

Notas de Clase: Series de Tiempo

Benjamín Oliva, Omar Alfaro Rivera & Emiliano Pérez Caullieres

2023-09-25

Contents

| | |
|---|-----------|
| Análisis de Series de Tiempo | 5 |
| Contexto del documento | 5 |
| 1 Introducción | 7 |
| 1.1 La naturaleza de los datos de Series de Tiempo | 8 |
| 1.2 Ejemplos y aplicaciones de las Series de Tiempo | 8 |
| 2 Elementos de Ecuaciones en Diferencia | 17 |
| 2.1 Ecuaciones en Diferencia para procesos deterministas | 17 |
| 2.2 Operador de rezago L | 33 |
| 3 Modelos de Series de Tiempo Estacionarias | 37 |
| 3.1 Definición de ergodicidad y estacionariedad | 37 |
| 3.2 Función de autocorrelación | 44 |
| 4 Procesos estacionarios univariados | 51 |
| 4.1 Procesos Autoregresivos (AR) | 51 |
| 4.2 Procesos de Medias Móviles (MA) | 76 |
| 4.3 Procesos ARMA(p, q) y ARIMA(p, d, q) | 81 |
| 4.4 Función de Autocorrelación Parcial | 87 |
| 4.5 Selección de las constantes p, q, d en un AR(p), un MA(q), un ARMA(p, q) o un ARIMA(p, d, q) | 89 |
| 4.6 Pronósticos | 94 |

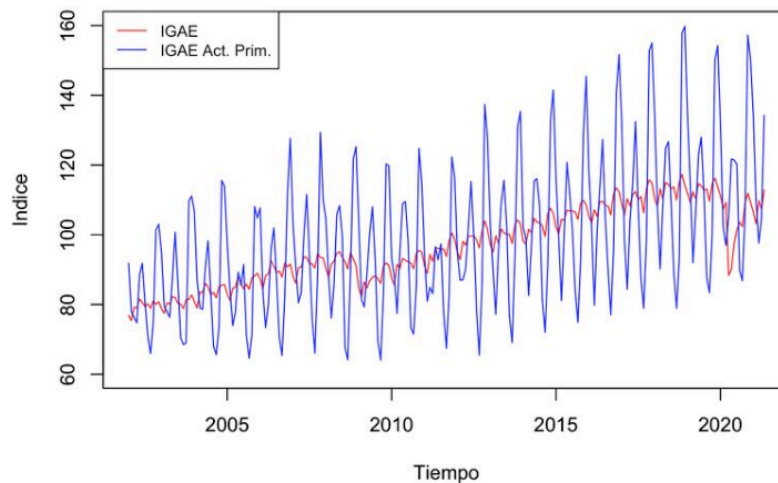
| | | |
|----------|--|-----------|
| 5 | Desestacionalización y filtrado de Series | 99 |
| 5.1 | Motivación | 99 |
| 5.2 | Filtro Hodrick-Prescott | 101 |

Análisis de Series de Tiempo

Notas de Clase

Otoño 2023

by Benjamín Oliva, Omar Alfaro Rivera & Emiliano Pérez Caullieres



Contexto del documento

Estas son notas de clase de la materia de Análisis de Series de Tiempo en la Facultad de Economía de la Universidad Nacional Autónoma de México.

Este es un trabajo siempre en proceso de mejora, para cualquier comentario o aclaración, contactar al correo benjov@ciencias.unam.mx o omarxalpha@gmail.com.

La versión PDF de estas notas se encuentra en: <https://github.com/benjov/Series-Tiempo/blob/main/docs/Notas-Series-Tiempo.pdf>

Las tablas de datos empleadas en los diferentes ejemplos se encuentran en: <https://github.com/benjov/Series-Tiempo/tree/main/BD>

Chapter 1

Introducción

Estas notas son un resumen, una síntesis comparativa y, en algunos casos, una interpretación propia de los libros de texto de Brooks (2019), Cowpertwait y Metcalfe (2009), Guerrero-Guzmán (2014), Enders (2015), Franses y van Dijk (2003), Kirchgassner, Wolters, y Hassler (2012), Lutkepohl (2005), Wei (2019), entre otros. En algunos casos se incorpora información adicional para efectos de dar contexto al tema analizado (ver sección de Bibliografía para mayores detalles).

El objetivo de este documento es proporcionar un conjunto de apuntes o notas que sirvan de apoyo para la clase de Series de Tiempo en la Facultad de Economía de la UNAM. Por esta razón, no deben considerarse como notas exhaustivas o como un sustituto de la clase y los laboratorios. Asimismo, es deseable que los alumnos puedan aportar sus observaciones y correcciones a estas notas, las observaciones a estas notas son esperadas y siempre serán bienvenidas y agradecidas.

Este es un trabajo siempre en proceso de mejora, para cualquier comentario o aclaración, contactar al correo benjov@ciencias.unam.mx o omarxalpha@gmail.com.

En estas notas se estudian los temas que típicamente son incluidos como parte de un curso estándar de análisis de series de tiempo y agrega otros tantos, los cuales son:

1. Modelos estacionarios univariados: $AR(p)$, $MA(q)$, $ARMA(p, q)$ y $ARIMA(p, d, q)$, y filtros para eliminar estacionalidad, entre otros;
2. Modelos no estacionarios univariados y Pruebas de raíz unitaria (o pruebas para determinar que una serie es estacionaria);
3. Modelos multivariados, entre lo que se incluye a los Vectores Autoregresivos (VAR) y los procedimientos de Cointegración

4. Modelación de series univariadas con errores con heterocedasticidad y autocorrelación: ARCH(r), GARCH(n), etc.;
5. Modelos multivariados con errores con heterocedasticidad y autocorrelación: M-GARCH y M-GARCH-M;
6. Casos particulares en los que las series incluidas en un modelo multivariado no son del mismo orden de integración, conocidos como modelos ADRL.
7. Modelos de Datos Panel en series de tiempo, y
8. Modelos no lineales como los de cambios de régimen.

1.1 La naturaleza de los datos de Series de Tiempo

El análisis de series de tiempo tiene muchas aplicaciones en diversos campos de la ciencia. Por ejemplo, en la economía continuamente se está expuesto a información recopilada de los mercados financieros, indicadores de empleo, índices o indicadores del nivel de producción, índices de precios, etc. En otros campos de las ciencias sociales se emplea el análisis de series de tiempo para analizar la evolución de la población, los nacimientos, o el número de personas con matrículas escolares. Finalmente, en las ciencias exactas se pueden encontrar casos como los de un epidemiólogo que puede estar interesado en el número de casos de influenza observados en algún periodo de tiempo dado y si a estos se les puede asociar con algún tipo de estacionalidad o si se trata del inicio de un fenómeno atípico.

La primera aproximación que se suele tener a las series de tiempo es mediante el examen de datos puestos en una gráfica, en la cual uno de los ejes es el tiempo y el otro es el valor tomado por la variable. No obstante, en este tipo de exámenes existen dos enfoques. Por un lado, existe el enfoque de la importancia del tiempo, el cual consiste en reconocer cómo lo que sucede hoy es afectado por lo que pasó ayer o, en general, en periodos pasados, o cómo lo que pasa hoy afectará los eventos futuros. Por otro lado, existe el enfoque del análisis frecuentista o de frecuencia, mediante el cual se busca reconocer la importancia que tiene para los investigadores los ciclos –por ejemplo, ciclos estacionales, momentos de crisis económicas, etc.–

1.2 Ejemplos y aplicaciones de las Series de Tiempo

Un primer ejemplo que puede ilustrar la presencia de los dos tipos de enfoques antes mencionadas es la Figura 1.1. En esta figura se muestra la evolución del

Indicador Global de la Actividad Económica (IGAE) en su versión global o del total de la economía y en su versión únicamente para las actividades primarias entre enero de 2002 y mayo de 2023.

```
library(readxl)

Base_1 <- read_excel("BD/Base_1_TimeSeries.xlsx")
IGAE_2013 <- ts(Base_1$IGAE_2013, start = 2002, freq = 12)
IGAE_PRIM_2013 <- ts(Base_1$IGAE_PRIM_2013, start = 2002, freq = 12)
ICC <- ts(Base_1$ICC, start = 2002, freq = 12)
ICC_LAG <- ts(Base_1$ICC_LAG, start = 2002, freq = 12)
IPC_BMV <- ts(Base_1$IPC_BMV, start = 2002, freq = 12)
TDC <- ts(Base_1$TDC, start = 2002, freq = 12)

plot(IGAE_2013, type = "l", lwd = 1, col = "red", ylab = "Indice",
     xlab = "Tiempo", ylim = c(60,160))
par(new = T)
# Indicador Global de la Actividad Económica, Act. Prim., base 2008
plot(IGAE_PRIM_2013, type = "l", lwd = 1, col = "blue",
     ylab = "Indice", xlab = "Tiempo", ylim = c(60,160))
# Leyenda
legend("topleft", c("IGAE","IGAE Act. Prim."), cex = 0.8,
      lty = 1:1, col = c("red", "blue"))
```

```
par(new = F)
```

Como se puede observar, el IGAE del total de la economía muestra, principalmente, que el enfoque del tiempo es más relevante. Es decir, que existe cierta persistencia en el indicador, lo que significa que la economía crece en razón del crecimiento reportado en periodos pasados. No obstante, lo que no podemos reconocer es que los eventos futuros tienen un efecto en el desempeño de la economía hoy día. Así, no es común observar cambios abruptos del indicador, salvo por la crisis global de 2008 y la reciente crisis causada por la Covid-19.

```
plot(ICC, type = "l", lwd = 1, col = "red", ylab = "Indice",
     xlab = "Tiempo", ylim = c(29, 50))
# Comando que indica a R que sin borrar la grafica anterior,
# grafique la siguiente.
par(new = T)
# Indice ??Como considera usted la situacion economica del pais
# hoy en dia comparada con la de hace 12 meses?, base enero 2003
plot(ICC_LAG, type = "l", lwd = 1, col = "blue", ylab = "Indice",
     xlab = "Tiempo", ylim = c(29,50))
# Leyenda
```

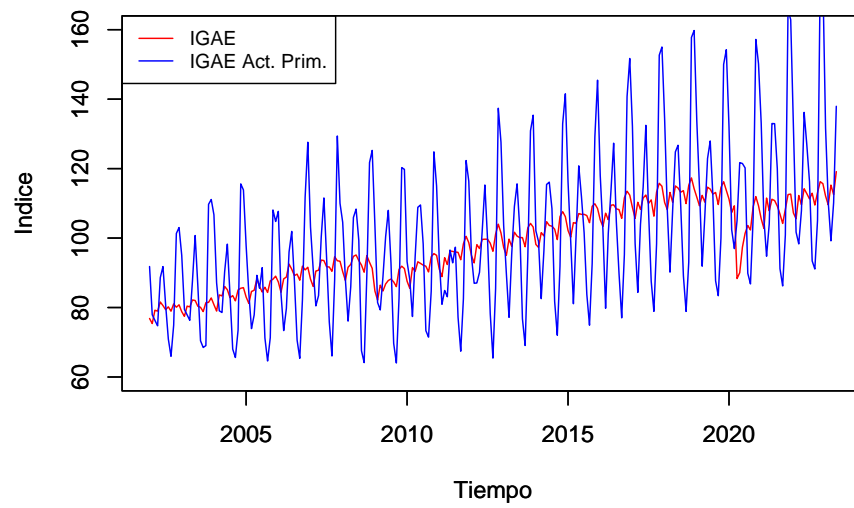


Figure 1.1: Indicador Global de Actividad Económica (IGAE) Global y para las Actividades Primarias (2008=100), Ene.2002 - May.2023

```
legend("bottomleft", c("ICC", "ICC lag"), cex = 0.8, lty = 1:1,
      col = c("red", "blue"))
```

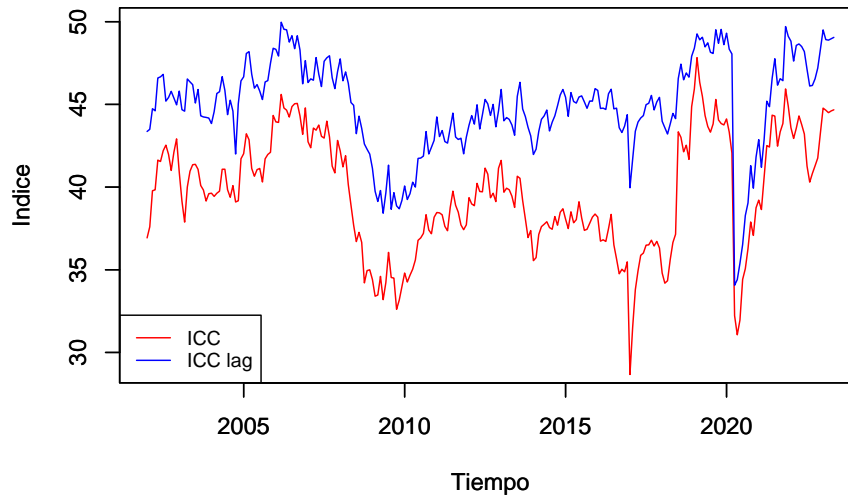


Figure 1.2: Índice de Confianza del Consumidor (ICC): General y resultado de ¿Cómo considera usted la situación económica del país hoy en día comparada con la de hace 12 meses? (puntos), Ene.2002-may.2023

```
par(new = F)
```

Por el contrario, el IGAE de las actividades primarias muestra una presencia significativa de la importancia de la frecuencia. No pasa desapercibido que existen muchos ciclos en la evolución del indicador. Algo que suena común en las actividades primarias, cuya producción depende de eventos que son cíclicos agrícolas asociados con el clima u otros factores determinantes de la oferta de productos agrícolas. Otro factor que puede incluir en el indicador son elementos de demanda, más que los de oferta. Por ejemplo, el consumo de alimentos típicos de algunas temporadas del año.

Como segundo ejemplo, en la Figura 1.2 se ilustra la evolución reciente del índice de Confianza del Consumidor (ICC) en dos de sus versiones: i) el Índice global y ii) el Índice de confianza de los consumidores cuando estos consideran la situación actual en la economía en relación el año anterior.

Destacamos que el ICC mide las expectativas de los consumidores en razón de la información pasada y de la esperada, según dichos consumidores.

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

# Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores
plot(IPC_BMV, type = "l", lwd = 1, col = "red", ylab = "Indice",
      xlab = "Tiempo", main = "Indice de Precios y \nCotizaciones BMV")
# Tipo de Cambio para Solventar Obligaciones en Moneda Extranjera
plot(TDC, type = "l", lwd = 1, col = "blue", ylab = "Pesos X Dolar",
      xlab = "Tiempo", main = "Tipo de Cambio")
```

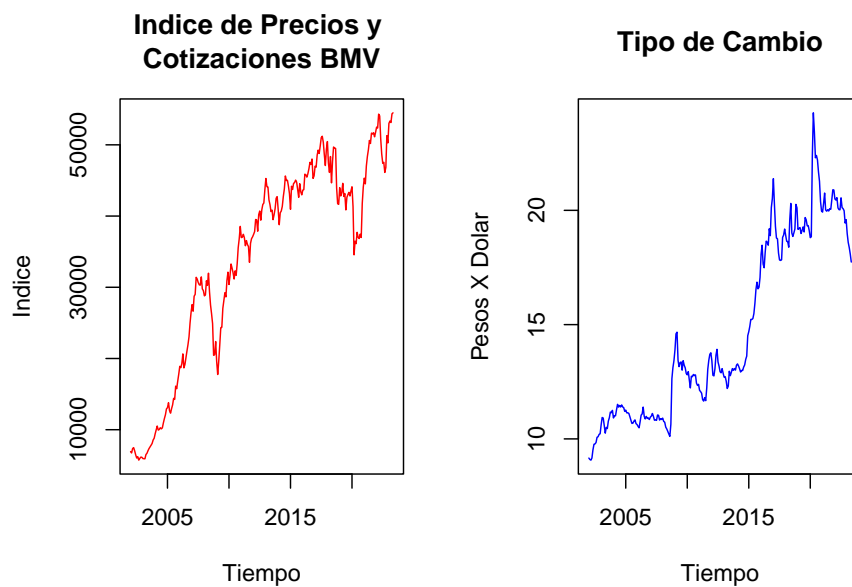


Figure 1.3: índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (Panel Derecho) y Tipo de Cambio para Solventar Obligaciones en Moneda Extranjera, pesos por dólar (Panel izquierdo), Ene.2002-May.2023

```
par(mfrow=c(1,1))
```

Así, es probable que las dos series de tiempo exhiban un gran peso para los eventos pasados, pero a la vez, un componente –probablemente menor– del componente de frecuencia. Esto último en razón de que los consumidores suelen considerar en sus expectativas de consumo los períodos cíclicos de la economía: temporadas navideñas, pagos de colegiaturas, etc. Este segundo ejemplo también ilustra que la confianza del consumidor no necesariamente está directamente correlacionada con el desempeño de la economía.

Como tercer ejemplo se muestra la evolución de dos series. La Figura 1.3 ilustra el comportamiento reciente de dos indicadores que son referencia para los inversionistas. Por un lado, se ubica el índice de Precios y Cotizaciones de la BMV (IPC), el cuál refleja el valor de las acciones de empresas que cotizan en la BMV y el volumen de acciones comercializadas, en conjunto. De esta forma, se ha interpretado que el IPC refleja el rendimiento del capital promedio invertido en las empresas que cotizan en la BMV.

Por otro lado, en la Figura 1.3 se presenta la evolución del Tipo de Cambio (TDC) –indicador financiero que se suele utilizar como medio de reserva de valor–. Esto, en razón de que el TDC es conocido como un instrumento que en momentos de crisis toma valores contracíclicos de la economía mexicana. No obstante, ambos indicadores no son comparables. Para hacerlos comparables en la Figura 1.4 se presentan en versión índice con una base en el primer mes de la muestra.

```
IPC_BMV_I <- 100*IPC_BMV/IPC_BMV[1]
TDC_I <- 100*TDC/TDC[1]
# Índice del índice de Precios y Cotizaciones de la Bolsa Mexicana
# de Valores
plot(IPC_BMV_I, type = "l", lwd = 1, col = "red", ylab = "Indice",
      xlab = "Tiempo", ylim = c(80,740))
# Comando que indica a R que sin borrar la grafica anterior,
# grafique la siguiente.
par(new = T)
# Índice del Tipo de Cambio para Solventar Obligaciones en
# Moneda Extranjera
plot(TDC_I, type = "l", lwd = 1, col = "blue", ylab = "Indice",
      xlab = "Tiempo", ylim = c(80,740))
# Leyenda
legend("topleft", c("Índice del IPC","Índice del TDC"), cex = 0.8,
      lty = 1:1, col = c("red", "blue"))
```

```
par(new = F)
```

En la perspectiva de la Figura 1.4 se puede apreciar que el TDC no es tan rentable, ya que una inversión en la BMV mediante el IPC, en el largo plazo, muestra más rendimientos. Asimismo, la Figura 1.4 ilustra que en ambas series se observa un dominio de la condición de tiempo y no uno de frecuencia. Es decir, tanto el IPC como el TDC no responden a condiciones como ciclos o temporadas que si son observables en actividades económicas como las primarias.

Finalmente, la Figura 1.5 ilustra una característica que también resulta de gran interés en el análisis de series de tiempo: los datos de alta frecuencia y de comportamiento no regular. Como se puede observar, en la Figura 1.5 se muestran

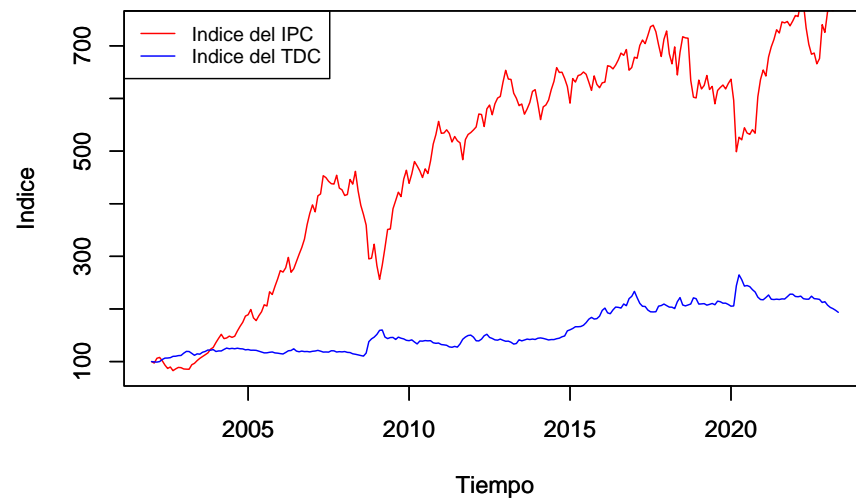


Figure 1.4: Índice del índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (Panel Derecho) e Índice del Tipo de Cambio para Solventar Obligaciones en Moneda Extranjera (ambos enero de 2002 = 100), pesos por dólar (Panel izquierdo), Ene.2002-May.2023

las diferencias logarítmicas de las series de IGAE de la actividad total, el IPC y el TDC.

```
par(mfrow=c(3,1))
# Indicador Global de la Actividad Econ?mica, base 2008
plot(diff(log(IGAE_2013), lag = 1), type = "l", lwd = 1,
     col = "red", ylab = "Var. %", xlab = "Tiempo",
     main = "Indicador Global de la Actividad Economica")
# Indice de Precios y Cotizaciones de la Bolsa Mexicana de Valores
plot(diff(log(IPC_BMV), lag = 1), type = "l", lwd = 1,
     col = "red", ylab = "Var. %", xlab = "Tiempo",
     main = "Indice de Precios y Cotizaciones BMV")
# Tipo de Cambio para Solventar Obligaciones en Moneda Extranjera
plot(diff(log(TDC), lag = 1), type = "l", lwd = 1,
     col = "blue", ylab = "Pesos X Dolar", xlab = "Tiempo",
     main = "Tipo de Cambio")
```

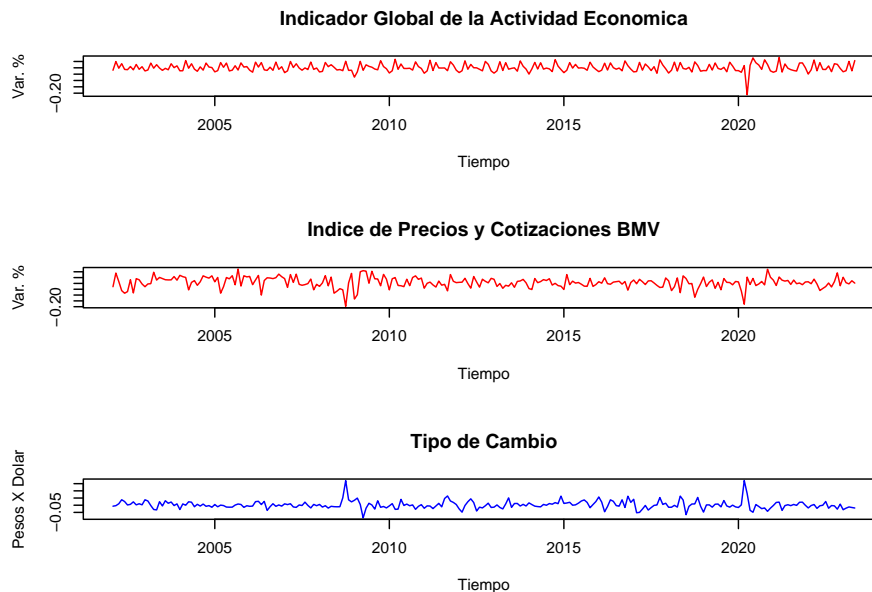


Figure 1.5: Tasas de Crecimiento mensuales (diferencias logarítmicas) de Indicador Global de la Actividad Económica, Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (Panel Derecho) y Tipo de Cambio para Solventar Obligaciones en Moneda Extranjera, Ene.2002-May.2023

```
par(mfrow=c(1,1))
```

Dichas diferencias se pueden interpretar como una tasa de crecimiento de las series por las siguientes razones. Consideremos una serie de tiempo dada por y_t , cuya versión logarítmica es $\ln(y_t)$. De esta forma, la diferencia logarítmica esta dada por la ecuación (1.1):

$$\Delta \ln(y_t) = \ln(y_t) - \ln(y_{t-1}) = \ln\left(\frac{y_t}{y_{t-1}}\right) \quad (1.1)$$

Ahora bien, si retomamos la definición de tasa de crecimiento (TC) de una serie de tiempo y_t entre el periodo t y $t - 1$ podemos obtener que:

$$TC = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 \quad (1.2)$$

De esta forma, si tomamos el logaritmo de la expresión de la ecuación (1.2) obtenemos la siguiente aproximación:

$$\frac{y_t}{y_{t-1}} - 1 \approx \ln\left(\frac{y_t}{y_{t-1}}\right) = \ln(y_t) - \ln(y_{t-1}) \quad (1.3)$$

La ecuación (1.3) es cierta cuando los valores de y_t y y_{t-1} son muy parecidos, es decir, cuando las variaciones no son tan abruptas. Otra forma de interpretar la ecuación (1.3) es que para tasas de crecimiento pequeñas, se puede utilizar como una buena aproximación a la diferencia logarítmica mostrada en la ecuación (1.1).

En la Figura 1.5 se reportan las diferencias logarítmicas del IGAE, IPC y TDC, todos, como una media de distintos tipos de redimientos. Es decir, podemos decir que un capitalista promedio (suponiendo que solo puede invertir en la actividad económica, en la bolsa o en el dólar), puede observar que le es más redituable en función de sus preferencias.

Notése que la dinámica de las variaciones de cada una de las series es significativamente diferente. Destaca que el TDC es una de las variables que, en general, no muestra grandes cambios a lo largo del tiempo. No obstante, se han observado cambios radicales, cuando menos en el año 2008. Lo anterior, son características que se han observado para el IPC. En cambio, el IGAE muestra un comportamiento más estable o estacionario.

Chapter 2

Elementos de Ecuaciones en Diferencia

2.1 Ecuaciones en Diferencia para procesos deterministas

En el capítulo previo se hizo una introducción al concepto de series de tiempo. En este Capítulo se pretende desarrollar la construcción de los procesos generadores de datos de las series de tiempo. En un sentido más formal, se expondrá que las series de tiempo se pueden considerar como una secuencia de variables aleatorias.

Para tales efectos, se desarrollará una introducción al concepto de ecuaciones en diferencia. Así, las preguntas que se pretende responder son:

1. ¿Cuál es la solución de la ecuación en diferencia que se estudia?
2. ¿Cuáles son las condiciones para que un proceso estocástico, representado mediante una ecuación en diferencia, llegue a alcanzar un punto de equilibrio en el largo plazo?

El término de *ecuación en diferencia* sirve para denominar un proceso similar o equivalente dentro de las ecuaciones diferenciales, dentro del cual se consideran a un conjunto de variables que están en función del tiempo. Así, si consideramos al tiempo como una variable continua, es decir, consideramos una variable $Z(t)$, podemos expresar las siguientes expresiones para la ecuación diferencial:

$$\frac{dZ(t)}{dt}; \frac{d^2Z(t)}{dt^2}; \dots; \frac{d^kZ(t)}{dt^k} \quad (2.1)$$

Por otro lado, suponiendo el caso del tiempo en forma discreta, es decir, con $t = \dots, -2, -1, 0, 1, 2, \dots$, entonces el comportamiento de la serie de variables dadas por Z_t , la cual se puede expresar como:

$$\Delta Z_t; \Delta^2 Z_t; \dots; \Delta^k Z_t \quad (2.2)$$

Observemos que una forma técnicamente más correcta es escribir las expresiones anteriores como:

$$\frac{\Delta Z_t}{\Delta t}; \frac{\Delta^2 Z_t}{\Delta t^2}; \dots; \frac{\Delta^k Z_t}{\Delta t^k} \quad (2.3)$$

No obstante, no pasa desapercibido que $\Delta t = 1$, por lo que resultan equivalentes ambos conjuntos de expresiones (2.2) y (2.3).

2.1.1 Ecuaciones en Diferencia Lineales de Primer Orden

El primer caso que se suele estudiar en relación a Ecuaciones en Diferencia es el de las Ecuaciones en Diferencia Lineales de Primer Orden. Al respecto, al igual que en el caso continuo, las variaciones de la variable Z_t se pueden expresar como se ilustra en el siguiente ejemplo. Consideremos la siguiente ecuación:

$$Z_t = a_0 + a_1 Z_{t-1} \quad (2.4)$$

Donde, $t = \dots, -2, -1, 0, 1, 2, \dots$, y a_0 y $a_1 \neq 0$ son números reales constantes. De (2.4) podemos despejar la variable Z_{t-1} y obtener una forma de ecuación en diferencia:

$$Z_t - a_1 Z_{t-1} = a_0 \quad (2.5)$$

Ahora denotemos a $LZ_t = Z_{t-1}$, es decir, mediante el operador L se puede rezagar una variable dada. En general, podemos decir que el operador tiene dos propiedades, la primera es que es lineal en el sentido de que abre sumas y saca escalares como se muestra en la siguiente expresión para el caso de un (1) rezago:

$$L(\alpha Z_t + \beta) = \alpha Z_{t-1} + \beta \quad (2.6)$$

Donde $\alpha, \beta \in \mathbb{R}$ y $\alpha, \beta \neq 0$. Otro resultado implícito en esta primera propiedad es que el operador rezago aplicado a cualquier escalar dará como resultado el escalar, puesto que este es una constante sin importa el momento t en el cual se encuentre la variable Z .

La segunda propiedad del operador es que se puede aplicar de forma consecutiva a una misma variable. Es decir, $L(Z_{t-1}) = LLZ_t = L^2 Z_t$, por lo que en general tendremos: $L^p Z_t = Z_{t-p}$ (con $p \in \mathbb{Z}$). Así, en el caso de p rezagos la propiedad de linealidad del operador rezago será:

$$L^p(\alpha Z_t + \beta) = \alpha Z_{t-p} + \beta \quad (2.7)$$

Dicho lo anterior podemos escribir la solución general de (2.5) como:

$$\begin{aligned}
 Z_t - a_1 L Z_t &= a_0 \\
 (1 - a_1 L) Z_t &= a_0 \\
 Z_t &= a_0 \frac{1}{1 - a_1 L} + s a_1^t \\
 Z_t &= a_0 \frac{1}{1 - a_1} + s a_1^t
 \end{aligned} \tag{2.8}$$

Donde $a_1 \neq 1$ y $t = \dots, -2, -1, 0, 1, 2, \dots$. Notése que la aplicación del operador rezago L a la constante a_1 dará como resultado el valor de la misma constante, ya que ésta no depende del momento t en el cuál observemos a la variable Z_t . En la ecuación (2.8) se adiciona un término $s a_1^t$ que permite ubicar la trayectoria inicial de la solución de la ecuación. El componente no significa un cambio respecto de la ecuación (2.5) original, ya que si buscáramos reconstruir a ésta ecuación tendríamos:

$$\begin{aligned}
 (1 - a_1 L) s a_1^t &= s a_1^t - a_1 s L a_1^t \\
 &= s a_1^t - a_1 s a_1^{t-1} \\
 &= s a_1^t - s a_1^t \\
 &= 0
 \end{aligned}$$

La ecuación (2.8) se suele interpretar como la solución de largo plazo. Ahora demostraremos por qué es cierta la ecuación y discutiremos algunas condiciones que se deben observar en esta solución para que sea una solución convergente. No obstante, primero discutiremos un método indirecto e incompleto para demostrar el resultado, dicho método es conocido como el método iterativo. Plantearemos las siguientes ecuaciones particulares donde suponemos la existencia del valor inicial Z_0 del proceso:

$$Z_1 = a_0 + a_1 Z_0$$

$$\begin{aligned}
 Z_2 &= a_0 + a_1 Z_1 \\
 &= a_0 + a_1 (a_0 + a_1 Z_0) \\
 &= a_0 + a_0 a_1 + a_1^2 Z_0 \\
 &= a_0 (1 + a_1) + a_1^2 Z_0
 \end{aligned}$$

$$\begin{aligned}
 Z_3 &= a_0 + a_1 Z_2 \\
 &= a_0 + a_1 (a_0 + a_0 a_1 + a_1^2 Z_0) \\
 &= a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 Z_0 \\
 &= a_0 (1 + a_1 + a_1^2) + a_1^3 Z_0
 \end{aligned}$$

De lo anterior se puede inferir que el método iterativo convergerá hacia una expresión como la siguiente en el momento t :

$$\begin{aligned} Z_t &= a_0 + a_1 Z_{t-1} \\ &= a_0(1 + a_1 + a_1^2 + \dots + a_1^{t-1}) + a_1^t Z_0 \\ &= a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t Z_0 \end{aligned} \quad (2.9)$$

Donde, es necesario que en la ecuación (2.9) se cumpla que $|a_1| < 1$ para que la suma sea convergente –más adelante detallaremos esta afirmación–. A este tipo de ecuaciones se les puede denominar como lineales. Esto en razón de que ningún término de la variable Z aparece elevado a ninguna potencia distinta a 1. También, son de primer orden, ya que el rezago de la variable Z es sólo de un período.

En adelante trabajaremos con ecuaciones en las que la variable Z se encuentra rezagada en cualquiera de los siguientes casos:

$$Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots, Z_{t-p}, \dots \quad (2.10)$$

Por lo que diremos que en adelante el curso versará sobre ecuaciones en diferencia lineales y de cualquier orden p .

Retomando la ecuación (2.9) y considerando la parte de la suma de los términos de a_1^i , de tal forma que buscaremos dar una expresión más compresible a dicho término. Definamos la siguiente expresión como:

$$S_{t-1} = \sum_{i=0}^{t-1} a_1^i \quad (2.11)$$

Por lo tanto, S_t estaría dado por la siguiente expresión:

$$\begin{aligned} S_t &= a_1 \sum_{i=0}^{t-1} a_1^i \\ &= a_1(1 + a_1 + a_1^2 + \dots + a_1^{t-1}) \\ &= a_1 + a_1^2 + a_1^3 + \dots + a_1^t \\ &= a_1 S_{t-1} \end{aligned} \quad (2.12)$$

Tomando los dos resultados de las ecuaciones (2.11) y (2.12) anteriores, podemos expresar que si a S_{t-1} le restamos S_t , y desarrollando ambos lados de la ecuación anterior podemos obtener:

$$\begin{aligned} S_{t-1} - a_1 S_{t-1} &= S_{t-1} - S_t \\ (1 - a_1) S_{t-1} &= (1 + a_1 + a_1^2 + \dots + a_1^{t-1}) - (a_1 + a_1^2 + a_1^3 + \dots + a_1^t) \\ (1 - a_1) S_{t-1} &= 1 - a_1^t \end{aligned}$$

Así, podemos concluir que:

$$S_{t-1} = \frac{1 - a_1^t}{1 - a_1} \quad (2.13)$$

Conjuntando éste último resultado de la ecuación (2.13) con la ecuación (2.9) tenemos la siguiente solución por el método de iteración:

$$Z_t = a_0 \left(\frac{1 - a_1^t}{1 - a_1} \right) + a_1^t Z_0 \quad (2.14)$$

De esta forma la ecuación (2.14) es una solución para la ecuación (2.9), que es una ecuación de un proceso de una Ecuación en Diferencia plantenado en la ecuación (2.4). Está solución aún no es general, en el sentido de que sea válida para cualquier tipo de proceso: convergente o divergente. Dicha convergencia o divengencia estará determinada por el parámetro a_1 . No debe pasar desapercibido que cuando $t \rightarrow \infty$ o cuando la muestra es muy grande (lo que es equivalente), podemos decir que la solución solo puede converger a la siguiente expresión cuando se considera que $|a_1| < 1$:

$$Z_t = a_0 \left(\frac{1}{1 - a_1} \right) \quad (2.15)$$

Retomemos ahora el caso general descrito en la ecuación (2.8) y determinemos una solución general en la cual $a_1 \neq 1$ y $t = \dots, -2, -1, 0, 1, 2, \dots$. Para ello observemos que el siguiente componente en la ecuación mencionada se puede interpretar como la suma infinita de términos descritos como:

$$\begin{aligned} \frac{1}{1 - a_1} &= 1 + a_1 + a_1^2 + \dots + a_1^t + \dots \\ &= \sum_{i=0}^{\infty} a_1^i \end{aligned} \quad (2.16)$$

Donde claramente es necesario que $|a_1| < 1$. Por lo tanto, sólo faltaría determinar el valor de la constante s en la ecuación (2.8) de la siguiente forma, supongamos que observamos el proceso en el momento inicial, por lo que es posible determinar el valor de la constante conociendo el valor inicial del proceso como sigue:

$$Z_0 = a_0 \frac{1}{1 - a_1} + s \quad (2.17)$$

De la ecuación (2.17) tenemos que:

$$s = Z_0 - a_0 \frac{1}{1 - a_1} \quad (2.18)$$

Así, juntando la ecuación (2.8) y ecuación (2.18) tenemos la expresión:

$$Z_t = a_0 \frac{1 - a_1^t}{1 - a_1} + a_1^t Z_0 \quad (2.19)$$

No debe pasar desapercibido que esta solución es la misma que la mostrada en la ecuación (2.14), por lo que en realidad ambas ecuaciones son una solución general indistintamente entre las ecuaciones (2.14) y (2.19). Ambas convergen a la expresión como la ecuación (2.15), con la misma condición de convergencia $|a_1| < 1$. Para ilustrar estas ecuaciones veamos algunos ejemplos al respecto.

Consideremos que tenemos un proceso Z_t que es descrito por una ecuación en diferencia lineal de primer orden dada por:

$$Z_t = 2 + 0.9Z_{t-1} \quad (2.20)$$

Siguiendo la expresión mostrada en la ecuación (2.19), obtenemos la expresión:

$$Z_t = 2 \left(\frac{1 - 0.9^t}{1 - 0.9} \right) + 0.9^t Z_0 \quad (2.21)$$

Donde asumiremos que el valor inicial es $Z_0 = 10$ y que la expresión debe converger al valor de 20, cuando t es muy grande o tiende a infinito. De forma similar tomemos otro ejemplo, en el cual asumimos la siguiente expresión:

$$Z_t = 2 - 0.5Z_{t-1} \quad (2.22)$$

Siguiendo la expresión mostrada en la ecuación (2.19), obtenemos:

$$Z_t = 2 \left(\frac{1 - (-0.5)^t}{1 + 0.5} \right) + (-0.5)^t Z_0 \quad (2.23)$$

Donde asumiremos que el valor inicial es $Z_0 = 10$ y que la ecuación converge al valor de 1.333333..., cuando t es muy grande o tiende a infinito. Ahora simulemos el comportamiento de ambos procesos y estableceremos los resultados del Cuadro 2.1. Notemos que el segundo proceso converge de una forma más rápida que el primero. El Cuadro 2.1 se ilustra en las siguientes dos Figura 2.1 y Figura 2.2.

```
library(knitr)
library(tidyverse)
library(kableExtra)

Tiempo = c(0:100)
Zt = rep(NA, 101)
Zt2 = rep(NA, 101)
```

```

Zt[1] = 10
Zt2[1] = 10

for (i in c(2:101)){
  Zt[i] = 2+0.9*Zt[i-1]
  Zt2[i] = 2 - 0.5*Zt2[i-1]
}
lista = c(1:16, 97:101)
Tiempo1 = Tiempo[lista]
Zt1=Zt[lista]
Zt21=Zt2[lista]

tabla1 = data.frame(Tiempo1, Zt1, Zt21)
colnames(tabla1) <- c("Tiempo", "$Z_t =2+0.9Z_{t-1}$",
                      "$Z_t = 2-0.5Z_{t-1}$")

kable(tabla1,
      caption = "Dos ejemplos de Ecuaciones Lineales de Primer Orden",
      format = "pandoc") %>%
  kable_styling(font_size = 10)

```

Table 2.1: Dos ejemplos de Ecuaciones Lineales de Primer Orden

| Tiempo | $Z_t = 2 + 0.9Z_{t-1}$ | $Z_t = 2 - 0.5Z_{t-1}$ |
|--------|------------------------|------------------------|
| 0 | 10.00000 | 10.000000 |
| 1 | 11.00000 | -3.000000 |
| 2 | 11.90000 | 3.500000 |
| 3 | 12.71000 | 0.250000 |
| 4 | 13.43900 | 1.875000 |
| 5 | 14.09510 | 1.062500 |
| 6 | 14.68559 | 1.468750 |
| 7 | 15.21703 | 1.265625 |
| 8 | 15.69533 | 1.367188 |
| 9 | 16.12580 | 1.316406 |
| 10 | 16.51322 | 1.341797 |
| 11 | 16.86189 | 1.329102 |
| 12 | 17.17570 | 1.335449 |
| 13 | 17.45813 | 1.332275 |
| 14 | 17.71232 | 1.333862 |
| 15 | 17.94109 | 1.333069 |
| 96 | 19.99960 | 1.333333 |
| 97 | 19.99964 | 1.333333 |
| 98 | 19.99967 | 1.333333 |
| 99 | 19.99970 | 1.333333 |

| Tiempo | $Z_t = 2 + 0.9Z_{t-1}$ | $Z_t = 2 - 0.5Z_{t-1}$ |
|--------|------------------------|------------------------|
| 100 | 19.99973 | 1.333333 |

```

tabla <- data.frame( Tiempo, Zt, Zt2 )

ggplot( data = tabla , aes(x = Tiempo, y = Zt) )+
  geom_line(col="blue4") +
  geom_point(col= "blue4") +
  labs(y=expression(Z[t]))

```

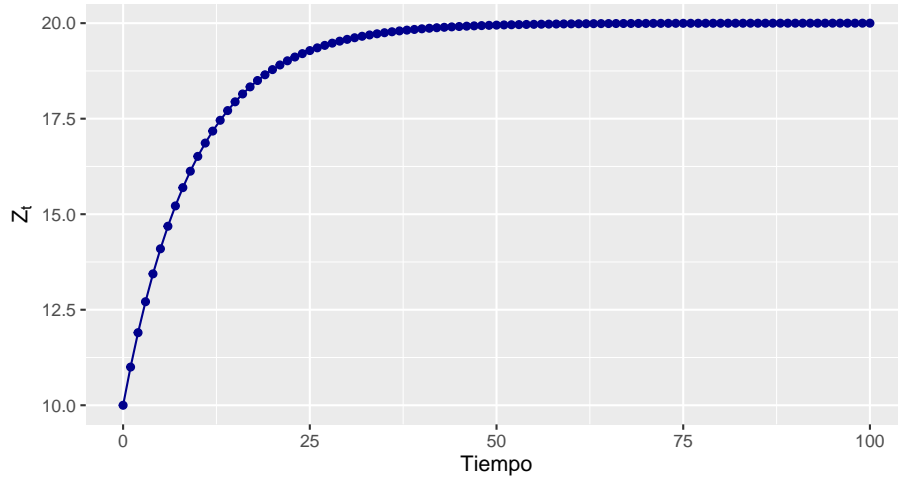


Figure 2.1: Evolución del proceso dado por $Z_t = 2 + 0.9Z_{t-1}$

```

ggplot(data = tabla , aes(x = Tiempo, y=Zt2)) +
  geom_line(col="red4") +
  geom_point(col= "red4") +
  labs(y=expression(Z[t]))

```

2.1.2 Ecuaciones en Diferencia Lineales de Segundo Orden y de orden superior

Como un segundo caso a estudiar se ubica el caso de las Ecuaciones en Diferencia Lineales de Segundo Orden y de orden superior. Primero, sea una ecuación como la siguiente, la cual es lineal y de segundo orden, ya que tiene asociado un término de Z_t rezagado dos períodos:

$$Z_t = a_0 + a_1 Z_{t-1} + a_2 Z_{t-2} \quad (2.24)$$

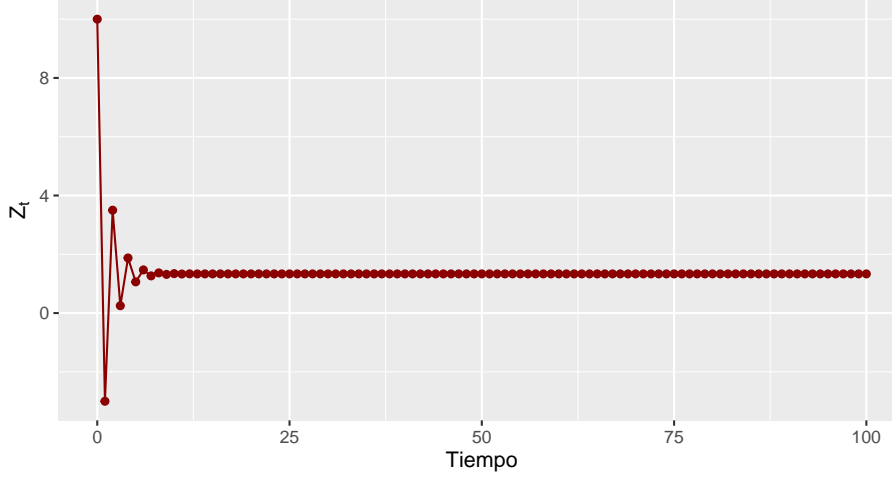


Figure 2.2: Evolución del proceso dado por $Z_t = 2 - 0.5Z_{t-1}$

Donde $t = \dots, -2, -1, 0, 1, 2, \dots$ y $a_1, a_2 \neq 0$. Reordenando la ecuación (2.24) podemos escribir:

$$\begin{aligned} Z_t - a_1 Z_{t-1} - a_2 Z_{t-2} &= a_0 \\ Z_t - a_1 L Z_t - a_2 L^2 Z_t &= a_0 \\ (1 - a_1 L - a_2 L^2) Z_t &= a_0 \end{aligned} \quad (2.25)$$

Así, la solución general propuesta para la ecuación (2.25) es la siguiente, la cual es una forma análoga a una Ecuación Lineal en Diferencia de Primer Orden:

$$Z_t = \frac{a_0}{1 - a_1 - a_2} + s_1 g_1^t + s_2 g_2^t \quad (2.26)$$

En donde s_1 y s_2 son constantes que se determinan mediante dos condiciones iniciales –por lo que para resolver este tipo de ecuaciones requerimos conocer dos condiciones iniciales–. Los valores de g_1 y g_2 están relacionados con los coeficientes a_1 y a_2 , de esta forma:

$$a_1 = g_1 + g_2 \quad (2.27)$$

$$a_2 = -g_1 g_2 \quad (2.28)$$

Lo anterior surge del siguiente procedimiento y recordando que siempre es posible descomponer una ecuación cuadrática en expresiones como las siguientes:

$$\begin{aligned} (1 - a_1 L - a_2 L^2) &= (1 - g_1 L)(1 - g_2 L) \\ &= 1 - g_1 L - g_2 L + g_1 g_2 L^2 \\ &= 1 - (g_1 + g_2)L + g_1 g_2 L^2 \end{aligned} \quad (2.29)$$

Donde se observa la equivalencia mostrada en las ecuaciones (2.27) y (2.28). Así, considerando la ecuación (2.26) tenemos que:

$$\begin{aligned} (1 - a_1 L - a_2 L^2) Z_t &= (1 - g_1 L)(1 - g_2 L) Z_t \\ &= a_0 + (1 - g_1 L)(1 - g_2 L) s_1 g_1^t \\ &\quad + (1 - g_1 L)(1 - g_2 L) s_2 g_2^t \end{aligned} \quad (2.30)$$

Por lo tanto, buscamos que para que el proceso sea equivalente y podamos interpretar que la ecuación (2.26) sea una solución general deberá pasar lo siguiente:

$$(1 - g_1 L)(1 - g_2 L) s_1 g_1^t + (1 - g_1 L)(1 - g_2 L) s_2 g_2^t = 0 \quad (2.31)$$

O, escrito de otra forma:

$$(1 - g_1 L) s_1 g_1^t = (1 - g_2 L) s_2 g_2^t = 0 \quad (2.32)$$

Ahora determinemos cuáles son los valores g_1 y g_2 dados los valores a_1 y a_2 que nos permitan determinar si el proceso será convergente. Para ello debemos resolver la siguiente ecuación que se deriva de la ecuación (2.29):

$$1 - a_1 x - a_2 x^2 = (1 - g_1 x)(1 - g_2 x) = 0 \quad (2.33)$$

Donde, claramente existen dos raíces: $x_1 = g_1^{-1}$ y $x_2 = g_2^{-1}$. Así, la solución estará dada por las raíces de la ecuación característica:

$$\begin{aligned} 1 - a_1 x - a_2 x^2 &= 0 \\ a_2 x^2 + a_1 x - 1 &= 0 \end{aligned} \quad (2.34)$$

Cuya solución es:

$$x = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2} \quad (2.35)$$

Es importante distinguir tres diferentes casos en relación con las raíces que surgen como solución de la ecuación (2.34), estos son:

Caso I. Si $a_1^2 + 4a_2 > 0$, la ecuación (2.34) proporcionará dos valores de raíces reales y distintos, eso es $x_1 = g_1^{-1} \neq x_2 = g_2^{-1}$. Si por ahora suponemos que $|g_1| < 1$ y que $|g_2| < 1$, entonces tendremos que:

$$\begin{aligned} (1 - g_1 L)^{-1} (1 - g_2 L)^{-1} a_0 &= \left(\sum_{j=0}^{\infty} g_1^j L^j \right) \left(\sum_{j=0}^{\infty} g_2^j L^j \right) a_0 \\ &= \left(\sum_{j=0}^{\infty} g_1^j \right) \left(\sum_{j=0}^{\infty} g_2^j \right) a_0 \\ &= \frac{a_0}{(1 - g_1)(1 - g_2)} \\ &= \frac{a_0}{1 - a_1 - a_2} \end{aligned} \quad (2.36)$$

Esto último es el punto de equilibrio de la ecuación (2.26); considerando que $|g_1| < 1$ y que $|g_2| < 1$ –notemos que los demás casos son divergentes, ya que la suma anterior no convergería–. De esta forma la solución de la ecuación estará dada por:

$$\lim_{t \rightarrow \infty} Z_t = \frac{a_0}{1 - a_1 - a_2} \quad (2.37)$$

Caso II. Si $a_1^2 + 4a_2 < 0$ en la ecuación (2.34), entonces las raíces serán números complejos conjugados, es decir:

$$g_i^{-1} = a \pm ib \quad (2.38)$$

$$g_i = u \pm iv \quad (2.39)$$

Dichas raíces las podemos escribir en coordenadas polares como:

$$g_1^{-1} = re^{i\theta} = r(\cos(\theta) + i\sin(\theta)) \quad (2.40)$$

$$g_2^{-1} = re^{-i\theta} = r(\cos(\theta) - i\sin(\theta)) \quad (2.41)$$

Donde: $r = \sqrt{u^2 + v^2}$, a esta expresión también se le conoce como modulo. Alternativamente, podemos escribir que $r = \sqrt{g_1 g_2}$. La única condición es que $r < 1$ para que el proceso descrito en la ecuación (2.26) sea convergente.

Al igual que en el **Caso I**, el punto de equilibrio de la ecuación se debería ubicar al rededor (2.37), siempre que $r < 1$, por lo que el factor que determina la convergencia es el modulo, ya que si el modulo es mayor a 1, el proceso será divergente, pero si es menor a 1 convergerá a (2.37). Para ilustrar, el caso contrario es divergente puesto que representa trayectorias senoidales (oscilatorias) que sólo pueden converger si a medida que pasa el tiempo, las ondas son menos amplias.

Caso III. Ahora revisemos el caso en el que $a_1^2 + 4a_2 = 0$, de esta forma las raíces serán idénticas:

$$g = g_1^{-1} = g_2^{-1} = \frac{-a_1}{2a_2} \quad (2.42)$$

Así, el punto de equilibrio será dado por la solución descrita como:

$$\begin{aligned} (1 - gL)^2 Z_t &= a_0 \\ Z_t &= \frac{a_0}{(1 - gL)^2} + s_1 g^t + s_2 t g^t \\ &= a_0 \sum_{i=0}^{\infty} (1 + i) g^i + s_1 g^t + s_2 t g^t \end{aligned} \quad (2.43)$$

Donde la expresión anterior es resultado de considerar el siguiente procedimiento. Sea:

$$f(g) = \frac{1}{(1-g)} = \sum_{j=0}^{\infty} g^j$$

Por lo que si hacemos la primer derivada de la expresión anterior tenemos que:

$$\begin{aligned} f'(g) &= \frac{1}{(1-g)^2} \\ &= \sum_{j=0}^{\infty} jg^{j-1} \\ &= 0 + g^0 + 2g^1 + 3g^2 + \dots \\ &= \sum_{j=0}^{\infty} (1+j)g^j \end{aligned}$$

Ahora veamos un ejemplo de una Ecuación Lineal en Diferencia de Segundo Orden. Supongamos la ecuación y el desarrollo siguientes:

$$\begin{aligned} Z_t &= 3 + 0.9Z_{t-1} - 0.2Z_{t-2} \\ (1 - 0.9L + 0.2L^2)Z_t &= 3 \end{aligned}$$

La solución dada por una ecuación similar a la expresión (2.34), obtendríamos la solución dada por las ecuaciones equivalentes a:

$$\begin{aligned} 1 - 0.9x + 0.2x^2 &= 0 \\ -0.2x^2 + 0.9x - 1 &= 0 \end{aligned}$$

De donde las raíces del polinomio característico $x_1 = g_1^{-1}$ y $x_2 = g_2^{-1}$ se obtienen de la expresión dada por:

$$\begin{aligned} x &= \frac{-0.9 \pm \sqrt{0.81 + (4)(-0.2)}}{(2)(-0.2)} \\ &= \frac{0.9 \pm 0.1}{0.4} \end{aligned}$$

Dado que el componente $a_1^2 + 4a_2$ es positivo, obtendremos dos raíces reales. Las raíces estarán dadas por $x_1 = 2.5$ y $x_2 = 2.0$, de lo cual podemos determinar que $g_1 = 0.4$ y $g_2 = 0.5$. De esta forma tenemos que $|g_1| < 1$ y $|g_2| < 1$, así la ecuación converge a la expresión dada por las siguientes expresiones:

$$\begin{aligned} Z_t &= \frac{3}{1 - 0.9L + 0.2L^2} + s_1(0.4)^t + s_2(0.5)^t \\ &= \frac{3}{1 - 0.9 + 0.2} + s_1(0.4)^t + s_2(0.5)^t \\ &= \frac{3}{(1 - 0.4)(1 - 0.5)} + s_1(0.4)^t + s_2(0.5)^t \end{aligned}$$

Al final, la ecuación que describe la solución general será:

$$z_t = 10 + s_1(0.4)^t + s_2(0.5)^t \quad (2.44)$$

Para determinar los valores de s_1 y s_2 necesitamos obtener dos valores iniciales de la ecuación para lo cual iniciaremos como $t = 0$ y luego obtenemos el valor de $t = 1$, consideremos el valor de $Z_0 = 0$ y $Z_1 = 50$:

$$\begin{aligned} Z_0 &= 10 + s_1(0.4)^0 + s_2(0.5)^0 \\ 0 &= 10 + s_1 + s_2 \\ Z_1 &= 10 + s_1(0.4)^1 + s_2(0.5)^1 \\ 50 &= 10 + 0.4s_1 + 0.5s_2 \end{aligned}$$

Por lo que la solución es: $s_1 = -450$ y $s_2 = 440$, de donde podemos expresar la ecuación como:

$$Z_t = 10 - 450(0.4)^t + 440(0.5)^t \quad (2.45)$$

La ecuación (2.45) anterior convergerá al valor de 10 cuando $t \rightarrow \infty$. Para ilustrar la trayectoria de esta ecuación tomemos un cuadro similar al de los ejemplos anteriores. En el Cuadro 2.2 y la Figura 2.3 mostramos los resultados de la trayectoria para 100 periodos.

Finalmente, discutiremos la solución para las Ecuaciones Lineales en Diferencia de Orden p , donde $p \geq 2$. En general una ecuación de este tipo se puede escribir como:

$$Z_t = a_0 + a_1 Z_{t-1} + a_2 Z_{t-2} + \dots + a_p Z_{t-p} \quad (2.46)$$

Donde $t = \dots, -2, -1, 0, 1, 2, \dots$ y $a_p \neq 0$. La ecuación (2.46) se puede escribir como:

$$\begin{aligned} Z_t - a_1 Z_{t-1} - a_2 Z_{t-2} - \dots - a_p Z_{t-p} &= a_0 \\ Z_t - a_1 L Z_t - a_2 L^2 Z_t - \dots - a_p L^p Z_t &= a_0 \\ (1 - a_1 L - a_2 L^2 - \dots - a_p L^p) Z_t &= a_0 \end{aligned} \quad (2.47)$$

```
t = ts(c(0:100))

Zt = 10-450*(0.4^t)+440*(0.5^t)

tabla_2 <- data.frame( t, Zt)

lista = c(1:16, 97:101)
t1 = t[lista]
Zt1=Zt[lista]
```

```

tabla1 = data.frame(Tiempo1, Zt1)
colnames(tabla1) <- c("Tiempo", "$Z_t = 10-450(0.4)^t+440(0.5)^t$")

kable(tabla1,
      caption = "Un ejemplo de Ecuación de Segundo Orden",
      format = "pandoc")%>%
kable_styling(font_size = 10)

```

Table 2.2: Un ejemplo de Ecuación de Segundo Orden

| Tiempo | $Z_t = 10 - 450(0.4)^t + 440(0.5)^t$ |
|--------|--------------------------------------|
| 0 | 0.00000 |
| 1 | 50.00000 |
| 2 | 48.00000 |
| 3 | 36.20000 |
| 4 | 25.98000 |
| 5 | 19.14200 |
| 6 | 15.03180 |
| 7 | 12.70022 |
| 8 | 11.42384 |
| 9 | 10.74141 |
| 10 | 10.38250 |
| 11 | 10.19597 |
| 12 | 10.09987 |
| 13 | 10.05069 |
| 14 | 10.02565 |
| 15 | 10.01294 |
| 96 | 10.00000 |
| 97 | 10.00000 |
| 98 | 10.00000 |
| 99 | 10.00000 |
| 100 | 10.00000 |

```

ggplot(data = tabla_2, aes(x = t, y=Zt)) +
  geom_line(col="green4") +
  geom_point(col= "green4") +
  labs(y=expression(Z[t]))

```

Por el Teorema Fundamental del Álgebra es posible escribir a la ecuación (2.47) como:

$$(1 - g_1 L)(1 - g_1 L) \dots (1 - g_p L) Z_t = a_0 \quad (2.48)$$

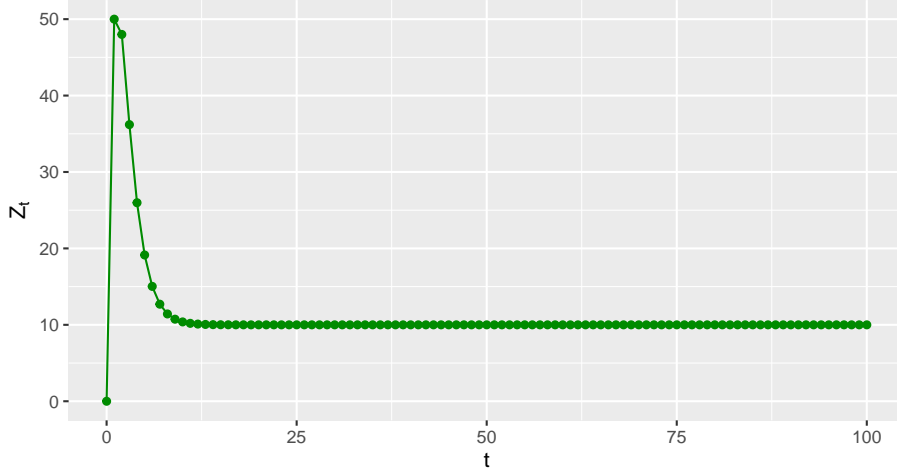


Figure 2.3: Evolución del proceso dado por $Z_t = 3 + 0.9Z_{t-1} - 0.2Z_{t-2}$

Utilizando la ecuación (2.47) y la ecuación (2.48) tenemos que la solución general de una ecuación como la descrita en (2.46) se puede escribir como:

$$Z_t = \frac{a_0}{1 - a_1 - a_2 - \dots - a_p} + s_1 g_1^t + s_2 g_2^t + \dots + s_p g_p^t \quad (2.49)$$

$$Z_t = \frac{a_0}{(1 - g_1)(1 - g_1) \dots (1 - g_p)} + s_1 g_1^t + s_2 g_2^t + \dots + s_p g_p^t \quad (2.50)$$

Donde s_1, s_2, \dots, s_p son constantes que se determinan utilizando p valores particulares de Z_t , y la solución general descrita en las ecuaciones (2.49) y (2.50) implica encontrar p raíces: $x_1 = g_1^{-1}, x_2 = g_2^{-1}, \dots, x_p = g_p^{-1}$ de los siguientes polinomios equivalentes:

$$(1 - g_1)(1 - g_1) \dots (1 - g_p) = 0 \quad (2.51)$$

$$1 - a_1 x - a_2 x^2 - \dots - a_p x^p = 0 \quad (2.52)$$

$$a_p x^p + \dots + a_2 x^2 + a_1 x - 1 = 0 \quad (2.53)$$

Antes de plantear la solución general, analicemos una solución particular cuando un conjunto de las p raíces, digamos un total de m , son iguales, es decir, cuando

sucede que $g_1 = g_2 = \dots = g_m = g$ (con $1 < m \leq p$). En este caso la solución general en la ecuación (2.50) se escribe como:

$$Z_t = \frac{a_0}{(1-g)^m(1-g_{m+1}) \dots (1-g_p)} + s_1 g^t + s_2 t g^t + \dots + s_m t^{m-1} g^t + s_{m+1} g_{m+1}^t + \dots + s_p g_p^t \quad (2.54)$$

Definamos:

$$f(g) = \frac{1}{1-g} = \sum_{j=0}^{\infty} g^j \quad (2.55)$$

Si retomamos el método descrito párrafos arriba tenemos las siguientes expresiones. Cuando $m = 2$:

$$f'(g) = \frac{1}{(1-g)^2} = \sum_{j=0}^{\infty} j g^{j-1} = \sum_{j=0}^{\infty} (1+j) g^j$$

En el otro extremo, cuando $m = p$:

$$f^{(p-1)}(g) = \frac{p-1}{(1-g)^p} = \sum_{j=0}^{\infty} \frac{(p-1+j)(p-2+j) \dots (2+j)(1+j)}{(p-1)!} g^j \quad (2.56)$$

Así, en el extremo cuando $m = p$ la solución general podría estar dada por:

$$Z_t = a_0 \sum_{j=0}^{\infty} \frac{(p-1+j)(p-2+j) \dots (2+j)(1+j)}{(p-1)!} g^j + g^t \sum_{i=0}^p s_i t^{i-1} \quad (2.57)$$

Donde $|g| < 1$, $t = \dots, -2, -1, 0, 1, 2, \dots$. Para finalizar esta sección, plantearemos la expresión de polinomio característico que nos permitirá hacer el análisis de convergencia de los procesos. Partamos de que la ecuación (2.53) se puede escribir como:

$$(x^{-1})^p - a_1(x^{-1})^{p-1} - a_2(x^{-1})^{p-1} - \dots - a_p = 0 \quad (2.58)$$

La ecuación (2.58) permite interpretar las raíces del polinomio característico de forma directa ya que $x_1^{-1} = g_1$, $x_2^{-1} = g_2$, ..., $x_p^{-1} = g_p$. Así, siempre que $p \geq 1$ en la ecuación (2.46), diremos que el proceso descrito en esa ecuación dará como resultado un proceso convergente si se cumplen las dos condiciones (2.59) y (2.60):

$$|a_p| < 1 \quad (2.59)$$

$$a_1 + a_2 + \dots + a_p < 1 \quad (2.60)$$

Alternativamente, cuando las raíces son reales lo anterior es equivalente a la expresión (2.61):

$$|g_i| < 1 \quad (2.61)$$

Para $\forall i = 1, 2, \dots, p$. Cuando las raíces son imaginarias, las dos condiciones (2.59) y (2.60) son equivalentes a la expresión (2.62):

$$\sqrt{g_i g_j} = \sqrt{u^2 + v^2} < 1 \quad (2.62)$$

Para $\forall i \neq j$ y $i, j = 1, 2, \dots, p$. Cuando $g_1 = g_2 = \dots = g_p = g$, la condición de la ecuación (2.61) se resume a que $|g| < 1$. En resumen, las condiciones descritas en las ecuaciones (2.61) y (2.62) se pueden ilustrar con un círculo unitario como el de la Figura 2.4 en que si las raíces se ubican dentro de éste, podemos decir que el proceso es convergente en el largo plazo.

```
# Crear datos para el círculo
theta <- seq(0, 2*pi, length.out=100)
x <- cos(theta)
y <- sin(theta)
circle_data <- data.frame(x, y)

# Dibujar el círculo
ggplot(circle_data, aes(x=x, y=y)) +
  geom_polygon(fill=NA, color="black") +
  coord_fixed(ratio=1) +
  xlim(-1.5, 1.5) +
  ylim(-1.5, 1.5)
```

2.2 Operador de rezago L

Denotemos, como se ha mencionado con anterioridad, con L al operador de rezago, el cual nos permitirá construir una relación entre diferencias y medias móviles como se verá más adelante en los procesos univariados $AR(p)$, $MA(q)$ y, en general, $ARIMA(p, d, q)$. Sean X , Y o Z variables con las que denotaremos a una serie de tiempo (note que hasta el momento no hemos definido qué es una serie de tiempo, no obstante no es necesario definirla para hacer uso del operador).

En esta sección resumiremos algunas propiedades usadas en el capítulo y en capítulos más adelante. Así, si a dicha serie le aplicamos el operador rezago antes definido, el resultado deberá ser que cada uno de los valores de la serie es retardado o regresado un período. Es decir:

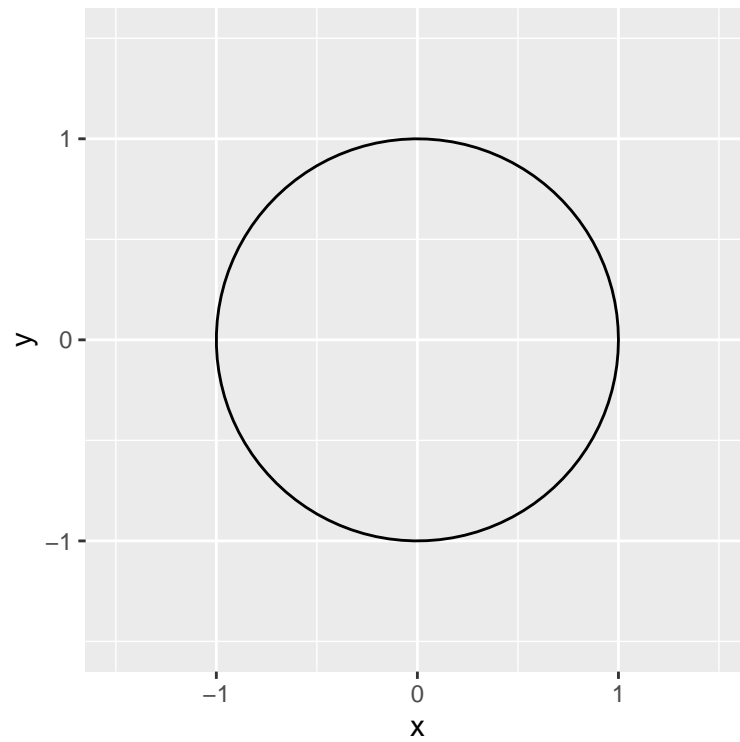


Figure 2.4: Circulo unitario en el que se cumple que $|g_i| < 1$ y $(g_i g_j)^{1/2} = (u^2 + v^2)^{1/2} < 1$

$$LZ_t = Z_{t-1} \quad (2.63)$$

De esta forma, si aplicamos el operador rezago L a la nueva serie de tiempo dada por Z_{t-1} podemos obtener Z_{t-2} , haciendo uso de la ecuación (2.63) podemos obtener:

$$LZ_{t-1} = L(LZ_t) = L^2Z_t = Z_{t-2} \quad (2.64)$$

Mediante una generalización podemos obtener:

$$L^k Z_t = Z_{t-k} \quad (2.65)$$

Para $k = \dots, -2, -1, 0, 1, 2, \dots$. Así, para $k = 0$ obtenemos la identidad dado que $L^0 Z_t = Z_t$, de tal forma que siempre asumiremos que $L^0 = 1$. En otro caso, cuando $k > 0$ a la serie de tiempo a la cual se le aplique el operador rezago L se le deberá aplicar un rezago de k periodos a cada uno de los elementos de la serie. Por el contrario, cuando $k < 0$ el operador rezago significa que se deberá adelantar $|k|$ veces a cada elemento de la serie. Por ejemplo, $L^{-3}Z_t = Z_{t+3}$.

Las reglas descritas en lo subsecuente se mantienen indistintamente cuando aplican para el caso de rezagar como para cuando se adelanta una serie. Como primera propiedad tomemos a la siguiente propiedad:

$$L^m Z_{t-n} = L^m (L^n Z_t) = L^{m+n} Z_t = Z_{t-(n+m)} \quad (2.66)$$

De lo anterior podemos inferir el siguiente resultado:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - L)Z_t \quad (2.67)$$

En el caso de la diferencia de orden cuatro o cuarta diferencia se puede expresar como:

$$\Delta_4 Z_t = Z_t - Z_{t-4} = (1 - L^4)Z_t \quad (2.68)$$

Al respecto, vale la pena aclarar que en ocasiones se hará uso de una notación alternativa dada por: Δ^k o Δ_k , donde $k = 1, 2, 3, \dots$, indistintamente, ya que en ambos casos se referirá a una diferencia de orden k . Esta notación resulta de gran utilidad cuando se quiere comparar periodos equivalentes como, por ejemplo, el mismo trimestre pero de un año anterior. De forma similar, para el caso de logaritmos podemos escribir a la ecuación (2.68) como:

$$\Delta^4 \ln(Z_t) = \Delta_4 \ln(Z_t) = \ln(Z_t) - \ln(Z_{t-4}) = (1 - L^4) \ln(Z_t) \quad (2.69)$$

Para el caso de una serie de tiempo que se le ha transformado mediante medias móviles, digamos de 4 periodos, podemos escribirla como:

$$Zs_t = \frac{1}{4}(Z_t + Z_{t-1} + Z_{t-2} + Z_{t-3}) = \frac{1}{4}(1 + L + L^2 + L^3)Z_t \quad (2.70)$$

Una generalización del anterior caso puede ser escrito como un polinomio de orden p con el operador rezago L dado como:

$$\begin{aligned}\alpha(L)Z_t &= (1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p)Z_t \\ &= Z_t - \alpha_1 Z_{t-1} - \alpha_2 Z_{t-2} - \dots - \alpha_p Z_{t-p}\end{aligned}\quad (2.71)$$

Donde α_i puede ser remplazada por cualquier constante a_i , con $i = 1, 2, 3, \dots$, para escribir ecuaciones como las anteriores. Adicionalmente, podemos decir que la ecuación (2.71) es una generalización del caso de medias móviles, el cual admite una ponderación distinta para cada uno de los elementos rezagados.

Existe la posibilidad de operar más de un polinomio a la vez. Para múltiples polinomios (digamos, los polinomios $\alpha(L)$ y $\beta(L)$) podemos escribir el siguiente resultado:

$$\alpha(L)\beta(L) = \beta(L)\alpha(L) \quad (2.72)$$

Tales polinomios del operador rezago también son llamados *filtros lineales*. A manera de ejemplo tomemos el siguiente caso de diferencias para una serie de Z_t :

$$\Delta Z_t = (1 - L)Z_t = Z_t - Z_{t-1} \quad (2.73)$$

y un proceso de medias móviles para la misma serie de Z_t :

$$Zs_t = \frac{1}{4}(1 + L^1 + L^2 + L^3)Z_t = \frac{1}{4}(Z_t + Z_{t-1} + Z_{t-2} + Z_{t-3}) \quad (2.74)$$

De tal forma que el producto de ambos procesos se puede escribir como:

$$(1 - L) \times \frac{1}{4}(1 + L^1 + L^2 + L^3)Z_t = \frac{1}{4}(1 - L^4)Z_t \quad (2.75)$$

Es decir, que el producto de dos polinomios, uno de diferencias y otro más de medias móviles, resulta en uno de diferencias pero de mayor grado, en este caso de grado 4.

Chapter 3

Modelos de Series de Tiempo Estacionarias

3.1 Definición de ergodicidad y estacionariedad

A partir de esta sección introduciremos mayor formalidad matemática al análisis de las series de tiempo. Por ello cambiaremos un poco la notación y ocuparemos a X_t en lugar de Z_t como objeto de nuestro análisis. Con X_t denotaremos a una serie de tiempo, ya que con Z_t denotaremos a una variable, sin que ella fuera necesariamente una serie de tiempo en los términos que a continuación discutimos. Asimismo, iniciaremos por establecer una serie de definiciones.

De esta forma, definiremos a una *serie de tiempo* como un vector de variables aleatorias de dimensión T , dado como:

$$X_1, X_2, X_3, \dots, X_T \quad (3.1)$$

Cada una de las X_t ($t = 1, 2, \dots, T$) consideradas como una variable aleatoria. Así, también podemos denotar a la serie de tiempo como:

$$\{X_t\}_{t=1}^T \quad (3.2)$$

Es decir, definiremos a una *serie de tiempo como una realización de un proceso estocástico* –o un Proceso Generador de Datos (PGD). Consideremos una muestra de los múltiples posibles resultados de muestras de tamaño T , la colección dada por:

$$\{X_1^{(1)}, X_2^{(1)}, \dots, X_T^{(1)}\} \quad (3.3)$$

Digamos que la ecuación (3.3) es una de las tantas posibles resultantes del proceso estocástico o PGD. Eventualmente podríamos estar dispuestos a observar

este proceso indefinidamente, de forma tal que estemos interesados en observar a la secuencia dada por $\{X_t^{(1)}\}_{t=1}^{\infty}$, lo cual no dejaría de ser sólo una de las tantas realizaciones o secuencias del proceso estocástico original.

Tan solo por poner un ejemplo, podríamos observar las siguientes realizaciones del mismo PGD:

$$\begin{aligned} &\{X_1^{(2)}, X_2^{(2)}, \dots, X_T^{(2)}\} \\ &\{X_1^{(3)}, X_2^{(3)}, \dots, X_T^{(3)}\} \\ &\{X_1^{(4)}, X_2^{(4)}, \dots, X_T^{(4)}\} \\ &\vdots \\ &\{X_1^{(j)}, X_2^{(j)}, \dots, X_T^{(j)}\} \end{aligned}$$

Donde $j \in \mathbb{Z}$. En lo subsecuente, diremos que una serie de tiempo es una realización del proceso estocástico subyacente. Considerando, en consecuencia, al proceso estocástico con todas sus posibilidades de realización.

Para hacer más sencilla la notación no distinguiremos entre el proceso en sí mismo y una de sus realizaciones, es decir, siempre escribiremos a una serie de tiempo como la secuencia mostrada en la ecuación (3.2), o más precisamente como la siguiente realización:

$$\{X_1, X_2, \dots, X_T\} \quad (3.4)$$

O simplemente:

$$X_1, X_2, \dots, X_T \quad (3.5)$$

El proceso estocástico de dimensión T puede ser completamente descrito por su función de distribución multivariada de dimensión T . No obstante, esto no resulta ser práctico cuando se opere más adelante en el curso. Por ello, en el curso, y en general casi todos los textos lo hacen, sólo nos enfocaremos en sus primer y segundo momentos, es decir, en sus medias o valores esperados:

$$\mathbb{E}[X_t]$$

Para $t = 1, 2, \dots, T$; o:

$$\begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_T] \end{bmatrix}$$

o,

$$[\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_T]]$$

De sus varianzas:

$$Var[X_t] = \mathbb{E}[(X_t - \mathbb{E}[X_t])^2]$$

Para $t = 1, 2, \dots, T$, y de sus $T(T-1)/2$ covarianzas:

$$\text{Cov}[X_t, X_s] = \mathbb{E}[(X_t - \mathbb{E}[X_t])(X_s - \mathbb{E}[X_s])]$$

Para $t < s$. Por lo tanto, en la forma matricial podemos escribir lo siguiente:

$$\begin{aligned} & \begin{bmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \cdots & \text{Cov}[X_1, X_T] \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \cdots & \text{Cov}[X_2, X_T] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[X_T, X_1] & \text{Cov}[X_T, X_2] & \cdots & \text{Var}[X_T] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \rho_{12} & \cdots & \rho_{1T} \\ \rho_{21} & \sigma_2^2 & \cdots & \rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T1} & \rho_{T2} & \cdots & \sigma_T^2 \end{bmatrix} \end{aligned} \quad (3.6)$$

Donde es claro que en la matriz de la ecuación (3.6) existen $T(T-1)/2$ covarianzas distintas, ya que se cumple que $\text{Cov}[X_t, X_s] = \text{Cov}[X_s, X_t]$, para $t \neq s$.

A menudo, esas covarianzas son denominadas como autocovarianzas puesto que ellas son covarianzas entre variables aleatorias pertenecientes al mismo proceso estocástico pero en un momento t diferente. Si el proceso estocástico tiene una distribución normal multivariada, su función de distribución estará totalmente descrita por sus momentos de primer y segundo orden.

Ahora introduciremos el concepto de ergodicidad, el cual indica que los momentos muestrales, los cuales son calculados en la base de una serie de tiempo con un número finito de observaciones, en la medida que $T \rightarrow \infty$ sus correspondientes momentos muestrales, tienden a los verdaderos valores poblacionales, los cuales definiremos como μ , para la media, y σ_X^2 para la varianza.

Este concepto sólo es cierto si asumimos que, por ejemplo, el valor esperado y la varianza son como se dice a continuación para todo $t = 1, 2, \dots, T$:

$$\mathbb{E}[X_t] = \mu_t = \mu \quad (3.7)$$

$$(3.8)$$

$$\text{Var}[X_t] = \sigma_X^2 \quad (3.9)$$

Mas formalmente, se dice que el PGD o el proceso estocástico es ergódico en la media si:

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=1}^T (X_t - \mu) \right)^2 \right] = 0 \quad (3.10)$$

y ergódico en la varianza si:

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=1}^T (X_t - \mu)^2 - \sigma_X^2 \right)^2 \right] = 0 \quad (3.11)$$

Estas condiciones se les conoce como *propiedades de consistencia* para las variables aleatorias. Sin embargo, éstas no pueden ser probadas. Por ello se les denomina como un supuesto que pueden cumplir algunas de las series. Más importante aún: **un proceso estocástico que tiende a estar en equilibrio estadístico en un orden ergódico, es estacionario.**

Podemos distinguir dos tipos de estacionariedad. Si asumimos que la función común de distribución del proceso estocástico no cambia a lo largo del tiempo, se dice que el proceso es *estrictamente estacionario*. Como este concepto es difícil de aplicar en la práctica, solo consideraremos a la *estacionariedad débil* o estacionariedad en sus momentos.

Definiremos a la estacionariedad por sus momentos del correspondiente proceso estocástico dado por $\{X_t\}$:

1. *Estacionariedad en media:* Un proceso estocástico es estacionario en media si $E[X_t] = \mu_t = \mu$ es constante para todo t .
2. *Estacionariedad en varianza:* Un proceso estocástico es estacionario en varianza si $Var[X_t] = E[(X_t - \mu_t)^2] = \sigma_X^2 = \gamma(0)$ es constante y finita para todo t .
3. *Estacionariedad en covarianza:* Un proceso estocástico es estacionario en covarianza si $Cov[X_t, X_s] = E[(X_t - \mu_t)(X_s - \mu_s)] = \gamma(|s - t|)$ es sólo una función del tiempo y de la distancia entre las dos variables aleatorias. Por lo que no depende del tiempo denotado por t (no depende de la información contemporánea).
4. *Estacionariedad débil:* Como la estacionariedad en varianza resulta de forma inmediata de la estacionariedad en covarianza cuando se asume que $s = t$, un proceso estocástico es débilmente estacionario cuando es estacionario en media y covarianza.

Puesto que resulta poco factible asumir una estacionariedad diferente a la débil, es adelante siempre que digamos que un proceso es estacionario se referirá al caso débil y sólo diremos que el proceso es estacionario, sin el apelativo de débil.

Ahora veamos un ejemplo de lo anterior. Supongamos una serie de tiempo denotada por: $\{U_t\}_{t=0}^T$. Decimos que el proceso estocástico $\{U_t\}$ es un *proceso estocástico puramente aleatorio* o es un *proceso estocástico de ruido blanco o caminata aleatoria*, si éste tiene las siguientes propiedades:

1. $E[U_t] = 0, \forall t$;
2. $Var[U_t] = E[(U_t - \mu_t)^2] = E[(U_t - \mu)^2] = E[(U_t)^2] = \sigma^2, \forall t$, y
3. $Cov[U_t, U_s] = E[(U_t - \mu_t)(U_s - \mu_s)] = E[(U_t - \mu)(U_s - \mu)] = E[U_t U_s] = 0, \forall t \neq s$.

En otras palabras, un proceso U_t es un ruido blanco si su valor esperado (promedio) es cero (0), tiene una varianza finita y constante, y además no le importa la historia pasada. Así, su valor presente no se ve influenciado por sus valores pasados no importando respecto de que periodo se tome referencia.

En apariencia, por sus propiedades, este proceso es débilmente estacionario –o simplemente, estacionario–. Todas las variables aleatorias tienen una media de cero, una varianza σ^2 y no existe correlación entre ellas.

Propongamos un ejemplo para ilustrar la teoría previa. Supongamos que definimos un nuevo proceso estocástico $\{X_t\}$ como:

$$X_t = \begin{cases} U_0 & \text{para } t = 0 \\ X_{t-1} + U_t & \text{para } t = 1, 2, 3, \dots \end{cases} \quad (3.12)$$

Donde $\{U_t\}$ es un proceso puramente aleatorio. Este proceso estocástico, o caminata aleatoria sin tendencia (ajuste - drift), puede ser reescrito como:

$$X_t = \sum_{j=0}^t U_j \quad (3.13)$$

Tratemos de dar más claridad al ejemplo, para ello asumamos que generamos a $\{U_t\}$ por medio del lanzamiento de una moneda. Donde obtenemos una cara (águila) con una probabilidad de 0.5, en cuyo caso decimos que la variable aleatoria U_t tomará el valor de +1, y una cruz (sol) con una probabilidad de 0.5, en cuyo caso decimos que la variable aleatoria U_t toma el valor de -1.

Este planteamiento cumple con las propiedades enunciadas ya que:

1. $\mathbb{E}[U_t] = 0.5 \times -1 + 0.5 \times 1 = 0, \forall t$
2. $Var[U_t] = \mathbb{E}[(U_t - 0)^2] = \frac{1}{2}((-1)^2) + \frac{1}{2}((1)^2) = 1, \forall t$
3. $Cov[U_t, U_s] = \mathbb{E}[(U_t - 0)(U_s - 0)] = \mathbb{E}[U_t \cdot U_s] = 0, \forall t \neq s.$

Retomando a nuestro proceso X_t , diremos que el caso de $X_0 = 0$, para $t = 0$. Si verificamos cuáles son sus primeros y segundos momentos de $\{X_t\}$ tenemos:

$$\mathbb{E}[X_t] = \mathbb{E}\left[\sum_{j=1}^t U_j\right] = \sum_{j=1}^t \mathbb{E}[U_j] = 0 \quad (3.14)$$

En cuanto a la varianza:

$$\begin{aligned}
 Var[X_t] &= Var\left[\sum_{j=1}^t U_j\right] \\
 &= \sum_{j=1}^t Var[U_j] + 2 * \sum_{j \neq k} Cov[U_j, U_k] \\
 &= \sum_{j=1}^t 1 \\
 &= t
 \end{aligned} \tag{3.15}$$

Lo anterior, dado que hemos supuesto que en la caminata aleatoria todas las variables aleatorias son independientes, es decir, $Cov[U_t, U_s] = E[U_t \cdot U_s] = 0$. Por su parte, la covarianza del proceso estocástico se puede ver como:

$$\begin{aligned}
 Cov[X_t, X_s] &= \mathbb{E}\left[\left(\sum_{j=1}^t U_j - 0\right)\left(\sum_{i=1}^s U_i - 0\right)\right] \\
 &= \mathbb{E}[(U_1 + U_2 + \dots + U_t)(U_1 + U_2 + \dots + U_s)] \\
 &= \sum_{j=1}^t \sum_{i=1}^s \mathbb{E}[U_j U_i] \\
 &= \mathbb{E}[U_1^2] + \mathbb{E}[U_2^2] + \dots + \mathbb{E}[U_k^2] \\
 &= \sigma^2 + \sigma^2 + \dots + \sigma^2 \\
 &= 1 + 1 + 1 + 1 \\
 &= \min(t, s)
 \end{aligned}$$

```

set.seed(1234)
# Utilizaremos una función guardada en un archivo a parte
# Llamamos a la función:
source("Caminata.R")

# Definimos argumentos de la función
Opciones <- c(-1, 1)
#
Soporte <- 10000

# Vamos a replicar el proceso con estos parámetros
Rango <- 200
#
Caminos <- 10

#

```

```

for(i in 1:Camino){
  TT <- data.matrix(data.frame(Caminata(Opciones, Soporte)[1]))
  #
  G_t <- data.matrix(data.frame(Caminata(Opciones, Soporte)[2]))
  #
  plot(TT, G_t, col = "blue", type = "l", ylab = "Ganancias",
        xlab = "Tiempo", ylim = c(-Rango,Rango))
  #
  par(new = TRUE)
  #
  i <- i +1
}
#
par(new = FALSE)

```

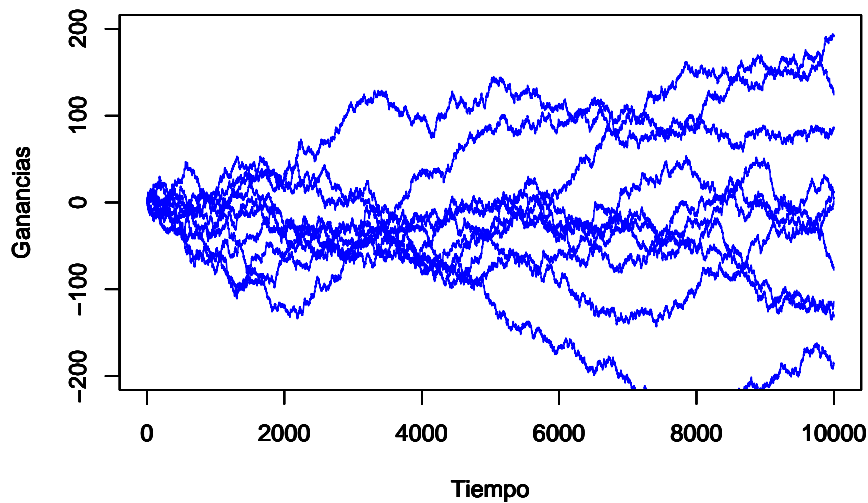


Figure 3.1: Ejemplo de 10 trayectorias de la caminata aleatoria, cuando sólo es posible cambios de +1 y -1

Así, el proceso estocástico dado por la caminata aleatoria sin un término de ajuste es estacionario en media, pero no en varianza o en covarianza, y consecuentemente, en general no estacionario, condición que contraria al caso del proceso simple descrito en U_t .

Es fácil ver que muchas de las posibilidades de realización de este proceso es-

toestocástico (series de tiempo) pueden tomar cualquiera de las rutas consideradas en el Figura 3.1.

3.2 Función de autocorrelación

Para ampliar la discusión, es posible calcular la fuerza o intensidad de la dependencia de las variables aleatorias dentro de un proceso estocástico, ello mediante el uso de las autocovarianzas. Cuando las covarianzas son normalizadas respecto de la varianza, el resultado es un término que es independiente de las unidad de medida aplicada, y se conoce como la *función de autocorrelación*.

Para procesos estacionarios, dicha función de autocorrelación esta dada por:

$$\rho(\tau) = \frac{\mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)]}{\mathbb{E}[(X_t - \mu)^2]} = \frac{\gamma(\tau)}{\gamma(0)} \quad (3.16)$$

Donde $\tau = \dots, -2, -1, 0, 1, 2, \dots$. Dicha función tiene las siguientes propiedades:

1. $\rho(0) = 1$. Es fácil demostrar que la función $\rho(0)$ es:

$$\rho(0) = \frac{\mathbb{E}[(X_t - \mu)(X_{t+0} - \mu)]}{\mathbb{E}[(X_t - \mu)^2]} = \frac{\mathbb{E}[(X_t - \mu)^2]}{\mathbb{E}[(X_t - \mu)^2]} = 1 \quad (3.17)$$

2. $\rho(\tau) = \rho(-\tau)$. Partiendo de la definición de $\rho(\tau)$ podemos ver que la distancia que existe entre t y $t+\tau$ es τ , de esta forma la autocorrelación de la variable X entre los periodos antes señalados debería ser la misma para el caso en que $\rho(-\tau)$. Partamos de la ecuación para ver más claramente:

$$\rho(\tau) = \frac{\mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)]}{\mathbb{E}[(X_t - \mu)^2]} = \frac{\mathbb{E}[(X_t - \mu)(X_{t-\tau} - \mu)]}{\mathbb{E}[(X_t - \mu)^2]} = \rho(-\tau) \quad (3.18)$$

3. $|\rho(\tau)| \leq 1$, para todo τ .

Derivado de las propiedades 1 y 2 antes descritas se puede concluir que sólo es necesario conocer la función de autocorrelación para el caso de $\tau = 1, 2, 3, \dots$, ya que de estos casos podemos derivar los valores de la función de autocorrelación complementarios de $\tau = \dots, -3, -2, -1$.

Partiendo de los supuestos de ergodicidad en relación a la media, varianza y covarianzas de un proceso estacionario, podemos estimar dichos parámetros con las siguientes formulaciones o propuestas de estimadores puntuales:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T X_t \quad (3.19)$$

$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (X_t - \hat{\mu})^2 = \hat{\sigma}^2 \quad (3.20)$$

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} (X_t - \hat{\mu})(X_{t+\tau} - \hat{\mu}), \text{ para } \tau = 1, 2, \dots, T-1 \quad (3.21)$$

No hacemos la demostración en estas notas –sería deseable que el alumno re-visará la afirmación– pero estos últimos son estimadores consistentes de μ , $\gamma(0)$ y $\gamma(\tau)$. Por su parte, un estimador consistente de la función de autocorrelación estará dado por:

$$\hat{\rho}(\tau) = \frac{\sum_{t=1}^{T-\tau} (X_t - \hat{\mu})(X_{t+\tau} - \hat{\mu})}{\sum_{t=1}^T (X_t - \hat{\mu})^2} = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} \quad (3.22)$$

El estimador de la ecuación (3.22) es asintóticamente insesgado. Por ejemplo, para el caso de un proceso de ruido blanco o caminata aleatoria, su varianza puede ser aproximada por el valor dado $1/T$. Ésta tiene, asintóticamente, una distribución normal. Dado esto, el intervalo de confianza al 95% será el dado por $\pm 2/\sqrt{T}$, en el cual se encuentra la mayoría de los coeficientes de autocorrelación estimados.

Ahora discutamos algunos ejemplos o aplicaciones. Cuando se realiza la evaluación de la estimación de un modelo de series de tiempo es importante saber si los residuales del modelo realmente tienen propiedades de un proceso puramente aleatorio, en particular, si ellos no están correlacionados entre sí. Así, la hipótesis a probar será:

$$H_0 : \rho(\tau) = 0, \text{ para todo } \tau = 1, 2, \dots, m \text{ y } m < T \quad (3.23)$$

Esta expresión se puede interpretar como una prueba respecto de si la correlación entre la información de periodos atrás es cero con la información contemporánea. Para hacer una prueba global de la hipótesis de si un número m de coeficientes de autocovarianzas son cero Box y Pierce (1970) desarrollarán la siguiente estadística:

$$Q^* = T \sum_{j=1}^m \hat{\rho}(j)^2 \quad (3.24)$$

Bajo la hipótesis nula esta estadística se distribuye asintóticamente como una chi cuadrado (χ^2) con $m-k$ grados de libertad y con k que representa al número de parámetros estimados.

Haciendo una aplicación estricta de la distribución de esta estadística, sabemos que esta se mantiene asintóticamente. Greta, Ljung y Box (1978) propusieron

la siguiente modificación de la estadística para muestras pequeñas:

$$Q = T(T+2) \sum_{j=1}^m \frac{\hat{\rho}(j)^2}{T-j} \quad (3.25)$$

La cual también se distribuye asintóticamente como χ^2 con $m - k$ grados de libertad.

También es intuitivamente claro que la hipótesis nula de no autocorrelación de residuales debería ser rechazada si alguno de los valores $\hat{\rho}(j)$ es muy grande, es decir, si Q o Q^* es muy grande. O más precisamente, si estas estadísticas son más grandes que los correspondientes valores críticos de la distribución χ^2 con $m - k$ grados de libertad a algún grado dado de significancia.

Una alternativa para esta prueba es una del tipo Multiplicadores de Lagrange (o LM) desarrollada por Breusch (1978) y Godfrey (1978). La cual, al igual que las estadísticas Q y Q^* , la hipótesis nula está dada por:

H_0 : Los residuales no están autocorrelacionados.

H_a : Los residuales muestran alguna autocorrelación de forma autoregresiva o de medias móviles.

La prueba consiste en realizar una regresión auxiliar en la cual los residuales se estiman en función de las variables explicativas del modelo original y en los residuales mismos pero rezagados hasta el término m (regresión auxiliar). La prueba resulta en una estadística con una distribución χ^2 con m grados de libertad la cual está dada por la expresión:

$$LM = T \times R^2 \quad (3.26)$$

Donde R^2 es el resultante de la regresión auxiliar y T es el número de observaciones totales.

En comparación con una prueba Durbin - Watson que es comúnmente usada en la econometría tradicional, para probar autocorrelación de los residuales, las estadísticas Q , Q^* y LM tienen las siguientes ventajas:

1. Permiten corroborar la existencia de autocorrelación para cualquier orden, y no solo para un primer orden (es decir, para cualquier valor de $\tau = 1, 2, 3, \dots$);
2. Los resultados se mantienen aún y cuando exista una probable variable endógena en forma rezagada, y
3. No depende del orden o la forma en que se acomoden las observaciones, algo que es muy probable que ocurra en la econometría tradicional.

El hecho de los residuales no estén autocorrelacionados no implica que estos sean independientes y normalmente distribuidos. La ausencia de autocorrelación no implica una independencia estocástica si las variables son normalmente distribuidas.

A menudo se asume que estos residuales están distribuidos normalmente, ya que la mayoría de las pruebas estadísticas tienen este supuesto detrás. No obstante, ello también depende de los otros momentos de la distribución, específicamente del tercer y cuarto momento. Los cuales expresan como:

$$\mathbb{E}[(X_t - \mathbb{E}[X_t])^i], i = 3, 4$$

El tercer momento es necesario para determinar el sesgo, el cual esta dado como:

$$\hat{S} = \frac{1}{T} \frac{\sum_{t=1}^T (X_t - \hat{\mu})^3}{\sqrt{\hat{\gamma}(0)^3}} \quad (3.27)$$

Para distribuciones simétricas (como en el caso de la distribución normal) el valor teórico para el sesgo es cero.

La curtosis, la cual esta dada en función del cuarto momento, se puede expresar como:

$$\hat{K} = \frac{1}{T} \frac{\sum_{t=1}^T (X_t - \hat{\mu})^4}{\hat{\gamma}(0)^2} \quad (3.28)$$

Para el caso de una distribución normal, esta estadística toma el valor de 3. Valores más grandes que 3 indican que la distribución tienen colas anchas. En tales casos se ubican a los datos financieros.

Usando el valor de las estadísticas para medir el sesgo y la curtosis, S y K , respectivamente, Jarque y Bera (1980) propusieron una prueba de normalidad, la cual puede ser aplicada a series de tiempo en niveles o en diferencias indistintamente. Dicha prueba se expresa como:

$$JB = \frac{T}{6} \left(\hat{S} + \frac{1}{4}(\hat{K} - 3)^2 \right) \quad (3.29)$$

La cual tiene una distribución χ^2 con 2 grados de libertad y donde T es el tamaño de la muestra. La hipótesis de que las observaciones están distribuidas de forma normal se rechaza si los valores de la estadística de prueba es más grande que los correspondientes valores críticos en tablas.

Veamos un ejemplo para ilustrar el uso de la función de autocorrelación. Tomemos como variable al número de pasajeros transportados por el sistema de transporte del metro de la CDMX.¹ Los datos empleados fueron tomados

¹Los datos y algoritmo está disponible en el repositorio de GitHub y corresponde a la Clase 3.

del INEGI y son una serie de tiempo en el periodo que va de enero de 2000 a junio de 2019, es decir, 234 observaciones. Como se puede apreciar en la Figura 3.2, el número de pasajeros por mes ha oscilado significativamente a lo largo de tiempo. Incluso podemos observar un cambio estructural de la serie entre 2011 y 2012. Asimismo, podemos ubicar una caída atípica que ocurrió en septiembre de 2017.

```
library(ggplot2)
library(dplyr)
library(readxl)

Datos <- read_excel("BD/Base_Transporte.xlsx",
                    sheet = "Datos", col_names = TRUE)

ggplot(data = Datos, aes(x = Periodo, y = Pax_Metro)) +
  geom_line(linewidth = 0.5, color = "darkblue") +
  #geom_point(size = 1.0, color = "darkblue") +
  #theme_bw() +
  xlab("Tiempo") +
  ylab("Millones de pasajeros") +
  theme(plot.title = element_text(size = 11, face = "bold", hjust = 0)) +
  theme(plot.subtitle = element_text(size = 10, hjust = 0)) +
  theme(plot.caption = element_text(size = 10, hjust = 0)) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  labs(
    title = "Pasajeros Transportados en el Metro de la CDMX",
    subtitle = "(Ene-2000 a Jul-2021)",
    caption = "Fuente: Elaboración propia con información del INEGI, \nhttps://www.inegi.org.mx/contenidos/resultados/indicadores/transporte/indicador_1001.aspx"
  )

#
ggsave("Pax_Metro.png", width = 20, height = 15, units = "cm")
```

A esta serie de tiempo le calculamos los principales estadísticos hasta ahora estudiados y obtenemos el Cuadro 3.1. En dicho cuadro se destaca que se muestra la función de autocorrelación para los tres primeros rezagos. Para mayor detalle, en la Figura 3.3 se muestra la función de autocorrelación, en donde las bandas descritas por las líneas azules son el intervalo de confianza dentro de las cuales no se puede rechazar la hipótesis nula de que $H_0 : \hat{\rho}(p) = 0$, para todo $\tau = 1, 2, \dots, T - 1$.

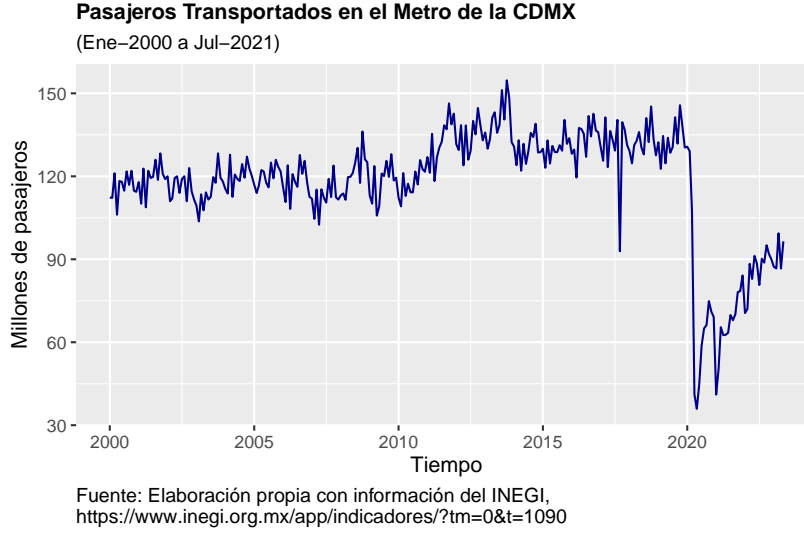


Figure 3.2: Evolución del número de pasajeros en el Metro de la CDMX, enero 2000 a mayo 2023

Table 3.1: Estadísticas descriptivas del número de pasajeros en el Metro de la CDMX, enero de 2000 a junio de 2019

| Estadística | Valor |
|--|----------|
| $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T X_t$ | 124.3000 |
| $\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (X_t - \hat{\mu})^2$ | 103.6400 |
| $\hat{\gamma}(1) = \frac{1}{T} \sum_{t=1}^{T-1} (X_t - \hat{\mu})(X_{t+1} - \hat{\mu})$ | 63.1100 |
| $\hat{\gamma}(2) = \frac{1}{T} \sum_{t=1}^{T-2} (X_t - \hat{\mu})(X_{t+2} - \hat{\mu})$ | 72.9100 |
| $\hat{\gamma}(3) = \frac{1}{T} \sum_{t=1}^{T-3} (X_t - \hat{\mu})(X_{t+3} - \hat{\mu})$ | 63.6900 |
| $\hat{\rho}(1) = \frac{\sum_{t=1}^{T-1} (X_t - \hat{\mu})(X_{t+1} - \hat{\mu})}{\sum_{t=1}^T (X_t - \hat{\mu})^2} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}$ | 0.6089 |
| $\hat{\rho}(2) = \frac{\sum_{t=1}^{T-2} (X_t - \hat{\mu})(X_{t+2} - \hat{\mu})}{\sum_{t=1}^T (X_t - \hat{\mu})^2} = \frac{\hat{\gamma}(2)}{\hat{\gamma}(0)}$ | 0.7035 |
| $\hat{\rho}(3) = \frac{\sum_{t=1}^{T-3} (X_t - \hat{\mu})(X_{t+3} - \hat{\mu})}{\sum_{t=1}^T (X_t - \hat{\mu})^2} = \frac{\hat{\gamma}(3)}{\hat{\gamma}(0)}$ | 0.6145 |
| $Q^* = T \sum_{j=1}^1 \hat{\rho}(j)^2$ | 86.7577 |
| $Q^* = T \sum_{j=1}^2 \hat{\rho}(j)^2$ | 290.9279 |

```

Pax_Metro <- ts(Datos$Pax_Metro,
                start = 2000,
                freq = 12)

acf(Pax_Metro,
    lag.max = 150,
    xlab = 'Resagos k en meses',
    main = "Funcion de Autocorrelación del número de pasajeros del metro")

```

Funcion de Autocorrelación del número de pasajeros del metro

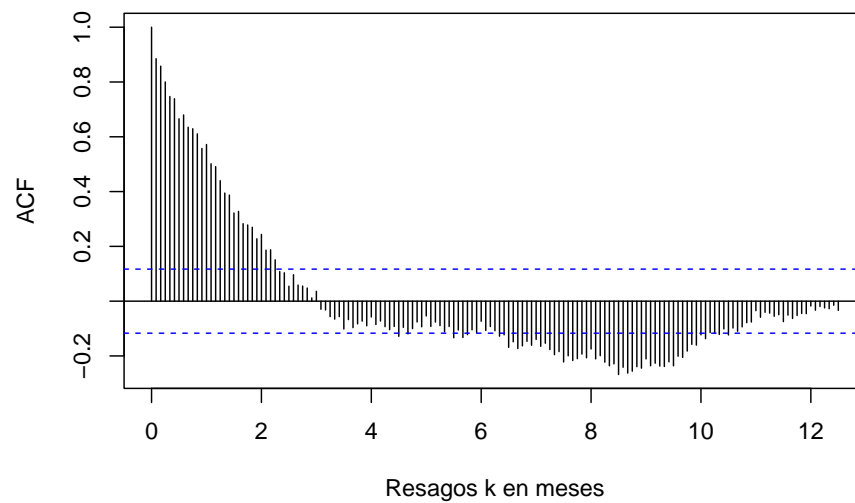


Figure 3.3: Función de Autocorrelación: 150 rezagos del número de pasajeros en el Metro de la CDMX, enero de 2000 a mayo de 2023

Chapter 4

Procesos estacionarios univariados

En este capítulo analizaremos el método o metodología de análisis de series de tiempo propuesto por Box y Jenkins (1970). Los modelos propuestos dentro de esta metodología o conjunto de métodos se han vuelto indispensables para efectos de realizar pronósticos de corto plazo.

En este sentido, se analizarán los métodos más importantes en series de tiempo: procesos autoregresivos (AR) y procesos de medias móviles (MA). Asimismo, se realizará un análisis de los procesos que resultan de la combinación de ambos, conocida como ARMA, los cuales son más comúnmente usados para realizar pronósticos.

4.1 Procesos Autoregresivos (AR)

Los procesos autoregresivos tienen su origen en el trabajo de Cochrane y Orcutt de 1949, mediante el cual analizaron los residuales de una regresión clásica como un proceso autoregresivo. Puede consultarse el apéndice para la discusión del modelo de regresión clásica.

4.1.1 AR(1)

Como primer caso analizaremos al proceso autoregresivo de primer orden, $AR(1)$, el cual podemos definir como una Ecuación Lineal en Diferencia Estocástica de Primer Orden. Diremos que una Ecuación Lineal en Diferencia de Primer Orden es estocástica si en su representación analítica considera un

componente estocástico como en la ecuación (4.1) descrita a continuación:

$$X_t = a_0 + a_1 X_{t-1} + U_t \quad (4.1)$$

Donde a_0 es un término constante, U_t es un proceso estacionario, con media cero (0), una varianza finita y constante (σ^2) y una covarianza que depende de la distancia entre t y cualquier $t-s$ (γ_s)—que no depende de los valores pasados o futuros de la variable—, X_0 es el valor inicial del proceso X_t . No obstante, en ocasiones vamos a asumir que la covarianza será cero (0), por lo que en esos casos tendremos un proceso puramente aleatorio. Considerando la ecuación (4.1) y un proceso de sustitución sucesivo podemos establecer lo siguiente, empezando con X_1 :

$$X_1 = a_0 + a_1 X_0 + U_1$$

Para X_2 :

$$\begin{aligned} X_2 &= a_0 + a_1 X_1 + U_2 \\ &= a_0 + a_1(a_0 + a_1 X_0 + U_1) + U_2 \\ &= a_0 + a_1 a_0 + a_1^2 X_0 + a_1 U_1 + U_2 \end{aligned}$$

Para X_3 :

$$\begin{aligned} X_3 &= a_0 + a_1 X_2 + U_3 \\ &= a_0 + a_1(a_0 + a_1 a_0 + a_1^2 X_0 + a_1 U_1 + U_2) + U_3 \\ &= a_0 + a_1 a_0 + a_1^2 a_0 + a_1^3 X_0 + a_1^2 U_1 + a_1 U_2 + U_3 \end{aligned}$$

Así, para cualquier X_t , $t = 1, 2, 3, \dots$, obtendríamos:

$$\begin{aligned} X_t &= a_0 + a_1 X_{t-1} + U_t \\ &= a_0 + a_1(a_0 + a_1 a_0 + a_1^2 a_0 + \dots + a_1^{t-2} a_0 + a_1^{t-1} X_0 \\ &\quad + a_1^{t-2} U_1 + \dots + a_1 U_{t-2} + U_{t-1}) + U_t \\ &= a_0 + a_1 a_0 + a_1^2 a_0 + a_1^3 a_0 + \dots + a_1^{t-1} a_0 + a_1^t X_0 \\ &\quad + a_1^{t-1} U_1 + \dots + a_1^2 U_{t-2} + a_1 U_{t-1} + U_t \\ &= (1 + a_1 + a_1^2 + a_1^3 + \dots + a_1^{t-1}) a_0 + a_1^t X_0 \\ &\quad + a_1^{t-1} U_1 + \dots + a_1^2 U_{t-2} + a_1 U_{t-1} + U_t \\ &= \frac{1 - a_1^t}{1 - a_1} a_0 + a_1^t X_0 + \sum_{j=0}^{t-1} a_1^j U_{t-j} \end{aligned} \quad (4.2)$$

De esta forma en la ecuación (4.2) observamos un proceso que es explicado por dos partes: una que depende del tiempo y otra que depende de un proceso estocástico. Asimismo, debe notarse que la condición de convergencia es idéntica

que en el caso de ecuaciones en diferencia estudiadas al inicio del curso: $|a_1| < 1$, por lo que cuando $t \rightarrow \infty$, la expresión (4.2) será la siguiente:

$$X_t = a_0 \frac{1}{1 - a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-j} \quad (4.3)$$

Así, desaparece la parte dependiente del tiempo y únicamente prevalece la parte que es dependiente del proceso estocástico. Esta es la solución de largo plazo del proceso $AR(1)$, la cual depende del proceso estocástico. Notemos, además, que esta solución implica que la variable o la serie de tiempo X_t es también un proceso estocástico que hereda las propiedades de U_t . Así, X_t es también un proceso estocástico estacionario, como demostraremos más adelante.

Observemos que la ecuación (4.3) se puede reescribir si consideramos la formulación que en la literatura se denomina como la descomposición de Wold, en la cual se define que es posible asumir que $\psi_j = a_1^j$ y se considera el caso en el cual $|a_1| < 1$, de esta forma tendremos que por ejemplo cuando:

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} a_1^{2j} = \frac{1}{1 - a_1^2}$$

Alternativamente y de forma similar a las ecuaciones en diferencia estudiadas previamente podemos escribir el proceso $AR(1)$ mediante el uso del operador rezago como:

$$\begin{aligned} X_t &= a_0 + a_1 L X_t + U_t \\ X_t - a_1 L X_t &= a_0 + U_t \\ (1 - a_1 L) X_t &= a_0 + U_t \\ X_t &= \frac{a_0}{1 - a_1 L} + \frac{1}{1 - a_1 L} U_t \end{aligned} \quad (4.4)$$

En esta última ecuación retomamos el siguiente término para reescribirlo como:

$$\frac{1}{1 - a_1 L} = 1 + a_1 L + a_1^2 L^2 + a_1^3 L^3 + \dots \quad (4.5)$$

Tomando este resultado para sustituirlo en ecuación (4.4), obtenemos la siguiente expresión:

$$\begin{aligned} X_t &= (1 + a_1 L + a_1^2 L^2 + a_1^3 L^3 + \dots) a_0 + (1 + a_1 L + a_1^2 L^2 + a_1^3 L^3 + \dots) U_t \\ &= (1 + a_1 + a_1^2 + a_1^3 + \dots) a_0 + U_t + a_1 U_{t-1} + a_1^2 U_{t-2} + a_1^3 U_{t-3} + \dots \\ &= \frac{a_0}{1 - a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-j} \end{aligned} \quad (4.6)$$

Donde la condición de convergencia y estabilidad del proceso descrito en esta ecuación es que $|a_1| < 1$. Por lo que hemos demostrado que mediante el uso del operador de rezago es posible llegar al mismo resultado que obtuvimos mediante el procedimiento de sustituciones iterativas.

La ecuación (4.6) se puede interpretar como sigue. La solución o trayectoria de equilibrio de un $AR(1)$ se divide en dos partes. La primera es una constante que depende de los valores de a_0 y a_1 . La segunda parte es la suma ponderada de las desviaciones o errores observados y acumulados en el tiempo hasta el momento t .

Ahora obtendremos los momentos que describen a la serie de tiempo cuando se trata de un proceso $AR(1)$. Para ello debemos obtener la media, la varianza y las covarianzas de X_t . Para los siguientes resultados debemos recordar y tener en mente que si U_t es un proceso puramente aleatorio, entonces:

1. $\mathbb{E}[U_t] = 0$ para todo t
2. $Var[U_t] = \sigma^2$ para todo t
3. $Cov[U_t, U_s] = 0$ para todo $t \neq s$

Dicho lo anterior y partiendo de la ecuación (4.6), el primer momento o valor esperado de la serie de tiempo será el siguiente:

$$\begin{aligned}
 \mathbb{E}[X_t] &= \mathbb{E}\left[\frac{a_0}{1-a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-j}\right] \\
 &= \frac{a_0}{1-a_1} + \sum_{j=0}^{\infty} a_1^j \mathbb{E}[U_{t-j}] \\
 &= \frac{a_0}{1-a_1} = \mu
 \end{aligned} \tag{4.7}$$

Respecto de la varianza podemos escribir la siguiente expresión a partir de la

ecuación (4.6):

$$\begin{aligned}
 Var[X_t] &= \mathbb{E}[(X_t - \mu)^2] \\
 &= \mathbb{E}\left[\left(\frac{a_0}{1-a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-j} - \frac{a_0}{1-a_1}\right)^2\right] \\
 &= \mathbb{E}[(U_t + a_1 U_{t-1} + a_1^2 U_{t-2} + a_1^3 U_{t-3} + \dots)^2] \\
 &= \mathbb{E}[U_t^2 + a_1^2 U_{t-1}^2 + a_1^4 U_{t-2}^2 + a_1^6 U_{t-3}^2 + \dots \\
 &\quad + 2a_1 U_t U_{t-1} + 2a_1^2 U_t U_{t-2} + \dots] \\
 &= \mathbb{E}[U_t^2] + a_1^2 \mathbb{E}[U_{t-1}^2] + a_1^4 \mathbb{E}[U_{t-2}^2] + a_1^6 \mathbb{E}[U_{t-3}^2] + \dots \\
 &= \sigma^2 + a_1^2 \sigma^2 + a_1^4 \sigma^2 + a_1^6 \sigma^2 + \dots \\
 &= \sigma^2(1 + a_1^2 + a_1^4 + a_1^6 + \dots) \\
 &= \sigma^2 \frac{1}{1-a_1^2} = \gamma(0)
 \end{aligned} \tag{4.8}$$

Previo a analizar la covarianza de la serie recordemos que para el proceso puramente aleatorio U_t su varianza y covarianza puede verse como $\mathbb{E}[U_t, U_s] = \sigma^2$, para $t = s$, y $\mathbb{E}[U_t, U_s] = 0$, para cualquier otro caso, respectivamente.

Dicho lo anterior, partiendo de la ecuación (4.6) la covarianza de la serie estará dada por:

$$\begin{aligned}
 Cov(X_t, X_{t-\tau}) &= \mathbb{E}[(X_t - \mu)(X_{t-\tau} - \mu)] \\
 &= \mathbb{E}\left[\left(\frac{a_0}{1-a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-j} - \frac{a_0}{1-a_1}\right) \right. \\
 &\quad \left. \times \left(\frac{a_0}{1-a_1} + \sum_{j=0}^{\infty} a_1^j U_{t-\tau-j} - \frac{a_0}{1-a_1}\right)\right] \\
 &= a_1^\tau \mathbb{E}[U_{t-\tau}^2 + a_1 U_{t-\tau-1}^2 + a_1^2 U_{t-\tau-2}^2 + a_1^3 U_{t-\tau-3}^2 + \dots] \\
 &= a_1^\tau \sigma^2 \frac{1}{1-a_1^2} = \gamma(\tau)
 \end{aligned} \tag{4.9}$$

Notése que con estos resultados en las ecuaciones (4.8) y (4.9) podemos construir la función de autocorrelación teórica como sigue:

$$\begin{aligned}
 \rho(\tau) &= \frac{\gamma(\tau)}{\gamma(0)} \\
 &= a_1^\tau
 \end{aligned} \tag{4.10}$$

Donde $\tau = 1, 2, 3, \dots$ y $|a_1| < 1$. Este último resultado significa que cuando el proceso autoregresivo es de orden 1 (es decir, AR(1)) la función de autocorrelación teóricamente es igual al parámetro a_1 elevado al número de rezagos considerados. No obstante, note que esto no significa que la autocorrelación

observada sea como lo expresa en planteamiento anterior. Por el contrario, una observación sencilla mostraría que la autocorrelación observada sería ligeramente distinta a la autocorrelación teórica.

Ahora veámos algunos ejemplos. En el primer ejemplo simularemos una serie y mostraremos el análisis de un proceso construido considerando un proceso puramente aleatorio como componente U_t . Por su parte, en un segundo ejemplo aplicaremos el análisis a una serie de tiempo de una variable económica observada.

Para el primer ejemplo consideremos un proceso dado por la forma de un $AR(1)$ como en la ecuación (4.4) cuya solución esta dada por la ecuación (4.6). En específico, supongamos que el término o componente estocástico U_t es una serie generada a partir de numeros aleatorios de una función normal con media 0 y desviación estándar 4. Los detalles del proceso simulado se muestra en las siguientes gráficas.

La Figura 4.1 ilustra el comportamiento que se debería observar en una serie considerando el procedimiento iterativo de construcción. Por su parte, la Figura 4.2 ilustra el proceso o trayectoria de la solución de la serie de tiempo. Finalmente, las Figuras 4.3 y 4.4 muestran el correlograma calculado considerando una función de autocorrelación aplicada al porceso real y una función de autocorrelación aplicada al proceso teórico, respectivamente.

```
library(ggplot2)
library(dplyr)
library(latex2exp)

a0 <- 5; a1 <- 0.9; X_0 <- (a0/(1 - a1)); T <- 1000

X_t <- data.frame(Tiempo = c(0:T))

set.seed(12345)

# Agregamos un término estocástico al data frame

X_t$U_t <- rnorm(T+1, mean = 0, sd = 4)

# Agregamos columnas con NA's para un proceso teorico y uno real
X_t$X_t <- NA
X_t$XR_t <- NA

# La serie teórica inicia en un valor inicial X_0
X_t$X_t[1] <- X_0

# La serie real inicia en un valor inicial X_0
```



```

X_t$XR_t[1] <- X_0

# Agregamos una columna para la función de Autocorrelación teórica:
X_t$rho <- NA

for (i in 2:(T + 1)) {
  # Real:
  X_t$XR_t[i] = a0 + a1*X_t$XR_t[i-1] + X_t$U_t[i-1]

  # Teórico:
  X_t$X_t[i] = X_t$X_t[i-1] + (a1^(i-1))*X_t$U_t[i-1]

  # Autocorrelación:
  X_t$rho[i-1] = a1^(i-1)
}

ggplot(data = X_t, aes(x = Tiempo, y = XR_t)) +
  geom_line(size = 0.5, color = "darkred") +
  #theme_bw() +
  xlab("Tiempo") +
  ylab(TeX("$X_t$")) +
  theme(plot.title = element_text(size = 11, face = "bold",
                                   hjust = 0)) +
  theme(plot.subtitle = element_text(size = 10, hjust = 0)) +
  theme(plot.caption = element_text(size = 10, hjust = 0)) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  labs(
    title = "Comportamiento del Proceso Real ('Estimado')",
    subtitle = "Con un error con Distribución Normal (media = 0, desviación estándar = 4)",
    caption = "Fuente: Elaboración propia."
  )

```

```

ggsave("G_AR_1_Real.png", width = 20, height = 10, units = "cm")

```

```

ggplot(data = X_t, aes(x = Tiempo, y = X_t)) +
  geom_line(size = 0.5, color = "darkblue") +
  #theme_bw() +
  xlab("Tiempo") +
  ylab(TeX("$X_t$")) +
  theme(plot.title = element_text(size = 11, face = "bold",
                                   hjust = 0)) +
  theme(plot.subtitle = element_text(size = 10, hjust = 0)) +
  theme(plot.caption = element_text(size = 10, hjust = 0)) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +

```

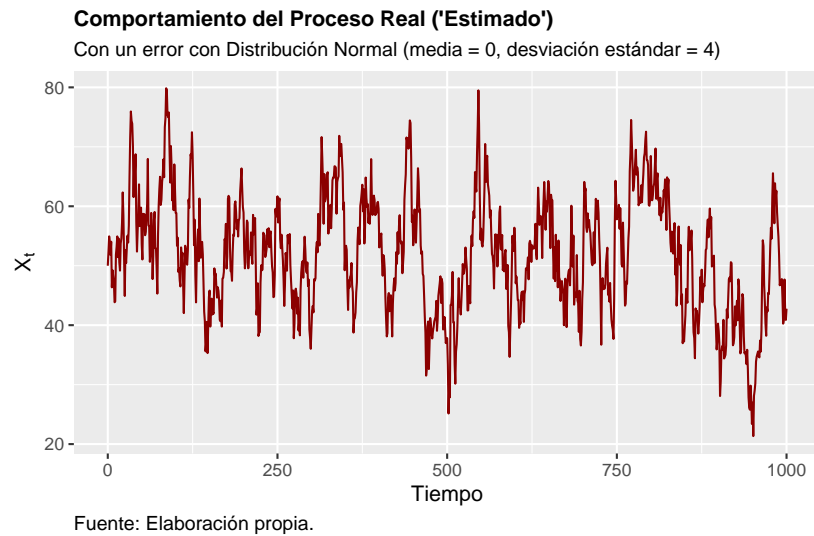


Figure 4.1: Comportamiento del Proceso Real ('Estimado')

```
labs(
  title = "Comportamiento del Proceso Teórico",
  subtitle = "Con un error con Distribución Normal (media = 0, desviación estándar = 4)",
  caption = "Fuente: Elaboración propia."
)

ggsave("G_AR_1_Teo.png", width = 20, height = 10, units = "cm")

acf(X_t$XR_t, lag.max = 30, col = "blue",
  ylab = "Autocorrelacion",
  xlab="Rezagos",
  main="Funcion de Autocorrelacion Real")

barplot(X_t$rho[1:30], names.arg = c(1:30), col = "blue",
  border="blue", density = c(10,20),
  ylab = "Autocorrelacion",
  xlab="Rezagos",
  main="Funcion de Autocorrelacion Teórica")
```

Recordemos que una trayectoria de equilibrio o solución de un $AR(1)$ es como se muestra en la ecuación (4.6). Así, nuestra serie simulada cumple con la

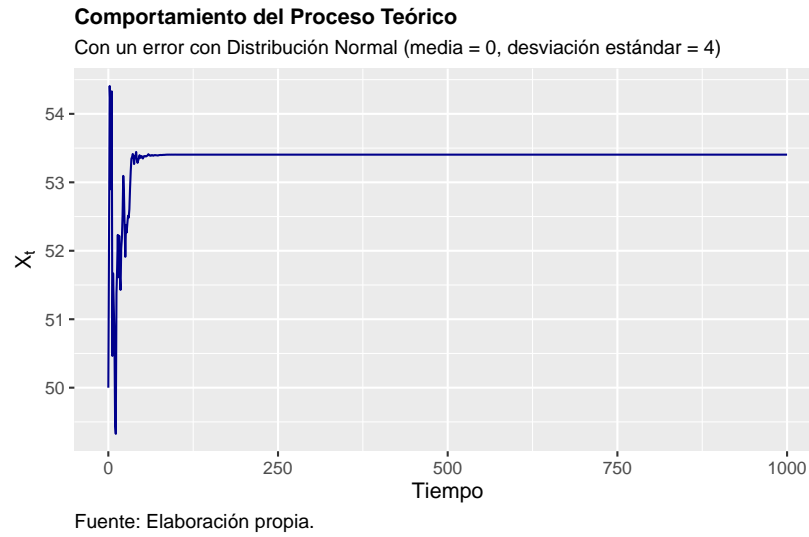


Figure 4.2: Comportamiento del Proceso Teórico

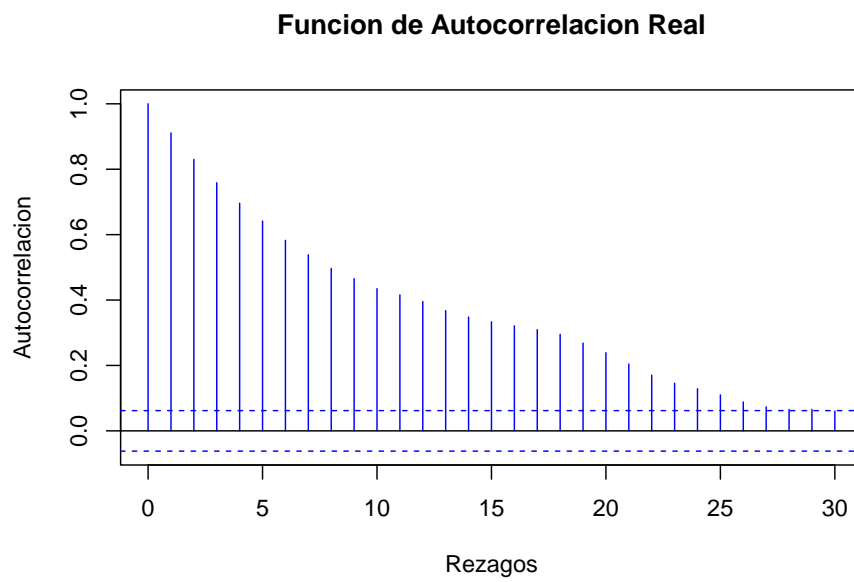


Figure 4.3: Funcion de Autocorrelacion Real

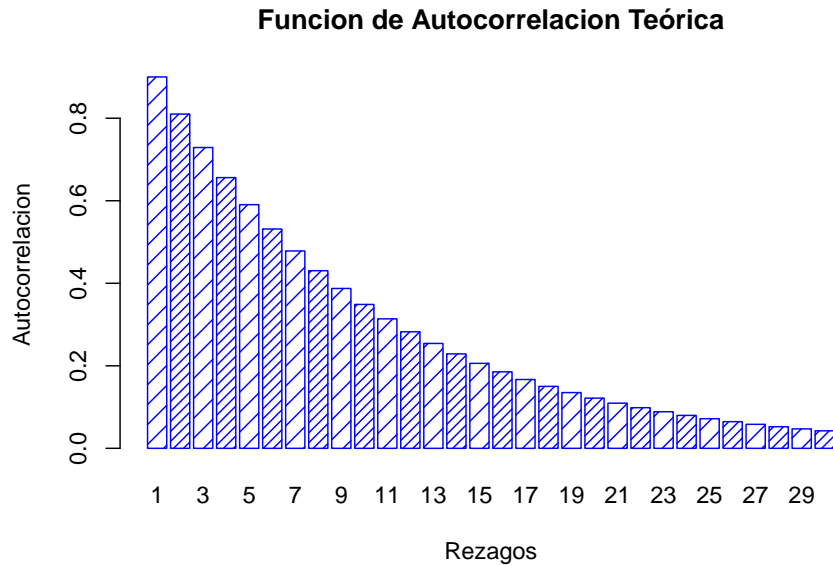


Figure 4.4: Funcion de Autocorrelacion Teórica

característica de que los errores son más relevantes cuando la serie es corta. Por el contrario, los errores son menos relevantes, cuando la serie es muy larga. La Figura 4.5 ilustra esta observación de la trayectoria de equilibrio.

```
ggplot(data = X_t, aes(x = Tiempo)) +
  geom_line(aes(y = XR_t), size = 0.5, color = "darkred") +
  geom_line(aes(y = X_t), size = 0.5, color = "darkblue") +
  #theme_bw() +
  xlab("Tiempo") +
  ylab(TeX("$X_t$")) +
  theme(plot.title = element_text(size = 11, face = "bold",
                                   hjust = 0)) +
  theme(plot.subtitle = element_text(size = 10, hjust = 0)) +
  theme(plot.caption = element_text(size = 10, hjust = 0)) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  labs(
    title = "Comportamiento de los Procesos Real y Teórico",
    subtitle = "Con un error con Distribución Normal (media = 0, desviación estándar =",
    caption = "Fuente: Elaboración propia."
  )
)
```

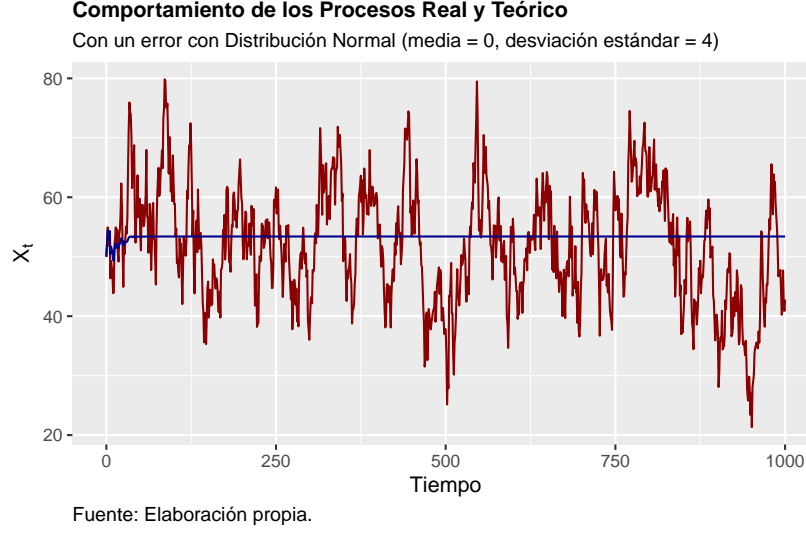


Figure 4.5: Comportamiento de los Procesos Real y Teórico

```
ggsave("G_AR_1_Comb.png", width = 20, height = 10, units = "cm")
```

Para el segundo ejemplo consideremos una aplicación a una serie de tiempo en específico: Pasajeros transportados mensualmente en el Sistema de Transporte Colectivo Metro (pasajeros medidos en millones).¹

A la serie se le aplicará una metodología de estimación dada por el método de Máxima Verosimilitud (ML, por sus siglas en inglés). Antes de realizar el proceso de estimación consideremos una transformación de diferencias logarítmicas, con el objeto de obtener una serie de tiempo expresada en tasas de crecimiento² y con un comportamiento parecido a un proceso estacionario.

Así, para cada una de las series que analicemos en diferencias logarítmicas respecto del momento k las expresaremos bajo la siguiente transformación:

$$DLX_t = \log(X_t) - \log(X_{t-k})$$

Donde $k = 1, 2, 3, \dots$ y $\log(\cdot)$ es la función logaritmo natural. Esta expresión se puede interpretar como una tasa de crecimiento puesto que asumimos variaciones pequeñas para las cuales se cumple que: $\log(X_t) - \log(X_{t-k}) \approx \frac{X_t - X_{t-k}}{X_t}$.

¹Fuente: INEGI, <https://www.inegi.org.mx/app/indicadores/?tm=0&t=1090>.

²Estas tasas no son porcentuales, para hacerlas porcentuales faltaría multiplicar por 100 cada valor de la serie.

Primero, para realizar el análisis de una serie de tiempo deberemos decidir si éste se realizará para la serie en niveles o en diferencias. Por convención, decimos que la serie está en niveles si ésta se analiza sin hacerle ninguna transformación o si se analiza aplicando solo logaritmos. Cuando la serie se analiza en diferencias significa que la diferencia se hace sin aplicar logaritmos o aplicando logaritmos. Sin embargo, la convención es hacer un análisis en diferencias logarítmicas.

Para decidir cómo analizar la serie de pasajeros en el metro de la CDMX en la Figura 4.6} se muestra la gráfica de la serie en niveles (sin transformación logarítmica y con transformación logarítmica) y en diferencias logarítmicas mensuales (es decir, con $k = 1$).

```
library(ggplot2)
library(dplyr)
library(readxl)
library(stats)

Datos <- read_excel("BD/Base_Transporte.xlsx",
                    sheet = "Datos", col_names = TRUE)

# En Niveles
Pax_Metro <- ts(Datos$Pax_Metro, start = c(2000, 1),
               freq = 12)

# En Logaritmos:
Pax_LMetro <- ts(log(Datos$Pax_Metro), start = c(2000, 1),
               freq = 12)

# Diferencias mensuales:
Pax_DLMetro <- ts( log(Datos$Pax_Metro) -
                  lag( log(Datos$Pax_Metro), 1 ),
                  start = c(2000, 1), freq = 12)

#
par(mfrow = c(3,1))

plot(Pax_Metro, xlab = "Tiempo",
     main = "Pasajeros transportados (Millones) en el SCM",
     col = "darkgreen")

plot(Pax_LMetro, xlab = "Tiempo",
     main = "LN Pasajeros transportados (Millones) en el SCM",
     col = "darkblue")

plot(Pax_DLMetro, xlab = "Tiempo",
     main = "Diff LN Pasajeros transportados (Millones) en el SCM",
```

```
col = "darkred")
```

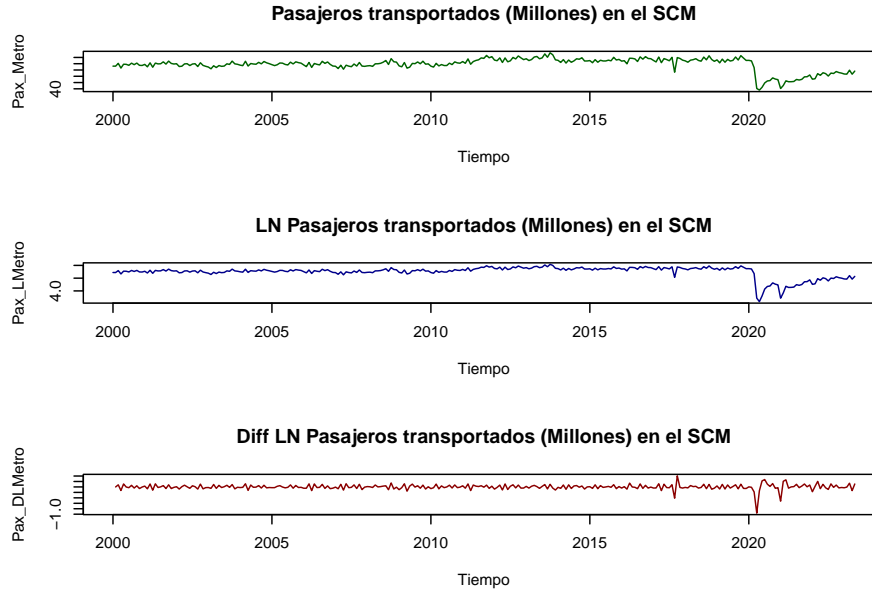


Figure 4.6: Pasajeros transportados (Millones) en el metro de la CDM en niveles y en diferencias logarítmicas

```
par(mfrow=c(1,1))
```

A continuación, estimaremos una $AR(1)$ para la serie en niveles bajo la transformación logarítmica ($PaxLMetro_t$) y en diferencias logarítmicas ($PaxDLMetro_t$). Para el primer caso obtenemos el siguiente resultado:

$$PaxLMetro_t = \begin{matrix} 4.7419 \\ (0.0105) \end{matrix} + \begin{matrix} 0.5916 \\ (0.0526) \end{matrix} PaxLMetro_{t-1}$$

$$\hat{\sigma}^2 = 0.004335 \quad y \quad AIC = -602.73$$

Para el segundo caso obtenemos el siguiente resultado:

$$PaxDLMetro_t = \begin{matrix} 0.0007 \\ (0.0023) \end{matrix} - \begin{matrix} 0.6194 \\ (0.0511) \end{matrix} PaxDLMetro_{t-1}$$

$$\hat{\sigma}^2 = 0.003344 \quad y \quad AIC = -660.53$$

En ambos casos observamos que el parámetro asociado al componente AR es significativo y cumple con la restricción de ser en valor absoluto menor a 1, por lo que la solución asociada al proceso será convergente. También en ambos casos se reporta la estadística o Criterio de Información de Akaike (AIC, por sus siglas en inglés), misma que más adelante discutiremos su importancia y aplicación.

4.1.2 AR(2)

Una vez analizado el caso de $AR(1)$ analizaremos el caso del $AR(2)$. La ecuación generalizada del proceso autoregresivo de orden 2 (denotado como $AR(2)$) puede ser escrito como:

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + U_t \quad (4.11)$$

Donde U_t denota un proceso puramente aleatorio con media cero (0), varianza constante (σ^2) y autocovarianza cero ($Cov(U_t, U_s) = 0$, con $t \neq s$), y un parámetro $a_2 \neq 0$. Así, utilizando el operador rezago podemos reescribir la ecuación (4.11) como:

$$\begin{aligned} X_t - a_1 X_{t-1} - a_2 X_{t-2} &= a_0 + U_t \\ (1 - a_1 L^1 - a_2 L^2) X_t &= a_0 + U_t \end{aligned}$$

Donde, vamos a denotar a $\alpha(L) = (1 - a_1 L^1 - a_2 L^2)$, y lo llamaremos como un polinomio que depende del operador rezago y que es distinto de cero. De esta forma podemos reescribir a la ecuación (4.11) como:

$$\alpha(L) X_t = a_0 + U_t \quad (4.12)$$

Ahora, supongamos que existe el inverso multiplicativo del polinomio $\alpha(L)$, el cual será denotado como: $\alpha^{-1}(L)$ y cumple con que:

$$\alpha^{-1}(L) \alpha(L) = 1 \quad (4.13)$$

Así, podemos escribir la solución a la ecuación (4.11) como:

$$X_t = \alpha^{-1}(L) \delta + \alpha^{-1}(L) U_t$$

Si utilizamos el hecho que $\alpha^{-1}(L)$ se puede descomponer a través del procedimiento de Wold en un polinomio de forma similar el caso de $AR(1)$, tenemos que:

$$\alpha^{-1}(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots \quad (4.14)$$

Por lo tanto, el inverso multiplicativo $\alpha^{-1}(L)$ se puede ver como:

$$1 = (1 - a_1 L^1 - a_2 L^2)(\psi_0 + \psi_1 L + \psi_2 L^2 + \dots) \quad (4.15)$$

Desarrollando la ecuación (4.15) tenemos la siguiente expresión:

$$\begin{array}{ccccccccc}
1 & = & \psi_0 & + & \psi_1 L & + & \psi_2 L^2 & + & \psi_3 L^3 & + & \dots \\
& & & - & a_1 \psi_0 L & - & a_1 \psi_1 L^2 & - & a_1 \psi_2 L^3 & - & \dots \\
& & & & & - & a_2 \psi_0 L^2 & - & a_2 \psi_1 L^3 & - & \dots
\end{array}$$

Ahora, podemos agrupar todos los términos en función del exponente asociado al operador rezago L . La siguiente es una solución particular y es una de las múltiples que podrían existir que cumpla con la ecuación (4.15). Sin embargo, para efectos del análisis sólo necesitamos una de esas soluciones. Utilizaremos las siguientes condiciones que deben cumplirse en una de las posibles soluciones:

$$\begin{array}{llll}
L^0 : & & \Rightarrow & \psi_0 = 1 \\
L : & \psi_1 - a_1 \psi_0 = 0 & \Rightarrow & \psi_1 = a_1 \\
L^2 : & \psi_2 - a_1 \psi_1 - a_2 \psi_0 = 0 & \Rightarrow & \psi_2 = a_1^2 + a_2 \\
L^3 : & \psi_3 - a_1 \psi_2 - a_2 \psi_1 = 0 & \Rightarrow & \psi_3 = a_1^3 + 2a_1 a_2 \\
\vdots & \vdots & \vdots & \vdots
\end{array}$$

De esta forma podemos observar que en el límite siempre obtendremos una ecuación del tipo $\psi_j - a_1 \psi_{j-1} - a_2 \psi_{j-2} = 0$ asociada a cada uno de los casos en que exista un L^j , donde $j \neq 0, 1$, y la cual siempre podremos resolver conociendo que las condiciones iniciales son: $\psi_0 = 1$ y $\psi_1 = a_1$.

Así, de las relaciones antes mencionadas y considerando que $\alpha^{-1}(L)$ aplicada a una constante como a_0 , tendrá como resultado otra constante. De esta forma podemos escribir que la solución del proceso $AR(2)$ en la ecuación (4.11) será dada por una expresión como sigue:

$$X_t = \frac{\delta}{1 - a_1 - a_2} + \sum_{j=0}^{\infty} \psi_{t-j} U_{t-j} \quad (4.16)$$

Donde todos los parámetros ψ_i está determinado por los parámetros a_0 , a_1 y a_2 . En particular, $\psi_0 = 1$ y $\psi_1 = a_1$ como describimos anteriormente. Al igual que en el caso del $AR(1)$, en la ecuación (4.16) las condiciones de estabilidad estarán dadas por las soluciones del siguiente polinomio característico:³

$$\lambda^2 - \lambda a_1 - a_2 = 0 \quad (4.17)$$

Así, la condición de estabilidad de la trayectoria es que $|\lambda_i| < 1$, para $i = 1, 2$. Es decir, es necesario que cada una de las raíces sea, en valor absoluto, siempre menor que la unidad. Estas son las condiciones de estabilidad para el proceso $AR(2)$.

Finalmente, al igual que en un $AR(1)$, a continuación determinamos los momentos de una serie que sigue un proceso $AR(2)$. Iniciamos con la determinación de la media de la serie:

$$\mathbb{E}[X_t] = \mu = \frac{a_0}{1 - a_1 - a_2} \quad (4.18)$$

³Note que raíces son equivalentes al inverso de las del polinomio dado por $\lambda^2 a_2 - \lambda a_1 - 1 = 0$.

Lo anterior es cierto puesto que $\mathbb{E}[U_{t-i}] = 0$, para todo $i = 0, 1, 2, \dots$. Para determinar la varianza utilizaremos las siguientes relaciones basadas en el uso del valor esperado, varianza y covarianza de la serie. Adicionalmente, para simplificar el trabajo asumamos que $a_0 = 0$, lo cual implica que $\mu = 0$. Dicho lo anterior, partamos de:

$$\begin{aligned}\mathbb{E}[X_t X_{t-\tau}] &= \mathbb{E}[(a_1 X_{t-1} + a_2 X_{t-2} + U_t) X_{t-\tau}] \\ &= a_1 \mathbb{E}[X_{t-1} X_{t-\tau}] + a_2 \mathbb{E}[X_{t-2} X_{t-\tau}] + \mathbb{E}[U_t X_{t-\tau}]\end{aligned}$$

Donde $\tau = 0, 1, 2, 3, \dots$ y que $\mathbb{E}[U_t X_{t-\tau}] = 0$ para todo $\tau \neq 0$.⁴ Dicho esto, podemos derivar el valor del valor esperado para diferentes valores de τ :

$$\begin{aligned}\tau = 0 : \quad \gamma(0) &= \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma^2 \\ \tau = 1 : \quad \gamma(1) &= \alpha_1 \gamma(0) + \alpha_2 \gamma(1) \\ \tau = 2 : \quad \gamma(2) &= \alpha_1 \gamma(1) + \alpha_2 \gamma(0)\end{aligned}$$

Donde debe ser claro que $\mathbb{E}[(X_t - \mu)(X_{t-\tau} - \mu)] = \mathbb{E}[X_t X_{t-\tau}] = \gamma(\tau)$. Así, en general cuando $\tau \neq 0$:

$$\gamma(\tau) = a_1 \gamma(\tau - 1) + a_2 \gamma(\tau - 2) \quad (4.19)$$

Realizando la sustitución recursiva y solucionando el sistema respectivo obtenemos que las varianzas y covarianzas estaran determinadas por:

$$Var[X_t] = \gamma(0) = \frac{1 - a_2}{(1 + a_2)[(1 - a_2)^2 - a_1^2]} \sigma^2 \quad (4.20)$$

$$\gamma(1) = \frac{a_1}{(1 + a_2)[(1 - a_2)^2 - a_1^2]} \sigma^2 \quad (4.21)$$

$$\gamma(2) = \frac{a_1^2 + a_2 - a_2^2}{(1 + a_2)[(1 - a_2)^2 - a_1^2]} \sigma^2 \quad (4.22)$$

Recordemos que las funciones de autocorrelación se obtienen de la división de cada una de las funciones de covarianza ($\gamma(\tau)$) por la varianza ($\gamma(0)$). Así, podemos construir la siguiente expresión:

$$\rho(\tau) - a_1 \rho(\tau - 1) - a_2 \rho(\tau - 2) = 0 \quad (4.23)$$

Ahora veámos un ejemplo. Utilizaremos la serie de Pasajeros en vuelos nacionales (en vuelos de salidas) para estimar un $AR(2)$ mediante el método de

⁴Es fácil demostrar esta afirmación, sólo requiere de desarrollar la expresión y utilizar el hecho de que U_t es un proceso puramente aleatorio, por lo que la covarianza es cero (0).

máxima verosimilitud (ML, por sus siglas en inglés). Antes de realizar el proceso de estimación consideremos una transformación de la serie en logaritmos y una más en diferencias logarítmicas; lo anterior con el objeto de obtener un conjunto de series de tiempo suavizada y expresada en tasas de crecimiento, con un comportamiento parecido a un proceso estacionario.

Así, para cada una de las series que analicemos en diferencias logarítmicas las expresaremos bajo la siguiente transformación:

$$DLX_t = \log(X_t) - \log(X_{t-k})$$

Donde $k = 1, 2, 3, \dots$ y $\log(\cdot)$ es la función logaritmo natural. Por convención, decimos que la serie está en niveles si ésta se analiza sin hacerle ninguna transformación o se analiza en logaritmos. Cuando la serie se analiza en diferencias significa que la diferencia se hace sin aplicar logaritmos. Y cuando la serie analizada está en diferencias logarítmicas también diremos que esta en diferencias. Sin embargo, lo común es hacer un análisis en logaritmos y en diferencias logarítmicas.

Primero, para decidir si se realizará un AR(2) para la serie en niveles o en diferencias analizaremos su gráfica. La serie en niveles, en niveles bajo una transformación logarítmica y en diferencias logarítmicas mensuales de los pasajeros en vuelos nacionales se muestra en la Figura 4.7.

```
library(ggplot2)
library(dplyr)
library(readxl)
library(stats)

Datos <- read_excel("BD/Base_Transporte.xlsx",
                    sheet = "Datos", col_names = TRUE)

# En Niveles
Pax_Nal <- ts(Datos$Pax_Nal,
              start = c(2000, 1),
              freq = 12)

# Logaritmos:
LPax_Nal <- ts(log(Datos$Pax_Nal),
               start = c(2000, 1),
               freq = 12)

# Diferencias mensuales:
DLPax_Nal <- ts(log(Datos$Pax_Nal) -
                lag(log(Datos$Pax_Nal), 1),
                start = c(2000, 1), freq = 12)
```

```
#
par(mfrow=c(3,1))

plot(Pax_Nal, xlab = "Tiempo", ylab = "Pasajeros",
     main = "Pasajeros en vuelos nacionales de salida",
     col = "darkgreen")

plot(LPax_Nal, xlab = "Tiempo", ylab = "LN Pasajeros",
     main = "LN Pasajeros en vuelos nacionales de salida",
     col = "darkblue")

plot(DLPax_Nal, xlab = "Tiempo", ylab = "DLN Pasajeros",
     main = "Diff LN Pasajeros en vuelos nacionales de salida",
     col = "darkred")
```

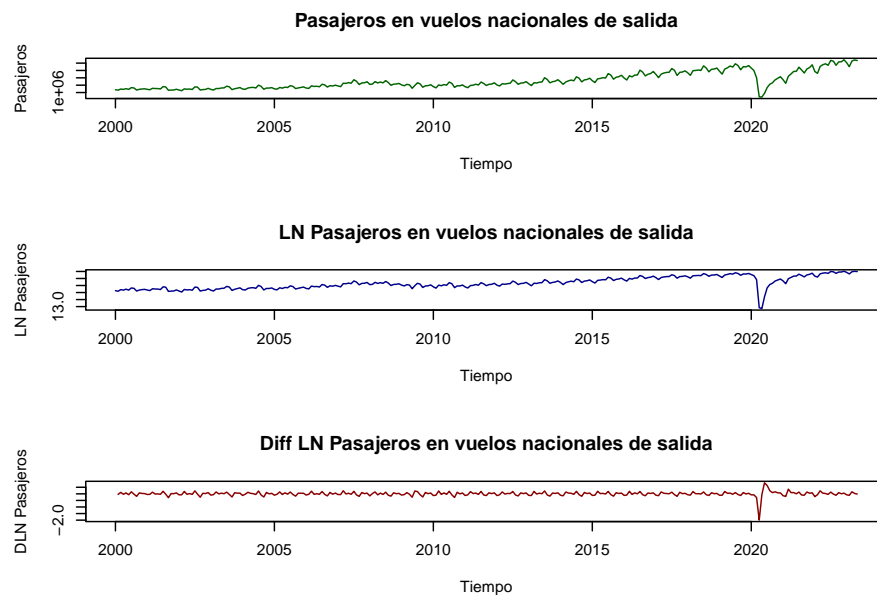


Figure 4.7: Pasajeros en vuelos de salidas nacionales en niveles y en diferencias logarítmicas

```
par(mfrow=c(1,1))
```

A continuación, estimaremos un $AR(2)$ para la serie en niveles bajo una transformación logarítmica ($LPaxNal_t$) y en diferencias logarítmicas ($DLPaxNal_t$). Para el primer caso obtenemos el siguiente resultado:

$$\begin{aligned}
 LPaxNal_t &= 14.6267 + 0.7637 LPaxNal_{t-1} \\
 &\quad (0.1816) \quad (0.0637) \\
 &\quad + 0.2025 LPaxNal_{t-2} \\
 &\quad (0.0646)
 \end{aligned}$$

$$\hat{\sigma}^2 = 0.01138 \quad \text{y} \quad AIC = -372.64$$

Para el segundo caso obtenemos el siguiente resultado:

$$\begin{aligned}
 DLPaxNal_t &= 0.0050 - 0.3205 DLPaxNal_{t-1} \\
 &\quad (0.0036) \quad (0.0592) \\
 &\quad - 0.4242 DLPaxNal_{t-2} \\
 &\quad (0.0591)
 \end{aligned}$$

$$\hat{\sigma}^2 = 0.009378 \quad \text{y} \quad AIC = -418.3$$

Para ambos casos entre parentésis indicamos los errores estándar y reportamos el estadístico de Akaike, AIC. Finalmente, podemos determinar si las soluciones serán convergentes, para ello en la Figura 4.8 mostramos las raíces asociadas a cada uno de los polinomios. De la inspección visual podemos concluir que ambas propuesta de AR(2) representan una solución convergente y estable.

```

source("arroots.R")

source("plot.armacroots.R")

AR_LPax_Nal <- arima(LPax_Nal, order = c(2, 0, 0),
                     method = "ML")

AR_DLPax_Nal <- arima(DLPax_Nal, order = c(2, 0, 0),
                     method = "ML")

par(mfrow=c(1,2))

plot.armacroots(arroots(AR_LPax_Nal),
               main="Inverse AR roots of \nAR(2): LN Pax Nal")

#
plot.armacroots(arroots(AR_DLPax_Nal),
               main="Inverse AR roots of \nAR(2): Diff LN Pax Nal")

par(mfrow=c(1,1))

```

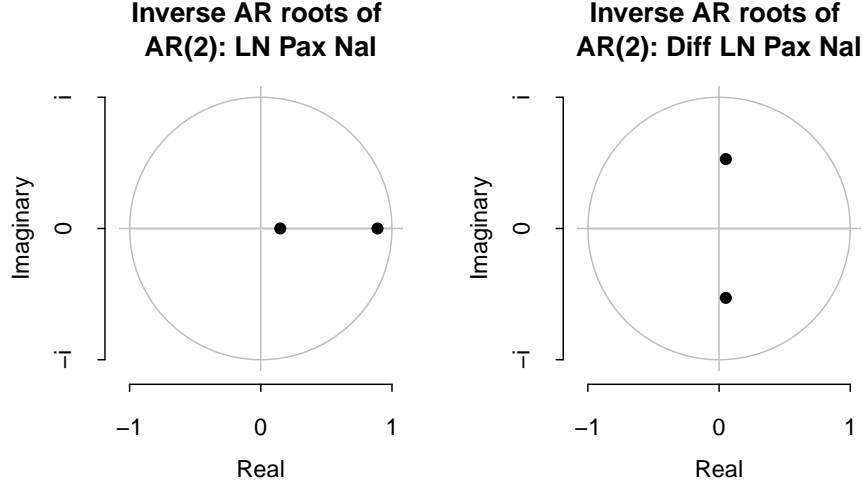


Figure 4.8: Inverso de las Raíces del polinomio característico

4.1.3 AR(p)

Veremos ahora una generalización de los procesos autoregresivos (AR). Esta generalización es conocida como un proceso $AR(p)$ y que puede ser descrito por la siguiente ecuación en diferencia estocástica:

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + \dots + a_p X_{t-p} + U_t \quad (4.24)$$

Donde $a_p \neq 0$, y U_t es un proceso puramente aleatorio con media cero (0), varianza constante (σ^2) y covarianza cero (0). Usando el operador rezago, L^k , para $k = 0, 1, 2, \dots, p$, obtenemos la siguiente expresión de la ecuación (4.24):

$$(1 - a_1 L - a_2 L^2 - a_3 L^3 - \dots - a_p L^p) X_t = a_0 + U_t \quad (4.25)$$

Definamos el polinomio $\alpha(L)$ como:

$$\alpha(L) = 1 - a_1 L - a_2 L^2 - a_3 L^3 - \dots - a_p L^p \quad (4.26)$$

De forma similar que en los procesos $AR(1)$ y $AR(2)$, las condiciones de estabilidad del proceso $AR(p)$ estarán dadas por la solución de la ecuación característica:

$$\lambda^p - a_1 \lambda^{p-1} - a_2 \lambda^{p-2} - a_3 \lambda^{p-3} - \dots - a_p = 0 \quad (4.27)$$

Así, solo si el polinomio anterior tiene raíces cuyo valor absoluto sea menor a uno ($|\alpha_i| < 1$) y si $1 - a_1L - a_2L^2 - a_3L^3 - \dots - a_pL^p < 1$ podremos decir que el proceso es convergente y estable. Lo anterior significa que la ecuación (4.26) puede expresarse en términos de la descomposición de Wold o como la suma infinita de términos como:

$$\frac{1}{1 - a_1L - a_2L^2 - a_3L^3 - \dots - a_pL^p} = \psi_0 + \psi_1L + \psi_2L^2 + \psi_3L^3 + \dots \quad (4.28)$$

Donde, por construcción de $\alpha(L)\alpha^{-1}(L) = 1$ implica que $\psi_0 = 1$. De forma similar a los proceso AR(1) y AR(2), es posible determinar el valor de los coeficientes ψ_j en términos de los coeficientes a_i . Así, la solución del proceso $AR(p)$ estará dada por:

$$X_t = \frac{a_0}{1 - a_1 - a_2 - a_3 - \dots - a_p} + \sum_{j=0}^{\infty} \psi_j U_{t-j} \quad (4.29)$$

Considerando la solución de la ecuación ((4.24)) expresada en la ecuación (4.29) podemos determinar los momentos del proceso y que estarán dados por una media como:

$$\mathbb{E}[X_t] = \mu = \frac{a_0}{1 - a_1 - a_2 - a_3 - \dots - a_p} \quad (4.30)$$

Lo anterior, considerando que $\mathbb{E}[U_t] = 0$, para todo t . Para determinar la varianza del proceso, sin pérdida de generalidad, podemos definir una ecuación: $\gamma(\tau) = \mathbb{E}[X_{t-\tau}X_t]$, la cual (omitiendo la constante, ya que la correlación de una constante con cualquier variable aleatoria que depende del tiempo es cero (0)) puede ser escrita como:

$$\gamma(\tau) = \mathbb{E}[(X_{t-\tau}) \cdot (a_1X_{t-1} + a_2X_{t-2} + a_3X_{t-3} + \dots + a_pX_{t-p} + U_t)] \quad (4.31)$$

Donde $\tau = 0, 1, 2, \dots, p$ y $a_0 = 0$, lo que implica que $\mu = 0$. De lo anterior obtenemos el siguiente conjunto de ecuaciones mediante sustituciones de los valores de τ :

$$\begin{aligned} \gamma(0) &= a_1\gamma(1) + a_2\gamma(2) + \dots + a_p\gamma(p) + \sigma^2 \\ \gamma(1) &= a_1\gamma(0) + a_2\gamma(1) + \dots + a_p\gamma(p-1) \\ &\vdots \\ \gamma(p) &= a_1\gamma(p-1) + a_2\gamma(p-2) + \dots + a_p\gamma(0) \end{aligned}$$

De esta forma, es fácil observar que la ecuación general para $p > 0$ estará dada por:

$$\gamma(p) - a_1\gamma(\tau-1) - a_2\gamma(\tau-2) - \dots - a_p\gamma(\tau-p) = 0 \quad (4.32)$$

Dividiendo la ecuación (4.32) por $\gamma(0)$, se obtiene la siguiente ecuación:

$$\rho(p) - a_1\rho(\tau-1) + a_2\rho(\tau-2) + \dots + a_p\rho(\tau-p) = 0 \quad (4.33)$$

Así, podemos escribir el siguiente sistema de ecuaciones:

$$\begin{aligned}\rho(1) &= a_1 + a_2\rho(1) + a_3\rho(2) + \dots + a_p\rho(p-1) \\ \rho(2) &= a_1\rho(1) + a_2 + a_3\rho(1) + \dots + a_p\rho(p-2) \\ &\vdots \\ \rho(p) &= a_1\rho(p-1) + a_2\rho(p-2) + \dots + a_p\end{aligned}$$

Lo anterior se puede expresar como un conjunto de vectores y matrices de la siguiente forma:

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \rho(2) & \rho(1) & \dots & \rho(p-3) \\ \vdots & \vdots & \dots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} \quad (4.34)$$

De lo anterior podemos escribir la siguiente ecuación que es una forma alternativa para expresar los valores de los coeficientes a_i de la solución del proceso $AR(p)$:

$$\rho = \mathbf{R}\mathbf{a} \quad (4.35)$$

Es decir, podemos obtener la siguiente expresión:

$$\mathbf{a} = \mathbf{R}^{-1}\rho \quad (4.36)$$

Ahora veámos un ejemplo. Utilizaremos la serie de Pasajeros en vuelos internacionales de salida para estimar un $AR(p)$ mediante el método de máxima verosimilitud (ML). Antes de realizar el proceso de estimación consideremos una transformación de la serie en logaritmos y una más en diferencias logarítmicas; lo anterior con el objeto de obtener un conjunto de series de tiempo suavizada y expresada en tasas de crecimiento, con un comportamiento parecido a un proceso estacionario.

Primero, para decidir si se realizará un $AR(p)$ para la serie en niveles o en diferencias analizaremos su gráfica. La serie de Pasajeros en vuelos internacionales de salidas se muestra en la Figura 4.9. En está se muestra la gráfica de la serie en niveles (sin transformación logarítmica y con transformación logarítmica) y en diferencias logarítmicas mensuales (es decir, con diferencia respecto del mes inmediato anterior).

```
library(ggplot2)
library(dplyr)
library(readxl)
library(stats)
```



```

Datos <- read_excel("BD/Base_Transporte.xlsx",
                    sheet = "Datos", col_names = TRUE)

# En Niveles
Pax_Int <- ts(Datos$Pax_Int,
              start = c(2000, 1),
              freq = 12)

# Logaritmos:
LPax_Int <- ts(log(Datos$Pax_Int),
               start = c(2000, 1),
               freq = 12)

# Diferencias mensuales:
DLPax_Int <- ts(log(Datos$Pax_Int) - lag(log(Datos$Pax_Int), 1),
                start = c(2000, 1),
                freq = 12)

#
par(mfrow=c(3,1))

plot(Pax_Int, xlab = "Tiempo", ylab = "Pasajeros",
     main = "Pasajeros en vuelos internacionales de salida",
     col = "darkgreen")

plot(LPax_Int, xlab = "Tiempo", ylab = "LN Pasajeros",
     main = "LN Pasajeros en vuelos internacionales de salida",
     col = "darkblue")

plot(DLPax_Int, xlab = "Tiempo", ylab = "DLN Pasajeros",
     main = "Diff LN Pasajeros en vuelos internacionales de salida",
     col = "darkred")

par(mfrow=c(1,1))

```

De la gráfica en la Figura 4.9 observamos que quizá la mejor forma de estimar un $AR(p)$ es mediante la serie en diferencias, ya que ésta es la que parece ser una serie estacionaria. A continuación, estimaremos una $AR(4)$ para la serie en diferencias logarítmicas ($DLPaxInt_t$):

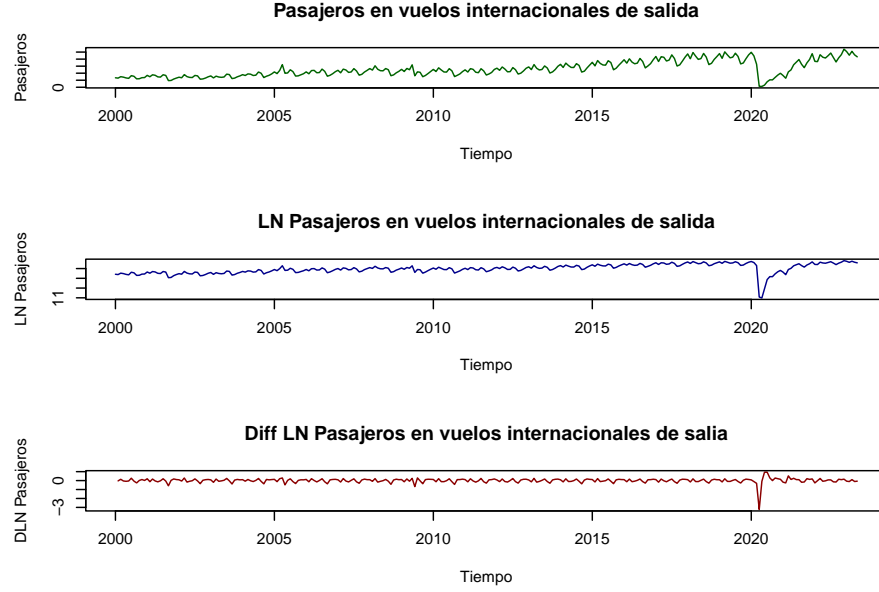


Figure 4.9: Pasajeros en vuelos internacionales de salida en niveles y en diferencias logarítmicas

$$\begin{aligned}
 DLPaxInt_t = & 0.0050 - 0.2701 DLPaxInt_{t-1} \\
 & (0.0052) \quad (0.0655) \\
 & - 0.4326 DLPaxNal_{t-2} \\
 & \quad (0.0664) \\
 & - 0.1956 DLPaxNal_{t-3} \\
 & \quad (0.0664) \\
 & - 0.0316 DLPaxNal_{t-4} \\
 & \quad (0.0653)
 \end{aligned}$$

$$\hat{\sigma}^2 = 0.02371 \quad \text{y} \quad AIC = -198.16$$

Entre parentésis indicamos los errores estándar y reportamos el estadístico de Akaike, AIC. Finalmente, podemos determinar si las soluciones serán convergentes, para ello en la Figura 4.10 mostramos las raíces asociadas a cada uno de los polinomios. De la inspección visual podemos concluir que el AR(4) representan una solución convergente y estable.

```
source("arroots.R")
source("plot.armarroots.R")
```

```

AR_LPax_Int <- arima(LPax_Int, order = c(4, 0, 0),
                     method = "ML")

AR_DLPax_Int <- arima(DLPax_Int, order = c(4, 0, 0),
                     method = "ML")

par(mfrow=c(1,2))

plot.armacoots(aroots(AR_LPax_Nal),
               main="Inverse AR roots of \nAR(2): LN Pax Int")

#
plot.armacoots(aroots(AR_DLPax_Nal),
               main="Inverse AR roots of \nAR(2): Diff LN Pax Int")

```

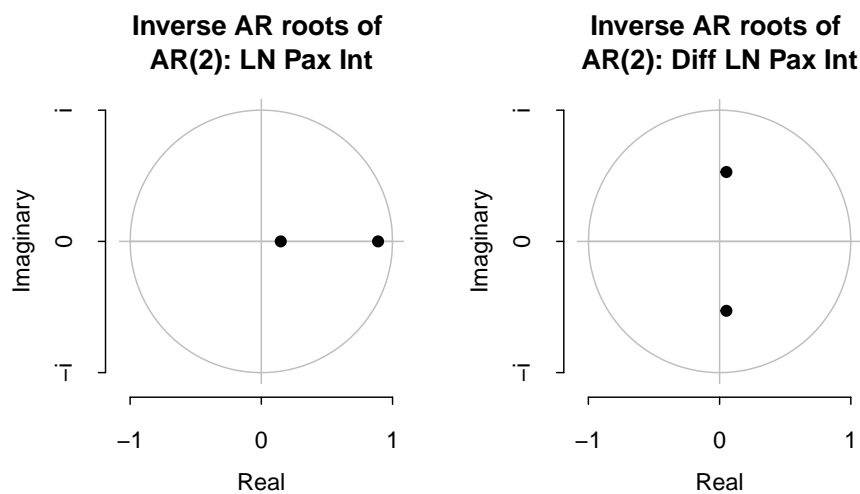


Figure 4.10: Inverso de las Raíces del polinomio característico

```

par(mfrow=c(1,1))

```

4.2 Procesos de Medias Móviles (MA)

4.2.1 MA(1)

Una vez planteado el proceso generalizado de $AR(p)$, iniciamos el planteamiento de los procesos de medias móviles, denotados como $MA(q)$. Iniciemos con el planteamiento del proceso $MA(1)$, que se puede escribir como una ecuación como la siguiente:

$$X_t = \mu + U_t - b_1 U_{t-1} \quad (4.37)$$

O como:

$$X_t - \mu = (1 - b_1 L)U_t \quad (4.38)$$

Donde U_t es un proceso puramente aleatorio, es decir, con $\mathbb{E}[U_t] = 0$, $Var[U_t] = \sigma^2$, y $Cov[U_t, U_s] = 0$. Así, un proceso $MA(1)$ puede verse como un proceso AR con una descomposición de Wold en la que $\psi_0 = 1$, $\psi_1 = -b_1$ y $\psi_j = 0$ para todo $j > 1$.

Al igual que los procesos autoregresivos, determinaremos los momentos de un proceso $MA(1)$. En el caso de la media observamos que será:

$$\begin{aligned} \mathbb{E}[X_t] &= \mu + \mathbb{E}[U_t] - b_1 \mathbb{E}[U_{t-1}] \\ &= \mu \end{aligned} \quad (4.39)$$

Por su parte la varianza estará dada por:

$$\begin{aligned} Var[X_t] &= \mathbb{E}[(X_t - \mu)^2] \\ &= \mathbb{E}[(U_t - b_1 U_{t-1})^2] \\ &= \mathbb{E}[U_t^2 - 2b_1 U_t U_{t-1} + b_1^2 U_{t-1}^2] \\ &= \mathbb{E}[U_t^2] - 2b_1 \mathbb{E}[U_t U_{t-1}] + b_1^2 \mathbb{E}[U_{t-1}^2] \\ &= \sigma^2 + b_1^2 \sigma^2 \\ &= (1 + b_1^2) \sigma^2 = \gamma(0) \end{aligned} \quad (4.40)$$

De esta forma, la varianza del proceso es constante en cualquier periodo t . Para determinar la covarianza utilizaremos la siguiente ecuación:

$$\begin{aligned} \mathbb{E}[(x_t - \mu)(x_{t+\tau} - \mu)] &= \mathbb{E}[(U_t - b_1 U_{t-1})(U_{t+\tau} - b_1 U_{t+\tau-1})] \\ &= \mathbb{E}[U_t U_{t+\tau} - b_1 U_t U_{t+\tau-1} - b_1 U_{t-1} U_{t+\tau} \\ &\quad + b_1^2 U_{t-1} U_{t+\tau-1}] \\ &= \mathbb{E}[U_t U_{t+\tau}] - b_1 \mathbb{E}[U_t U_{t+\tau-1}] \\ &\quad - b_1 \mathbb{E}[U_{t-1} U_{t+\tau}] + b_1^2 \mathbb{E}[U_{t-1} U_{t+\tau-1}] \end{aligned} \quad (4.41)$$

Si hacemos sustituciones de diferentes valores de τ en la ecuación (4.41) notaremos que la covarianza será distinta de cero únicamente para el caso de $\tau = 1, -1$. En ambos casos tendremos como resultado:

$$\begin{aligned}
 \mathbb{E}[(x_t - \mu)(x_{t+1} - \mu)] &= \mathbb{E}[(x_t - \mu)(x_{t-1} - \mu)] \\
 &= -b_1 \mathbb{E}[U_t U_t] \\
 &= -b_1 \mathbb{E}[U_{t-1} U_{t-1}] \\
 &= -b_1^2 \sigma^2 = \gamma(1)
 \end{aligned} \tag{4.42}$$

De esta forma tendremos que las funciones de autocorrelación estarán dadas por los siguientes casos:

$$\begin{aligned}
 \rho(0) &= 1 \\
 \rho(1) &= \frac{-b_1}{1 + b_1^2} \\
 \rho(\tau) &= 0 \text{ para todo } \tau > 1
 \end{aligned}$$

Ahora regresando a la ecuación (4.37), su solución la podemos expresar como:

$$\begin{aligned}
 U_t &= -\frac{\mu}{1 - b_1} + \frac{1}{1 - b_1 L} X_t \\
 &= -\frac{\mu}{1 - b_1} + X_t + b_1 X_{t-1} + b_1^2 X_{t-2} + \dots
 \end{aligned}$$

Donde la condición para que se cumpla esta ecuación es que $|b_1| < 1$. La manera de interpretar esta condición es como una condición de estabilidad de la solución y como una condición de invertibilidad. Notemos que un $MA(1)$ (y en general un $MA(q)$) es equivalente a un $AR(\infty)$, es decir, cuando se invierte un MA se genera un AR con infinitos rezagos.

En esta sección no desarrollaremos un ejemplo, primero explicaremos en qué consiste una modelación del tipo $MA(q)$ y después platearemos un ejemplo en concreto.

4.2.2 MA(q)

En general, el proceso de medias móviles de orden q , $MA(q)$, puede ser escrito como:

$$X_t = \mu + U_t - b_1 U_{t-1} - b_2 U_{t-2} - \dots - b_q U_{t-q} \tag{4.43}$$

Podemos reescribir la ecuación (4.43) utilizando el operador rezago, así tendremos el proceso de $MA(q)$ como:

$$\begin{aligned}
 X_t - \mu &= (1 - b_1 L - b_2 L^2 - \dots - b_q L^q) U_t \\
 X_t - \mu &= \beta(L) U_t
 \end{aligned} \tag{4.44}$$

Donde U_t es un proceso puramente aleatorio con $\mathbb{E}[U_t] = 0$, $Var[U_t] = \mathbb{E}[U_t^2] = \sigma^2$ y $Cov[U_t, U_s] = \mathbb{E}[U_t U_s] = 0$, y $\beta(L) = 1 - b_1 L - b_2 L^2 - \dots - b_q L^q$ es un polinomio del operador rezago L . la ecuación (4.44) puede ser interpretada como un proceso $AR(q)$ sobre la serie U_t .

Ahora determinemos los momentos de un proceso $MA(q)$:

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[\mu + U_t - b_1 U_{t-1} - b_2 U_{t-2} - \dots - b_q U_{t-q}] \\ &= \mu + \mathbb{E}[U_t] - b_1 \mathbb{E}[U_{t-1}] - b_2 \mathbb{E}[U_{t-2}] - \dots - b_q \mathbb{E}[U_{t-q}] \\ &= \mu\end{aligned}\tag{4.45}$$

En el caso de la varianza tenemos que se puede expresar como:

$$\begin{aligned}Var[X_t] &= \mathbb{E}[(X_t - \mu)^2] \\ &= \mathbb{E}[(U_t - b_1 U_{t-1} - b_2 U_{t-2} - \dots - b_q U_{t-q})^2] \\ &= \mathbb{E}[U_t^2 + b_1^2 U_{t-1}^2 + b_2^2 U_{t-2}^2 + \dots + b_q^2 U_{t-q}^2 \\ &\quad - 2b_1 U_t U_{t-1} - \dots - 2b_{q-1} b_q U_{t-q+1} U_{t-q}] \\ &= \mathbb{E}[U_t^2] + b_1^2 \mathbb{E}[U_{t-1}^2] + b_2^2 \mathbb{E}[U_{t-2}^2] + \dots + b_q^2 \mathbb{E}[U_{t-q}^2] \\ &\quad - 2b_1 \mathbb{E}[U_t U_{t-1}] - \dots - 2b_{q-1} b_q \mathbb{E}[U_{t-q+1} U_{t-q}] \\ &= \sigma^2 + b_1^2 \sigma^2 + b_2^2 \sigma^2 + \dots + b_q^2 \sigma^2 \\ &= (1 + b_1^2 + b_2^2 + \dots + b_q^2) \sigma^2\end{aligned}\tag{4.46}$$

En el caso de las covarianzas podemos utilizar una idea similar al caso del $AR(p)$, construir una expresión general para cualquier rezago τ :

$$\begin{aligned}Cov[X_t, X_{t+\tau}] &= \mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)] \\ &= \mathbb{E}[(U_t - b_1 U_{t-1} - b_2 U_{t-2} - \dots - b_q U_{t-q}) \\ &\quad (U_{t+\tau} - b_1 U_{t+\tau-1} - b_2 U_{t+\tau-2} - \dots - b_q U_{t+\tau-q})]\end{aligned}$$

La expresión anterior se puede desarrollar para múltiples casos de $\tau = 1, 2, \dots, q$. De esta forma tenemos el siguiente sistema:

$$\begin{aligned}\tau = 1 & : \gamma(1) = (-b_1 + b_1 b_2 + \dots + b_{q-1} b_q) \sigma^2 \\ \tau = 2 & : \gamma(2) = (-b_2 + b_1 b_3 + \dots + b_{q-2} b_q) \sigma^2 \\ & \vdots \\ \tau = q & : \gamma(q) = b_q \sigma^2\end{aligned}$$

Donde $\gamma(\tau) = 0$ para todo $\tau > q$. Es decir, todas las autocovarianzas y autocorrelaciones con ordenes superiores a q son cero (0). De esta forma, esta característica teórica permite identificar el orden de $MA(q)$ visualizando la función de autocorrelación y verificando a partir de cual valor de rezago la autocorrelación es no significativa.

Regresando al problema original que es el de determinar una solución para la ecuación (4.43), tenemos que dicha solución estará dada por un $AR(\infty)$ en términos de U_t :

$$U_t = -\frac{\mu}{1-b_1-b_2-\dots-b_q} + \beta(L)^{-1}X_t - \frac{\mu}{1-b_1-b_2-\dots-b_q} + \sum_{j=0}^{\infty} c_j X_{t-j} \quad (4.47)$$

Donde se cumple que: $1 = (1 - b_1L^1 - b_2L^2 - \dots - b_qL^q)(1 - c_1L - c_2L^2 - \dots)$ y los coeficientes c_j se pueden determinar por un método de coeficientes indeterminados y en términos de los valores b_i . De igual forma que en el caso de la ecuación (4.24), en la ecuación (4.47) se deben cumplir condiciones de estabilidad asociadas con las raíces del polinomio característico dado por:

$$1 - b_1x - b_2x^2 - \dots - b_qx^q = 0 \quad (4.48)$$

El cual debe cumplir que $|x_i| < 1$ y que $1 - b_1 - b_2 - \dots - b_q < 1$.

Ahora veamos un ejemplo del proceso $MA(q)$, para lo cual retomaremos la serie de Pasajeros transportados en el metro de la CDMX (*PaxMetro*). Estimaremos el $MA(q)$ mediante el método de máxima verosimilitud (ML). Antes de realizar el proceso de estimación consideremos una transformación de la serie en logaritmos y una más en diferencias logarítmicas; lo anterior con el objeto de obtener un conjunto de series de tiempo suavizada y expresada en tasas de crecimiento, con un comportamiento parecido a un proceso estacionario.

La serie de Pasajeros transportados en el metro de la CDMX se muestra en la Figura 4.6 se muestra la gráfica de la serie en niveles (sin transformación logarítmica y con transformación logarítmica) y en diferencias logarítmicas mensuales (es decir, con una diferencia respecto del mes inmediato anterior). Utilizaremos la serie en diferencias, ya que es la que parece ser estacionaria. Esta serie tiene la peculiaridad de que tiene un salto a la baja y uno al alza entre septiembre de 2017 y octubre de 2017. Para controlar ese efecto, en nuestro modelo $MA(q)$ incluiremos dos variables dummies para dichos meses.

A continuación, estimaremos una $MA(4)$ para la serie en diferencias:

$$\begin{array}{rcll}
DLPaxMetro_t & = & 0.0000 & - \quad 0.7804 \quad U_{t-1} \\
& & (0.0000) & (0.0661) \\
& & & + \quad 0.3591 \quad U_{t-2} \\
& & & (0.0826) \\
& & & - \quad 0.2775 \quad U_{t-3} \\
& & & (0.0908) \\
& & & - \quad 0.1120 \quad U_{t-4} \\
& & & (0.0769) \\
& & & - \quad 0.3789 \quad D_{Sep2017} \\
& & & (0.0436) \\
& & & + \quad 0.3695 \quad D_{Oct2017} \\
& & & (0.0434)
\end{array}$$

$$\hat{\sigma}^2 = 0.002176 \quad \text{y} \quad AIC = -749.94$$

Entre parentésis indicamos los errores estándar y al final reportamos el estadístico de Akaike, AIC. Finalmente, podemos determinar si la solución serán convergente, para ello en la Figura 4.11 mostramos las raíces asociadas a cada uno de los polinomios. De la inspección visual podemos concluir que ambas propuesta de AR(2) representan una solución convergente y estable.

```

library(ggplot2)
library(dplyr)
library(readxl)
library(stats)

Datos <- read_excel("BD/Base_Transporte_ARIMA.xlsx",
                    sheet = "Datos", col_names = TRUE)

source("maroots.R")

source("plot.armaroots.R")

Pax_Metro <- ts(Datos$Pax_Metro,
               start = c(2000, 1),
               freq = 12)

LPax_Metro <- ts(log(Datos$Pax_Metro),
               start = c(2000, 1),
               freq = 12)

DLPax_Metro <- ts(log(Datos$Pax_Metro) -
               lag(log(Datos$Pax_Metro), 1),
               start = c(2000, 1),
               freq = 12)

```



```

D_Sep2017 <- ts(Datos$D_Sep2017,
               start = c(2000, 1),
               freq = 12)

D_Abr2020 <- ts(Datos$D_Abr2020,
               start = c(2000, 1),
               freq = 12)

D_May2020 <- ts(Datos$D_May2020,
               start = c(2000, 1),
               freq = 12)

MA_LPax_Metro <- arima(LPax_Metro, order = c(0, 0, 4),
                      method = "ML")

MA_LPax_Metro_2 <- arima(LPax_Metro, order = c(0, 0, 4),
                      xreg = cbind(D_Abr2020, D_Sep2017,
                                   D_May2020),
                      method = "ML")

par(mfrow=c(1,2))

plot.armacoots(maroots(MA_LPax_Metro),
               main="Inverse MA roots of \nMA(p): LN PAX Metro")

plot.armacoots(maroots(MA_LPax_Metro_2),
               main="Inverse MA roots of \nMA(p): LN PAX Metro con Dummy")

par(mfrow=c(1,1))

```

4.3 Procesos ARMA(p, q) y ARIMA(p, d, q)

Hemos establecido algunas relaciones las de los porcesos AR y los procesos MA, es decir, cómo un $MA(q)$ de la serie X_t puede ser reexpresada como un $AR(\infty)$ de la serie U_t , y viceversa un $AR(p)$ de la serie X_t puede ser reexpresada como un $MA(\infty)$.

En este sentido, para cerrar esta sección veámos el caso de la especificación que conjunta ambos modelos en un modelo general conocido como $ARMA(p, q)$ o $ARIMA(p, d, q)$. La diferencia entre el primero y el segundo es las veces que su tuvo que diferenciar la serie analizada, registro que se lleva en el índice d de los parámetros dentro del concepto $ARIMA(p, d, q)$. No obstante, en general nos referiremos al modelo como $ARMA(p, q)$ y dependerá del analista si modela

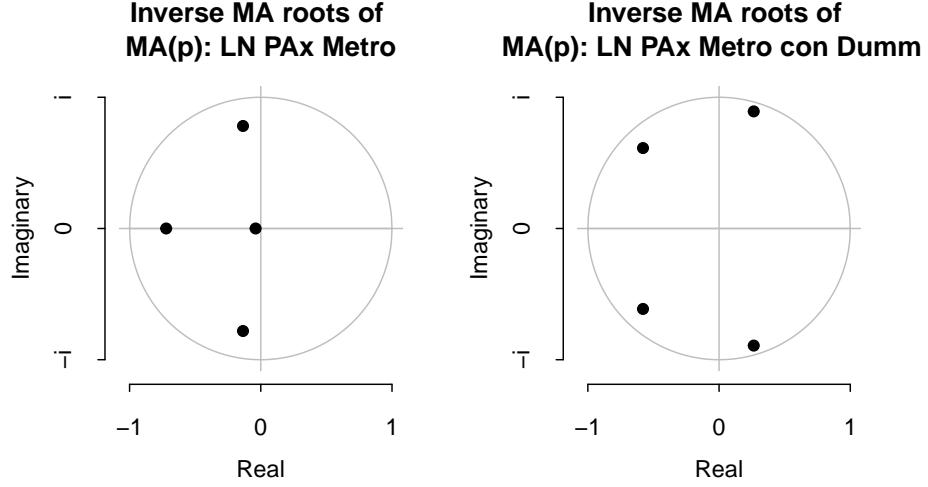


Figure 4.11: Inverso de las Raíces del polinomio característico

la serie en niveles (por ejemplo, en logaritmos) o en diferencias logarítmicas (o diferencias sin logaritmos).

4.3.1 ARMA(1, 1)

Dicho lo anterior vamos a empezar con el análisis de un $ARMA(1,1)$. Un proceso $ARMA(1,1)$ puede verse como:

$$X_t = \delta + a_1 X_{t-1} + U_t - b_1 U_{t-1} \quad (4.49)$$

Aplicando el operado rezago podemos reescribir la ecuación (4.49) como:

$$(1 - a_1 L)X_t = \delta + (1 - b_1 L)U_t \quad (4.50)$$

Donde U_t es un proceso puramente aleatorio como en los casos de $AR(p)$ y $MA(q)$, y X_t puede ser una serie en niveles o en diferencias (ambas, en términos logarítmicos).

Así, el modelo $ARIMA(p,q)$ también tiene una representación de Wold que estará dada por las siguientes expresiones:

$$X_t = \frac{\delta}{1 - a_1} + \frac{1 - b_1 L}{1 - a_1 L} U_t \quad (4.51)$$

Donde $a_1 \neq b_1$, puesto que en caso contrario X_t sería un proceso puramente aleatorio con una media $\mu = \frac{\delta}{1-a_1}$. Así, podemos reescribir la descomposición de Wold a partir del componente de la ecuación (4.51):

$$\frac{1-b_1L}{1-a_1L} = \psi_0 + \psi_1L + \psi_2L^2 + \psi_3L^3 + \dots \quad (4.52)$$

Está ecuación es equivalente a la expresión:

$$\begin{aligned} (1-b_1L) &= (1-a_1L)(\psi_0 + \psi_1L + \psi_2L^2 + \psi_3L^3 + \dots) \\ &= \psi_0 + \psi_1L + \psi_2L^2 + \psi_3L^3 + \dots \\ &\quad -a_1\psi_0L - a_1\psi_1L^2 - a_2\psi_2L^3 - a_1\psi_3L^4 - \dots \end{aligned}$$

De esta forma podemos establecer el siguiente sistema de coeficientes indeterminados:

$$\begin{array}{lll} L^0 : & & \Rightarrow \psi_0 = 1 \\ L^1 : & \psi_1 - a_1\psi_0 = -b_1 & \Rightarrow \psi_1 = a_1 - b_1 \\ L^2 : & \psi_2 - a_1\psi_1 = 0 & \Rightarrow \psi_2 = a_1(a_1 - b_1) \\ L^3 : & \psi_3 - a_1\psi_2 = 0 & \Rightarrow \psi_3 = a_1^2(a_1 - b_1) \\ \vdots & \vdots & \vdots \\ L^j : & \psi_j - a_1\psi_{j-1} = 0 & \Rightarrow \psi_j = a_1^{j-1}(a_1 - b_1) \end{array}$$

Así, la solución a la ecuación (4.49) estará dada por la siguiente generalización:

$$X_t = \frac{\delta}{1-a_1} + U_t + (a_1-b_1)U_{t-1} + a_1(a_1-b_1)U_{t-2} + a_1^2(a_1-b_1)U_{t-3} + \dots \quad (4.53)$$

En la ecuación (4.53) las condiciones de estabilidad y de invertibilidad del sistema (de un MA a un AR, y viceversa) estarán dadas por: $|a_1| < 1$ y $|b_1| < 1$. Adicionalmente, la ecuación (4.53) expresa cómo una serie que tiene un comportamiento $ARMA(1, 1)$ es equivalente a una serie modelada bajo un $MA(\infty)$.

Al igual que en los demás modelos, ahora determinaremos los momentos del proceso $ARMA(1, 1)$. La media estará dada por:

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[\delta + a_1X_{t-1} + U_t - b_1U_{t-1}] \\ &= \delta + a_1\mathbb{E}[X_{t-1}] \\ &= \frac{\delta}{1-a_1} \\ &= \mu \end{aligned} \quad (4.54)$$

Donde hemos utilizado que $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}] = \mu$. Es decir, la media de un $ARMA(1, 1)$ es idéntica a la de un $AR(1)$.

Para determinar la varianza tomaremos una estrategia similar a los casos de $AR(p)$ y $MA(q)$. Por lo que para todo $\tau \geq 0$, y suponiendo por simplicidad que $\delta = 0$ (lo que implica que $\mu = 0$) tendremos:

$$\begin{aligned}\mathbb{E}[X_{t-\tau}X_t] &= \mathbb{E}[(X_{t-\tau}) \cdot (a_1X_{t-1} + U_t - b_1U_{t-1})] \\ &= a_1\mathbb{E}[X_{t-\tau}X_{t-1}] + \mathbb{E}[X_{t-\tau}U_t] - b_1\mathbb{E}[X_{t-\tau}U_{t-1}]\end{aligned}\quad (4.55)$$

De la ecuación (4.55) podemos determinar una expresión para el caso de $\tau = 0$:

$$\begin{aligned}\mathbb{E}[X_tX_t] &= \gamma(0) \\ &= a_1\gamma(1) + \mathbb{E}[U_tX_t] - b_1\mathbb{E}[X_tU_{t-1}] \\ &= a_1\gamma(1) + \sigma^2 + b_1\mathbb{E}[U_{t-1}(a_1X_{t-1} + U_t - b_1U_{t-1})] \\ &= a_1\gamma(1) + \sigma^2 - b_1a_1\sigma^2 + b_1\sigma^2 \\ &= a_1\gamma(1) + (1 - b_1(a_1 - b_1))\sigma^2\end{aligned}\quad (4.56)$$

Para el caso en que $\tau = 1$:

$$\begin{aligned}\mathbb{E}[X_{t-1}X_t] &= \gamma(1) \\ &= a_1\gamma(0) + \mathbb{E}[X_{t-1}U_t] - b_1\mathbb{E}[X_{t-1}U_{t-1}] \\ &= a_1\gamma(0) - b_1\sigma^2\end{aligned}\quad (4.57)$$

Estas últimas expresiones podemos resolverlas como sistema para determinar los siguientes valores:

$$\gamma(0) = \frac{1 + b_1^2 - 2a_1b_1}{1 - a_1^2}\sigma^2 \quad (4.58)$$

$$\gamma(1) = \frac{(a_1 - b_1)(1 - a_1b_1)}{1 - a_1^2}\sigma^2 \quad (4.59)$$

En general para cualquier valor $\tau \geq 2$ tenemos que la autocovarianza y la función de autocorrelación serán:

$$\gamma(\tau) = a_1\gamma(\tau - 1) \quad (4.60)$$

$$\rho(\tau) = a_1\rho(\tau - 1) \quad (4.61)$$

Por ejemplo, para el caso de $\tau = 1$ tendríamos:

$$\rho(1) = \frac{(a_1 - b_1)(1 - a_1b_1)}{1 + b_1^2 - 2a_1b_1} \quad (4.62)$$

De esta forma, la función de autocorrelación oscilará en razón de los valores que tome a_1 y b_1 .

4.3.2 ARMA(p, q)

La especificación general de un $ARMA(p, q)$ (donde $p, q \in \mathbb{N}$) puede ser descrita por la siguiente ecuación:

$$\begin{aligned} X_t = & \delta + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} \\ & + U_t - b_1 U_{t-1} - b_2 U_{t-2} - \dots - b_q U_{t-q} \end{aligned} \quad (4.63)$$

Donde U_t es un proceso puramente aleatorio, y X_t puede ser modelada en niveles o en diferencias (ya sea en logaritmos o sin transformación logarítmica).

Mediante el uso del operador rezago se puede escribir la ecuación (4.63) como:

$$(1 - a_1 L - a_2 L^2 - \dots - a_p L^p) X_t = \delta + (1 - b_1 L - b_2 L^2 - \dots - b_q L^q) U_t \quad (4.64)$$

En la ecuación (4.64) definamos dos polinomios: $\alpha(L) = (1 - a_1 L - a_2 L^2 - \dots - a_p L^p)$ y $\beta(L) = (1 - b_1 L - b_2 L^2 - \dots - b_q L^q)$. Así, podemos reescribir la ecuación (4.64) como:

$$\alpha(L) X_t = \delta + \beta(L) U_t \quad (4.65)$$

Asumiendo que existe el polinomio inverso tal que: $\alpha(L)^{-1} \alpha(L) = 1$. La solución entonces puede ser escrita como:

$$\begin{aligned} X_t &= \alpha(L)^{-1} \delta + \alpha(L)^{-1} \beta(L) U_t \\ &= \frac{\delta}{1 - a_1 - a_2 - \dots - a_p} + \frac{\beta(L)}{\alpha(L)} U_t \\ &= \frac{\delta}{1 - a_1 - a_2 - \dots - a_p} + U_t + \psi_1 L U_t + \psi_2 L^2 U_t + \dots \end{aligned} \quad (4.66)$$

Donde la ecuación (4.66) nos permite interpretar que un $ARMA(p, q)$ se puede reexpresar e interpretar como un $MA(\infty)$ y donde las condiciones para la estabilidad de la solución y la invertibilidad es que las raíces de los polinomios característicos $\alpha(L)$ y $\beta(L)$ son en valor absoluto menores a 1.

Adicionalmente, la fracción en la ecuación (4.66) se puede descomponer como en la forma de Wold:

$$\frac{\beta(L)}{\alpha(L)} = 1 + \psi_1 L + \psi_2 L^2 + \dots \quad (4.67)$$

Bajo los supuestos de estacionariedad del componente U_t , los valores de la media y varianza de un proceso $ARMA(p, q)$ serán como describimos ahora. Para el caso de la media podemos partir de la ecuación (4.66) para generar:

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E} \left[\frac{\delta}{1 - a_1 - a_2 - \dots - a_p} + U_t + \psi_1 U_{t-1} + \psi_2 U_{t-2} + \dots \right] \\ &= \frac{\delta}{1 - a_1 - a_2 - \dots - a_p} \\ &= \mu \end{aligned} \quad (4.68)$$

Esta expresión indica que en general un proceso $ARMA(p, q)$ converge a una media idéntica a la de un proceso $AR(p)$. Para determinar la varianza utilizaremos la misma estrategia que hemos utilizado para otros modelos $AR(p)$ y $MA(q)$.

Sin pérdida de generalidad podemos asumir que $\delta = 0$, lo que implica que $\mu = 0$, de lo que podemos establecer una expresión de autocovarianzas para cualquier valor $\tau = 0, 1, 2, \dots$:

$$\begin{aligned}
 \gamma(\tau) &= \mathbb{E}[X_{t-\tau}X_t] \\
 &= \mathbb{E}[X_{t-\tau}(\delta + a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} \\
 &\quad + U_t - b_1U_{t-1} - b_2U_{t-2} - \dots - b_qU_{t-q})] \\
 &= a_1\gamma(\tau-1) + a_2\gamma(\tau-2) + \dots + a_p\gamma(\tau-p) \\
 &\quad + \mathbb{E}[X_{t-\tau}U_t] - b_1\mathbb{E}[X_{t-\tau}U_{t-1}] - \dots - b_q\mathbb{E}[X_{t-\tau}U_{t-q}] \quad (4.69)
 \end{aligned}$$

Ahora veámos un ejemplo. Utilizaremos la serie de Pasajeros en vuelos nacionales de salida para estimar un $ARMA(p, q)$ mediante el método de máxima verosimilitud (ML). Antes de realizar el proceso de estimación consideremos una transformación de la serie en diferencias logarítmicas, ya que según la gráfica en la Figura 4.7 esa es la que puede ser estacionaria.

A continuación, estimaremos una $ARMA(1, 1)$ para la serie en diferencias logarítmicas ($DLPaxNal_t$). También incorporaremos al análisis variables exógenas tales como dummies de estacionalidad. En particular, utilizaremos los meses de enero, febrero, julio y diciembre. No debe pasar desapercibido que un análisis de estacionalidad más formal debería considerar todos los meses para separar del término de error la parte que puede ser explicada por los ciclos estacionales.

Así obtenemos el siguiente resultado:

$$\begin{array}{rclcl}
 DLPaxNal_t & = & -0.0025 & + & 0.6018 & DLPaxNal_{t-1} \\
 & & (0.0045) & & (0.0964) & \\
 & & & - & 0.9064 & U_{t-1} \\
 & & & & (0.0397) & \\
 & & & - & 0.0867 & DEne_t \\
 & & & & (0.0230) & \\
 & & & - & 0.0409 & DFeb_t \\
 & & & & (0.0263) & \\
 & & & + & 0.1529 & DJul_t \\
 & & & & (0.0245) & \\
 & & & + & 0.0628 & DDic_t \\
 & & & & (0.0223) &
 \end{array}$$

$$\hat{\sigma}^2 = 0.007096 \quad \text{y} \quad AIC = -475.12$$

Donde entre parentésis indicamos los errores estándar. Adicionalmente, reportamos el estadístico de Akaike (AIC). Finalmente, podemos determinar si las soluciones serán convergentes, para ello en la Figura 4.12 mostramos las raíces asociadas a cada uno de los polinomios. De la inspección visual podemos concluir que tenemos una solución convergente y estable. Por su parte la Figura 4.13 muestra los residuales de la estimación del $ARMA(1,1)$.

```
knitr::include_graphics("Plots/G_Roots_ARMA11.png")
```

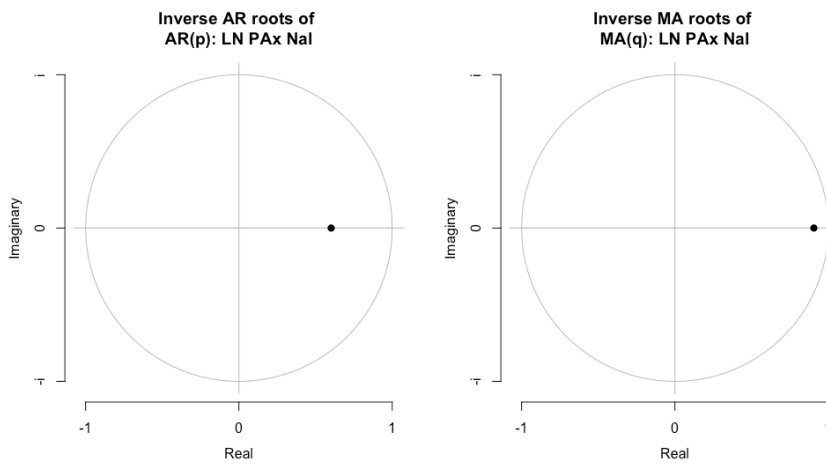


Figure 4.12: Inverso de las Raíces del polinomio característico de un $ARMA(1,1)$

```
knitr::include_graphics("Plots/G_Residuals_ARMA11.png")
```

En lo que resta de este capítulo, utilizaremos la serie en diferencias logarítmicas de los pasajeros en vuelos nacionales de salida, $DLPaxNaI_t$, para discutir los ejemplos que ilustran cada uno de los puntos teóricos que a continuación exponemos.

4.4 Función de Autocorrelación Parcial

Ahora introduciremos el concepto de Función de Autocorrelación Parcial (PACF, por sus siglas en inglés). Primero, dadas las condiciones de estabilidad y de convergencia, si suponemos que un proceso AR, MA, ARMA o ARIMA tienen toda la información de los rezagos de la serie en conjunto y toda la información de los promedios móviles del término de error, resulta importante

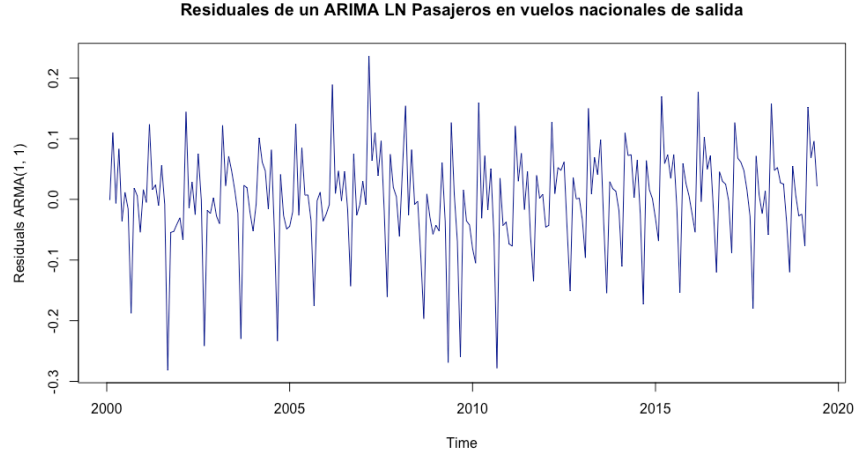


Figure 4.13: Residuales de un $ARMA(1, 1)$ de la serie $DLPaxNal_t$

construir una métrica para distinguir el efecto de $X_{t-\tau}$ o el efecto de $U_{t-\tau}$ (para cualquier τ) sobre X_t de forma individual.

La idea es construir una métrica de la correlación que existe entre las diferentes variables aleatorias, si para tal efecto se ha controlado el efecto del resto de la información. Así, podemos definir la ecuación que puede responder a este planteamiento como:

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \dots + \phi_{kk}X_{t-k} + U_t \quad (4.70)$$

Donde ϕ_{ki} es el coeficiente de la variable dada con el rezago i si el proceso tiene un orden k . Así, los coeficientes ϕ_{kk} son los coeficientes de la autocorrelación parcial (considerando un proceso $AR(k)$). Observemos que la autocorrelación parcial mide la correlación entre X_t y X_{t-k} que se mantiene cuando el efecto de las variables X_{t-1} , X_{t-2} , ... y X_{t-k-1} en X_t y X_{t-k} ha sido eliminado.

Dada la expresión considerada en la ecuación (4.70), podemos resolver el problema de establecer el valor de cada ϕ_{ki} mediante la solución del sistema que se representa en lo siguiente:

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & \dots & \rho(k-3) \\ \vdots & \vdots & \dots & \vdots \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} \quad (4.71)$$

Del cual se puede derivar una solución, resolviendo por el método de cramer, o cualquier otro método que consideremos y que permita calcular la solución de sistemas de ecuaciones.

4.5. SELECCIÓN DE LAS CONSTANTES P, Q, D EN UN $AR(P)$, UN $MA(Q)$, UN $ARMA(P, Q)$ O UN $ARIMA(P, D, Q)$

| Modelo si: | Función de Autocorrelación | Función de Autocorrelación Parcial |
|---------------|----------------------------|---------------------------------------|
| MA(q) | Rompimientos en la función | No hay rompimientos |
| AR(p) | No hay rompimientos | Rompimientos en la función |

Table 4.1: Relación entre la Función de Autocorrelación y la Función de Autocorrelación Parcial de una serie X_t .

Posterior al análisis analítico plantearemos un enfoque para interpretar las funciones de autocorrelación y autocorrelación parcial. Este enfoque pretende aportar al principio de parcimonia, en el cual podemos identificar el número de parámetros que posiblemente puede describir mejor a la serie en un modelo $ARMA(p, q)$.

En el siguiente cuadro mostramos un resumen de las características que debemos observar para determinar el número de parámetros de cada uno de los componentes AR y MA. Lo anterior por observación de las funciones de autocorrelación y autocorrelación parcial. Este enfoque no es el más formal, más adelante implementaremos uno más formal y que puede ser más claro de cómo determinar el número de parámetros.

Continuando con el ejemplo en la Figura 4.14 mostramos tanto la Función de Autocorrelación como la Función de Autocorrelación Parcial. En esta identificamos que ambas gráficas muestran que el modelo que explica a la variable $DLPaxNal_t$ tiene tanto componentes AR como MA. Sin embargo, dado lo errático del comportamiento de ambas funciones, resulta complicado determinar cuál sería un buen número de parámetros p y q a considerar en el $ARMA(p, q)$. Por esta razón a continuación plantearemos algunas pruebas más formales para determinar dichos parámetros.

```
knitr::include_graphics("Plots/G_ACF_PACF.png")
```

4.5 Selección de las constantes p, q, d en un $AR(p)$, un $MA(q)$, un $ARMA(p, q)$ o un $ARIMA(p, d, q)$

Respecto de cómo estimar un proceso $ARMA(p, q)$ –en general utilizaremos este modelo para discutir, pero lo planteado en esta sección es igualmente aplicable en cualquier otro caso como aquellos modelos que incluyen variables exógenas– existen diversas formas de estimar los parámetros a_i y b_i : i) por máxima verosimilitud y ii) por mínimos cuadrados ordinarios. El primer caso requiere que conozcamos la distribución del proceso aleatorio U_t . El segundo, por el contrario, no requiere

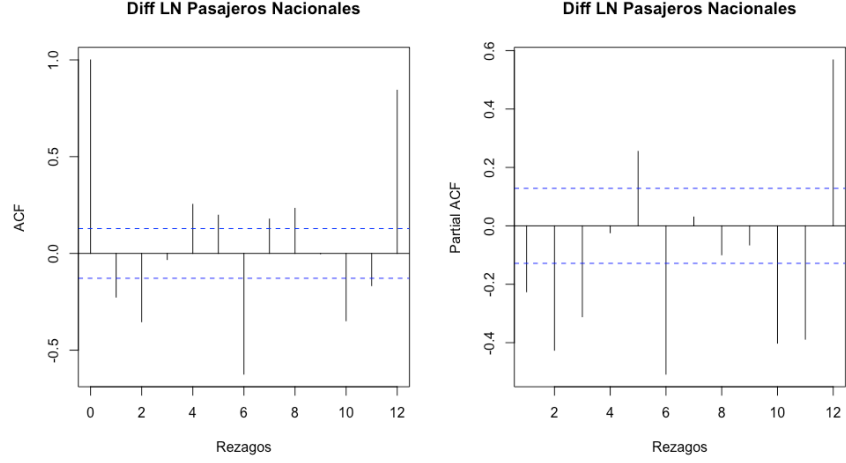


Figure 4.14: Función de Autocorrelación y la Función de Autocorrelación Parcial de una serie $DLPaxNal_t$

el mismo supuesto. No obstante, para el curso utilizaremos el método de máxima verosimilitud.

Otra duda que debe quedar hasta el momento es ¿cómo determinar el orden p y q del proceso ARMA(p , q)? La manera más convencional y formal que existe para tal efecto es utilizar los criterios de información. Así, el orden se elige de acuerdo a aquel criterio de información que resulta ser el mínimo. En el caso de d se selecciona revisando la gráfica que parezca más estacionaria—más adelante mostraremos un proceso más formal para su selección.

Los criterios de información más comunes son los siguientes:

1. FPE (Final Prediction Error):

$$FPE = \frac{T+m}{T-m} \frac{1}{T} \sum_{t=1}^T (\hat{U}_t^{(p)})^2 \quad (4.72)$$

2. Akaike:

$$AIC = \ln \left[\frac{1}{T} \sum_{t=1}^T (\hat{U}_t^{(p)})^2 \right] + m \frac{2}{T} \quad (4.73)$$

3. Schwarz:

$$SC = \ln \left[\frac{1}{T} \sum_{t=1}^T (\hat{U}_t^{(p)})^2 \right] + m \frac{\ln(T)}{T} \quad (4.74)$$

4.5. SELECCIÓN DE LAS CONSTANTES P, Q, D EN UN $AR(P)$, UN $MA(Q)$, UN $ARMA(P, Q)$ O UN $ARIMA(P, I, Q)$

4. Hannan - Quinn:

$$HQ = \ln \left[\frac{1}{T} \sum_{t=1}^T \left(\hat{U}_t^{(p)} \right)^2 \right] + m \frac{2 \ln(\ln(T))}{T} \quad (4.75)$$

Donde $\hat{U}_t^{(p)}$ son los residuales estimados mediante un proceso ARIMA y m es el número de parámetros estimados: $m = p + q + 0 + 1$ (ya que asumimos que $d = 0$). Una propiedad que no se debe perder de vista es que los criterios de información cumplen la siguiente relación:

$$\text{orden}(SC) \leq \text{orden}(HQ) \leq \text{orden}(AIC) \quad (4.76)$$

Por esta razón, durante el curso solo utilizaremos el criterio de Akaike para determinar el orden óptimo del proceso ARMA, ya que ello garantiza el orden más grande posible.

Ahora veamos un ejemplo de estimación del número de rezagos óptimo de un $ARMA(p, q)$. Retomemos la serie en diferencias logarítmicas de los pasajeros en vuelos nacionales de salidas, pero ahora incluiremos las variables exógenas de dummies estacionales: enero, febrero, julio y diciembre.

Como mencionamos, las gráficas de las funciones de autocorrelación permiten observar el valor de la correlación existente entre la variable en el momento t con cada uno de los rezagos. Incluso la Función de Autocorrelación Parcial puede ayudar a determinar el número máximo de rezagos que se debe incluir en el proceso $AR(p)$. No obstante, una métrica más formal es el uso de los criterios de información. En nuestro caso, dado lo discutido, sólo utilizaremos el criterio de Akaike.

Al respecto, en el siguiente cuadro reportamos el criterio de Akaike que resultan de aplicar dicho criterio a los residuales resultantes de cada combinación de procesos $ARMA(p, q)$. La forma de escoger será aquel modelo que reporta el criterio de Akaike menor. En la cuarta columna de la tabla se señala el valor del criterio de información que resulta ser el mínimo de todos los posibles.

El cuadro anterior reporta los resultados para 36 diferentes modelos, todos incluyen variables exógenas. Como resultado del análisis concluimos que el modelo más adecuado es el 24, el cual considera un $ARMA(4, 6)$, con variables dummies para controlar la estacionalidad de los meses de enero, febrero, julio y diciembre. Más adelante mostramos los resultados del modelo.

No obstante, una inspección de los residuales del modelo nos permite sospechar que requiere de incluir un par de dummies más. Ambas, asociadas con la caída del transporte aéreo en 2009, principalmente asociado con la crisis mundial de ese año. La Figura 4.15 muestra los residuales mencionados.

| Modelo | AR(p) | MA(p) | Akaike (AIC) | Óptimo |
|--------|-------|-------|--------------|--------|
| 1 | 1 | 1 | -475.1190 | |
| 2 | 1 | 2 | -473.4397 | |
| 3 | 1 | 3 | -483.1239 | |
| 4 | 1 | 4 | -482.4932 | |
| 5 | 1 | 5 | -506.9880 | |
| 6 | 1 | 6 | -533.9411 | |
| 7 | 2 | 1 | -473.4758 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 24 | 4 | 6 | -752.8996 | * |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 36 | 6 | 6 | -752.2632 | |

Table 4.2: Criterio de Akaike para diferentes modelos $ARMA(p, q)$ de la serie $DLPaxNal_t$.

```
knitr::include_graphics("Plots/G_Residuals_ARMA46.png")
```

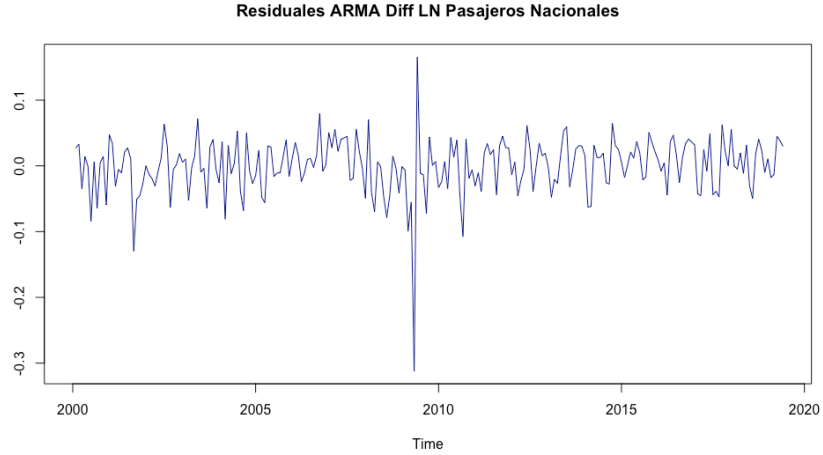


Figure 4.15: Residuales del $ARMA(4, 6)$ de una serie $DLPaxNal_t$

Una vez incluidas dos dummies más para mayo y junio de 2009, analizamos un total de 36 modelos ARMA y determinamos que el orden que minimiza el criterio de Akaike es un $ARMA(4, 6)$. El siguiente cuadro muestra los resultados para este nuevo modelo. No lo mostramos en esta sección, pero ambos modelos reportados tienen raíces de sus respectivos polinomios característicos menores a 1 en valor absoluto. En la Figura 4.16 mostramos los residuales ahora ajustados por las dummies de mayo y junio de 2009.

4.5. SELECCIÓN DE LAS CONSTANTES P, Q, D EN UN $AR(P)$, UN $MA(Q)$, UN $ARMA(P, Q)$ O UN $ARIMA(P, D, Q)$

| Variable | Modelo 1 | | Modelo 2 | |
|------------------|-------------|------------|-------------|------------|
| | Coeficiente | Error Est. | Coeficiente | Error Est. |
| Constante | 0.0134 | 0.0033 | 0.0145 | 0.0029 |
| $DLPaxNal_{t-1}$ | 0.0004 | 0.0011 | 0.0014 | 0.0016 |
| $DLPaxNal_{t-2}$ | 0.9997 | 0.0019 | 0.9988 | 0.0024 |
| $DLPaxNal_{t-3}$ | 0.0001 | 0.0012 | 0.0002 | 0.0014 |
| $DLPaxNal_{t-4}$ | -0.9997 | 0.0004 | -0.9993 | 0.0009 |
| \hat{U}_{t-1} | -0.2877 | 0.0670 | -0.1472 | 0.0633 |
| \hat{U}_{t-2} | -1.2028 | 0.0608 | -1.3323 | 0.0625 |
| \hat{U}_{t-3} | 0.2302 | 0.0765 | 0.1201 | 0.0756 |
| \hat{U}_{t-4} | 1.2085 | 0.0645 | 1.3142 | 0.0642 |
| \hat{U}_{t-5} | -0.2634 | 0.0708 | -0.1034 | 0.0717 |
| \hat{U}_{t-6} | -0.2171 | 0.0606 | -0.3864 | 0.0657 |
| $DEne_t$ | -0.2904 | 0.0172 | -0.2865 | 0.0156 |
| $DFeb_t$ | -0.1312 | 0.0170 | -0.1273 | 0.0140 |
| $DJul_t$ | 0.3303 | 0.0173 | 0.3164 | 0.0156 |
| Dic_t | -0.0118 | 0.0169 | -0.0138 | 0.0137 |
| $DMay2009_t$ | | | -0.3378 | 0.0360 |
| $DJun2009_t$ | | | 0.2371 | 0.0359 |
| $\hat{\sigma}^2$ | 0.001841 | | 0.001283 | |
| AIC | -752.9 | | -835.73 | |

Nota: El método de estimación fue máxima verosimilitud.

Table 4.3: Criterio de Akaike para diferentes modelos $ARMA(p, q)$ de la serie $DLPaxNal_t$.

```
knitr::include_graphics("Plots/G_Residuals_ARMA46_D.png")
```

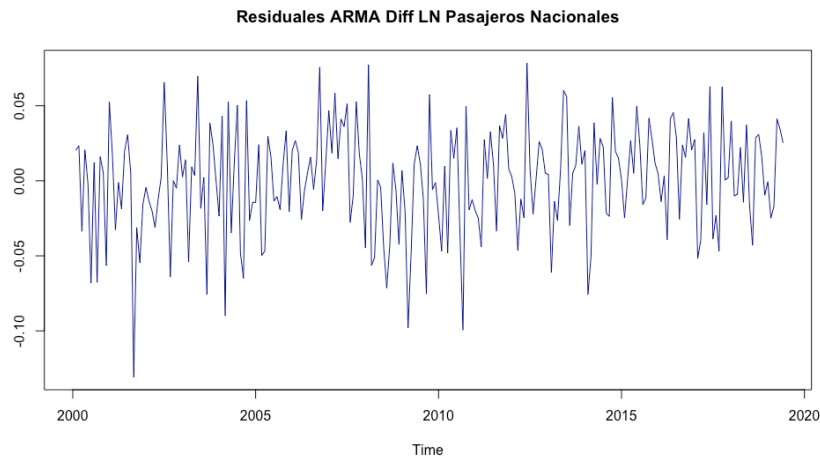


Figure 4.16: Residuales del ARMA(4, 6) de una serie $DLPaxNal_t$

4.6 Pronósticos

Para pronósticar el valor de la serie es necesario determinar cuál es el valor esperado de la serie en un momento $t + \tau$ condicional en que ésta se comporta como un $AR(p)$, un $MA(q)$ o un $ARMA(p, q)$ y a que los valores antes de t están dados. Por lo que el pronóstico de la serie estará dado por una expresión:

$$\mathbb{E}_t[X_{t+\tau}] = \delta + a_1 \mathbb{E}_t[X_{t+\tau-1}] + a_2 \mathbb{E}_t[X_{t+\tau-2}] + \dots + a_p \mathbb{E}_t[X_{t+\tau-p}] \quad (4.77)$$

Lo anterior para todo $\tau = 0, 1, 2, \dots$ y considerando que los componentes $MA(q)$ en la ecuación ((4.77)) son cero dado que para todo valor $t + \tau$ es cierto que $\mathbb{E}_t[U_{t+\tau}] = 0$.

Continuando con el ejemplo, en la Figura 4.17 mostramos el resultado del pronóstico de la serie a partir del modelo ARMA(4, 6) que hemos discutido anteriormente.

```
library(ggplot2)
library(dplyr)
library(readxl)
library(stats)
```

```

Datos <- read_excel("BD/Base_Transporte_ARIMA.xlsx",
                    sheet = "Datos", col_names = TRUE)

Pax_Nal <- ts(Datos$Pax_Nal,
              start = c(2000, 1),
              freq = 12)

LPax_Nal <- ts(log(Datos$Pax_Nal),
               start = c(2000, 1),
               freq = 12)

DLPax_Nal <- ts(log(Datos$Pax_Nal) - lag(log(Datos$Pax_Nal), 1),
                start = c(2000, 1),
                freq = 12)

D_Mar2020 <- ts(Datos$D_Mar2020,
                start = c(2000, 1),
                freq = 12)

D_Abr2020 <- ts(Datos$D_Abr2020,
                start = c(2000, 1),
                freq = 12)

D_Jun2020 <- ts(Datos$D_Jun2020,
                start = c(2000, 1),
                freq = 12)

D_Jul2020 <- ts(Datos$D_Jul2020,
                start = c(2000, 1),
                freq = 12)

D_Mar2021 <- ts(Datos$D_Mar2021,
                start = c(2000, 1),
                freq = 12)

D_Ene <- ts(Datos$D_Ene,
            start = c(2000, 1),
            freq = 12)

D_Feb <- ts(Datos$D_Feb,
            start = c(2000, 1),
            freq = 12)

D_Jul <- ts(Datos$D_Jul,
            start = c(2000, 1),

```

```

        freq = 12)

D_Dic <- ts(Datos$D_Dic,
           start = c(2000, 1),
           freq = 12)

ARMA_Ex_DLPax_Nal_2 <- arima(DLPax_Nal, order = c(6, 0, 6),
                           xreg = cbind(D_Ene, D_Feb, D_Jul,
                                         D_Dic, D_Mar2020,
                                         D_Abr2020, D_Jun2020,
                                         D_Jul2020, D_Mar2021),
                           method = "ML")

## Warning in arima(DLPax_Nal, order = c(6, 0, 6), xreg = cbind(D_Ene, D_Feb, :
## possible convergence problem: optim gave code = 1

Predict_Datos <- read_excel("BD/Predict_Base_Transporte_ARIMA.xlsx",
                           sheet = "Datos", col_names = TRUE)

D_Mar2020_f <- ts(Predict_Datos$D_Mar2020,
                 start = c(2022, 7),
                 freq = 12)

D_Abr2020_f <- ts(Predict_Datos$D_Abr2020,
                 start = c(2022, 7),
                 freq = 12)

D_Jun2020_f <- ts(Predict_Datos$D_Jun2020,
                 start = c(2022, 7),
                 freq = 12)

D_Jul2020_f <- ts(Predict_Datos$D_Jul2020,
                 start = c(2022, 7),
                 freq = 12)

D_Mar2021_f <- ts(Predict_Datos$D_Mar2021,
                 start = c(2022, 7),
                 freq = 12)

D_Ene_f <- ts(Predict_Datos$D_Ene,
              start = c(2022, 7),
              freq = 12)

D_Feb_f <- ts(Predict_Datos$D_Feb,
              start = c(2022, 7),

```



```

      freq = 12)

D_Jul_f <- ts(Predict_Datos$D_Jul,
             start = c(2022, 7),
             freq = 12)

D_Dic_f <- ts(Predict_Datos$D_Dic,
             start = c(2022, 7),
             freq = 12)

DLPax_Nal_f <- predict(ARMA_Ex_DLPax_Nal_2, n.ahead = 24,
                     newxreg = cbind(D_Ene_f, D_Feb_f, D_Jul_f,
                                       D_Dic_f, D_Mar2020_f,
                                       D_Abr2020_f, D_Jun2020_f,
                                       D_Jul2020_f, D_Mar2021_f))

Pronostico_Arima <- read_excel("BD/Pronostico_Arima.xlsx",
                              sheet = "Datos", col_names = TRUE)

Pronostico_Arima$Pax_Nal_f <- Pronostico_Arima$Pax_Nal

for(i in 1:24){
  Pronostico_Arima$Pax_Nal_f[281 + i] <-
    Pronostico_Arima$Pax_Nal_f[281 + i - 1]*(1 + DLPax_Nal_f$pred[i])
}

options(scipen = 999) # NO notacion cientifica

ggplot(data = Pronostico_Arima, aes(x = Periodo)) +
  geom_line(aes(y = Pax_Nal_f, color = "Pax_Nal_f")) +
  geom_line(aes(y = Pax_Nal, color = "Pax_Nal")) +
  scale_color_brewer(type = "qual", palette = 2) +
  #theme_bw() +
  theme(legend.position = "bottom") +
  theme(legend.title = element_blank()) +
  guides(col = guide_legend(nrow = 1, byrow = TRUE)) +
  xlab("Tiempo") +
  ylab("Pasajeros") +
  theme(plot.title = element_text(size = 11, face = "bold",
                                   hjust = 0)) +
  theme(plot.subtitle = element_text(size = 10, hjust = 0)) +
  theme(plot.caption = element_text(size = 10, hjust = 0)) +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  labs(
    title = "Pasajeros en vuelos nacionales (Salidas)",

```

```

subtitle = "(Ene-2000 a Jun-2019)",
caption = "Fuente: Elaboración propia"
)

```

```
## Warning: Removed 24 rows containing missing values (`geom_line()`).
```

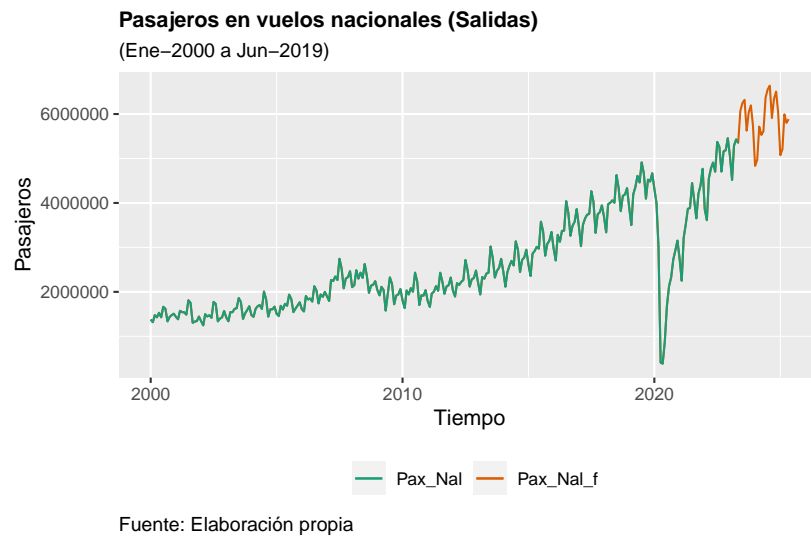


Figure 4.17: Pronóstico de la serie $PaxNAL_t$ a partir de una ARMA(4,6) en diferencias logarítmicas

```
ggsave("Pax_Nal_f.png", width = 20, height = 15, units = "cm")
```

```
## Warning: Removed 24 rows containing missing values (`geom_line()`).
```

Chapter 5

Desestacionalización y filtrado de Series

5.1 Motivación

Existen múltiples enfoques para la desestacionalización de series. Algunos modelos, por ejemplo, pueden estar basados en modelos ARIMA como un conjunto de dummies. No obstante, en el caso particular que discutiremos en este curso, estará basado en un modelo ARIMA de la serie. Este enfoque está basado en el modelo X11 de la oficina del censo de Estados Unidos (Census Bureau) el cual es conocido como el modelo X13-ARIMA-SEATS.¹ El modelo X13-ARIMA-SEATS es, como su nombre lo indica, la combinación de un modelo ARIMA con componentes estacionales (por la traducción literal de la palabra: seasonal).

Un modelo ARIMA estacional emplea la serie en diferencias y como regresores los valores rezagados de las diferencia de la serie tantas veces como procesos estacionales s existan en ésta, con el objeto de remover los efectos aditivos de la estacionalidad. Sólo para entender qué significa este mecanismo, recordemos que cuando se utiliza la primera diferencia de la serie respecto del periodo inmediato anterior se remueve la tendencia. Por su parte, cuando se incluye la diferencia respecto del mes s se está en el caso en que se modela la serie como una media móvil en términos del rezago s .

El modelo ARIMA estacional incluye como componentes autoregresivos y de medias móviles a los valores rezagados de la serie en el periodo s en diferencias. El ARIMA(p, d, q)(P, D, Q) estacional puede ser expresado de la siguiente manera utilizando el operador rezago L :

$$\Theta_P(L^s)\theta_p(L)(1-L^s)^D(1-L)^dX_t = \Psi_Q(L^s)\psi_q(L)U_t(\#eq : ARIMA_{seas}) \quad (5.1)$$

¹La información y material respecto del modelo esta disponible en la dirección <https://www.census.gov/srd/www/x13as/>

Donde $\Theta_P(\cdot)$, $\theta_p(\cdot)$, $\Psi_Q(\cdot)$ y $\psi_q(\cdot)$ son polinomios de L de orden P , p , Q y q respectivamente. En general, la representación es de una serie no estacionaria, aunque si $D = d = 0$ y las raíces del polinomio característico (de los polinomios del lado izquierdo de la ecuación @ref(eq:ARIMA_seas)) todas son más grandes que 1 en valor absoluto, el proceso modelado será estacionario.

Si bien es cierto que existen otras formas de modelar la desestacionalización, como la modelación en diferencias con dummies para identificar ciertos patrones regulares, en los algoritmos disponibles como el X11 o X13-ARIMA-SEATS se emplea la formulación de la ecuación @ref(eq:ARIMA_seas). A continuación implementaremos la desestacionalización de una serie.

Como ejemplo utilizaremos la serie del Índice Nacional de Precios al Consumidor (INPC). Podemos ver que la serie original del INPC y su ajuste estacional bajo una metodología X13-ARIMA-SEATS son como se muestra en la Figura 5.1.

```
knitr::include_graphics("Plots/G_INPC_Adj.png")
```

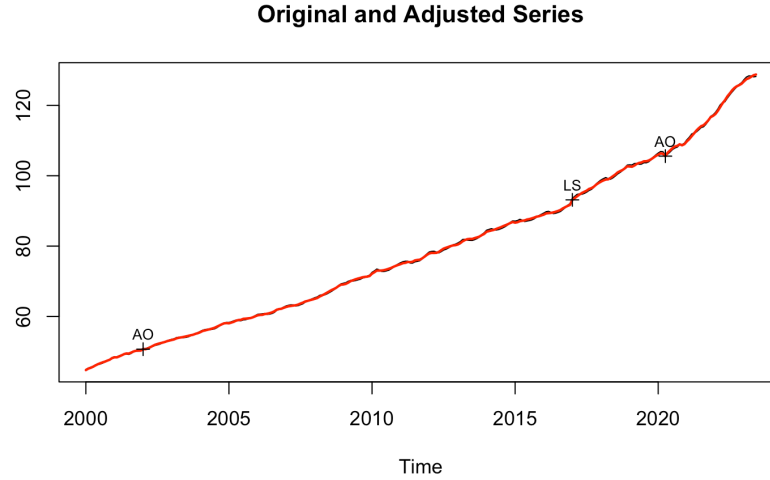


Figure 5.1: Índice Nacional de Precios al Consumidor ($INPC_t$) y su serie desestacionalizada utilizando un proceso X13-ARIMA-SEATS

El mismo procesamiento puede ser seguido para todas las series que busquemos analizar. En particular, en adelante, además del INPC que incluimos en la lista, utilizaremos las siguientes series, así como su versión desestacionalizada:

- Índice Nacional de Precios al Consumidor (base 2QJul2018 = 100), $INPC_t$.

- Tipo de Cambio FIX, TC_t
- Tasa de rendimiento promedio mensual de los Cetes 28, en por ciento anual, $CETE28_t$
- Indicador global de la actividad económica (base 2013 = 100), $IGAE_t$
- Industrial Production Index o Índice de Producción Industrial de los Estados Unidos (base 2012 = 100), IPI_t

5.2 Filtro Hodrick-Prescott

Como último tema de los procesos univariados y que no necesariamente aplican a series estacionarias, a continuación desarrollaremos el procedimiento conocido como filtro de Hodrick y Prescott (1997). El trabajo de estos autores era determinar una técnica de regresión que permitiera utilizar series agregadas o macroeconómicas para separarlas en dos componentes: uno de ciclo de negocios y otro de tendencia. En su trabajo original Hodrick y Prescott (1997) utilizaron datos trimestrales de alguna series como el Gross Domestic Product (GNP, por sus siglas en Inglés), los agregados montearios M1, empleo, etc., de los Estados Unidos que fueron observados posteriormente a la Segunda Guerra Mundial.

El marco conceptual de Hodrick y Prescott (1997) parte de suponer una serie X_t que se puede descomponer en la suma de componente de crecimiento tendencial, g_t , y su componente de ciclo de negocios, c_t , de esta forma para $t = 1, 2, \dots, T$ tenemos que:

$$X_t = g_t + c_t(\#eq : HP_{Eq}) \quad (5.2)$$

En la ecuación @ref(eq:HP_Eq) se asume que el ajuste de la ruta seguida por g_t es resultado de la suma de los cuadrados de su segunda diferencia. En esa misma ecuación sumiremos que c_t son las desviaciones de g_t , las cuales en el largo plazo tienen una media igual a cero (0). Por esta razón, se suele decir que el filtro de Hodrick y Prescott representa una descomposición de la serie en su componente de crecimiento natural y de sus desviaciones transitorias que en promedio son cero, en el largo plazo.

Estas consideraciones que hemos mencionado señalan que el procesamiento de Hodrick y Prescott (1997) implican resolver el siguiente problema minimización para determinar cada uno de los componentes en que X_t se puede descomponer:

$$\min_{\{g_t\}_{t=1}^T} \left[\sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [\Delta g_t - \Delta g_{t-1}]^2 \right] \quad (5.3)$$

Donde $\Delta g_t = g_t - g_{t-1}$ y $\Delta g_{t-1} = g_{t-1} - g_{t-2}$; $c_t = X_t - g_t$, y el parámetro λ es un número positivo que penaliza la variabilidad en el crecimiento de las series. De acuerdo con el trabajo de Hodrick y Prescott (1997) la constante λ

debe tomar valores específicos de acuerdo con la periodicidad de la series. Así, λ será:

- 100 si la serie es de datos anuales
- 1,600 si la serie es de datos trimestrales
- 14,400 si la serie es de datos mensuales

En resumen, podemos decir que el filtro de Hodrick y Prescott (1997) es un algoritmo que minimiza las distancias o variaciones de la trayectoria de largo plazo. De esta forma, determina una trayectoria estable de largo plazo, por lo que las desviaciones respecto de esta trayectoria serán componentes de ciclos de negocio o cambios transitorios (tanto positivos como negativos).

A continuación, ilustraremos el filtro de Hodrick y Prescott (1997) para dos series desestacionalizadas: $INPC_t$ y TC_t . Las Figura 5.2 y Figura 5.3 muestran los resultados de la implementación del filtro.

```
knitr::include_graphics("Plots/G_INPC_HP.png")
```

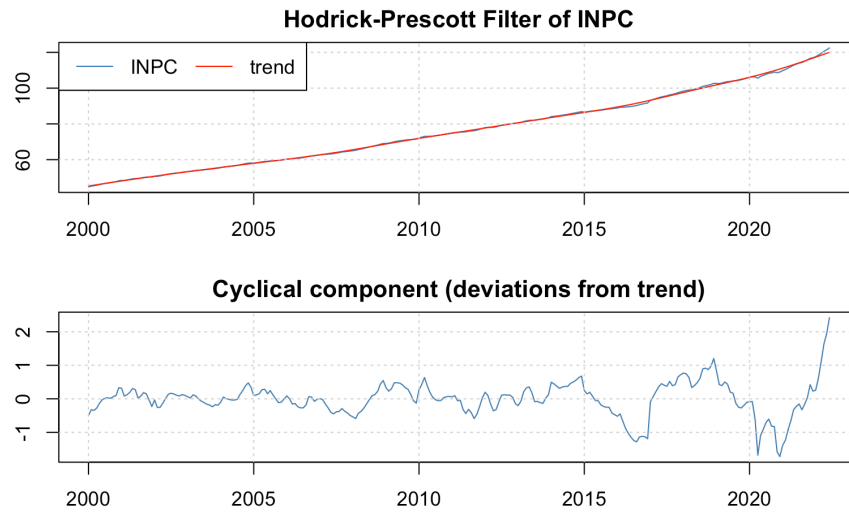


Figure 5.2: Descomposición del Índice Nacional de Precios al Consumidor ($INPC_t$) en su tendencia o trayectoria de largo plazo y su ciclo de corto plazo utilizando un filtro Hodrick y Prescott (1997)

```
knitr::include_graphics("Plots/G_TC_HP.png")
```

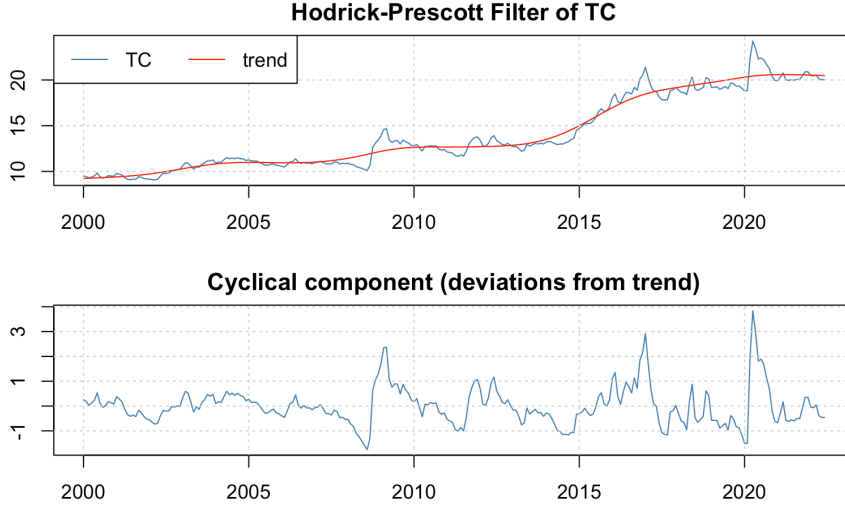


Figure 5.3: Descomposición del Tipo de Cambio FIX (TC_t) en su tendencia o trayectoria de largo plazo y su ciclo de corto plazo utilizando un filtro Hodrick y Prescott (1997)

5.2.1 Filtro Hodrick-Prescott planteado por St-Amant & van Norden

Método tradicional de HP consiste en minimizar la serie $\{\tau_t\}_{t=-1}^T$:

$$\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=1}^T [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2$$

Donde λ es un parámetro fijo (determinado ex-ante) y τ_t es un componente de tendencia de y_t .

Sin pérdida de generalidad, asumiremos que τ_{-1} y τ_0 son cero (0). De esta manera, la forma matricial del filtro HP es:

$$(Y - G)'(Y - G) + \lambda G'K'KG$$

La derivada de los anteriores:

$$-2Y + 2G + \lambda 2K'KG = 0$$

Despejando:

$$G_{hp} = [I_T + \lambda K'K]^{-1}Y$$

Donde G es el vector de tendencia, Y es el vector de la serie de datos, λ es la constante tradicional, y K es de dimensión $T \times T$ y está dada por la expresión:

$$K = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -2 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Así:

$$K' = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Método modificado de HP consiste en minimizar los valores de la serie $\{\tau_t\}_{t=1}^T$:

$$\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 + \lambda_{ss} \sum_{t=T-j}^T [\Delta \tau_t - u_{ss}]$$

Donde λ es una parámetro fijo (determinado ex-ante), τ_t es un componente de tendencia de y_t , y los nuevos parámetros son u_{ss} y λ_{ss} ajustadas por el procedimiento de Marcet y Ravn (2004).

Este procedimiento asume que parte del filtro HP y que esta versión tiene el problema de pérdida de información al final y al principio de la muestra. La razón es que es un procedimiento univariado que requiere de mucha información futura y pasada para mejorar el ajuste.

El componente adicional al filtro HP es un componente de castigo por desviaciones de la tasa de crecimiento de largo plazo, u_{ss} .

El proceso de selección de λ_{ss} es e propuesto por Marcet y Ravn (2004), el cual consiste en utilizar un λ convencional y el filtro HP convencional para estimar la siguiente función:

$$F(\lambda) = \frac{\sum_{t=2}^{T-1} ((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2}{\sum_{t=1}^T (y_t - \tau_t)^2}$$

Entonces el valor de λ_{ss} será aquel que:

$$F(\lambda_{ss}) = \frac{\sum_{t=2}^{T-1} ((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2}{\sum_{t=1}^T (y_t - \tau_t)^2} = F(\lambda)$$

nota: Antón (2009) estimó $\lambda_{ss} = 1096$ para datos trimestrales del PIB.

La forma matricial del filtro HP-SAVN es:

$$(Y - G)'(Y - G) + \lambda G'K'KG + \lambda_{ss}(L^jG + \bar{u}_{ss}M^j)$$

Donde $L^j = (0, 0, \dots, 0, -1, 0, \dots, 0, 1)$, en el cual el valor -1 es en la posición $T-j-1$ -ésima, y M^j es un vector que toma valores de cero hasta antes de $T-j$ y de 1 después.

La derivada de los anteriores:

$$-2Y + 2G + \lambda 2K'KG + \lambda_{ss}L'^j = 0$$

Despejando:

$$G_{SAVN} = \frac{1}{2}[I_T + \lambda K'K]^{-1}(2Y - \lambda_{ss}L'^j)$$

$$G_{SAVN} = [I_T + \lambda K'K]^{-1}Y - \frac{1}{2}[I_T + \lambda K'K]^{-1}\lambda_{ss}L'^j$$

Donde G es el vector de tendencia, Y es el vector de la serie de datos, λ es la constante tradicional, y K es de dimensión $T \times T$ y está dada por la expresión:

$$K' = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Dicho lo anterior, podemos modificar $F(\lambda)$ para el filtro HP convencional como en forma matricial:

$$F(\lambda) = \frac{G'K'KG}{(Y - G)'(Y - G)}$$