF de la serie Y_t

La figura 3.4 muestra las funciones ACF y PACF de la serie Y_t . En series donde hay una clara tendencia observaríamos en la ACF que en todos los retrasos los valores serían significativos y decrecerían muy poco con el tiempo, pero en esta serie ocurre que los valores de ACF son prácticamente cero excepto por los retrasos estacionales. Probablemente esto nos indica que no se necesitan diferencias para que el modelo sea estacionario, o bien que no hay raíces unitarias.

Para corroborar puede ser ayuda la prueba DFA como se vio en la sección 3.6, esta prueba determina si existen raíces unitarias en un modelo AR(p), en general los programas calculan el estadístico de prueba para un valor de p asignado por default.

```
> library(tseries)
> adf.test(serie3)
```

Augmented Dickey-Fuller Test

```
data: serie3
Dickey-Fuller = -3.6689, Lag order = 3, p-value = 0.03562
alternative hypothesis: stationary
```

El test DFA aplicado a la serie Y_t muestra un valor p de 0.03562, es decir, si se rechaza que el proceso es no estacionario (hipótesis nula) la probabilidad de equivocarse es del 3.562 %, por lo que el proceso puede asumirse como estacionario. Sin embargo, debe tomarse con ciertas precauciones ya que no se sabe si los errores ε_t se comportan como ruido blanco, de cualquier modo se aceptará el resultado pues concuerda con la inspección visual de la ACF. Como se mencionó anteriormente cuando se consiguen media y varianza constantes no es necesario aplicar una diferenciación de orden D . Para corroborar se aplicó una diferencia estacional, con lo que se obtuvo un proceso cuya media y varianza no eran constantes, y cuyos pronósticos mostraban intervalos de confianza muy amplios. Esto a su vez, está relacionado con el hecho de que el comportamiento estacional es muy estable con el tiempo, como se vio en el capítulo anterior por lo que no es probable que exista alguna raíz unitaria estacional que deba removerse. De lo anterior se concluye que $d = 0$ y $D = 0$ es decir, se usará un modelo SARMA.

Pudiera parecer contradictorio que en el capítulo anterior, se observó algo parecido a una tendencia cuadrática (en cuyo caso $d = 2$), pues en el suavizamiento por medias móviles centrado se observó una tendencia a la alza hasta 2012 y luego decrecimiento desde 2013 (véase la figura 2.4), sin embargo, es muy posible que ese movimiento no haya reflejado la tendencia, pudo deberse al componente cíclico y a los eventos ocurridos en el negocio. Además en ejercicios adicionales se vio que los modelos SARMA tuvieron mejor ajuste y predicciones más estables que modelos SARIMA con $d = 2$ en donde las predicciones no reflejaban correctamente el comportamiento estacional de la serie el cual es el movimiento más importante en la serie.

Paso 3: Identificar los parámetros p , q , P y Q

La elección de los parámetros es probablemente una de las tareas más difíciles en la modelación, para empezar uno puede apoyarse de las gráficas de ACF y PACF en la figura 3.4, de donde se obtiene la siguiente información: sobre los parámetros estacionales lo más seguro es que $P = 1$ pues solo en el retraso 12 la correlación parcial es estadísticamente significativa, por otro lado Q podría tomar los valores uno o dos, pues en la diferencia 36 ya no parece significativa; en cuanto a los parámetros no estacionales se observa que ambos p y q podrían tomar valores entre cero y tres. Para este caso no se utilizó la función EACF pues Wei (2006) no recomienda su uso en modelos estacionales, ya que adquiere patrones complicados, y sólo proporciona información sobre p y q ¹⁸. Esta función puede encontrarse en la librería TSA de R.

La naturaleza cambiante de estos datos dificultó aún más la detección del proceso, por lo que se aprovecharon las herramientas de programación que brinda R, utilizando para la estimación de los parámetros la función `Arima` de la librería `forecast`, y un bucle que permitió extraer en

¹⁸Wei, W., *op. cit.*, p. 169.

menor tiempo los resultados del estadístico AICC, pruebas de normalidad y rachas de varios modelos a la vez, con el propósito de obtener un panorama general del comportamiento de un parámetro específico.

```
valoresaicc<-numeric(length(16))
normal<-numeric(length(16))
aleat<-numeric(length(16))
i=1
for (p in 0:3){
  for (q in 0:3) {modelo<- Arima(serie3,order=c(p,0,q),seasonal =
    list(order=c(1,0,1),period=12),include.mean=TRUE)
    valoresaicc[i]<-modelo$aicc
    z<-modelo$residuals
    x<-ad.test(z)
    normal[i]<-x$p.value
    y<-runs.test(as.factor(z>median(z)))
    aleat[i]<-y$p.value
    i<-i+1}
}
valoresaicc
normal
aleat
```

Con el bucle se corrieron varios modelos cambiando en cada uno p y q , que por tener mayor cantidad de valores posibles, tienen más combinaciones, mientras que P y Q se cambiaron manualmente. A partir de las corridas se determinaron los parámetros de la siguiente forma:

- Sobre el parámetro P no hay duda que su valor es 1, pues cualquier cambio provocó menores ajustes (mayores valores de AICC) y residuales no distribuidos normalmente.
- Al respecto de Q también resultó claro que su valor deber ser uno, caso en el que se obtuvieron mejores ajustes y los residuales resultaron en su mayoría normalmente distribuidos, lo cual no ocurre con $Q = 2$ ni $Q = 0$.
- Una vez detectados P y Q , la elección de p y q se hizo con base en las estimaciones de los modelos $SARIMA(p, 0, q) \times (1, 0, 1)$ los valores de AICC de estos modelos pueden verse en el cuadro 3.4. En este caso los valores de AICC son negativos, pero el criterio sigue siendo el mismo: elegir el modelo con menor AICC, pues su valor tiene una escala arbitraria. Se deduce entonces que los mejores modelos en ajuste son el $SARIMA(0, 0, 2) \times (1, 0, 1)_{12}$ y el $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$ pues al tener los menores valores de AICC son lo que proporcionan mayor información con el menor número de parámetros.
- El modelo con menor AICC es el $SARIMA(0, 0, 2) \times (1, 0, 1)_{12}$, con ello es el que mejor explica los datos con el mínimo número de parámetros, sin embargo, presenta un problema, los parámetros θ_1 y Θ_1 no resultan significativos, esto es no son mayores que 1.96 veces su error típico, por lo que podrían ser cero. Del mismo modo ocurre en el modelo $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$ (el segundo con mejor ajuste) con los parámetros

p	q	$AICC$	p	q	$AICC$
0	0	-119.98	2	0	-122.94
0	1	-120.50	2	1	-121.26
0	2	-123.13	2	2	-118.57
0	3	-120.55	2	3	-115.74
1	0	-122.20	3	0	-121.22
1	1	-121.24	3	1	-118.56
1	2	-120.57	3	2	-115.75
1	3	-118.00	3	3	-114.58

Cuadro 3.4: Valores de AICC para distintos modelos $SARIMA(p, 0, q) \times (1, 0, 1)_{12}$

ϕ_1 y Θ_1 . Dado que en segundo punto se estableció que el mejor ajuste se obtuvo con $Q = 1$, el problema no está en Θ_1 , es más probable que $\theta_1 = 0$ o $\phi_1 = 0$. En R es posible fijar un parámetro en cero, en la función `Arima` con el comando `fixed`.

- Finalmente se encontró que el modelo con mejor ajuste fue el $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$ fijando en cero el parámetro ϕ_1 , en los resultados se observa que, gracias a este cambio, todos los parámetros son significativos.

```
> library(forecast)
> modelo2<-Arima(serie3,order=c(2,0,0), seasonal = list(order=c(1,0,1),
+ period=12),include.mean=TRUE, fixed=c(0,NA,NA,NA,NA),
+ transform.pars=FALSE)
> modelo2
```

```
Series: serie3
ARIMA(2,0,0)(1,0,1)[12] with non-zero mean
```

```
Coefficients:
      ar1      ar2      sar1      sma1  intercept
      0  0.3257  0.9906  -0.7895      5.7413
s.e.    0  0.1366  0.0244   0.2607      0.0389
```

```
sigma^2 estimated as 0.003266:  log likelihood=67.03
AIC=-124.07  AICc=-122.82  BIC=-114.12
```

Paso 4: Verificación de supuestos

Una vez verificado que todos los parámetros son significativos, en este paso se pretende probar que los errores se comportan estadísticamente como un ruido blanco, pues de ello depende la veracidad de las inferencias hechas sobre los parámetros.

La figura 3.5 muestra que la distribución de los residuales es bastante parecida a una distribución normal, y en la gráfica Q-Q plot los puntos se acercan bastante bien a la línea de identidad, sólo con ligeras desviaciones en la parte derecha de la cola, aún así los cuantiles

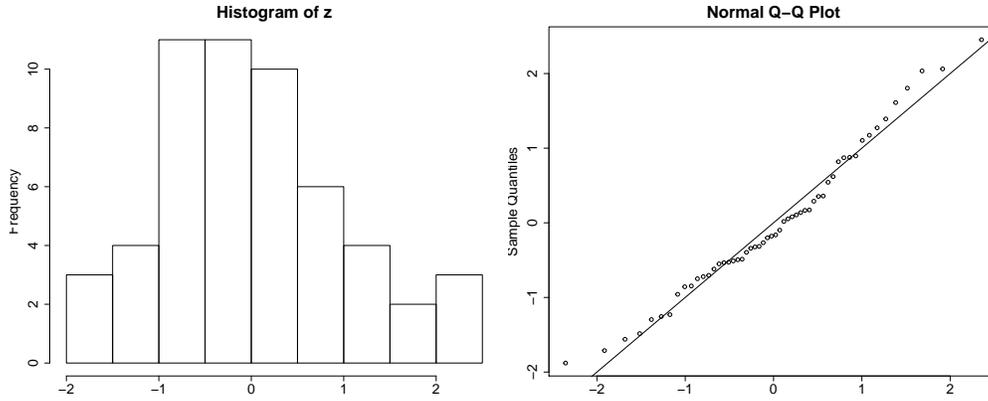


Figura 3.5: Histograma y Q-Q plot de los residuales estandarizados del modelo $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$ con parámetro $\phi_1 = 0$

empíricos se acercan bastante a los cuantiles teóricos. También el comportamiento se verifica al aplicar la prueba de Anderson-Darling,

```
> library(nortest)
> ad.test(z)
```

Anderson-Darling normality test

```
data: z
A = 0.47427, p-value = 0.232
```

en la que se obtiene un *valor p* mucho mayor a 0.05 con lo que no se rechaza la hipótesis nula de normalidad. Por otro lado para corroborar que los residuales no estén autocorrelacionados

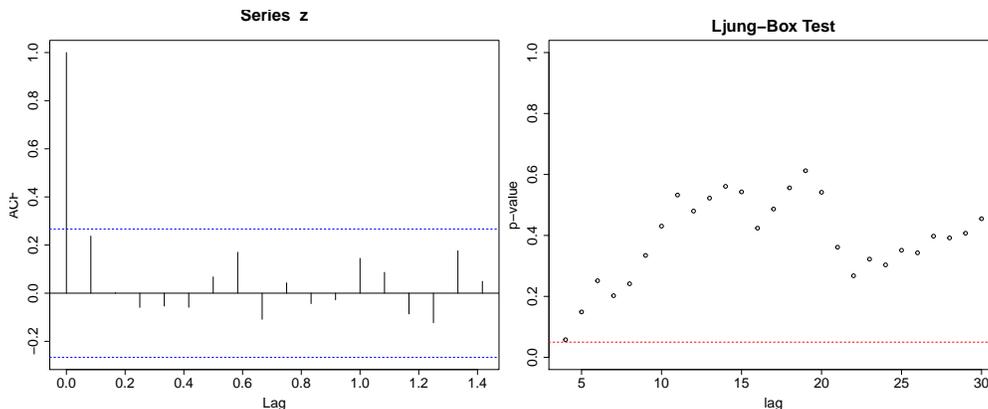


Figura 3.6: ACF y prueba Ljung-Box-Pierce (para $H > 3$) de los residuales estandarizados del modelo $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$ con parámetro $\phi_1 = 0$.

puede verse a partir de la gráfica ACF (Figura 3.6, lado izquierdo) que no hay autocorrelaciones muestrales significativas en ningún retraso. Al respecto del test de Ljung-Box-Pierce, no hay un valor H específico para el cual probar que la suma de las autocorrelaciones es cero,

por ello es común probarlo para varios valores de H y obtener una gráfica con los *valores* p de cada prueba, como la gráfica del lado derecho en la Figura 3.6, en donde se utilizó un valor $H = 30$. El gráfico se generó con la siguiente instrucción `LBQPlot(z, lag.max = 30, StartLag = 4, k = 3, SquaredQ = TRUE)`, en este caso estamos estimando 3 parámetros por lo que la prueba puede iniciarse desde $H = 4$, y aunque en este punto la hipótesis nula se acepta por muy poco, en los siguientes valores de H se acepta que no hay autocorrelación.

Por último la prueba de rachas nos indica que los residuales son aleatorios, con todos estos argumentos puede concluirse que los residuales del modelo elegido sí se comportan como un ruido blanco y los parámetros del modelo no son cero.

```
> runs.test(as.factor(z>median(z)))

Runs Test

data:  as.factor(z > median(z))
Standard Normal = -1.3739, p-value = 0.1695
alternative hypothesis: two.sided
```

Paso 5: Predicción

Haciendo uso de la función `forecast`, en la figura 3.7 se aprecian las estimaciones de los primeros 12 valores futuros de la serie (línea azul) y sus intervalos al 80 % y 95 % de confianza (bandas gris oscuro y gris claro respectivamente). En ella se observa que la predicción sigue el efecto estacional promedio de la serie, caídas en los meses de enero a marzo y de septiembre a noviembre, y alzas de abril a agosto y diciembre. Es de resaltar el hecho de que los intervalos de confianza tengan una razonable distancia de la media pues se observó en otros modelos, algunos incluso con buen ajuste, que los intervalos crecían mucho con el tiempo, eso parecería indicar que los parámetros elegidos fueron adecuados para la serie.

A pesar de que la predicción se hizo a la serie transformada Y_t pueden obtenerse ya algunas

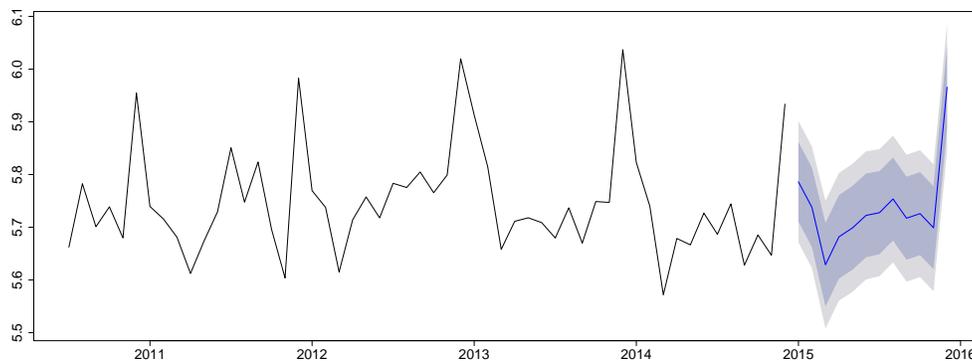


Figura 3.7: Predicción de la serie para el año 2015

conclusiones de la figura 3.7, como el hecho de que en todo el año 2015 se espera una venta

ligeramente mayor contra la de 2014, esto puede explicarse por el hecho de que se estimó un modelo sin diferenciación, por lo que se espera que el nivel de la serie sea siempre el mismo, y como en 2014 las ventas fueron en general bajas, en 2015 deberían subir para regresar a la media habitual.

Para tener la serie y sus predicciones en la escala original las transformaciones aplicadas antes de ajustar el modelo deben invertirse por lo que a la serie se aplicaron la función exponencial, el inverso de los pesos del efecto calendario y el inverso de los pesos por el efecto inflacionario. El efecto de la inflación, es susceptible de predecirse aplicando la metodología Box Jenkins a la serie de la inflación. Ya que la inflación no es la serie de interés en este trabajo, sólo se comentará que fue posible modelarla con un proceso SARIMA(0, 1, 2) × (1, 0, 1). Después de este procedimiento las ventas esperadas para 2015 pueden verse en la figura 3.8. Los cambios principales con respecto a la figura 3.7 es que se observa en la escala original, y que el efecto inflacionario muestra como las ventas han ido subiendo por el alza de precios, efecto ya tomado en cuenta en las predicciones de la serie. Básicamente este pronóstico tuvo el objetivo de informar a los dueños el comportamiento esperado de sus ventas para que revisaran su presupuesto de compras. No tendría mucho sentido describir a la serie a través de esta gráfica pues la tendencia no puede verse claramente por el efecto inflacionario, por ejemplo, en 2015 se esperan las ventas más altas que en cualquier año anterior pero, por lo visto en los análisis anteriores, esto no refleja una tendencia creciente sino el alza de precios.

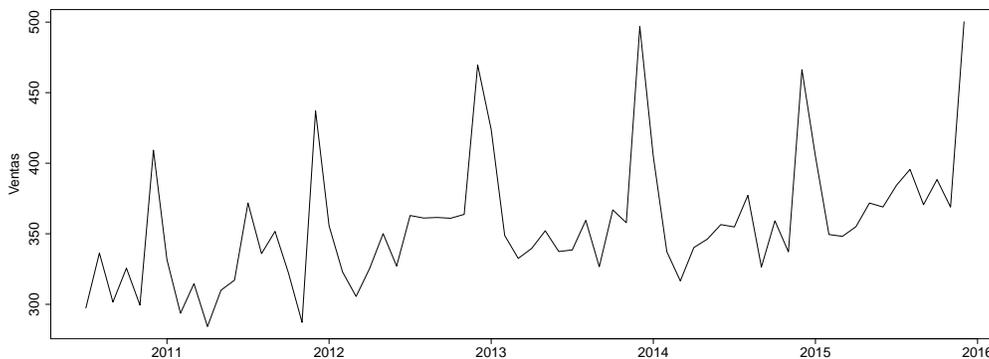


Figura 3.8: Predicción de la serie para el año 2015 expresado en la escala original (moneda nacional)

Debe tomarse en cuenta el uso de estos pronósticos asume que las condiciones serán las mismas entre periodos, sin embargo, este último punto puede no cumplirse del todo, ya que este establecimiento ha tenido varios cambios en periodos cortos de tiempo, como se mencionó a finales de 2012 cuando se registró el robo del vehículo usado para transportar mercancía, y entre 2012 y 2014 los hijos de la familia, encargados de resurtir el local, se iniciaron en otras actividades profesionales por lo que se redujo el tiempo trabajado en el negocio.

Capítulo 4

Modelos de regresión dinámicos (Regresión con errores ARMA)

4.1. Conceptos básicos de los modelos de regresión con errores ARMA

Este capítulo pretende conjuntar los conceptos vistos en los capítulos 1 y 3, pues los modelos dinámicos unen los conceptos de regresión lineal con los conceptos de series de tiempo según la metodología Box-Jenkins.

En un modelo de regresión $Y_i = \beta_1 + \beta_2 X_i + u_i$ se asume que los errores u_i son independientes e idénticamente distribuidos, pero cuando las variables dependiente y explicativas son recogidas a través del tiempo, es común es que estos supuestos no se cumplan. Hyndman y Athanasopoulos (2003) explican que las consecuencias de incumplir con estos supuestos son las siguientes:

- Los coeficientes de regresión estimados ya no son los mejores estimadores pues se ha omitido cierta información en su cálculo.
- Las pruebas asociadas al modelo (por ejemplo, las pruebas t para verificar la significancia de los coeficientes) no dan resultados confiables.

Por ello se considera que los errores pueden ser modelados de manera más apropiada como un proceso ARMA(p, q). El modelo de regresión con errores ARMA es el siguiente:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + w_t, \\ \phi(B)w_t &= \theta(B)e_t, \end{aligned} \tag{4.1}$$

en donde la variable dependiente y_t está en función de k variables explicativas $x_{i,t}$ (con $i = 1, \dots, k$), y un término de error, el error del modelo de regresión w_t que a su vez, al ser un proceso ARMA está afectado por un ruido blanco e_t . De hecho, también es posible considerar que los errores w_t pueden ser modelados con procesos ARIMA lo que se verá en la siguiente sección.

Para poder utilizar el modelo de regresión con errores ARMA Hyndman y Athanasopoulos (2012) recomiendan que tanto la variable dependiente como las explicativas sean estacionarias¹. Por tanto deben someterse a diferenciación las variables que lo necesiten para ser estacionarias, pero debido a que es deseable conservar la relación entre la variable dependiente y las explicativas, si se diferencia una variable deben diferenciarse todas. En este caso se le conoce como un *modelo en diferencias*, y es equivalente a utilizar un modelo de regresión con errores ARIMA². Cuando el modelo incluye variables que son todas estacionarias y no necesitan diferenciación se le conoce como *modelo en niveles*.

Este modelo es susceptible de utilizar variables explicativas con retraso, como en (4.2), en donde w_t es el proceso ARMA y x_{t-i} es la variable x_t retrasada i lugares con $i = 1, \dots, j$.

$$y_t = \beta_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \dots + \gamma_j x_{t-j} + w_t \quad (4.2)$$

El número de variables con retraso j puede seleccionarse utilizando el estadístico AICC junto con los valores p y q para el proceso ARMA. Es decir, se toman las j variables retrasadas y parámetros p y q para los cuales el valor de AICC sea el menor. Este modelo puede ser útil al pensar en series de tiempo económicas en donde variables como el gasto en publicidad, compra de mercancía u otras variables pueden no impactar de manera inmediata en las ventas. Por ejemplo, los gastos en publicidad (variable explicativa) pueden surtir efecto en la cantidad de clientes (variable dependiente) uno o varios meses después de haber sido efectuados.

4.2. Tendencia en modelos de regresión dinámicos

Según Brockwell y Davis (2002) las variables $x_{i,t}$ pueden utilizarse para modelar alguna tendencia en los datos, por ejemplo, usando $x_{1,t} = t$ y $x_{3,t} = t^2$ se puede ajustar una tendencia cuadrática. Al respecto Hyndman y Athanasopoulos (2003) explican con mayor detalle las tendencias lineales diferenciando las tendencias deterministas de las estocásticas, una tendencia determinista es obtenida cuando se usa un modelo de regresión de la forma:

$$y_t = \beta_0 + \beta_1 t + w_t, \quad (4.3)$$

donde w_t es un proceso ARMA. Mientras que una tendencia estocástica es obtenida cuando w_t en (4.3) es un proceso ARIMA con $d = 1$ (con ello, $w_t - w_{t-1}$ es un proceso ARMA), y si a (4.3) se le resta $y_{t-1} = \beta_0 + \beta_1(t-1) + w_{t-1}$ puede visualizarse la expresión como:

$$\begin{aligned} (y_t - y_{t-1}) &= \beta_1 + w'_t, & \text{o bien} \\ y_t &= y_{t-1} + \beta_1 + w'_t \end{aligned} \quad (4.4)$$

¹Aunque no lo hacen explícito, parece ser que Hyndman y Athanasopoulos se refieren a variables estocásticas en este punto, pues no aplican el mismo criterio cuando las variables explicativas son funciones del tiempo t .

²Hyndman, R.J. & Athanasopoulos G. 2012, *Forecasting: principles and practice*, Recuperado de: <https://www.otexts.org/fpp/9/1>.

donde $w'_t = w_t - w_{t-1}$ es un proceso ARMA. La primera igualdad en (4.4) permite observar que los modelos con errores ARIMA son modelos en diferencias con errores ARMA ya que se modela la serie diferenciada $Y_t - Y_{t-1}$ y w' es un proceso ARMA, mientras que la segunda igualdad permite interpretar el modelo como una caminata aleatoria donde el error no es un ruido blanco sino un proceso ARMA.

Ambos modelos de tendencia son tan similares que incluso producen estimaciones muy cercanas, la diferencia está en que el intervalo de predicción de la tendencia estocástica es más amplio pues los errores son no estacionarios, esto no es necesariamente una desventaja. Hyndman y Athanasopoulos explican que (4.3) tiene implícito el supuesto de que la tendencia no cambiará a través del tiempo, mientras que la tendencia de (4.4) se interpreta como la tendencia promedio de la serie en el periodo observado y no necesariamente la tasa de crecimiento que será observada en el futuro, por tanto es más seguro hacer predicciones tomando en cuenta una tendencia estocástica.

4.3. Estimación de los coeficientes de regresión

En el primer capítulo se vio como estimar los parámetros de un modelo de regresión por mínimos cuadrados ordinarios, sin embargo, este método presenta el problema de que en presencia de residuales correlacionados no proporciona estimadores de varianza mínima. Un mejor método de estimación de los parámetros $\beta = (\beta_0, \dots, \beta_k)$ es el método de mínimos cuadrados generalizados (*GLS*, por sus siglas en inglés), el cual permite agregar al modelo el supuesto de que los errores de la regresión están correlacionados. Para entenderlo se examinará el modelo de regresión lineal con k variables explicativas en forma matricial

$$\mathbf{Y} = X\beta + \mathbf{W},$$

donde $\mathbf{Y} = (Y_1, \dots, Y_n)$, X es la matriz con las variables explicativas en la que cada renglón $x'_t = (x_{t1}, \dots, x_{tk})$, $\beta = (\beta_1, \dots, \beta_k)$ y $\mathbf{W} = (W_1, \dots, W_n)$. En la estimación por mínimos cuadrados ordinarios se parte del supuesto de que los errores \mathbf{W} no están correlacionados, esto es $E(\mathbf{W}\mathbf{W}') = \sigma^2\mathbf{I}$, en cambio la estimación por GLS permite que los residuales estén correlacionados de forma que:

$$E(\mathbf{W}\mathbf{W}') = \Gamma_n,$$

donde $\Gamma_n \neq \mathbf{I}$ es la matriz de varianzas y covarianzas de \mathbf{W} . En el método de MCO se obtenían los estimadores de β al minimizar la suma de los residuales al cuadrado que en forma matricial se expresa como $(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta)$, mientras que la estimación por mínimos cuadrados generalizados se obtiene al minimizar la suma de cuadrados ponderados:

$$(\mathbf{Y} - X\beta)' \Gamma_n^{-1} (\mathbf{Y} - X\beta).$$

Los estimadores que se obtienen son de la forma

$$\hat{\beta}_{GLS} = (X' \Gamma_n^{-1} X)^{-1} X' \Gamma_n^{-1} \mathbf{Y}, \quad (4.5)$$

y la gran ventaja del método de mínimos cuadrados generalizados es que proporciona los mejores estimadores lineales insesgados, en el sentido de que las combinaciones lineales de β

son las de menor varianza, es decir,

$$\text{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_{GLS}) \leq \text{Var}(\mathbf{c}'\hat{\boldsymbol{\beta}})$$

por lo que la estimación por GLS se prefiere contra la estimación por OLS³.

También es posible estimar $\boldsymbol{\beta}$ por máxima verosimilitud⁴, en donde la función de verosimilitud es de la forma:

$$L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = (2\pi)^{-n/2}(\det\Gamma_n)^{-1/2}\exp\{-(1/2)(\mathbf{Y} - X\boldsymbol{\beta})'\Gamma_n^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\}, \quad (4.6)$$

a través de la cual se obtienen los parámetros $(\beta_1, \dots, \beta_k)$, (ϕ_1, \dots, ϕ_p) y $(\theta_1, \dots, \theta_q)$. En la práctica los parámetros no se obtienen de (4.6) con un solo cálculo algebraico sino que, como explican Brockwell y Davis, para minimizar la log-verosimilitud (el logaritmo de la ecuación (4.6)) se requieren los errores de predicción de un paso del proceso $W = Y - X\boldsymbol{\beta}$, y para obtener estos últimos se requieren estimaciones iniciales de $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ y $\boldsymbol{\theta}$ usando el algoritmo innovaciones, sin embargo no se detalla como incluir en el algoritmo los parámetros $\boldsymbol{\beta}$.

Debe notarse que el modelo tiene un problema tanto en la estimación de los parámetros como en la identificación del modelo ARMA de los residuales, y es que para obtenerlos deben conocerse los coeficientes de regresión, o bien debe conocerse el modelo de ARMA para obtener los coeficientes $\boldsymbol{\beta}$.

4.4. Identificación del modelo

Un primer método de identificación del modelo es una extensión del método propuesto por Cochrane y Orcutt (1949)⁵ el cual parte del supuesto que si se conocen los parámetros $\boldsymbol{\phi}$ y $\boldsymbol{\theta}$ de los errores w_t , no es necesario calcular numéricamente los parámetros $\boldsymbol{\beta}$, sino que pueden obtenerse algebraicamente por GLS como en (4.5), el procedimiento de este método es el siguiente:

- 1) Calcular $\hat{\boldsymbol{\beta}}_{OLS}$ y los residuales estimados $Y_t - X_t'\hat{\boldsymbol{\beta}}_{OLS}$, para $t = 1, \dots, n$.
- 2) Ajustar un modelo ARMA(p, q) por máxima verosimilitud a los residuales obtenidos.
- 3) Calcular algebraicamente el estimador $\hat{\boldsymbol{\beta}}_{GLS}$ haciendo uso del modelo ARMA ajustado a los residuales.
- 4) Calcular los residuales con los coeficientes GLS $Y_t - X_t'\hat{\boldsymbol{\beta}}_{GLS}$, $t = 1, \dots, n$, y regresar al paso (2), parar cuando los estimadores se hayan estabilizado.⁶

³Brockwell & Davis, 2002, *op. cit.*, p. 212.

⁴Brockwell & Davis, 2002, *op. cit.*, p. 213.

⁵Brockwell & Davis, 2002, *op. cit.*, p. 214.

⁶Brockwell & Davis, 2002, *op. cit.*, p. 214.

Otro método que explican Hyndman y Athanasopoulos, se basa en utilizar un modelo *proxy* para los errores ARMA, esto es, modelarlos con un proceso ARMA preestablecido o aproximado sin importar que no sea el que mejor ajuste a los errores. Aunque en primera instancia el modelo no sea el más adecuado, permite tomar en cuenta buena parte de la autocorrelación de los errores, de manera que los coeficientes β estimados se acerquen a los del modelo definitivo. Posteriormente los coeficientes β permiten encontrar un mejor modelo para los w_t con el cual volver a estimar el modelo. El procedimiento completo lo explican de esta forma:

- A) Verificar que la variable explicada y todas las predictivas sean estacionarias, en caso contrario aplicar diferenciación a todas las variables para preservar la interpretabilidad, hasta que se obtenga la estacionariedad.
- B) Ajustar la regresión con el modelo *proxy* para los residuales w_t , el modelo proxy que se sugiere utilizar es un AR(2) para series no estacionales, o bien un SARIMA(2, 0, 0) \times (1, 0, 0)_m para series estacionales, donde m es el periodo estacional.
- C) Calcular los errores w_t del modelo de regresión ajustado e identificar un modelo ARMA apropiado para los mismos.
- D) Volver a ajustar el modelo completo con el nuevo modelo ARMA para los errores.
- E) Verificar que los otros errores e_t se comporten como un proceso de ruido blanco.

Obtenido el modelo definitivo puede usarse para hacer predicciones de la variable dependiente, tomando en cuenta que para lograrlo también deben estimarse a futuro las variables explicativas y de los errores w_t con los algoritmos mencionados en el capítulo anterior. Cuando las variables X_t son funciones del tiempo t las predicciones se obtienen simplemente sustituyendo valores de t . Pero cuando son series de tiempo se tienen dos opciones, la primera es tratar y predecir cada serie por separado, y la segunda es asumir ciertos valores. Puede considerarse como una desventaja que las predicciones de la variable dependiente no toman en cuenta la incertidumbre de las predicciones de las variables explicativas.

4.5. Aplicación

En esta sección se aplicarán tres modelos a los datos, en el primero de ellos se analizará la tendencia de la serie determinista y estocástica, en el segundo se harán predicciones tomando como variable explicativa las compras del negocio, y en el tercero se probará si las compras de meses anteriores tienen algún efecto sobre las ventas.

En el primer caso el modelo a probar es $Y_t = \beta_0 + \beta_1 t + w_t$, donde Y_t son las ventas mensuales y w_t es un proceso ARMA, básicamente es el modelo probado en el primer capítulo en donde el incumplimiento de supuestos no permitió obtener conclusiones válidas, pero con una metodología que si permite hacer inferencia sobre los coeficientes β_0 y β_1 . El modelo fue ajustado con el comando `Arima` a través del parámetro `xreg` permite agregar una variable explicativa, en este caso la variable tiempo. En un primer momento se utilizó el modelo proxy propuesto en el inciso B) (es decir un SARIMA $(2, 0, 0) \times (1, 0, 0)_{12}$) y ajustar un modelo ARMA más adecuado a w_t , en este caso se optó por un SARIMA $(0, 0, 2) \times (1, 0, 1)_{12}$ los resultados fueron los siguientes:

```
> time_model1<-Arima(serie3,xreg=seq(1:54),order=c(0,0,2),
+ seasonal = list(order = c(1,0,1), period = 12),
+ fixed=c(0,NA,NA,NA,NA,NA),transform.pars=FALSE)
> time_model1
```

```
Series: serie3
ARIMA(0,0,2)(1,0,1)[12] with non-zero mean
```

Coefficients:

	ma1	ma2	sar1	sma1	intercept	seq(1:54)
	0	0.3143	0.9923	-0.8106	5.7489	-2e-04
s.e.	0	0.1326	0.0170	0.2002	0.0381	6e-04

```
sigma^2 estimated as 0.003238: log likelihood=67.01
AIC=-122.01 AICc=-120.23 BIC=-110.08
```

En este primer modelo no se hará un análisis a fondo ya que es muy parecido a la aplicación del capítulo anterior, véase que los coeficientes obtenidos son muy parecidos, si acaso cambiaron fue por el efecto de la variable tiempo que resultó tener un coeficiente β_1 resultó prácticamente cero y no significativo. Si se eliminara esta variable estaríamos exactamente en el caso del modelo utilizado en el capítulo tres.

Si bien se aplicó el modelo también modelando los errores con un proceso SARIMA con $d = 1$ (es decir, una tendencia estocástica), no es necesario mostrar los resultados pues del mismo modo la variable tiempo no fue significativa y su inclusión no hace sentido. Del hecho de que la variable tiempo no tenga un efecto significativo en las ventas puede concluirse que no existe en los datos una tendencia lineal creciente ni decreciente.

El modelo que principalmente pretende probarse en este capítulo es el que tiene una variable explicativa diferente del tiempo, para este caso además de las ventas mensuales (variable dependiente) se pudieron recabar también los datos de las compras de mercancía (variable explicativa). La figura 4.1 muestra ambas series deflactadas, sin efecto calendario y con la transformación Box-Cox aplicada. Recuérdese que deflactar se refiere a remover el efecto de la inflación, mientras que quitar el efecto calendario se refiere a ponderar la serie para que todas las observaciones reflejen ventas de 30 días por mes. Por tanto este segundo modelo a probar tiene la siguiente forma:

$$Y_t = \beta_0 + \beta_1 X_t + w_t, \quad (4.7)$$

en donde Y_t son las ventas, X_t las compras (ambas series con las transformaciones arriba descritas) y w_t un proceso ARMA (en este caso estacional).

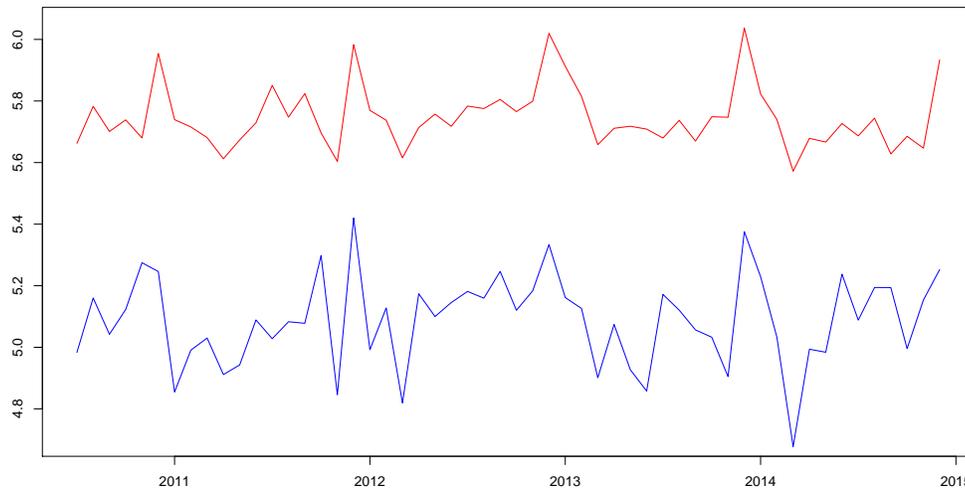


Figura 4.1: Serie de ventas (y_t) en rojo y serie de compras (x_t) en azul, ambas sin efecto calendario, deflactadas y con transformación Box-Cox.

Siguiendo el procedimiento de Hyndman y Athanasopoulos, el inciso A) habla de verificar la estacionariedad de Y_t y X_t , para la serie de ventas ya fue corroborado en el capítulo anterior, por lo que ahora necesitamos saber si la serie de compras es estacionaria. Gráficamente puede verse que no parece tener ninguna tendencia, pero el test Dickey-Fuller aumentado indica lo contrario:

```
> adf.test(compra3)
```

```
Augmented Dickey-Fuller Test
```

```
data: compra3
```

```
Dickey-Fuller = -3.2837, Lag order = 3, p-value = 0.08345
```

```
alternative hypothesis: stationary
```

el valor $p=0.083$ muestra que la serie tiene una raíz unitaria, pero recuérdese que entre las desventajas de esta prueba (y de las pruebas de raíces unitarias en general) está el encontrar raíces unitarias donde no las hay realmente. Es muy posible que este sea el caso, pues el valor p es inferior a 0.1 y la serie no presenta un comportamiento a la alza ni a la baja, además la función de ACF en la figura 4.2 muestra que las autocorrelaciones de la serie tienen valores pequeños desde los primeros retrasos, lo cual no ocurriría de tener alguna tendencia. Por lo anterior se considera a la serie de compras como estacionaria, y dado que ambas series lo son, se utilizará un modelo en niveles.

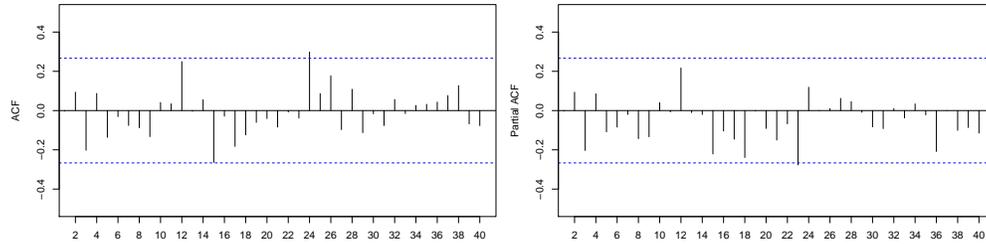


Figura 4.2: Funciones muestrales ACF (izquierda) y PACF (derecha) de serie de compras (x_t)

Al ajustar por primera vez el modelo dinámico, se modelaron los errores w_t con el proceso recomendado en el inciso B), el SARIMA(2, 0, 0) \times (1, 0, 0)₁₂ por ser una serie estacional. Con este ejercicio se obtuvieron los errores w_t , su comportamiento se muestra en la figura 4.3 donde se aprecian los errores y las funciones ACF y PACF.

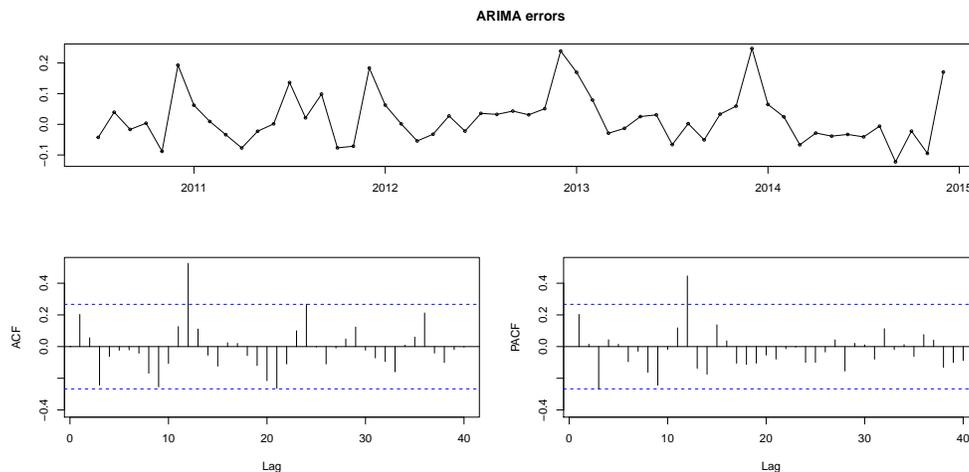


Figura 4.3: Gráficas de errores w_t , ACF y PACF obtenidos con el modelo *proxy*

A partir de las gráficas de la figura 4.3 se pretende identificar los valores de los parámetros P , Q , p y q que describan a la serie w_t más adecuadamente que el modelo *proxy*. Sobre P y Q debe ser 1 en ambos casos, pues sólo en el retraso 12 la correlación y la autocorrelación parcial, tienen un valor que sobresale de las bandas de confianza. Por otro lado las funciones ACF y PACF no tienen valores significativos antes del retraso 12, lo que indicaría que $p = 0$ y $q = 0$, sin embargo este modelo no cumplió ciertas pruebas sobre la no autocorrelación de los

residuales. Al usar el bucle mostrado en el capítulo anterior, en este caso con la modificación de insertar como variable predictiva las compras, fue posible obtener los valores de AICC sobre el ajuste de varios modelos. El modelo con mejor ajuste para w_t (tomando en cuenta que cumpliera los supuestos sobre los residuales e_t) fue el SARIMA(0, 0, 2) \times (1, 0, 1)₁₂. Se realizó nuevamente la estimación del modelo dinámico con el proceso SARIMA(0, 0, 2) \times (1, 0, 1)₁₂ para los errores w_t .

Al usar el modelo *proxy* para los errores w_t se eliminó parte de la autocorrelación de la regresión, y con ello los coeficientes β fueron lo suficientemente buenos para elegir un mejor modelo a los errores w_t . Esto a su vez permitió re-estimar adecuadamente los coeficientes β . Aunque suena simple, el proceso resulta un poco largo, por lo que a continuación se presenta el código en R con el que se llegó al modelo final.

```
#Modelo 1 (Ecuación 4.7)
#La serie w_t fue ajustada con el modelo proxy indicado en el inciso B)
#Los objetos serie3 y compra3 son las ventas y compras después de las
#transformaciones (inflación, calendario y Box-Cox)
#El modelo se guarda en un objeto tipo Arima al que se le llamó "proxy"

proxy<-Arima(serie3, xreg=compra3,order=c(2,0,0),
             seasonal = list(order =c(1, 0, 0), period = 12))
#Siguiendo con el inciso C), el comportamiento de la serie w_t antes
#inobservable, puede ahora observarse en la figura 4.3 con el comando:

tsdisplay(arima.errors(proxy), main="ARIMA errors",lag.max = 40)

#Las gráficas sugieren que los w_t pueden verse como un proceso
#SARIMA con P=1, Q=1, p=0 y q=0.

#Modelo 2
#Según el inciso D) se vuelve a ajustar el modelo completo
#con el nuevo proceso identificado

nvo<- Arima(serie3, xreg=compra3,order=c(0,0,0),
            seasonal = list(order = c(1,0,1), period = 12))

Series: venta3
ARIMA(0,0,0)(1,0,1)[12] with non-zero mean

Coefficients:
      sar1      sma1  intercept  compra3
      0.9576  -0.6978      4.4407   0.2562
s.e.   0.1369   0.4834      0.3574   0.0702

sigma^2 estimated as 0.003457:  log likelihood=70.86
```

```
AIC=-131.72  AICc=-130.47  BIC=-121.77
```

```
#Continuando con el punto E) se prueba el cumplimiento de  
#los residuales e_t
```

```
z<-nvo$residuals  
> ad.test(z)
```

```
Anderson-Darling normality test
```

```
data: z  
A = 0.6345, p-value = 0.09308
```

```
> runs.test(as.factor(z>median(z)))
```

```
Runs Test
```

```
data: as.factor(z > median(z))  
Standard Normal = -2.4729, p-value = 0.0134  
alternative hypothesis: two.sided
```

```
> Box.test(z, lag = 5, type = "Ljung-Box", fitdf = 2)
```

```
Box-Ljung test
```

```
data: z  
X-squared = 8.1213, df = 3, p-value = 0.04357
```

```
#No hubo problemas con la estimación  
#Los residuales cumplieron con la prueba de normalidad  
#Pero no cumplieron la prueba de rachas  
#La prueba Ljung-Box no se acepta para H=5 (ni para H<5)  
#H=6 es el primer valor donde se acepta la hipótesis nula  
#Al parecer los residuales e_t aún conservan correlación  
#Deben probarse otros modelos
```

```
#Modelo 3 (Definitivo)
```

```
#Después de correr otros modelos se determinó que P=1,Q=1, p=0 y q=2
```

```
def<-Arima(serie3, xreg=compra3,order=c(0,0,2),  
          seasonal = list(order = c(1, 0, 1), period = 12))
```

Se deja hasta este punto el programa en R cuyo objetivo fue dar a conocer los principales

pasos de cómo se obtuvo el modelo definitivo. Los parámetros de este modelo, así como las pruebas de cumplimiento de supuestos de los residuales e_t , se presentan a continuación. La estimación produjo los siguientes resultados:

```
> def<-Arima(serie3, xreg=compra3,order=c(0,0,2),
+           seasonal = list(order = c(1, 0, 1), period = 12))
> def
```

```
Series: serie3
ARIMA(0,0,2)(1,0,1)[12] with non-zero mean
```

```
Coefficients:
          ma1      ma2      sar1      sma1  intercept  compra3
      0.2906  0.4130  0.9612 -0.6470     4.6279   0.2188
s.e.  0.1269  0.1377  0.0854  0.3722     0.2737   0.0534
```

```
sigma^2 estimated as 0.00272:  log likelihood=75.72
AIC=-137.43  AICc=-135  BIC=-123.51
```

Puede verse que todos los coeficientes, a excepción de uno, son significativos pues rebasan 1.96 veces su desviación estándar. El parámetro media móvil estacional Θ_1 (`sma1`) es el que resulta no significativo al 95 % de confianza, sin embargo, su estadístico $t=(-0.6470/0.3722)=-1.738$ indica que el valor absoluto de Θ_1 es 1.738 veces más grande que su desviación estándar. Si se compara contra el cuantil de la normal al 90 % de confianza (1.64 para la prueba de dos colas), vemos que es todavía significativo a un 90 % de confianza, por tanto seguirá considerándose significativo. Además otras corridas mostraron que al fijarlo en cero, las predicciones no seguían adecuadamente el patrón estacional.

Retomando el inciso E) para el modelo definitivo, toca el turno de verificar los supuestos hechos sobre los errores e_t . Empezando por el supuesto de normalidad, la prueba Anderson-Darling muestra, con un *valor p* de 0.5369, que se acepta la hipótesis de normalidad. Además de que los la figura 4.4 muestra una distribución de los e_t normal en el histograma y en el Q-Q plot.

```
> ad.test(def$residuals)

Anderson-Darling normality test

data:  def$residuals
A = 0.31347, p-value = 0.5369
```

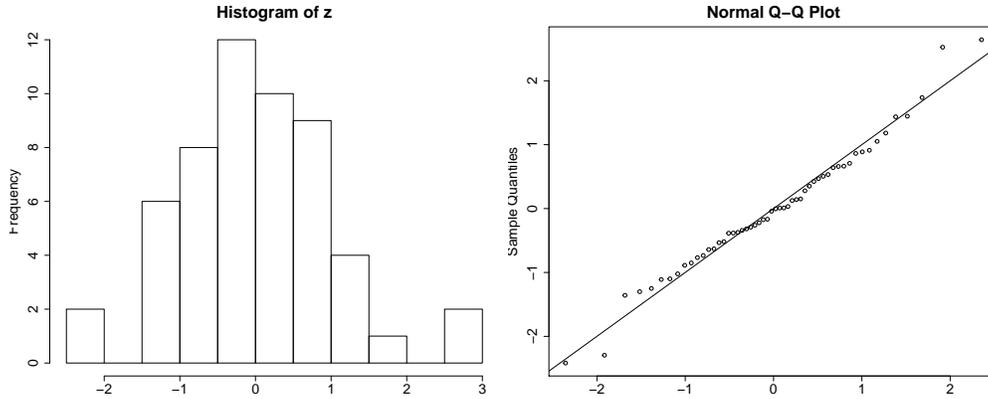


Figura 4.4: Histograma y Q-Q plot de los residuales e_t estandarizados del modelo definitivo

Sobre el supuesto de no autocorrelación, en la siguiente prueba de rachas se acepta la hipótesis de aleatoriedad con un valor p de 0.78.

```
> runs.test(as.factor(def$residuals>median(def$residuals)))
```

Runs Test

```
data: as.factor(def$residuals > median(def$residuals))
Standard Normal = -0.27477, p-value = 0.7835
alternative hypothesis: two.sided
```

La figura 4.5 muestra la función ACF cuyos valores son todos no significativos, y los *valores* p de las pruebas de Ljung-Box, todos mayores a 0.05, indican que los errores e_t no están correlacionados, por lo que podemos asumir que los errores se comportan como un ruido blanco.

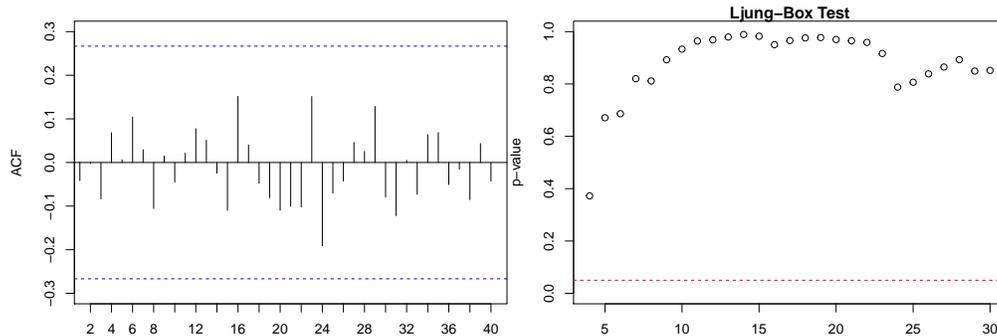


Figura 4.5: Gráficas de ACF y Ljung-Box Test de los errores e_t del modelo dinámico definitivo

El modelo ya puede usarse para hacer predicciones de la serie de ventas, para lo cual, es necesario aplicar por separado los pasos de la metodología Box-Jenkins y obtener predicciones de la serie de compras. Vale la pena mencionar algunos aspectos sobre la identificación de este modelo, pues si bien se siguió la metodología aplicada en el capítulo anterior, esta serie

tuvo algunos aspectos particulares.

Nótese que a pesar de que el comportamiento de la serie de compras sea un poco errático, aún conserva un componente estacional en el que las compras son bajas los primeros meses del año y altas en los últimos. Esta combinación de componente errático y estacional trae consigo un problema: varios modelos, estacionales o no, una vez ajustados parecen cumplir los supuestos de los residuales. Ahora bien, no sólo se busca que el modelo a aplicar cumpla los supuestos para los residuales, sino también que no tenga problemas de estimación e incluso que las predicciones reproduzcan el comportamiento estacional de la serie.

Si se observa la figura 4.2, parece tener sentido que p , q , P y Q sean cero, o bien cabe la posibilidad de que $Q = 2$, pues pareciera que la única autocorrelación significativa está en el retraso 24 de la ACF. Sin embargo, teniendo en cuenta lo descrito en el párrafo anterior, no parece compatible que la serie provenga de un proceso de ruido blanco. Otros modelos fueron probados y se encontró que un modelo con $P = 1$ y $Q = 1$ es un posible candidato. De los tres modelos a probar se encontró lo siguiente:

- El modelo de ruido blanco cumplió los supuestos para los residuales (normalidad y no autocorrelación), pero las proyecciones en este modelo son simplemente la media de las observaciones, con lo que no se reprodujo el comportamiento estacional de la serie. Tuvo un AICC de -49.71.
- El modelo SARIMA(0, 0, 0) \times (0, 0, 2)₁₂ ($P = 0$, $Q = 2$) cumplió los supuestos para los residuales, sin embargo el parámetro Θ_2 no resultó significativo, su error de estimación resultó más grande que el propio parámetro. Por otro lado, las predicciones no reflejaron el comportamiento estacional de la serie. Tuvo un AICc de -57.99.
- El modelo SARIMA(0, 0, 0) \times (1, 0, 1)₁₂ ($P = 1$, $Q = 1$) cumplió los supuestos para los residuales, en cuando estimación el parámetro Θ_1 no es significativo a un 95% de confianza, pero su desviación es menor que el parámetro, por último las predicciones sí reflejaron el comportamiento estacional de la serie. Tuvo un AICC de -53.73.

A pesar de no tener el menor AICC, el SARIMA(0, 0, 0) \times (1, 0, 1)₁₂ tuvo un mejor comportamiento en cuanto estimación y predicción de la serie, por lo que se ajustó a la serie de compras. Calculadas las predicciones de las compras para 2015, fue posible estimar las ventas del año 2015, ahora con el modelo dinámico definitivo, usando la función `forecast` de la librería `forecast`. Las estimaciones de ambas series se observan en la figura 4.6, en donde se observa que el intervalo de confianza de las compras es más grande que el de las ventas, de hecho el intervalo de confianza de las ventas es más estrecho comparado con el del modelo SARIMA en el capítulo anterior, es decir, considerar las compras redujo la variación de las estimaciones.

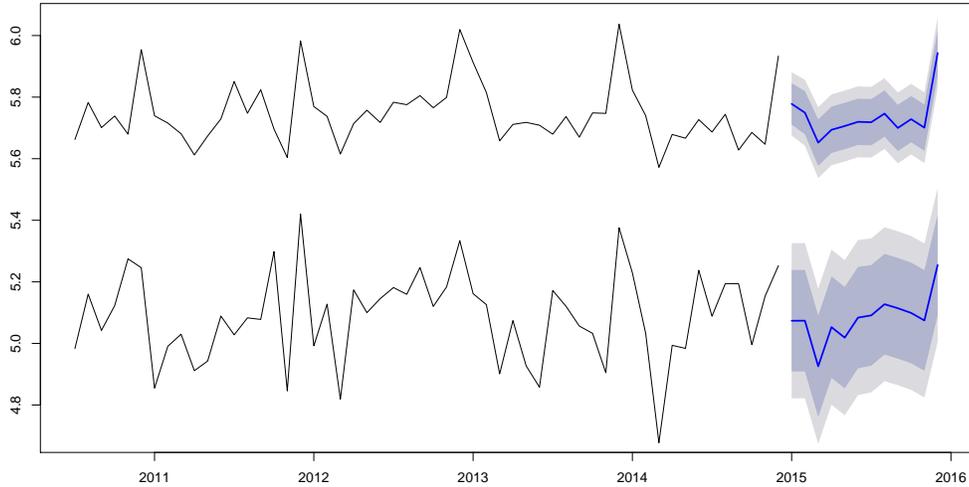


Figura 4.6: Predicción e intervalos de confianza de la serie de ventas y_t (arriba) y de la serie de compras x_t (abajo)

En la figura 4.7 se visualiza la predicción para 2015 de ambas variables en la escala original. Un hecho que vale la pena resaltar es que la variable compras tiene un patrón no tan bien definido como el de la variable ventas, incluso el modelo con mejor ajuste muestra que fuera del comportamiento estacional no hay mucha interacción entre los meses del mismo año. De tener un patrón menos errático probablemente se pudiera reducir aun más la varianza de las predicciones, es un punto que los dueños podrían tomar en cuenta para su planeación de compras.

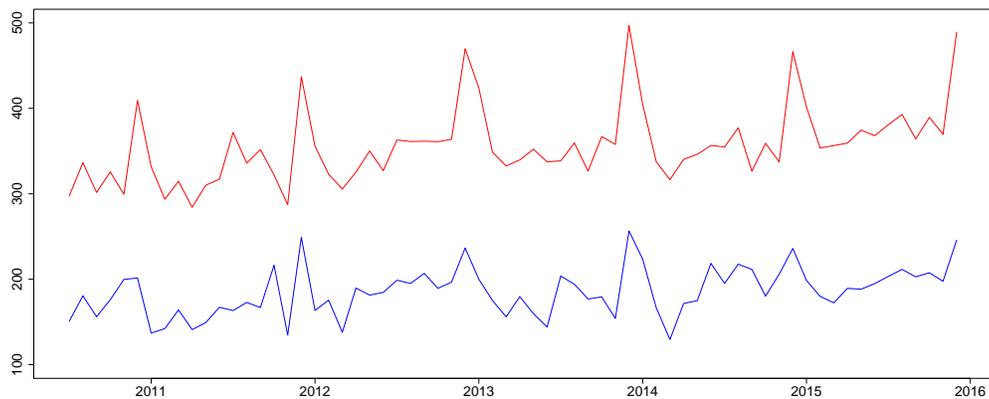


Figura 4.7: Predicción de la serie para el año 2015 expresado en la escala original (moneda nacional)

Como último análisis de este capítulo se verá la posibilidad de ajustar el modelo dinámico con la variable compras inmediatas y compras retrasadas. No hay una regla de cuántos predictores retrasados elegir inicialmente, y según Hyndman y Athanasopoulos el mejor modelo es el que presente menor AICC, sin embargo, se hizo un ejercicio considerando la variable X_t hasta con cuatro retrasos y se encontró que el modelo con menor AICC es el que únicamente tiene la variable control X_t . No tendría sentido repetir el análisis anterior, además de que en cierto sentido es cuestionable utilizar solamente el estadístico AICC para determinar el modelo a elegir sin considerar si se cumplen los supuestos para los residuales.

Por otro lado, las correlaciones de las compras retrasadas contra las ventas en general no son significativas, y dado que el negocio no tiene ninguna bodega tal vez no tenga sentido preguntarse por el impacto de las compras de dos, tres o cuatro meses anteriores. De hecho, algunas empresas que se dedican a la compraventa de mercancía, tienen políticas sobre la cantidad de mercancía a comprar, la cual debe ser suficiente para que se termine en no más de un mes. Teniendo en cuenta la información anterior, hace más sentido preguntarse si las ventas se ven afectadas por las compras hasta con un mes de retraso. Por tanto se ajustó un modelo como el siguiente:

$$y_t = \beta_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + w_t, \quad (4.8)$$

en donde las ventas y_t son explicadas por las compras del mismo mes x_t y las compras con un mes de retraso x_{t-1} . Se encontró que el modelo en el que ajustaron mejor los errores w_t , según las gráficas de ACF, PACF y valores de AICC fue el SARIMA(0, 0, 2) \times (1, 0, 1). También se cumplieron los supuestos de los errores e_t a una variable normal, y se obtuvieron los siguientes coeficientes:

Call:

```
arima(x = serie3[2:54], order = c(0, 0, 2), seasonal = list(order = c(1, 0,
1), period = 12), xreg = mercancia[2:54, 1:2])
```

Coefficients:

	ma1	ma2	sar1	sma1	intercept	comp_lag0	comp_lag1
	0.2895	0.4053	0.8897	-0.3823	4.7110	0.2095	-0.0073
s.e.	0.1252	0.1409	0.1466	0.4059	0.4284	0.0539	0.0522

sigma^2 estimated as 0.002787: log likelihood = 75.12, aic = -134.24

con lo cual se descarta el uso de la variable retrasada pues se observa que el coeficiente $\gamma_1 = -0.0073$ de la variable x_{t-1} (identificado como `comp_lag1`) es prácticamente cero, es decir que las compras del mes anterior no influyen en las ventas del mes actual y por tanto tampoco tienen información que aportar en la predicción. Es posible que este resultado tenga que ver con el hecho de que se compre la mercancía semanalmente, y que como se mencionó, no se cuenta con el espacio para comprar mercancía para más de un mes.

Capítulo 5

Conclusiones

A lo largo del trabajo se obtuvieron estimaciones de las ventas del establecimiento para el año 2015 con distintos métodos. Como se mencionó al principio del documento, estos métodos varían en complejidad tanto en su construcción teórica como en el cálculo de estimaciones, pues algunos podrían hacerse con ayuda de una hoja de cálculo mientras que otros requieren el uso de paquetes estadísticos. En cada método fue posible medir el ajuste del modelo, pero también al finalizar el año fue posible comparar la exactitud de cada método con el que se realizó la predicción. Si bien en el trabajo se manejó una escala arbitraria, para medir el error de estimación se obtuvieron las predicciones en la escala de la moneda nacional, se revirtió la transformación del logaritmo, cuando se aplicó, y se compararon contra las ventas de 2015 deflactadas y sin efecto calendario.

<i>Modelo</i>	<i>ECM</i>
Holt-Winters	3,084.39
SARIMA	1,813.71
Modelo dinámico	1,807.97

Cuadro 5.1: Error cuadrático medio de las predicciones de 2015 de cada modelo

El error cuadrático medio en cada método puede verse en el cuadro 5.1, en este se observa que el modelo con las estimaciones menos precisas fue el modelo Holt-Winters, mientras que el más preciso fue el modelo dinámico. De esto podría concluirse que las estimaciones fueron más precisas entre más complejo fue el modelo usado. Véase que el error del modelo dinámico fue apenas un poco menor que el error del modelo SARIMA, sin embargo, el AICC obtenido en ambos modelos en los capítulos 3 y 4, mostró que el modelo dinámico tuvo un mejor ajuste pues su valor AICC fue de -135, menor que el del modelo SARIMA de -122. Por otro lado, si bien el error muestra qué método proporcionó mejores estimaciones, cada método también proporcionó información relevante del comportamiento de la serie.

Con los métodos clásicos vistos en el capítulo 2, se estimó el patrón estacional de la serie, mediante los índices estacionales normalizados y con estos se logró obtener la serie desestacionalizada. En este primer ejercicio se vio que el comportamiento estacional suele variar muy poco año con año. La tendencia mostró tener un comportamiento menos estable en el tiempo,

pues hasta 2012 parecía subir pero en adelante parecía bajar, por lo que la estimación de los ingresos se hizo con el modelo exponencial Holt-Winters, en el que se tomó en cuenta los resultados de la desestacionalización para estimar valores iniciales. Con este método se confirmó que el comportamiento estacional es poco variable pues el parámetro de suavizamiento exponencial estacional γ fue muy cercano a cero, es decir, se daba prácticamente todo el peso de la estimación estacional a los valores reales de la serie y poco o nada a los valores estimados.

Una ventaja del método Holt-Winters que es un método relativamente sencillo, en su interpretación y en el cálculo de las predicciones. Es decir, tiene lógica pensar que la serie está compuesta por tres elementos nivel, pendiente y estacionalidad, y que estos se combinan de alguna manera para crear la serie observada. El cálculo de las estimaciones podría ser realizado sin problemas por los mismos dueños, pues como se observa en el capítulo 2, se obtuvieron resultados muy similares con la función `ets` de la librería `forecast` de R, lo que se hace en una sola instrucción teniendo las ventas sin inflación ni efecto calendario.

La desventaja de este método es que al dar más peso a las observaciones más recientes, en los datos de este negocio, tomó como tendencia de las predicciones la tendencia a la baja de los últimos dos años, cuando la tendencia no siempre fue así, y además es posible que haya tenido que ver con el robo registrado. De hecho las predicciones con este método tienen menor error al principio de 2015 que al final. En conclusión es un método sencillo, pero cuyas predicciones deben tomarse con cautela sobre todo para predicciones de largo plazo.

En el análisis con el modelo de SARIMA se obtuvo información a partir de los ejercicios realizados para la identificación del modelo. Se observó que no hay una clara tendencia a la alza ni a la baja, si bien en los ejercicios del Capítulo 2 se notó un comportamiento de subida y bajada, las funciones ACF y PACF mostraron que las observaciones no están fuertemente correlacionadas con las observaciones anteriores por lo que se descartó la posibilidad de que exista alguna tendencia en los datos que fuera necesario remover.

Contrariamente, el comportamiento estacional tuvo un marcado patrón. Para modelar tanto el comportamiento estacional como no estacional se usó un modelo $SARIMA(2, 0, 0) \times (1, 0, 1)_{12}$, en donde uno de los dos parámetros autorregresivos p se fijó en cero. Sobre los parámetros P , Q , p y q seleccionados, no hay una clara interpretación pues no hay una teoría que relacione las características de la actividad del negocio con el tipo de proceso que deba seguir su serie de ventas, simplemente el proceso encontrado es el que se ajusta mejor a los datos, de manera que se infiere que las ventas fueron generadas por dicho proceso.

Como desventaja se tiene que la identificación del modelo no es fácil, pues existe una amplia variedad de modelos a elegir, cada uno con una medida de ajuste y comportamiento de residuales distintos. En ocasiones si el ajuste de un modelo mejora con respecto a otro que se tuviese contemplado, lleva a considerar cambiar de opinión. Esto hizo necesario probar muchos modelos y apoyarse en herramientas de programación para obtener información de todos estos modelos posibles a elegir. Al final, en la elección del modelo no se tomó en cuenta exclusivamente el ajuste, sino que también se observó el cumplimiento de los supuestos, descartando aquellos modelos que, a pesar de tener buen ajuste, incumplían alguno de los

supuestos de los residuales.

Otra desventaja del modelo es que no tiene una interpretación sencilla, a diferencia del modelo anterior tipo Holt-Winters, en este no es tan intuitivo ver a la serie como una combinación de choques aleatorios, a veces de sus mismos valores con retraso, a veces de un ruido blanco o ambas situaciones. Más aún, la interpretación en los modelos ARIMA y SARIMA es un poco más difícil al incluir parámetros para modelar la tendencia y estacionalidad. Además conforme se desarrolla el análisis aparecen algunos aspectos en la identificación en los que es necesario profundizar en conocimientos técnicos, por ejemplo, las pruebas de raíces unitarias. La ventaja de toda esta teoría detrás del modelo es que proporciona predicciones mucho más cercanas a las ventas reales de 2015, comparadas con las obtenidas con Holt-Winters, como puede verse en el cuadro 5.1, pues el ECM es considerablemente más bajo.

El tercer modelo probado, el modelo dinámico, mostró adecuarse muy bien a los datos, y tuvo también la utilidad de corroborar la hipótesis de que el monto de las compras hechas en un mes afecta de manera positiva a la venta esperada del mismo mes. Esta es una de las conclusiones más importantes del trabajo, pues dado que el coeficiente de regresión de la variable compras fue significativamente mayor que cero, fue posible informar a los dueños del negocio dos aspectos, el primero es que sus compras afectan a sus ventas de manera significativa, y el segundo cuánto sube la venta esperada por cada peso de compra de mercancía. También se comprobó que las compras de un mes no tienen un peso significativo en las ventas del mes siguiente, esto se relacionó a que no se cuenta con el suficiente espacio para guardar mercancía para más de un mes.

Este modelo puede considerarse más complejo que el anterior, pues utiliza también los modelos SARIMA con todos los conceptos asociados, sin embargo, la identificación del modelo del error fue menos compleja gracias a que se utilizaron modelos *proxy*, los cuales permitieron llegar de forma más rápida al modelo con mejor ajuste. Por otro lado, las predicciones del modelo resultaron un poco más precisas que las obtenidas con el modelo SARIMA, pero de mayor utilidad fue obtener intervalos de confianza más reducidos que con el modelo SARIMA.

En el futuro es probable que pudieran mejorarse las estimaciones obtenidas con este modelo, es decir, las compras siguen a grandes rasgos el comportamiento estacional de las ventas pero con mayor variación, por lo que no eliminan tanta variación en la serie de ventas. Para obtener estimaciones más precisas, a los dueños se les recomienda que tengan más control en el dinero destinado a compras, si bien es cierto que dependen en gran medida del monto vendido, un patrón de compras más claro podría disminuir el error en la estimación. También con el mismo fin, el registro de esta variable debería hacerse de forma más detallada, pues se observó que su registro es menos riguroso que el de los ingresos, lo cual crea la posibilidad de error en la recolección de los datos, que a su vez aporta en la variabilidad de las compras.

Por último, no siempre el mejor método es el más preciso, por ejemplo, si se requieren estimaciones de las ventas futuras para decidir el monto a comprar el mes siguiente, se sabe entonces que se tiene poco tiempo para el análisis, y que podría sacrificarse un poco de precisión usando el método Holt-Winters sin tanto problema pues el error no es grande para estimaciones

a corto plazo. Por otro lado, si más bien se requiere una predicción a largo plazo para tomar decisiones financieras, o bien se quiere medir el impacto de alguna otra variable en las ventas, entonces lo que se busca es precisión en las predicciones o en los coeficientes estimados, en tal caso sería mejor utilizar los modelos SARIMA o dinámicos. Ambos modelos requieren mayor tiempo de análisis, no solo porque deba elegirse uno entre una gran variedad sino también porque son modelos que se deben ir actualizando conforme se recaben nuevas observaciones, pues estas pueden provocar cambios en los valores de los parámetros al realizar nuevamente los cálculos, o incluso cambiar número de parámetros a estimar o el tipo de modelo. En el modelo Holt-Winters las nuevas estimaciones también provocan cambios en los valores de los parámetros, pero el cálculo es exactamente el mismo y el modelo no tendría por qué cambiar.

En resumen, la precisión de las estimaciones es siempre un objetivo del análisis pero la elección de algún método no sólo depende de la precisión, sino también de circunstancias particulares de cada situación, como la intención y tiempo con que se cuente para hacer el análisis.

Bibliografía

1. Anderson, D. R. (2008). *Model Based Inference in the Life Sciences: A Primer on Evidence*. Colorado, United States: Springer.
2. Anderson, D. R., Sweeney, D. J., Williams, T.A., Camm, J.D., Cochran J.J. (2015). *Statistics for Business & Economics*. USA: Cengage Learning, 13a ed.
3. Brockwell, P. J. & Davis, R. A.(1991). *Time Series: Theory and Methods*. New York: Springer-Verlag, 2nd ed.
4. Brockwell, P. J. & Davis, R. A.(2002). *Introduction to time deries and forecasting*. New York: Springer-Verlag, 2nd ed.
5. Dougherty, C. (2011). *Introduction to econometrics*. New York, USA: Oxford University Press.
6. Guerrero-Guzmán, V. M. (1991). *Análisis estadístico de series de tiempo económicas*. México: Thomson, 2a ed.
7. Gujarati, D. N. y Porter, D. C. (2010). *Econometría*. (Pilar Carril, trad.). México: Mc Graw Hill. 5a edición. (Obra original publicada en 2009).
8. Hyndman, R.J. & Athanasopoulos G. (2012), *Forecasting: principles and practice*. Recuperado de: <https://www.otexts.org/fpp>
9. Hyndman, R.J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3).
Recuperado desde: <https://www.jstatsoft.org/article/view/v027i03>
10. Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin, Germany: Springer.
11. Hyndman, R. J. (29 de Noviembre de 2010). Initializing the Holt-Winters method [Mensaje en un blog]. Recuperado de: <http://www.r-bloggers.com/initializing-the-holt-winters-method/>
12. Hyndman, R. J. (27 de Enero de 2014). New in forecast 5.0 [Mensaje en un blog]. Recuperado de: <http://robjhyndman.com/hyndsight/forecast5/>
13. Hyndman, R. J. (29 de Mayo de 2011). Comparing HoltWinters() and ets() [Mensaje en un blog]. Recuperado de: <http://robjhyndman.com/hyndsight/estimation2/>.

14. Jain, D. R. & Jhunjhunwala. (2007). *Business Statistics*. New Delhi: Tata McGraw-Hill.
15. Montero, J. M. (2007). *Estadística Descriptiva*. Madrid, España: Thompson Ediciones Paraninfo.
16. Morilla, C. (2000). *Análisis de Series Temporales*. Madrid, España: La Muralla.
17. Sabau H. (2011). *Análisis Económico Dinámico*. México: Universidad Iberoamericana.
18. Shumway, R. H. & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications: With R Examples*. New York, USA: Springer. 3rd ed.
19. Tsay, R. (2005). *Analysis of financial time series*. New Jersey, USA: Wiley-Interscience. 2nd ed.
20. Wei W. (2006), *Time Series Analysis Univariate and Multivariate Methods*. Boston, USA: Pearson Addison Wesley, 2nd ed.