# Técnicas de Pronósticos
# Aplicaciones con R

# TÉCNICAS DE PRONÓSTICOS
## Aplicaciones con R

NORMAN GIRALDO GÓMEZ
Profesor Asociado
Escuela de Estadística
Universidad Nacional de Colombia
Medellín

**Universidad Nacional de Colombia**
Medellín

Copyright ©2006 Norman Diego Giraldo Gómez

No está permitido reproducir esta publicación o transmitirla por cualquier forma o medio, electrónico, mecánico, fotocopiado, escaneo ó de otro tipo excepto para citas cortas, sin el permiso del autor.

Técnicas de Pronósticos. Aplicaciones con R / Norman Diego Giraldo Gómez.

p. cm.—(Colección Notas de Clase)

"Universidad Nacional de Colombia."

Incluye referencias bibliográficas e índice.

ISBN 0-000-00000-0 (pbk.)

1. Probabilidades—Teoría. 2. Matemáticas

Ciencias—Investigación—Teoría. I. Giraldo, Norman D. II. Series.

519.2

G887c

Diagramación en LaTeX.
Impresión
Editorial

# Índice general

**1. Introdución al Curso** **1**

1.1. Objetivos del Curso . . . . . 1

1.2. Metodología, Requisitos y Evaluación . . . . . 2

1.3. Por qué Pronósticos? . . . . . 3

1.4. Ejemplos de Pronósticos . . . . . 4

1.5. Programa del Curso: Capítulo 2 . . . . . 5

1.6. Programa del Curso: Capítulos 3,4,5 . . . . . 5

1.7. Programa del Curso: Secciones 6,7,8 . . . . . 5

1.8. Programa del Curso: Capítulos 9,10 . . . . . 6

1.9. Introducción al R . . . . . 7

1.10. Detalles de R: "Task View" . . . . . 8

1.11. Bibliografía en Dropbox . . . . . 9

**2. Introdución al Lenguaje R** **11**

2.1. Introducción al Curso . . . . . 11

2.2. Introducción al R . . . . . 11
2.2.1. Uso de R de Manera Interactiva versus Ejecución desde un Programa 13
2.2.2. Ejecución desde un Programa . . . . . 14
2.2.3. Lectura de Datos . . . . . 17
2.2.4. Manejo de Series de Tiempo con R . . . . . 19
2.3. Repaso del Modelo de Regresión Lineal . . . . . 20
2.3.1. Estimación de los Parámetros . . . . . 22
2.3.2. Pruebas de Ajuste del Modelo . . . . . 26
2.3.3. Estadísticos de Ajuste y de Selección de Modelos . . . . . 28
2.3.4. Mínimos Cuadrados Nolineales . . . . . 30

**3. Modelado y Pronóstico de la Tendencia 35**

3.1. El Modelo Aditivo de Componentes de series de tiempo . . . . . 35
3.2. Estimación de la Tendencia . . . . . 39
3.3. Pronósticos con base en la Tendencia . . . . . 40
3.4. Medidas de Precisión en los Pronósticos . . . . . 41
3.5. Caso de Estudio: Pronóstico de Ventas al Menudeo . . . . . 42
3.5.1. Descripción de los Datos . . . . . 42
3.5.2. Programa en R . . . . . 43
3.6. Problemas . . . . . 48

**4. Modelado y Pronóstico incluyendo la Componente Estacional 51**

4.1. Introducción . . . . . 51
4.2. Estimación de la Componente Estacional . . . . . 54
4.2.1. Modelos con Variables Indicadoras . . . . . 54
4.2.2. Modelos con Funciones Trigonométicas . . . . . 58

**5. Modelos con Base en Suavizadores** **61**

5.1. Regresión Local Loess . . . . . 61

5.2. Método híbrido Loess y componente estacional . . . . . 65

5.3. Método de descomposición STL . . . . . 65

5.4. Modelo de Holt-Winters . . . . . 67

5.5. El Modelo de Espacio de Estados . . . . . 69

5.6. Pruebas de Estabilidad Estructural en Modelos de Componentes . . . . . 71

5.6.1. Estimación Recursiva de los Parámetros . . . . . 72

5.6.2. Pruebas CUSUM . . . . . 73

5.6.3. Implementación en R: la librería strucchange . . . . . 76

**6. Medias Móviles y Filtros** **83**

6.1. Medias Móviles . . . . . 83

6.2. Método de descomposición con base en Medias móviles . . . . . 90

6.3. Librerías para filtrado de series en R . . . . . 92

6.4. Problemas . . . . . 93

**7. Validación de los Supuestos sobre los Errores** **95**

7.1. Introducción . . . . . 95

7.2. Series Estacionarias en Covarianza . . . . . 97

7.2.1. Estimación de las funciones de Autocovarianza y Autocorrelación. . 100

7.3. Pruebas de Incorrelación . . . . . 104

7.3.1. Prueba Ljung-Box (LB) . . . . . 105

7.3.2. Prueba Durbin-Watson (DW) . . . . . 108

7.3.3. Prueba Durbin-Watson Generalizada . . . . . 111

7.4. Alternativas cuando los Residuos Estructurales muestran Autocorrelación . 113

7.5. Pruebas de Heterocedasticidad . . . . . 116

**8. Modelos ARMA para la Componente Aleatoria** **119**

8.1. Introducción . . . . . 119

8.2. Procesos de Medias Móviles de orden q, MA(q) . . . . . 120

8.3. Procesos Autorregresivos de Orden p, AR(p) . . . . . 129

8.4. Procesos Autoregresivos y de Medias Móviles ARMA(p,q) . . . . . 140

8.5. Modelos ARMA estacionales, SARMA . . . . . 148

**9. Raíces Unitarias y Tendencias Estocásticas (ARIMA)** **155**

9.1. Introducción . . . . . 155

9.2. Modelos ARIMA . . . . . 155

9.3. Los Modelos de Marcha Aleatoria . . . . . 157

9.4. Raíces Unitarias, Estimación y Prueba de Hipótesis . . . . . 164

9.5. Prueba Dickey-Fuller (DF) . . . . . 166

9.6. La Potencia de las pruebas Dickey-Fuller . . . . . 167

9.7. Implementación en R de la Prueba Dickey-Fuller . . . . . 168

9.8. Prueba Dickey-Fuller Aumentada . . . . . 169

**10. Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)** **173**

10.1. Modelos SARIMA . . . . . 173

10.2. Pruebas de Raíz Unitaria Estacional . . . . . 179

10.3. La prueba Canova-Hansen . . . . . 180

10.4. La prueba HEGY . . . . . 184

10.5. Prueba OCSB . . . . . 189

10.5.1. Ejemplo de resumen . . . . . 191

**11. Modelos de Rezagos Distribuídos y Cointegración** **195**

11.1. Modelo de rezagos distribuídos . . . . . 195

11.2. *Cointegración . . . . . 200

**A. Datos de Series de Tiempo** **201**

A.1. Series con Tendencia y Estacionalidad . . . . . 201

A.1.1. Ejemplo 1. Cinco Series . . . . . 201

A.1.2. Ejemplo 2. Consumo domiciliario agua potable en Medellín . . . . 203

# Índice de figuras


1.1. potencia de energía: entrada agua en metros cúbicos diarios en embalses . . 3

1.2. logprecio kwh bolsa energía colombia . . . . . 3

2.1. Gráfica de una serie mensual . . . . . 20

3.1. Descomposición de la Serie $Y_t$ . . . . . 36

3.2. Indices de vivienda usada nominal y real . . . . . 38

3.3. Tasa de Empleo, Australia, 02/1978 - 09/2014 . . . . . 39

3.4. Pronósticos (—-) versus Observados (-o-o-) . . . . . 48

3.5. Ajuste Logístico a Población de Medellín . . . . . 49

4.1. Average Monthly Temperatures at Nottingham, 1920-1939 . . . . . 52

4.2. Ajuste con Variables Indicadoras . . . . . 58

5.1. Suavizamiento Loess de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06 . . . . . 63

5.2. Descomposición STL número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06 . . . . . 66

5.3. Estimación de la tendencia con los Métodos de Loess y Holt-Winters . . . 70
5.4. Estimadores Recursivos para el Modelo de Producción de Cemento . . . . 74
5.5. Región de Rechazo de la Prueba $CUSUM$. . . . . 75
5.6. Resultados prueba CUSUM gráfica . . . . . 78
5.7. Resultados prueba MOSUM gráfica . . . . . 78
5.8. Punto de Quiebre 1974-Q4 en la Serie y en los Residuos . . . . . 82

6.1. Precios de la Acción de Microsoft Suavizados con SES . . . . . 86

7.1. Ejemplo Función de Autocovarianza . . . . . 97
7.2. Ejemplo Función de Autocorrelación. . . . . 99
7.3. Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario . . . 100
7.4. Ejemplos de Fac y Bandas de Bartlett . . . . . 103
7.5. log(igbc) y diff(log(igbc)), a izquierda, con sus fac muestrales, a derecha . . 103
7.6. Fac de los Residuos Estructurales Serie Cementos . . . . . 104
7.7. Densidad $\chi^2_m$. . . . . 106
7.8. Fac teóricas de un $AR(1)$. . . . . 108
7.9. Y estimada versus residuales estructurales . . . . . 118

8.1. Fac teóricas de modelos $MA(q)$ . . . . . 122
8.2. fac y facp de un $MA(3)$. . . . . 125
8.3. Pronosticos Estructurales versus Estructurales + MA(6) . . . . . 129
8.4. Círculo Unitario . . . . . 131
8.5. fac de ar(2) . . . . . 134
8.6. Aportes hidrológicos en Gw/h, 01/07/98 - 31/05/14 . . . . . 138
8.7. fac y facp de Aportes hidrológicos . . . . . 140
8.8. Fac Muestral de $ARMA(p, q)$. . . . . 145
8.9. Posibles submodelos (1,1)=AR, (1,2),(2,1)=ARMA, (2,2)=SARMA . . . . 151

8.10. Pronósticos con tres modelos para producción de cemento . . . . . 154

9.1. Trayectorias de Marchas Aleatorias . . . . . 158

9.2. Pronósticos de la Serie PNB-USA. . . . . 161

9.3. (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada . . . . . 161

9.4. Pronósticos usd/yen con ARIMA(3,1,2)(continua) y Tendencia Lineal+AR(2) (punteada) . . . . . 162

9.5. Serie USD por Libra . . . . . 163

9.6. fac y facp muestrales de proceso con raíz unitaria . . . . . 165

10.1. Aplicación de diferencia simple y luego diferencia estacional a la serie de producción de cemento . . . . . 174

10.2. Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12] . . . . . 177

10.3. Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12] . . . . . 179

10.4. Serie de Producción Trimestral de Cemento, Australia . . . . . 191

10.5. Serie Diferenciada de Producción Trimestral de Cemento . . . . . 192

10.6. Pronósticos a 8 trimestres de la Producción Trimestral de Cemento . . . . . 193

10.7. Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA . . . . . 193

# Listings

# CAPÍTULO 1

# Introdución al Curso

## 1.1. Objetivos del Curso

1. El objetivo es desarrollar modelos estadísticos de series de tiempo que permitan el cálculo de pronósticos.
2. De igual manera, introducir el manejo de software R para implementarlos.
3. El curso se apoya en el texto de Notas de Clase, que desarrolla tres temas, articulados, similares al texto de Diebold [1999], implementándolos en el software R.
   - *a*) Modelos de descomposición
   - *b*) Modelos para los residuos de la descomposición
   - *c*) Modelos alternos
   - *d*) Comparaciones
4. Ejemplos con énfasis en el cálculo de pronósticos de series reales.

## 1.2. Metodología, Requisitos y Evaluación

1. Metodología: clase de exposición seguida de clase de práctica con R en la sala de cómputo del bloque 43, tercer piso

2. Las exposiciones en clase se harán siguiendo las Notas de Clase, con posibles modificaciones en el transcurso del semestre y textos de consulta (incluyendo Wikipedia)

3. Pre-requisitos: Cálculo, probabilidad, inferencia estadística, análisis de regresión.

4. Texto guía: Notas de clase del profesor.

5. Bibliografía: Existen otros textos que desarrollan la teoría de series de tiempo utilizando el R como el software básico, por ejemplo, Cowpertwait and Metcalfe [2009], Shumway and Stoffer [2005], Cryer and Chan [2008], Aragon [2008], **?**. En Dropbox hay varios textos similares a las notas de clase.

6. Programación en lenguajes R, SAS.

7. Evaluación: 3 trabajos, de igual valor, en grupos de máximo 2 personas, con base en las prácticas en R. Trabajos tipo Proyecto: los trabajos 1,2,3 son tres fases de un análisis que tiene como objetivo comparar los pronósticos de las metodologías expuestas.

   Las evaluaciones de los trabajo tienen en cuenta la calidad de la argumentación de los análisis tanto como los resultados.

.

## 1.3. Por qué Pronósticos?

Los pronósticos se utilizan en forma constante en diversos campos: economía, finanzas, mercadeo, medio ambiente, ingeniería, salud.

Recientemente, técnicas de cluster para conjuntos grandes de datos de series de tiempo (bid data)

Una finalidad es proveer una guía para las decisiones que deben tomarse.

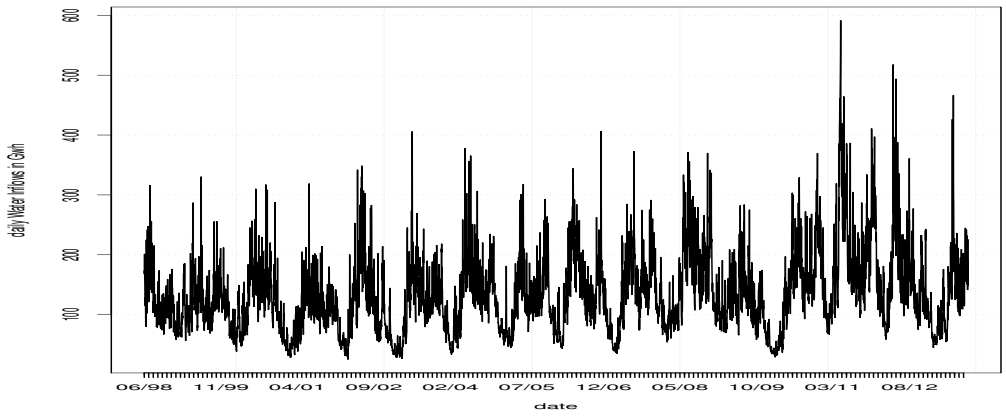



Figura 1.1: potencia de energía: entrada agua en metros cúbicos diarios en embalses

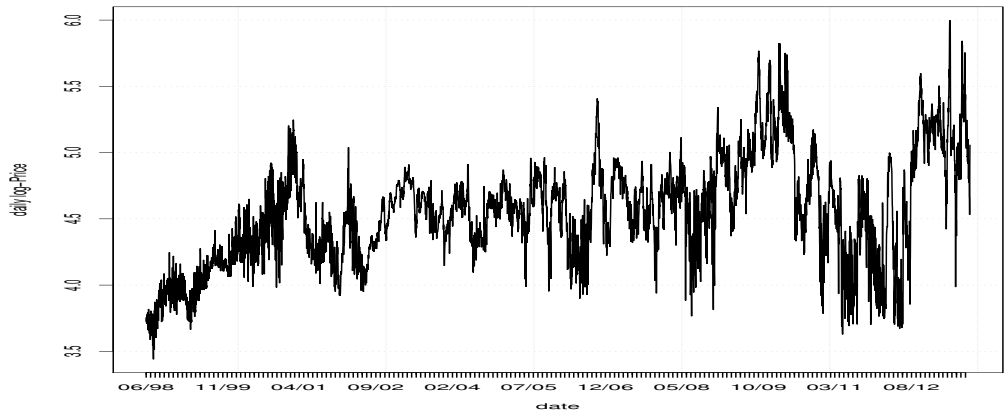



Figura 1.2: logprecio kwh bolsa energía colombia

Algunos ejemplos en estos campos pueden ser los siguientes.

## 1.4. Ejemplos de Pronósticos

1. Planeación y Control de Operaciones. Por ejemplo, las decisiones de producción de un artículo con base en los pronósticos de ventas. Es posible por ejemplo, detectar una disminución en la tendencia de ventas que conlleve a reducir la producción, ó al contrario.

2. En Mercadeo la decisión de invertir en publicidad puede depender de pronósticos de ventas.

3. Caso: Leonisa SA. Problema: pronósticos de ventas de diferentes modelos en el Catálogo.

4. En Economía. Las decisiones del Banco de la República, por ejemplo para el control de la inflación, requieren el pronóstico y el examen del comportamiento de ciertas variables macroeconómicas, como el PIB, la tasa de desempleo, el IPC, las tasas de interés a distintos plazos, activas y pasivas.

5. En Economía los pronósticos sobre ciclos económicos de recesión y expansión como guía de aumento ó disminución de las tasas interbancarias.

6. En Planeación. Los pronósticos de demanda de viviendas, en los diferentes rangos, para orientar políticas de uso del suelo en los diferentes POT.

7. En Planeación. Los pronósticos de gastos sirven para definir los presupuestos futuros.

8. En Planeación. Los pronósticos de evolución de casos de enfermedades ó contingencias de salud. Aunque en casos como epidemias los modelos son series no lineales.

9. En Planeación. El pronóstico de consumo de energía eléctrica domiciliaria es fundamental para las decisiones de generación.

10. En Turismo. El pronósticos de número de turistas mensuales para determinar la demanda hotelera.

11. En Epidemiología y Medio ambiente. La vigilancia de los niveles de contaminantes en el aire tiene como herramienta fundamental las series de tiempo. Pero adicionalmente el efecto de estos niveles sobre la salud.

## 1.5. Programa del Curso: Capítulo 2

- El Capítulo 2 es introducción al R y repaso de estadística básica: regresión lineal múltiple, pruebas de hipótesis, diagnósticos.
- Con respecto a R se requiere tener el programa instalado. Y conocer los procedimientos básicos: lectura de datos, operaciones básicas
- Introducción de las operaciones básicas con datos de series de tiempo: manejo de la frecuencia de muestreo, fechas (calendario), acceso a bases de datos.
- Repaso de regresión lineal múltiple con R. Es básica para los modelos de descomposición, que son regresiones lineales (en algunos casos también regresión no lineal)

## 1.6. Programa del Curso: Capítulos 3,4,5

Se introduce el modelo de descomposición aditivo clasico

$$Y_n = T_n + S_n + Z_n$$

Serie = Tendencia + Estacionalidad + Residuo aleatorio

1. 3: modelos para la tendencia
2. 4,5 : modelos para la estacionalidad
3. Primer Trabajo: en grupos de máximo 2 personas (40 $\%$)

## 1.7. Programa del Curso: Secciones 6,7,8

Se introduce el análisis del Residuo aleatorio $Z_n$

$$Y_n = T_n + S_n + Z_n$$

1. 6: herramientas estadísticas para analizar $Z_n$
2. 7: modelos para la $Z_n$
3. 8: ensamblar el modelo completo de componentes y pronosticar
4. Segundo Trabajo: (30 $\%$)

## 1.8. Programa del Curso: Capítulos 9,10

Se introducen dos modelos alterno al modelo de componentes: los modelos ARIMA y SARIMA (auto-regresivo de media móvil integrado - estacional)

1. Sección 9: modelos ARIMA ($\Longleftrightarrow Y_n = T_n + Z_n$)
2. Sección 10: modelos SARIMA ($\Longleftrightarrow Y_n = T_n + S_n + Z_n$)
3. Tercer Trabajo: (30 $\%$)

## 1.9. Introducción al R

Qué es R? En Wikipedia ([1]) responden:

> Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo o graficación.

Cómo se obtiene R y se instala?.

1. Entrar a `http://www.cran.r-project.org/`.
2. Seleccionar: "Download R for Windows" → "base".
3. Descargar el archivo .exe más reciente.
4. Ejecutar el archivo .exe
5. Detalles de instalación: R instala unas librerías por defecto. El resto de las 8884 librerías se puede instalar de manera inmediata dando click en (Paquetes→Instalar paquete(s)) donde se establece una conexión a Internet con los servidores de R en el mundo. A continuación se escoge uno, por ejemplo, Austria y luego aparece la lista de paquetes que se pueden instalar.

[1] https://es.wikipedia.org/wiki/R_(lenguaje_de_programación)

## 1.10. Detalles de R: “Task View”

CRAN Task View: Time Series Analysis
Maintainer: Rob J Hyndman ([2])

- Forecasting and Univariate Modeling
- Frequency analysis
- Decomposition and Filtering
- Seasonality
- Stationarity, Unit Roots, and Cointegration
- Analysis of large groups of time series
- Continuous time models
- Resampling

[2] http://robjhyndman.com/hyndsight/

## 1.11. Bibliografía en Dropbox


- Giraldo, N. *Notas de Clase. Series de Tiempo con R*. En Dropbox
- Diebold, F. *Elementos de Pronósticos*. International Thomson Editores, 1999.
- Klaus Neusser *Time Series Analysis in Economics*.
  *www.neusser.ch/downloads/TimeSeriesBook.pdf*
- R. Adhikari and R. K. Agrawal *An Introductory Study on Time Series Modeling and Forecasting*.
  *https://arxiv.org/abs/1302.6613*
- Ihaka, Ross *Time Series Analysis. Lecture Notes*.
  *https://www.stat.auckland.ac.nz/∼ihaka/726/notes.pdf*
- R.H. Shumway and D.S. Stoffer *Time Series Analysis and Applications Using the R Statistical Package.*
- R.H. Shumway and D.S. Stoffer: A gentle introduction to time series analysis is now available (for free). *http://www.stat.pitt.edu/stoffer/tsa4/tsaEZ.pdf*
- M. Dettling *Applied Time Series Analysis*
  *https://stat.ethz.ch/education/semesters/ss2013/atsa/ATSA-Scriptum-SS2013_130415.pdf*
- Mahmood Arai *A Brief Guide to R for Beginners in Econometrics.*.
  *homes.chass.utoronto.ca/∼jmaheu/4050/Rintro.pdf*

# CAPÍTULO 2

# Introdución al Lenguaje R

## 2.1. Introducción al Curso

El curso de Series de Tiempo consiste de dos temas: el modelo de descomposición y el modelo ARIMA-SARIMA. El objetivo es desarrollar modelos de series de tiempo que permitan el cálculo de pronósticos. El curso está organizado con base en el texto de Diebold [1999]. Existen otros textos que desarrollan la teoría de series de tiempo utilizando el R como el software básico, por ejemplo, Cowpertwait and Metcalfe [2009], Shumway and Stoffer [2005], Cryer and Chan [2008], Aragon [2008], Hyndman et al. [2008].

## 2.2. Introducción al R

1. Qué es R?. Es un software libre, de código abierto, para programación orientada a objetos, dedicado a cómputos estadísticos, en bioestadística, econometría y finanzas computacionales, entre muchos otros.

2. La información sobre R está en el sitio web `http://www.r-project.org/`, incluyendo el ejecutable, manuales, descripción de librerías, etc.. Es un sistema de librerías (toolboxs) de R para realizar finanzas computacionales e ingeniería financiera. R posee aproximadamente 1600 librerías sobre diversas aplicaciones. R-metrics posee 10

librerías esplecializadas para manejo de riesgo, portafolios, valoración, econometría,...

3. Cómo se obtiene R y se instala?.

   *a*) Entrar a `http://www.cran.r-project.org/`.

   *b*) Seleccionar: "Download R for Windows" → "base".

   *c*) Descargar el archivo .exe más reciente: (Download R x.xx.x for Windows).

   *d*) Ejecutar el archivo .exe desde el escritorio.

4. Detalles de instalación: R instala unas librerías por defecto. El resto de las (≈) 8000 librerías se puede instalar de manera inmediata dando click en (Paquetes→Instalar paquete(s)) donde se establece una conexión a Internet con los servidores de R (mirrors) en el mundo (el principal está en Viena, Austria). A continuación se escoge uno, por ejemplo, Colombia (Cali) y luego aparece la lista de paquetes que se pueden instalar. Esta es la vía para las librerías de R-metrics.

5. Algunas librerías:

   - `actuar`: cálculos actuariales.
   - `bbmle`: estimaciónpor máxima verosimilitud.
   - `fBonds`: cálculos con bonos.
   - `fuzzyOP`: lógica difusa.
   - `nnet`: redes neuronales.

6. Algunos manuales:

   *a*) Entrar a `http://www.cran.r-project.org/`.

   *b*) Seleccionar: "Documentation".

   *c*) Seleccionar: "Manuals"

   *d*) Seleccionar: "contributed documentation", "Non-English Documents".

   *e*) "R para Principiantes", the Spanish version of "R for Beginners", translated by Jorge A. Ahumada (PDF).

   *f*) A Spanish translation of "An Introduction to R" by Andrés González and Silvia González (PDF, Texinfo sources).

   *g*) "Metodos Estadisticos con R y R Commander" by Antonio Jose Saez Castillo (PDF, ZIP, 2010-07-08).

### 2.2.1. Uso de R de Manera Interactiva versus Ejecución desde un Programa

Al ejecutar R aparece una ventana denominada *R Console* ó ventana interactiva. El símbolo ">" indica que se espera un comando.

- `> a = 7.5`
  equivale a los dos siguientes; todos crean un objeto numérico "a" con el valor 7.5.
- `a <- 7.5`
- `7.5 -> a`
- `> b = c(3,2,5)`
  Crea un vector columna.
- `> tb = t(b)`
  Crea el vector traspuesto de "b", es decir, como vector fila.
- `> b1 = c(b,a)`
  Concatena y crea un vector columna: $b_1 = (3, 2, 5, 7.5)'$.
- `> b2 = seq(1,length(b1),1)`
- `> b3 = cbind(b1,b2)`
  Produce una matriz $4 \times 2$.
- `> b4 = rbind(b1,b2)`
  Produce una matriz $2 \times 4$.
- `> (b34 = b3%*%b4)`
  Produce producto matricial

```
> b34
     [,1] [,2] [,3]  [,4]
[1,] 10.0    8 18.0 26.50
[2,]  8.0    8 16.0 23.00
[3,] 18.0   16 34.0 49.50
[4,] 26.5   23 49.5 72.25
```

- `> h = matrix(0,4,4)`
  Crea una matriz $4 \times 4$ de ceros.

- `diag(h) = 1` Asigna unos a la diagonal de la matriz $h$, es decir, la vuelve una matriz identidad de $4 \times 4$.
- `b34i = solve(b34 + h)`
  Encuentra $(b_{34} + h)^{-1}$.
- `bb = b34*h`
  Multiplica elemento a elemento.

```
> bb
     [,1] [,2] [,3]  [,4]
[1,]   10    0    0  0.00
[2,]    0    8    0  0.00
[3,]    0    0   34  0.00
[4,]    0    0    0 72.25
```

### 2.2.2. Ejecución desde un Programa

En lugar de ó al mismo tiempo del manejo interactivo se pueden usar programas. Los programas en R tienen extención *.R* ó *.r*, por ejemplo, *calculo.2.r*. Los pasos para generar un programa se pueden describir como sigue.

1. Escoger un directorio base:
   Escogemos en la barra de menú (Archivo→Cambiar dir...), luego se escoge o crea la ruta desde donde se quiera trabajar. Luego (Archivo→Nuevo script) y aparece otra ventana "Sin nombre - Editor R".
2. En la barra menú seleccionamos (Ventanas→Divididas verticalmente) para divir la pantalla.
3. Escribimos el programa. Por ejemplo "ejemplo1.r"

```
# Ejemplo del uso de programas.
# Muestra el uso de for, if, ifelse.

b = 3
x = c(2,3,-4.2,5,10,7,3,2,3,0)
y = numeric(length(x))

for(i in 1:length(x)){
```

```
if(x[i] == b) y[i] = 0
else y[i] = 1
}

z = ifelse(x==b,0,1)
(cbind(y,z))
```

4. Al terminar el código, en la pestaña (Archivo→Guardar como...), escribimos: ejemplo1.r

5. Para ejecutar el programa en la pestaña (Editar→Ejecutar todo)

Suponga que se inicia una sesión nueva en R y se corre en la consola
`> source("ejemplo1.r")`
automáticamente se ejecutan los comandos en el programa. En la memoria quedan las variables generadas: `b, x, y, z`.
`> ls.str()` muestra lo que hay en la memoria.
`> z = NULL` elimina la variable `z`.
Una parte importante de la programación con R es escribir funciones. Para hacerlo se escribe el programa en el editor. Por ejemplo.

```
# Función para calcular el estadístico t para la igualdad de medias

prueba.t2 = function(y1,y2){
n1 = length(y1)
n2 = length(y2)
yb1 = mean(y1)
yb2 = mean(y2)
s1 = var(y1)
s2 = var(2)
s = ((n1-1)*s1 + (n2-2)*s2)/(n1 + n2 -2)
tst = (yb1 - yb2)/sqrt(s*(1/n1 + 1/n2))
return(tst)
}
```

Se salva el archivo con el mismo nombre de la función, es decir, `prueba.t2.r`
Para cargarlo en una sesión interactiva ó en otro programa se corre: `source("prueba.t2.r")`

**Ejemplo 2.2.1.** `# Prueba de igualdad de medias asumiendo varianzas iguales.`

```
source("prueba.t2.r")
s1 = 2; s2 = 2
u1 = 4; u2 = 6
n1 = 100
n2 = 30
x1 = rnorm(n=n1,mean=u1,sd=s1)
x2 = rnorm(n2,u2,s2)
tst = prueba.t2(x1,x2)

# Cálculo del valor crítico de una t con v=n1+n2-2
# Para alpha = 0.05/2 = 0.025

v = n1 + n2 - 2
tcrit = abs(qt(0.025,v))
abs(tst) < tcrit # Si TRUE no se rechaza H0:u1=u2
```

**Concepto de data.frame (Hojas de Datos)**

Un objeto data.frame es una clase de objeto en R que es una colección de columnas que contienen datos, que no tienen que ser del mismo tipo, con el mismo número de filas. Las columnas pueden tener nombres. (ver Cap 6 de Manual ó sección 2.2.2 de Farnsworth).

**Ejemplo 2.2.2.** *Si D se ha definido como data.frame, y contiene las columnas edad, genero, salario, entoces al ingresar* `D$edad` *en la consola, escribe los datos de la columna edad.*

**Concepto de list**

Una lista (list) es otro tipo de objeto en R, más general que un data.frame. En una lista se coloca un grupo de otros objetos que pueden ser: escalares, data.frames, cadenas de símbolos. Las funciones en R generalmente retornan un objeto que puede ser un escalar,o también vector, o matríz, o un data.frame, o un list.

**Ejemplo 2.2.3.**

```
nombre = c("Cecilia", "Olga", "Beatriz")
ab = 1:3
cd = 3:5
# crear un data.frame
```

```
(B = data.frame(nombre=nombre, ab=ab, cd=cd))
B$nombre[2] # Olga
# crear un list con un data.frame incluído
R = list(B=B, ab=ab, cd=cd)
R$B$nombre[1] # Cecilia
```

### 2.2.3. Lectura de Datos

R posee varias funciones para lectura de datos. Incluso, tiene capacidad para leer archivos que se encuentran en un sitio Web determinado, dado que exista una conexión de internet.

#### La función read.table

La función `G = read.table("nombre")` se utiliza para leer datos en R de tal forma que retorna un objeto data.frame, en este ejemplo indicado por la letra G, y donde "nombre" es el nombre del archivo en donde están los datos, por ejemplo, `nombre = base.dat`. Las extensiones pueden ser ".dat", ".txt", ".prn". En particular, ".prn" es la adecuada cuando se quieren leer datos de excel. En este caso se guarda el archivo excel con el formato ".prn", que corresponde a columnas separadas por espacios. No es conveniente salvar un archivo excel con extensión ".txt" porque este formato guarda controles de tabuladores invisibles que impiden la lectura de R.

Si cada columna en el archivo de lectura de datos tiene un nombre entonces el comando se modifica colocando `G = read.table("nombre", header = TRUE)`. Entonces conviene añadir la instrucción `attach(G)`, la cual hace que los nombres de las columnas pasen a ser los nombres de las variables activas en R. Ocasionalmente una de las columnas es la fecha de la observación, con el formato día-mes-año, por ejemplo, "23/10/2010". En este caso es conveniente colocar la instrucción en la forma

```
G = read.table("nombre", header = TRUE, stringsAsFactors=FALSE)
```

El efecto es que toma los datos de las fechas como alfanuméricos y no como factores, que es un tipo de datos para utilizar en diseño de experimentos.

**Ejemplo 2.2.4.** *En este ejemplo se muestra cómo utilizar la función* `read.table` *para leer datos de un archivo en una página web. El archivo tiene dos columnas con los nombres:* `fecha` *y* `x`*. Los datos de la variable* `fecha` *son alfanuméricos.*

```
#---------------------------------------
```

```
archivo = "http://www.unalmed.edu.co/~ndgirald/
Datos/Datos%20curso%20Series%20II/fechaydatos.prn"
G = read.table(archivo, header = TRUE,
stringsAsFactors=FALSE)
attach(G)
np = length(x)
fechas  = strptime(fecha, "%d/%m/%y")
plot(fechas,x, xaxt="n", panel.first = grid(),type='l',ylab='')
axis.POSIXct(1, at=seq(as.Date(fechas[1]),as.Date(fechas[np]),
"months"), format="%m/%y")
axis.POSIXct(1, at=seq(as.Date(fechas[1]),as.Date(fechas[np]),
"years"), labels = FALSE, tcl = -0.2)
```

**Funciones de lectura de archivos Excel**

Hay otras funciones de lectura, como `G = read.csv("nombre", header = TRUE)`, que lee archivos excel "delimitados por comas", de extensión .csv.

Los archivos con extensión ".xls", ".xlsx" se puenden leer con la función `read_excel()` de la librería `readxl`.

```
#---------------------------------------
install.packages("readxl")
library(readxl)

res <- read_excel("international-petroleum-world-cr.xlsx")
attach(res)
x = golfo
y = mexico
```

**Funciones para escritura de datos**

Hay funciones para escribir datos en un archivo, como la función
`G = write.matrix(data.frame, archivo)`.

```
#---------------------------------------
D = data.frame(fechas=fecha,x=x)
require(MASS)
```

```
write.matrix(D,"h.dat")
```

### 2.2.4. Manejo de Series de Tiempo con R

En esta sección consideramos una serie de tiempo $T_t,\ t = 1, 2, \ldots, T$, como un vector columna. En R se ingresaría inicialmente como un objeto tipo vector numérico. Pero para los análisis en los capítulos siguientes se requiere redefinir la serie como un objeto de clase "time series", mediante la función `ts()`.

En esta sección se muestran algunos procedimientos en R para el manejo inicial de series de tiempo, tales como lectura, graficación y especificación de formatos de fechas.

**Ejemplo 2.2.5.** *Un primer procedimiento consiste en ingresar los datos de la serie de tiempo directamente en el programa mediante la instrucción* `structure()`.

Utilizando el comando `ts()` se puede añadir información acerca de las fechas y la frecuencia de muestreo, es decir, cuántas observaciones anuales se tomaron, así como la fecha inicial. En este ejemplo se asigna una frecuencia de 12, es decir, datos mensuales. Además, se especifica que la serie inicia en Enero (01) de 1990.

```
# leer la serie de tiempo como un vector numerico
y=structure(c(1574,1368,1387,1109,1257,1376,2143,1208,
2007,1876,1702,1819,1802,1205,1684,1682,1991,2394,1914,
2499,2130,2529,2328,2076,2496,1647,2518,2205,2395,2891,
2712,2427,2477,2860,2505,3355,1760,2318,3111,2570,2868,
3042,2749,2839,3140,2909,2982,3667,2814,2732,3265,3166,
2792,3742,3099,3278,4120,3553,3675,3799,3427,3234,3733,
3642,3553,3647,3624,2973,3597,3731,4092,4100,2762,3953,
4152,4229,4419,4774,4313,4060,4664,4374,4419,4908,4321,
4772,4361,4969,5111,5014,4858,5159,5086,5379,5605,5269))

# convertir a objeto ts
y=ts(y,frequency=12,start=c(1990,1))

# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1990/3/1"), length.out = length(y), by =  "months")

# grafica con fechas
```

```
ts.plot(y,main="serie F")
```

El comando `plot.ts()` grafica varias series (hasta 10 máximo), que deben estar en una matriz ó data.frame.

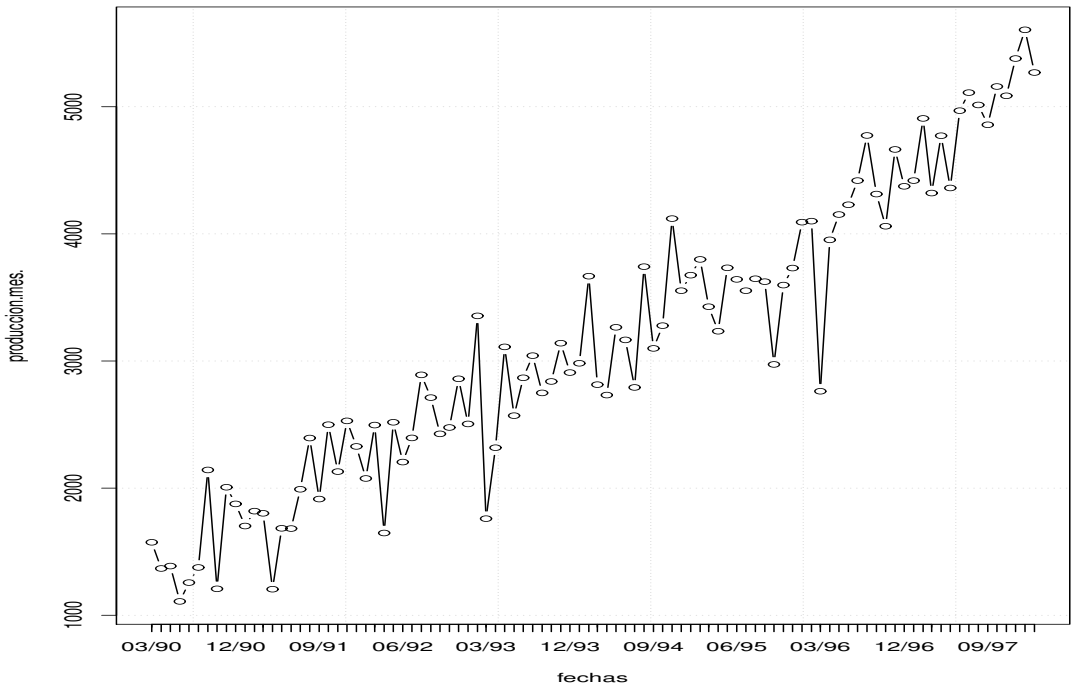



Figura 2.1: Gráfica de una serie mensual

## 2.3. Repaso del Modelo de Regresión Lineal

El modelo de regresión lineal múltiple es básico para los modelos de componentes de series de tiempo. Una exposición más completa está en varios textos, por ejemplo, en Kleinbaum et al. [1988] y Sheather [2009, caps. 5,6]. El modelo siguiente es lo suficientemente general para introducir los conceptos básicos. Suponga que se tiene una muestra de tamaño $T$, de la variable aleatoria $Y$, y $T$ mediciones de las variables $X_1$, $X_2$, que no se consideran aleatorias,

$$Y_t, X_{1,t}, X_{2,t}, \quad t = 1, \ldots, T.$$

Además suponga que $\varepsilon_t \sim iid\,(0, \sigma^2)$ es una muestra de una variable aleatoria continua de media cero y varianza $\sigma^2$. Si se cumple que

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \tag{2.1}$$

entonces se dice que las variables$Y, X_1, X_2$ satisfacen un modelo de regresión lineal. Donde $Y_t$ es la variable dependiente y $X_{1,t},\ X_{2,t}$ son las variables explicativas y el término $\varepsilon_t$ es el error ó residuo. Las variables $\beta_0, \beta_1, \beta_2, \sigma^2$ se denominan los parámetros.

Si se define la matriz de diseño $X \in \mathbb{R}^{T\times 3}$, y los vectores $\underline{Y}$, $\underline{\varepsilon}$

$$X = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} \\ 1 & X_{1,2} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{1,T} & X_{2,T} \end{bmatrix},\ \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix},\ \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}\ \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

entonces el modelo se puede expresar en forma matricial como

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}. \tag{2.2}$$

Los supuestos del modelo lineal son

1. Las $\varepsilon_t$ no están correlacionadas.
2. Las $\varepsilon_t$ tienen varianzas constante.
3. El rango de la matriz $X$ es 3, es decir, no existen constantes $a, b$ tal que

   $$aX_{1,t} + bX_{2,t} = 1.$$

   De manera equivalente, las columnas de X son linealmente independientes.

El modelo se dice que es lineal porque $Y_t$ se relaciona linealmente con $X_{1,t},\ X_{2,t}$, es decir, $Y_t$ es una combinación lineal de $X_{1,t}, X_{2,t}$. Una relación de la forma, por ejemplo

$$Y_t = \alpha_0 + \frac{1}{1 + e^{\beta_0+\beta_1 X_{1,t}+\beta_2 X_{2,t}}} + \varepsilon_t, \tag{2.3}$$

es no lineal. Este modelo (2.3) es un ejemplo simple de una red neuronal.

La forma general de una red neuronal con $p$ nodos de entrada y $q$ nodos ocultos es

$$Y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j f\left(\beta_{0,j} + \sum_{k=1}^{p} \beta_{j,k} X_{k,t}\right) + \varepsilon_t, \tag{2.4}$$

donde $f(x) = (1 + e^x)^{-1}$. Para el caso de (2.3) se tiene $p = 2, q = 1$. En Venables and Ripley [2002], pag. 243:

Las redes neuronales proveen una manera flexible de generalización de modelos lineales. Son modelos de regresión no lineal pero con muchos parámetros, lo que las hace muy flexibles, capaces de aproximar cualquier función suave.

### 2.3.1. Estimación de los Parámetros

Los parámetros de los modelos (2.1) y (2.3), son: $\beta_0,\ \beta_1,\ \beta_2,\ \sigma^2$ y se busca estimarlos a partir de la muestra $Y_t, X_{1,t}, X_{2,t}, t = 1, 2, \ldots, T$. Algunos métodos de estimación son:

1. Mínimos cuadrados ordinarios. (MCO)
2. Mínimos cuadrados robustos. (MCR)
3. Máxima verosimilitud. (MLE)
4. Mínimos cuadrados no lineales.
5. Momentos.

En el caso MCO, los estimadores de los coeficientes $\beta_i$ se definen como el vector $\underline{\hat{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ que minimiza la distancia

$$G(\underline{\beta}) = \sum_{t=1}^{T} \left(Y_t - (\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t})\right)^2, \tag{2.5}$$

es decir, $\underline{\hat{\beta}}$ cumple que

$$\forall \underline{\beta},\ G(\underline{\hat{\beta}}) \leq G(\underline{\beta}).$$

Y se escribe

$$\underline{\hat{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}}\ G(\underline{\beta}).$$

Minimizar la función objetivo $G(\underline{\beta})$ es resolver el problema de estimación por MCO. El estimador $\underline{\hat{\beta}}$ se calcula mediante una fórmula. El estimador de MCO está dado por

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y}. \tag{2.6}$$

Este estimador $\underline{\hat{\beta}}$ es un estimador lineal insesgado, es decir, $\mathbb{E}(\underline{\hat{\beta}}) = \underline{\beta}$.

$$\begin{aligned} \mathbb{E}(\underline{\hat{\beta}}) &= \mathbb{E}((X'X)^{-1}X'\underline{Y}) \\ &= (X'X)^{-1}X'\mathbb{E}(X\underline{\beta} + \underline{\varepsilon}) \end{aligned}$$

$$\begin{aligned} &= (X'X)^{-1}X'(X\underline{\beta} + \mathbb{E}(\underline{\varepsilon})) \\ &= (X'X)^{-1}X'X\underline{\beta} = \underline{\beta}. \end{aligned}$$

Y es de mínima varianza, osea, $Var(\hat{\beta}_i) \leq Var(\hat{\beta}_i^*)$, $i = 0, 1, 2$, para $\hat{\beta}_i^*$ cualquier estimador diferente del de MCO.

El problema de estimación por MCO no requiere que $\varepsilon_t \sim iid\, N(0, \sigma^2)$. En cambio, la estimación MLE sí. El supuesto $\varepsilon_t \sim iid\, N(0, \sigma^2)$ implica que los estimadores MCO se distribuyen normal, $\widehat{\beta}_i \sim N(\cdot, \cdot)$, lo que permite definir los estadísticos de ajuste como la t y la F.

Una vez definidos los $\hat{\beta}_i$, se definen los pronósticos dentro de la muestra (valores ajustados) como

$$\widehat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \hat{\beta}_2 X_{2,t}, \quad t = 1, \ldots, T. \tag{2.7}$$

Nótese que se cumple $\underline{\widehat{Y}} = X\underline{\hat{\beta}}$. Reemplazando (2.6) es esta expresión se obtiene

$$\underline{\widehat{Y}} = X(X'X)^{-1}X'\underline{Y} = H\underline{Y}, \tag{2.8}$$

donde la matriz

$$H = X(X'X)^{-1}X' \tag{2.9}$$

se denomina la matriz "hat" ó "de proyección". Los pronósticos fuera de la muestra (predicciones) se definen como

$$\widehat{Y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,T+h} + \hat{\beta}_2 X_{2,T+h}, \quad h = 1, \ldots, m, \tag{2.10}$$

donde $X_{1,T+h},\ X_{2,T+h}$ hay que proveerlos.

Se define los residuos estimados como

$$\hat{\varepsilon}_t = Y_t - \widehat{Y}_t, \quad t = 1, \ldots, T. \tag{2.11}$$

En forma vectorial es

$$\underline{\hat{\varepsilon}} = \underline{Y} - X\underline{\hat{\beta}} = \underline{Y} - H\underline{Y} = (I - H)\underline{Y}. \tag{2.12}$$

De esta identidad se obtiene

$$\begin{aligned} \mathbb{E}(\underline{\hat{\varepsilon}}) &= \mathbb{E}(\underline{Y} - X\underline{\hat{\beta}}) \\ &= \mathbb{E}(\underline{Y}) - \mathbb{E}(X\underline{\hat{\beta}}) \\ &= \mathbb{E}(X\underline{\beta}) - X\mathbb{E}(\underline{\hat{\beta}}) \\ &= X\underline{\beta} - X\underline{\beta} = \underline{0}. \end{aligned}$$

Además, por definición de Varianza de un vector aleatorio,

$$\begin{aligned} Var(\underline{\hat{\varepsilon}}) &= \mathbb{E}(\underline{\hat{\varepsilon}}\,\underline{\hat{\varepsilon}}') \in \mathbb{R}^{T\times T} \\ &= \mathbb{E}((I-H)\underline{Y}\,[(I-H)\underline{Y}]') \\ &= (I-H)\mathbb{E}(\underline{Y}\,\underline{Y}')(I-H)' \\ &= (I-H)\mathbb{E}(\underline{Y}\,\underline{Y}')(I-H)'. \end{aligned}$$

Colocando

$$\mathbb{E}(\underline{Y}\,\underline{Y}') = Var(\underline{Y}) = Var(X\underline{\beta}+\underline{\varepsilon}) = \mathbb{E}(\underline{\varepsilon}\,\underline{\varepsilon}')$$

entonces la hipótesis $\varepsilon_t \sim iid(0,\sigma^2)$ implica $\mathbb{E}(\underline{\varepsilon}\,\underline{\varepsilon}') = \sigma^2 I$, con $I \in \mathbb{R}^{T\times T}$ la matriz identidad. Se puede comprobar que $H^2 = H$ y por tanto $(I-H)(I-H)' = I-H$ (ejercicio!), y por tanto

$$Var(\underline{\hat{\varepsilon}}) = \sigma^2(I-H)(I-H)' = \sigma^2(I-H). \tag{2.13}$$

Los elementos de la diagonal de $\sigma^2(I-H)$ son las varianzas $Var(\hat{\varepsilon}_t) = \sigma^2(1-H_{t,t})$. Los residuos estandarizados se definen como

$$r_t = \frac{\hat{\varepsilon}_t}{\widehat{\sigma}\sqrt{1-H_{t,t}}}, \tag{2.14}$$

donde $\widehat{\sigma}$ es el estimador de $\sigma$, definido como la raíz cuadrada del MSE. El MSE "error cuadrático medio estimado" se define como

$$\hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T}(Y_t-\hat{Y}_t)^2 = \frac{1}{T-k}\sum_{t=1}^{T}\hat{\varepsilon}_t^2, \tag{2.15}$$

donde $k$ es el número de coeficientes $\beta$ en el modelo.

**Ejemplo 2.3.1.** *Los datos de tres variables $Y, X1, X2$, para ajustar el modelo de regresión (2.1) están en la página web*
`http://www.stat.umn.edu/geyer/5102/data/ex5-3.txt`

```
#-------------------------------------------------------------
# programa en R
library(fRegression) # para estadisticos especiales de regresion

archivo = "http://www.stat.umn.edu/geyer/5102/data/ex5-3.txt"
D = read.table(archivo,header=T)
attach(D)
```

```
m2 =  lm(y ~ x1 + x2)
summary(m2)
anova(m2) #  tabla anova
#----------------------------------
(beta = coefficients(m2)) # coefficientes
confint(m2, level=0.95) # Intervalos de confianza para los parámetros

#----------------------------------
X = model.matrix(m2) # matriz de diseño

H = X%*%solve(t(X)%*%X)%*%t(X) # matrix de proyeccion

yhat = H%*%y
# yhat = fitted(m2) : valores ajustados

#----------------------------------

ehat = y-yhat
# ehat = residuals(m2) : residuos

#----------------------------------
(sigma.hat = sqrt(sum(ehat*ehat)/(nrow(H)-3)))

#----------------------------------
Id = mat.or.vec(nrow(H),ncol(H))

diag(Id) = rep(1,nrow(H))
r = ehat/(sigma.hat*sqrt(diag(Id-H)))

#----------------------------------
par(mfrow=c(1,1))
plot(density(ehat))
lines(density(r),col='red')
lines(density(rnorm(5000,0,1)),col='blue')

#----------------------------------
vcov(m2) # matriz de covariazas para los parametros
cov2cor(vcov(m2))
```

### 2.3.2. Pruebas de Ajuste del Modelo

**Pruebas de Significación de los parámetros $\beta_i$**

Si $\varepsilon_t \sim iidN(0, \sigma^2)$ se demuestra que

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\hat{\beta}_i}), \tag{2.16}$$

donde $\widehat{\sigma}_{\hat{\beta}_i} = \sqrt{\widehat{Var}(\hat{\beta}_i)} =$ error estándar. Además se cumple que

$$t_i = \hat{\beta}_i / \widehat{\sigma}_{\hat{\beta}_i} \sim t_{T-k}, \tag{2.17}$$

donde $k$ es el número de parámetros $\beta_i$, donde $t_i$ es el estadístico t-Student de la prueba de significancia de los parámetros $\beta_i$. Esta prueba se describe así. Para $i = 1, \ldots, k$,

1. HIpótesis nula versus alterna: $H_0: \ \beta_i = 0$ versus $H_1: \ \beta_i \neq 0$.
2. Estadístico de prueba $t_i := \hat{\beta}_i / \widehat{\sigma}_{\hat{\beta}_i}$.
3. Si $H_0$ es cierta entonces $t_i \sim t_{T-k}$ (T-Student con $T - k$ grados de libertad).
4. Decisión. Se plantean 3 maneras equivalentes para decidir acerca de la hipótesis nula $H_0$.

   1 Si $|t_i| > 1.96$ entonces se rechaza $H_0: \ \beta_i = 0$ con un nivel de significancia de $5\,\%$, es decir, hay una probabilidad de $5\,\%$ de rechazar $H_0$ siendo cierta.

   2 Se rechaza la nula a un nivel de significancia de $5\,\%$ si

   $$0 \notin [\hat{\beta}_i - 1.96\,\widehat{\sigma}_{\hat{\beta}_i}, \hat{\beta}_i + 1.96\,\widehat{\sigma}_{\hat{\beta}_i}]$$

   Note que, con R se pueden obtener los intervalos anteriores con `confint(m2, level=0.95)`

   3 Defina

   $$Valor.p := \mathbb{P}(|t_{T-k}| > |t_{obs}|) = \mathbb{P}(t_{T-k} > |t_{obs}|) + \mathbb{P}(t_{T-k} < -|t_{obs}|).$$

   Si Valor.p $< 0.05$ se rechaza $H_0$ a un nivel de significancia del $5\,\%$, de lo contrario no se rechaza $H_0$.

**Ejemplo 2.3.2.** *Continuando con el Ejemplo 2.3.1, las pruebas t-Student se pueden programar, como se muestra a continuación, pero resulta más eficiente utilizar el comando* `summary(modelo)`.

```
# matriz de covariazas para los parametros estimados
V=vcov(m2)
# matriz de correlaciones entre los parametros estimados
cov2cor(vcov(m2))
# estadisticos t-Student
(tStudent = beta/sqrt(diag(V)))
nu=nrow(H)-3 # es T-k
# valores p. Note la función pt()
(valor.p = pt(abs(tStudent),nu,lower.tail=FALSE)+
(pt(-abs(tStudent),nu,lower.tail=TRUE)))
# tabla resumen
err.std = sqrt(diag(V))
cbind(beta,err.std,tStudent,valor.p)
#-----------------------------------------------------
                beta   err.std  tStudent      valor.p
(Intercept) 0.4342085 0.9093713 0.4774821 6.352320e-01
x1          1.4178999 0.1719226 8.2473141 1.091459e-10
x2          0.6742687 0.1688444 3.9934327 2.274572e-04
#-----------------------------------------------------
# alternativa: summary
summary(m2)
```

**Prueba F**

Se define la suma de cuadrados total $SST$ como la suma de cuadrados de la regresión $SSR$ más suma de cuadrados de errores $SSE$, dados por

$$SST = \sum_{t=1}^{T}(Y_t - \bar{Y})^2 \tag{2.18}$$

$$= \sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2 \tag{2.19}$$

$$= SSR + SSE. \tag{2.20}$$

Se define la hipótesis

$$H_0 : \forall j = 1, 2, \ldots, k, \beta_j = 0$$
$$H_1 : \exists j, \ni b_j \neq 0$$

y el estadístico para esta prueba se define como

$$F = \frac{\frac{SSR}{k-1}}{\frac{SSE}{T-k}} \tag{2.21}$$

el cual se distribuye como una F-Fisher bajo $H_0$

$$F \overset{H_0}{\sim} F_{k-1,T-k}. \tag{2.22}$$

Si se rechaza la nula $H_0$, las variables $X_{1,t}, \ldots, X_{k-1,t}$ tienen capacidad para predecir los valores de $Y_t$. Si el valor observado de F es $F_{obs}$ y $F_{obs} > F_{k-1,T-k,\alpha}$, se rechaza $H_0$. De manera equivalente, si le Valor p, definido como Valor.p = $\mathbb{P}(F_{obs} > F_{k-1,T-k})$, cumple con ser menor que el nivel de significación $\alpha$, se rechaza la nula.

**Ejemplo 2.3.3.** *Continuando con el Ejemplo 2.3.1, con $k = 3$, $T = 50$, $\alpha = 0,05$, se tiene que $F_{k-1,T-k,\alpha} = F_{2,47,0.95} = 3.195056$ (Use:* `qf(0.95,2,47)`*). Luego, dado $F_{obs} = 66.48$, como $F_{obs} > 3.2$, se rechaza la nula $H_0$, es decir, $X_{1,t}$ y $X_{2,t}$ tienen capacidad predictiva sobre $Y_t$.*

```
summary(m2)
#--------
F-statistic: 66.48 on 2 and 47 DF,  p-value: 1.987e-14
```

### 2.3.3. Estadísticos de Ajuste y de Selección de Modelos

Se define el estadístico de ajuste R cuadrado, $R^2$, como

$$R^2 = 1 - \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} \tag{2.23}$$

como el porcentaje de la varianza de $Y_t$ que es explicado (atribuible) por las variables explicativas. Observe que $k$ no aparece.

$R^2 \geq 0.7$ buena

$R^2 = 0.6$ buena pero no mucho

$R^2 = 0.4$ regular-mala

$R^2 = 0.2$ desechar, seguramente la prueba $F$ y las pruebas $t$ no rechazan la nula

Se define el $R$ cuadrado ajustado, $\bar{R}^2$ como

$$\bar{R}^2 = 1 - \frac{(T-1)SSE}{(T-k)SST}. \tag{2.24}$$

Note que $R^2 = 1 - SSE/SST$, por tanto, $\bar{R}^2$ se puede escribir en terminos de $R^2$ de la siguiente forma

$$\bar{R}^2 = 1 - \frac{T-1}{T-k}(1-R^2). \tag{2.25}$$

Si $k \nearrow$, entonces $T - k \searrow$ y $\frac{T-1}{T-k} \nearrow$, luego $\bar{R}^2 \searrow$. Es decir, $\bar{R}^2$ penaliza aumentar el número de parámetros.

Se define el criterio de información de Akaike, AIC, como

$$AIC = e^{2k/T}\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T} = \frac{(T-k)\hat{\sigma}^2}{T}e^{2k/T}. \tag{2.26}$$

donde $\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = (T-k)\widehat{\sigma}^2$. El AIC es un estimador de $\sigma^2$ pero penalizado por el número de grados de libertad, es decir, aumenta cuando $k$ aumenta. El AIC también se define como el logaritmo de (2.26).

$$AIC = \log\left(\frac{(T-k)\widehat{\sigma}^2}{T}\right) + \frac{2k}{T}. \tag{2.27}$$

De la definición del criterio $AIC$ se observa:

- Si $k \nearrow$, entonces $AIC \nearrow$.
- Si $\widehat{\sigma}(= MSE) \searrow$, entonces $AIC \searrow$.

En Diebold [1999, pág. 73]

> "Como en el caso de $S^2$, muchos de los criterios más importantes para la selección de modelos de pronóstico tienen la forma de factor de penalización multiplicado por $MSE$".

Se define el criterio de información de Schwarz BIC

$$BIC = T^{\frac{k}{T}}\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T} = T^{\frac{k}{T}}\frac{(T-k)\hat{\sigma}^2}{T} \tag{2.28}$$

El BIC también se define como el logaritmo de (2.28).

$$BIC = \log\left(\frac{(T-k)\widehat{\sigma}^2}{T}\right) + \frac{k}{T}\log(T). \tag{2.29}$$

De la definición del criterio $BIC$ se observa:

- Si $k \nearrow$, entonces $BIC \nearrow$.

- Si $\widehat{\sigma}(= MSE) \searrow$, entonces $BIC \searrow$.

La regla para utilizar AIC y BIC es que para escoger entre varios modelos de regresión con respecto a la misma variable dependiente, todos lineales, ó todos no lineales pero con la misma estructura, es decir, modelos anidados, se escoge el de menor AIC ó menor BIC.

Diebold [1999, pág. 74–75], introduce dos criterios para elegir entre AIC y BIC.

1. Consistencia.
2. Eficiencia asintótica.

**Definición 2.3.1 (Consistencia).** *Si se examinan varios modelos, y el modelo generador de los datos (MGD) está incluído, el criterio se dice consistente si a medida que $T$ aumenta, la probabilidad de que el criterio seleccione el MGD tiende a 1.*

**Resultado:** $\bar{R}^2$, $AIC$ no son consistentes pero $BIC$ sí es.

**Definición 2.3.2 (Eficiencia Asintótica).** *Un criterio es eficiente asintóticamente si a medida que $T$ aumenta, elige varios modelos cuyas varianzas de error de pronóstico a un paso tienden a la que se obtendría con el MGD. Donde el error de pronóstico a un paso está dado por $Y_{T+1} - \widehat{Y}_{T+1}$*

**Resultado:** $AIC$ es eficiente asintóticamente pero $BIC$ no. En Diebold [1999, pág. 75]:

> "Muchos autores recomiendan usar el modelo más parsimonioso que selecciona el $BIC$ en igualdad de circunstancias."

### 2.3.4. Mínimos Cuadrados Nolineales

En algunos casos es necesario ajustar un modelo no lineal. R posee funciones que permiten ajustar por mínimos cuadrados no lineales modelos de la forma

$$Y_t = g(\underline{X}_t, \beta) + \epsilon_t, t = 1, 2, \ldots, T. \tag{2.30}$$

Nótese que debe ser un modelo con errores aditivos. La función en R para mínimos cuadrados no lineales es `nls()` y se aplica de la misma forma que `lm()`. Por ejemplo, para el modelo no lineal (2.3)

$$Y_t = \frac{1}{1 + e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}}} + \varepsilon_t, \tag{2.31}$$

Se puede estimar con la función `nls()`, con la instrucción

```
m = nls(Y ~ 1/(1+exp(b0+b1*X1+b2*X2)),
start=list(b0=0.1, b1=1, b2=1))
```

Nótese que la función requiere unos valores iniciales para los parámetros. Con la transformación $Z = \log(1/Y - 1)$ se puede linearizar el modelo y colocar $Z = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}$. Con este modelo se podrían obtener estimadores iniciales de los parámetros. En el capítulo siguiente hay un ejemplo concreto de estimación de este modelo.

**Ejemplo 2.3.4.** *Continuando con el Ejemplo 2.3.1, se examinan tres modelos lineales:*

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \varepsilon_t, \tag{2.32a}$$

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \tag{2.32b}$$

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \tag{2.32c}$$

*Los resultados aparecen en el Cuadro 2.1. Las instrucciones en R se dan a continuación.*

Tabla 2.1: Resultados de los Modelo Regresión

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 2.3778 | 0.8797 | 2.70 | 0.0095 |
| x1 | 1.6979 | 0.1798 | 9.45 | 0.0000 |
| | Estimate | Std. Error | t value | Pr(>\|t\|) |
| (Intercept) | 0.4342 | 0.9094 | 0.48 | 0.6352 |
| x1 | 1.4179 | 0.1719 | 8.25 | 0.0000 |
| x2 | 0.6743 | 0.1688 | 3.99 | 0.0002 |
| | Estimate | Std. Error | t value | Pr(>\|t\|) |
| x1 | 1.4623 | 0.1435 | 10.19 | 0.0000 |
| x2 | 0.7174 | 0.1415 | 5.07 | 0.0000 |

```
#-------------------------------
m1 =  lm(y ~ x1)
summary(m1)

m2 =  lm(y ~ x1 + x2)
summary(m2)

m3 =  lm(y ~ -1 + x1 + x2)
summary(m3)
```

Las Tablas Anova para `m2`

```
#-----------------------------------------------------
anova(m2)
Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1 110.593 110.593 117.005 2.391e-14 ***
x2         1  15.074  15.074  15.947 0.0002275 ***
Residuals 47  44.424   0.945
```

Del help de anova:

> When given a sequence of objects, anova tests the models against one another in the order specified.

```
#-----------------------------------------------------
anova(m1,m2,m3)

Analysis of Variance Table

Model 1: y ~ x1
Model 2: y ~ x1 + x2
Model 3: y ~ -1 + x1 + x2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     48 59.498
2     47 44.424  1   15.0736 15.947 0.0002275 ***
3     48 44.640 -1   -0.2155  0.228 0.6352320
```

Calcular estadísticos de selección de modelos

```
#--------------------------
# funcion:
medidas = function(m,y,k){
# m = objeto producido con lm()
# y = variable dependiente
```

```
# k = número de coeficientes beta
T = length(y)
yest = fitted(m)
sse = sum((yest-y)^2)
ssr = sum((y-mean(y))^2)
mse = sse/(T-k)
R2 = 1 - sse/ssr
Ra2 = 1 - (T-1)*(1-R2)/(T-k)
aic = log((T-k)*exp(2*k/T)*mse/T)
bic = log(T^(k/T)*(T-k)*mse/T)
M = c(mse,Ra2,  aic, bic)
names(M) = c("mse","R2-ad","log.aic","log.bic")
return(M)
}
#-------------------------
# también: source("medidas.r")

N1 = medidas(m1,y,k1)
N2 = medidas(m2,y,k2)
N3 = medidas(m3,y,k3)

N = cbind(N1,N2,N3)
rownames(N) = c("mse", "R2 ajus", "AIC", "BIC")
colnames(N) = c("m1","m2","m3")
(N)
#---------------------------------------
              m1          m2          m3
mse     1.2395425 0.945200935  0.92999874
R2 ajus 0.6429115 0.727705658  0.73208512
AIC     0.2539204 0.001766852 -0.03339404
BIC     0.3304013 0.116488232  0.04308688
```

Se podría escoger el modelo `m2` ya que los valores del R cuadrado ajustado para el modelo 3 no son válidos por ser un modelo sin intercepto. Además, los valores estimados de la variable $Y_t$, dentro de la muestra, son muy similares para ambos, el modelo 2 y el modelo3.

# CAPÍTULO 3

## Modelado y Pronóstico de la Tendencia

Una serie de tiempo es una sucesión de variables aleatorias ordenadas de acuerdo a una unidad de tiempo, $\{Y_1, \ldots, Y_T\}$. Por ejemplo, la concentración en el aire de dióxido de azufre en ppm (partes por millón), ($100ppm = 262mg/m^3$), medida semanalmente en una estación de monitoreo, es importante para vigilar la calidad del aire en una ciudad.

En este capítulo se introduce el Modelo Aditivo de Componentes para series de tiempo, y se analizan modelos específicos para la componente de la tendencia.

## 3.1. El Modelo Aditivo de Componentes de series de tiempo

Dada una serie $\{Y_t, t = 1, \ldots, T\}$, el Modelo Aditivo de Componentes consiste en asumir que $Y_t$ se puede descomponer en tres componentes:

$$Y_t = T_t + S_t + \varepsilon_t, \tag{3.1}$$

donde $T_t$ es la tendencia, $S_t$ es la componente estacional y $\varepsilon_t$ es la componente de errores. Las componentes $T_t$ y $S_t$ son funciones determinísticas de $t$. Su evolución es completamente predecible.

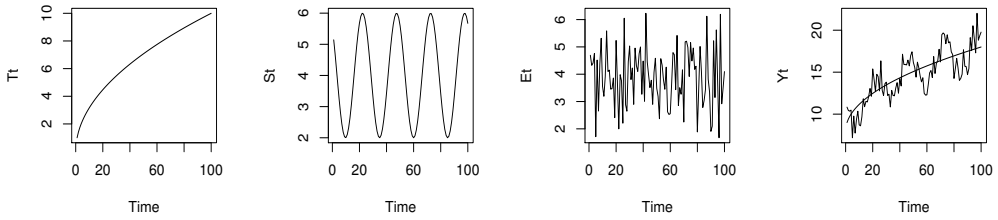
El modelo (3.1) se denomina también modelo Estructural Básico,

La componente $T_t$ en algunos casos puede ser estacional (es decir, periódica, de período grande). Por ejemplo, en una serie diaria, $S_t$ puede tener período 30 días, y $T_t$ período 360 días.

El modelo multiplicativo consiste en asumir que $Y_t$ se puede descomponer en tres componentes:

$$Y_t = T_t S_t e^{\varepsilon_t}. \tag{3.2}$$

En lo que sigue solamente se tratará el modelo aditivo. En la Figura (3.1) siguiente se muestra la idea de la descomposición. Al superponer las series en los paneles (a), (b) y (c) se obtiene la serie en el panel (d).




(a) Tendencia (b) Patrón Estacional (c) Ruido (d) Serie

Figura 3.1: Descomposición de la Serie $Y_t$

El análisis consiste en modelar y estimar $T_t$ y $S_t$ y luego extraerlas de $Y_t$ para obtener

$$\hat{\varepsilon}_t = Y_t - \hat{T}_t - \hat{S}_t.$$

La serie $\hat{\varepsilon}_t$ se modela y estima para finalmente reconstruír $Y_t$, $\hat{Y}_t = \hat{T}_t + \hat{S}_t + \hat{\varepsilon}_t$, y poder realizar el pronóstico $\hat{Y}_{T+h} = \hat{T}_{T+h} + \hat{S}_{T+h} + \hat{\varepsilon}_{T+h}$, utilizando la información disponible $Y_1, \ldots, Y_T$ con $h = 1, 2, \ldots, m$. Sin embargo, puede suceder que la serie $\hat{\varepsilon}_t$ sea incorrelacionada, es decir, $Corr(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+s}) = 0$, para $s \neq 0$. En este caso $\hat{\varepsilon}_{T+h} = 0, \forall h > 0$.

**Definición 3.1.1 (Tendencia).** *Se define como una función $T_t$ de $t$ que describe la evolución lenta y a largo plazo del nivel medio de la serie.*

La función $T_t$ puede depender de parámetros, que deben estimarse. Pero también es posible utilizar modelos no-paramétricos, como filtros, medias móviles y regresiones locales. Los

modelos paramétricos se pueden denominar “modelos globales”, en contraposición a los no-paramétricos, denominados “modelos locales”.

**Modelos:** Una lista de posibles modelos paramétricos para $T_t$ es:

$$\text{Lineal:} \quad T_t = \beta_0 + \beta_1 t, \tag{3.3a}$$

$$\text{Cuadrático:} \quad T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \tag{3.3b}$$

$$\text{Cúbico:} \quad T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3, \tag{3.3c}$$

$$\text{Exponencial:} \quad T_t = \exp(\beta_0 + \beta_1 t). \tag{3.3d}$$

$$\text{Logístico:} \quad T_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)}, \tag{3.3e}$$

En la tendencia cuadrática (3.3b) se cumple:

1. Si $\beta_1, \beta_2 > 0$, $T_t$ es monótona creciente.
2. Si $\beta_1, \beta_2 < 0$, $T_t$ es monótona decreciente.
3. Si $\beta_1 > 0$ y $\beta_2 < 0$, $T_t$ es cóncava hacia abajo (o cóncava).
4. Si $\beta_1 < 0$ y $\beta_2 > 0$, $T_t$ es cóncava hacia arriba (o convexa).

**Ejemplo 3.1.1.** *El IPVU, índice de precios de la vivienda usada es un indicador macroeconómico calculado por el Banco de la República, que se obtiene en el sitio:*
`http://www.banrep.gov.co/es/ipvu`.
*En el archivo Excel generado se lee:*

> *El Índice nominal de precios de la vivienda usada se refiere al resultado del cálculo del IPVU contemplando el efecto de la inflación*
>
> *El Índice real de precios de la vivienda usada se refiere al resultado del cálculo del IPVU eliminando el efecto de la inflación (i.e IPC promedio ponderado de las ciudades incluidas en el cálculo).*
>
> *El Índice total de precios de la vivienda usada se refiere al cálculo del IPVU para las ciudades de Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Cúcuta, Manizales, Neiva y Villavicencio. La ciudad de Bogotá incluye el municipio de Soacha y la ciudad de Medellín incluye los municipios de Bello, Envigado e Itagüí.*

*Las gráficas del IPVU nominal y real se pueden ver en la Figura 3.2*

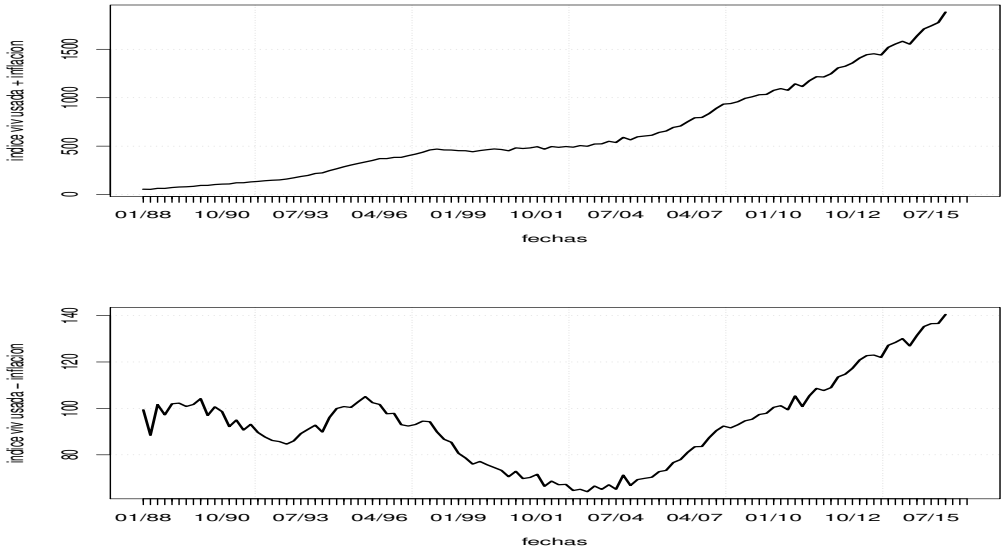



Figura 3.2: Indices de vivienda usada nominal y real

**Ejemplo 3.1.2.** *Datos sobre tasa de empleo en varios países de la OECD (Organisation for Economic Co-operation and Development ([1])) se pueden consultar en la página web* `https://data.oecd.org/emp/employment-rate.htm`. *La definición de tasa de empleo se define así:*

> *Employment rates are defined as a measure of the extent to which available labour resources (people available to work) are being used. They are calculated as the ratio of the employed to the working age population... Employed people are those aged 15 or over who report that they have worked in gainful employment for at least one hour in the previous week or who had a job but were absent from work during the reference week. The working age population refers to people aged 15 to 64. This indicator is seasonally adjusted and it is measured in terms of thousand persons aged 15 and over; and as a percentage of working age population.*

[1] Organización para la Cooperación y el Desarrollo Económicos (OCDE)

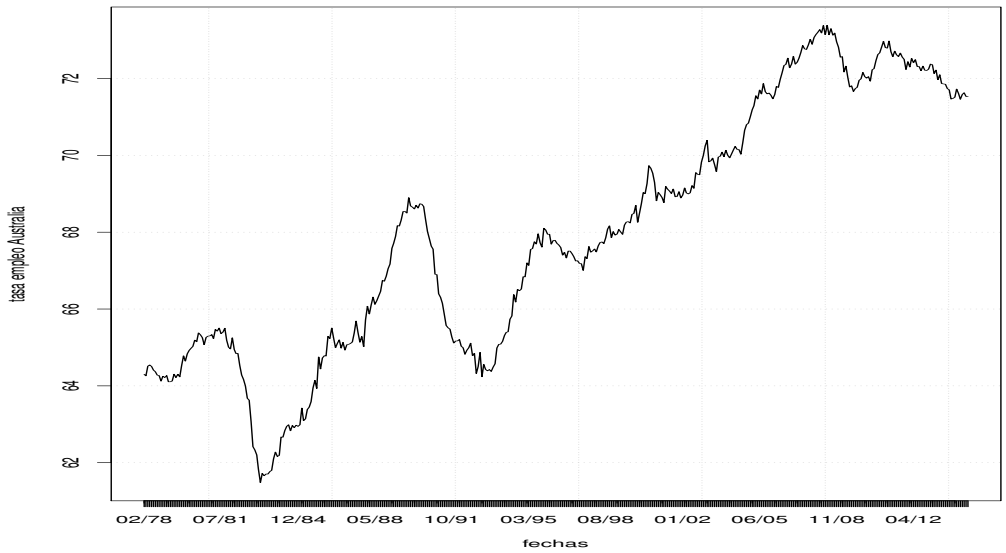



Figura 3.3: Tasa de Empleo, Australia, 02/1978 - 09/2014

*Por ejemplo, los datos mensuales de Australia entre 02/1978 y 08/2104 se pueden ver en la Figura 3.3. Un análisis de la tendencia requiere considerar el cambio de tendencia que parece darse desde finales de 2008.*

## 3.2. Estimación de la Tendencia

En este capítulo se introduce la estimación de la tendencia mediante modelos de regresión lineal y no lineal. En un capítulo posterior se introducen modelos no paramétricos para estimar la tendencia. Son modelos de regresión no paramétrica que incluyen suavizadores, filtros lineales y no lineales y medias móviles. Hay otros métodos que no se consideran en este curso, por ejemplo, wavelets. En ocasiones la expresión "suavizar una serie" es equivalente a "extracción de la tendencia de una serie", y ambas equivalen a la estimación de la tendencia.

Para la estimación de los parámetros $\underline{\beta} = (\beta_0, \beta_1, \beta_2)'$ en los modelos lineales (3.3a),(3.3b), (3.3c) y (3.5) se utiliza el método de mínimos cuadrados ordinarios (MCO). En este método el vector de parámetros estimados $\hat{\underline{\beta}}$ es el vector que produce el valor mínimo de la suma de errores cuadrados. Es decir $\hat{\underline{\beta}}$ es el valor en el cual $G(\underline{\beta}) = \sum_{t=1}^{T} \left(Y_t - T_t(\underline{\beta})\right)^2$ toma el

valor mínimo.

$$\underline{\hat{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}} G(\underline{\beta}). \tag{3.4}$$

Para los modelos (3.3d) y (3.3e) se usa el método de mínimos cuadrados no lineales, que también minimiza la suma de errores cuadrados $G(\underline{\beta}) = \sum_{t=1}^{T} \left(Y_t - T_t(\underline{\beta})\right)^2$, pero $T_t(\underline{\beta})$ es una función no lineal de $\underline{\beta}$.

Aunque la serie tenga una componente estacional $S_t$, $Y_t = T_t + S_t + \varepsilon_t$, solamente consideramos un modelo de regresión entre $Y_t$ y $T_t$, tal que $Y_t = T_t + \eta_t$, donde $\eta_t$ es el término de error, de forma que $\eta_t = S_t + \varepsilon_t$. Por ejemplo.

1. En el caso lineal $\underline{\beta} = (\beta_0, \beta_1)'$, con $T_t = \beta_0 + \beta_1 t$ se ajusta el modelo de regresión lineal: $Y_t = \beta_0 + \beta_1 t + \eta_t$.

2. En el caso cuadrático $\underline{\beta} = (\beta_0, \beta_1, \beta_2)'$, con $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ se ajusta el modelo de regresión lineal $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \eta_t$. Nótese que en este caso hay que definir variable explicativa adicional $t^2$.

3. El modelo con tendencia exponencial (3.3d),

   $$Y_t = \exp(\beta_0 + \beta_1 t) + \varepsilon_t$$

   es no lineal y es casi equivalente a un modelo con tendencia lineal para el logaritmo de $Y_t$, denominado modelo Logarítmico Lineal ó Log-Lineal, definido como

   $$\ln Y_t = \beta_0 + \beta_1 t + \varepsilon_t. \tag{3.5}$$

   Es posible estimar por mínimos cuadrados ordinarios el modelo Log-Lineal (3.5) y utilizar los parámetros estimados $\hat{\beta}_0, \hat{\beta}_1$ como valores iniciales en la estimación del modelo con tendencia exponencial por mínimos cuadrados no lineales. Pero los parámetros estimados en ambos modelos no necesariamente coinciden.

## 3.3. Pronósticos con base en la Tendencia

Suponga la serie con tendencia

$$Y_t = T_t + \eta_t,\ t = 1, \ldots, T,$$

con $(\eta_t, t \geq 1)$ una sucesión $iid(0, \sigma^2)$. Los pronósticos de $Y_t$ en los tiempos $T+1, T+2, \ldots, T+h,\ h \geq 1$ se definen como

**Definición 3.3.1.**

$$\widehat{Y}_{T+j} = \widehat{T}_{T+j}, j = 1, \ldots, h. \tag{3.6}$$

donde $\widehat{T}_t$ es la función estimada de la tendencia. Por ejemplo, en el modelo lineal

$$Y_t = \beta_0 + \beta_1 t + \eta_t,$$

$\widehat{T}_t = \hat{\beta}_0 + \hat{\beta}_1 t$. Al reemplazar $t$ por $T + h$ se obtiene $\widehat{T}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1(T + h) + \hat{\eta}_{T+h}$. Los pronósticos $\hat{\eta}_{T+h}$, pueden o no ser cero. Son cero si los residuos $\hat{\eta}_t,\ t = 1, \ldots, T$ son incorrelacionados. Para decidir esto se realiza una prueba de incorrelación y una prueba de normalidad. Por el contrario, puede suceder que los residuos estimados $\hat{\eta}_t$ sean autocorrelacionados y por tanto, diferentes de cero. En este caso el pronóstico es

$$\widehat{Y}_{T+j} = \widehat{T}_{T+j} + \hat{\eta}_{T+j}.$$

La definición general de pronóstico, para una serie $Y_t, t \in \mathbb{Z}$, con base en la información $Y_1, \ldots, Y_T$, es una esperanza condicional, como sigue.

**Definición 3.3.2.**

$$\widehat{Y}_{T+j} = \mathbb{E}(Y_{T+j}|Y_1, \ldots, Y_T), j = 1, \ldots, h. \tag{3.7}$$

Otros tipos de pronósticos son

- El pronóstico por intervalo se obtiene si se asume que $\eta_t \sim iid\, N(0, \sigma^2)$. Entonces se cumple que $\hat{Y}_{T+h} \pm 1.96\hat{\sigma}$ es un intervalo de confianza del 95 %.
- Si se ignora los errores en $\hat{\beta}_0$ y $\hat{\beta}_1$ el IC es $\hat{\beta}_0 + \hat{\beta}_1(T + h) \pm 1.96\hat{\sigma}$, donde $\hat{\sigma}^2 = MSE = \frac{1}{T-2}\sum_{t=1}^{T} e_t^2$
- El pronóstico de densidad es $Y_{T+h} \sim N(\hat{Y}_{T+h}, \hat{\sigma}^2)$.

## 3.4. Medidas de Precisión en los Pronósticos

Suponga que los últimos $m$ datos se utilizan para constrastar los pronósticos realizados a partir del modelo estimado con los $T - m$ primeros datos. Este procedimiento se denomina "validación cruzada".

Asuma que $t = T-m+1, \ldots T$. Suponiendo el valor del pronóstico $\widehat{Y}_t$ y el residuo estimado $\hat{\eta}_t = Y_t - \widehat{Y}_t$ se pueden definir varias medidas para la precisión del pronóstico, ver Hyndman and Koehler [2006]. Con base en

$$p_t = 100\frac{\hat{\eta}_t}{Y_t} = 100\frac{Y_t - \widehat{Y}_t}{Y_t},$$

se definen tres medidas.

1. El error absoluto porcentual medio ó MAPE:

$$MAPE = \frac{1}{m} \sum_{t=T-m+1}^{T} |p_t|. \tag{3.8}$$

2. La raíz cuadrada del error cuadrático medio, RMSE:

$$RMSE = \sqrt{\frac{1}{m} \sum_{t=T-m+1}^{T} \hat{\eta}_t^2}. \tag{3.9}$$

## 3.5. Caso de Estudio: Pronóstico de Ventas al Menudeo

Este caso está analizado en Diebold [1999, sección 4.5]. El objetivo aquí es repetir el análisis utilizando R. Los Modelos a utilizar son los modelos lineal, cuadrático, cúbico y exponencial. Se aplicarán los criterios AIC y BIC para escoger el más adecuado.

### 3.5.1. Descripción de los Datos

- Ventas al menudeo en USD a precios de USD en 1999
- Periodicidad: Mensual
- Fechas: 01/1955 - 12/1994
- Número de observaciones: 468
- Datos ajustados estacionalmente. Es decir, si se tiene $Z_t = T_t + S_t + \varepsilon_t$, se dice que $Y_t$ está ajustada estacionalmente o desestacionalizada si $Y_t = Z_t - S_t = T_t + \varepsilon_t$. Es decir, es una serie en la que se eliminó la componente estacional.
- Usar el período 01/1955 - 12/1993 para estimar los modelos (período de entrenamiento) y el periodo 01/1994 - 12/1994 para examinar la eficiencia de los pronósticos fuera de la muestra (“validación cruzada”).

### 3.5.2. Programa en R

**Lectura de Datos, Graficación con fechas**

```
D = read.table("ventas_al_menudeo.dat",header=T)
attach(D) # utiliza el nombre de las columnas como variables

# no hay variable con fechas : mensual 01/1955 hasta 12/1994
# para 469 obs, des-estacionalizadas
# RTTR es el volumen de ventas en grandes almacenes en USD de 1999
# La variable RTRR del archivo tiene datos faltantes NA al inicio
# y al final

y = na.omit(RTRR)/10000

# Convertir los datos en un objeto tipo ts
y = ts(y,frequency=12,start=c(1955,01))

# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1955/1/1"), length.out = length(y), by =  "months")

# grafica con fechas
ts.plot(y,main="Ventas al menudeo en UDS de 1999")

# otros comandos para graficar con fechas con mas detalle: mes-año

np = length(y)
ejex.mes = seq(fechas[1],fechas[np], "months")
ejex.año = seq(fechas[1],fechas[np],"years")

plot(fechas,y, xaxt="n", panel.first = grid(),type='l',
ylab='ventas.mes', lwd = 2)
axis.Date(1, at=ejex.mes, format="%m/%y")
axis.Date(1, at=ejex.año, labels = FALSE, tcl = -0.2)
```

**Ajuste de los Modelos utilizando Regresión Lineal**

Ajustar cuatro modelos: lineal, cuadrático, cúbico y exponencial, mediante regresión lineal, utilizando la función de R, lm().

```
# Generar datos para validacion cruzada: dejar el ultimo año

T = length(y)
yi = y[1:(T-12)]
yf = y[(T-12+1):T]

# Ajustar 4 modelos: lineal, cuadrático, cúbico, log-lin

t = seq(1:(T-12))
t2 = t^2
t3 = t^3
lyi = log(yi)

# estimacion por minimos cuadrados ordinarios
mod.lin = lm(yi~t)
mod.cuad = lm(yi~t+t2)
mod.cub = lm(yi~t+t2+t3)
mod.llin = lm(lyi~t) # auxiliar para el exponencial

summary(mod.lin)
summary(mod.cuad)
summary(mod.cub)
```

**Ajuste del Modelo Exponencial Lineal**

```
# paso 1) estimar el modelo auxiliar log - linear
mod.llin = lm(lyi~t)

# paso 2) guardar los parametros del log-lineal
b0.est = mod.llin$coefficient[1]
b1.est = mod.llin$coefficient[2]
```

```
# paso 3) guardar los datos en un data.frame
Ds = data.frame(yi,t)

# paso 4) usar la funcion nls
mod.exp = nls(yi~exp(beta0+beta1*t),
data=Ds,
start=list(beta0=b0.est, beta1=b1.est))

# paso 5) resultados
summary(mod.exp)
```

**Ajuste de los Modelos**

Tabla 3.1: Ajuste de los Modelos Lineal, Cuadrático, Cúbico y Exponencial

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -17537.673441 | 1503.384304 | -11.665463 | 0.000000 |
| t | 340.738727 | 5.405165 | 63.039465 | 0.000000 |
| (Intercept) | 18886.034127 | 383.923338 | 49.192201 | 0.000000 |
| t | -111.729690 | 3.678499 | -30.373717 | 0.000000 |
| t2 | 0.938731 | 0.007390 | 127.020988 | 0.000000 |
| (Intercept) | 19386.588312 | 513.151618 | 37.779455 | 0.000000 |
| t | -124.127295 | 9.210396 | -13.476868 | 0.000000 |
| t2 | 1.002967 | 0.044379 | 22.600006 | 0.000000 |
| t3 | -0.000089 | 0.000061 | -1.467890 | 0.142793 |
| (Intercept) | 9.390e+00 | 1.432e-02 | 655.8 | 0.000000 |
| t | 5.769e-03 | 3.536e-05 | 163.1 | 0.000000 |

Nótese de los resultados del cuadro 3.1 que el modelo cúbico no ajusta ya que el coeficiente de $t^3$ no dá significativo al nivel de 5 $\%$. Se procede a calcular los estadísticos AIC y BIC.

**Calcular los Estadísticos de Selección del Modelo**

Para calcular los estadísticos de Selección: R2 ajustado, MSE, log(AIC), log(BIC) se usará la función `medidas()`, que los calcula para cada modelo.

```
medidas = function(m,y,k){
# y = serie, m = modelo, k = numero parametros
T = length(y)
yest = fitted(m)
sse = sum((yest-y)^2)
ssr = sum((y-mean(y))^2)
mse = sse/(T-k)
R2 = 1 - sse/ssr
Ra2 = 1 - (T-1)*(1-R2)/(T-k)
aic = log((T-k)*exp(2*k/T)*mse/T)
bic = log(T^(k/T)*(T-k)*mse/T)
M = c(Ra2, sqrt(mse), aic, bic)
names(M) = c("R2-ajus","raiz(MSE)","logAIC","logBIC")
return(M)}

M.lin = medidas(mod.lin,yi,2)
M.cuad = medidas(mod.cuad,yi,3)
M.cub = medidas(mod.cub,yi,4)
M.exp =  medidas(mod.exp,yi,2)

M = cbind(M.lin,M.cuad,M.cub,M.exp)
(M)
```

| | M.lin | M.cuad | M.cub | M.exp |
|---|---|---|---|---|
| R2-ajus | 0.892 | 0.997 | 0.997 | 0.989 |
| MSE | 2.709 | 0.078 | 0.078 | 0.264 |
| logAIC | 1.001 | -2.543 | -2.544 | -1.327 |
| logBIC | 1.018 | -2.517 | -2.509 | -1.309 |

Tabla 3.2: Estadísticos de Selección de Modelos

Del cuadro 3.2 se puede concluír que el modelo que mejor ajusta los datos es el modelo cuadrático (3.3b) ya que presenta el menor BIC.

**Chequeo de las Hipótesis del Modelo de Regresión**

En este momento el diagnóstico de incorrelación y normalidad de los errores se hace mediante algunos procedimientos gráficos, que se programan con las instrucciones siguientes

```
r = mod.cuad$residuals

par(mfrow=c(2,2))
plot(t,r,type='o',ylab='residuo')
abline(h=0,lty=2)
plot(density(r),xlab='x',main= '')
qqnorm(r)
qqline(r,col=2)
acf(r,ci.type="ma",60)
```

**Cálculo de Pronósticos con el Modelo Cuadrático**

En esta última parte se examina cómo es la calidad de los pronósticos con el modelo cuadrático (3.3b). Se utiliza la técnica de validación cruzada que consiste en calcular los pronósticos con el modelo ajustado con el primer grupo de datos y compararlos con los datos reales del segundo grupo. Se hace con las siguientes instrucciones.

```
# Pronósticos
tt = seq(482,493,1)
tt2 = tt^2
pr2 = predict(mod.cuad,data.frame(t=tt,t2=tt2))

plot(tt,yf,type="b")
lines(tt,pr2,col="red")
```

Los pronósticos están en la variable "pr2" generada en la última parte del programa. A partir del examen de la figura (3.4) se puede concluír que el modelo cuadrático genera pronósticos confiables a corto plazo.

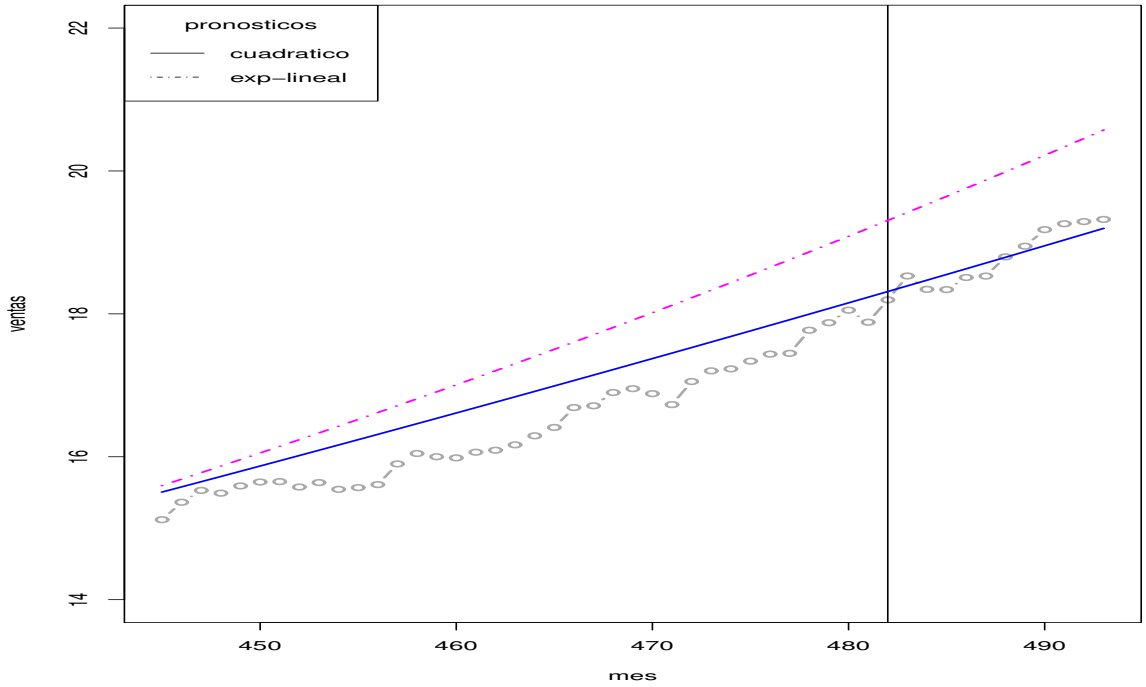



Figura 3.4: Pronósticos (—-) versus Observados (-o-o-)

## 3.6. Problemas

1. Con base en los datos de la tasa de empleo en Australia, en la Figura 3.3, ajuste varios modelos de tendencia, dejando como período de validación cruzada los 3 últimos años (36 meses). Los datos se pueden descargar de la página web mencionada en el Ejemplo 3.1.2.

2. En el Cuadro 3.3 se muestran los datos del total de habitantes en Medellín, según los Censos desde 1905 hasta 2005, según el DANE ([2]). En Poveda [1982] se planteó ajustar los datos de desde 1938 hasta 1973 mediante un modelo de componentes con tendencia logística, (3.3e).

$$Y_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)} + \varepsilon_t. \tag{3.10}$$

   Realice el ajuste de la tendencia utilizando un modelo de tendencia logístico y pronostique a 10 años.

[2] http://es.wikipedia.org/wiki/Demografía_de_Medellín

Tabla 3.3: Población de Medellín censos 1905 - 2005

| | Año | Población |
|---|---|---|
| 1 | 1905 | 59.810 |
| 2 | 1912 | 70.550 |
| 3 | 1918 | 79.150 |
| 4 | 1928 | 120.040 |
| 5 | 1938 | 168.270 |
| 6 | 1951 | 358.190 |
| 7 | 1964 | 772.890 |
| 8 | 1973 | 1.077.250 |
| 9 | 1985 | 1.468.090 |
| 10 | 1993 | 1.630.010 |
| 11 | 2005 | 2.223.080 |

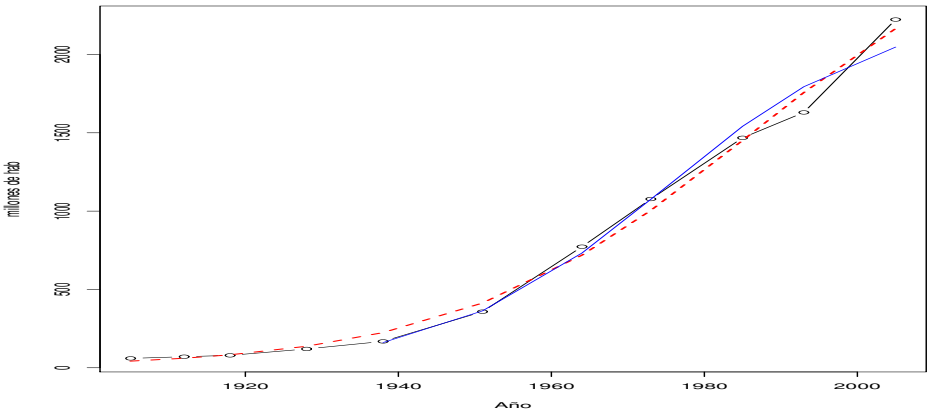



Figura 3.5: Ajuste Logístico a Población de Medellín

3. Consumo de agua mensual en el área metropolitana en metros cúbicos, por estrato, para los estratos 1,2,3,4,5,6, incluyendo el consumo total. Desde enero de 2002. En total son 7 series que se identificarán con los nombres : consumo 1, consumo 2,..., consumo 6, consumo total. Los datos están en el archivo consumo.dat, en la página del curso. Diseñe un plan inicial de análisis y descríbalo la introducción del trabajo. Por ejemplo, analizar las tendencias de todos los estratos, determinar si hay estacionalidad en todos o algunos.

*a*) Lea los datos a partir del archivo asignado que se encuentra en la página web del curso. Grafique la serie. Desarrolle los siguientes puntos y reporte los resultados de cada uno. Con estos resultados elabore el reporte del trabajo.

*b*) Use la función de descomposición stl para analizar lo siguiente:

1) Hay una tendencia global que puede describirse por una función lineal, cuadrática ó cúbica?
2) La tendencia se puede describir mejor por secciones? Es decir, no hay tendencia global?
3) Se puede observar una componente estacional?. Cuál es el período?.

*c*) Si observa que es factible que existan tendencia y estacionalidad, estímelas conjuntamente, con la restricción que se definión para la estacionalidad de considerar solamente $s - 1$ componentes estacionales (variables indicadoras), donde $s$ es el período estacional. Reporte una tabla con los parámetros estimados, errores estándar, estadísticos t y valores p. Reporte el AIC y el R-cuadrado ajustado.

*d*) Si observa solamente tendencia ó estacionalidad, estímelas y reporte los estadísticos del punto anterior. En el caso de estacionalidad reporte el gráfico el patrón estacional, es decir, los coeficientes $\widehat{\delta}_i$ versus $i = 1, 2, \ldots, s$.

*e*) Reporte qqplot normal e histograma de los residuos, así como la prueba Jarque-Bera de normalidad. Comente sobre estos resultados.

*f*) Calcule pronósticos con base en el modelo estimado. Para esto escoja un horizonte de pronósticos adecuado a su serie.

*g*) Reporte las conclusiones del trabajo: un resumen de los resultados que encontró, y de los problemas que se presentaron en la elaboración del modelo. Por ejemplo, un comentario acerca de lo que Usted crea que logró el modelo: capturó la dinámica de la serie?, su tendencia?, los pronósticos parecen realistas y confiables?. Qué otras alternativas podrían haberse propuesto?. Cuál es una de las críticas al modelo de descomposición?.

# CAPÍTULO 4

# Modelado y Pronóstico de Series Estacionales

## 4.1. Introducción

En este capítulo se introducen modelos para la componente estacional $S_t$, que forma parte de la descomposición de una serie $Y_t = T_t + S_t + \varepsilon_t$, con $T_t$ la tendencia y $\varepsilon_t$ la componente aleatoria. Además se exponen procedimientos de estimación de $S_t$ con base en regresión lineal múltiple.

**Definición 4.1.1 (Componente Estacional).** *La componente estacional $S_t$ se define como una función no aleatoria, periódica, de período $s = 2, 3, \ldots$. El conjunto de valores de $S_t$ para $t = 1, \ldots, s$ se denomina "el patrón estacional". El período estacional $s$ es el número mínimo de períodos que tarda el patrón estacional en repertirse.*

Una definición de Estacionalidad está en Hylleberg et al. [1990]:

> "La estacionalidad es el movimiento intra-anual, sistemático, aunque no necesariamente regular, causado por cambios en el clima, el calendario, y los momentos de las decisiones, hechas por los agentes de la economía, directa ó indirectamente, en sus decisiones de producción y consumo. Estas decisiones están influenciadas por las dotaciones, las expectativas y preferencias de los agentes, y las técnicas de producción disponibles en la economía".

La unidad de tiempo $t$ de $Y_t$ es día, semana, mes, trimestre, semestre, año. Definir el período estacional, por ejemplo, como $s = 12$, significa que el patrón estacional consiste de doce meses, para una serie mensual. De manera similar con $s = 4$, serie trimestral. Un período semanal es $s = 52$ y se muestrea 52 veces al año. También se puede definir un modelo con período en días. Por ejemplo, si $s = 7$, entonces el patrón estacional que se repite cada 7 días. También es posible un período de $s = 365$ días. En los dos últimos casos la serie se muestrea diario.

**Ejemplo 4.1.1.** *labelnotten Una ejemplo de series con componente estacional es la serie* ***nottem****, en la librería stats de R. Corresponde a la serie mensual "Average Monthly Temperatures at Nottingham, 1920-1939", que contiene las temperaturas promedio de grados Fahrenheit tomadas en el Castillo de Nottingham (UK), por 20 años, entre 1920 y 1939. Datos del libro de Anderson, O. D. (1976) Time Series Analysis and Forecasting: The Box-Jenkins Approach. La componente estacional de período* $s = 12$ *se observa claramente.* ([1])

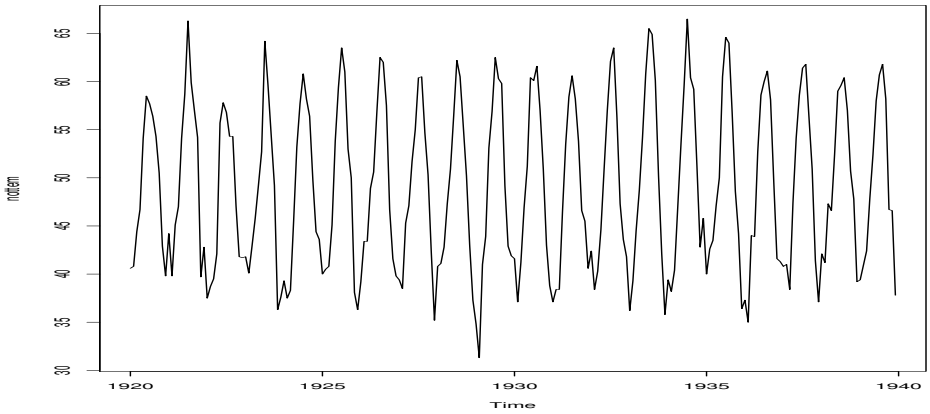



Figura 4.1: Average Monthly Temperatures at Nottingham, 1920-1939

## Propiedades de $S_t$

En todo lo que sigue $S_t,\ t = 1, 2, \ldots$ es una función real en tiempo discreto, es decir, es una sucesión. Que cumple las propiedades siguientes.

1. $S_t$ es una función periódica con periodo $s$, es decir cumple: $S_{t+s} = S_t$ para $t = 1, 2, \ldots$. Por tanto, sólo se requiere definir $S_t$ en los primeros $s$ valores, $S_t,\ t =$

[1] http://www.astrostatistics.psu.edu/datasets/R/html/datasets/html/00Index.html

$1, 2, \ldots, s$. Es decir, basta con definir el patrón estacional.

2. Si $S1_t$ y $S2_t$ son funciones estacionales con periodo $s$ entonces $aS1_t + bS2_t$, para $a, b \in \mathbb{R}$, es también una función estacional de período $s$. Luego, el conjunto de funciones periódicas de período $s$ es un espacio vectorial de dimensión $s$.

3. Una base para este espacio vectorial está conformada por las $s$ variables indicadoras estacionales, $I_j(t) \in \{0, 1\}$, para $j = 1, \ldots, s, t = 1, 2, \ldots$, definidas por

$$I_j(t) = \begin{cases} 1\,, & t = j, j+s, j+2s, \ldots \\ 0\,, & \text{en otro caso} \end{cases} \tag{4.1}$$

Como conclusión, si $S_t$ es periódica de período $s$, existen constantes $\delta_j, j = 1, \ldots, s$ tales que

$$S_t = \sum_{j=1}^{s} \delta_j I_j(t),\ t = 1, 2, \ldots. \tag{4.2}$$

La condición para $I_j(t) = 1$ se puede dar de dos formas

$$I_j(t) = 1 \Leftrightarrow \begin{cases} t = j, j+s, j+2s, \ldots \\ t \equiv j (\text{mod}\, s) \Leftrightarrow t - j = s\lfloor t/s \rfloor. \end{cases} \tag{4.3}$$

donde $\lfloor x \rfloor$ es la parte entera de $x$, p.ej. $\lfloor 2.34 \rfloor = 2$. La segunda forma se lee " t congruente con j, módulo s", significa que $t - j$ es divisible por $s$. Lo cual se expresa como $t - j = s\lfloor t/s \rfloor$, ya que esto quiere decir que el cociente $\frac{t-j}{s}$ es un entero: $\lfloor t/s \rfloor$.

Por ejemplo, si $s = 4, j = 3, t = 11$, entonces $I_3(11) = 1, 0?$. Utilizando la segunda forma, como se verifica que

$$11 = 4\lfloor 11/4 \rfloor + 3 = 4\lfloor 2.75 \rfloor + 3,$$

por tanto, $I_3(11) = 1$. También, $11 - 3 = 8$ es divisible por $4$, luego $I_3(11) = 1$. En cambio, $I_3(10) = 0$, que se verifica de igual manera.

4. También se puede utilizar el conjunto

$$\{sen(2\pi jt/s), cos(2\pi jt/s), j = 1, \ldots, s\}.$$

Para $S_t$ periódica de período $s$, escogemos $m \geq 1$ y constantes $b_{1,j}, b_{2,j}, j = 1, \ldots, m$ tales que,

$$S_t \approx \sum_{j=1}^{m} \beta_{1,j} sin(2\pi jt/s) + \beta_{2,j} cos(2\pi jt/s),$$

para $t \geq 1$. El valor de $m$ se escoge

$$m = \begin{cases} 1 \leq m \leq \frac{s}{2} - 1, & \text{si } s \text{ es par} \\ 1 \leq m \leq \frac{s-1}{2}, & \text{si } s \text{ es impar} \end{cases} \tag{4.4}$$

Valores $m = 1, 2$ son recomendables por parsimonia.

5. Nótese que en el modelo con variables indicadoras el número de parámetros aumenta con el período $s$. Para períodos grandes, por ejemplo $s = 360$ no sería un modelo práctico. Una alternativa más práctica sería el modelo con variables trigonométricas

$$\{sen(2\pi jt/360), cos(2\pi jt/360), j = 1, \ldots, m\}.$$

## 4.2. Estimación de la Componente Estacional

### 4.2.1. Modelos con Variables Indicadoras

Si se define un modelo para la componente de tendencia $T_t$, por ejemplo,

$$T_t = \sum_{j=0}^{k} \beta_j t^j, k = 1, 2, 3,$$

y el modelo para la componente estacional con variables indicadoras (4.2), entonces el modelo de regresión lineal múltiple para $Y_t$ está dado por:

$$Y_t = \beta_0 + \sum_{j=1}^{k} \beta_j t^j + \sum_{j=1}^{s} \delta_j I_j(t) + \varepsilon_t, \tag{4.5}$$

donde $k = 1, 2, 3$, $\beta_j,\ j = 1, \ldots, k$ y $\delta_j,\ j = 1, \ldots, s$ son parámetros a estimar. Es decir, se puede escribir $Y_t$ de la forma $Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{i,t} + \varepsilon_t$, donde $X_{i,t}$ es de la forma $t_i$ ó $I_i(t)$, y se pueden utilizar todos los recursos de regresión lineal múltiple para la estimación y pronóstico con este modelo.

A continuación se enumeran los pasos para estimar el modelo con tendencia y estacionalidad $S_t$ con variables indicadoras, (4.5).

1. Identificar el período de la serie $Y_t$, $s$, si es que existe. Al declarar en R la serie como objeto "ts" con frecuencia, por ejemplo 12, se está identificado el posible período como $s = 12$. Sin embargo, el caso $s = 360$ no se incluye con indicadoras. En este caso sería preferible usar funciones trigonométricas, como se explica más adelante.

2. Generar las $s$ variables indicadoras $I_j(t)$ para $t = 1, 2, \ldots, T$. Es decir, generar la matriz $I_j(t)$, donde al variar $j = 1, \ldots, s$ varían las columnas y al variar $t = 1, \ldots, T$ varían las filas.

3. Estimar los parámetros en el modelo (4.5) mediante regresión lineal. Nótese que se trata de estimar los parámetros de las componentes de tendencia y estacional, conjuntamente.

La función en (4.5)

$$T_t + S_t = \beta_0 + \sum_{j=1}^{k} \beta_j t^j + \sum_{j=1}^{s} \delta_j I_j(t)$$

se denominará "componente estructural" , y los residuales $\varepsilon_t$ en (4.5) se denominarán "residuos estructurales" .

### El Problema de la Multicolinealidad con Variables Indicadoras

Se deben incluír solamente $s - 1$ variable indicadoras en el modelo en lugar del conjunto completo de $s$, para evitar problemas de multicolinealidad, ya que las variables indicadoras cumplen

$$\sum_{j=1}^{s} I_j(t) = 1, \forall t,$$

por lo que estarían perfectamente correlacionadas con el intercepto $\beta_0$, pues, aunque no se hace explícito, esta constante está asociada con una variable explicativa que es constante e igual a 1 (ver Diebold [1999, pág. 90]). En consecuencia, el modelo (4.5) se modifica y queda

$$Y_t = \beta_0 + \sum_{j=1}^{k} \beta_j t^j + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t, \tag{4.6}$$

### Estimación en R del modelo con Variables Indicadoras

En la librería `forecast` hay varios recursos y funciones para estimar y pronosticar con el modelo (4.6). Suponga que $Y_t$ tiene una componente estacional de período $s$, y que se declara como objeto `ts` en R, con frecuencia $s$.

El primer paso consiste en generar las variables indicadoras $I_j(t), j = 1, \ldots, s, t = 1, \ldots, T$. La función `seasonaldummy` de la librería `forecast` calcula una matriz de ceros y unos, denotada por ejemplo, $It$, de dimensión $T \times (s - 1)$, con $s - 1$ columnas, a partir de la

información sobre la serie $Y_t$ de frecuencia $s$. Si el dato de $Y_t$ corresponde a la estación $j$ entonces la primer fila de la matriz $It$ será un vector de ceros excepto en la posición $j$, que corresponde a un 1. De acuerdo a este valor inicial se genera la matriz $It$. Esta información se genera con el comando `ts()` al incluír la opción (por ejemplo) `start=c(1995,2)`, que indica que la serie mensual inicia en febrero de 1995. Ejemplo:
`y = ts(y,frequency=12, start=c(1995,2))`

Adicionalmente, el comando `seasonaldummyf` genera otra matriz de ceros y unos, denotada, por ejemplo, $Itp$, con $s-1$ columnas y con un número de filas igual al número de pronósticos, a partir de un tiempo específico. El patrón de las estaciones en esta matriz está de acuerdo a la información de la serie, como se indicó.

Antes de un ejemplo con `seasonaldummy`, como ejercicio, correr el siguiente código en R que calcula la matriz $I_j(t)$ de dimensión $T \times s$.

**Código R 4.2.1.**

```
# funcion para calcular la matriz It
#---------------------------
dummy=function(s,n,i){
# s: estaciones, n: numero de datos, i: estación de inicio
A=diag(s)
for(j in 1:(floor(n/s))){
A=rbind(A,diag(s))}
A=A[i:(n+i-1),]
return(A)}
```

**Ejemplo 4.2.1.** *Considere la modelación de la serie de producción de cemento Portland, trimestral, en toneladas por mil, entre Q1 1956 y Q3 1994, en Australia. El modelo de componentes que se asume para esta serie es de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{4-1} \delta_j I_t(t) + \varepsilon_t. \tag{4.7}$$

*Es decir, el modelo a estimar es lineal para la tendencia, más una componente estacional con base en variables indicadoras.*

**Código R 4.2.2.**

```
# -----Ejemplo con producción de cemento
library(forecast)
```

```
library(xtable)
E = read.table("cementq.dat", header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
#----modelo con tendencia lineal y estacionalidad
#----con variables indicadoras estacionales
t = seq(1,length(y))
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)
print(xtable(mod1),digits=6)
r1 = mod1$residuals
yhat1 = mod1$fitted.values
#----------------------------
```

*Los resultados del modelo lineal con variables estacionales (4.7) está en el Cuadro 4.1 siguiente. Una comparación entre la serie ajustada y la estimada se puede ver en la Figura (4.2). Con un R-cuadrado ajustado* $\overline{R}^2 = 0.8966$*, el ajuste resulta aceptable. También se observa que los coeficientes de las variables indicadoras ItQ2, ItQ3 no son significativos, sin embargo, ItQ1 sí, y por tanto, concluímos existe una componente estacional en la serie. Sin embargo, la componente estacional parece no ser constante (no es un modelo global).*

Tabla 4.1: Resultados del Ajuste del Modelo con Variables Estacionales

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 665.2149 | 24.1428 | 27.55 | 0.0000 |
| t | 7.2364 | 0.2023 | 35.78 | 0.0000 |
| ItQ1 | -163.0881 | 25.6796 | -6.35 | 0.0000 |
| ItQ2 | -33.5553 | 25.6788 | -1.31 | 0.1933 |
| ItQ3 | 12.0287 | 25.6796 | 0.47 | 0.6402 |

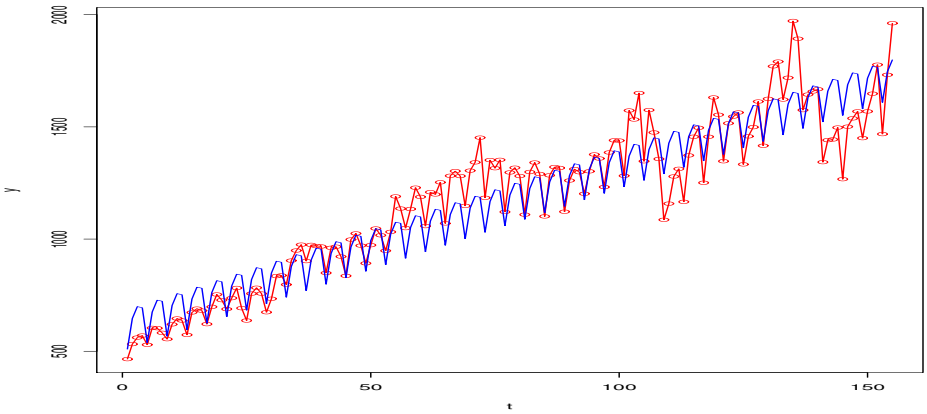



Figura 4.2: Ajuste con Variables Indicadoras

**Ejercicio 4.2.1.** *Cambiar el modelo (4.7) por uno con tendencia cuadrática. Cómo se compara con el modelo (4.7) utilizando las medidas de ajuste (AIC, BIC, R2 ajustado, MSE), y las pruebas F parciales?.*

### 4.2.2. Modelos con Funciones Trigonométicas

En este caso el modelo a estimar es:

$$Y_t = \beta_0 + \sum_{j=1}^{m} \beta_j t^j + \sum_{j=1}^{k} \beta_{1,j} sen(2\pi jt/s) + \beta_{2,j} cos(2\pi jt/s) + \varepsilon_t, \tag{4.8}$$

donde el límite $k$ debe estar acotado por: $1 \leq k \leq \frac{s}{2} - 1$, si $s$ es par, o por $1 \leq k \leq \frac{s-1}{2}$ si $s$ es impar. El procedimiento de estimación empieza con la generación de las variables explicativas $sen(2\pi it/s)$, $cos(2\pi it/s)$. Para calcular estas variables se utiliza la función `It.trig = fourier(y,k)`, la cual calcula los armónicos $sen(2\pi jt/s), cos(2\pi jt/s)$, para cada $j = 1, 2, \ldots, k$.

**Ejemplo 4.2.2.** *Continuando con el Ejemplo (4.2.1), ahora se estima el modelo (4.7), pero utilizando funciones trigonométricas. Nótese que al ser $s = 4$ para, entonces el número de armónicos es como máximo $s/2 - 1 = 1$.*

$$Y_t = \beta_0 + \beta_1 t + \beta_1 sin(2\pi t/4) + \beta_2 cos(2\pi t/4) + \varepsilon_t. \tag{4.9}$$

*Se incluye el cálculo de pronósticos a 8 trimestres con los modelos de variables indicadoras y de funciones trigonométricas. Los resultados están en el Cuadro 4.2.*

**Código R 4.2.3.**

```
#----modelo con tendencia lineal y estacionalidad
#----con variables trigonometricas

It.trig = fourier(y,1)
mod2 = lm(y ~ t + It.trig)
summary(mod2)
print(xtable(mod2),digits=6)
r2 = mod2$residuals
yhat2 = mod2$fitted.values
# pronosticos 8 trimestres
# con el modelo con variables indicadoras
T = length(yi)
Itp = seasonaldummyf(y,8)
tp = seq(T+1,T+8,1)
prons.vi = predict(mod1,
data.frame(t = tp, It=I(Itp)))
# con el modelo con funciones trigonometricas
Itp.f = fourierf(y,1,8)
prons.vt = predict(mod2,
data.frame(t = tp, It.trig=I(Itp.f)))
print(xtable(rbind(prons.vi,prons.vt)))
#-------------------------------------
```

Tabla 4.2: Resultados de ocho pronósticos con Holt-Winters, Indicadoras y Trigonométricas

| Trimestre | Obs | H-W | Indicadora | Trigonom |
|---|---|---|---|---|
| 1992 Q4 | 1569.00 | 1572.62 | 1752.06 | 1750.25 |
| 1993 Q1 | 1450.00 | 1320.77 | 1599.55 | 1598.36 |
| 1993 Q2 | 1569.00 | 1488.50 | 1733.36 | 1738.47 |
| 1993 Q3 | 1648.00 | 1570.55 | 1780.69 | 1771.06 |
| 1993 Q4 | 1777.00 | 1602.48 | 1781.63 | 1793.91 |
| 1994 Q1 | 1468.00 | 1350.63 | 1629.12 | 1632.41 |
| 1994 Q2 | 1732.00 | 1518.35 | 1762.93 | 1763.56 |
| 1994 Q3 | 1962.00 | 1600.40 | 1810.26 | 1805.11 |

El método Holt-Winters se introducirá en un Capítulo posterior. La conclusión que puede obtenerse es que en este caso es indiferente utilizar variables indicadoras ó funciones trigonométricas, dado que los pronósticos son muy similares. A esta conclusión se puede llegar también observando que los valores de la raíz del error cuadrático medio, $111.1831$, $111.1278$, para el modelo de variables indicadoras y para el de funciones trigonométricas, respectivamente, son prácticamente idénticos.

**Ejemplo 4.2.3.** *Un modelo para la demanda diaria de energía del mercado Nord Pool (países escandinavos) es el siguiente. La serie de precios es diaria para un período de 30/12/1996-26/04/2000.*

*Esta serie diaria presenta dos componentes estacionales: una con período semanal y otra con período anual. Además se asume una tendencia lineal. Denotando la componente estacional semanal por $s_t$ y la componente estacional anual por $S_t$, se escribe el modelo como sigue.*

$$\begin{aligned} Y_t &= a + bt + s_t + S_t + e_t, \\ s_t &= \sum_{j=1}^{6} \delta_j I_j(t), \\ S_t &= \alpha\, cos\left(\frac{2\pi t}{365}\right) + \beta\, \mathrm{sen}\left(\frac{2\pi t}{365}\right). \end{aligned} \tag{4.10}$$

*En este modelo la componente con período anual se modela con funciones trigonométricas y la componente con período semanal con variables indicadoras. El siguiente código en R implementa el modelo (4.10).*

**Código R 4.2.4.**

```
# Código en R --------------------
library(forecast)
D = read.table("serie.dat",header=T)
attach(D)
y = ts(y, frequency = 7)
n = length(y)
t = 1:n
St = seasonaldummy(y)
x1 = cos(2*pi*t/365)
x2 = sin(2*pi*t/365)
modelo1 = ln(y ~ t + St + x1 + x2)
summary(modelo1)
```

# CAPÍTULO 5

# Modelos con Base en Suavizadores

Este capítulo introduce varios modelos alternos a los de descomposición para la tendencia y la estacionalidad. Son modelos no paramétricos, es decir, las estimaciones se basan en algoritmos aplicados a los datos, sin asumir un modelo paramétrico específico. Una característica importante es que son mucho más flexibles para modelar la tendencia y la estacionalidad. En algunos casos las estimaciones se basan en una ventana (intervalo) de datos móvil. Por tanto, cabe denominarlos estimadores locales de tendencia y estacionalidad.

## 5.1. Regresión Local Loess

La Regresión Loess (denominada en R inicialmente Lowess: Locally weighted scatterplot smoothing) es un modelo de regresión no paramétrica que regresa $y_i$ versus $x_i$, pero no asume un modelo global fijo, es decir, no asume un intercepto y una pendiente fijas sino variables, de manera local. Aquí local significa una ventana móvil que contiene un número determinado de datos de la variable independiente. La motivación para utilizar Loess es que se puede estimar la tendencia de la serie $T_t$ sin utilizar un modelo global paramétrico, que en muchos casos resulta muy restrictivo.

La descripción del algoritmo de estimación Loess es el siguiente. Suponga datos bivariados: $(x_i, y_i),\ i = 1, \ldots, n$. El objetivo de Loess es calcular una función de regresión local $g(x)$

de forma que,

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{5.1}$$

que es el análogo de $a + bx_i$ en $y_i = a + bx_i + \varepsilon_i$. Suponga $x_{(1)} \leq x \leq x_{(n)}$, entonces $g(x)$ se calcula así:

1. Se escoge $1 \leq q \leq n$.

2. Se escogen los $q$ valores $x_i$ más cercanos a $x$. Luego $\frac{|x_i - x|}{\lambda_q(x)} \leq 1$.

3. Defina $\lambda_q(x)$ la distancia máxima entre $x$ y los $q$ $x_i$ escogidos.

4. Defina $w(x) = \begin{cases} (1 - x^3)^3 & , \ 0 \leq x \leq 1 \\ 0 & , \ x > 1 \end{cases}$

5. Defina $w_i(x) = w\left(\frac{|x_i - x|}{\lambda_q(x)}\right)$, para cada $x_i$ escogido.

6. Ajuste $y_i = a + bx_i$ ó $y_i = a + bx_i + cx_i^2$ con MCO, tomando solamente los $q$ valores más cercanos a $x$, ponderando cada $x_i$ con $w_i(x)$.

7. Defina $g(x)$ como el valor estimado en $x$ mediante la regresión local anterior.

En series de tiempo se toma, $(t_i, y_i)$, con $x_i = t_i$. En cada $t$ se estima la tendencia $T_t$ con $\widehat{T}_t = \hat{g}(t)$. En R se puede implementar la regresión Loess mediante la función `loess()`.

Una característica importante y útil de la función `predict()` es que admite objetos generados con `loess()` para calcular pronósticos. En el Ejemplo siguiente se muestra cómo implementar estos procedimientos.

**Ejemplo 5.1.1.** *Se consideran los datos del DANE sobre estadísticas de edificación, correspondientes a número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06.*

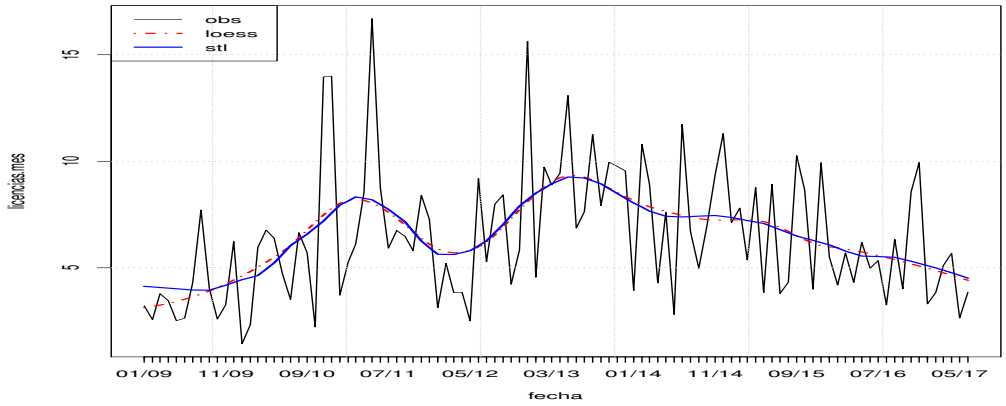



Figura 5.1: Suavizamiento Loess de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

Este es el código en R para suavizamiento con Loess, ver Figura 5.1.

**Código R 5.1.1.**

```
# Analisis de número de unidades aprobadas para vivienda
# 88 municipios 2009/01 - 2017/06
# Puede asumirse una tendencia global?
library(Kendall)
# leer

G = read.table("series_histo_88_mpio_jun17.prn",
header = TRUE, stringsAsFactors = FALSE)
attach(G)

# definir como serie de tiempo
y = ts(G$Totalviv,frequency=12,start=c(1986,1))

# detectar tendencia monotona con la prueba Mann-Kendall
# Ho: no hay tendencia, Ha: tendencia monotona
MannKendall(y)

# estimacion loess
T = length(y)
```

```
t = seq(1,T)
yw = loess(y ~ t, control = loess.control(surface = "direct"))
yest.yw = yw$fitted

# grafica con fechas en el eje x
np = length(y)

fecha = seq(as.Date("2009/01/01"), as.Date("2017/06/01"), by="months")

ejex.mes = seq(fecha[1],fecha[np], "months")
ejex.año = seq(fecha[1],fecha[np],"years")

plot(fecha,y, xaxt="n", panel.first = grid(),type='l',
ylab='licencias.mes', lwd = 1)
axis.Date(1, at=ejex.mes, format="%m/%y")
axis.Date(1, at=ejex.año, labels = FALSE, tcl = -0.2)

lines(fecha,yest.yw, xaxt="n", panel.first = grid(),
type='l',col='red',lty=2,lwd=2)

legend("topleft", c("obs","loess"), lty = c(1,4),
lwd=c(1,2),col=c('black','red') )

# pronosticos a un mes
tt = seq((T-12+1),T,1)
pr.y = predict(yw, data.frame(t = tt))
```

### Prueba Mann-Kendall para detectar tendencia monótona

Si se tiene una serie de tiempo $\{Y_t, t = 1, 2, \ldots, T\}$ se define la prueba de hipótesis de Mann-Kendall con las hipótesis nula $H_0$ : los datos $Y_t$ son iid, versus la alterna $H_1$ : los datos tienen una tendencia monótona. Bajo $H_0$ el estadístico Mann-Kendall se define como

$$S = \sum_{k=1}^{T-1} \sum_{j=k+1}^{T} signo(Y_j - Y_k). \tag{5.2}$$

donde $signo(x) = \pm 1$ si $x > 0, x < 0$, y $signo(0) = 0$. $S$ se distribuye aproximadamente Normal, con media cero. Un valor positivo de $S$ indica una tendencia creciente, y uno

negativo una tendencia decreciente. La prueba está implementada en la librería `kendall`.

## 5.2. Método híbrido Loess y componente estacional

En el modelo de componentes con variables indicadoras (4.5), pág 55,

$$Y_t = \beta_0 + \sum_{j=1}^{k} \beta_j t^j + \sum_{j=1}^{s} \delta_j I_j(t) + \varepsilon_t,$$

la tendencia y la estacionalidad son modelos paramétricos globales, lo que quiere decir que sus parámetros no cambian con $t$. Para introducir una mayor flexibilidad se propone estimar la tendencia mediante Loess, y conservar la parte paramétrica de la componente estacional. Se tendría un modelo semi-paramétrico ("híbrido") de la forma

$$Y_t = g(t) + \sum_{j=1}^{s} \delta_j I_j(t) + \varepsilon_t, \tag{5.3}$$

para el cual se podría estimar el conjunto total de coeficientes $\delta_j$, porque no habría, estrictamente, un problema de multicolinealidad.

El procedimiento de estimación del modelo "híbrido" (5.3) se puede realizar en dos pasos. Estimar por Loess la tendencia, obteniendo $\widehat{T}_t = \widehat{g}(t)$, y luego estimar la componente estacional y el error estructural mediante regresión lineal con respecto a la variable $Y_t - \widehat{T}_t$.

## 5.3. Método de descomposición STL

La descomposición STL es un procedimiento para estimar las componentes del modelo $Y_t = T_t + S_t + \varepsilon_t$, desarrollado por Cleveland et al. [1990], con base en Regresión Loess. Consiste de una secuencia de dos aplicaciones iteradas de Regresión Loess. Para aplicar este método se debe especificar el período (frequency=s) de la componente estacional, mediante la función `ts(x,frequency=s)`, donde (por ejemplo) $s = 4, 12, 52, 360$.

**Ejemplo 5.3.1.** *Al aplicar el método stl a los datos del Ejemplo (5.1.1) se obtiene la descomposición en la Figura (5.2). El tercer panel es la tendencia estimada por Loess. El segundo la componente estacional. El tercero es la componente del error. Hay que observar que la componente estacional tiene un rango estimado* $[-2, 6] + 7 = [5, 13]$ *y la serie un rango* $[5, 10]$. *Una conclusión es que la componente estacional podría ser significativa porque su rango es similar al de la serie. Un modelo paramétrico posiblemente la detectaría.*

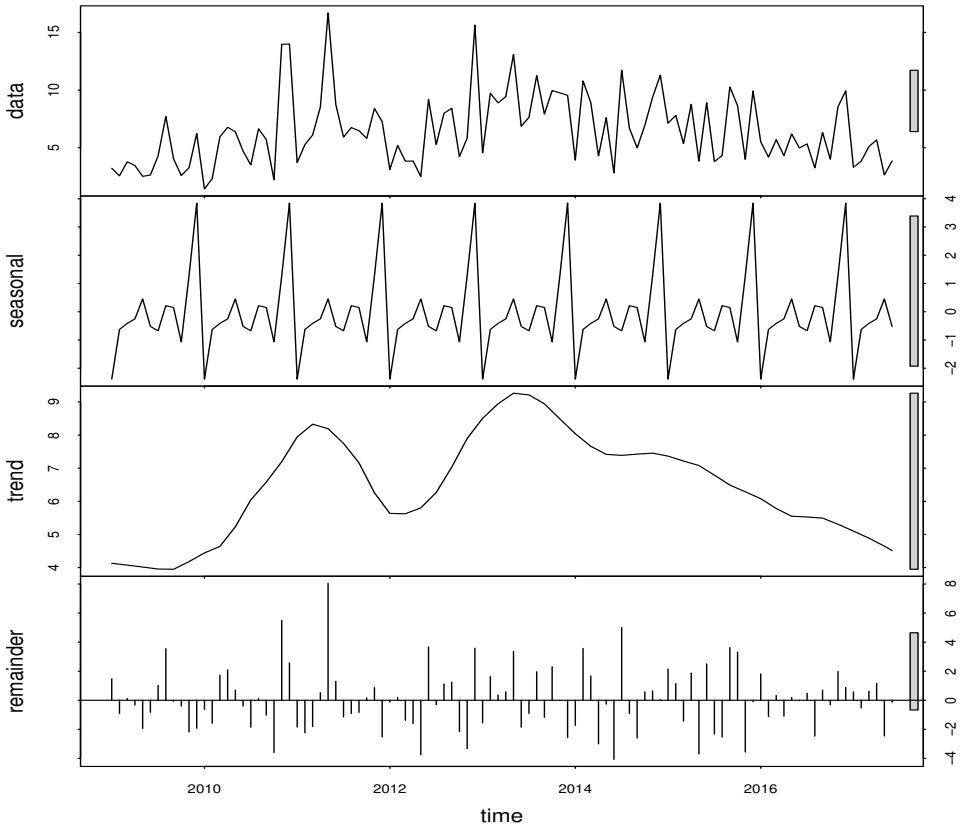



Figura 5.2: Descomposición STL número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

**Código R 5.3.1.**

```
#---------implementar descomposicion con stl
m = stl(y, s.window = 'per', t.window = 50)
# t.window = 50 : ventana de 50 observaciones para la
# tendencia
St = m1$time.series[,1]
Tt = m1$time.series[,2]
et = m1$time.series[,3]
plot(m)
```

Otras librerías en R también tienen funciones que realizan la descomposición de $T_t, S_t, \varepsilon_t$,

son, por ejemplo,

1. La función `decompose()`.
2. La librería `mFilter`, que implementa el filtro Hodrick-Prescott, el cual estima $T_t$ y $S_t$, mediante la función `hpfilter`. ([1])
3. La librería `timsac` tiene la función `decomp()` que realiza la descomposición incluyendo una componente autoregresiva y otra para fechas de intervenciones, $Y_t = T_t + S_t + R_t + TA_t + \varepsilon_t$.
4. La función `ma.filter()` de la librería `rmaf` (Refined Moving Average Filter, Qiu [2015]).

## 5.4. Modelo de Holt-Winters

En el modelo de componentes con variables indicadoras (4.5), pág 55,

$$Y_t = \beta_0 + \sum_{j=1}^{k} \beta_j t^j + \sum_{j=1}^{s} \delta_j I_j(t) + \varepsilon_t,$$

la tendencia y la estacionalidad son modelos paramétricos globales, lo que quiere decir que sus parámetros no cambian con $t$. Como en el caso del modelo híbrido con Loess, para introducir una mayor flexibilidad se pueden permitir parámetros variables con $t$. Por ejemplo, tomando $k = 1$ en el modelo anterior, y haciendo depender cada parámetro del tiempo $t$, se tendría un modelo que se denomina de parámetros variables:

$$Y_t = \beta_{0,t} + \beta_{1,t} t + \sum_{j=1}^{s} \delta_{j,t} I_j(t) + \varepsilon_t, \tag{5.4}$$

En este modelo los parámetros se asume que varían de acuerdo a marchas aleatorias, definidas como

$$\begin{aligned} \beta_{0,t} &= \beta_{0,t-1} + \varepsilon_{\beta_0,t} \\ \beta_{1,t} &= \beta_{1,t-1} + \varepsilon_{\beta_1,t} \\ \delta_{j,t} &= \delta_{j,t-1} + \varepsilon_{\delta_j,t}, \end{aligned}$$

donde $\varepsilon_{\beta_0,t} \sim iidN(0, \sigma^2_{\beta_0})$, $\varepsilon_{\beta_1,t} \sim iidN(0, \sigma^2_{\beta_1})$, $\varepsilon_{\delta_j,t} \sim iidN(0, \sigma^2_{\delta_j})$, son variables aleatorias distribuídas iid Normales. Se dice que los parámetros en (5.4) "varían lentamente".

[1] http://www.bondeconomics.com/2015/04/why-you-should-never-use-hodrick.html

El modelo de Holt-Winters es un algoritmo para estimar recursivamente las cantidades

$$\begin{aligned} \mu_t &= \beta_{0,t}, \\ \beta_t &= \beta_{1,t}t, \\ S_t &= \sum_{j=1}^{s} \delta_{j,t} I_j(t), \end{aligned}$$

a partir de unos valores iniciales $\mu_0, \beta_0, S_{0,j} = \delta_{j,0}, j = 1, 2, \ldots, s$. Fué introducido por Winters [1960].

**Definición 5.4.1.** *El algoritmo Holt-Winters consiste en el sistema de ecuaciones recursivas para $\mu_t$, $\beta_t$ y $S_t$, dado por*

$$\begin{aligned} \mu_t &= \alpha(Y_t - S_{t-s}) + (1-\alpha)(\mu_{t-1} + \beta_{t-1}), \\ \beta_t &= \beta(\mu_t - \mu_{t-1}) + (1-\beta)\beta_{t-1}, \\ S_t &= \gamma(Y_t - \mu_t) + (1-\gamma)S_{t-s}, \end{aligned}$$

*donde $t = 2, \ldots, T$.*

Las constantes $\alpha, \beta, \gamma \in (0, 1)$ son los parámetros de suavizamiento (tunnig parameters). Se escogen de manera que se minimize el MSE. O también, por tanteo. Según Abraham and Ledolter [2000, pag. 169],

> "Si las componentes de tendencia y estacionalidad cambian rápidamente, las constantes $\alpha_j$ deben tomarse cercanas a 1.0. Si las componentes son estables y varían lentamente, se escogen cercanas a 0.0".

Los valores iniciales se toman como los parámetros estimados de las regresiones

$$Y_t = \sum_{j=1}^{s} \delta_j I_j(t) + \varepsilon_t, \tag{5.5}$$

$$\widehat{\varepsilon_t} = \beta_0 + \beta_1 t + \eta_t, \ t = 1, 2, \ldots, T. \tag{5.6}$$

es decir, $S_{0,j} = \widehat{\delta}_j$, $\beta_0 = \widehat{\beta}_1$, $\mu_0 = \widehat{\beta}_0 + \widehat{\beta}_1$.

El valor suavizado de la serie en $t = 1, 2, \ldots, T$ es

$$\widehat{Y_t} = \mu_t + \beta_t + S_t.$$

El pronóstico para un horizonte $j = 1, 2, \ldots, m$

$$\widehat{Y}_{T+j} = (\mu_T + j\beta_T) + S_{T-s+j}$$

En: Harvey [2006], pag. 358

> "El procedimiento Holt-Winters está considerado como una aproximación a un filtro de Kalman aplicado al modelo de componentes; el por qué algunos continúan utilizándolo es algo misterioso. Discusiones adicionales sobre diferentes formas de las recursiones aditiva y multiplicativas Holt-Winters se pueden encontrar en Ord, Koehler and Snyder (1997)".

La programación en R del método Holt-Winters se puede hacer de manera que se estiman las constantes $\alpha, \beta, \gamma$ automáticamente, como los valores que minimizan el MSE. O también, se puede ingresar un valor particular para cada constante.

```
# ---------ejemplo programación de Holt-Winters
# y es un objeto 'ts', con período s
m = HoltWinters(y)
(c(m$alpha,m$beta,m$gamma))
# la tendencia, componente estacional y y estimada
Tt = m$fitted[,2] + m$fitted[,3]
St = m$fitted[,4]
Yt.hat = m$fitted[,1]
# y 12 pronosticos se calculan como
ypron = predict(m,12, prediction.interval = TRUE)
plot(m, p)
```

**Ejemplo 5.4.1.** *Continuando con el Ejemplo de la serie de producción trimestral de cemento, se incluye 8 pronósticos con el método de Holt-Winters. Los pronósticos se mostraron en el Cuadro 4.2 anterior.*

**Ejemplo 5.4.2.** *Continuando con el Ejemplo 5.1.1, de los datos del DANE sobre estadísticas de edificación, correspondientes a número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06. Se puede observar una comparación entre Loess y Holt-Winters en la Figura (5.3)*

A

## 5.5. El Modelo de Espacio de Estados

El Modelo de Espacio de Estados en Hyndman et al. [2008], pag. 232, es similar la modelo Holt-Winter, que incorpora la hipótesis de variación lenta de acuerdo a marchas aleatorias.

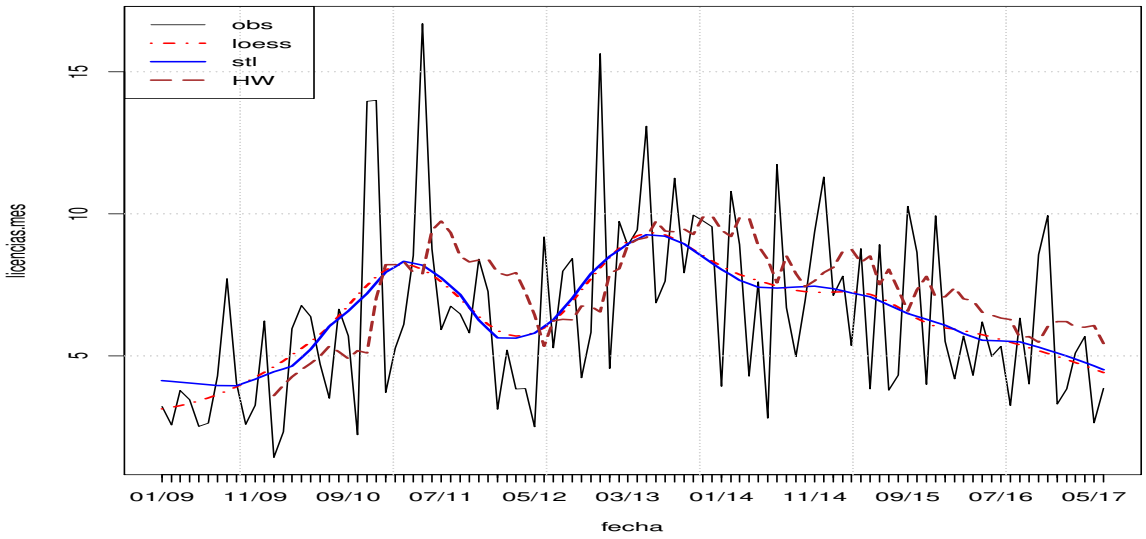



Figura 5.3: Estimación de la tendencia con los Métodos de Loess y Holt-Winters

Los parámetros $(\alpha, \delta, \gamma)$ controlan la varianza de un solo error aleatorio $\epsilon_t$, que afecta cada ecuación del sistema. Concretamente,

$$y_t = \mu_t + \beta_t + S_{t-s} + \epsilon_t, \tag{5.7}$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \alpha\epsilon_t, \tag{5.8}$$
$$\beta_t = \beta_{t-1} + \delta\epsilon_t, \tag{5.9}$$
$$S_t = S_{t-s} + \gamma\epsilon_t, \tag{5.10}$$

con $\epsilon_t \sim iidN(0, \sigma^2)$.

Este modelo se define equivalente a Holt-Winters, y se denota `ETS(A,A,A)`. La estimación de los parámetros se realiza mediante la función `ets()` de la librería `forecast`.

```
m1 = ets(y,model="AAA",damped=FALSE)
yhat1 = m1$fitted
#---- h pronosticos
yf.ets = forecast(m1,h)$mean
```

Ver también la función `StructTS(y,type="BSM")`, que estima un modelo similar a `ETS(A,A,A)`.

## 5.6. Pruebas de Estabilidad Estructural en Modelos de Componentes

La idea básica de este tipo de pruebas es esta observación en Diebold [1999, pág. 92]

> "Con frecuencia, las relaciones comerciales y económicas varían con el tiempo. Los procedimientos recursivos de estimación nos permiten evaluar y rastrear los parámetros variables en el tiempo, y, en consecuencia, son útiles para elaborar y evaluar diversos modelos de pronósticos."

Suponga el modelo de regresión lineal

$$Y_t = \sum_{i=1}^{k} \beta_i X_{i,t} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{5.11}$$

donde $\varepsilon_t \sim iid\, N(0, \sigma^2)$. Si lo parámetros $\beta_i$ no fueran constantes sino que cambiaran con el tiempo $t$, de manera aleatoria o no, se colocaría

$$Y_t = \sum_{i=1}^{k} \beta_{i,t} X_{i,t} + \varepsilon_t, \tag{5.12}$$

un modelo con coeficientes variables.

Note que (5.11) también se puede escribir como $Y_t = \underline{X}_t'\underline{\beta}+\varepsilon_t$ con $\underline{X}_t' = (X_{t,1}, X_{t,2}, \ldots, X_{t,k})$ vector fila en $\mathbb{R}^k$ y $\underline{\beta} \in \mathbb{R}^k$, vector de parámetros. En el caso que los parámetros varíen en el tiempo el modelo quedaría

$$Y_t = \underline{X}_t'\underline{\beta}_t + \varepsilon_t.$$

**Definición 5.6.1 (Estabilidad Estructural).** *Si se asume el modelo de coeficientes variables (5.12), la estabilidad estructural del modelo se define como la hipótesis nula de que los coeficientes se mantengan constantes en el tiempo $\beta_{i,t} \equiv \beta_i$. Lo contrario es que algún $\beta_{i,t}$ cambie con t, que es la hipótesis alterna.*

$$\begin{aligned} H_0 &: \forall i = 1, \ldots, k, \beta_{i,t} \equiv \beta_i \\ H_1 &: \exists i, \beta_{i,t} \not\equiv \beta_i. \end{aligned} \tag{5.13}$$

**Ejemplo 5.6.1.** *Algunos ejemplos de cambio estructural en la literatura, debidos a intervenciones en fechas específicas, son*

- *EL nivel del rio Nilo, serie anual, con un cambio en 1898 con la primera represa de Ashwan.*

- *El número de muertes en accidentes automovilísticos mensuales en Gran Bretaña. La introducción del uso obligatorio del cinturón de seguridad en Enero 31 de 1983 cambió la serie mediante una intervención que disminuyó el nivel. En Harvey and Durbin [1986] se hace un análisis completo utilizando modelos estructurales.*
- *En Box and Tiao [1975, pag. 70] se exponen dos intervenciones en la ciudad de Los Angeles: la apertura del la autopista Golden State Freeway y la entrada en vigencia de la ley 63 que reducía la proporción de hidrocarburos en la gasolina. Ambas intervenciones se esperaba que produjeran un efecto en los niveles de contaminación del aire en los inicios de la década de 1960, medidos en la serie de ppm horaria de $O_3$.*
- *El efecto de los atentados de 11 de Septiembre de 2001 en New York y Washington, en el número de turistas que ingresaron a varios países (ver Haywood and Randal [2006])*
- *El efecto de medidas de restricción de porte de armas de fuego en el número de homicidios por tal causa en varias ciudades.*

Dos técnicas para chequear el supuesto de estabilidad estructural son

1. La estimación recursiva de los parámetros del modelo.
2. Las pruebas CUSUM, CUSUMSQ, MOSUM.

El objetivo al aplicar estas técnicas es asegurar que el modelo para pronósticos es estable. En caso de inestabilidad estructural, los pronósticos con un modelo de regresión global podrían no ser confiables. En este caso sería recomendable utilizar un procedimiento adaptativo (recursivo) como Holt-Winters ó el modelo de espacio de estados en (5.7), pag.70. También un modelo de regresión no paramétrica como Loess.

### 5.6.1. Estimación Recursiva de los Parámetros

**Definición 5.6.2 (Estimadores Recursivos).** *Asumiendo el modelo de coeficientes variables (5.12)*

$$Y_t = \sum_{i=1}^{k} \beta_{i,t} X_{i,t} + \varepsilon_t,$$

*en lugar de utilizar todos los datos para la estimación, se utilizan los $k$ primeros datos y se estiman los parámetros $\hat{\beta}_{i,k}$. Luego se utilizan los primeros $k+1$, y se estiman los $\hat{\beta}_{i,k+1}$,*

*y así sucesivamente, hasta utilizar los $T$ datos. Los parámetros obtenidos $\hat{\beta}_{i,t}$ $t = k, \ldots, T$ se denominan Estimadores Recursivos.*

Para implementar en R la estimación recursiva se utilizan las siguientes instrucciones. Suponga que se trata de un modelo con tendencias lineal y estacional, dado por $Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_j(t) + \epsilon_t, t = 1, \ldots, T$. El total de parámetros es $k = 2 + s - 1$. Se toma un número inicial de datos $k$ un poco mayor de $2 + s - 1$, por ejemplo, $k = 2 + s - 1 + 10$, para tratar de evitar fluctuaciones en los estimadores por poca información, y se corre el siguiente programa.

**Código R 5.6.1.**

```
#--------programación de los estimadores recursivos
k = 2+s-1+10
n = T-k
parm = mat.or.vec(n,(2+s-1))
for(j in 1:n){
yj = y[1:(k+j)]
tj = t[1:(k+j)]
Itj = It[1:(k+j),]
mod.j = lm(yj ~ tj + Itj)
parm[j,] = t(mod.j$coefficient)
}
colnames(parm)=c("beta.0","beta.1",
"delta.1","delta.2","delta.3")
plot.ts(parm)
```

**Ejemplo 5.6.2.** *Con base en el Ejemplo 4.2.1, pag. 56, de la serie de producción trimestral de cemento, la gráfica de los estimadores recursivos de los 5 parámetros para el modelo con tendencia lineal y variables indicadoras estacionales (4.7), pag. 56, están en la Figura 5.4.*

*Lo que puede concluírse de esta figura es que se aprecia un cambio en las trayectorias de los estimadores, por ejemplo en $\beta_0$, en $\beta_1$ y en $\delta_3$. Por tanto, se puede concluír de este diagnóstico que existe evidencia de inestabilidad estructural en el modelo.*

### 5.6.2. Pruebas CUSUM

Las pruebas CUSUM y CUSUM cuadrado fueron introducidas por Brown et al. [1975] para chequear la hipótesis nula de estabilidad estructural de un modelo de regresión lineal,

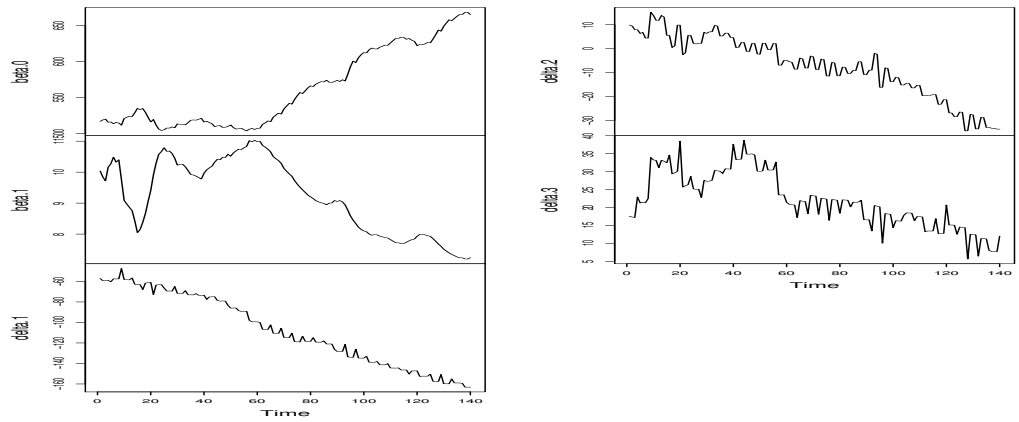



Figura 5.4: Estimadores Recursivos para el Modelo de Producción de Cemento

dada en (5.13). Son herramientas estándar en el análisis de series económicas. El estadístico $CUSUM_t$ de la prueba se define como sigue.

Para cada $t = k, k+1, \ldots, T-1$ se calcula el valor estimado,

$$\widehat{Y}_t = \sum_{j=1}^{k} \hat{\beta}_{j,t} X_{j,t},$$

donde $\hat{\beta}_{j,t}$ son los estimadores recursivos, y se forma el Residual Recursivo

$$\hat{\varepsilon}_{t+1,t} = Y_{t+1} - \hat{Y}_t. \tag{5.14}$$

Si $\sigma^2$ es la varianza de error $\varepsilon_t$, se puede probar que

$$\hat{\varepsilon}_{t+1,t} \sim N(0, \sigma^2 r_t), \tag{5.15}$$

donde $r_t = 1 + \underline{X}_{t+1}\big(\sum_{j=1}^{t} \underline{X}'_j \underline{X}_j\big)^{-1} \underline{X}'_{t+1}$, con $\underline{X}_t$ vectores fila.

**Definición 5.6.3 (Residuales Recursivos.).** *Los residuales recursivos estandarizados se definen como*

$$W_{t+1,t} = \frac{\hat{\varepsilon}_{t+1,t}}{\hat{\sigma}\sqrt{r_t}}, \tag{5.16}$$

*donde* $\hat{\sigma}^2 = MSE$ *y se tiene que* $W_{t+1,t} \sim i.i.d.\ N(0,1)$.

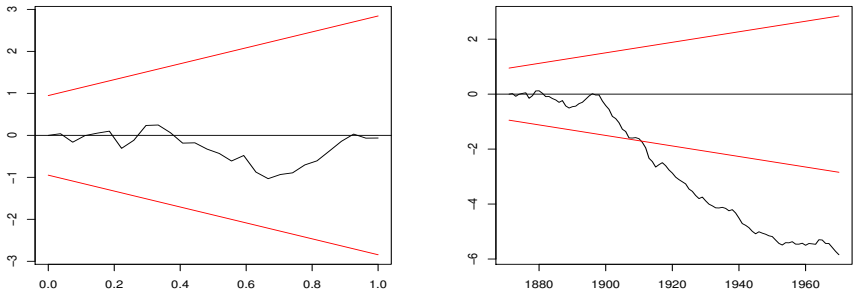
**Definición 5.6.4 (CUSUM).** *La suma acumulada de los residuos recursivos estandarizados ó CUSUM, se define como*

$$CUSUM_t = \sum_{i=1}^{t} \frac{W_{i+1,i}}{\sigma_W}, \quad t = k, k+1, \ldots, T-1, \tag{5.17}$$

*donde*

$$\sigma_W = \sqrt{\frac{\sum_{i=k}^{T-1}\left(W_{i+1,i} - \overline{W}\right)^2}{T-k}}, \qquad \overline{W} = \frac{1}{T-k}\sum_{i=k}^{T-1} W_{i+1,i}.$$

La región de aceptación de la prueba CUSUM para la hipótesis (5.13) es el interior de una región (ó banda) limitada por 2 líneas rectas. Si la gráfica de los residuales CUSUM permanece dentro de esta banda entonces no se rechaza la hipótesis nula de estabilidad estructural. La expresión para las rectas es




(a) No se rechaza $H_0$ (b) Se rechaza $H_0$

Figura 5.5: Región de Rechazo de la Prueba $CUSUM$.

$$\pm\, a\left(\sqrt{T-k} + 2\frac{t-k}{\sqrt{T-k}}\right) \tag{5.18}$$

con $a = 0.948$ para un nivel $\alpha = 0.05$. En la Figura 5.5 se muestran los dos casos posibles.

**Definición 5.6.5.** *El estadístico CUSUMSQ, la suma acumulada de residuos recursivos cuadrados, se define como*

$$CUSUMSQ_t = \frac{\sum_{i=k}^{t} W_{i+1,i}^2}{\sum_{i=k}^{T} W_{i+1,i}^2}, \quad t = k, k+1, \ldots, T \tag{5.19}$$

Si existe estabilidad estructural la gráfica de $CUSUMSQ_t$ versus $t$ es una línea recta. Períodos en los cuales aumente la inestabilidad se notarán apartados de esta línea. Sin embargo, el siguiente comentario de A. Zeileis es importante ([2]):

[2] https://stat.ethz.ch/pipermail/r-help/2005-January/064022.html

"The CUSUMSQ test is more appropriate if you want to detect changes in the variance (rather than the conditional mean)".

Aunque el estadístico $CUSUMSQ_t$ es fácil de programar, los valores críticos no están disponibles en R.

**Definición 5.6.6.** *El estadístico MOSUM, basado en una media de residuales recursivos (ó residuales de MCO), sobre una ventana móvil, se define con las cantidades:* $m = T-k-1$ *y* $0 < h < 1$*, la proporción de datos en la ventana móvil, como*

$$MOSUM_j = \frac{1}{\hat{\sigma}\sqrt{\lfloor hm \rfloor}} \max_{0 \leq j \leq m-\lfloor hm \rfloor} \left| \sum_{t=k+j+1}^{k+j+\lfloor hm \rfloor} W_{t+1,t} \right|, \quad j = 0, 1, \ldots, m - \lfloor hm \rfloor, \tag{5.20}$$

La región de aceptación de la prueba MOSUM para la hipótesis (5.13) es el interior de una región (ó banda) limitada por 2 líneas horizontales. Si la gráfica de MOSUM permanece dentro de esta banda entonces no se rechaza la hipótesis nula de estabilidad estructural. Los detalles están en Chu et al. [1995, pag. 606]. En este artículo se hace la anotación de que la prueba MOSUM puede ser más sensible a los cambios estructurales por el hecho de estar basada en una ventana móvil y no en la totalidad de los datos.

### 5.6.3. Implementación en R: la librería strucchange

La librería strucchange, ver Zeileis et al. [2002] y Zeileis et al. [2003], implementa las pruebas CUSUM y MOSUM de cambios estructurales en modelos de regresión. Sin embargo, no implementa la prueba CUSUMSQ.

**Funciones de la librería**

1. **recresid:** Calcula los residuales recursivos ó errores de predicción a un paso estandarizados, en un modelo de regresión lineal. La suma acumulada de éstos es el CUSUM. La instrucción en R para calcularlos es `rrc = recresid(y ~ t + It).`

2. **efp:** Calcula los residuales recursivos junto con la región de aceptación (Bandas). Hay dos versiones. Una utiliza los residuales recursivos y otra con base en los residuales de la regresión por Mínimos Cuadrados (no se introdujeron éstos). Las instrucciones en R son, respectivamente,

```
prueba.cusum1 = efp( y ~ t + It, type = "Rec-CUSUM")
```

```
plot(prueba.cusum1)

prueba.cusum2 = efp(y ~ t + It, type = "OLS-CUSUM")
plot(prueba.cusum2)
```

3. **sctest:** Calcula la prueba formal de la hipótesis nula de estabilidad estructural, y depende del tipo de residuales que se utilicen para calcular, residuales recursivos ó residuales de Mínimos Cuadrados.

```
sctest(prueba.cusum1)
sctest(prueba.cusum2)
```

**Ejemplo 5.6.3.** *Con base en el Ejemplo (4.2.1), pag. 56, de la serie de producción trimestral de cemento, la gráfica del estadístico CUSUM para el modelo con tendencia lineal y variables indicadoras estacionales (4.7), pag. 56, está en la Figura 5.6.*

A partir de los resultados en la Figura 5.6 se rechazaría la hipótesis nula de estabilidad estructural en el modelo, ya que el estadístico CUSUM se sale de las bandas de confianza. Igual conclusión con la prueba MOSUM, que está en la Figura 5.7. La conclusión es que el modelo (4.7), pag. 56 detecta un cambio estructural en la serie.

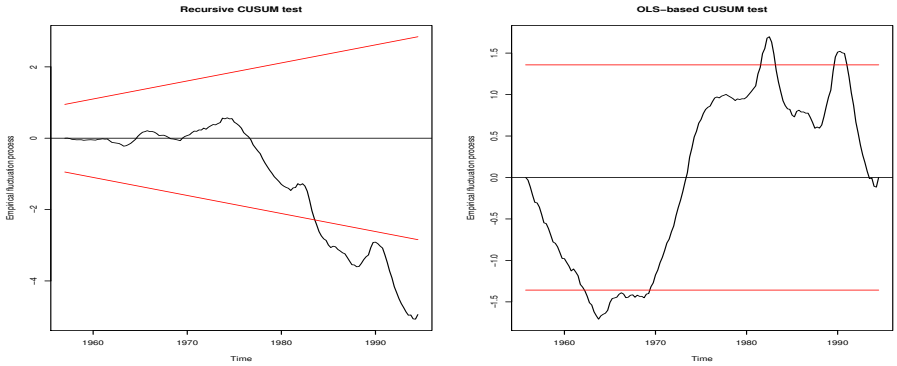
La prueba formal, con un nivel de significación de 5 %, confirma el rechazo de la hipótesis nula de estabilidad.

```
          estadistico      valor.p
Rec-CUSUM    1.703042 1.781954e-05
OLS-CUSUM    1.708886 5.814298e-03
Rec-MOSUM    2.246572 1.000000e-02
OLS-MOSUM    2.098502 1.000000e-02
```

El código R siguiente muestra la programación de los comandos con los cuales se generaron los análisis y las gráficas del ejemplo anterior.

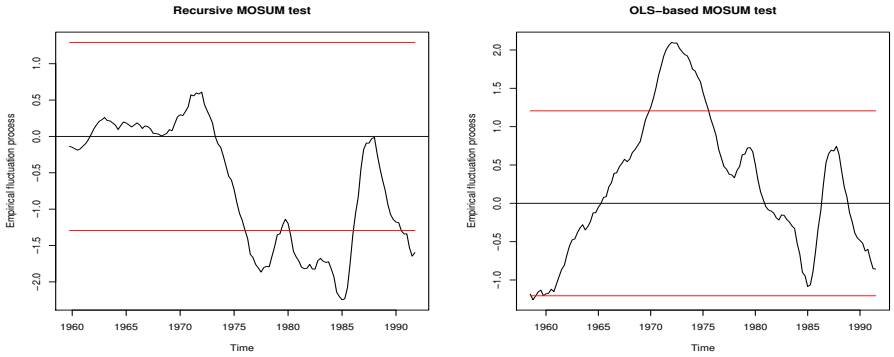
**Código R 5.6.2.**

```
#----programa en R para analisis de estabilidad estructural
#----con la serie de produccion de cemento en Australia, trimestral
library(strucchange)
library(forecast)
```




(a) prueba CUSUM residuales recursivos (b) prueba CUSUM residuales ols

Figura 5.6: Resultados prueba CUSUM gráfica




(a) prueba MOSUM residuales recursivos (b) prueba MOSUM residuales mco

Figura 5.7: Resultados prueba MOSUM gráfica

```
library(lmtest)

E = read.table("cementq.dat",
header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))

#----modelo con tendencia lineal y estacionalidad
T = length(y)
```

```
t = seq(1,T)
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)

#--- pruebe Durbin-Watson
dwtest(y ~ t + It)

#----estimacion recursiva de parametros
k = 2 + frequency(y) -1 + 10
n = T-k
parm = mat.or.vec(n, 5)
for(j in 1:n){
yj = y[1:(k+j)]
tj = t[1:(k+j)]
Itj = It[1:(k+j),]
mod.j = lm(yj ~ tj + Itj)
parm[j,] = t(mod.j$coefficient)
}
#-----grafica de las trayectorias
colnames(parm)=c("beta.0",
"beta.1","delta.1","delta.2",
"delta.3")
plot.ts(parm,main="")

#-----pruebas cusum graficas

prueba.cusum1 = efp(y ~ t + It, type = "Rec-CUSUM")
plot(prueba.cusum1)

prueba.cusum2 = efp(y ~ t + It, type = "OLS-CUSUM")
plot(prueba.cusum2)

prueba.cusum3 = efp(y ~ t + It, type = "Rec-MOSUM")
plot(prueba.cusum3)

prueba.cusum4 = efp(y ~ t + It, type = "OLS-MOSUM")
plot(prueba.cusum4)
```

```
#----pruebas cusum formales con valor p
```

```
#----pruebas cusum formales con valor p
B1=sctest(prueba.cusum1)
B2=sctest(prueba.cusum2)
B3=sctest(prueba.cusum3)
B4=sctest(prueba.cusum4)
B = rbind(c(B1$statistic,B1$p.value),
c(B2$statistic,B2$p.value),
c(B3$statistic,B3$p.value),
c(B4$statistic,B4$p.value))
colnames(B)=c("estadistico","valor.p")
rownames(B)=c("Rec-CUSUM","OLS-CUSUM","Rec-MOSUM","OLS-MOSUM")
(B)
```

**Observación:** La prueba CUSUM asume el supuesto de incorrelación de los errores, $Corr(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+k}) = 0$ para $k \neq 0$. Para decidir si se cumple el supuesto se utiliza el valor del estadístico Durbin-Watson. Si está cercano a 2 se acepta el supuesto, si no, se rechaza. En caso de rechazar el supuesto, Diebold [1999, pág 101, nota 10] afirma que no conviene realizar la prueba de estabilidad estructural. Una razón es que cuando hay autocorrelación de $\hat{\varepsilon}_t$, los errores estandar de los coeficientes, $\widehat{\sigma}_{\hat{\beta}_i}$, no son confiables. En el Ejemplo siguiente la prueba Durbin-Watson produce el resultado

```
DW = 0.3871, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

luego, los residuos estructurales están autocorrelacionados y las pruebas CUSUM pueden perder validez. Nótese que el estadístico DW es mucho menor de 2.0.

**Identificación del punto de quiebre**

La función `breakpoints` permite identificar los puntos de quiebre ó cambio estructural. Suponga el modelo $Y_t = \underline{X_t}\underline{\beta} + \varepsilon_t, \quad t = 1, \ldots, T$. Se asumen $m$ puntos de quiebre o de cambio estructural en los cuales los coeficientes cambian de un valor a otro, permaneciendo constantes entre los puntos. Los puntos de quiebre son tales que

$$Y_t = \underline{X_i}\beta_t + \mu_i, \quad i_{j-1} \leq t \leq i_j, \quad j = 1, \ldots, m. \tag{5.21}$$

Los índices $i_j$ de los puntos de quiebre se estima con la función `breakpoints`. Pero el número de puntos de quiebre, $m$, es arbitrario. Por ejemplo, para que estime $m = 2$ puntos

se utilizan las instrucciones:

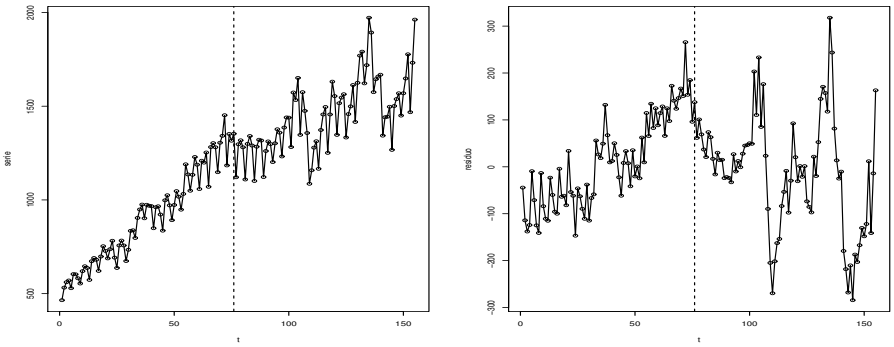
```
bp.n = breakpoints(y2 ~ t + It, breaks=2)
summary(bp.n)
B = bp.n$extract.breaks(bp.n$RSS.table,breaks=2)
```

En el vector B están las observaciones que corresponden a cambios estructurales.

**Código R 5.6.3.**

```
#----encontrar el punto de quiebre: determinar 2 puntos posibles
bp.n = breakpoints(y ~ t + It, breaks=2)
summary(bp.n)

B= bp.n$extract.breaks(bp.n$RSS.table,breaks=2)

#---- grafica de cusum con punto de quiebre
rcres = recresid(y ~ t + It)
plot(cumsum(rcres),type='l')
abline(v=B[1],lty=2,lwd=2)
abline(v=B[2],lty=2,lwd=2)

r1 = mod1$residuals
#----grafica de residuos OLS con punto de quiebre
plot(t,r1,type='o',ylab='residuo')
abline(v=B[1],lty=2,lwd=2)
abline(v=B[2],lty=2,lwd=2)

#----grafica de la serie con punto de quiebre
plot(t,y,type='o',ylab='serie')
abline(v=B[1],lty=2,lwd=2)
abline(v=B[2],lty=2,lwd=2)
```

El punto de quiebre se detecta en el cuarto trimestre de 1974 (la observación número 76). Ver el código en R anterior para la aplicación de la función `breakpoints`. En este caso concluímos que los pronósticos con este modelo pueden no ser válidos. Entonces preferimos los pronósticos con un modelo adaptativo como Holt-Winters, ó el Modelo de Componentes Estructurales de Harvey (5.7). También es posible ampliar el modelo de descomposición




(a) Con la Serie (b) Con los residuos

Figura 5.8: Punto de Quiebre 1974-Q4 en la Serie y en los Residuos

para incluír una intervención que modele el cambio en la observación 76. Pero el análisis de intervenciones es tema de otro capítulo.

# CAPÍTULO 6

# Medias Móviles y Filtros

## 6.1. Medias Móviles

En esta sección se introduce el concepto de Media Móvil. Hay muchos tipos de Medias Móviles con diferentes aplicaciones.

**Definición 6.1.1.** *Una Media Móvil es un operador lineal que transforma una serie* $\{Y_t,\ t = 1, 2, \ldots, T\}$*, en otra* $\overline{Y}_t$ *definida como*

$$\overline{Y}_t = \sum_{j=-m}^{m} w_j Y_{t-j},\ t = m+1, \ldots, T-m \tag{6.1}$$

*donde* $w_j \in \mathbb{R}, j = 0, \pm 1, \ldots, \pm m$*, son pesos predeterminados.*

En algunos casos para determinar la tendencia y reducir las fluctuaciones locales (suavizar) se los pesos restringen a $\sum_{r=-m}^{m} w_r = 1$. Una opción muy utilizada es pesos simétricos: $w_{-j} = w_j$.

Nótese que en la serie transformada $\overline{Y}_t$ se pierden $2m$ datos. Las Medias Móviles se utilizan como estimadores de la tendencia, es decir, $\widehat{T}_t = \overline{Y}_t$. Esto permite remover la tendencia al formar la serie sin tendencia $X_t = Y_t - \overline{Y}_t$. Las Medias Móviles bilaterales se utilizan también para estimar la componente estacional. En este caso $\widehat{S}_t = \overline{Y}_t$, y la serie $X_t = Y_t - \overline{Y}_t$

se denomina la serie des-estacionalizada. Algunos ejemplos de Medias Móviles son: (ver Diebold [1999, pág. 84])

**Ejemplo 6.1.1.** *La Media Móvil Bilateral con $w_j = w_{-j}$, $j = 0, 1, \ldots, m$ se define como*

$$\overline{Y}_t = \sum_{j=-m}^{m} w_j Y_{t-j}. \tag{6.2}$$

*El término $Y_{t-j}$ se llama "rezago de orden $j$". Si $j < 0$ el rezago $Y_{t-j}$ es un valor "futuro". Hay varias maneras en R de generar los pesos $w_j$ y de calcular la Media Móvil. Para el caso de pesos iguales*

$$w_j = \frac{1}{2m+1},\ j = -m, \ldots, m, \tag{6.3}$$

*se puede hacer con las funciones* `kernel()` *y* `kernapply()`*. Por ejemplo, una Media Móvil con $m = 4$ tiene $2(4) + 1 = 9$ pesos, y la instrucción para calcularla es*

```
k4 = kernel("daniell", m=4)
# el parámetro m=4 es el ancho de ventana de la Media Móvil
Yt.s = kernapply(Yt, k4)
plot(Yt)
lines(Yt.s,col='red')
```

**Ejemplo 6.1.2.** *La Media Móvil Unilateral*

$$w_j = \frac{1}{m+1},\ j = 0, 1, \ldots, m se define como \tag{6.4}$$

*El entero $2m + 1$ es el ancho de ventana de la media móvil, $m = 2, 3, \ldots$. La Media Móvil unilateral (6.4) utiliza valores presentes y pasados de la serie. Es un ejemplo de filtro lineal "causal" . La Media Móvil Bilateral (6.2) es un filtro lineal no causal.*

**Ejemplo 6.1.3.** *Un ejemplo de Media Móvil unilateral es el Suavizamiento Exponencial Simple (SES), definido por*

$$\overline{Y}_t = \sum_{j=0}^{m} \alpha(1-\alpha)^j Y_{t-j},\ \alpha \in (0,1), m > 1. \tag{6.5}$$

*El método SES también se denomina con las siglas EWMA (Exponentially Weighted Moving Average). En R se coloca $m = \lfloor \frac{2-\alpha}{\alpha} \rfloor$, ó también $\alpha = \frac{2}{m+1}$.*

La ecuación (6.5) se puede obtener como solución de la relación recursiva

$$\bar{Y}_t = \alpha Y_t + (1-\alpha)\bar{Y}_{t-1}$$

para $t = 2, 3, \ldots, T$. Se comprueba que

$$\bar{Y}_t = (1-\alpha)^{m+1}\bar{Y}_{t-m-1} + \alpha\sum_{j=0}^{m}(1-\alpha)^j Y_{t-j}.$$

Si $m$ es grande se puede aproximar

$$(1-\alpha)^{m+1} \approx 0$$

y por tanto

$$\overline{Y}_t = \sum_{j=0}^{m}\alpha(1-\alpha)^j Y_{t-j}.$$

- Si $\alpha \approx 0$, entonces el filtro reduce la serie casi a una constante. Se comprueba a partir de la relación

  $$\overline{Y}_t = (1-\alpha)^{m+1}\overline{Y}_{t-m-1} + \alpha\sum_{j=0}^{m}(1-\alpha)^j Y_{t-j},$$

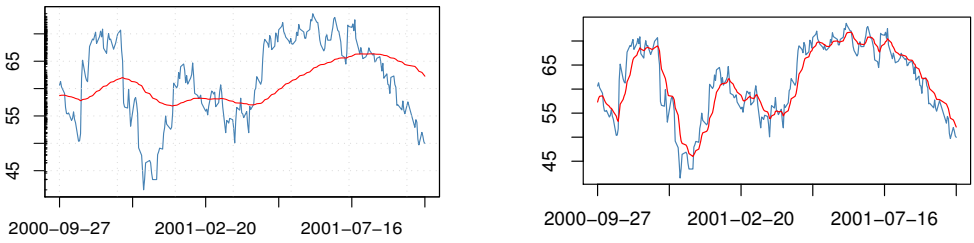
  ya que si $\alpha \approx 0$, entonces $\overline{Y}_t \approx \overline{Y}_{t-m-1}$ para todo $m > 1$. Con lo cual $\overline{Y}_t \approx \overline{Y}_0$.

- Si $\alpha \approx 1$, entonces $\overline{Y}_t = Y_t$, es decir, se reproduce la serie y no la suaviza. Con $\alpha = 0.2$ ya se observa esta característica.

En R la función `emaTA(x, lambda = alfa , startup = 0)` de la librería `fTrading` implementa EWMA. Se usa la fórmula $\alpha = \frac{2}{m+1}$, y a la función se le puede ingresar $m$ en lugar de $\alpha$. El siguiente es un ejemplo de código en R para EWMA.

**Ejemplo 6.1.4.** *Código en R para SES (EWMA) con la función* `emaTA()` *de la librería* `fTrading`. *Los resultados están en la Figura 6.1. Nótese que en "lambda" se ingresa el valor de* $\alpha$.

```
library(fTrading)
x = MSFT # Base de Datos de Microsoft OHLC
x = x[,"Close"] # Precios de Cierre
y = emaTA(x, lambda = 0.189)
seriesPlot(x)
lines(y, col="red")
```




(a) $\alpha = 0.019$ (b) $\alpha = 0.189$

Figura 6.1: Precios de la Acción de Microsoft Suavizados con SES

## La función filter de R

Mediante la función `filter()` de la librería stats se pueden implementar varios tipos de medias móviles, por ejemplo, medias móviles de tipo particular como medias Henderson y filtros recursivos. En esta sección se explica su funcionamiento y aplicaciones.

En R existen dos versiones de `filter`: La primera está en la librería `stats` que es cargada por defecto en R, y la segunda está en la librería `signal`. Cuando se utilice la de la librería signal se escribe `signal::filter()`. Esta última se describe al final de esta sección. La exposición siguiente es sobre la función `filter` de la librería stats. La forma de utilizarla es la siguiente.

```
x = filter(y, w, method = "convolution" ó "recursive",
sides = 1 ó 2, circular = TRUE ó FALSE, init)
```

Los argumentos se definen así.

| | |
|---|---|
| `y`: | Es la serie que se quiere suavizar. Puede ser multivariada. |
| `w`: | Es el vector de pesos $\underline{w}$, con pesos $w_j$. |
| `method`: | Cualquiera de las dos `"convolution"` o `"recursive"`. |
| | Si es `"convolution"` la función calcula una media móvil. |
| | Si es `"recursive"` calcula un filtro recursivo. |
| `sides`: | Si se usa `"convolution"`, y `sides=1` calcula $\sum_{j=0}^{m} w_j Y_{t-j}$ |
| | Si `sides=2` calcula $\sum_{j=-m}^{m} w_{\|j\|} Y_{t-j}$. |
| `circular`: | Solo cuando se usa `"convolution"`, es TRUE ó FALSE. |
| `init`: | Solo cuando se usa `"recursive"`. |

**La opción: method = "convolution"**. Se utiliza para calcular Medias Móviles.

Para calcular una media móvil unilateral $\overline{Y}_t = \sum_{j=0}^{m} w_j Y_{t-j}$ se programa:

```
w = c(w0, w1,...,wm)
x = filter(y, w, "conv",1,TRUE,NULL)
```

Nótese que el primer valor de $w$, $w[1]$ se aplica al último valor de la serie, $Y_t$. Un ejemplo más simple, para $\overline{Y}_t = \frac{1}{m+1} \sum_{j=0}^{m} Y_{t-j}$, se programa

```
x = filter(y,rep(1,m+1)/(m+1),"conv",1,TRUE,NULL)
```

Para calcular una media móvil doble $\overline{Y}_t = \sum_{j=-m}^{m} w_j Y_{t-j}$ se ingresa el vector de pesos $\underline{w} = (w_k, w_{k-1}, \ldots, w_0, \ldots, w_{k-1}, w_k)$ con las opciones `sides=2` y `circular=TRUE`.

```
w = c(wk,wk1,...,w0,...,wk1,wk)
x = filter(y,w,"conv",2,TRUE,NULL)
```

Por ejemplo, para extraer la tendencia de la serie de temperaturas mensuales, "nottem" (ver Ejemplo 4.1, pag. 52) se programa

```
#---------Ejemplo media móvil doble
Yt = nottem
f <- frequency(Yt)
w <- c(0.5, rep(1, f - 1), 0.5)/f
Tt = filter(Yt,w,"conv",2,T,NULL)
Xt = Yt - Tt
t=seq(1,length(Yt))
par(mfrow = c(3,1))
```

```
plot(t,Yt,type='l')
plot(t,Tt,type='l')
plot(t,Xt,type='l')
```

**La opción: circular=TRUE**. En la Media Móvil unilateral se calcula $\overline{Y}_t = \sum_{j=0}^{m} w_j Y_{t-j}$. Si se escribe la serie en la forma $Y_1, \ldots, Y_T$ y se tienen los $m+1$ pesos $w_0, w_1, \ldots, w_m$, al aplicar `X=filter(Y,w,"conv",1,TRUE,NULL)` se obtiene una serie $X = (X_1, \ldots, X_T)$, calculada de la siguiente manera.

$$
\begin{aligned}
X_T &= w_0 Y_T + w_1 Y_{T-1} + \ldots + w_m Y_{T-m} \\
X_{T-1} &= w_0 Y_{T-1} + w_1 Y_{T-2} + \ldots + w_m Y_{T-m-1} \\
X_{T-2} &= w_0 Y_{T-2} + w_1 Y_{T-3} + \ldots + w_m Y_{T-m-2} \\
&\vdots \\
X_{m+1} &= w_0 Y_{m+1} + w_1 Y_m + \ldots + w_m Y_1 \\
X_m &= w_0 Y_m + w_1 Y_{m-1} + \ldots + w_{m-1} Y_1 + (w_m Y_T) \\
X_{m-1} &= w_0 Y_{m-1} + w_1 Y_{m-2} + \ldots + w_{m-2} Y_1 + (w_{m-1} Y_T + w_m Y_{T-1}) \\
&\vdots \\
X_2 &= w_0 Y_2 + w_1 Y_1 + (w_2 Y_T + w_3 Y_{T-1} + \ldots + w_m Y_{T-m+2}) \\
X_1 &= w_0 Y_1 + (w_1 Y_T + w_2 Y_{T-1} + \ldots + w_m Y_{T-m+1})
\end{aligned}
$$

El efecto de la opción `circular=TRUE` está en la última línea. Consiste en utilizar las sumas que están en paréntesis. Pero los valores que forman estas sumas son los últimos, luego, suaviza los valores iniciales con los últimos. Si se usa esta opción la serie suavizada $\overline{Y}_t$ no pierde datos. Pero puede ser inconveniente si la serie es, por ejemplo, creciente. De manera similar se reemplaza en las medias bilaterales. Por lo que no siempre es aconsejable usarla.

**Ejemplo 6.1.5.** *Cálculo de EWMA con la función* `filter`. *Como en EWMA (6.5) se tiene*

$$X_t = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^m Y_{t-m},$$

*se puede generar $X$ así:*

```
m = floor(1/a - 1)
w = a*(1-a)^seq(0,m,1)
x = filter(y,w,"conv",1,TRUE,NULL)
```

**Ejemplo 6.1.6.** *La función* `kernel()` *se utiliza con* `kernapply()`. *Para utilizar* `filter()` *para una media móvil bilateral con ancho de ventana $m = 10$ y pesos iguales, se programa así:*

```
k4 =  kernel("daniell", m=10)
xs = filter(y,k4,"convolution",2,FALSE,NULL)
```

**La opción: method = recursive"**. Calcula una serie $X$ definida por

$$X_t = \sum_{j=1}^{m} w_j X_{t-j} + Y_t, \tag{6.6}$$

donde los pesos $w_j$ se deben ingresar en un vector de la forma $(w_1, \ldots, w_m)$. A partir de la serie $Y = (Y_{t-T+1}, Y_{t_T}, \ldots, Y_{t-1}, Y_t)$ calcula $X = (X_{t-T+1}, \ldots, X_{t-1}, X_t)$. Nótese que $X_{t-T+1} = \sum_{j=1}^{m} w_j X_{t-j-T+1} + Y_{t-T+1}$ requiere los valores $X_{t-j-T+1}, j = 1, \ldots, m$. Pero estos $m$ valores no están disponibles y deben proporcionarse en un vector de valores iniciales, por ejemplo $Z = (Z_1, \ldots, Z_m)$. El programa se muestra a continuación.

**Ejemplo 6.1.7.** *Para calcular* $X_t = \sum_{j=1}^{m} w_j X_{t-j} + Y_t$

```
w = c(w1,...,wm)
z = c(z1,...,zm)
x = filter(y, w, "rec",init=z)
```

**Ejemplo 6.1.8.** *Para calcular* $X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j Y_{t-j}$ *se hace una combinación de "conv" y "rec" como sigue.*

```
a = c(a1,...,ap)
b = c(b0,...,bq)
z = c(z1,...,zp)
u = filter(y, b, "conv",1,T,NULL)
x = filter(u, a, "rec", init=z)
```

### La función `filter()` de la librería `signal`

La librería `signal` tiene otra versión de la función `filter`. La forma de aplicar filter es `x = signal::filter(b,a,y)`. Los vectores $a$ y $b$ se definen como:

$$a = c(1, a_1, \ldots, a_p)$$
$$b = c(b_0, b_1, \ldots, b_q)$$

El vector $a$ se denomina de coeficientes recursivos, el vector $b$ se denomina de coeficientes de media móvil. Note que la primera componente del vector $a$ es 1. Al aplicar

`x=signal::filter(b,a,y)` a $Y = (Y_{t-T+1}, Y_{t_T}, \ldots, Y_{t-1}, Y_t)$ se obtiene una serie filtrada $X = (X_{t-T+1}, \ldots, X_{t-1}, X_t)$ definida por:

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} = \sum_{j=0}^{q} b_j Y_{t-j}. \tag{6.7}$$

**Ejemplo 6.1.9.** *Suponga que se quiere calcular $X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=0}^{q} b_j Y_{t-j}$. Se puede utilizar la función filter de la librería signal ó la función filter de la librería stat. Con la librería stat el programa es el mismo del ejemplo (6.1.8). La programación con la función filter de signal es como sigue.*

```
a = c(a1,...,ap)
b = c(b0,...,bq)
z = c(z1,...,zp)
x = signal::filter(b, c(1,-a),y)
```

*Nótese que se cambia el signo al vector a para que al reemplazar en la definición en (6.7) se obtenga el resultado deseado. Además, nótese que al cargar la librería signal se debe diferenciar la función filter indicando de cuál librería se trata colocando* `x=signal::filter`. *No parece que haya una ventaja muy grande al utilizar esta versión de filter.*

## 6.2. Método de descomposición con base en Medias móviles

En esta sección se exponen dos procedimientos para estimar las componentes $T_t$ y $S_t$ de la serie $Y_t = T_t + S_t + \varepsilon_t, t = 1, \ldots, T$, con base en medias móviles dobles de la forma

$$\sum_{j=-m}^{m} w_j Y_{t+j}, \quad t = m+1, \ldots, T-m. \tag{6.8}$$

Cuando se combinan estos dos procedimientos, producen resultados muy similares a la función de descomposición `stl()`. Los procedimientos se basan en un algoritmo que está en Brockwell and Davis [2002, pág. 24] y se describe a continuación.

La primera parte del procedimiento consistes en estimar la tendencia con una media móvil diseñada para eliminar la componente estacional y disminuír el ruido aleatorio, definida como sigue.

Si el período $s$ es par, se coloca $s = 2q$ y para $q + 1 \leq t \leq T - q$ se define

$$\widehat{T_t} = \frac{1}{s}\left(\frac{1}{2}Y_{t-q} + Y_{t-q+1} + \cdots + Y_t + \cdots + Y_{t+q-1} + \frac{1}{2}Y_{t+q}\right), \tag{6.9}$$

Si el período $s$ es impar, se coloca $s = 2q + 1$, y para $q + 1 \leq t \leq T - q$,

$$\widehat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t-j}, \quad . \tag{6.10}$$

La segunda parte del procedimiento consiste en estimar la componente estacional. Primero se calcular $\widehat{T}_t$ con el filtro (6.9) ó (6.10). Segundo, se estima $\widehat{S}_k$ para cada $k = 1, 2, \ldots, s$ (el patrón estacional) promediando los valores de $Y_t - \widehat{T}_t$ para todos los $t$ que corresponden a la estación $k$ (son los $t = s\lfloor t/s \rfloor + k$). Con $q = \lfloor s/2 \rfloor$ se define

$$\hat{S}_k^{(1)} = \frac{1}{b-a+2} \sum_{j=a}^{b} Y_{k+sj} - \widehat{T}_{k+sj}, \quad k = 1, 2, \ldots, s, \tag{6.11}$$

donde $a = \left\lfloor \frac{(q-k)_+}{s} \right\rfloor + 1$ y $b = \left\lfloor \frac{(T-q-k)}{s} \right\rfloor$.

Tercero, se coloca

$$\hat{S}_k = \hat{S}_k^{(1)} - \frac{1}{q} \sum_{k=1}^{s} \hat{S}_k^{(1)}, \quad k = 1, 2, \ldots, s, \tag{6.12}$$

para garantizar $\sum_{k=1}^{s} \hat{S}_k = 0$ que es una condición equivalente a eliminar la última estación, y tiene como objeto garantizar la identificabilidad de la componente $S_t$. Finalmente se repite el patrón obtenido el número de veces necesario para formar $\hat{S}_t,\ t = 1, \ldots, T$.

El algoritmo se puede implementar en una función en R como sigue.

**Código R 6.2.1.**

```
#----------algoritmo Brockwell-Davis
alg.bro.dav = function(x,p){
n <- length(x)
q <- floor(p/2)
a <- rep(1,2*q+1)

if(q==p/2){
    a[2*q+1] <- 0.5
    a[1] <- 0.5
    a <- a/p
    m <- stats::filter(x,a,"conv",2,F,NULL)
}
else{
    a <- a/p
```

```
    m <- stats::filter(x,a,"conv",2,F,NULL)
}

w <- double(p)

for(k in 1:p){
    j <- seq(floor(max(0,q-k)/p)+1, floor((n-q-k)/p), 1)
    w[k] <- sum(x[k+j*p] - m[k+j*p])/(floor((n-q-k)/p) -
            floor(max(0,q-k)/p) + 1)
}

s1 <- w - mean(w)
s <- c(rep(s1,floor(n/p)),s1[1:(n%%p)])
D <- list(T=m, S=s)
return(D)
}
```

**Ejemplo 6.2.1.** *La programación en R del filtro (6.9) se puede hacer con la función* `filter()`.

$$T_t = \frac{1}{12}\Big(\frac{1}{2}Y_{t-6} + \sum_{u=-5}^{5} Y_{t-u} + \frac{1}{2}Y_{t+6}\Big), \qquad t = 7, \ldots, 45, \tag{6.13}$$

```
#-----Programa R para estimar la tendencia
#-----en una serie con componente estacional
s = frequency(y)
w = c(0.5, rep(1, s - 1), 0.5)/s
Tt = filter(y,w,"conv",2,T,NULL)
# Notese que el filtro estima las componente de tendencia y estacionalidad
# exactamente como lo hace la función stl()
m = stl(y, "per")
Tt.stl = m$time.series[,2]
par(mfrow = c(1,1))
plot(t,Tt,type='l'); lines(t,Tt.stl,col='red');
```

## 6.3. Librerías para filtrado de series en R

Algunas librerías en R para filtrado de series son las siguientes.

La librería `robfilter` tiene siete filtros

1. `robreg.filter` Simple regression filters
2. `hybrid.filter` Median and repeated median hybrid filters
3. `dw.filter` Two-step location-regression-based filters
4. `wrm.filter` Weighted repeated median filters
5. `robust.filter` Regression filters with additional rules (outlier and level shift detection)
6. `adore.filter` Adaptive repeated median filters
7. `madore.filter` Multivariate adaptive repeated median filters

La librería `mFilter` (Decomposition of a time series into trend and cyclical components using various filters) de M. Barcilar, tiene los filtros clásicos

1. `bkfilter` Baxter-King filter of a time series
2. `bwfilter` Butterworth filter of a time series
3. `cffilter` Christiano-Fitzgerald filter of a time series
4. `hpfilter` Hodrick-Prescott filter of a time series
5. `trfilter` Trigonometric regression filter of a time series

### Otras técnicas de ajuste estacional

Existen otras técnicas de ajuste estacional que son técnicas ad-hoc para calcular indices estacionales y luego usar estos para destacionalizar. Un ejemplo es el método X11-ARIMA utilizado para desestacionalizar series por el *U.S. Census Bureau* ahora es X12-ARIMA en `www.census.gov/srd/www/x12a`

## 6.4. Problemas

1. (ver Brockwell and Davis [2002, pag.43, ex. 1.14]) Muestre que el filtro con coeficientes `[a-2, a-1, a0, a1, a2] = [-1, 4, 3, 4,-1]` pasa polinomios de grado 3 y elimina componentes estacionales de período 3.

# CAPÍTULO 7

# Validación de los Supuestos sobre los Errores

## 7.1. Introducción

En el Modelo de Componentes Estructurales aditivo la serie se asume que se puede descomponer en las componentes de tendencia, estacionalidad y error:

$$Y_t = T_t + S_t + \varepsilon_t, \quad t \in \mathbb{Z}.$$

La componente que interesa analizar en este capítulo es $\varepsilon_t$. La serie estimada $\widehat{\varepsilon}_t$, denominada "residuo estructural", resulta de sustraer de $Y_t$ las estimaciones de las dos primeras componentes, es decir

$$\widehat{\varepsilon}_t = Y_t - \widehat{T}_t - \widehat{S}_t.$$

Para calcular los pronósticos $\widehat{Y}_{T+j}$, $j = 1, \ldots, h$, se debe analizar la serie $\widehat{\varepsilon}_t$ para detectar si tiene una dinámica autocorrelacionada ó, por el contrario, es solamente ruido aleatorio. En el primer caso es posible pronosticar los residuos estructurales e incorporar éstos a los pronósticos con la tendencia y la componente estacional. En el segundo caso el pronóstico de un ruido aleatorio es cero. Para clarificar la idea de ruido aleatorio se introduce la siguiente definición básica.

**Definición 7.1.1 (Ruido Blanco).** *Una serie ($\varepsilon_t,\ t \in \mathbb{Z}$) se dice que es Ruido Blanco si cumple*

$$\mathbb{E}(\varepsilon_t) \equiv 0, \tag{7.1a}$$
$$Var(\varepsilon_t) \equiv \sigma^2, \tag{7.1b}$$
$$\forall k \neq 0, Cov(\varepsilon_t, \varepsilon_{t+k}) = 0. \tag{7.1c}$$

*Se denota* $\varepsilon_t \sim RB(0, \sigma^2)$.

Si además cumple que $\varepsilon_t \sim N(0, \sigma^2)$ se dice que $\varepsilon_t$ es Ruido Blanco Gaussiano y se denota $\varepsilon_t \sim RBG(0, \sigma^2)$. Nótese que en este caso la sucesión debe ser iid, $\varepsilon_t \sim iid\, N(0, \sigma^2)$, ya que una sucesión incorrelacionada que se distribuya Normal debe ser iid. En este caso también se cumple $\mathbb{E}(\varepsilon_{t+k}|\varepsilon_t) = \mathbb{E}(\varepsilon_{t+k}) = 0$, luego, en una sucesión $\varepsilon_t \sim iid\, N(0, \sigma^2)$ los pronósticos a $k$ pasos son cero.

Obsérvese que si no se cumple $\forall k \neq 0,\ Cov(\varepsilon_t, \varepsilon_{t+k}) = 0$ entonces se tendría $\exists k \neq 0,\ Corr(\varepsilon_t, \varepsilon_{t+k}) \neq 0$. Lo cual implica que los errores estarán autocorrelacionados, y por tanto, podrían ser pronosticados. Por ejemplo, con $T_t = \beta_0 + \beta_1 t$,

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, \ldots, T, \quad \varepsilon_t \sim RBG(0, \sigma^2)$$
$$\hat{Y}_{T+k} = \hat{\beta}_0 + \hat{\beta}_1 (T + k) + \hat{\varepsilon}_{T+k},\ k = 1, 2, \ldots, h,$$

Verificar las condiciones (7.1a), (7.1b) y (7.1c) se denomina: "Validación de los supuestos de los errores". La condición (7.1a) se cumple automáticamente si $\varepsilon_t$ es el residuo de la regresión $Y_t = T_t + S_t + \varepsilon_t$. Las otras dos condiciones se comprueban aplicando pruebas de hipótesis adecuadas. El resto del capítulo se dedicará a exponer pruebas para (7.1c), denominadas "pruebas de incorrelación".

Para la condición de varianza constante, (7.1b), las pruebas disponibles en la literatura, por ejemplo, White, Breusch-Pagan y Goldfeld-Quandt, se denominan "pruebas de heterocedasticidad". Asumen que la hipótesis alterna equivale a suponer que $Var(\varepsilon_t)$ es una función de una combinación de las variables explicativas, por ejemplo, de $t$, $t^2$, las variables indicadoras en las columnas de la matriz $It$, ó de los valores estimados $\widehat{Y}_t$, que es una combinación particular de las anteriores.

Esta forma de la hipótesis alterna no es útil si se asume el modelo de componentes, ya que lo que se busca es precisamente eliminar el efecto de tales variables en los residuos estructurales. Rechazar la hipótesis nula de varianza constante llevaría a tratar de corregir el efecto de heterocedasticidad y a modificar el modelo de manera sustancial. Por lo que optamos por no realizar la validación de este supuesto en esta etapa del modelo y concentrarnos en las pruebas de incorrelación.

## 7.2. Series Estacionarias en Covarianza

Para determinar si $\widehat{\varepsilon}_t, t = 1, \ldots, T$ presenta autocorrelación o no, se introducen dos pruebas de hipótesis y el estadístico función de autocovarianza. El concepto de serie estacionaria es básico en la teoría de series de tiempo. Es diferente del concepto de serie estacional.

Al considerar una serie de tiempo $Y_t$ se asumió que el índice $t$ toma valores enteros, negativos y positivos, $t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, escribiendo $(Y_t,\ t \in \mathbb{Z})$. Los valores observados de $Y_t$ forman un conjunto finito que se denomina una *trayectoria muestral*, denotada $\{Y_1, Y_2, \ldots, Y_T\}$.

**Definición 7.2.1 (Serie de Tiempo Estacionaria en Covarianza).** *Una serie* $(Y_t,\ t \in \mathbb{Z})$ *se dice estacionaria en covarianza si cumple dos condiciones.*

$$\mathbb{E}(Y_t) \equiv \mu, \tag{7.2a}$$

$$Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1). \tag{7.2b}$$

*donde* $Cov(Y_{t_1}, Y_{t_2}) = E((Y_{t_1} - \mu)(Y_{t_2} - \mu))$ *y* $R(t)$ *debe ser una función par.*

Es decir, la covarianza entre $Y_{t_1}$ y $Y_{t_2}$ depende únicamente de la distancia entre los tiempo $t_2$ y $t_1$, y no de si $t_1 < t_2$ ó $t_1 > t_2$. Es decir,

$$Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1) = R(t_1 - t_2) = Cov(Y_{t_2}, Y_{t_1}).$$

Por ejemplo,

$$Cov(Y_{-4}, Y_3) = R(3 - (-4)) = R(7) = Cov(Y_0, Y_7),$$
$$Cov(Y_3, Y_4) = R(4 - 3) = R(1) = Cov(Y_0, Y_1).$$

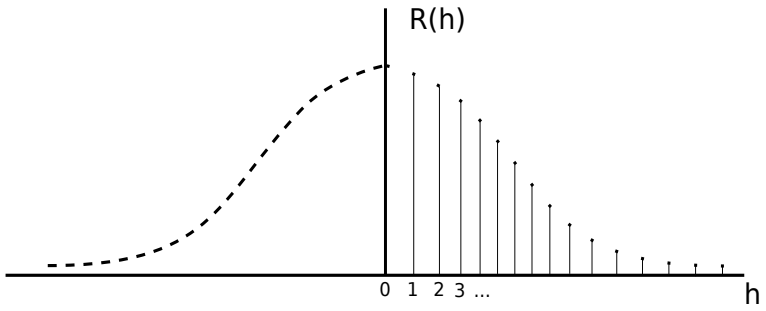



Figura 7.1: Ejemplo Función de Autocovarianza

Nótese que por ser funciones pares las autocovarianzas solamente se grafican en el semieje positivo. Las siguientes identidades son inmediatas.

$$\begin{aligned} Cov(Y_t, Y_{t+h}) &= E((Y_t - \mu)(Y_{t+h} - \mu)) \\ &= E(Y_t Y_{t+h}) - \mu E(Y_t) - \mu E(Y_{t+h}) + \mu^2 \\ &= E(Y_t Y_{t+h}) - \mu^2 - \mu^2 + \mu^2 \\ &= E(Y_t Y_{t+h}) - \mu^2. \end{aligned}$$

Algunos ejemplos de funciones de autocovarianza para series estacionarias en covarianza se muestran a continuación.

$$R(h) = \sigma^2 e^{-\alpha|h|},\ \alpha > 0, \tag{7.3a}$$

$$R(h) = \sigma^2 e^{-\alpha|h|}\, cos(\beta h),\ \alpha > 0, \beta \in \mathbb{R}, \tag{7.3b}$$

$$R(h) = \sigma^2 e^{-\alpha|h|}(\, cos(\beta h) + \frac{\alpha}{\beta}\sin(\beta|h|)),\ \alpha > 0, \beta \in \mathbb{R}, \tag{7.3c}$$

$$R(h) = 2\alpha\sigma^2(1 - 2\alpha h^2)e^{-\alpha h^2},\ \alpha > 0, \tag{7.3d}$$

$$R(h) = \sigma^2 e^{-\alpha^2 h^2},\ \alpha > 0, \tag{7.3e}$$

$$R(h) = \sigma^2 e^{-\alpha|h|}(1 + \alpha|h| + \frac{1}{3}\alpha^2 h^2),\ \alpha > 0, \tag{7.3f}$$

**Algunas propiedades de las series estacionarias en covarianza**

**Definición 7.2.2.** *Un componente determinístico es una serie $Z_t$ que tiene la propiedad de que si se conoce su historia hasta un tiempo, por ejemplo, $t_0$, dada por $Z_s, s \leq t_0$, entonces se puede conocer $Z_t$ para $t > t_0$.*

Por ejemplo, si $A \sim N(\mu, \sigma^2)$ entonces $Z_t = A\, cos(\lambda t), t \geq 0$ es un componente determinístico.

Algunas propiedades permiten formarse una idea más clara de como es una serie estacionaria en covarianza.

1. Una serie estacionaria en covarianza con trayectorias monótonas (crecientes ó decrecientes) solamente puede ser una constante. Es decir,

$$Y_t \leq Y_{t+1},\ \forall t \in \mathbb{Z} \Rightarrow Y_t = Y_0,$$
$$Y_t \geq Y_{t+1},\ \forall t \in \mathbb{Z} \Rightarrow Y_t = Y_0.$$

2. Una serie estacionaria en covarianza toma valores en un intervalo $[a, b]$ infinitas veces ó nunca. Es decir, si

$$A = (\#\{t : Y_t \in [a, b]\} = 0),$$

$$B = (\#\{t : Y_t \in [a, b]\} = \infty),$$

entonces $A \cup B = \Omega$, $A \cap B = \phi$.

3. En vista de las dos propiedades anteriores, una serie estacionaria en covarianza fluctúa alrededor de la media $\mathbb{E}(Y_t) = \mu$. Y se mantiene dentro de un intervalo finito. Se dice entonces que presenta reversión en la media.

4. La función de autocovarianza de una serie estacionaria en covarianza sin componentes determinísticas, $R(h)$, tiende a cero cuando $h$ tiende a infinito:

$$\lim_{h\to\infty} R(h) = 0.$$

Como este resultado es de la forma $p \Rightarrow q$, se cumple la contra-recíproca, es decir, $no(q) \Rightarrow no(p)$. Por lo tanto, si la función de autocorrelación $R(h)$ no tiende a cero la serie no puede ser estacionaria en covarianza.

Una serie de tiempo $(X_t, t \in \mathbb{Z})$, se dice determinística si, dados sus valores para $t \leq t_0$, se pueden determinar complentamente sus valores para $t > t_0$.

En el modelo de componentes $Y_t = T_t + S_t + \epsilon_t$, lo que se espera es que los residuos $\epsilon_t$ sean una serie estacionaria en covarianza, incluyendo un ruido blanco.

**Definición 7.2.3** ( Función de Autocorrelación)**.** *La función de autocorrelación teórica de una serie estacionaria en covarianza* $(Y_t,\ t \in \mathbb{Z})$ *se define como*

$$\rho(h) := Corr(Y_t, Y_{t+h}) = \frac{R(h)}{R(0)}, \quad h = 0, 1, 2, \ldots$$

*La función de autocorrelación se indicará con las siglas "fac" (acf en inglés). ,*

Como $|\rho(h)| \leq 1$ se cumple entoces que $\left|\frac{R(h)}{R(0)}\right| \leq 1$, es decir $|R(h)| \leq |R(0)| = |\sigma^2| = \sigma^2$, con $R(0) = Var(Y_t) = \sigma^2 > 0$, entonces $|R(h)| \leq \sigma^2,\ \forall h$, donde se asume también que $\sigma^2 < \infty$.

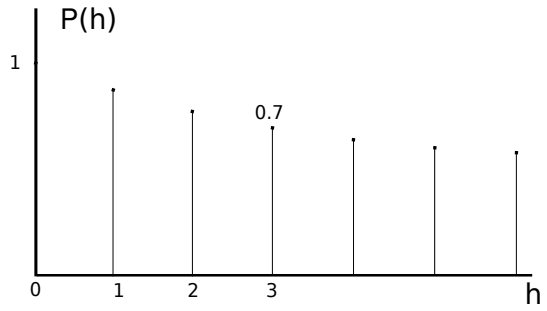
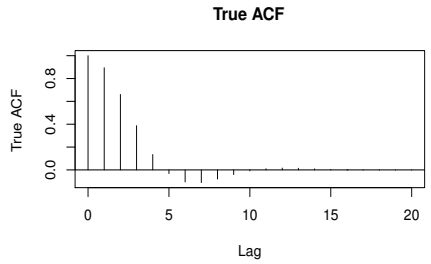


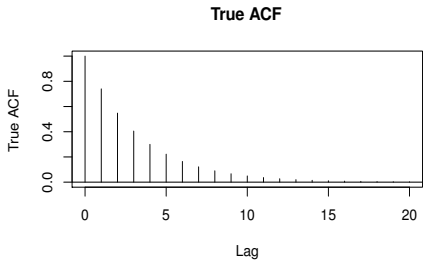

Figura 7.2: Ejemplo Función de Autocorrelación.

Los valores de $\rho(h)$ indican qué grado de asociación lineal existe entre dos valores de la serie $Y_t$ separados una distancia en el tiempo de $h$. Por ejemplo, $\rho(3) = 0.7$ se interpreta como que $Y_t$ y $Y_{t+3}$ tienen una correlación de $0.7$.

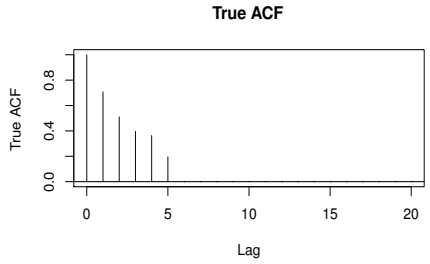
Si se observan las gráficas de $\rho(k)$ en las figuras (7.2), éstas parecen tender a cero, luego se puede aceptar que hay evidencia de que las series correspondientes son estacionarias en covarianza.




(a) Fac 1




(b) Fac 2




(c) Fac 3

Figura 7.3: Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario

### 7.2.1. Estimación de las funciones de Autocovarianza y Autocorrelación.

A partir de la expresión teórica para $\rho(k)$

$$\rho(k) = \frac{E((Y_t - \mu)(Y_{t+k} - \mu))}{E(Y_t - \mu)^2} \tag{7.4}$$

se puede obtener el estimador (de momentos), $\hat{\rho}(k)$, basado en la trayectoria $\{Y_1, Y_2, \ldots, Y_T\}$

$$\hat{\rho}(k) = \frac{\sum_{j=k+1}^{T}(Y_j - \bar{Y})(Y_{j-k} - \bar{Y})}{\sum_{j=1}^{T}(Y_j - \bar{Y})^2}, \tag{7.5}$$

donde $\bar{Y} = \frac{1}{T}\sum_{j=1}^{T} Y_j$ es la media de la trayectoria. Una cuestión práctica es decidir el rango $m$ de los valores de $k = 1, 2, \ldots, m$. Una posibilidad es $m = T/4$.

También se utiliza

$$\hat{\rho}(k) = \frac{\dfrac{1}{T-k+1}\displaystyle\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\dfrac{1}{T-1}\displaystyle\sum_{t=1}^{T}(Y_t - \bar{Y})^2}. \tag{7.6}$$

El cálculo de la fac estimada se realiza en el lenguaje R mediante la función `acf(x)`. La opción `ci.type = "ma"` calcula las "bandas de Bartlett", que se definen más adelante.

```
acf(x, lag.max = m, ci.type = "ma",
type = c("correlation", "covariance", "partial"))
```

**Pregunta:** Cómo debería ser la Función de Autocorrelación para un Proceso Ruido Blanco?.

Si $(\varepsilon_t,\ t \in \mathbb{Z})$ es ruido blanco, como $E(\varepsilon_t) = 0$, se cumple que la media de $\varepsilon_t$ es constante y como $Cov(\varepsilon_t, \varepsilon_{t+k}) = \sigma^2$, para $k = 0$, entonces la función de autocovarianza cumple:

$$R(k) = \begin{cases} \sigma^2, & \text{si } k = 0 \\ 0, & \text{si } k \geq 1. \end{cases}$$

La función de autocorrelación teórica $\rho(k)$ (fac) de un ruido blanco es cero excepto en $k = 0$. Sin embargo, las autocorrelaciones muestrales $\hat{\rho}(k)$ de un ruido blanco tienen la propiedad siguiente.

**Proposición 7.2.1** (Teorema de Bartlett)**.** *Si $(\varepsilon_t,\ t \in \mathbb{Z})$ es ruido blanco entonces las autocorrelaciones muestrales $\hat{\rho}(k) = \widehat{Cov}(\varepsilon_t, \varepsilon_{t+k}),\ k = 1, 2, \ldots, m$, donde $1 < m < T$ es un entero arbitrario, con base en una muestra $\{\varepsilon_1, \ldots, \varepsilon_T\}$, cumplen*

*(i) $\widehat{\rho}(k)$ son idependientes.*

*(ii)* $\widehat{\rho}(k) \overset{a}{\sim} N(0, 1/T)$ *para* $T$ *grande.*

*donde* $\overset{a}{\sim}$ *significa: se distribuye aproximadamente.*

Es decir, si $T$ es grande y $\varepsilon_t$ es ruido blanco entonces todas las $\widehat{\rho}(k)$ se distribuyen Normales de media cero y varianza $\frac{1}{T}$. Por propiedad de la Normal que dice que para $X \sim N(0, \sigma^2)$ entonces

$$P(-2\sigma \leq X \leq 2\sigma) \approx 0.95,$$

las autocorrelaciones $\widehat{\rho}(k)$ deben estar en el intervalo $[-2/\sqrt{T}, 2/\sqrt{T}]$, con una probabilidad de $0.95$ aproximadamente. El valor de $m$ se puede tomar, como se mencionó, como $T/4$.

**Definición 7.2.4.** *La región en* $\mathbb{R}^2$ *dada por* $\{(x, y) \in [0, m] \times [-2/\sqrt{T}, 2/\sqrt{T}]\}$ *se denomina "banda de Bartlett".*

Se puede realizar un diagnóstico gráfico, con base en $\hat{\rho}(k)$, la fac estimada de los residuos estructurales estimados $\hat{\epsilon}_t$, con base en la banda de Bartlett, para decidir si los residuos son ruido blanco. Si se observa

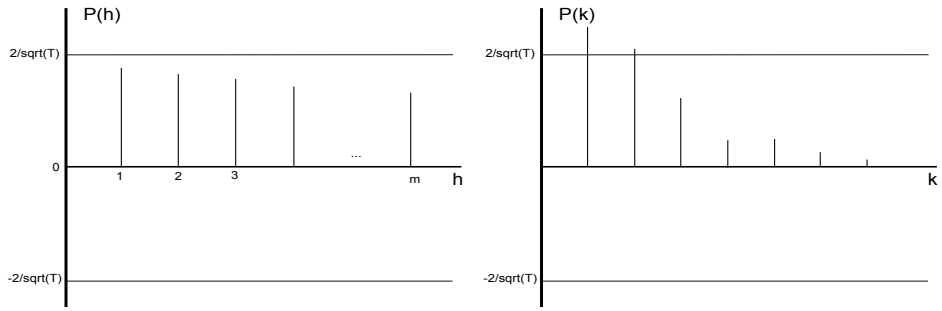
$$\forall k \leq m,\ \widehat{\rho}(k) \in [-2/\sqrt{T}, 2/\sqrt{T}] \tag{7.7}$$

entonces hay evidencia de que los residuos estructurales son ruido blanco y por tanto se puede pasar a realizar pronósticos de la serie con la parte estructural. Si se observan algunos $k$ para los cuales $\widehat{\rho}(k) \notin [-2/\sqrt{T}, 2/\sqrt{T}]$, entonces se tiene evidencia de que el residuo estructural no es ruido blanco y tiene una dinámica autocorrelacionada que se puede modelar, por ejemplo mediante un proceso ARMA, con el fin de mejorar los pronósticos estructurales.

Como ilustación, al observar la Figura (7.4(a)) , se puede concluír que hay evidencia de que $Y_t$ es ruido blanco. En la Figura (7.4(b)) no se cumple que $\forall k,\ \widehat{\rho}(k) \in [-2/\sqrt{T}, 2/\sqrt{T}]$, por tanto, hay evidencia de que $Y_t$ no es ruido blanco.

**Ejemplo 7.2.1.** *El codigo siguiente calcula la fac de la serie* `y=log(igbc)` *entre 2009-02-02 y 2010-12-30, y de su primera diferencia* `r=diff(log(y),1,1)`. *El resultado se muestra en la Figura (7.5).*

*Puede apreciarse que la fac de log(igbc) aparece muy persistente, tendiendo a cero lentamente, lo cual se puede interpretar como que no es estacionaria en covarianza. En el capítulo 7 se analizará este hecho mediante pruebas de hipótesis. En cambio, la fac de la serie de las diferencias aparece con todos sus valores dentro de las bandas de Bartlett, por lo que puede concluírse que puede tratarse de un ruido blanco. El valor* $m = 60$ *corresponde a 60 días, y se escogió arbitrariamente.*




(a) Fac 1. Caso de ruido blanco (b) Fac 2. Caso de serie autocorrelacionada

Figura 7.4: Ejemplos de Fac y Bandas de Bartlett

```
# cálculo de la fac
t = seq(1,length(y))
r = diff(y,1,1)
par(mfrow=c(2,2))
plot(t,y,type='l')
acf(y,60)
plot(t[-1],r,type='l')
acf(r,60)
```

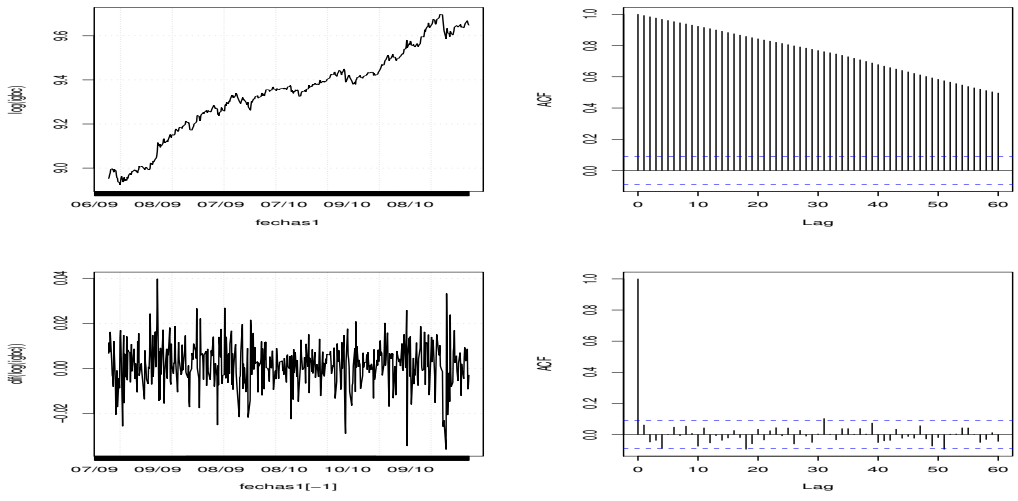



Figura 7.5: log(igbc) y diff(log(igbc)), a izquierda, con sus fac muestrales, a derecha

**Ejemplo 7.2.2.** *Retomando el Ejemplo (4.2.1) en la pag. 56, para modelación de la serie de producción de cemento Portland, trimestral, en toneladas por mil, entre Q1 1956 y Q3 1994, en Australia. Para la serie se asumió un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t.$$

*Es decir, un modelo lineal para la tendencia, más una componente estacional, de período $s = 4$. Al estimar los residuos estructurales $\widehat{\varepsilon}_t$ se obtienen las gráficas siguientes.*

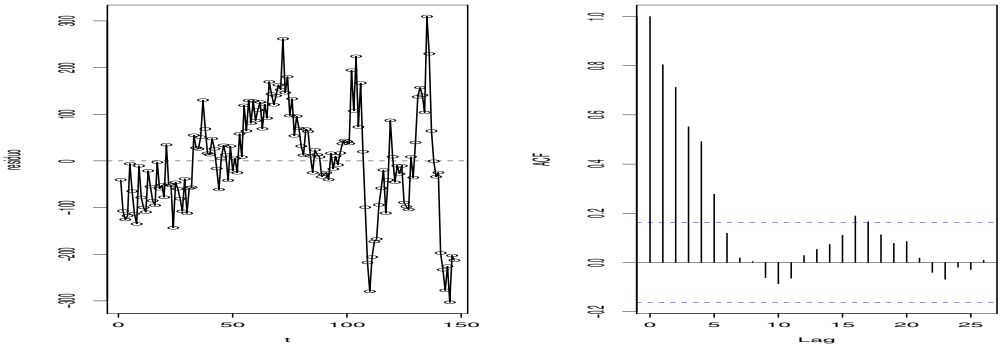



Figura 7.6: Fac de los Residuos Estructurales Serie Cementos

*Es evidente que los residuos no son ruido blanco. En la gráfica de la fac muestral, en el panel derecho de la Figura 7.6, se puede apreciar cómo las primeras cinco autorcorrelaciones se salen de las Bandas de Bartlett. Además, en la gráfica del panel izquierdo, la gráfica de $\widehat{\varepsilon}_t$ muestra períodos de incremento y luego de decrecimiento, que puede tomarse como evidencia de autocorrelación.*

## 7.3. Pruebas de Incorrelación

Decidir si una serie es ruido blanco usando la función de autocorrelación estimada $\hat{\rho}(k)$ es equivalente a decidir si las autocorrelaciones estimadas son todas estadísticamente cero, para $k \geq 1$. Además de la gráfica de $\hat{\rho}(k)$ con las bandas de Bartlett, se utilizan las siguientes pruebas de hipótesis para incorrelación.

1. Prueba Ljung-Box.

2. Prueba Durbin-Watson.

En las pruebas de hipótesis consideramos cuatro pasos:

1. Hipótesis de la Prueba.
2. Estadístico de la Prueba.
3. Distribución del Estadístico bajo $H_0$.
4. Decisión.

El Error Tipo I es rechazar la Hipótesis nula $H_0$ siendo ésta cierta. Se trata de un falso positivo porque en ocasiones la alterna $H_1$ es el resultado de interés. El nivel de significación $\alpha$ se define como

$$\mathbb{P}(\text{Error Tipo I}) = \alpha.$$

El Error Tipo II es no rechazar la nula $H_0$ siendo la alterna $H_1$ la cierta. Se trata de un falso negativo. La probabilidad de cometer este error es

$$\mathbb{P}(\text{Error Tipo I}) = \beta.$$

Además, la probabilidad $1 - \beta$, denominada la potencia de la prueba, se interpreta como la probabilidad de rechazar la nula correctamente, es decir, la probabilidad de rechazar $H_0$ cuando $H_1$ es cierta.

Las pruebas que tienen una baja potencia tienden a no rechazar la nula $H_0$. No tienen capacidad para detectar $H_1$. En ocasiones, la alterna $H_1$ puede estar muy cerca de la nula $H_0$, y se requiere una prueba con alta potencia para detectarla.

### 7.3.1. Prueba Ljung-Box (LB)

A partir del Teorema de Bartlett, (7.2.1), si $\varepsilon_t \sim RB(0, \sigma^2),\ t = 1, \ldots, T$ y $\hat{\rho}(k) = \widehat{Corr}(\varepsilon_t, \varepsilon_{t+k}),\ k = 1, 2, \ldots, m$, entonces, para $T$ grande, se tiene

1. $\hat{\rho}(k) \overset{a}{\sim} N(0, 1/T)$.
2. $\hat{\rho}(k)$ son idenpendientes.

Luego $\sqrt{T}\hat{\rho}(k) \overset{a}{\sim} N(0, 1)$, y $(\sqrt{T}\hat{\rho}(k))^2 \overset{a}{\sim} \chi^2_1$, por lo tanto

$$Q = T \sum_{k=1}^{m} \hat{\rho}^2(k) \ \overset{a}{\sim}\ \chi^2_m. \tag{7.8}$$

La prueba de Ljung-Box utiliza este resultado para definir el estadístico de la prueba.

1. **Hipótesis de la Prueba**

$$H_0 : \varepsilon_t \sim RB(0, \sigma^2)$$
$$H_1 : no(H_0)$$

2. **Estadístico de Prueba**

   Ljung y Box modificaron el estadístico $Q$ en (7.8) para mejorar la aproximación de $Q$ a la $\chi^2_m$ en muestras pequeñas, usando

$$Q_{LB} = T(T+2)\sum_{k=1}^{m} \frac{1}{T-k}\hat{\rho}^2(k). \tag{7.9}$$

3. **Distribución del Estadístico de la Prueba bajo** $H_0$

   Si $H_0 : \varepsilon_t \sim RB(0, \sigma^2)$ es cierta se cumple $Q_{LB} \overset{a}{\sim} \chi^2_m$.

4. **Decisión**

   *a*) Colocando $Q_{\text{obs}}$ el estadístico observado y $Valor\ p = P(\chi^2_m \geq Q_{\text{obs}}|H_0$ cierto), si $Valor\ p < 0.05$, se rechaza $H_0$. Si $Valor\ p > 0.05$, no se rechaza $H_0$.

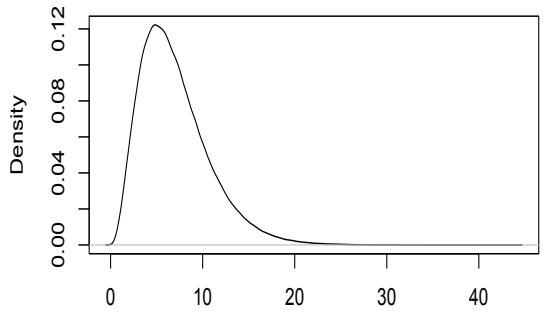



Figura 7.7: Densidad $\chi^2_m$.

   Un Valor muy pequeño de $Q_{\text{obs}}$ significa que *todas* las $\hat{\rho}^2(k)$ son pequeñas, luego no se rechaza $H_0$.

   *b*) **Usando tablas:** Si $\chi^2_{m,\alpha}$ es valor crítico de $\chi^2_m$ correspondiente a la probabilidad $\alpha$, $\alpha$ nivel de significancia, entonces si $Q_{\text{obs}} < \chi^2_{m,\alpha}$, no se rechaza $H_0$ y si $Q_{\text{obs}} > \chi^2_{m,\alpha}$ se rechaza $H_0$.

**Ejemplo 7.3.1.** *Continuando con el Ejemplo 7.2.2, la prueba para incorrelación de los residuos del modelo lineal con estacionalidad de los datos de producción de cemento,*

*resulta en:*
`X-squared = 18.0052, df = 25, p-value = 0.8422,`
*por tanto, no rechaza la hipótesis nula de que los residuos son Ruido Blanco. Igual sucede con el caso* `df = 35`.

**Ejemplo 7.3.2.** $m = 6$, $\alpha = 0.05$, $\chi^2_{6,0.05} = 12.54$ *Si el estadístico observado fuera* $Q_{\text{obs}} = 66.93$ *entonces* $Q_{\text{obs}} > \chi^2_{6,0.05}$, *por lo tanto, se rechaza* $H_0$

**Ejemplo 7.3.3.** *Ejemplo con datos de la accion de Siemens AG. Calcular la prueba Ljung-Box para los rendimientos logarítmicos*

**Código R 7.3.1.**

```
#-------------ejemplo accion Siemens bolsa Frankfurt
# https://es-us.finanzas.yahoo.com/q/hp?s=SIE.F+Precios+históricos

D = read.csv("siemens.SIE.F.csv", header=TRUE, stringsAsFactors=FALSE)
Xn = D$Close
Zn = diff(log(Xn),1,1)

#-------------graficar autocorrelacion (fac)
par(mfrow=c(2,2))
ts.plot(Zn)
ts.plot(Xn)
acf(Zn,30,ci.type="ma")
acf(Xn,30,ci.type="ma")

#-------------pruebas Ljung-Box
Box.test(Zn, lag = 15 , type =  "Ljung-Box")
Box.test(Zn, lag = 30 , type =  "Ljung-Box")
Box.test(Zn, lag = 90 , type =  "Ljung-Box")
Box.test(Zn, lag = 180 , type =  "Ljung-Box")

#-------------graficar autocorrelacion sin cero
par(mfrow=c(1,1))
acf(Zn,90,ci.type="ma")

B=acf(Zn,90,ci.type="ma")
plot(B$lag[-1],B$acf[-1],type='h',lwd=2)
abline(h=-2/sqrt(length(Zn)),col='blue',lty=2)
```

```
abline(h=2/sqrt(length(Zn)),col='blue',lty=2)

#-------------autocorrelacion
N = length(Zn)
cor(Zn[1:(N-40)],Zn[41:N])
```

### 7.3.2. Prueba Durbin-Watson (DW)

La prueba DW se basa en la definición del modelo $AR(1)$.

**Definición 7.3.1** ( Proceso $AR(1)$)**.** *Una serie* $(Y_t,\ t \in \mathbb{Z})$ *se dice* $AR(1)$ *con media* $\mu$ *si cumple que*

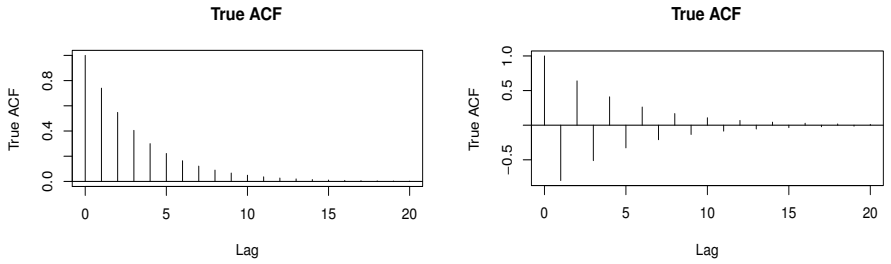
$$Y_t = \mu(1-\phi) + \phi Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{7.10}$$

*donde* $|\phi| < 1$*,* $\mu \in \mathbb{R}$ *y* $\varepsilon_t \sim RB(0, \sigma^2)$.

**Propiedades.** Si $Y_t \sim AR(1)$ entonces es un proceso estacionario en covarianza, tal que

- $E(Y_t) \equiv \mu$
- $Var(Y_t) \equiv \dfrac{\sigma^2}{1-\phi}$
- $Cov(Y_t, Y_{t+k}) = R(k) = \dfrac{\sigma^2\phi^k}{1-\phi^2},\ k = 0, 1, \ldots$
- $\rho(k) = \phi^k,\ k = 0, 1, 2, \ldots$

luego la fac teórica de un $AR(1)$ tiene las formas generales




(a) $\phi > 0$ (b) $\phi < 0$

Figura 7.8: Fac teóricas de un $AR(1)$.

Ahora si se tiene una serie $(Y_t,\ t = 1, \ldots, T)$ y se quiere probar que es incorrelacionada, asumimos que $\mathbb{E}(Y_t) = 0$, o si no, se transforma $Y_t^* = Y_t - \bar{Y}_t$.

La prueba DW se basa en asumir primero que $Y_t \sim AR(1)$, es decir, $Y_t$ sigue el modelo (7.10), con $\mu = 0$.

1. **Hipótesis de la Prueba**

$$H_0 : \phi = 0$$
$$H_1 : \phi \neq 0.$$

Luego si $H_0$ es cierta, $Y_t = \varepsilon_t$ y como $\varepsilon_t \sim RB(0, \sigma^2)$ entonces $Y_t \sim RB(0, \sigma^2)$

2. **Estadístico de Prueba**

**Definición 7.3.2.** *El estadístico DW se define como*

$$d = \frac{\sum_{t=2}^{T}(Y_t - Y_{t-1})^2}{\sum_{t=1}^{T} Y_t^2}. \tag{7.11}$$

Entonces como $\sum_{t=2}^{T}(Y_t - Y_{t-1})^2 = \sum_{t=2}^{T} Y_t^2 + \sum_{t=2}^{T} Y_{t-1}^2 - 2\sum_{t=2}^{T} Y_t Y_{t-1}$, se aproxima

$$\begin{aligned} d &= \frac{\sum_{t=2}^{T} Y_t^2 + \sum_{t=1}^{T-1} Y_t^2 - 2\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_t^2} \\ &\approx \frac{\sum_{t=2}^{T} Y_t^2 + \sum_{t=2}^{T} Y_t^2 - 2\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_t^2} \\ &= 2 - 2\hat{\rho}(1). \end{aligned}$$

es decir,

$$d \approx 2(1 - \hat{\rho}(1)). \tag{7.12}$$

A partir de (7.12) se puede concluír que, como $-1 < \hat{\rho}(1) < 1$ entonces $0 < d < 4$. Y si $\hat{\rho}(1) \approx 0$ entonces $d \approx 2$. Además,

- $d < 2 \Leftrightarrow 1 - \hat{\rho}(1) < 1 \Leftrightarrow \hat{\rho}(1) > 0$
- $d > 2 \Leftrightarrow \hat{\rho}(1) < 0$

luego un $d < 2$ indica posible autocorrelación positiva en la serie y $d > 2$ posible autocorrelación negativa.

3. **Distribución del estadístico bajo $\mathbf{H_0}$**. En el caso de $H_0$ cierta, $Y_t$ es un ruido blanco. Sin embargo, para la distribución de $d$ solamente se conocen los valores críticos para varios niveles de significación.

4. **Decisión**. Colocando $d_{\text{obs}}$ el estadístico observado y $Valor\ p = P(d \geq d_{\text{obs}}|H_0$ cierto), si $Valor\ p < 0.05$, se rechaza $H_0$. Si $Valor\ p > 0.05$, no se rechaza $H_0$.

4. **Prueba DW en R**

   *a*) La función `dwtest` en la librería `lmtest`. Permite probar hipótesis alternas $H_1 : \phi > 0$, $H_1 : \phi \neq 0$, $H_1 : \phi < 0$, colocando
   `dwtest(y ~ t + It, alternative = c("greater", "two.sided","less"))`.
   Calcula $d$ y el $Valor\ p$.

   *b*) La función `durbin.watson` en la librería `car`. Se escribe
   `durbinWatsonTest(y ~ t + It)`.

**Ejemplo 7.3.4.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t.$$

*Al estimar los residuos estructurales $\widehat{\varepsilon}_t$ se observó que posiblemente muestran una dinámica autocorrelacionada. Se puede confirmar este hecho con las pruebas Ljung-Box y Durbin-Watson (con las dos funciones de R mencionadas). Los resultados de R se muestran a continuación del código siguiente.*

*Es claro que los residuos estructurales presentan autocorrelación. Es decir, el supuesto de incorrelación no es válido.*

**Código R 7.3.2.**

```
# ejemplo de pruebas LB, DW
# para la serie de produccion de cemento
library(car)
library(forecast)
library(lmtest)
#----
E = read.table(cementq.dat, header = TRUE)
attach(E)
```

```
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
#----modelo con tendencia lineal y estacionalidad
#----con variables indicadoras estacionales
t = seq(1,length(y))
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)
et = residuals(mod1)
# pruebas Ljung-Box
Box.test(et, lag = 25, type = "Box-Pierce")
# pruebas Durbin-Watson
dwtest(mod1) # de la libreria lmtest
durbinWatsonTest(mod1) # de la libreria car


# resultados -------------------------------------
Box-Ljung test
X-squared = 293.9028, df = 25, p-value < 2.2e-16

#--------con dwtest
Durbin-Watson test
DW = 0.3646, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0

#--------con durbin.watson
lag Autocorrelation D-W Statistic p-value
1   0.8043803       0.3645806     0
Alternative hypothesis: rho != 0
-------------------------------------------------
```

### 7.3.3. Prueba Durbin-Watson Generalizada

Las pruebas Ljung-Box y Durbin-Watson coincidirían en su decisión siempre que la auto correlación de orden 1 sea dominante. Es decir, mayor que las restantes. Pero la prueba DW puede no rechazar la hipótesis nula de incorrelación, siendo falsa debido a autocorrelaciones significativas de órdenes altos. Para incorporar esta posibilidad, Vinod [1973] introdujo la prueba Durbin-Watson Generalizada (DWG). Consiste en una serie de pruebas en secuencia, para $j = 1, 2, \ldots$, asumiendo en cada $j$ un modelo de la forma $Y_t = \varphi_j Y_{t-j} + \eta_t$. La prueba

para el paso $j$ es

$$H_0 : \phi_j = 0$$
$$H_1 : \phi_j \neq 0.$$

dado que en las pruebas anteriores $j-1, \ldots, 1$, no se rechazó la hipótesis nula. El estadístico Durbin-Watson generalizado se define como

$$d_j = \frac{\sum_{t=j+1}^{T}(Y_t - Y_{t-j})^2}{\sum_{t=1}^{T} Y_t^2}. \tag{7.13}$$

La prueba DWG se encuentra implementada en la librería `car`. Por ejemplo, colocando

```
durbinWatsonTest(mod,max.lag=10)
```

se obtienen los valores de los estadísticos $d_j$, para $j = 1, \ldots, 10$, los correspondientes valores de $\widehat{\rho}(j)$ y los valores p. Esta secuencia de pruebas permite determinar el orden de la menor auto correlación significativa.

**Ejemplo 7.3.5.** *Retomando el Ejemplo (7.3.4), la prueba Durbin-Watson Generalizada, implementada en el código R siguiente, con los resultados en la Tabla 7.1, muestra que, efectivamente, la autocorrelación en los residuos no se limita al rezago $k = 1$ sino que también incluye hasta el rezago $k = 6$.*

**Código R 7.3.3.**

```
# ejemplo de prueba DWG
# para la serie de produccion de cemento
library(car)
library(forecast)
library(lmtest)

E = read.table(cementq.dat, header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))

#----modelo con tendencia lineal y estacionalidad
#----con variables indicadoras estacionales
t = seq(1,length(y))
It = seasonaldummy(y)
```

```
mod1 = lm(y ~ t + It)
summary(mod1)
r = residuals(mod1)
# prueba Durbin-Watson Generalizada
# de la libreria car
durbinWatsonTest(mod1,max.lag=10)
```

Tabla 7.1: Resultados de la Prueba DW Generalizada

| | Lag | rho.hat | dw | Valor.p |
|---|---|---|---|---|
| 1 | 1.00 | 0.80 | 0.39 | 0.00 |
| 2 | 2.00 | 0.71 | 0.55 | 0.00 |
| 3 | 3.00 | 0.60 | 0.76 | 0.00 |
| 4 | 4.00 | 0.53 | 0.89 | 0.00 |
| 5 | 5.00 | 0.34 | 1.26 | 0.00 |
| 6 | 6.00 | 0.20 | 1.53 | 0.01 |
| 7 | 7.00 | 0.10 | 1.71 | 0.21 |
| 8 | 8.00 | 0.07 | 1.74 | 0.19 |
| 9 | 9.00 | -0.01 | 1.89 | 0.93 |
| 10 | 10.00 | -0.06 | 1.96 | 0.53 |

## 7.4. Alternativas cuando los Residuos Estructurales muestran Autocorrelación

En el caso de detectarse autocorrelación en los residuos estructurales el procedimiento consistirá en buscar un modelo tipo ARMA, autorregresivo de media móvil, que se tratará en el Capítulo 6. Pero es posible utilizar otras alternativas.

**Transformación Cochran-Orcutt**

Una es la transformación Cochran-Orcutt, que permite re-estimar los parámetros cuando existe una autocorrelación de tipo AR(1) en los residuos.

Si la prueba DW detecta la alterna $\phi = \rho(1) \neq 0$, entonces se procede a transformar el modelo para la serie $Y_t$. Escribiendo el modelo estructural como un modelo de regresión lineal en (5.11), $Y_t = \underline{X}_t'\underline{\beta} + \varepsilon_t$, donde la prueba DW asume $\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$ y $\eta_t$ es ruido

blanco, la transformación Cochran-Orcutt se define como

$$Y_t - \phi Y_{t-1} = \alpha(1 - \widehat{\rho}(1)) + \beta(\underline{X}_t - \widehat{\rho}(1)\underline{X}_{t-1}) + \eta_t. \tag{7.14}$$

De esta manera, los residuos $\eta_t$ son ruido blanco. Esta transformación depende de una autocorrelación en los residuos tipo AR(1). No sería de esperar que funcionara en casos de estructuras de autocorrelación más complejas, como será el caso de la mayoría de las series de tiempo a considerar.

**Ejemplo 7.4.1.** *Retomando el Ejemplo (7.3.4), para modelación de la serie de producción de cemento Portland, trimestral, al aplicar la transformación Cochran-Orcutt mediante la función* `cochrane.orcutt` *de la librería* `orcutt`*, se obtiene que se rechaza la nula de ruido blanco. La función* `cochrane.orcutt`*, desafortunadamente, no es compatible con las funciones para Durbin-Watson y DWG.*

```
# transformacion Cochran-Orcutt
require(orcutt)
mod2 = cochrane.orcutt(mod1)
r = mod2$Cochrane.Orcutt$residuals
Box.test(r, lag = 25, type = "Ljung-Box")
# pruebas Durbin-Watson
dwtest(mod2) # no funciona!
durbinWatsonTest(mod2) # no funciona!
```

### Incluír rezagos de la variable dependiente en el modelo

Otra alternativa consiste en incluír en el modelo una ó varias componentes autorregresivas, es decir, aumentar el número de variables explicativas incluyendo rezagos de la variables dependiente. Supongamos que el modelo inicial es un modelo con tendencia lineal y componente estacional dado por

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \tag{7.15}$$

Y se incluyen en el modelo, por ejemplo, dos rezagos de la variable dependiente

$$Y_t = \beta_0 + \beta_1 t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \tag{7.16}$$

Con este cambio se trata de "capturar" la autocorrelación en los residuos estructurales estimados $\widehat{\varepsilon}_t$. En la librería `dynlm` se implementa este modelo mediante la función `dynlm`. La forma de aplicarla es con base en operadores de rezago. Por ejemplo, para el modelo (7.16) se coloca

```
require(dynlm)
mod2 = dynlm(y ~ t + It + L(y,1)+ L(y,2))
summary(mod2)
```

Sin embargo, se sabe que cuando se agregan rezagos de la variable dependiente, el estadístico Durbin-Watson se sesga hacia 2 y tiene baja potencia.

**Ejemplo 7.4.2.** *Retomando el Ejemplo (7.3.4), para modelación de la serie de producción de cemento Portland, trimestral. Se puede preguntar si el modelo (7.16) logra capturar la estructura autoregresiva, detectada anteriormente, tanto como para que la prueba Durbin-Watson no rechace la hipótesis nula. El programa siguiente muestra los pasos en la estimación de los modelos (7.15) y (7.16).*

**Código R 7.4.1.**

```
# ejemplo de pruebas LB, DW y DWG
# para la serie de produccion de cemento
# ademas ejemplo con rezagos con libreria dynlm
library(car)
library(forecast)
library(lmtest)
# modelo incorporando terminos rezagados de la serie
# utilizando la libreria dynlm
require(dynlm)
mod2 = dynlm(y ~ t + It + L(y,1)+ L(y,2))
summary(mod2)
r2 = residuals(mod2)
# pruebas Ljung-Box
Box.test(r2, lag = 15, type = "Ljung-Box")
# pruebas Durbin-Watson y Durbin-Watson Generalizada
durbinWatsonTest(mod2)
durbinWatsonTest(mod2,max.lag=10)
```

Los resultados de la estimación con el modelo (7.16) se muestran en la Tabla 7.2 siguiente. Sin embargo, las pruebas LB y DW para los residuos de este modelo, rechazan la hipótesis de ruido blanco.

```
#----------------resultados pruebas DW
Durbin-Watson test
```

Tabla 7.2: Resultados del Modelo estructural con dos rezagos 1,2 de $Y_t$

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 100.1147 | 35.5231 | 2.82 | 0.0055 |
| t | 0.8695 | 0.4035 | 2.15 | 0.0329 |
| ItQ1 | -159.4054 | 15.5466 | -10.25 | 0.0000 |
| ItQ2 | 74.9556 | 21.9855 | 3.41 | 0.0009 |
| ItQ3 | 65.1675 | 17.0202 | 3.83 | 0.0002 |
| L(yi, 1) | 0.6505 | 0.0828 | 7.86 | 0.0000 |
| L(yi, 2) | 0.2180 | 0.0839 | 2.60 | 0.0103 |

```
DW = 1.9078, p-value = 0.5338
alternative hypothesis: true autocorrelation is not 0
#----------------
durbinWatsonTest(mod2)
 lag Autocorrelation D-W Statistic p-value
   1      0.01205702      1.907794   0.498
 Alternative hypothesis: rho != 0
```

La prueba DWG con 10 rezagos rechaza la hipótesis nula. Aunque las auto correlaciones de órdenes 1,2,3 no sean significativas y no rechazan la nula, la de orden 4 sí es significativa, y se rechaza la hipótesis de ruido blanco para este modelo. Ver la Tabla 7.3. La estrategia de incluír variables rezagadas no produce el resultado deseado en este caso particular.

Como se anotó, la prueba DW se afecta por el termino $\phi_1 Y_{t-1}$, y el estadístico se sesga hacia 2, es decir, la prueba tiende a aceptar $H_0$. La prueba tiene baja potencia para detectar la hipótesis alternativa $H_1$. En este caso se modifica el estadístico DW a

$$h = \left(1 - \frac{d}{2}\sqrt{\frac{T}{1 - T S^2_{\phi_1}}}\right) \tag{7.17}$$

## 7.5. Pruebas de Heterocedasticidad

**Ejemplo 7.5.1.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral.*

*Al estimar los residuos estructurales $\widehat{\varepsilon}_t$ es evidente que muestran heterocedasticidad, como se aprecia a partir de la Figura 7.9.*

Tabla 7.3: Resultados Prueba DWG para el modelo con rezagos 1,2

| | auto correlacion | dw | Valor p |
|---|---|---|---|
| 1 | 0.02 | 1.95 | 0.75 |
| 2 | 0.08 | 1.83 | 0.30 |
| 3 | -0.08 | 2.11 | 0.41 |
| 4 | 0.38 | 1.20 | 0.00 |
| 5 | -0.07 | 2.09 | 0.42 |
| 6 | -0.24 | 2.40 | 0.00 |
| 7 | -0.18 | 2.20 | 0.11 |
| 8 | 0.08 | 1.68 | 0.08 |
| 9 | -0.12 | 2.08 | 0.26 |
| 10 | -0.20 | 2.21 | 0.04 |

*Se pueden aplicar las pruebas de heterocedasticidad, Breusch-Pagan, Prueba de Scores para Varianza no constante y Goldfeld-Quandt. Los resultados se muestran a continuación.*

```
-----------------------------------------------
studentized Breusch-Pagan test
BP = 19.9866, df = 4, p-value = 0.0005025

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 23.03938    Df = 1     p = 1.587171e-06

Goldfeld-Quandt test
GQ = 6.8213, df1 = 69, df2 = 68, p-value = 7.005e-14
-----------------------------------------------
```

Sin embargo, la estrategia consiste en buscar un modelo tipo ARMA(p,q) para los residuos estructurales y no una transformación como la que se sugeriría a partir de los resultados de estas pruebas. Por lo que la verificación de la homocedasticidad de la varianza de los residuos estructurales no se incluirá en las pruebas de validación de los residuos.

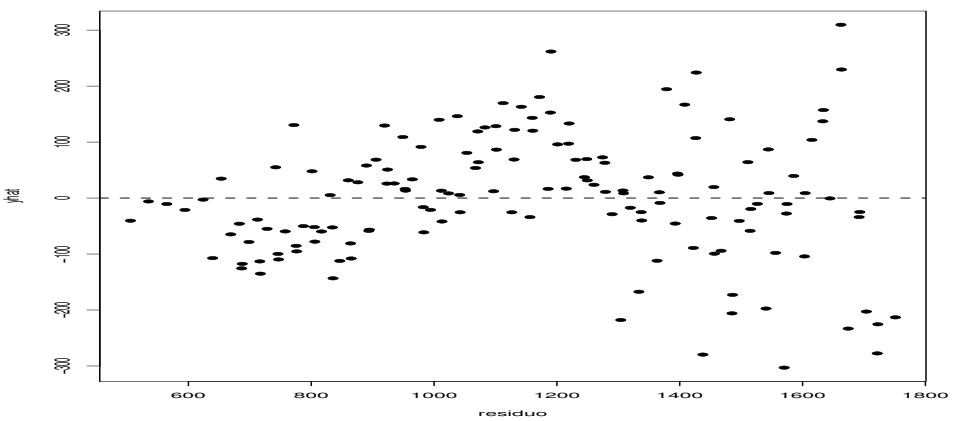



Figura 7.9: Y estimada versus residuales estructurales

# CAPÍTULO 8

# Modelos ARMA para la Componente Aleatoria

## 8.1. Introducción

En los modelos de descomposición $Y_t = T_t + S_t + \varepsilon_t,\ t = 1, 2, \ldots$ se estima $\widehat{\varepsilon}_t$ y se determina si es o nó ruido blanco mediante las pruebas Ljung-Box y Durbin-Watson. En caso de encontrar que $\widehat{\varepsilon}_t$ no es ruido blanco, el siguiente paso es modelar esta componente mediante cuatro posibles modelos.

1. Medias Móviles de orden $q$, $MA(q)$.
2. Autoregresivos de orden $q$, $AR(p)$.
3. Autoregresivos de Medias Móviles, $ARMA(p, q)$.
4. Autoregresivos de Medias Móviles Estacionales, $SARMA(p, q)(p_s, q_s)$.

Con relación a los tres primeros modelos Diebold [1999](ver pag. 129) anota lo siguiente : "Los tres modelos varían en su capacidad de capturar distintos tipos de comportamiento de autoregresión. Comenzaremos dando las características de las funciones de autocorrelación y las cantidades relacionadas con cada modelo; éstas no tiene nada que ver con los datos ni con la estimación, pero son fundamentales para desarrollar una comprensión básica de las propiedades de los modelos, necesarias para llevar a cabo pronósticos inteligentes".

Anotamos que Diebold [1999] no incluye el caso de Autoregresivos de Medias Móviles Estacionales, $SARMA(p,q)(p_s,q_s)$.

**Definición 8.1.1.** *El Operador de Rezago se denota por $L$ (lag, en inglés) y es tal que $L(Y_t) = Y_{t-1},\ t \in \mathbb{Z}$. Es decir, $L$ opera sobre una serie rezagándola un período hacia atrás. De igual manera $L(Y_{t-1}) = Y_{t-2}$, luego $L(L(Y_t)) = L^2(Y_t) = Y_{t-2}$ y en general $L^p(Y_t) = Y_{t-p}$. Se define también $L^0 = I$, el operador identidad.*

Un polinomio de grado $p$ en el operador $L$ se define como el operador formado por una combinación lineal de potencias de $L$. Por ejemplo, para constantes $\beta_j, j = 0, 1, \ldots, p$ si se considera el polinomio en $L$

$$B_p(L) = \beta_0 I + \beta_1 L + \beta_2 L^2 + \cdots + \beta_p L^p, \tag{8.1}$$

entonces, al aplicarlo a $Y_t$ se obtiene

$$\begin{aligned} B_p(L)(Y_t) &= (\beta_0 I + \beta_1 L + \beta_2 L^2 + \cdots + \beta_p L^p) Y_t, \\ &= \sum_{j=0}^{p} \beta_j L^j Y_t, \\ &= \sum_{j=0}^{p} \beta_j Y_{t-j}, \\ &= \beta_0 Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p}. \end{aligned}$$

## 8.2. Procesos de Medias Móviles de orden q, MA(q)

**Definición 8.2.1 (Proceso MA(q)).** *Se dice que una serie $Y_t, t \in \mathbb{Z}$ sigue un proceso $MA(q),\ q = 1, 2, \ldots$ de media móvil de orden q, si se cumple que*

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad t \in \mathbb{Z}, \tag{8.2}$$

*para constantes $\theta_1, \ldots, \theta_q$ y $\varepsilon_t \sim RB(0, \sigma^2)$. La expresión para (8.2) con el operador $L$ es $Y_t = \Theta_q(L)(\varepsilon_t)$, donde*

$$\Theta_q(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{8.3}$$

*se denomina el polinomio de media móvil.*

**Ejemplo 8.2.1.** *Defina $Y_t \sim MA(2)$ dado por*

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},$$

$$\theta_1 = -0.4,\ \theta_2 = 0.4,$$
$$\varepsilon_t \sim iid\, N(0, 9), \quad t \in \mathbb{Z}.$$

**Propiedades de los procesos MA(q)**

Suponga que $Y_t$ cumple (8.2). Entonces

1. $\mathbb{E}(Y_t) \equiv 0$. Se cumple inmediatamente pues $\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t) + \theta_1\mathbb{E}(\varepsilon_{t-1}) + \cdots + \mathbb{E}(\theta_q\varepsilon_{t-q}) = 0$, ya que $\varepsilon_t$ es ruido blanco. Se puede definir procesos MA$(q)$ con media $\mu$ diferente de cero colocando $Y_t = \mu + \Theta_q(L)(\varepsilon_t)$.

2. $Var(Y_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$. Debido a que $\varepsilon_t$ es incorrelacionado se tiene que

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

   Nótese que $Var(Y_t) > Var(\varepsilon_t) = \sigma^2$, en general.

3. $Cov(Y_t, Y_{t+k}) = R(k)$, donde

$$R(k) = \begin{cases} \sigma^2 \displaystyle\sum_{j=0}^{q-k} \theta_j\theta_{j+k}, & 0 \le k \le q \\ 0, & k \ge q+1 \end{cases} \tag{8.4}$$

   con $\theta_0 = 1$. La ecuación (8.4) se puede interpretar como una indicación de que un $MA(q)$ es un proceso débilmente correlacionado, ya que su autocovarianza es cero a partir de un cierto rezago. Por esta razón se pueden ver los procesos MA(q) como alternativas al Ruido Blanco, que es completamente incorrelacionado.

4. Un $MA(q)$ siempre es un proceso estacionario con fac, $\rho(k) = \frac{R(k)}{R(0)},\ k \in \mathbb{Z}$.

**Ejemplo 8.2.2.** *Con el proceso $MA(2)$ del Ejemplo 8.2.1*

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \qquad \varepsilon_t \sim iid\, N(0, 9), \quad t \in \mathbb{Z},$$

*con*

$$\theta_1 = -0.4,\ \theta_2 = 0.4,\ \sigma_2 = 9,$$

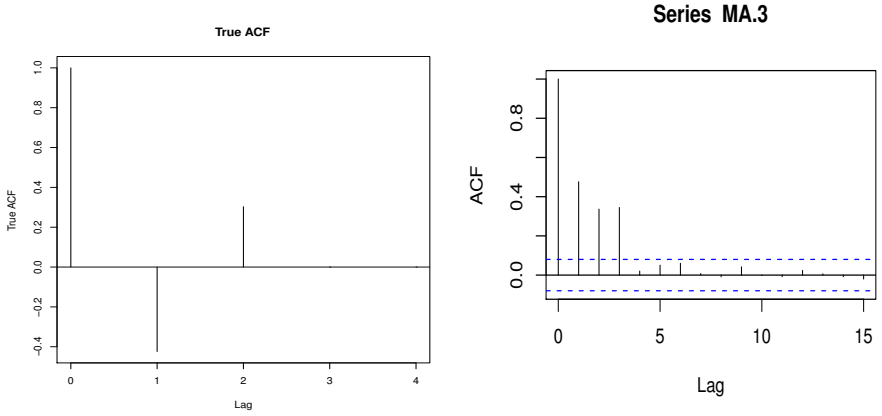
*entonces*

$$R(0) = \left(1 + 0.4^2 + 0.4^2\right)9 = 11.88$$

$$
\begin{aligned}
R(1) &= 9\sum_{j=0}^{2-1}\theta_j\theta_{j+1} = 9(\theta_0\theta_1 + \theta_1\theta_2) \\
&= 9\big(-0.4 + (-0.4)(0.4)\big) = -5.04 \\
R(2) &= 9\sum_{j=0}^{2-2}\theta_j\theta_{j+2} = 9(\theta_0\theta_2) = 9(0.4) = 3.6.
\end{aligned}
$$

*Entonces la fac es*

$$
\begin{aligned}
&\rho(0) = 1, \quad \rho(1) = -\frac{5.04}{11.88} = -0.42, \quad \rho(2) = \frac{3.6}{11.88} = 0.30 \\
&\rho(3) = \rho(4) = \cdots = 0
\end{aligned}
$$




(a) Función de Autocorrelación del Ejemplo 8.2.1

(b) fac muestral de un $MA(3)$

Figura 8.1: Fac teóricas de modelos $MA(q)$

**Conclusión.** De acuerdo con (8.4), si la fac muestral de una serie $Y_t$ termina abruptamente puede tratarse de un $MA(q)$. Por ejemplo, en la Figura 8.1(b) sería factible un modelo MA(3).

**Definición 8.2.2 (Función de Autocorrelación Parcial (facp)).** *Suponga que* $(Y_t,\ t \in \mathbb{Z})$ *es estacionaria. La facp es una función de* $k$, $\alpha(k),\ k = 1, 2, \ldots$ *definida por*

1. $\alpha(1) = \rho(1)$

2. $\alpha(k) = Corr(\varepsilon_1, \varepsilon_k)$ *donde*

$$\varepsilon_1 = Y_1 - E(Y_1|Y_2, \ldots, Y_{k-1})$$
$$\varepsilon_k = Y_k - E(Y_k|Y_2, \ldots, Y_{k-1}), \qquad k = 2, \ldots$$

*Y la facp muestral se define por* $\hat{\alpha}(k)$

1. $\hat{\alpha}(1) = \hat{\rho}(1)$
2. $\hat{\alpha}(2)$ : *se regresa* $Y_t$ *sobre* $Y_{t-1}$ *y* $Y_{t-2}$ *tal que* $Y_t = \phi_{21}Y_{t-1} + \phi_{22}Y_{t-2} + \varepsilon_t$ *entonces* $\hat{\alpha}(2) = \hat{\phi}_{22}$
3. $\hat{\alpha}(k)$ : *se regresa* $Y_t$ *sobre* $Y_{t-1}, \ldots, Y_{t-k}$ *tal que* $Y_t = \phi_{k1}Y_{t-1} + \cdots + \phi_{kk}Y_{t-k} + \varepsilon_t$ *entonces* $\hat{\alpha}(k) = \hat{\phi}_{kk}$

La facp de un proceso $Y_t \sim MA(q)$ se puede encontrar si se asume la condición de invertibilidad para un $MA(q)$

## Condición de Invertibilidad del Proceso MA(q)

Los procesos MA(q) pueden presentar un problema de falta de identificabilidad. Puede ocurrir que existan dos conjuntos de parámetros $\underline{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_q^{(i)})$, para $i = 1, 2$, con $\underline{\theta}^{(1)} \neq \underline{\theta}^{(2)}$ y sin embargo se cumple que las respectivas autocovarianzas (8.4) son iguales, $R(k, \underline{\theta}^{(1)}) \equiv R(k, \underline{\theta}^{(2)})$. Si se cumple la condición siguiente, "de invertibilidad", no ocurre este problema.

**Definición 8.2.3.** *Dado un proceso MA(q),* $Y_t = \theta_q(L)(\varepsilon_t)$ *donde*

$$\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q,$$

*si las q raíces de* $\Theta_q(z) = 0$, $z_1, z_2, \ldots, z_q \in \mathbb{C}$, *cumplen* $|z_j| > 1$, *se dice que el proceso* $Y_t$ *es invertible.*

Es decir, si $\Theta_q(z)$ cumple que "sus raíces están fuera del círculo unitario", lo cual es equivalente a la condición:

$$\Theta_q(z) \neq 0, \ \forall z \in \mathbb{C}, \ |z| \leq 1. \tag{8.5}$$

**Ejemplo 8.2.3.** *Consideremos el proceso* $Y_t \sim MA(2)$ *en el Ejemplo 8.2.1,*

$$Y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.4\varepsilon_{t-2},$$

*y veamos si es invertible. Hallamos las raíces del polinomio* $\theta_q(z)$

$$\theta_2(z) = 1 - 0.4z + 0.4z^2 = 0,$$
$$z = \frac{0.4 \pm \sqrt{0.4^2 - 4(0.4)(1)}}{2(0.4)} = \frac{1}{2} \pm \frac{3}{2}\mathrm{i}$$

*por tanto*

$$|z| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2} = \sqrt{\frac{1}{4} + 9} \; > \; 1,$$

*luego* $Y_t$ *es invertible.*

## Función fac parcial de un Proceso MA(q) invertible

Suponga que se cumple la condición de invertibilidad (8.5) para el polinomio

$$\Theta_q(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$$

Entonces la función $1/\Theta_q(z)$ se puede desarrollar en una serie de Taylor alrededor de $z = 0$, de la forma

$$\Theta_q(z)^{-1} = \sum_{j=0}^{\infty} \psi_j z^j, \tag{8.6}$$

para $|z| < 1$ y tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\psi_0 = 1$.

Si se reemplaza $z$ por $L$ se tiene un operador en el rezago $L$ de la forma $\Theta_q(L)^{-1} = \sum_{j=0}^{\infty} \psi_j L^j$, el cual, aplicado a $Y_t$ produce

$$\Theta_q(L)^{-1} Y_t = \sum_{j=0}^{\infty} \psi_j L^j(Y_t) = \sum_{j=0}^{\infty} \psi_j Y_{t-j}.$$

Ahora, multiplicando ambos miembros de $Y_t = \Theta_q(L)(\varepsilon_t)$ por $\Theta_q(L)^{-1}$ se obtiene

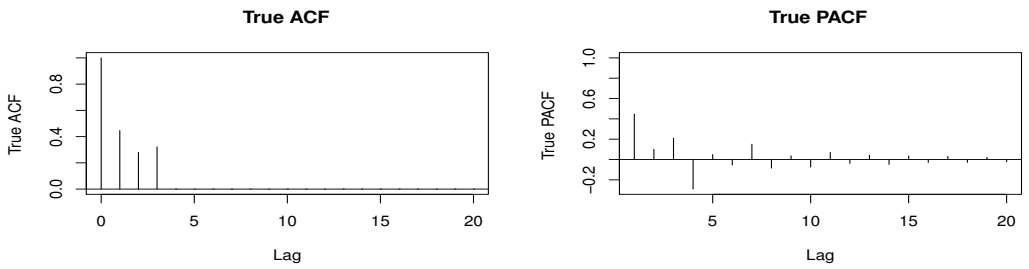
$$\Theta_q(L)^{-1} Y_t = \Theta_q(L)^{-1} \Theta_q(L) \varepsilon_t = \varepsilon_t.$$

Por tanto, $\sum_{j=0}^{\infty} \psi_j Y_{t-j} = \varepsilon_t$. Como $\psi_0 = 1$ se puede despejar $Y_t$ de la identidad anterior y se obtiene

$$Y_t = -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \cdots + \varepsilon_t, \tag{8.7}$$

de donde concluímos que si se hace la regresión de $Y_t$ sobre los primeros $k$ rezagos $Y_{t-j}, j = 1, \ldots, k$, entonces el k-ésimo coeficiente es $\alpha(k) = \psi(k)$ y como $\psi(k) \to 0$ porque la serie $\sum_{k=0}^{\infty} \psi_k$ es convergente, entonces $\alpha(k) \to 0$ cuando $k \to \infty$. Por tanto, la facp de un MA(q) decrece a cero.

**Ejemplo 8.2.4.** *En las Figuras siguientes 8.2 se observa la fac y la facp de un MA(3).*




(a) $fac$ (b) $facp$

Figura 8.2: fac y facp de un $MA(3)$.

**Estimación de procesos MA(q)**

La estimación de procesos $Y_t \sim MA(q)$ no es directa. Una de las razones es que en $Y_t = \theta_q(L)(\varepsilon_t)$ los valores del ruido blanco $\varepsilon_t$ no son observables. Supongamos una muestra de tamaño $T$ del proceso: $Y_1, \ldots, Y_T$.

En el caso de estimadores de mínimos cuadrados ordinarios ver (2.5), pag.22, el estimador de $\underline{\theta} = (\theta_1, \ldots, \theta_q)'$ se define como el vector $\hat{\underline{\theta}}$ que minimiza la función

$$G(\underline{\theta}) = \sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} \big(Y_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\big)^2, \tag{8.8}$$

Pero los $\varepsilon_{t-j}$ no son observables. Se puede recurrir a la invertibilidad, y utilizar la serie en (8.7) pero truncada hasta un cierto $n$. Permitiendo seguir con el signo igual, se escribe

$$Y_t = -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \cdots - \psi_n Y_{t-n} + \varepsilon_t. \tag{8.9}$$

Considerando la ecuación

$$\Theta_q(L)^{-1}\Theta_q(L) = (1 + \psi_1 z + \psi_2 z^2 + \ldots)(1 + \theta_1 + \ldots + \theta_q) = 1$$

se pueden despejar los $\psi_j$ en función del vector $\underline{\theta} = (\theta_1, \ldots, \theta_q)$. Escribiendo $\psi_j = \psi_j(\underline{\theta})$, la minimización se hace con base en (8.9) con la función objetivo

$$G(\underline{\theta}) = \sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} \big(Y_t + \sum_{j=1}^{n} \psi_j(\underline{\theta}) Y_{t-j}\big)^2. \tag{8.10}$$

El correspondiente estimador $\underline{\hat{\theta}}$ se denomina estimador de mínimos cuadrados condicionales. Es asintóticamente eficiente y tiene las mismas propiedades asintóticas del estimador de máxima verosimilitud. El estimador de máxima verosimilitud se define en la sección siguiente.

En R hay varias funciones para estimar procesos MA(q). Entre éstas la función `arma` de la librería `tseries` y la función `arima`. La primera calcula los valores p de los coeficientes, y la segunda permite calcular pronósticos. Ambas utilizan estimadores de mínimos cuadrados condicionales y de máxima verosimilitud. El siguiente código muestra cómo estimar y pronosticar un proceso MA(q).

**Ejemplo 8.2.5.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t.$$

*Al estimar los residuos estructurales $\widehat{\varepsilon}_t$ se observó que posiblemente muestran una dinámica autocorrelacionada. Se confirmó este hecho con las pruebas Ljung-Box y Durbin-Watson. En este Ejemplo se plantea investigar si los residuos podrían seguir una dinámica MA(q). En el código R siguiente se estima secuencialmente varios modelos MA(4), MA(5), etc. hasta encontrar un modelo con parámetros significativos, que presente el menor AIC.*

*La posibilidad de encontrar un MA(q) óptimo (aunque el orden q puede ser alto) está garantizada por una propiedad que dice que dado un proceso estacionario en covarianza que cumple cierta condición técnica (su densidad espectral deber ser continua en $[-\pi, \pi]$), entonces siempre existirá un MA(q) que lo aproxime con un grado deseado de error (ver Brockwell and Davis [2002], pag. 132, Corollary 4.4.1). En la Figura 8.3 se muestran pronósticos con el modelo estructural y el modelo conjunto: estructural con residuos MA(6), que es el MA(q) óptimo.*

```
#---------analisis de residuos estructurales con MA(q)
library(car)
library(forecast)
library(lmtest)
archivo = "cementq.dat"
E = read.table(archivo, header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
#----modelo con tendencia lineal y estacionalidad
```

```
#----con variables indicadoras estacionales
T = length(y)
t = seq(1,T)
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)
r = mod1$residuals
```

Estimar varios MA(q=4,5,6) para los residuos con dos funciones

```
#----------
require(tseries)

mod2.1 = arma(r,order = c(0,4))
summary(mod2)
mod2.2 = arima(r,order = c(0,0,4))
summary(mod3)

#-----
mod3.1 = arma(r,order = c(0,5))
summary(mod3.1)
mod3.2 = arima(r,order = c(0,0,5))
summary(mod3.2)

#-----modelo escogido
mod4.1 = arma(r,order = c(0,6))
summary(mod4.1)
mod4.2 = arima(r,order = c(0,0,6))
summary(mod4.2)
#-----
r4.2 = residuals(mod4.2)
```

Examinar residuos del MA(6)

```
par(mfrow=c(3,2))
plot(t,r4.2,type='o',ylab='residuo')
abline(h=0,lty=2)
```

```
plot(density(r4.2),xlab='x',main= '')
acf(r4.2,60,ci.type="ma",main="")
pacf(r4.2,60,main="")
qqnorm(r4.2)
qqline(r4.2,col=2)
# grafica residuos
cpgram(r4.2)

#------- pruebas Ljung-Box
Box.test(r4.2, lag = 6, type = "Ljung-Box")
Box.test(r4.2, lag = 16, type = "Ljung-Box")
```

Calcular pronosticos estructurales + MA(6)

```
m = 6
tp = seq(T+1,T+m)
Itp = seasonaldummyf(y,m)

y.p = predict(mod1,data.frame(t=tp,It=I(Itp)))
r.p = predict(mod4.2,n.ahead=m)$pred
#------- pronosticos conjuntos
yr.p = y.p+r.p


#------- graficar
m0 = 40
N = seq(T-m0,T)
par(mfrow=c(1,1))
plot(seq(T-m0,T),y[N],type='b',
ylim=c(1200,2200),xlim=c(T-m0,T+m),
ylab='produccion trim. tons.',
xlab='trimestre')
lines(seq(T+1,T+m),yr.p,lty=7,lwd=2,col='red')
lines(seq(T+1,T+m),y.p,lty=4,lwd=2,col='blue')

legend("topleft",
legend=c("observada","E+MA(6)","Estructural"),
col=c('black','red','blue'),lty=c(1,7,4))
```

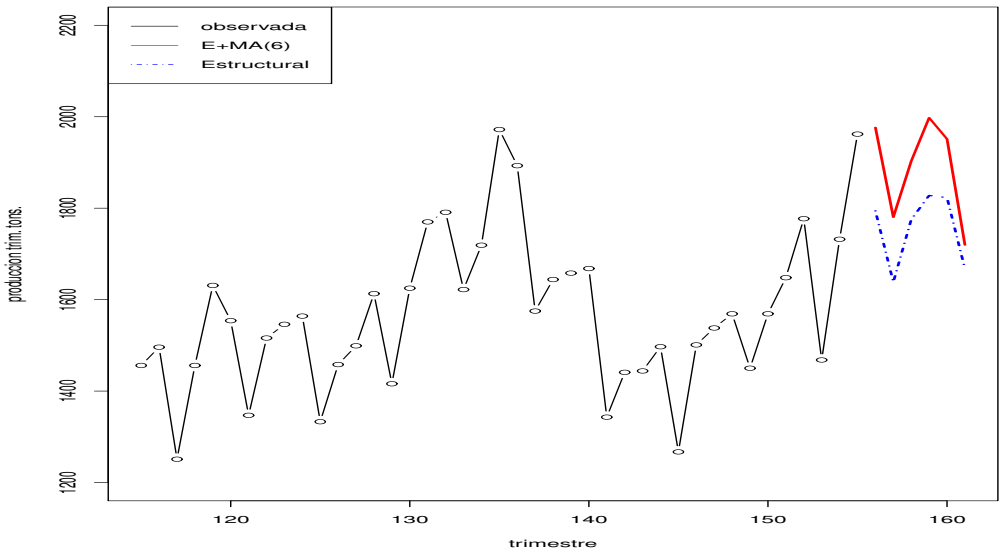



Figura 8.3: Pronosticos Estructurales versus Estructurales + MA(6)

**Simulación de procesos MA(q) con R**

La simulación de un proceso MA(q) se puede hacer con la función `filter()`, descrita en la sección 6.1, pag.86, ya que el proceso es una media móvil unilateral de un ruido blanco. Si se denota por `et` el vector de ruido blanco con varianza $\sigma^2$ y se indica por `theta` un vector con los parámetros $\theta_1, \ldots, \theta_q$, se simulan $n$ valores de $Y_t \sim MA(q)$ ejecutando el comando `y=filter(et,theta,"conv",sides=1,circular=TRUE)`. Produce el mismo resultado que la función `arima.sim()`, cuando se escribe `y=arima.sim(list=(order(0,0,q),ma=theta),n=n,sd=sigma)`.

## 8.3. Procesos Autorregresivos de Orden p, AR(p)

**Definición 8.3.1 (Proceso AR(p)).** *Se dice que un proceso* $(Y_t, t \in \mathbb{Z})$ *sigue un modelo AR(p) de media cero si es la solución de la ecuación recursiva siguiente.*

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \tag{8.11}$$

*donde* $\varepsilon_t \sim RB(0, \sigma^2)$, $\varphi_j, j = 1, 2, \ldots, p$ *son los coeficientes (parámetros) y* $p = 0, 1, 2, \ldots$ *es el orden.*

El caso $p = 0$ se interpreta como $Y_t = \varepsilon_t$. Se define

$$\Phi_p(z) = 1 - \sum_{j=1}^{p} \varphi_j z^j \tag{8.12}$$

el polinomio autorregresivo. Usando el operador de rezago $L$ se puede escribir (8.11) como

$$\Phi_p(L)(Y_t) = \varepsilon_t.$$

Un proceso $Y_t \sim AR(p)$ con media diferente de cero, $\mathbb{E}(Y_t) = \mu \neq 0$, se define mediante la ecuación

$$\Phi_p(L)(Y_t - \mu) = \varepsilon_t. \tag{8.13}$$

Obsérvese que

$$\begin{aligned}
\Phi_p(L)(Y_t - \mu) &= \Phi_p(L)(Y_t) - \Phi_p(L)(\mu) \\
&= \Phi_p(L)(Y_t) - \mu(1 - \sum_{j=1}^{p} \varphi_j), \\
&= \varepsilon_t.
\end{aligned}$$

Luego, $\Phi_p(L)(Y_t) = \varphi_0 + \varepsilon_t$, con $\varphi_0 = \mu(1 - \sum_{j=1}^{p} \varphi_j)$. La constante $\varphi_0$ se denomina intercepto.

## Condición Suficiente para que un AR(p) sea Estacionario en Covarianza

**Proposición 8.3.1.** *Una condición suficiente para que $Y_t \sim AR(p)$ sea estacionario en covarianza es que las p raíces de la ecuación $\varphi_p(z) = 0$, $z_i,\ i = 1, 2 \ldots, p$, cumplan*

$$|z_i| > 1. \tag{8.14}$$

Es decir, si $\varphi_p(z)$ cumple que "sus raíces están fuera del círculo unitario", lo cual es equivalente a la condición:

$$\varphi_p(z) \neq 0,\ \forall z \in \mathbb{C},\ |z| \leq 1. \tag{8.15}$$

En la Figura 8.4 se observa el círculo unitario junto con una la raíz $z_j$ y su conjugado $\bar{z}_j$, ambas fuera del círculo.

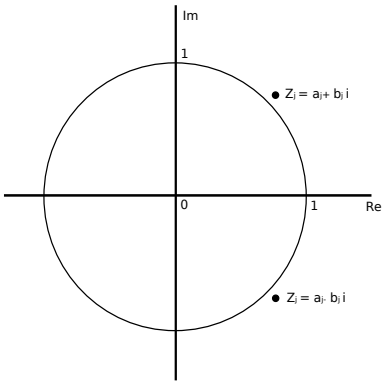



Figura 8.4: Círculo Unitario

Sin embargo, la condición (8.15) no es necesaria. Puede suceder que algunas de las raíces de $\varphi_p(z) = 0$ estén adentro y otras afuera del círculo unitario, pero si se cumple que $\varphi_p(z) \neq 0,\ \forall z \in \mathbb{C},\ |z| = 1$, en este caso $Y_t$ es estacionario en covarianza.

**Ejemplo 8.3.1.** *El proceso* $Y_t \sim AR(2)$*,*

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t, \tag{8.16}$$

*cumple que es estacionario en covarianza ya que las raíces de su polinomio autorregresivo* $\Phi_2(z) = 1 - 1.5z + 0.9z^2 = 0$ *están fuera del círculo unitario. Se puede hacer la comprobación con las funciones* `InvertibleQ()` *de la librería* `FitAR`*, y* `armaRoots()` *de la librería* `fArma`*, como sigue.*

```
phi = c(1.5,-0.9)
InvertibleQ(phi)
[1] TRUE
armaRoots(phi)
      re      im   dist
1 0.8333  0.6455 1.0541
2 0.8333 -0.6455 1.0541
```

## Algunas Propiedades de los Procesos AR(p) Estacionarios

Para un proceso $Y_t \sim AR(p)$ de media cero, que cumple (8.15), se cumple:

$$\mathbb{E}(Y_t) = 0, \tag{8.17}$$

$$\sum_{j=1}^{p} \varphi_j < 1. \tag{8.18}$$

*Demostración.* Para comprobar (8.17), si $Y_t$ es estacionario en covarianza entonces $\mathbb{E}(Y_t) = \mu$. Además,

$$\mathbb{E}(Y_t) = \varphi_1 E(Y_{t-1}) + \varphi_2 E(Y_{t-2}) + \cdots + \varphi_p E(Y_{t-p}) + 0,$$

pero todas las esperanzas son iguales a $\mu$, luego

$$\mu = \varphi_1 \mu + \varphi_2 \mu + \cdots + \varphi_p \mu.$$

Si $\mu \neq 0$ entonces $1 = \varphi_1 + \cdots + \varphi_p$, por tanto el polinomio autorregresivo evaluado en $z = 1$ se anula, es decir, $\varphi_p(1) = 0$, lo cual es una contradicción ya que $\forall\, z \in \mathbb{C},\ |z| \leq 1$ $\varphi_p(z) \neq 0$. Luego debe tenerse que $\mu = 0$, es decir, el proceso definido en (8.11) es de media cero.

Para comprobar (8.18), si se asume que $\sum_{j=1}^{p} \varphi_j \geq 1$, el caso $z = 1$ implica nuevamente que $\varphi_p(1) = 0$, lo cual es una contradicción con la condición (8.15). Si se asume $\sum_{j=1}^{p} \varphi_j > 1$ entonces $\varphi_p(1) < 0$. Pero $\varphi_p(0) = 1$. Como la función $\varphi_p(x)$ es continua en $[0, 1]$ debe existir un punto $0 < x < 1$ en el cual $\varphi_p(x) = 0$. Pero eso es nuevamente una contradicción porque este $x$ es una raíz de $\varphi_p(z) = 0$ que cumple $|z| < 1$, lo cual no puede ocurrir. □

## La Función de Autocovarianza de un Proceso AR(p)

La función de autocovarianza de un proceso $Y_t \sim AR(p)$, $R(k) = Cov(Y_t, Y_{t+k})$, se puede calcular resolviendo una ecuación recursiva lineal denominada la ecuación de Yule–Walker.

**Proposición 8.3.2.** *Suponga un proceso AR(p),*

$$Y_t = \varphi_0 + \sum_{j=1}^{p} \varphi_j Y_{t-j} + \varepsilon_t,$$

*que satisface la condición* (8.14)*. Entonces su función de autocovarianza $R(k)$ satisface la ecuación recursiva*

$$R(k) = \sum_{j=1}^{p} \varphi_j R(k-j) + \sigma^2 \delta_{k,0},\ k = 0, 1, 2, \ldots. \tag{8.19}$$

*denominada ecuación de Yule–Walker, donde $\delta_{k,0} = 0$ si $k \neq 0$, y $\delta_{0,0} = 1$ (la función $\delta_{k,0}$ se denomina delta de Kronecker).*

*Demostración.* Con $\mu = \mathbb{E}(Y_t)$, la ecuación $Y_t = \varphi_0 + \sum_{j=1}^{p} \varphi_j Y_{t-j} + \varepsilon_t$ equivale a $Y_t - \mu = \sum_{j=1}^{p} \varphi_j (Y_{t-j} - \mu) + \varepsilon_t$. Multiplicando ambos miembros por $Y_{t-k} - \mu$, con $1 \leq k \leq t$, y tomando valor esperado $\mathbb{E}(.)$ se obtiene

$$\begin{aligned} R(k) &= \mathbb{E}((Y_t - \mu)(Y_{t-k} - \mu)) \\ &= \sum_{j=1}^{p} \varphi \mathbb{E}((Y_{t-j} - \mu)(Y_{t-k} - \mu)) + \mathbb{E}(\varepsilon_t (Y_{t-k} - \mu)) \\ &= \sum_{j=1}^{p} \varphi_j R(k-j). \end{aligned}$$

Ya que se cumple $\mathbb{E}(\varepsilon_t (Y_{t-k} - \mu)) = 0$ porque, a partir de la definición del proceso $Y_t$ en (8.11), $Y_{t-k}$ depende de $\varepsilon_s$ con $s \leq t - k$, que son variables incorrelacionadas con $\varepsilon_t$. El caso $k = 0$ es similar, con $\mathbb{E}(\varepsilon_t (Y_t - \mu)) = Var(\varepsilon) = \sigma^2$. □

A partir de la ecuación Yule-Walker (8.19) es inmediato, al dividir estas ecuaciones por $R(0)$, que la fac $\rho(k)$, cumple una ecuación similar dada por

$$\rho(k) = \varphi_1 \rho(k-1) + \varphi_2 \rho(k-2) + \cdots + \varphi_p \rho(k-p), \quad k = 1, 2, \ldots, \tag{8.20}$$

La ecuación (8.20) es una ecuación recursiva lineal de orden $p$, homogénea, para $\rho(k), k \geq 0$, con $p$ condiciones iniciales $\rho(0) = 1, \rho(1), \ldots, \rho(p-1)$, las cuales es necesario especificar para resolver la ecuación. Existe una fórmula general para expresar la solución de la ecuación, pero, utilizando las condiciones iniciales y la recursión, se puede calcular su valor para un rezago $k$ particular.

**Ejemplo 8.3.2.** *Retomando el Ejemplo 8.3.1, para el proceso* $Y_t \sim AR(2)$,

$$Y_t = 1.5 Y_{t-1} - 0.9 Y_{t-2} + \varepsilon_t,$$

*estacionario en covarianza, su fac* $\rho(k)$ *satisface la ecuación recursiva lineal*

$$\rho(k) = 1.5\rho(k-1) - 0.9\rho(k-2), \quad k \geq 2. \tag{8.21}$$

*Los casos* $k = 0, 1$ *en la ecuación anterior permiten obtener los valores iniciales*

$$\rho(0) = 1, \ \rho(1) = \frac{\varphi_1}{1 - \varphi_2} = \frac{15}{19}.$$

*Por ejemplo,*

$$\rho(2) = 1.5\rho(1) - 0.9\rho(0) = 1.5 \left( \frac{15}{19} \right) - 0.9.$$

*Por otro lado, se puede demostrar que la solución de (8.21) es de la forma*

$$\rho(k) = \frac{8\sqrt{51}}{57}\left(\frac{3}{\sqrt{10}}\right)^k cos\left(\arctan\left(\frac{\sqrt{15}}{5}\right)k - \arctan\left(\frac{\sqrt{15}}{57}\right)\right).$$

*Nótese que en este caso la fac muestra una semi-periodicidad, como se vé en la Figura 8.5, de período*

$$\frac{2\pi}{\arctan\left(\frac{\sqrt{15}}{5}\right)} = 9.5,$$

*amortiguada exponencialmente, tendiendo a cero.*

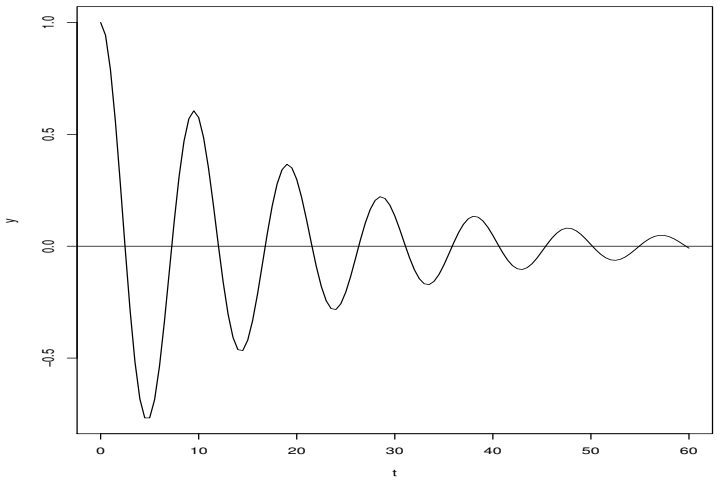



Figura 8.5: fac de ar(2)

La función de autocovarianza, $R_X(k)$, se pueden calcular en el lenguaje R con la función `ARMAacf()` siguiente

```
phi = c(1.5,-0.9)
Acf = ARMAacf(ar=phi,ma=numeric(0),lag.max=30)
 str(Acf)
 Named num [1:31] 1 0.789 0.284 -0.284 -0.682 ...
 - attr(*, "names")= chr [1:31] "0" "1" "2" "3" ...
```

**Función fac parcial de un Proceso AR(p)**

El razonamiento para determinar la facp de un AR(p) es directo. La FACP de un proceso $AR(p)$ es $\alpha(k)$ tal que $\widehat{\alpha}(k)$ es el coeficiente $\widehat{\beta}_{k,k}$ en la regresión

$$Y_t = \beta_0 + \beta_{k,1}Y_{t-1} + \cdots + \beta_{k,k}Y_{t-k} + \eta_t, \quad k = 2 \tag{8.22}$$

pero como $\beta_{k,k} = 0$ si $k \geq p+1$ entonces $\widehat{\alpha}(k) = 0$ si $k \geq p+1$. La fac parcial de un AR(p) es cero para rezagos $k \geq p+1$.

**Ejemplo 8.3.3.** *Retomando el Ejemplo 8.3.1, para el proceso* $Y_t \sim AR(2)$,

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t,$$

*estacionario en covarianza, la función de autocovarianza parcial,* $\alpha_X(k)$*, se pueden calcular en R con la función siguiente*

```
phi = c(1.5,-0.9)
PAcf = ARMAacf(ar=phi,
ma=numeric(0),lag.max=30,pacf=TRUE)
str(PAcf)
 num [1:30] 7.89e-01 -9.00e-01 3.10e-15 -8.67e-16 -9.40e-16 ...
```

**Simulación de procesos AR(p)**

La simulación de un proceso AR(p) se puede hacer mediante la función `filter()`, escribiendo `y=filter(et,phi,"recursive",init=y0)`, donde `phi=c(f1,...,fp)` es el vector de coeficientes $\varphi_j$ en la ecuación $Y_t = \sum_{j=1}^{p} \varphi_j Y_{t-j} + \varepsilon_t$, y $(\varepsilon_t) \sim RB(0, \sigma^2)$. El vector `y0=c(y1,...,yp)` consiste de los primeros $p$ valores necesarios para iniciar la recursión AR(p).

**Ejemplo 8.3.4.** *En el caso del Ejemplo 8.3.1 la instrucción es*

```
et = rnorm(1000,0,2)
y=filter(et,c(1.5,-0.9),"recursive",init=y0)
```

*donde* `y0=c(0,0)`*, si se toman los valores iniciales como la media del proceso, que es cero.*

La simulación de un proceso AR(p) de media `mu` diferente de cero, se realiza colocando `y=mu+filter(et,phi,"recursive",init=y0)`

**Estimación de procesos AR(p)**

La estimación de un proceso $Y_t \sim AR(p)$ es directa y existen varios métodos para realizarla. Si se tiene una muestra del proceso, $Y_1, \ldots, Y_T$, se pueden utilizar:

1. los estimadores Yule-Walker (con base en la fac muestral),
2. estimadores por mínimos cuadrados ordinarios,
3. estimadores de máxima verosimilitud, asumiendo que los errores cumple $(\varepsilon_t) \sim iidN(0, \sigma^2)$,
4. estimadores por el método de Burg.

En el caso de estimadores de mínimos cuadrados ordinarios, ver (2.5), pag.22, el estimador de $\underline{\varphi} = (\varphi_1, \ldots, \varphi_p)'$ se define como el vector $\widehat{\underline{\varphi}}$ que minimiza la función

$$G(\underline{\varphi}) = \sum_{t=1+p}^{T} \varepsilon_t^2 = \sum_{t=1+p}^{T} \big(Y_t - \sum_{j=1}^{p} \varphi_j Y_{t-j}\big)^2. \tag{8.23}$$

Un problema que surge con estos estimadores es que el vector estimado $\widehat{\underline{\varphi}}$ puede no cumplir la condición (8.15), para que la solución $Y_t$ sea estacionaria en covarianza. El método de Burg permite corregir esta posibilidad.

Los estimadores Yule-Walker se basan en la siguiente identidad. La fac de un $AR(p)$, $\rho(k) = Corr(Y_t, Y_{t+k}), \quad k = 1, 2, \ldots, p, p+1, \ldots$, cumple lo siguiente.

Dadas la matriz $A$ y los vectores $\underline{\varphi}, \underline{\rho}$

$$A = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(2) & \cdots & \rho(p-2) \\ \rho(2) & \rho(3) & \cdots & \rho(p-3) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{bmatrix}, \ \underline{\varphi} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}, \ \underline{\rho} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix},$$

entonces es válido el sistema lineal $p \times p$

$$A\underline{\varphi} = \underline{\rho}. \tag{8.24}$$

Luego, dados los estimadores de la fac $\widehat{\rho}(1), \ldots, \widehat{\rho}(p)$ se pueden reemplazar en la expresiones para $A$ y $\underline{\rho}$, y resolver (8.24) para obtener

$$\widehat{\underline{\varphi}} = \widehat{A}^{-1}\widehat{\underline{\rho}}. \tag{8.25}$$

Estos estadísticos $\underline{\widehat{\varphi}}$ así calculados se denominan los estimadores de Yule-Walker de $\underline{\varphi}$.

Los métodos de máxima verosimilitud y Burg se pueden consultar en Brockwell and Davis [2002] y Grandell [2014]

### Estimación de un proceso AR(p) en R

La estimación de un proceso AR(p) en R se puede hacer mediante varias funciones en R. Las función `ar()` de la librería `stats` provee la estimación por los métodos: mínimos cuadrados, Yule-Walker, máxima verosimilitud y Burg.

```
ar(x, aic = TRUE, order.max = NULL,
   method = c("yule-walker", "burg", "ols", "mle", "yw")
aic (del help ?ar)
Logical flag. If TRUE then the Akaike Information
Criterion is used to choose the order of the
autoregressive model. If FALSE, the model of order
order.max is fitted.
```

**Ejemplo 8.3.5.** *Retomando el Ejemplo 8.3.1, para el proceso* $Y_t \sim AR(2)$,

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t,$$

*estacionario en covarianza, los resultados de una estimación con base en una muestra simulada, utilizando* `ar()` *con las opciones "ols", "burg", "yw" y "mle" se muestran a continuación.*

| | ols | burg | yw | mle |
|---|---|---|---|---|
| 1 | 1.48 | 1.48 | 1.48 | 1.50 |
| 2 | -0.88 | -0.88 | -0.88 | -0.90 |

Tabla 8.1: Estimadores OLS, Burg, Y-W, MaxVer en un AR(2)

La función `FitAR()` de la librería `FitAR` utilizan un algoritmo para máxima verosimilitud. Por ejemplo, las instrucciones siguientes determinan un orden $p$ óptimo mediante el mínimo AIC, considerando todas las posibilidades hasta un máximo, en el ejemplo de 30.

```
p = SelectModel(y,lag.max=32, Best=1)
m2 = FitAR(y, p)
coef(m2)
```

La función `arima()` de `stats` utiliza máxima verosimilitud. Un inconveniente con esta función es que reporta la media con el nombre de intercepto. Así, podría generar confusión ya que intercepto se refiere a $\varphi_0$.

**Ejemplo 8.3.6.** *La serie de aportes hidrológicos diarios en Colombia muestra la cantidad de agua convertida en energía, en todas las represas para generación de electricidad en el país. La unidad es Gw/hora. Para los datos entre Julio 1 de 1998 y 31 de Mayo de 2014, la gráfica de los aportes, $Y_t$ está en la Figura 8.6. Inicialmente se propone un modelo de componentes lineal y trigonométricas, de la forma*

$$Y_t = b_0 + b_1 t + \sum_{j=1}^{k} \alpha_j \cos(2\pi jt/365) + \beta_j \sin(2\pi jt/365) + \varepsilon_t, \tag{8.26}$$

*donde $(\varepsilon_t)$ es una serie con autocorrelacionada, que posiblemente puede ser AR(p). Los análisis en el programa R siguiente muestran que este supuesto es correcto.*

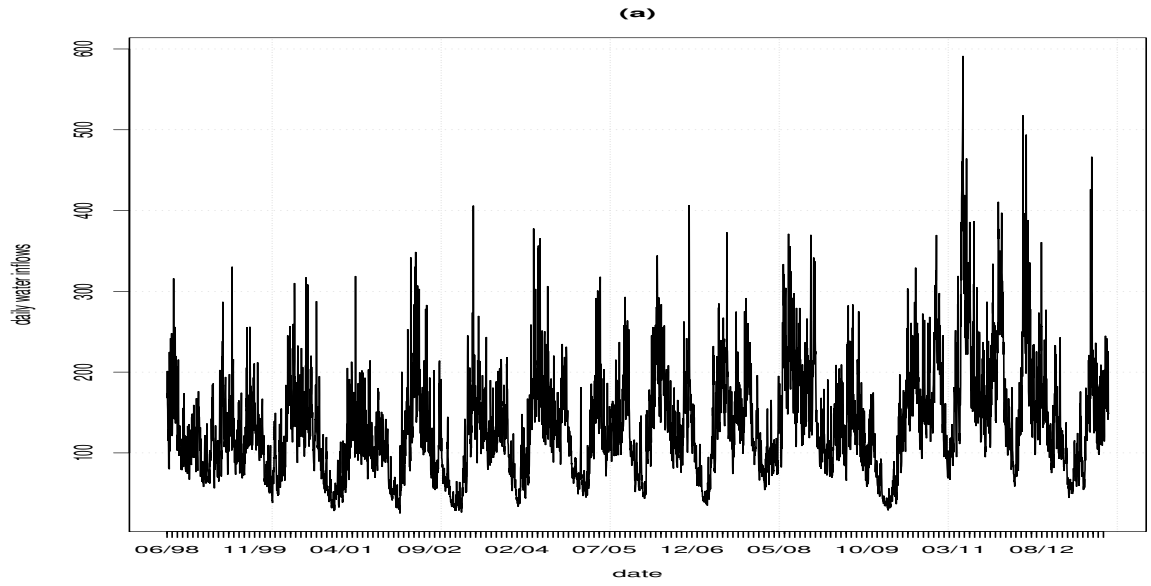



Figura 8.6: Aportes hidrológicos en Gw/h, 01/07/98 - 31/05/14

```
# Ejemplo de analisis AR(p)
# Serie de aportes hidrológicos diarios
# en Gw/h
#-----------------modela estacionalidad con sen, cos
#---------------- serie con frecuencia diaria
```

```
Ap = ts(Ap,frequency = 365)
require(forecast)
It.trig = fourier(Ap,4)
t = seq(1,length(Ap))/100
mod2 = lm(Ap ~ t + It.trig)
summary(mod2)
```

Los resultados son aceptables, aunque el $R^2$ está bajo.

```
Call: lm(formula = Ap  t + It.trig)
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   104.80070    1.34523  77.906  < 2e-16 ***
t               1.00769    0.04182  24.094  < 2e-16 ***
It.trigS1-365  32.45044    0.95059  34.137  < 2e-16 ***
It.trigC1-365  24.35062    0.95159  25.589  < 2e-16 ***
It.trigS2-365 -11.17645    0.95123 -11.749  < 2e-16 ***
It.trigC2-365  31.80005    0.95038  33.460  < 2e-16 ***
It.trigS3-365   0.43385    0.95037   0.457    0.648
It.trigC3-365  -3.70841    0.95124  -3.899 9.79e-05 ***
It.trigS4-365  -7.41607    0.95115  -7.797 7.52e-15 ***
It.trigC4-365   4.54505    0.95022   4.783 1.77e-06 ***
---
Multiple R-squared:  0.4021,    Adjusted R-squared:  0.4012
F-statistic: 415.6 on 9 and 5561 DF,  p-value: < 2.2e-16
```

Identificar los residuos...fac y fac parcial

```
#-----------------examen fac y fac parcial residuos
r = residuals(mod2)
require(TSA)
par(mfrow=c(1,2))
TSA::acf(r,60, drop.lag.0 = TRUE,ci.type="ma")
pacf(r,60)
```

Se puede concluír que los residuos estructurales pueden ser modelados mediante un AR(9). El proceso de análisis continúa estimando este modelo, examinando sus residuos y finalmente, si tales residuos son ruido blanco, calcular pronósticos, por ejemplo, a 30 días.

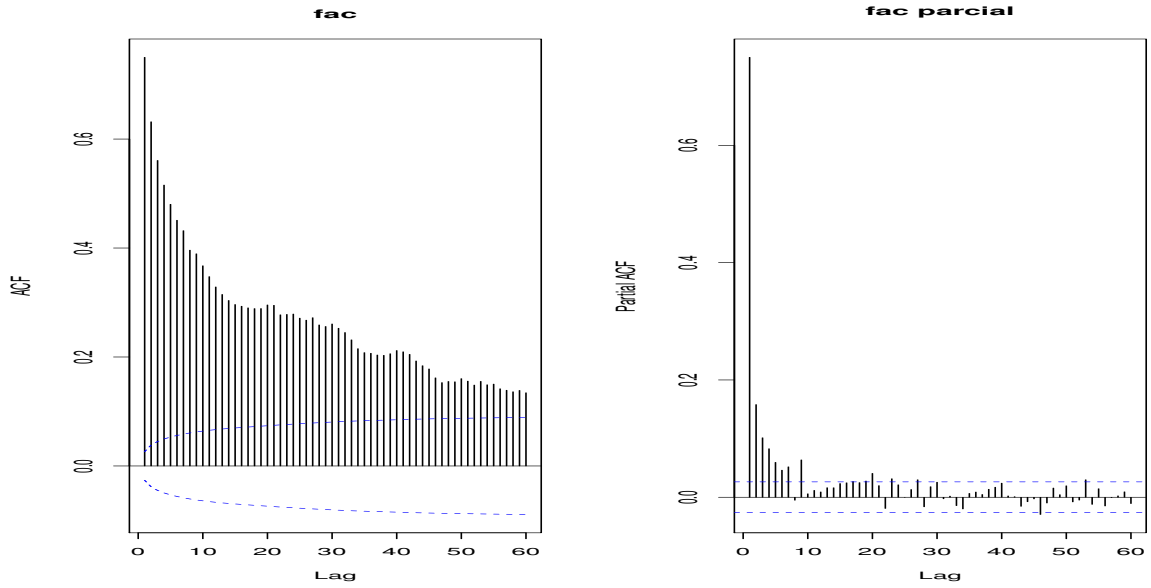



Figura 8.7: fac y facp de Aportes hidrológicos

## 8.4. Procesos Autoregresivos y de Medias Móviles ARMA(p,q)

Una manera de introducir los procesos ARMA(p,q) es la siguiente. Si en un proceso $Y_t \sim AR(p)$

$$Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \cdots - \varphi_p Y_{t-p} = \varepsilon_t, \qquad \varepsilon_t \sim R.B.(0, \sigma^2), \tag{8.27}$$

se reemplaza $\varepsilon_t$ por un proceso $MA(q)$, $Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$ entonces (8.27) queda

$$Y_t - \varphi_1 Y_{t-1} - \cdots - \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{8.28}$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$. El efecto de este cambio es que los errores no se toman incorrelacionados sino débilmente autocorrelacionados. Se define entonces un proceso ARMA(p,q) como un modelo que combina las propiedades de memoria larga de los AR(p) con las propiedades de ruido débilmente autocorrelacionado en los MA(q), y que tiene suficiente flexibilidad y parsimonia para representar una variedad grande de procesos estacionarios en covarianza. Usando la notación del operador de rezago (8.28) se puede definir el proceso $ARMA(p, q)$ por

**Definición 8.4.1.** *Un proceso $Y_t \sim ARMA(p, q)$ de media cero, se define como la solución de la ecuación recursiva (8.29)*

$$\Phi_p(L)(Y_t) = \Theta_q(L)(\varepsilon_t), \qquad t \in \mathbb{Z}, \tag{8.29}$$

*donde* $\varepsilon_t \sim RB(0, \sigma^2)$, *y* $\Phi_p(z) = 1 - \sum_{j=1}^{p} \varphi_j z^j$, $\Theta_q(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$ *son los polinomios autoregresivo y de media móvil respectivamente.*

Las condiciones de estacionariedad de la parte AR(p) y de invertibilidad de la parte MA(q) se asumen en el modelo ARMA(p,q), (8.29). Por lo tanto, se asume que las raíces de las ecuaciones $\Phi_p(z) = 0$ y $\Theta_q(z) = 0$ están fuera del círculo unitario. Además se asume que estos polinomios no tienen raíces en común. Si se cumplen estas condiciones el proceso $Y_t \sim ARMA(p, q)$ es estacionario e identificable. Un ARMA(p,q) de media diferente de cero se define por la ecuación $\Phi_p(L)(Y_t) = \varphi_0 + \Theta_q(L)(\varepsilon_t)$ donde $\varphi_0 = \mu(1 - \sum_{j=1}^{p} \varphi_j)$.

**Ejemplo 8.4.1.** *Suponga un proceso* $Y_t \sim ARMA(1, 1)$ *de media cero, dado por*

$$\Phi_1(L)(Y_t) = \Theta_1(L)(\varepsilon_t) \tag{8.30}$$

*donde* $\Phi_1(L) = 1 - \varphi L$ *y* $\Theta_1(L) = 1 + \theta L$. *Es decir* $Y_t = \varphi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. *Si* $|\varphi| < 1$ *y* $|\theta| < 1$ *es estacionario e invertible. Por ejemplo*

$$Y_t = 0.9Y_{t-1} + \varepsilon_t - 0.4\varepsilon_{t-1},$$

*con* $\varepsilon_t \sim RB(0, \sigma^2)$. *Si se asume* $E(Y_t) = \mu = 2$ *entonces el proceso queda*

$$\begin{aligned} Y_t &= (1 - \varphi)\mu + \varphi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \\ &= 2(1 - 0.9) + 0.9Y_{t-1} + \varepsilon_t - 0.4\varepsilon_{t-1}. \end{aligned}$$

**Ejemplo 8.4.2.** *Consideremos un modelo de descomposición con tendencia lineal y estacionalidad de período 12, modelada por variables indicadoras donde el residuo estructural es un proceso ARMA(1,1). Entonces el modelo se escribe como el sistema de dos ecuaciones siguiente.*

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \sum_{j=1}^{11} \delta_j I_j(t) + \varepsilon_t, \\ \varepsilon_t &= \varphi\varepsilon_{t-1} + a_t + \theta a_{t-1}, \end{aligned}$$

*con* $a_t \sim RB(0, \sigma^2)$, *el residuo del proceso ARMA(1,1). Nótese que el número de parámetros de este modelo es 16, incluyendo la varianza* $\sigma^2$.

**Ejemplo 8.4.3.** *Considere el proceso* $Y_t \sim ARMA(2, 1)$ *dado por*

$$Y_t = 2 + \frac{1 - 0.4L}{1 - 1.5L + 0.9L^2}\varepsilon_t, \qquad \varepsilon_t \sim R.B.(0, \sigma^2)$$

*que equivale a*

$$Y_t = 2(1 - 1.5 + 0.9) + 1.5Y_{t-1} - 0.9Y_{t-3} + \varepsilon_t - 0.4\varepsilon_{t-1}$$

*con ecuación característica*

$$1 - 1.5z + 0.9z^2 = 0$$

*y sus raíces dadas por*

$$z = \frac{1.5 \pm \sqrt{1.5^2 - 4(0.9)}}{2(0.9)} = 0.83 \pm 0.645\text{i}$$

*luego*

$$|z| = 1.05 > 1$$

*por tanto $Y_t$ es un proceso estacionario en covarianza. También es invertible.*

**Representación causal de los procesos ARMA**

Suponga que $Y_t \sim ARMA(p, q)$ y que las raíces de las ecuaciones $\Phi_p(z) = 0$ y $\Theta_q(z) = 0$ están fuera del círculo unitario entonces se puede aplicar la Proposición **??** y definir, para $|z| < 1$, la función de transferencia como el desarrollo en serie de potencias

$$\Psi(z) := \frac{\Theta_q(z)}{\Phi_p(z)} = \sum_{j=0}^{\infty} \psi_j z^j,$$

con $\psi_0 = 1$ y $\sum_{j=0}^{\infty} |\psi_j| < \infty$. A partir de $\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t$, se puede escribir

$$Y_t = \Psi(L)(\varepsilon_t) = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots, \tag{8.31}$$

La identidad (8.31) se denomina la representación causal de un ARMA(p,q). Y la capacidad de los procesos ARMA(p,q) de representar una amplia variedad de procesos estacionarios en covarianza se basa en que tales procesos satisfacen una identidad del tipo (8.31). Entonces, dada la función de transferencia del proceso, $\Psi(z)$, siempre será posible encontrar una función racional, es decir, un cociente de polinomios, de la forma $\Theta_q(z)/\Phi_p(z)$ que la aproxime bien en el sentido

$$|\Psi(z) - \Theta_q(z)/\Phi_p(z)| \leq C|z|^{p+q+1}, \tag{8.32}$$

para $C > 0$ una constante. Este resultado se conoce como "aproximación de Padé". En la práctica no se conoce la función $\Psi(z)$, pero se espera que, al encontrar un posible modelo ARMA(p,q), se cumpla la aproximación (8.32). Entonces el problema básico inicial es encontrar los órdenes p y q adecuados.

Nótese que de la representación causal se sigue inmediatamente que $E(Y_t) = 0$ y $Var(Y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$. Además, de la teoría de series $\sum_{j=0}^{\infty} |\psi_j| < \infty \Rightarrow \sum_{j=0}^{\infty} \psi_j^2 < \infty$.

**La función de autocovarianza de un proceso ARMA**

La función de autocovarianza de un proceso $Y_t \sim$ ARMA(p,q) estacionario de media cero. Si se indica por $R(k) = Cov(Y_t, Y_{t+k})$ su función de autocovarianza, para $k = 0, 1, \ldots$ un método para calcular esta función se basa en la representación causal (8.31). Multiplicando ambos miembros por $Y_{t-k}$ y tomando esperanza $\mathbb{E}(.)$ se obtienen las ecuaciones recursivas siguientes, similares a las ecuaciones Yule-Walker para AR(p), (8.19). Defina $n = max(p, q+1)$,

$$R(k) - \varphi_1 R(k-1) - \ldots - \varphi_p R(k-p) = \begin{cases} \sigma^2 \sum_{j=k}^{q} \theta_j \psi_{j-k}, & \text{si } k = 0, 1, \ldots, n-1 \\ 0, & \text{si } k = n, n+1, \ldots. \end{cases} \tag{8.33}$$

**Ejemplo 8.4.4.** *(tomado de Brockwell and Davis [2002], pag. 93). Considere el proceso ARMA(2,1) dado por* $(1 - L + \frac{1}{4}L^2)Y_t = (1 + L)\varepsilon_t$. *Entonces* $n = max(p, q+1) = 2$, *por tanto, para* $k = 0, 1$ *se tiene el sistema lineal*

$$\begin{aligned} R(0) - R(1) + \frac{1}{4}R(2) &= \sigma^2(\psi_0\theta_0 + \psi_1\theta_1) = \sigma^2(1 + \psi_1), \\ R(1) - R(0) + \frac{1}{4}R(1) &= \sigma^2(\theta_1\psi_0) = \sigma^2\psi_0 = \sigma^2. \end{aligned} \tag{8.34}$$

*Para calcular los coeficientes* $(\psi_j, j = 0, 1, \ldots)$ *se utiliza la función* `ARMAtoMA(a,b,m)`, *la cual calcula los coeficientes* $\psi_j, j = 1, 2, \ldots, m$, *dados los vectores* $a = (\varphi_1, \ldots, \varphi_p)$ *y* $b = (\theta_1, \ldots, \theta_q)$. *Entonces, escribiendo* `ARMAtoMA(c(1.0, -0.25), 1.0, 10)`, *se obtiene el vector*

```
[1] 2.00000000 1.75000000 1.25000000 0.81250000 0.50000000
[6] 0.29687500 0.17187500 0.09765625 0.05468750 0.03027344
```

*de donde* $\psi_1 = 2$. *Usando la segunda ecuación de (8.33), se obtiene* $R(2) = R(1) - \frac{1}{4}R(0)$, *luego, reemplazando en el sistema (8.34), y resolviendo, se obtienen* $R(0) = 32\sigma^2/3$, $R(1) = 28\sigma^2/3$. *Utilizando la segunda ecuación de (8.33),*

$$R(k) = R(k-1) - \frac{1}{4}R(k-2),\ k = 2, 3, \ldots$$

*se puede calcular la autocovarianza recursivamente.*

La varianza $R(0) = Var(Y_t)$ se calcula a partir de la primera ecuación en (8.33), la cual define un sistema lineal, una de cuyas incógnitas es $R(0)$, que se resuelve en función de $\sigma^2$.

**Estimación de modelos ARMA**

La estimación de procesos $Y_t \sim$ ARMA(p,q) se basa en un supuesto: que el vector $\underline{Y} = (Y_1, \ldots, Y_n)'$ se distribuye Normal multivariado con media $\underline{\mu}$, y matriz de covarianzas $\Sigma = [Cov(Y_i, Y_j)]_{n\times n}$. Como el proceso se asume estacionario, se cumple $Cov(Y_i, Y_j) = R(j-i)$, donde $R(k)$ es la función de autocovarianza de $Y_t$. La forma de $\Sigma$ es la de una matriz tipo Toeplitz: las diagonales descendentes son constantes:

$$\Sigma = \begin{bmatrix} R(0) & R(1) & \cdots & R(n-1) \\ R(1) & R(0) & \cdots & R(n-2) \\ R(2) & R(1) & \cdots & R(n-3) \\ \vdots & \vdots & & \vdots \\ R(n-1) & R(n-2) & \cdots & R(0) \end{bmatrix}. \tag{8.35}$$

Por ejemplo, para $Y_t$ un AR(p), $R(k)$ se calcula mediante las ecuaciones Yule-Walker, (8.19), $R(k) = \mu + \sum_{j=1}^{p} \varphi_j R(k-j)$. Por tanto, colocando $\underline{\beta} = (\mu, \sigma^2, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q)'$, la matriz $\Sigma$ depende del vector $\underline{\beta}$, y se escribe $\Sigma(\underline{\beta})$. Este supuesto permite implementar la estimación por máxima verosimilitud. Se escribe la densidad Normal Multivariada como

$$f(\underline{y}, \underline{\beta}) = \frac{1}{(2\pi)^{n/2}\sqrt{det(\Sigma(\underline{\beta}))}} \exp\left(-\frac{1}{2}(\underline{y}-\underline{\mu})\Sigma(\underline{\beta})^{-1}(\underline{y}-\underline{\mu})'\right) \tag{8.36}$$

donde $\underline{\mu} = (\mu, \ldots, \mu)' \in \mathbb{R}^n$. La función de log-verosimilitud se define a partir del logaritmo de la densidad (8.36), $\log(f(\underline{y}, \underline{\beta}))$, y está dada por

$$\begin{aligned} L(\underline{\beta}) &:= \sum_{j=1}^{n} \log f(\underline{y_j}, \underline{\beta}) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log det(\Sigma(\underline{\beta})) - \frac{1}{2}(\underline{y}-\underline{\mu})'\Sigma(\underline{\beta})^{-1}(\underline{y}-\underline{\mu}). \end{aligned} \tag{8.37}$$

El estimador ML de máxima verosimilitud de $\underline{\beta}$ se define como

$$\widehat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}}(-L(\underline{\beta})). \tag{8.38}$$

**Identificación de modelos ARMA**

No es fácil identificar los modelos $ARMA(p,q)$ mediante la FAC y FACP. Si $q \geq p \geq 1$ entonces para $k \geq q+1$ se cumple (8.33) y para $1 \leq k \leq p-1$, $\rho(k)$ no tiene un patrón general, luego la FAC muestral presenta un patrón definido solamente para $k \geq q+1$.

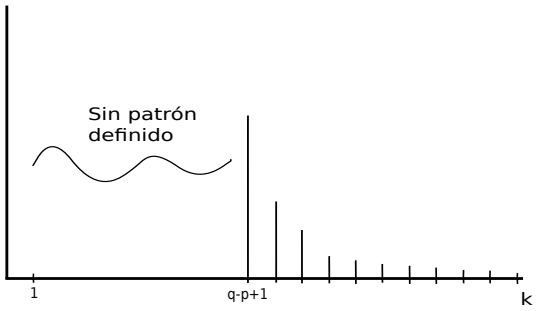



Figura 8.8: Fac Muestral de $ARMA(p, q)$.

Una alternativa consiste en buscar una pareja de órdenes $(p, q)$ dentro de un rango inicial, por ejemplo $p, q = 0, 1, \ldots, 10$, que minimice alguna función de $\hat{\varepsilon}^2$, como por ejemplo el AIC ó el BIC.

La identificación con base en el AIC se basa en comparar varios modelos $ARMA(p, q)$ para valores de, por ejemplo, $p = 0, 1, \ldots, 20$ y $p = 0, 1, \ldots, 20$ con base en el criterio $AIC$ el cual está definido por

$$AIC = -2L(\widehat{\underline{\beta}}) + 2k, \qquad k = p + q \tag{8.39}$$

Entonces identifica un posible modelo parsimonioso escogiendo el modelo con el mínimo $AIC$. El código siguiente en R muestra una programación básica con base en la función `arima()`.

```
#------------programa para escoger un posible
#             modelo ARMA(p,q),con p,q=0,1,2,3
nll = matrix(nrow=4,ncol=4,
dimnames=list(paste("p=",0:3,sep=""),
paste("q=",0:3,sep="")))
aic = nll
lb = nll
for (p in (0:3)) {
for (q in (0:3)) {
ARMAlabel = sprintf("ARMA(%d,%d)",p,q)
armamodel = arima(y,order=c(p, 0, q),include.mean = TRUE)
lbtest = Box.test(na.omit(armamodel$resid), lag = 12,type="Ljung")
nll[p+1,q+1] = armamodel$loglik;
aic[p+1,q+1] = armamodel$aic;
```

```
lb[p+1,q+1] = lbtest$p.value}}
cat("LogLik:\n")
print(nll)
cat("AIC:\n")
print(aic)
cat("Box-Ljung-Test:\n")
print(lb)
min(aic)
```

**Librerías en R para identificación, estimación y pronósticos de modelos ARMA**

El plan de análisis con procesos ARMA consiste en los siguientes pasos, con algunas librerías R para implementarlos.

1. Identificar el modelo ARMA(p,q). La función `auto.arima()` de la librería `forecast`, y la función `autoarmafit()` de la liberería `timsac`; el programa asume que el modelo ARMA(p,q) se define con $\theta_q(z) = 1 - \sum_{j=1}^{q} \theta_j z^j$, es decir, los coeficientes $\theta_j$ cambian de signo en esta librería con respecto a la convención utilizada en estas Notas. También tiene la función `armafit()` para ajuste de modelos ARMA.

2. Estimar el modelo. La función `arima()` estima primero por mínimos cuadrados condicionales y luego por máxima verosimilitud. La función `arma()`, de `tseries`, estima mediante mínimos cuadrados condicionales. La función `FitARMA()`, de `FitAR` para estimación de ARMA(p,q).

3. Chequeo del ajuste del modelo a los datos. La función `arma()`,

4. Pronósticos con el modelo. O simulación del modelo.

**Ejemplo 8.4.5.** *Suponga que se propone simular un $ARMA(3,2)$ tal que se escogen las raíces de $\varphi_3(z)$ y $\theta_2(z)$, con módulos fuera del círculo unitario. Las tres raíces son $z1[1], z1[2], z1[3]$ y los coeficientes del polinomio autorregresivo de grado 3, mónico, $\Phi_3(z)$,*

$$\begin{aligned}\Phi_3(z) &= (1 - z/z1[1])(1 - z/z1[2])(1 - z/z1[3]) \\ &= 1 - \varphi_1 z - \varphi_2 z^2 - \varphi_3 z^3 \\ &= 1 - 0.1466381z - 0.1430615z^2 - 0.3576538z^3\end{aligned}$$

*se calculan en el código siguiente en el vector a. El código incluye además, la identificación mediante la función* `auto.arima`*, la estimación con* `arima` *y el cálculo de pronósticos con* `predict`.

```
#simulación, identificación, estimación y predicción de un ARMA(3,2)
#----------
library(polynom)
z1 = complex(3)
z1[1] = -0.8 - 1.3i
z1[2] = Conj(z1[1])
z1[3] = 1.2

a = poly.calc(z1)
a = a/a[1]
#-----------raices del polinomio de media movil de grado 2
z2 = complex(2)
z2[1] = -1.2 -0.5i
z2[2] = Conj(z2[1])
#-------------- los coeficientes estan en el vector b
b = poly.calc(z2)
(b = b/b[1])
#------------- usar la funcion arima.sim con los coeficientes a y b.
n = 300
y = arima.sim(list(order=c(3,0,2), ar=-a[2:4], ma=b[2:3]), n=n,
sd=sqrt(0.3))
#--------------graficar
plot.ts(y)
#-------------- verificar el modelo
require(forecast)
auto.arima(y)
#-------------- auto.arima identifica un arma(3,2)
#-------------- para su estimacion se usa la instruccion
mod1 = arima(y1, c(3,0,2))
#-------------- los pronosticos con este arma(2,3)

py1 = predict(mod1, n.ahead=20)

b = c(y[(300-90+1):300],py1$pred)
plot(1:110,b, type='b', col=2)
points(91:110,py1$pred, type='b', col='blue')
points(91:110, py1$pred+1.64*py1$se, type='l', col='blue')
points(91:110, py1$pred-1.64*py1$se, type='l', col='blue')
```

## 8.5. Modelos ARMA estacionales, SARMA

Los modelos ARMA estacionales estacionarios se introducen con el fin de modelar estacionalidades remanentes en los residuos estructurales. Por tanto, dependen del parámetro $s$ definido como el período de la serie. Hasta el momento no se ha introducido un procedimiento para detectar y estimar el período en una componente estacional. Pero para el caso del residuo estructural $\epsilon_t$ sería útil contar con algunos. Se pueden mencionar, por ejemplo

1. La función `periodicity` de la librería `xts` (ver Ryan and Ulrich [2017]). De la página de ayuda: "Estimate the periodicity of a time-series-like object by calculating the median time between observations in days ".

```
et = resid(modelo)
et = ts(et, frequency=4)
periodicity(et)
```

2. Un programa de aplicación por R. Hyndman ([1])

> If you really have no idea what the periodicity is, probably the best approach is to find the frequency corresponding to the maximum of the spectral density. However, the spectrum at low frequencies will be affected by trend, so you need to detrend the series first. The following R function should do the job for most series... It will return 1 for data that have no strong periodicity, and the length of period otherwise

```
find.freq <- function(x)
{
    n <- length(x)
    spec <- spec.ar(c(x),plot=FALSE)
    if(max(spec$spec)>10) # Arbitrary threshold chosen by trial and error.
    {
        period <- round(1/spec$freq[which.max(spec$spec)])
        if(period==Inf) # Find next local maximum
        {
            j <- which(diff(spec$spec)>0)
            if(length(j)>0)
            {
```

[1] https://stats.stackexchange.com/questions/1207/period-detection-of-a-generic-time-series

```
                nextmax <- j[1] + which.max(spec$spec[j[1]:500])
                period <- round(1/spec$freq[nextmax])
            }
            else
                period <- 1
        }
    }
    else
        period <- 1
    return(period)
}
```

**Definición 8.5.1.** *Se dice que $Y_t$ sigue un modelo ARMA Estacional de período $s > 1$, ó S-ARMA, denotado por $Y_t \sim ARMA(p,q)(p_s,q_s)[s]$, donde $p, q, p_s, q_s$ son enteros no negativos, si para $\epsilon_t \sim RB(0,\sigma^2)$*

$$\Phi_{p_s}(L^s)\Phi_p(L)Y_t = \Theta_{q_s}(L^s)\Theta_q(L)\epsilon_t, \tag{8.40}$$

*donde*

$$\Phi_p(L) = 1 - \sum_{j=1}^{p} \varphi_j L^j, \tag{8.41a}$$

$$\Theta_q(L) = 1 + \sum_{j=1}^{q} \theta_j L^j, \tag{8.41b}$$

$$\Phi_{p_s}(L^s) = 1 - \sum_{j=1}^{p_s} \varphi_{s,j} L^{js}, \tag{8.41c}$$

$$\Theta_{q_s}(L^s) = 1 + \sum_{j=1}^{q_s} \theta_{s,j} L^{js}. \tag{8.41d}$$

En (8.40) $(p,q)$ son los órdenes AR y MA del modelo ARMA ordinario y $(p_s,q_s)$ los órdenes AR y MA estacionales. Nótese que en (8.41c) y (8.41d) las potencias de $L$ son de la forma $L^s, L^{2s}, \ldots$.

Una interpretación de (8.40) se puede obtener si se coloca

$$e_t^s = \frac{\Theta_{q_s}(L^s)}{\Phi_{p_s}(L^s)} e_t = e_t + \psi_1^s e_{t-s} + \psi_2^{2s} e_{t-2s} + \ldots,$$

luego, $e_t^s$ es un ruido que solamente tiene efectos no nulos cada $s$ períodos, y al colocar

$$Y_t = \frac{\Theta_q(L)}{\Phi_p(L)} e_t^s,$$

se obtiene un efecto de autocorrelación en $Y_t$ cada $s$ períodos que se asemeja a una estacionalidad de período $s$, y sirve para modelarla.

**Ejemplo 8.5.1.** *El modelo* $Y_t \sim ARMA(1,1)(1,2)[4]$ *corresponde a*

$$(1-\varphi_1 L)(1-\varphi_{4,1}L^4)Y_t = (1+\theta_1 L)(1+\theta_{4,1}L^4+\theta_{4,2}L^8)\epsilon_t, \tag{8.42}$$

*Y el modelo* $Y_t \sim ARMA(1,2)(1,1)[12]$ *corresponde a*

$$(1-\varphi_1 L)(1-\varphi_{12,1}L^{12})Y_t = (1+\theta_1 L+\theta_2 L^2)(1+\theta_{12,1}L^{12})\epsilon_t, \tag{8.43}$$

Nótes que en un S-ARMA, por ejemplo el (8.43), al multiplicar los polinomios autorregresivos de media móvil se obtiene

$$\begin{aligned}(1-\varphi_1 L)(1-\varphi_{12,1}L^{12}) &= 1-\varphi_1 L-\varphi_{12,1}L^{12}+\varphi_1\varphi_{12,1}L^{13}\\ (1+\theta_1 L+\theta_2 L^2)(1+\theta_{12,1}L^{12}) &= 1+\theta_1 L+\theta_2 L^2+\theta_1\theta_{12,1}L^{13}+\theta_2\theta_{12,1}L^{14}\end{aligned}$$

luego (8.43) es un ARMA en el cual hay un coeficiente autorregresivo $\varphi_1\varphi_{12}$ en el rezago $1+12=13$, y también hay un coeficiente de media móvil $\theta_1\theta_{12}$ en el rezago $1+12=13$ y otro $\theta_2\theta_{12,1}$ en el rezago $2+12=14$. Estos podrían verse en la fac y en la fac parcial, además del rezago 14. Esta observación es útil con referencia al siguiente procedimiento de identificación gráfico, en la librería `library(TSA)`, con la función `armasubsets()`. Calcula un panel para identificación para modelos ARMA y SARMA, que detecta estos rezagos. Sin embargo, tiene limitaciones gráficas para rezagos de órdenes altos. En el Ejemplo siguiente se muestra una aplicación e interpretación de los resultados de la función.

**Ejemplo 8.5.2.** *Para identificar un posible modelo SARMA en una serie con tendencia lineal y período* $s=7$ *en los residuos se programa un procedimiento de identificación gráfico con* `armasubsets()`. *Se trata es de escoger los órdenes* $p, q, p_s, q_s$ *con base en un panel que muestra cuáles son los modelos más opcionados con respecto al BIC. La última instrucción grafica un panel en la Figura 8.9.*

```
mod2 = lm(yi ~ t + It)
summary(mod2)
r = mod2$residuals
library(TSA)
res=armasubsets(y=r,nar=14,nma=14,y.name='r',
ar.method='ols')
plot(res)
```

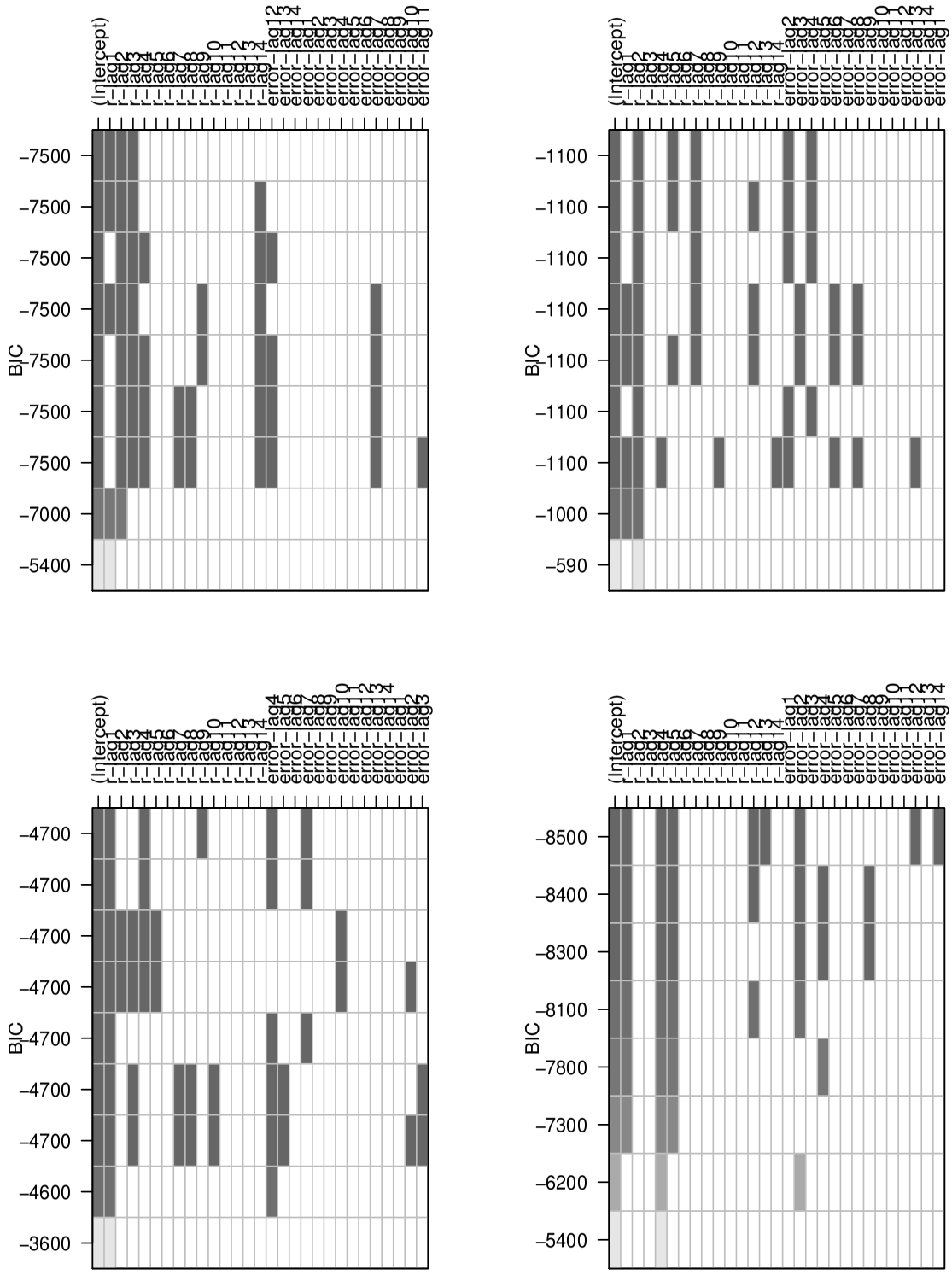



Figura 8.9: Posibles submodelos (1,1)=AR, (1,2),(2,1)=ARMA, (2,2)=SARMA

En la Figura 8.9 aparecen las identificaciones de un AR en la celda (1,1). En las celdas (1,2) y (2,1) aparecen modelos ARMA. Y en la (2,2) un SARMA.

En cada gráfica la información sobre los rezagos a incluír en el modelo está orgarnizada en dos sectores. A la izquierda los autorregresivos, identificada con "r-lag" y a la derecha los de media móvil, "error-lag". Los modelos más opcionados ó factibles son los que incluyen los rezagos con menor BIC, es decir, los cuadros oscuros de la primera línea.

En la gráfica (1,1) aparecen muy evidentes los rezagos autorregresivos hasta p=3. Entonces se puede identificar un AR(3).

En la celda (1,2), arriba-derecha, se observan rezagos autorregresivos p = 5,7, y de media móvil q = 2,4. Entonces un modelo factible es un ARMA(4,4). Nótese que con `auto.arima()` se identifica un ARMA(4,4).

En la celda (2,1), abajo-izquierda, se observan rezagos autorregresivos p = 4. Y se observan rezagos de media móvil q = 4,7. Un posible modelo es un ARMA(4,4). Con `auto.arima()` se identifica un ARMA(4,5).

En la celda (2,2), abajo-derecha, se observa algo que no aparece en las anteriores. Aparecen rezagos autorregresivos 1,12,13. Lo que dá lugar a suponer una estructura $(1-\varphi L)(1-\varphi_{12}L^{12})$. Porque así se justificaría el rezago 13 como el producto $L \times L^{12} = L^{13}$. Entonces una primera decisión es una estructura SARMA $(1,q)(1,q_{12})[12]$. En los rezagos de media móvil se observan rezagos de media móvil 2,12,14. Con lo cual una opción podría ser colocar $(1,2)(1,1)[12]$ porque así se justificaría el rezago 14 que proviene de 2+12 = 14. Es el modelo siguiente.

$$(1-\varphi_1 L)(1-\varphi_{12,1}L^{12})Y_t = (1+\sum_{j=1}^{2}\theta_j L^j)(1+\theta_{12,1}L^{12})\epsilon_t, \tag{8.44}$$

## La estimación de los modelos SARMA

Los modelos SARMA se estiman con la función `arima()`. Por ejemplo, la estimación del modelo (8.44), $Y_t \sim ARMA(1,2)(1,1)[12]$, se hace con los comandos siguientes.

```
mod4 = arima(r,order=c(1,0,2),
seasonal = list(order = c(1, 0, 1), period = 12))
```

Este modelo tiene residuos dados por `r4 = mod4$residuals`. El análisis de incorrelación de éstos se hace de la manera explicada en el curso, en los talleres y en las notas de clase. Igual con los pronósticos.

### Ejemplo de modelo ARMA estacional, S-ARMA

Estos modelos son una alternativa para los residuos en los modelos con componentes estacionales y de tendencia. Por ejemplo, un modelo con tendencia lineal y componente estacional, $S_t$, según el modelo de descomposición podría estar dado por

$$\begin{aligned} Y_t =& \beta_0 + \beta_1 t + S_t + \epsilon_t, \\ \epsilon_t \sim& \text{ S-ARMA}(p,q)(p_s,q_s)[s]. \end{aligned} \tag{8.45}$$

**Ejemplo 8.5.3.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t.$$

*Un esquema de análisis es el siguiente. Tomando en cuenta que anteriormente en el Ejemplo 8.2.5, en la pag. 126, se ajustó un MA(6) a los residuos estimados $\widehat{\varepsilon}_t$. Ahora interesa examinar si se puede tener también $\widehat{\varepsilon}_t \sim$ S-ARMA, y cuál de éstos modelos produce mejores pronósticos.*

*Un modelo posible es SARMA(1,0,2)(1,0,2)[4], con media cero. En la Tabla 8.2 se pueden ver los resultados de la estimación, utilizando la función* `arima()`*. El resultado de la prueba Ljung-Box para incorrelación de los residuos es*
`X-squared = 23.9558, df = 26, p-value = 0.5785`.

*En la Tabla 8.3 se pueden comparar las medidas de calidad de pronósticos para los modelos: Estructural (4.7), Estructural con residuos MA(6) y Estructural con residuos SARMA(1,0,2)(1,0,2)[4]. Este último modelo tiene los pronósticos de menor MAPE.*

Tabla 8.2: Coeficientes estimados del modelo SARMA(1,0,2)(3,0,1)[4]

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| ar1 | 0.6852 | 0.0991 | 6.9143 | 0.0000 |
| ma1 | 0.0234 | 0.1144 | 0.2047 | 0.8378 |
| ma2 | 0.4114 | 0.1097 | 3.7503 | 0.0002 |
| sar1 | 1.1032 | 0.1511 | 7.3023 | 0.0000 |
| sar2 | -0.4470 | 0.1400 | -3.1927 | 0.0014 |
| sar3 | 0.2762 | 0.0896 | 3.0833 | 0.0020 |
| sma1 | -0.5645 | 0.1461 | -3.8638 | 0.0001 |
| intercept | -44.0386 | 93.4644 | -0.4712 | 0.6375 |

Tabla 8.3: Comparación de medidas de calidad de pronósticos

| | ME | RMSE | MAE | MPE | MAPE | ACF1 | Theil's U |
|---|---|---|---|---|---|---|---|
| estructural | -84.3254 | 137.1416 | 122.2615 | -5.7245 | 7.6581 | 0.1503 | 0.6695 |
| E+ma6 | -40.1263 | 97.0538 | 79.2159 | -2.8445 | 4.8387 | -0.0503 | 0.5103 |
| E+sarma | 41.6997 | 96.8725 | 68.5528 | 2.1812 | 3.9090 | 0.0445 | 0.5023 |

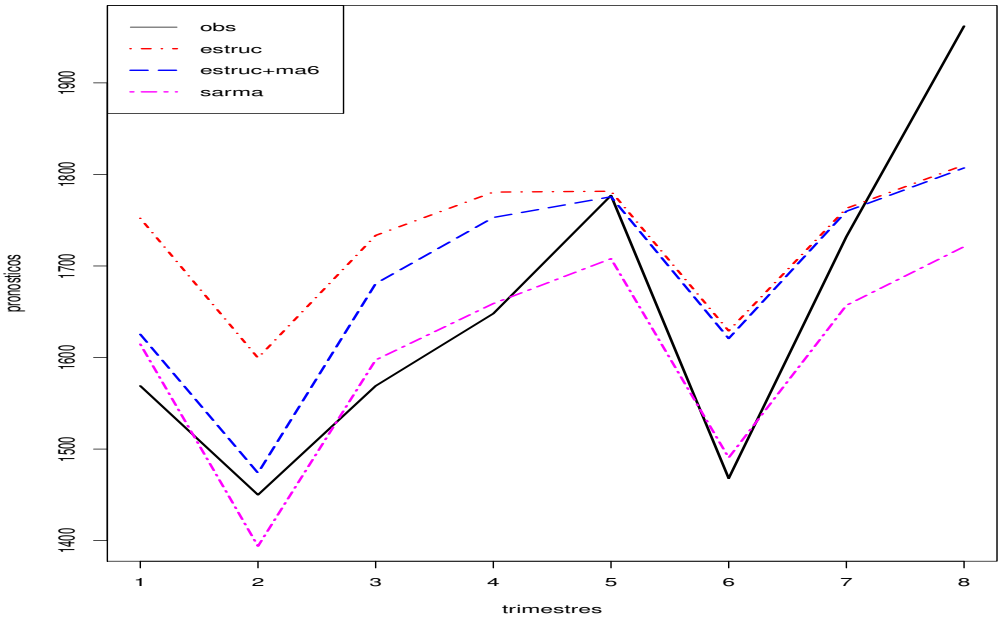



Figura 8.10: Pronósticos con tres modelos para producción de cemento

CAPÍTULO 9

# Raíces Unitarias y Tendencias Estocásticas (ARIMA)

## 9.1. Introducción

En este capítulo se presenta en modelo alterno al de componentes con errores ARMA, conocido como modelo ARIMA-SARIMA ó modelo de Box-Jenkins. El modelo ARIMA puede tomarse como análogo del modelo con tendencia, $Y_t = T_t + \varepsilon_t$, y el modelo SARIMA como análogo del modelo con tendencia y estacionalidad, $Y_t = T_t + S_t + \varepsilon_t$. Primero se introducen los modelos ARIMA, luego las pruebas de raíz unitaria de Dickey-Fuller para detectar estos modelos. Más adelante se introducen los modelos SARIMA y las pruebas de raíz unitaria estacional para detectarlos.

## 9.2. Modelos ARIMA

**Observación:** Si en $Y_t = a+bt+\varepsilon_t$, $\varepsilon_t \sim RB(0, \sigma^2)$ se toma $\Delta Y_t = (1-L)Y_t = Y_t - Y_{t-1}$, se obtiene

$$\begin{aligned} Y_t - Y_{t-1} &= a + bt + \varepsilon_t - (a + b(t-1) + \varepsilon_{t-1}) \\ &= b + \varepsilon_t - \varepsilon_{t-1} = b + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad \theta_1 = -1, \end{aligned}$$

luego, llamando $W_t = \Delta Y_t$ y $\eta_t = \varepsilon_t - \varepsilon_{t-1}$, se obtiene $W_t = b + \eta_t$. Nótese que $\eta_t$ es MA(1) y por tanto $W_t$ es un $MA(1)$ con media diferente de cero, estacionario en covarianza. Luego, diferenciar una serie no estacionaria, con tendencia lineal, puede producir una serie estacionaria en covarianza. Esta idea es la base de los modelos Box-Jenkins. Pero requiere algunos conceptos para formalizarla.

**Definición 9.2.1** (**Raíz Unitaria Autoregresiva).** *Suponga un modelo* $ARMA(p, q)$

$$\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2),$$

*tal que una de las* $p$ *raíces de la ecuación* $\Phi_p(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p = 0$ *es* $z = 1$*, entonces se dice que* $Y_t$ *tiene una raíz unitaria autorregresiva.*

En este caso el polinomio $\Phi_p(z)$ factoriza como $\Phi_p(z) = (1 - z)\Phi_{p-1}(z)$ donde $\Phi_{p-1}(z)$ es un polinomio de grado $p - 1$ y

$$\Phi_{p-1}(L)(1 - L)Y_t = \Theta_q(L)\varepsilon_t,$$

luego si $Y_t \sim ARMA(p, q)$ con una raíz unitaria entonces $\Delta Y_t \sim ARMA(p - 1, q)$.

**Ejemplo 9.2.1.** *Si* $\Phi(L) = (1 - 0.2L)(1 - L), \quad \Theta(L) = 1 - 0.3L$

$$\begin{aligned}(1 - 0.2L)(1 - L)Y_t &= (1 - 0.2L)\varepsilon_t \\ (1 - 1.2L + 0.2L^2)Y_t &= \varepsilon_t - 0.3\varepsilon_{t-1} \\ Y_t - 1.2Y_{t-1} + 0.2Y_{t-2} &= \varepsilon_t - 0.3\varepsilon_{t-1}.\end{aligned}$$

*Entonces* $Y_t \sim ARMA(2, 1)$*, y no es estacionario en covarianza pues* $1 - 1.2z + 0.2z^2 = 0$ *tiene raíces* $z_1 = 1,\ z_2 = 5$ *y sólo una cumple* $|z| > 1$*. Pero* $W_t = \Delta Y_t = Y_t - Y_{t-1}$ *cumple* $(1 - 0.2L)W_t = (1 - 0.3)\varepsilon_t \sim ARMA(1, 1)$ *es estacionario en covarianza e invertible.*

**Definición 9.2.2** (**Proceso ARIMA(p,d,q)).** *Una serie de tiempo* $Y_t, t \in \mathbb{Z}$ *sigue un modelo* $ARIMA(p, 1, q)$*, con tendencia, si* $(1 - L)Y_t$ *sigue un proceso ARMA(p,q) estacionario en covarianza, con media diferente de cero. Es decir, si*

$$\Phi_p(L)(1 - L)Y_t = \delta + \Theta_q(L)\varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2), \tag{9.1}$$

*y* $\Phi_p(z) = 0$ *tiene las raíces fuera del círculo unitario. Una serie de tiempo sigue un modelo* $ARIMA(p, d, q)$ *si*

$$\Delta^d Y_t = (1 - L)^d Y_t, \qquad d = 1, 2, \ldots$$

*es un proceso ARMA(p,q) estacionario en covarianza.*

En la práctica los casos más importantes son $d = 0, 1$, ver Diebold [1999, pág. 210].

## 9.3. Los Modelos de Marcha Aleatoria

Las marchas aleatorias son ejemplos básicos de modelos ARIMA.

**Definición 9.3.1.** *La marcha aleatoria sin tendencia se define como un proceso $Y_t \sim AR(1)$, de media cero, con coeficiente $\varphi = 1$, es decir*

$$Y_t = Y_{t-1} + \varepsilon_t,\ t \in \mathbb{Z}, \qquad \varepsilon_t \sim RB(0, \sigma^2). \tag{9.2}$$

*Se denota $Y_t \sim ARIMA(0, 1, 0)$. La marcha aleatoria con tendencia se define como un proceso $AR(1)$ de media diferente de cero, con coeficiente $\varphi = 1$, es decir*

$$Y_t = \mu + Y_{t-1} + \varepsilon_t,\ \mu \in \mathbb{R}. \tag{9.3}$$

La marcha aleatoria sin tendencia cumple, para $t \geq 0$, y $Y_0$ dado, no aleatorio,

$$Y_t = Y_0 + \sum_{i=0}^{t} \varepsilon_i, \tag{9.4a}$$

$$E(Y_t) = Y_0, \tag{9.4b}$$

$$Var(Y_t) = t\sigma^2. \tag{9.4c}$$

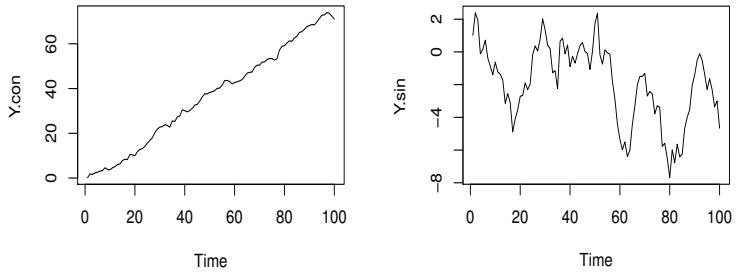
La marcha aleatoria con tendencia cumple

$$Y_t = t\delta + Y_0 + \sum_{i=0}^{t} \varepsilon_i, \tag{9.5a}$$

$$E(Y_t) = t\delta + Y_0, \tag{9.5b}$$

$$Var(Y_t) = t\sigma^2. \tag{9.5c}$$

Un ejemplo de trayectorias de marchas aletorias con y sin tendencia se muestran a continuación en la Figura 9.1.




(a) con tendencia (b) sin tendencia

Figura 9.1: Trayectorias de Marchas Aleatorias

Obsérvese que la marcha aleatoria con tendencia $Y_t$, tiene una característica similar a una serie con tendencia lineal de la forma $X_t = a + bt + \varepsilon_t$, $\varepsilon_t \sim RB(0, \sigma^2)$, porque en promedio crecen en cada período $[t-1, t]$, una cantidad constante, $E(Y_t - Y_{t-1}) = \delta$ y $E(X_t - X_{t-1}) = b$.

Suponga $Y_t = \mu + \varphi Y_{t-1} + \varepsilon_t,\ t = 1, 2, \ldots, T$, un proceso AR(1) con media diferente de cero. Los pronósticos a $j$ pasos a partir de $T$ son

$$\hat{Y}_T(j) = \mu + \varphi^j Y_T$$

para $j = 1, 2, \ldots, h$. En el caso $\varphi = 1$, se obtiene una marcha aleatoria con tendencia. En este caso $\hat{Y}_T(j) = \mu + Y_T,\ j \geq 1$, lo cual se interpreta como que una marcha aleatoria con tendencia no puede pronosticarse. Su mejor pronósticos es el último valor conocido.

Finalmente, a partir de $Y_t = Y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$, para $k = 1, 2, \ldots$,

$$\begin{aligned} Cov(Y_t, Y_{t+k}) &= \mathbb{E}((Y_t - \mathbb{E}(Y_t))(Y_{t+k} - \mathbb{E}(Y_{t+k}))) \\ &= \mathbb{E}(\sum_{j=1}^{t} \varepsilon_j \sum_{i=1}^{t+k} \varepsilon_i) = \sum_{j=1}^{t} \sum_{i=1}^{t+k} \mathbb{E}(\varepsilon_j \varepsilon_i) \\ &= \sigma^2 t, \end{aligned}$$

Como $Var(Y_t) = \sigma^2 t$ entonces

$$Corr(Y_t, Y_{t+k}) = \frac{\sigma^2 t}{\sqrt{\sigma^2 t \sigma^2 (t+k)}} = \frac{1}{\sqrt{1 + k/t}}. \tag{9.6}$$

Para $t$ grande tal que $k/t$ sea muy pequeño, se puede aproximar a cero y obtener que las autocorrelaciones en marchas aleatorias con tendencia son casi uno, $\rho(k) \approx 1$. Como la tendencia no determina el valor de la autocorrelación, la conclusión también es válida en el caso de marchas aleatorias sin tendencia. Las marchas aleatorias son series con fuertes autocorrelaciones.

**Ejemplo 9.3.1.** *Si* $Y_t \sim ARMA(1,1)$ *con* $E(Y_t) = 0$ *entonces*

$$(1 - \varphi L)Y_t = (1 + \theta L)\varepsilon_t$$

*y se puede comprobar que las autocorrelaciones de* $Y_t$ *satisfacen:*

$$\rho(1) = \frac{(1 + \varphi\theta)(\varphi + \theta)}{1 + \theta^2 + 2\varphi\theta}$$
$$\rho(k) = \rho(k-1), \qquad k \geq 2.$$

*Si* $\varphi \to 1$ *obtenemos un* $ARIMA(0,1,1) = IMA(1)$ *y*

$$\rho(1) \to \frac{(1+\theta)(1+\theta)}{1 + \theta^2 + 2\theta} = 1.$$

*Por lo tanto* $\rho(k) \equiv 1, \ \forall k \geq 1$.

Nótese que si $Y_t \sim ARIMA(p,1,q)$ entonces $Y_t - Y_{t-1} = W_t$, con $W_t$ estacionario. Si $Y_0$ se asume dada, entonces $Y_t = Y_0 + \sum_{j=1}^{t} W_j$. Esta identidad es similar a la marcha aleatoria, (9.4a), pag.157. Además, un IMA(1) es similar a la marcha aleatoria sin tendencia. Estas observaciones dan lugar a afirmar que: "los procesos $ARIMA(p,1,q)$ se comportan como marchas aleatorias en ciertos aspectos". Como se comprobó en (9.6), las marchas aleatorias son muy autocorrelacionadas. Los procesos ARIMA también muestran una fuerte autocorrelación. La fac muestral es muy persistente y cercana a uno.

**Definición 9.3.2.** *Si la serie* $Y_t$ *no estacionaria en covarianza es tal que* $\Delta Y_t$ *es estacionario, se dice que* $Y_t$ *es integrada de orden 1,* $I(1)$. *Si* $\Delta^2 Y_t = \Delta(\Delta Y_t) = (1-L)^2 Y_t$ *es estacionario en covarianza se dice que integrada de orden 2,* $I(2)$. *La definición de proceso integrado de orden* $d \geq 1$ *es inmediata.*

Luego un $ARIMA(p,d,q)$ es un proceso $I(d)$.

**Definición 9.3.3.** *Una serie* $(Y_t, t \in \mathbb{Z})$ *se dice que tiene tendencia estocástica si es integrada de orden* $d = 1, 2$.

La marcha aleatoria con tendencia (9.3) es un ejemplo de tendencia estocástica, en contraposición a una seria con tendencia lineal $T_t = a + bt$, la cual se dice que tiene tendencia

determinística. De igual manera, un proceso $ARIMA(p, 1, q)$ es un proceso con tendencia estocástica.

Una serie con tendencia estocástica presenta períodos de duración aleatoria durante los cuales crece linealmente.

Es el caso de los precios de las acciones. En un período de tiempo el precio puede aumentar por efecto de la demanda. Pero llega un momento en el cual la demanda se satura y los precios empiezan nuevamente a oscilar, o a tener otro comportamiento. Para el rendimiento ó retorno del precio de las acciones, la hipótesis de mercados eficientes asume que son ARIMA(0,1,0). Sin embargo, se dan casos en los que se encuentra un ARIMA(0,1,1), denominado IMA(1), ver, por ejemplo, Box et al. [1976], pag. 103.

La implementación de los modelos ARIMA empieza con series con tendencia estocástica. La metodología consiste en diferenciar la series $d = 1, 2$ veces, hasta obtener otra serie estacionaria en covarianza, identificable por un modelo ARMA, y estimar los parámetros de este modelo.

## Ejemplos de ARIMA

A continuación se muestran algunos ejemplos de estimación de modelos ARIMA con la función `arima()`. En la sección siguiente se trata las pruebas de hipótesis para detectar la presencia de raíces unitarias.

**Ejemplo 9.3.2.** *En Diebold [1999, pág. 12] se considere la serie del Producto Interno Bruto en EUA entre 1869-1933, con el período 1934-1993 para comparación. Se examinan dos posibles modelos para esta serie.*

1. **Modelo con tendencia determinística lineal y errores** $AR(2)$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \varepsilon_t \\ \varepsilon_t &= \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \eta_t, \qquad \eta_t \sim RB(0, \sigma^2). \end{aligned}$$

2. **Modelo ARIMA**

$$\begin{aligned} Y_t &\sim ARIMA(1, 1, 0) \textit{ con tendencia} \\ (1 - \varphi L)(1 - L) Y_t &= \delta + \varepsilon_t. \end{aligned}$$

*Entre 1932 y 1933 hubo una grave recesión en EUA y los pronósticos con base en 1933 se hacen en una posición por debajo de la tendencia*

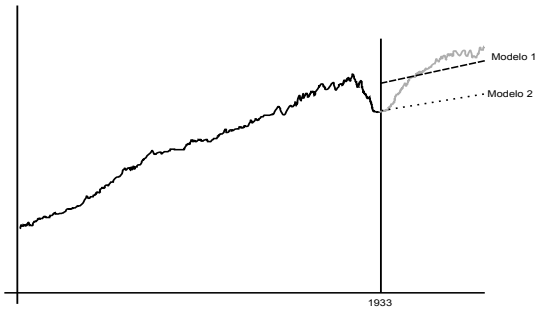



Figura 9.2: Pronósticos de la Serie PNB-USA.

*El modelo 1 produce mejores pronósticos que siguen la tendencia, en cambio, el modelo 2 no, pues subestima los valores.*

**Ejemplo 9.3.3.** *Análisis de la Tasa de cambio: USD/YEN, el precio de 1 US Dólar en Yen, ver Diebold [1999, pág. 221]. La serie es mensual desde 1973-01 hasta 1996-07. Por ejemplo,* $Y_t = 107$ *significa que en el mes* $t$ *se pagaban 107 Yen por 1 USD. El análisis siguiente es con el logaritmo de esta tasa. En lo que sigue* $y$ *denota el logaritmo de la tasa y* $dy$ *la serie diferenciada* `dy = diff(y,1,1)`.

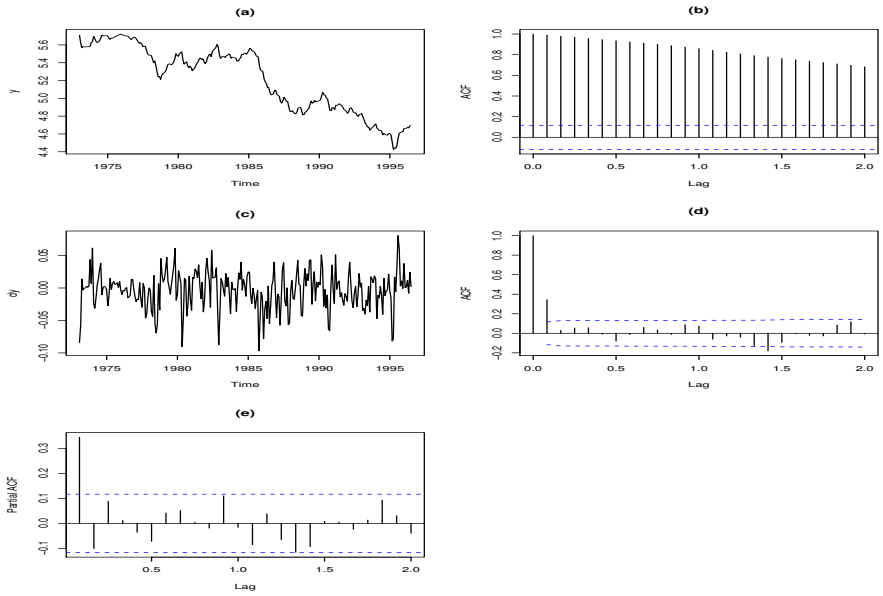



Figura 9.3: (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada

*La Figura 9.3 (b) sugiere un proceso integrado debido a que la fac es muy persistente; y*

*las Figuras (d) y (e) sugieren un ARMA para la serie diferenciada. Utilizando la función* `z = autoarmafit(dy)`*, de la librería* `timsac`*, con la serie diferenciada se obtiene un modelo ARMA(3,2). Utilizando la función* `arima(y,order=c(3,1,2))` *se obtienen los resultados de la Tabla 9.1. Finalmente, utilizando la prueba Ljung-Box, se comprueba*

Tabla 9.1: Parámetros del modelo ARIMA(3,1,2)

| | parametros | sd.dev | est t |
|---|---|---|---|
| ar1 | 0.18 | 0.06 | 2.91 |
| ar2 | -0.93 | 0.01 | -77.71 |
| ar3 | 0.37 | 0.06 | 6.17 |
| ma1 | 0.20 | 0.02 | 11.32 |
| ma2 | 0.99 | 0.03 | 32.43 |

*que los residuos de este modelo son Ruido Blanco. El resutado de la prueba es*

```
X-squared = X-squared = 15.7889, df = 26, p-value = 0.9411,
```

*y no rechaza la hipótesis de ruido blanco ya que el valor p es mayor de 0.05. Este modelo es válido para pronosticar el logaritmo de la tasa de cambio, a corto plazo. Hay que anotar que la función* `arima()` *de R tiene varias versiones, y que no es lo mismo estimar el modelo ARIMA(3,1,2) con la variable* $\log(Y_t)$*, que estimar el modelo ARMA(3,2) con la variable* $\Delta \log(Y_t)$*. Los pronósticos a 19 períodos se muestran en la Figura 9.4, comparados con los pronósticos de un modelo con tendencia lineal y error estructural AR(2).*

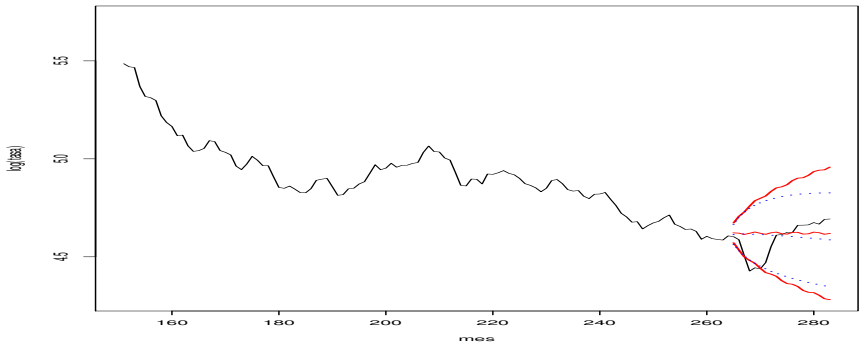
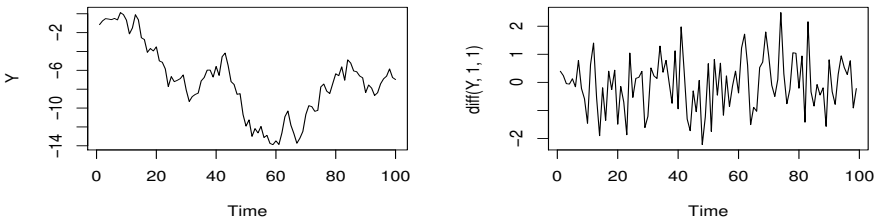



Figura 9.4: Pronósticos usd/yen con ARIMA(3,1,2)(continua) y Tendencia Lineal+AR(2) (punteada)

**Ejemplo 9.3.4.** *Considerando* $Y_t$ *la serie del logaritmo de la tasa de cambio USD/Libra, es decir, el log del precio de 1 libra esterlina en usd, mensual, en el período 01/80-12/88,*

*con 478 observaciones. La trayectoria parece una marcha aleatoria sin tendencia, como se observa en la Figura 9.6.*




(a) Serie USD/Libra

(b) Primera Diferencia de la Serie USD/Libra

Figura 9.5: Serie USD por Libra

*La primera diferencia parece estacionaria, casi ruido blanco. La fac de la primera diferencia confirma que es ruido blanco, obteniendo un modelo de marcha aleatoria con tendencia, dado por*

$$(1-L)Y_t = \mu + \varepsilon_t, \quad \mu = -0.0009.$$

*Este modelo se denomina Modelo de Black-Scholes ó Modelo Log-Normal para el precio de un activo.*

**Ejemplo 9.3.5.** *Considerando $Y_t$ la serie del empleo en Canadá de periodicidad trimestral en el período 01/1962-04/1993, desestacionalizada. Se propuso un modelo $AR(2)$. Pero usando la función* `auto.arima`*, se obtiene $Y_t \sim ARIMA(1,1,0)$ sin intercepto, es decir,*

$$(1-\varphi L)(1-L)Y_t = \varepsilon_t, \qquad \varepsilon_t \sim RB(0,\sigma^2), \tag{9.7}$$

*con $\hat{\varphi} = 0.4598,\ s.e. = 0.0758,\ \hat{\sigma} = 2.068,\ BIC = 491.22,\ AIC = 485.41,\ MAPE = 1.0453$.*

*Si se ajusta un modelo $AR(2)$ a la serie, es decir,*

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \tag{9.8}$$

*se obtiene $\hat{\varphi}_1 = 1.4505,\ \hat{\varphi}_2 = -0.4763,\ \hat{\sigma} = 2.022,\ \hat{\mu} = 97.498,\ AIC = 493.57,\ BIC = 505.22,\ MAPE = 1.0648$.*

*La conclusión es que sería preferible utilizar el modelo (9.7) por tener un mejor ajuste. Pero es posible utilizar otra herramienta para discriminar entre los posibles modelos.*

*La prueba de hipótesis Diebold-Mariano, en la función* `dm.test` *de la librería* `forecast` *compara las medias de los errores de pronóstico a h pasos, dentro de la muestra, para dos modelos. La hipótesis nula es que las medias son iguales y la alterna que no lo son. Es posible utilizar hipótesis de una cola, con la especificación de que el primer modelo tiene menor error medio que el segundo.*

*Se obtiene el valor-p para la prueba de dos colas:* `p-value = 0.8412`. *Luego no rechaza la hipótesis nula y ambos modelos tienen la misma capacidad predictiva.*

```
#### programa en R
library(forecast) # Para la función arima y la prueba dm.test
# Canadian employment index, seasonally adjusted, 1961:1-1994:4
# 136 obs
E = read.table("CAEMP.DAT", header = TRUE)
y = ts(E$caemp, frequency = 4, start = c(1961,01), end = c(1994,04))
f1 = arima(y, order=c(2,0,0))
f2 = arima(y, order=c(1,1,0))
accuracy(f1)
accuracy(f2)
dm.test(residuals(f1), residuals(f2), h=1)
----------resultado
data:  residuals(f1) residuals(f2)
DM = -0.2004, Forecast horizon = 1,
Loss function power = 2, p-value = 0.8412
alternative hypothesis: two.sided
```

## 9.4. Raíces Unitarias, Estimación y Prueba de Hipótesis

La discusión en esta sección es para un $Y_t \sim ARIMA(p, 1, q)$. Suponga un proceso $AR(1)$, $Y_t = \varphi Y_{t-1} + \varepsilon_t$, en el cual $\varphi = 1$ pero $\varphi$ se estima usando MCO. Este estimador $\widehat{\varphi}$, tiene dos propiedades

1. **Superconsistencia**

   *a*) En el caso $\varphi = 1$ el estimador de MCO $\hat{\varphi}_T$ tiene la propiedad de que

   $$T(\widehat{\varphi}_T - 1) \xrightarrow{d} Z, \qquad T \to \infty, \tag{9.9}$$

   donde $Z$ es una variable aleatoria no degenerada, y el símbolo $\xrightarrow{d}$ denota la convergencia en distribución.

   *b*) En el caso $|\varphi| < 1$, el estimador de MCO, $\hat{\varphi}_T$ cumple

   $$\sqrt{T}(\widehat{\varphi}_T - \varphi) \xrightarrow{d} Z \qquad T \to \infty, \tag{9.10}$$

   donde $Z$ es variable aleatoria no degenerada.

   como $\sqrt{T} < T$, se dice que la convergencia en el caso (9.9) es más rápida que en el caso (9.10), y a esto se denomina superconsistencia, es decir, el estimador de MCO, $\widehat{\varphi}_T$ de una raíz unitaria es superconsistente.

2. **Sesgo de $\widehat{\varphi}$**
   Si $\widehat{\varphi}_T$ es el estimador de MCO de $\varphi$ entonces $E(\widehat{\varphi}_T) < \varphi$ y el sesgo es mayor cuando $\varphi = 1$. El sesgo es $\varphi - \widehat{\varphi}_T$ y crece si se considera tendencia en la serie. Aunque $\widehat{\varphi}_T$ converge a $\varphi$, cuando $T \to \infty$, el sesgo puede ser apreciable en muestras no muy grandes.

## La fac y facp muestrales en presencia de Raíces Unitarias

En un proceso $Y_t$ con raíz unitaria, por ejemplo, $ARIMA(p, 1, q)$ la fac muestral converge a cero con mucha lentitud, la facp muestral presenta un valor cercano a 1 en el rezago $k = 1$ y los demás valores son aproximadamente cero. Por ejemplo la Figura 9.6 siguiente se muestra la fac y la facp típicas de un proceso con raíz unitaria.

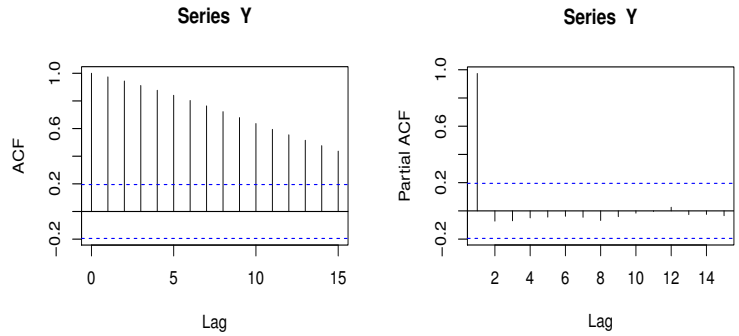



Figura 9.6: fac y facp muestrales de proceso con raíz unitaria

Hay que tener precaución con relación a tomar la decisión de diferenciar la serie para buscar un ARMA cuando se observan las características anteriores. Y es que es necesario aplicar una prueba de hipótesis para determinar si existe una raíz unitaria en lugar de una tendencia lineal determinística ó, incluso, un AR(p) muy persistente. Las siguientes observaciones son pertinentes.

1. Diebold [1999, pág. 221]: "Si no hay raíz unitaria es conveniente utilizar modelos de niveles (componentes determinísticas) y es adecuando diferenciar sólo en el caso de raíz unitaria; si la diferenciación es inadecuada puede ser dañina, incluso asintóticamente".

2. Cuando una serie tiene raíz unitaria la serie es no estacionaria y los estimadores MCO no se distribuyen Normal.

3. Soares and Medeiros [2008, pág. 4]: "La mayoría de los artículos sobre pronósticos de demanda de energía toman diferencias sin hacer una prueba previa de raíz unitaria. Esto es un error grande cuando la tendencia es deterministica, tomar diferencias introduce una componente MA no invertible el cual causa problemas serios de estimación."

## 9.5. Prueba Dickey-Fuller (DF)

La prueba Dickey-Fuller se basa en asumir que la serie se puede aproximar por un proceso AR(1) con tres variantes: media cero, media diferente de cero y tendencia lineal. Inicialmente se asume que $Y_t$ sigue un modelo AR(1) de media cero y se procede a transformar este modelo de la siguiente manera.

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varepsilon_t, \\ Y_t - Y_{t-1} &= (\varphi_1 - 1) Y_{t-1} + \varepsilon_t, \\ \Delta Y_t &= \rho Y_{t-1} + \varepsilon_t. \end{aligned}$$

donde $\rho = \varphi_1 - 1$. La existencia de una raíz unitaria $\varphi_1 = 1$ equivale a $\rho = 0$.

1. Prueba DF para el caso 1: $Y_t \sim AR(1)$ con media cero. Entonces

$$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t. \tag{9.11}$$

   La hipótesis nula es $H_0 : \rho = 0$ versus la alterna $H_a : \rho < 0$. El estadístico de la prueba se denota por $\tau$ y su distribución bajo $H_0$ permite calcular los valores críticos, por ejemplo, $\tau_{0.05}$ para un nivel de significación de 5 %.

El criterio de rechazo es $\widehat{\tau} < \tau_{0.05}$, con $\widehat{\tau}$ el valor calculado del estadístico. En R la prueba DF se implementa mediante la librería `urca` por medio de la función, para este caso, `ur.df(y,type="none",lags=0)`.

2. Prueba DF para el caso 2: $Y_t \sim AR(1)$ con media diferente de cero. Entonces se modifica (9.11) así:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \tag{9.12}$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y,type="drift",lags=0)`.

3. Prueba DF para el caso 3: suponiendo que $Y_t \sim AR(1)$ con media diferente de cero y con tendencia lineal, entonces

$$\Delta Y_t = \alpha + \beta\, t + \rho Y_{t-1} + \varepsilon_t, \tag{9.13}$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y,type="trend",lags=0)`.

## 9.6. La Potencia de las pruebas Dickey-Fuller

Con base en la Tabla 9.2 siguiente,

| | | $H_0$ real | |
|---|---|---|---|
| | | V | F |
| $H_0$ según prueba | V | OK | error II |
| | F | error I | OK |

Tabla 9.2: Tabla de decisiones en una prueba de hipótesis

recordamos la definición de potencia de una prueba, $1 - \beta$ donde $\beta = \mathbb{P}(errorII)$. El error tipo II se denomina "falso negativo". Es el evento de no rechazar $H_0$ siendo falsa. El error tipo I se denomina "falso positivo". Es el evento de rechazar $H_0$ siendo cierta. La probabilidad del error tipo I es $\alpha$, el nivel de significación. Fijado este nivel, y el tamaño de la muestra, la potencia puede ser baja ó alta. Una baja potencia significa un $\beta$ alto, es decir, una probabilidad alta de cometer error II, de no rechazar la hipótesis nula siendo falsa; es decir, de detectar un "falso negativo". También una prueba con baja potencia significa que es "conservadora". Es una prueba poco capaz de detectar la hipótesis alterna cuando esta es cierta.

Las pruebas DF se sabe que tienen baja potencia. Por tanto, rechazar la hipótesis nula de raíz unitaria no necesariamente significa que se deba asumir que existe. Como señala Diebold [1999, pág 220]: "Las pruebas de raíz unitaria tienen problemas de potencia y tamaño de muestra. De potencia porque las hipótesis alternas son muy cercanas a la hipótesis nula". La presencia de cambios estructurales en la serie se sabe que disminuye la potencia de las pruebas DF, pero no se desarrollará este tema.

## 9.7. Implementación en R de la Prueba Dickey-Fuller

**Ejemplo 9.7.1.** *Continuando con el Ejemplo 9.3.3, del logaritmo de la tasa de cambio USD/Yen, log del precio de 1 usd en yens, en donde se encontró que la serie* $\ln Y_t$ *puede ser modelada por un IMA(1), es decir por un ARIMA(0,1,1). A continuación se realizan las prueba DF y DF aumentada. El resultado es que en ambos casos no se rechaza la hipótesis de raíz unitaria, por lo que el modelo ARIMA(0,1,1) queda justificado. Se realiza la prueba de raíz unitaria tomando en cuenta la tendencia. Es decir, se aplica el caso 3: "trend".*

```
## analisis con pruebas dickey-fuller de usd/yen
library(forecast)
library(urca)
library(fUnitRoots)
uno = read.table("yen_usd.dat",header=T,stringsAsFactors=F)
attach(uno)
y = ts(y,frequency=12, start=c(1973,01), end = c(1996,07))
ly = log(y)
## prueba df caso 3 = trend, con libreria urca
df.trend = ur.df(y = ly, lags = 0, type = 'trend' )
summary(df.trend)
```

La prueba DF es para el caso 3 de un AR(1) con tendencia lineal, a un nivel de significación de 5 %. El estadístico DF observado fué -1.7764 mayor que el valor crítico -3.42, $\widehat{\tau} > \tau_{0.05}$, luego no rechaza la hipótesis nula de raiz unitaria. Aunque hay que tener en cuenta la baja potencia de la prueba. En conclusión, es factible un modelo ARIMA(p,1,q) para el logaritmo de la tasa de cambio usd/yen.

```
--------------- Resultados (salida de R)
Test regression trend
Call: lm(formula = z.diff ~ z.lag.1 + 1 + tt)
```

```
Value of test-statistic is: -1.7764 2.568 1.5811
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

## 9.8. Prueba Dickey-Fuller Aumentada

La prueba aumentada de Dickey-Fuller no es solamente una prueba sino que requiere una estrategia de análisis para su aplicación. Como señalan Elder and Kennedy [2001, pag. 139]:

> " Un ingrediente crucial en esta prueba, que no se reconoce bien en los libros de texto, es que se requiere una estrategia de prueba, en oposición al simple cálculo del estadístico de la prueba. Esta estrategia es necesaria para determinar si un intercepto, un intercepto más una tendencia con el tiempo, o ninguna de las dos anteriores debería incluírse al correr la regresión para la prueba de raiz unitaria. Incluír demasiados regresores puede generar una pérdida de potencia, mientras que no incluír suficientes puede generar resultados sesgados hacia el no rechazo de la hipótesis nula...incluír intercepto, ó intercepto más tendencia, es necesario para permitir una representación de la hipótesis alterna que pueda competir contra la hipótesis nula de raíz unitaria."

La estrategia para la prueba Dickey-Fuller Aumentada consiste en asumir que $Y_t \sim AR(p)$, y determinar cuál de los tres casos considerados anteriormente, con $p = 1$, determina una mejor aproximación a la serie original. Nótese que los casos a examinar dependen del orden autorregresivo $p$, por lo que la búsqueda para, por ejemplo, $p = 1, 2, 3$, requiere examinar $3 \times 3 = 9$ casos. La decisión se toma con base en el menor AIC.

**Caso 1.** Suponiendo que $Y_t \sim AR(p)$ con media cero, caso indicado por "none", entonces se realizan las transformaciones siguientes.

$$Y_t = \sum_{j=1}^{p} \varphi_j Y_{t-j} + \varepsilon_t = \varphi_1 Y_{t-1} + \sum_{j=2}^{p} \varphi_j Y_{t-j} + \varepsilon_t. \tag{9.14}$$

Defina

$$\rho_1 = \sum_{j=1}^{p} \varphi_j,\ \rho_i = -\sum_{j=i}^{p} \varphi_j,\ i = 2, \ldots, p,$$

entonces con $\rho = \rho_1 - 1$; la ecuación (9.14) se transforma en

$$\Delta Y_t = \rho Y_{t-1} + \sum_{j=2}^{p} \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \tag{9.15}$$

Si hay una raíz unitaria se cumple

$$1 - \sum_{j=1}^{p} \varphi_j = 1 - \rho_1 = 0,$$

por tanto, $\rho = 0$. En este caso el modelo (9.15) equivale a

$$\Delta Y_t = \sum_{j=1}^{p-1} \rho_{j+1} \Delta Y_{t-j} + \varepsilon_t, \tag{9.16}$$

es decir un $AR(p-1)$ en la variable $Z_t = \Delta Y_t$. La hipótesis nula es $H_0 : \rho = 0$ y la alterna $H_a : \rho < 0$. El estadístico DF, $\tau$ tiene la misma distribución asintótica que el estadístico DF en el caso AR(1) dado en (9.11). La hipótesis nula se rechaza si $\widehat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

> "Así, los resultados del proceso $AR(1)$ se generalizan asintóticamente en forma directa a procesos de orden superior". Diebold [1999, pág. 128]

**Caso 2** Suponiendo $Y_t \sim AR(p)$ con media diferente de cero, caso indicado por "drift". Con la misma notación del Caso 1 anterior, se realizan las transformaciones siguientes.

$$\begin{aligned} Y_t - \mu &= \sum_{j=1}^{p} \varphi_j (Y_{t-j} - \mu) + \varepsilon_t \\ Y_t &= \alpha + \varphi_1 Y_{t-1} + \sum_{j=2}^{p} \rho_j (Y_{t-j+1} - Y_{t-j}) + \varepsilon_t \\ \Delta Y_t &= \alpha + \rho Y_{t-1} + \sum_{j=2}^{p} \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \end{aligned} \tag{9.17}$$

donde $\alpha = \mu(1 - \sum_{j=1}^{p} \varphi_j)$, con los $\rho_j$ definidos como en el Caso 1 anterior.

La hipótesis nula es $H_0 : \rho = 0$ y la alterna $H_a : \rho < 0$. Bajo $H_0 : \rho = 0$ el término $\alpha$ se anula. La distribución asintótica del estadístico DF es igual a la del caso $AR(1)$ con media, (9.12). La hipótesis nula se rechaza si $\widehat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

**Caso 3** Suponiendo $Y_t \sim AR(p)$ con tendencia lineal, caso indicado por "trend". En este caso se define

$$Y_t = a + bt + \sum_{j=1}^{p} \varphi_j \big(Y_{t-j} - a - b(t-j)\big) + \varepsilon_t \tag{9.18}$$

que se puede reordenar de la siguiente forma

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^{p} \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \tag{9.19}$$

donde

$$\begin{aligned} k_1 &= a\left(1 - \sum_{i=1}^{p} \varphi_i\right) + b\sum_{i=1}^{p} i\varphi_i, \\ k_2 &= b\left(1 - \sum_{i=1}^{p} \varphi_i\right), \\ \rho_1 &= \sum_{i=1}^{p} \varphi_i,\ \rho = \rho_1 - 1. \end{aligned}$$

La hipótesis nula es $H_0:\ \rho = 0$ y la alterna $H_a:\ \rho < 0$. Bajo $H_0:\ \rho = 0$ se tiene que $k_2 = 0,\ k_1 = b\sum_{i=1}^{p} i\varphi_i$ y el estadístico DF tiene la misma distribución asintótica del estadístico en el caso AR(1) con tendencia lineal, (9.13). La hipótesis nula se rechaza si $\widehat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

**La Prueba Dickey-Fuller Aumentada**. A continuación se presenta una implementación de la estrategia de análisis mencionada en Elder and Kennedy [2001, pag. 139]. El objetivo es determinar cuál caso de los tres posibles con modelos AR(p) es el que mejor aproxima la serie para así lograr una mayor potencia en la prueba.

Para esto se utiliza la librería `dynlm` que extiende la función de regresión lineal `lm()` para series de tiempo, permitiendo incluír valores rezagados $Y_{t-j}$, con el comando `L(y,j)` y tendencia lineal $a+bt$ con el comando `trend(y)` como variables explicativas. Por ejemplo, para el modelo de la ecuación (9.19), en la pag. 171, con p = 3

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^{p} \rho_j \Delta Y_{t-j+1} + \varepsilon_t, \tag{9.20}$$

es decir $\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \rho_2 \Delta Y_{t-1} + \rho_3 \Delta Y_{t-2} + \varepsilon_t$, se programa, con `dly` la variable $\Delta Y_t$,

```
reg.trend1 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1)+ L(dly, 2)).
```

A continuación se programan las regresiones para $p = 1, 2$, excluyendo el caso "none" de media cero.

```
## estrategia de regresiones para la df aumentada
require(dynlm)

reg.drift1 = dynlm(dly ~ L(ly,1))
reg.drift2 = dynlm(dly ~ L(ly,1) + L(dly, 1))

reg.trend1 = dynlm(dly ~ trend(ly) + L(ly,1))
reg.trend2 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1))

(c(AIC(reg.drift1),AIC(reg.drift2)))
(c(AIC(reg.trend1),AIC(reg.trend2)))
```

Se detecta el caso reg.trend1, que corresponde a $p = 2$ como el de menor AIC.

```
-----resultados de los valores AIC para los 4 modelos
 -1202.376 -1240.019
 -1203.215 -1245.147
```

Y la correspondiente prueba DF aumentada se programa con la función `ur.df()` de la librería `urca` con: `ur.df(y = ly, lags = 2, type = 'trend')`.

```
#---------programa la prueba ADF con p = 2 y trend
df.trend = ur.df(y = ly, lags = 2, type = 'trend' )
summary(df.trend)
#-----resultados de la prueba DF aumentada
Test regression trend
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Value of test-statistic is: -2.58 2.9804 3.448
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

La hipótesis nula se rechaza si $\widehat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria. A partir de este último resultado se concluye que no se rechaza la hipótesis de raíz unitaria ya que el valor observado de la prueba -2.58 es mayor que el valor crítico al nivel de 5 %, -3.42, $\widehat{\tau} > \tau_{0.05}$. Luego, es válido diferenciar la serie. En el Ejemplo 9.3.3 se ajustó el modelo ARIMA(3,1,2) que produce buenos pronosticos del logaritmo de la tasa.

# CAPÍTULO 10

# Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)

## 10.1. Modelos SARIMA

En esta sección se define y aplica el operador rezago estacional, $L^s$, donde $s > 1$, entero, es el período de una componente estacional, que se asume está, latente, en la serie $Y_t$. Se definen las operaciones

$$L^s(Y_t) = Y_{t-s}, \tag{10.1}$$

$$(I - L^s)Y_t = \Delta_s Y_t = Y_t - Y_{t-s}, \tag{10.2}$$

$$(I - L^s)^2 Y_t = \Delta_s^2 Y_t = (I - 2L^s + L^{2s})Y_t. \tag{10.3}$$

Si la serie $Y_t$ tiene una componente estacional con período $s$ es posible eliminarla diferenciando una ó dos veces, con un rezago de orden $s$, es decir, si $D = 1, 2$, transformando $Y_t$ en la serie $W_t$:

$$W_t = \Delta_s^D Y_t. \tag{10.4}$$

En R la instrucción para calcular (10.4) es `diff(y,s,D)`. Se espera que la serie $W_t$ sea estacionaria en covarianza, con una posible estructura $SARMA$. En caso de existir una componente de tendencia y otra de estacionalidad es posible eliminar ambas mediante la

transformación

$$W_t = (1 - L^s)^D (1 - L)^d Y_t = \Delta_s^D \Delta^d Y_t \tag{10.5}$$

En R la transformación (10.5) se puede realizar con la instrucción siguiente

```
Wt = diff(diff(Yt,1,d),s,D)
```

La transformación (filtro lineal) $X_t = \Delta^d \Delta_s^D Y_t$ elimina la tendencia y la estacionalidad de $Y_t$, dejando una posible estructura $SARMA(p, q)(p_s, q_s)[s]$, definida en (8.40), pag. 149.

**Ejemplo 10.1.1.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t.$$

*En el Ejemplo 8.2.5, en la pag. 126, se ajustó un MA(6) a los residuos estimados $\widehat{\varepsilon}_t$. En el Ejemplo 8.5.3, en la pag. 153, se estimó un modelo alterno SARMA(2,1)(2,1)[4] para los residuos. En la Figura 10.1 se muestra el efecto de la diferenciación ordinaria,* `Ut = diff(Yt,1,d)`*, en el panel superior derecho. Luego el de la diferenciación estacional* `Wt = diff(Wt,s,D)`*, en el panel inferior derecho. El resultado es que la serie $W_t$ muestra una dinámica que puede corresponder a un proceso estacionario.*

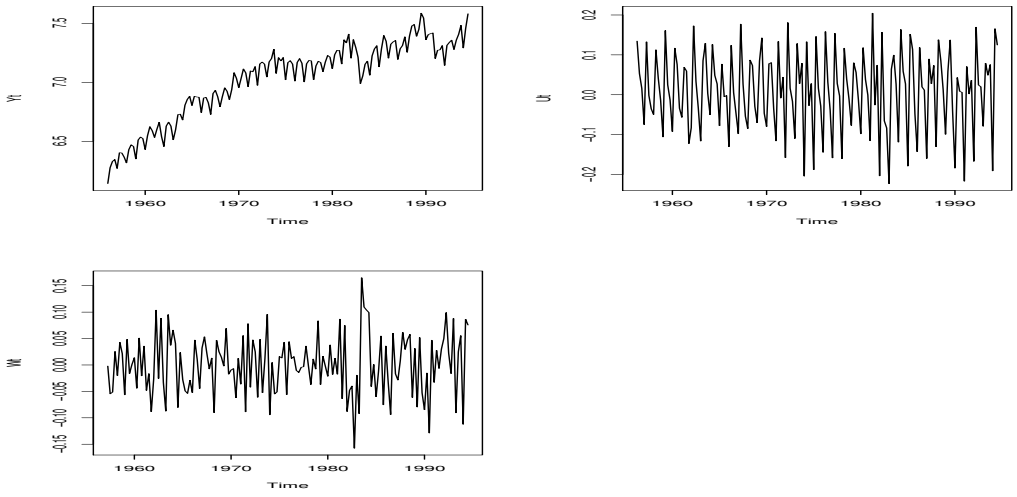



Figura 10.1: Aplicación de diferencia simple y luego diferencia estacional a la serie de producción de cemento

**Definición 10.1.1.** *Se dice que la serie* $(Y_t, t \in \mathbb{Z})$*, sigue un modelo* $SARIMA(p,d,q)(p_s,D,q_s)[s]$ *si cumple que*

$$\Phi_{p_s}(L^s)\Phi_p(L)\Delta_s^D\,\Delta^d\,Y_t = \Theta_{q_s}(L^s)\Theta_q(L)\varepsilon_t, \tag{10.6}$$

*donde* $d, D = 1, 2$*, y* $\varepsilon_t \sim RB(0,\sigma^2)$*. Y*

$$\Phi_p(L) = 1 - \sum_{j=1}^{p}\varphi_j L^j,$$

$$\Theta_q(L) = 1 + \sum_{j=1}^{q}\theta_j L^j,$$

$$\Phi_{p_s}(L^s) = 1 - \sum_{j=1}^{p_s}\varphi_{s,j} L^{js},$$

$$\Theta_{q_s}(L^s) = 1 + \sum_{j=1}^{q_s}\theta_{s,j} L^{js}.$$

Las notaciones anteriores no son las utilizadas en la mayoría de los textos. Están tomadas del libro Prado and West [2010]. La idea del modelo SARIMA es que la serie filtrada $W_t$ que satisface

$$\varphi_p(L)\Phi_P(L^s)W_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0,\sigma^2).$$

tiene un papel equivalente al residuo estructural $\varepsilon_t$ en el modelo de componentes

$$Y_t = \sum_{j=0}^{k}\beta_j t^j + \sum_{j=1}^{s-1}\delta_j I_j(t) + \varepsilon_t. \tag{10.7}$$

En (10.6), al diferenciar la serie $Y_t$, se eliminan la tendencia y estacionalidad aleatorias. Pero diferenciar en (10.7) para eliminarlas puede ser un procedimiento incorrecto ya que estas componentes son básicas para calcular los pronósticos. Este hecho llama la atención sobre la necesidad de realizar pruebas de hipótesis que permitar decidir entre ambos modelos cuando hay evidencia de estacionalidad.

**Ejemplo 10.1.2.** *Suponga un modelo* $SARIMA(1,1,0)(1,1,0)[12]$*, es decir,* $p = 1, d = 1, q = 0, p_s = 1, D = 1, q_s = 0, s = 12$*, dado por*

$$(1-\varphi_1 L)(1-\varphi_{12}L^{12})(1-L^{12})(1-L)Y_t = \varepsilon_t,$$

*Colocando* $X_t = (1-L)(1-L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ *se obtiene, de manera equivalente,* $(1-\varphi_1 L)(1-\varphi_{12}L^{12})X_t = \varepsilon_t$*. Cuando se desarrollan los polinomios autoregresivos del modelo se obtiene*

$$(1-\varphi_1 L - \varphi_{12}L^{12} + \varphi_1\varphi_{12}L^{13})X_t = \varepsilon_t,$$

*ó también,* $X_t = \varphi_1 X_{t-1} - \varphi_{12} X_{t-12} + \varphi_1 \varphi_{12} X_{t-13} + \varepsilon_t$.

*Note que este modelo se puede considerar como un* $AR(13)$ *con restricciones sobre ciertos coeficientes, que deben asumirse cero. Se podría indicar por* $AR([1, 12, 13])$. *El parámetro* $\varphi_1\varphi_{12}$ *se estima como un solo valor* $\varphi_{13}$ *y puede ser no significativo. En el caso que sea no significativo el modelo se dice sin interacción, y se expresa como*

$$(1 - \varphi_1 L - \varphi_{12} L^{12}) X_t = \varepsilon_t.$$

**Ejemplo 10.1.3.** *Consideremos el modelo del ejemplo anterior, con parámetros* $\varphi_1 = 0.8$, $\varphi_{12} = -0.137$, $\sigma^2 = 2$. *El siguiente código en R utiliza la función* `sarima.Sim()` *de la librería CombMSC para simular trayectorias de modelos SARIMA y así visualizar mejor esta clase de modelos.*

```
# simulacion sarima(1,1,0)(1,1,0)[12] con la libreria CombMSC
library(CombMSC)
library(forecast)
library(TSA)

ti=0.8; ti12=0.737;
sigma= sqrt(2)
y = sarima.Sim(n = 52, period =12,
model=list(order=c(1,1,0),ar = ti,
ma = NULL, sd = sigma),
seasonal=list(order=c(1,1,0),ar=ti12,ma = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)

dy = diff(diff(y,12,1),1,1)
layout(1:2)
ts.plot(y)
ts.plot(dy)

# identificar
auto.arima(dy)

# identificador armasubset() libreria TSA
res=armasubsets(y=dy,
nar=14,nma=14,
y.name='dy',
```

```
ar.method='ols')

plot(res)
```

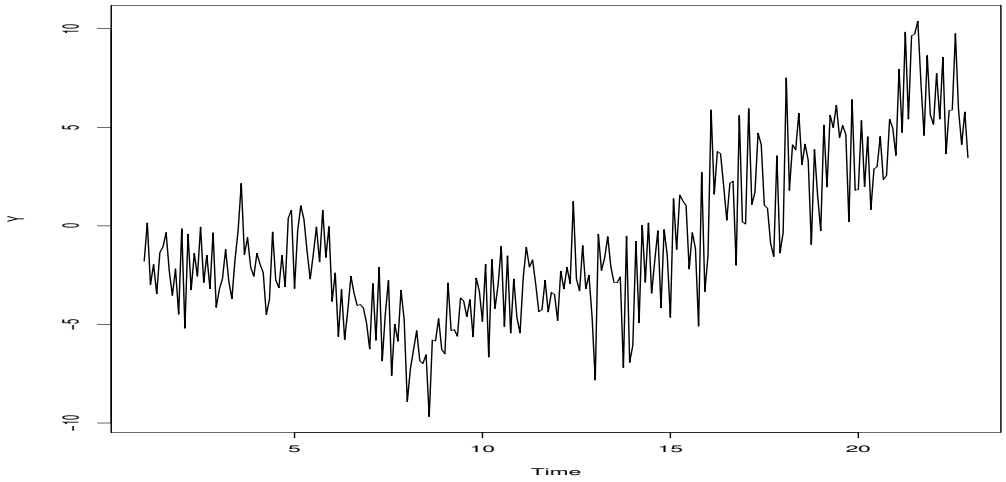



Figura 10.2: Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12]

**Ejemplo 10.1.4.** ***Modelo "Air-passengers"*** *Se define como un modelo multiplicativo de la forma* $SARIMA(0,1,1)(0,1,1)[12]$ *dado por*

$$(1-L)(1-L^{12})Y_t = (1+\theta_1)(1+\theta_{12}L^{12})\varepsilon_t, \qquad |\theta_1|<1,\ |\theta_{12}|<1. \tag{10.8}$$

*Es un modelo utilizado con frecuencia para modelar series con tendencia lineal y componente estacional. Por ejemplo la serie "airline", del texto de Box y Jenkins: "Monthly totals of international airline passengers, 1949 to 1960".*

**Nota sobre la identificación de este modelo**
*Al utilizar* $W_t = (1-L)(1-L^{12})Y_t$ *y examinar la fac de* $W_t$ *se observa que* $W_t = (1+\theta_1 L)(1+\theta_{12}L^{12})\varepsilon_t$ *es un* $MA(13)$. *Se puede comprobar que la fac de* $W_t$ *tiene valores diferentes de cero sólo en los rezagos 1,11,12,13.*
*La función de autocorrelación, fac, de* $W_t$ *es*

$$\begin{aligned}
\rho_0 &= (1+\theta_1^2)(1+\theta_{12}^2)\\
\rho_1 &= \theta_1(1+\theta_{12}^2)\\
\rho_{11} &= \theta_1\theta_{12}\\
\rho_{12} &= \theta_{12}(1+\theta_1^2)
\end{aligned}$$

*Suponga el modelo air-passengers con parámetros* $\theta_1 = -0.377$, $\theta_{12} = -0.572$, $\sigma^2 = 0.0014$*, el siguiente código simula una trayectoria de este modelo.*

```
#----- simular el modelo airline
ti=-0.377; ti12=-0.572;
sigma= sqrt(0.0014)

y = sarima.Sim(n = 12, period =12,
model=list(order=c(0,1,1),ma = ti,
ar = NULL, sd = sigma),
seasonal=list(order=c(0,1,1),ma=ti12,ar = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)

dy = diff(diff(y,12,1),1,1)
layout(1:2)
ts.plot(y)
ts.plot(dy)

# identificar
auto.arima(dy)

# identificador armasubset() libreria TSA
res=armasubsets(y=dy,
nar=14,nma=14,
y.name='dy',
ar.method='ols')
layout(1:1)
plot(res)

# estimar
m1 = arima(y,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=12))
summary(m1)

# pronosticar
pr = predict(m1,n.ahead=12)$pred
```

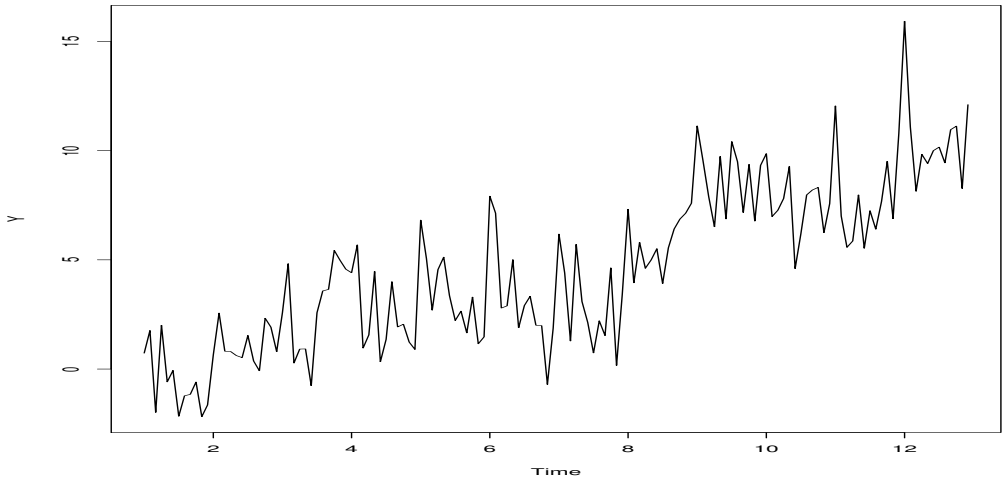



Figura 10.3: Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12]

**Ejercicio 10.1.1.** *Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA* $SARIMA(1,1,1)(1,1,1)[4]$ *definido por*

$$(1-\varphi_1 L)(1-\varphi_4 L^4)(1-L)(1-L^4)Y_t = (1-\theta_1 L)(1-\theta_4 L^4)\varepsilon_t.$$

*Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA con la serie* $Y_t$*. Es equivalente a estimar un modelo ARMA(1,1)(1,1) con la serie diferenciada,* $W_t = (1-L)(1-L^4)Y_t$*?. Use los siguientes valores para los parámetros.*

```
fi=-0.8; fi4=-0.37; ti=-0.64; ti4=-0.513; sigma= sqrt(0.014)
```

## 10.2. Pruebas de Raíz Unitaria Estacional

Se han considerado hasta el momento dos posibles modelos para una serie con componentes de tendencia y estacional:

1. Modelos con componentes determinísticas.
2. Modelos no estacionarios, integrados estacionalmente, es decir, $SARIMA(p,d,q)(p_s,D,q_s)[s]$, con $d, D = 0, 1$.

Se necesitan técnicas para decidir cuál modelo aplicar. Parte del desarrollo de la teoría de series de tiempo se ha dado en el área de las pruebas de raíces unitarias, como las pruebas

Dickey Fuller aumentadas, expuestas en el capítulo anterior. Pero también se has desarrollado pruebas para detectar raíces unitarias estacionales, como las pruebas HEGY, ver Hylleberg et al. [1990], y la prueba Canova-Hansen, indicada por CH, ver Canova and Hansen [1995], que se exponen a continuación,

Para introducir la idea de las pruebas de raíz unitaria estacional hay que recordar que una serie $Y_t \sim ARMA(p,q)$ se dice integrada, ó con una raíz unitaria si se cumple que el polinomio autorregresivo $\Phi_p(z)$ tiene una raíz igual a 1. Es decir, si se cumple que $\Phi_p(1) = 0$. En este caso el polinomio se puede factorizar como $\Phi_p(z) = (1-z)\Phi_{p-1}(z)$.

**Definición 10.2.1.** *Una serie de tiempo $Y_t \sim ARMA(p,q)$ se dice integrada estacional, de período $s$, con $s$ par, si se cumple que algunas ó todas las raíces de $z^s = 1$ son raíces del polinomio autorregresivo $\Phi_p(z)$.*

Las raíces de $z^s = 1$ se denominan las raíces de la unidad, y son $s$ números complejos, con $s = 4, 12$ por ejemplo. Para el caso $s = 4$ se tiene que las cuatro raíces de la unidad son $1, -1, i, -i$, donde $i = \sqrt{-1}$ es la unidad imaginaria.

Con la representación $z = re^{i\theta}$, con la fórmula de Euler $e^{i\theta} = cos(\theta) + i\sin(\theta)$, donde $r = |z|$ y $\theta$ es el ángulo que forma $|z|$ con el eje X, se pueden identificar las raíces unitarias con los ángulos $\theta = 2\pi j/s, j = 0, 1, \ldots, s-1$, ya que se tiene $r = 1$ en todos los casos, pues las raíces unitarias tienen módulo 1. Estos ángulos se denominan las "frecuencias estacionales".

Por ejemplo, para $s = 12$, las raíces son

$$\pm 1, \pm i, -(1 \pm i\sqrt{3})/2, (1 \pm i\sqrt{3})/2, -(\sqrt{3} \pm i)/2, (\sqrt{3} \pm i)/2.$$

y se identifican con los ángulos $\theta_j = 2\pi j/12, j = 0, 1, \ldots, 11$

## 10.3. La prueba Canova-Hansen

Los modelos de Descomposición con errores ARMA y los SARIMA se consideran, inicialmente, equivalentes. Pero, como señalan Canova and Hansen [1995, pag. 238],

> "Es difícil saber apriori cuál posibilidad [modelo] produce la mejor descripción de la serie. Algunas series muestran cambios en los patrones estacionales, por ejemplo, el consumo de energía eléctrica, las series de consumo e ingresos, la serie del producto interno bruto".

La prueba CH se basa en esta observación. La inestabilidad estructural de la componente estacional se toma como una característica equivalente a la existencia de una raíz unitaria estacional y opuesta a estacionario estacional.

En Ghysels and Osborn [2001] pag. 29 :

> "[...] for the sample sizes often observed in practice, it may be difficult to discriminate between deterministic seasonality and a seasonal unit root process."

pag. 42:

> "It is now well known that a series generated by a unit root process can wander widely and smoothly over time without any inherent tendency to return to its underlying mean value [...]. In the seasonal context, there are S unit root processes, none of which has an inherent tendency to return to a deterministic pattern. As a result, the values for the seasons can wander widely and smoothly in relation to each other [...]"

**Descripción de la Prueba**

El modelo que asume la prueba es un modelo sin tendencia y con componente estacional, descrita mediante funciones trigonométricas ó indicadoras, con el primer rezago de $y$ como variable exógena. Concretamente, la ecuación (29) de la página 243 en el artículo de Canova and Hansen [1995], es

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \, cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \operatorname{sen}\left(\frac{2\pi jt}{s}\right), \tag{10.9}$$

donde se asume $s$ es par, $t = 1, 2, \ldots$. Nótese que cuando $j = s/2$ entonces $\operatorname{sin}\left(\frac{2\pi jt}{s}\right) = \operatorname{sin}(\pi t) \equiv 0$. Luego, el número de parámetros en el modelo (10.9) es $s - 1$, el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores

$$\underline{\gamma} = (\gamma_1, \ldots, \gamma_{s/2})',$$
$$\underline{\gamma}^* = (\gamma_1^*, \ldots, \gamma_{s/2-1}^*)'.$$

La hipótesis nula en la prueba CH es que los coeficientes $\underline{\gamma}, \underline{\gamma}^*$ no cambian con $t$ versus que varían con $t$, según el modelo de marcha aleatoria:

$$\underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \tag{10.10}$$

$$\underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t, \tag{10.11}$$

donde $\underline{u}_t, \underline{v}_t$ son vectores iid, independientes de media cero. Esta hipótesis alterna es la forma en la que se establece la presencia de raíces unitarias estacionales, equivalente a inestabilidad estructural en la componente estacional. Las hipótesis de la prueba se pueden escribir como sigue.

$$\begin{aligned} H_0: &\quad \underline{\gamma}_t \equiv \underline{\gamma}, \underline{\gamma}_t^* \equiv \underline{\gamma}^*, \\ H_a: &\quad \underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t. \end{aligned}$$

En Canova and Hansen [1995, pag. 240, sec. 2.2], se re-escribe la prueba con base en cierto parámetro $\tau^2$, colocando $H_0 : \tau^2 = 0$, versus $H_a : \tau^2 > 0$. El estadístico de la prueba $L$, es un tipo de estadístico denominado Multiplicador de Lagrange (LM).

$$LM = \frac{1}{T^2}\, tr\left( (A'\Omega A)^{-1} A' \left( \sum_{j=1}^{T} \underline{F}_j \underline{F}_j' \right) A \right) \tag{10.12}$$

requiere resultados avanzados para su definición, por lo que no se considerarán aquí y se remite al artículo original. Además, en el artículo se proveen los valores críticos para varios niveles de significación. La nula se rechaza si el valor observado del estadístico es mayor que el valor crítico

$$\widehat{L} > L_\alpha.$$

## Implementación en R

La prueba Canova-Hansen está implementada en la librería `uroot`, con la función `CH.test`, que tiene los parámetros siguientes

```
ch.test(x, type = c("dummy", "trigonometric"),
lag1 = FALSE, NW.order = NULL,   sid = NULL,
xreg = NULL, pvalue = c("RS", "raw"),
rs.nobsreg = 13),
```

la serie debe estar en la variable $x$, y debe haberse declarado como objeto "ts", con `frequency=4,7,12,...`.

El parámetro `type = c("dummy", "trigonometric")` indica si se utilizan variables trigonométricas ó indicadoras para modelar la estacionalidad.

El parámetro `lag1` es variable lógica. Si TRUE se incluye el primer rezago de la serie como variable explicativa en la regresión. Por defecto se coloca FALSE.

El parámetro `NW.order` es un entero que determina cuántos rezagos se utilizan para la estimación de la matriz de covarianza en la fórmula de Newey y West.

El parámetro `sid` es un vector numérico optional que especifica cuáles frecuencias deberían incluírse en la prueba. Por defecto se escogen todas las frecuencias incluyendo la cero.

El parámetro `xreg` es un vector ó matrix opcional de variables exógenas con el mismo número de filas que la serie x.

Nótese, sin embargo, que para aplicar la prueba CH es necesario eliminar la tendencia lineal de la serie $Y_t$, en caso de existir. Para aplicar la prueba se prefirió filtrar previamente esta tendencia, colocando $Y_t - \widehat{T}_t = \widehat{S}_t + \widehat{\varepsilon}_t$, donde $\widehat{T}_t$ se estimó mediante el filtro `stl()` ó Loess.

**Ejemplo 10.3.1.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral, para decidir entre los modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA, se aplicará la prueba Canova-Hansen para determinar la posibilidad de una raíz unitaria estacional que apoye un posible modelo SARIMA.*

*La prueba se implementa con los comandos siguientes. Nótese que se utilizó el procedimiento `stl()` para extraer la tendencia de la serie.*

```
#-----implementacion de la prueba Canova-Hansen
require(uroot)
#-----eliminar la tendencia con stl()
m1 = stl(y, s.window = 'per')
s1 = m1$time.series[,1]
t1 = m1$time.series[,2]
e1 = m1$time.series[,3]
y1 = s1+e1

res = ch.test(y1, type = "dummy", sid = 7, pvalue='raw')
res
----------------------
Canova and Hansen test for seasonal stability
     statistic pvalue
[1,]    0.3391 0.1126
```

Como el valor p es mayor de 0.05, luego no se rechaza la hipótesis nula y no se detecta una raíz unitaria estacional, que correspondería a inestabilidad estructural en la componente estacional. Por tanto, el modelo SARIMA no sería preferible para modelar la serie.

## 10.4. La prueba HEGY

La prueba HEGY (Hylleberg, Engle, Granger, Yoo, también indicada EGHY), en Hylleberg et al. [1990], es una generalización de la prueba aumentada de Dickey-Fuller. Permite decidir cuáles de las frecuencias estacionales corresponden a raíces estacionales significativas. En caso de detectarse al menos una, exceptuando la frecuencia que corresponde a cero, se concluirá que la serie en cuestión está integrada estacionalmente, que es la hipótesis nula.

Esta sección describe la idea básica de la prueba, que utiliza una regresión lineal múltiple. Los estadísticos para detectar las raíces unitarias estacionales son tipo t-Student y F, correspondientes a los estadísticos para pruebas de hipótesis sobre los parámetros de la regresión. Esta sección es con base en el artículo de Hylleberg et al. [1990].

Inicialmente se asume que la serie $Y_t$ tiene una componente estacional de período s par, y se puede aproximar por un proceso AR(s), tal que

$$\Phi_s(L)Y_t = \varepsilon_t,$$

para $\varepsilon_t$ ruido blanco. Se asume que están dados s números complejos $\theta_j \in \mathbb{C}$, $j = 1, \ldots, s$, en los cuales $\Phi_s(\theta_j)$ es un número finito. Estos s números son las s raíces estacionales. Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - z/\theta_k, \\ \Delta(z) &= \prod_{k=1}^{s} \delta_k(z), \\ \lambda_k &= \Phi_s(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k). \end{aligned} \tag{10.13}$$

Entonces se cumple, por una identidad debida a Lagrange, que

$$\Phi_s(z) = \sum_{k=1}^{s} \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\Phi_s^*(z), \tag{10.14}$$

donde $\Phi_s^*(z)$ es un polinomio posiblemente infinito ó racional. Nótese que, por la definición (10.13) se cumple

$$\Phi_s(\theta_k) = 0 \Leftrightarrow \lambda_k = 0 \tag{10.15}$$

por lo tanto, para determinar si $\theta_k$ es raíz estacional se debe comprobar $\lambda_k = 0$. Pero estas $\lambda_k$ se pueden asimilar a los coeficientes en una regresión lineal múltiple que se define a partir de la identidad (10.14). Y por tanto, probar $\lambda_k = 0$ se puede hacer, en principio, mediante una prueba t-Student de significación de parámetros, ó con una prueba F.

Por ejemplo, para el caso $\Phi_s(z) = 1 - z^4$, donde $\theta_1 = 1$, $\theta_2 = -1$,$\theta_3 = i$,$\theta_4 = -i$, la identidad (10.14) se transforma en

$$\begin{aligned}\Phi_s(z) = \lambda_1 z(1+z)(1+z^2) + \lambda_2(-z)(1-z)(1+z^2)\\ +\lambda_3(-iz)(1-z)(1+z)(1-iz) + \lambda_4(iz)(1-z)(1+z)(1+iz)\\ +\Phi_s^*(z)(1-z^4).\end{aligned} \tag{10.16}$$

Si se definen nuevas constantes $\pi_1, \pi_2, \pi_3, \pi_4$ mediante las relaciones:

$$\pi_1 = -\lambda_1, \pi_2 = -\lambda_2, 2\lambda_3 = -\pi_3 + i\pi_4, 2\lambda_4 = -\pi_3 - i\pi_4$$

entonces la identidad (10.16) queda

$$\begin{aligned}\Phi_s(z) &= -\pi_1 z(1+z+z^2+z^3) - \pi_2(-z)(1-z+z^2-z^3)\\ &\quad -(\pi_4+\pi_3 z)(-z)(1-z^2) + \Phi_s^*(z)(1-z^4).\end{aligned} \tag{10.17}$$

Finalmente, en $\Phi_s(L)Y_t = \varepsilon_t$ se reemplaza $\Phi_s(L)$ por la correspondiente expresión obtenida en (10.17). Si se definen las variables

$$\begin{aligned}X_{1,t} &= (1+L+L^2+L^3)Y_t,\\ X_{2,t} &= -(1-L+L^2-L^3)Y_t,\\ X_{3,t} &= -(1-L^2)Y_t,\\ X_{4,t} &= (1-L^4)Y_t,\end{aligned}$$

y se reemplaza $\Phi_s^*(L) = 1$, se obtiene la ecuación

$$X_{4,t} = \pi_1 X_{1,t} + \pi_2 X_{2,t} + \pi_3 X_{3,t} + \varepsilon_t. \tag{10.18}$$

La ecuación (10.18) es una regresión lineal múltiple. A partir de esta ecuación se plantea la prueba de hipótesis HEGY para el caso de período $s = 4$.

## Descripción de la Prueba

La hipótesis en la prueba HEGY, para el caso $s = 4$, se compone de tres hipótesis, para las raíces unitarias estacionales: $\pm 1, \pm i$, definidas de la manera siguiente.

$$H_0 : \Phi_s(1) = 0 \quad \Leftrightarrow \quad \pi_1 = 0, \;\; H_a : \Phi_s(1) > 0 \Leftrightarrow \pi_1 < 0,$$

$$H_0 : \Phi_s(-1) = 0 \quad \Leftrightarrow \quad \pi_2 = 0, \;\; H_a : \Phi_s(2) > 0 \Leftrightarrow \pi_2 < 0,$$
$$H_0 : |\Phi_s(i)| = 0 \quad \Leftrightarrow \quad \pi_3 = \pi_4 = 0, \;\; H_a : no(H_0).$$

Nótese que la raíz $z = 1$ corresponde a una raíz unitaria para tendencia aleatoria. Si esta hipótesis nula no se rechaza se asume que existe raíz unitaria ordinaria.

Los estadísticos de las pruebas son t-Student para $z = 1, -1$, y F para $\pm i$. Las decisiones se toman con base en los valores p correspondientes.

En el caso $s = 12$, hay 7 hipótesis, dos para $z = 1, -1$, y una para cada una de las cinco parejas de raíces unitarias estacionales conjugadas.

Nótese que no rechazar una de las hipótesis nulas equivale a aceptar que es una raíz unitaria estacional. En este caso hay que tener en cuenta la potencia de la prueba. Recordando que la potencia es la probabilidad: Prob(rechazar $H_0$ | $H_1$ es cierta), una baja potencia significa que la prueba no es capaz de detectar la alterna cuando ésta es cierta, o también, que la probabilidad de no rechazar $H_0$ cuando es falsa, es alta.

En caso de no rechazar una de las hipótesis nula, cabría esperar que el modelo adecuado para la series sea de tipo SARIMA integrado, es decir, después de diferenciar la serie se obtiene un proceso ARMA estacionario.

En caso de rechazar todas las hipótesis nulas la serie no tiene raíces unitarias estacionales y entonces se clasifica como una serie "estacionaria estacional", es decir, se puede modelar como un modelo de tendencia lineal más componentes estacionales con residuos posiblemente SARMA.

**Implementación en R**

La prueba HEGY está implementada en la función `hegy.test`, de la librería `uroot`, ver López-de LaCalle [2017]. Al colocar

```
res = hegy.test(x,
deterministic = c(1,0,0),
lag.method = c("fixed", "AIC", "BIC", "AICc"),
maxlag = 0,
pvalue = c("RS", "bootstrap", "raw"),
rs.nobsreg = 15,
boot.args = list(seed = 123,
lag.method = lag.method[1], maxlag = maxlag,
byseason = FALSE, nb = 1000, BTdim = c(100, 10), debug.tid = -1))
```

la serie debe estar en la variable $y$, y debe haberse declarado como objeto "ts". El parámetro `deterministic=c(1,1,1)` se refiere a si se desea incluír en el modelo un intercepto, tendencia lineal y componente estacional con variables indicadoras. El modelo se define en el caso que se anota como

$$\Phi_s(L)(Y_t - (\beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_j(t))) = \varepsilon_t. \tag{10.19}$$

El parámetro `pvalue`, según la ayuda:

```
a character specifying the method employed to compute
p-values: "RS", the default, interpolation based on
response surface regressions; "bootstrap", bootstrap;
"raw" interpolation in the tables provided in the
reference papers.
```

Dos ejemplos de la ayuda:

```
hegy.test(x, deterministic = c(1,1,1), lag.method = "fixed", maxlag = 1)
hegy.test(x, deterministic = c(1,1,1), lag.method = "fixed", maxlag = 1,
  pvalue = "bootstrap")
```

**Ejemplo 10.4.1.** *Retomando el Ejemplo (10.3.1) en la pag. 183, para modelación de la serie de producción de cemento Portland trimestral, donde se aplicó la prueba Canova-Hansen, se va a aplicar ahora la prueba HEGY con el mismo objetivo de determinar si el modelo SARIMA es adecuado. Los comandos en R son los siguientes.*

```
# prueba hegy con intercepto, tendencia y variables indicadoras.
require(uroot)
hegy.out1 = hegy.test(x=y,
deterministic = c(1,1,1),
lag.method = "fixed", maxlag = 1)
hegy.out1
statistic p-value
t_1      -4.153  0.0046 **
t_2     -3.1308  0.0206 *
F_3:4   14.9083       0 ***
F_2:4    12.533       0 ***
F_1:4   15.6967       0 ***
```

```
Deterministic terms: constant + trend + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions
```

El resultado anterior muestra que se rechaza la hipótesis nula de raíces unitarias en $1, -1, \pm i$, indicadas por `tpi_1, tpi_2, Fpi_3:4`. La última corresponde a las raíces en las frecuencias $\pi/2, 3\pi/2$, la primera corresponde a la frecuencia cero, y es una raíz unitaria para la tendencia, similar a la prueba Dickey-Fuller. Luego, no sería apropiado ajustar un modelo SARIMA a la serie.

Si se repite la prueba con la serie sin tendencia, rechaza nuevamente la hipótesis de raíces unitarias estacionales.

```
# prueba hegy sin tendencia.
require(uroot)
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
s1 = m1$time.series[,1]
t1 = m1$time.series[,2]
e1 = m1$time.series[,3]
y1 = s1+e1
hegy.out1 = hegy.test(x=y1,
deterministic = c(0,0,1),
lag.method = "fixed", maxlag = 1)
hegy.out1
      statistic p-value
t_1     -5.4194       0 ***
t_2      -3.064  0.0257 *
F_3:4   14.9748       0 ***
F_2:4    12.475       0 ***
F_1:4   19.6446       0 ***
Deterministic terms: constant + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions
```

## 10.5. Prueba OCSB

La prueba Osborn, Chui, Smith, And Birchenhall para raíz unitaria estacional. Ver Osborn et al. [1988] y Ghysels and Osborn [2001] pag. 66. La prueba se basa en la regresión

$$\Delta\Delta_s Y_t = \beta_1 \Delta_s Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t. \tag{10.20}$$

Nótese que, utilizando la identidad

$$\Delta_s = \Delta(I + L + \ldots + L^{s-1}),$$

se tiene

$$\Delta_s Y_{t-1} = (I + L + \ldots + L^{s-1})\Delta Y_{t-1}$$

y todas la variables en la regresión (10.20) incluyen la primera diferencia $\Delta$. Por tanto, como afirman Ghysels and Osborn [2001] pag. 66:

> An interpretation of the OCSB test, therefore, is that it tests seasonal integration in a process that has been first differenced.

En Rodrigues and Osborn [1999] se propone la regresión

$$\Delta_s Y_t = \beta_1 (I + L + \ldots + L^{s-1}) Y_{t-1} + \beta_2 Y_{t-s} + \epsilon_t. \tag{10.21}$$

como vehículo para la prueba de hipótesis nula de raíz unitaria estacional, incluyendo la raíz unitaria ordinaria. La expresión de la nula y la alterna en Osborn et al. [1988], pag. 364 es

$$H_0 : Y_t \sim I(1,1) \tag{10.22}$$

$$H_1 : Y_t \sim I(0,0) \vee I(1,0) \vee I(0,1) \tag{10.23}$$

El estadístico de la prueba se define como el estadístico ADF $t_{\widehat{\beta}_2}$, con respecto a la regresión (10.21). La hipótesis equivale en este caso a $H_0 : \beta_2 = 0$, vs. $H_1 : \beta_2 < 0$.

Se rechaza la nula cuando el estadístico es menor que el valor crítico $t_{\widehat{\beta}_2} < t_\alpha$.

Además, el estadístico de la prueba ADF $t_{\widehat{\beta}_1}$ se puede utilizar para la hipótesis de raíz unitaria ordinaria $H_0 : Y_t \sim I(1)$. La hipótesis equivale en este caso a $H_0 : \beta_1 = 0$, vs. $H_1 : \beta_1 < 0$.

La distribución asintótica de estos estadísticos bajo $H_0$ requiren conceptos avanzados y no se incluyen aquí.

Pero ambas pruebas no pueden equipararse a la prueba Dickey-Fuller, porque, como Ghysels and Osborn [2001] pag. 66 observan, los valores críticos ADF no son estrictamente válidos

para éstas, en consecuencia para la prueba OCSB se requieren valores críticos calculados por simulación.

La prueba OCSB está implementada en la librería `forecast`, con la función `ocsb.test`. Los valores críticos en esta implementación se calculan mediante simulación. La aplicación se hace con la instrucción `ocsb.test(x)`. Esta función calcula solamente $t_{\widehat{\beta}_2}$ y el valor crítico para el nivel de significación de 5 %. Obsérvese en este Ejemplo, que se rechaza la nula de raíz unitaria estacional si la serie está sin tendencia, y no se rechaza si se incluye ésta.

**Ejemplo 10.5.1.** *Retomando los Ejemplos (10.3.1) y (10.4.1) en las pags. 183 y 187, para modelación de la serie de producción de cemento Portland trimestral, donde se aplicaron las pruebas Canova-Hansen y HEGY, se va a aplicar ahora la prueba OCSB con el mismo objetivo de determinar si el modelo SARIMA es adecuado. Los comandos en R son los siguientes.*

```
#----------serie con tendencia y estacionalidad
require(forecast)
ocsb.test(y)
Test statistic: -2.092, 5% critical value: -1.8927
alternative hypothesis: stationary
#--------------eliminar tendencia
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
s1 = m1$time.series[,1]
t1 = m1$time.series[,2]
e1 = m1$time.series[,3]
y1 = y-y1
ocsb.test(y1)
Test statistic: -2.1177, 5% critical value: -1.8927
```

En el primer caso el estadístico observado es menor que el valor crítico por lo que rechaza la nula $H_0 : Y_t \sim I(1,1)$. Y por tanto acepta una de las alternas $H_1 : Y_t \sim I(0,0) \vee I(1,0) \vee I(0,1)$.

Al eliminar la tendencia, se rechaza la nula $H_0 : Y_t \sim I(1,1)$. Como no hay tendencia, la alterna debe ser $Y_t \sim I(0,0)$.

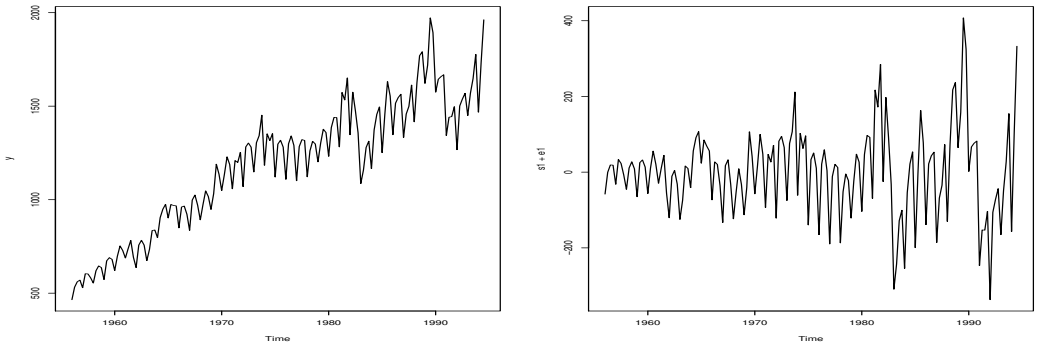
### 10.5.1. Ejemplo de resumen

**Ejemplo 10.5.2.** *Retomando el Ejemplo (7.2.2) en la pag. 104, para modelación de la serie de producción de cemento Portland, trimestral, para decidir entre los modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA.*

$$1)\ Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{3} \delta_j I_t(t) + \varepsilon_t,\ \varepsilon_t \sim MA(6), \tag{10.24}$$

$$2)\ Y_t = SARIMA(3,1,2)(1,1,2)[4]. \tag{10.25}$$

En la Figura 10.4 se puede ver la serie con y sin tendencia. Se observa que la componente estacional no es constante, por lo que es posible que se tenga un modelo SARIMA (10.25) en lugar del modelo de componentes (10.24). Al examinar la fac de la serie diferenciada

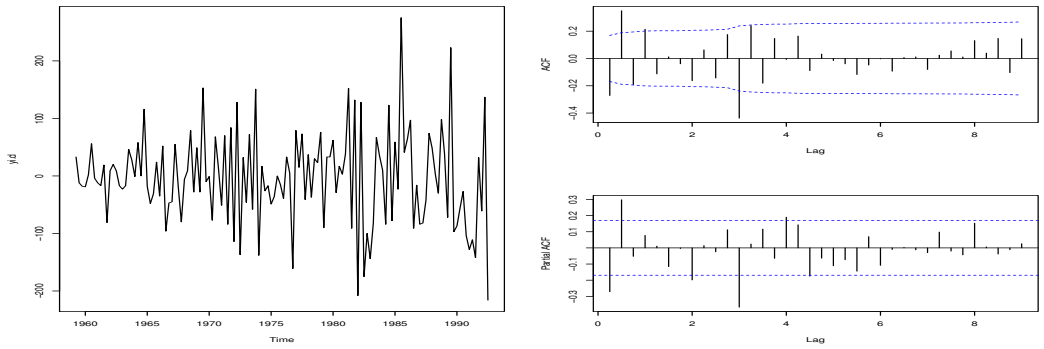



(a) serie con tendencia

(b) serie sin tendencia

Figura 10.4: Serie de Producción Trimestral de Cemento, Australia

$W_t = (1-L)(1-L^4)Y_t$, para $Y_t$ la serie excluyendo los últimos 8 datos para realizar validación cruzada, se observa posibles modelos ARMA, por ejemplo, con rezagos 1, 12 y 13 significativos, según la Figura 10.5.




(a) $W_t = (1-L)(1-L^4)Y_t$ (b) fac y facp de $W_t$

Figura 10.5: Serie Diferenciada de Producción Trimestral de Cemento

Mediante el uso de la función `armasubSets()` de la librería `TSA`, aplicada a la serie $W_t$ se llega al modelo `SARIMA(3,1,2)(1,1,2)[4]`. Los parámetros estimados están en la Tabla 10.1.

Tabla 10.1: Parámetros estimados de SARIMA(3,1,2)(1,1,2)[4]

| | parametros | sd.dev | est t |
|---|---|---|---|
| ar1 | 0.55 | 0.30 | 1.85 |
| ar2 | 0.57 | 0.20 | 2.82 |
| ar3 | -0.31 | 0.12 | -2.60 |
| ma1 | -0.73 | 0.32 | -2.28 |
| ma2 | -0.23 | 0.31 | -0.73 |
| sar1 | 0.90 | 0.25 | 3.56 |
| sma1 | -1.56 | 0.32 | -4.83 |
| sma2 | 0.67 | 0.18 | 3.80 |

La prueba Ljung-Box para los residuos de este modelo arroja el resultado
`X-squared = 34.672, df = 26, p-value = 0.1189`,
por lo que puede aceptarse este modelo para pronosticar. Los pronósticos con el modelo SARIMA versus los obtenidos con el modelo componentes (10.24) se pueden observar en la Figura 10.6. Se llegó a que el SARIMA no se detecta como modelo factible por las prueba CH, HEGY y OCSB. El MAPE de sus pronósticos, a 8 trimestres, es 7.89, mayor que el

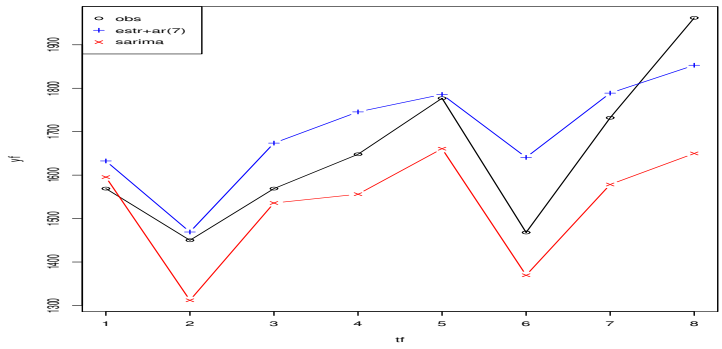



Figura 10.6: Pronósticos a 8 trimestres de la Producción Trimestral de Cemento

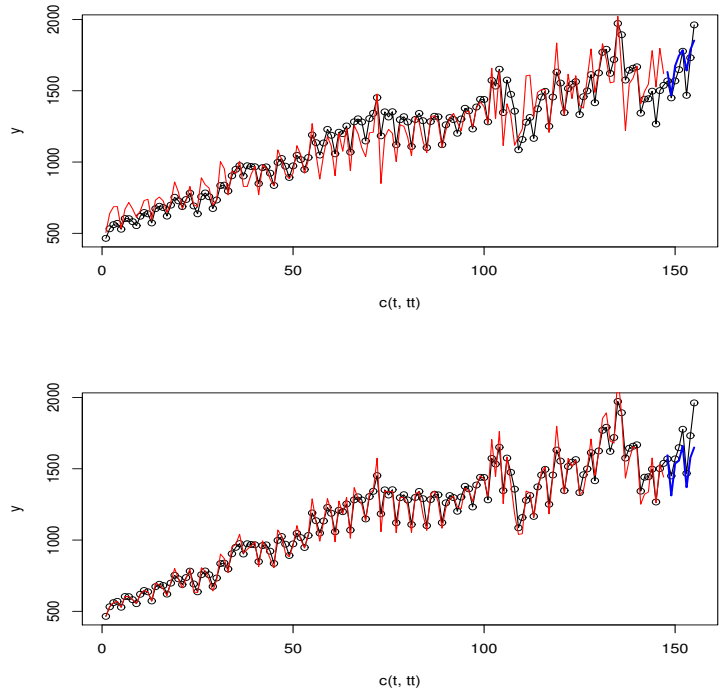



Figura 10.7: Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA

del modelo de Descomposición con errores MA(6), que es 4.64. Sin embargo, el modelo SARIMA tiene mejor ajuste dentro de la muestra. En la Figura 10.7 se puede comparar tanto los ajustes dentro de la muestra como los pronósticos.

Cuál modelo escoger?. Desde el punto de vista de la validación cruzada a 8 trimestres, el modelo de Descomposición. Desde el punto de vista del ajuste dentro de la muestra, el modelo SARIMA.

# CAPÍTULO 11

# Modelos de Rezagos Distribuídos y Cointegración

## 11.1. Modelo de rezagos distribuídos

El objetivo de este capítulo es desarrollar la idea de la regresión lineal entre dos ó más procesos estacionarios. Si se consideran $(X_n, Y_n, Z_n, n \in \mathbb{Z})$ tres procesos estacionarios en covarianza, de media cero, y $(\beta_j, j = 0, 1, \ldots) \in \ell_1$ es una sucesión de reales con $\beta_0 = 1$ entonces se puede definir el proceso estacionario en covarianza $D_n = \sum_{j=0}^{\infty} \beta_j X_{n-j}$, y definir un modelo de regresión lineal de la forma

$$Y_n = \alpha_0 + \alpha_1 D_n + Z_n, \tag{11.1}$$

para $\alpha_0, \alpha_1$ parámetros. Reemplazando $D_n$ en función de $(X_n, n \in \mathbb{Z})$ se obtiene

$$Y_n = \alpha_0 + \alpha_1 \sum_{j=0}^{\infty} \beta_j X_{n-j} + Z_n, \tag{11.2}$$

donde los $(\beta_j)$ son también parámetros. Si se tienen muestras finitas de los procesos $(Y_n, X_n)$ entonces no es posible estimar el conjunto infinito de parámetros en (11.2) por lo que se hace necesario introducir algunas restricciones. Una de éstas consiste en asumir que la función de transferencia $\beta(z) = \sum_{j=0}^{\infty} \beta_j z^j$ se puede reemplazar (aproximar) por un filtro causal tipo ARMA(p,q) de la forma

$$\beta(z) = \frac{\Theta_q(z)}{\Phi_p(z)} = \frac{1 + \sum_{j=1}^{q} \theta_j z^j}{1 - \sum_{j=1}^{p} \varphi_j z^j}, \tag{11.3}$$

para $|z| < 1$, donde $\Phi_p(z) \neq 0, \Theta_q(z) \neq 0 \; \forall z, |z| \leq 1$. También en (11.2) se asume que $(Z_n) \sim ARMA(p_z, q_z)$ y $(X_n) \sim ARMA(p_x, q_x)$, incluyendo la posibilidad de S-ARMA estacional, son independientes. Entonces (11.2) se escribe

$$Y_n = \alpha_0 + \alpha_1 \frac{\Theta_q(L)}{\Phi_p(L)} X_n + Z_n. \tag{11.4}$$

Si se redefinen los parámetros $\theta_j$ de tal forma que $\alpha_1(1 + \sum_{j=1}^{q} \theta_j) \equiv \sum_{j=0}^{q} \theta_j$ y se reemplaza $X_n$ por $X_{n-b}$, donde $b = 0, 1, \ldots$, entonces (11.4) queda

$$Y_n = \alpha_0 + \frac{\Theta_q(L)}{\Phi_p(L)} X_{n-b} + Z_n. \tag{11.5}$$

El modelo (11.5) se define como "modelo de función racional con rezagos distribuídos, con intercepto". El rezago $X_{n-b}$ indica la demora que tarda en surtir efecto un cambio de las $X_n$ en las $Y_n$.

El principal problema del modelo (11.5) es la identificación de los órdenes $p, q, p_z, q_z$. Dos posibilidades consisten en: 1) el método de Box-Jenkins con base en pre-blanqueado y 2) búsqueda directa ensayando varios valores de $p, q, p_z, q_z$ y escogiendo el modelo de mínimo AIC. Esta última requiere un método de estimación. Los métodos de estimación se presentan más adelante.

**Método de identificación de Box-Jenkins**

El método se describe en cuatro pasos. A partir del modelo (11.5) se coloca $\alpha_0 = 0$ y se re-definen los coeficientes $\beta_j$ mediante la identidad

$$\frac{\Theta_q(z) z^b}{\Phi_p(z)} = \sum_{j=0}^{\infty} \beta_j z^j = \beta(z). \tag{11.6}$$

Los pasos son los siguientes; más adelante se dan las correspondientes instrucciones en R.

**Paso 1**. Se estima un modelo para $(X_n) \sim ARMA(p_x, q_x)$ tal que

$$X_n = \frac{\Theta_{q_x}(L)}{\Phi_{p_x}(L)} U_n, \tag{11.7}$$

donde $(U_n) \sim RB(0, \sigma_u^2)$. Se calcula el filtro causal inverso

$$\nu(L) = \frac{\Phi_{p_x}(L)}{\Theta_{q_x}(L)} = 1 + \nu_1 L + \nu_2 L^2 + \ldots \tag{11.8}$$

**Paso 2**. Se aplica el filtro inverso (11.8) a izquierda en la ecuación (11.5). Se calculan $W_n = \nu(L) Y_n$ y $U_n = \nu(L) X_n$. Definiendo $V_n = \nu(L) Z_n$ se obtiene

$$W_n = \beta(L) U_n + V_n. \tag{11.9}$$

En esta ecuación $(V_n)$ y $(U_n)$ son procesos independientes por hipótesis, ya que $(Z_n)$ y $(X_n)$ lo son; y son observables (calculables), en cambio $V_n$ no lo es, pero esto no causa inconveniente debido a que se elimina al tomar esperanzas en el paso siguiente.

**Paso 3**. Se multiplica a derecha ambos miembros de la ecuación (11.9) por $U_{n-k}$, para $k = 0, 1, 2, \ldots$, y luego se toma esperanza. En el miembro de la izquierda se obtiene la covarianza cruzada $R_{w,u}(k) = Cov(W_n, U_{n+k}) = \mathbb{E}(W_n U_{n+k})$. Se tiene además

$$\begin{aligned} R_{w,u}(k) &= \mathbb{E}(\beta(L)U_n U_{n-k}) + \mathbb{E}(V_n U_{n-k}) \\ &= \sum_{j=0}^{\infty} \beta_j \mathbb{E}(U_n U_{n-k}) \\ &= \beta_k \sigma_u^2. \end{aligned}$$

Se estiman los $\beta_k$ mediante la relación

$$\widehat{\beta}_k = \frac{\widehat{R}_{w,u}(k)}{\widehat{\sigma}_u^2} = \frac{\widehat{\rho}_{w,u}(k)\widehat{\sigma}_w}{\widehat{\sigma}_u}. \tag{11.10}$$

**Paso 4**. Según (11.6), a cada par de órdenes $(p, q)$ y polinomios $\Phi_p(z), \Theta_q(z)$, le corresponde una sucesión $(\beta_k)$. E inversamente, dados los $(\widehat{\beta}_k)$ les debe corresponder ciertos $(p, q)$ y polinomios $\Phi_p(z)$, $\Theta_q(z)$. En la Figura 11.1 a la derecha de cada patrón $(\beta_k)$ se muestra el correspondiente modelo. Comparando $(\beta_k)$ con $(\widehat{\beta}_k)$, se elige el $(p, q)$ del caso con mejor correspondencia, y se puede iniciar el proceso de estimación.

**Implementación en R**

Para implementar el método de identificación Box-Jenkins en R es conveniente primero explicar las funciones: `arima()`, `arima.sim()`, `poly.calc()`, `ARMAtoMA()` e `ImpulseVMA`.

**La función arima**. Estima modelos SARIMA. La especificación del modelo a estimar depende de cómo R representa los procesos ARMA(p,q), que es como sigue.

$$Y_n = \varphi_0 + \sum_{j=1}^{p} \varphi_j Y_{n-j} + Z_n + \sum_{j=1}^{q} \theta_j Z_{n-j}. \tag{11.11}$$

Al aplicar `m1 = arima(y,order=c(p,0,q))` el objeto `m1` contiene un nodo (slot) denominado `coef` tal que `m1$coef` es el vector $(\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_1, \varphi_0)$.

**La función poly.calc**. Para definir un modelo ARMA estacionario en covarianza e invertible se puede comenzar definiendo las raíces de los polinomios autorregresivo, $\Phi_p(z)$ y de media móvil, $\Theta_q(z)$, de tal forma que se garantice que están por fuera del círculo unitario.

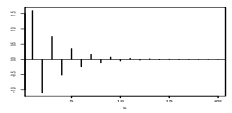
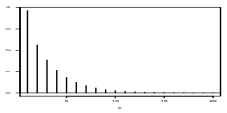
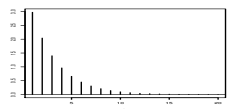
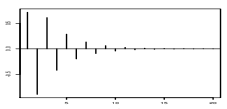
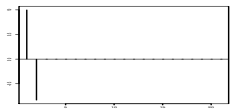
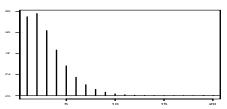
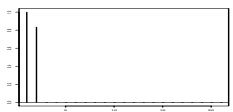
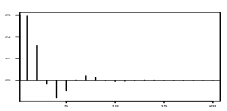
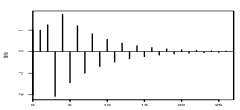
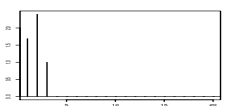
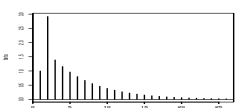
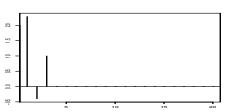
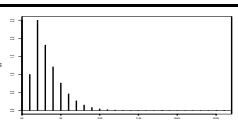
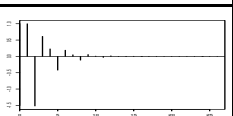
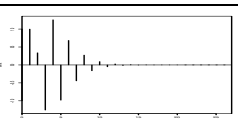
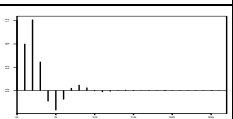
| $(p,q)$ | $\beta_j, j=0,1,\ldots$ | Modelo | $(p,q)$ | $\beta_j, j=0,1,\ldots$ | Modelo |
|---|---|---|---|---|---|
| (1,0) |  | $\frac{\theta_0}{1-\varphi_1 L}$ | (1,1) |  | $\frac{\theta_0+\theta_1 L}{1-\varphi_1 L}$ |
| (1,0) |  | $\frac{\theta_0}{1-\varphi_1 L}$ | (1,1) |  | $\frac{\theta_0+\theta_1 L}{1-\varphi_1 L}$ |
| (0,1) |  | $\frac{\theta_0+\theta_1 L}{1}$ | (2,0) |  | $\frac{\theta_0}{1+\varphi_1 L+\varphi_2 L^2}$ |
| (0,1) |  | $\frac{\theta_0+\theta_1 L}{1}$ | (2,0) |  | $\frac{\theta_0}{1-\varphi_1 L-\varphi_2 L^2}$ |
| (1,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L}$ | (0,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1}$ |
| (1,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L}$ | (0,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1}$ |
| (2,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L-\varphi_2 L^2}$ | (2,1) |  | $\frac{\theta_0+\theta_1 L}{1-\varphi_1 L-\varphi_2 L^2}$ |
| (2,2) |  | $\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L-\varphi_2 L^2}$ | (2,1) |  | $\frac{\theta_0+\theta_1 L}{1-\varphi_1 L-\varphi_2 L^2}$ |

Tabla 11.1: Modelos de Transferencia

Si el vector `r` contiene $p$ raíces, la función `a=poly.calc(r)` de la librería `polynom`, calcula los coeficientes de un polinomio `a`, con las raíces `r`. El objeto `a` es un vector de coeficientes correspondientes a las potencias de $z$ en orden ascendente. Por ejemplo, `-2.796 + 0.41*z + 0.4*z^2 + z^3`. Sin embargo, los polinomios $\Phi_p(z)$ y $\Theta_q(z)$ deben ser mónicos, es decir, el término constante debe ser la unidad. Si se normaliza escribiendo $a = a/a[1]$, el resultado es un polinomio mónico con raíces `r`, por ejemplo, `1 - 0.146*z - 0.143*z^2 - 0.357*z^3`. Si este polinomio fuera $\Phi_3(z) = 1 - \sum_{j=1}^{3} \varphi_j z^j$, el vector de los coeficientes $\varphi_j$ se obtienen como `-a[2:4]`. Si el polinomio fuera $\Theta_q(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$, el vector de coeficientes $\theta_j$ se obtiene como `a[2:4]`; el detalle es el signo menos en los coeficientes autorregresivos.

**La función arima.sim**. Si los polinomios `a=poly.cal(ra)`, `b=poly.cal(rb)` son los polinomios $\Phi_p(z)$ y $\Theta_q(z)$, respectivamente, se pueden simular n valores de un proceso ARMA utilizando `arima.sim()`,
`Xn = arima.sim(list(order=c(p,0,q),ar = -a[2:p],ma = b[2:q]), n=n,sd=sigma).`

**Las funciones ARMAtoMA, ImpulseVMA**. Dado el proceso $(X_n) \sim ARMA(p_x, q_x)$ en (11.7) para obtener su representación causal, dada por la sucesión $(\psi_j, j = 0, 1, 2, \ldots)$,

$$\frac{\Theta_{q_x}(z)}{\Phi_{p_x}(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \tag{11.12}$$

con $\psi_0 = 1$, se utilizan las funciones `ARMAtoMA` de la librería `stats` e `ImpulseVMA` de la librería `portes`. En adelante solamente se utilizará esta última. Si los vectores `phi,theta` corresponden a $(\varphi_1, \ldots, \varphi_p)$ y $(\theta_1, \ldots, \theta_1)$, respectivamente, se escribe `ImpulseVMA(phi,theta,m)`, y el parámetro `m` es el número de términos en el desarrollo (11.12).

Si se utiliza la función `arima()` para estimar un modelo ARMA, con
`m1 = arima(x,order=c(p,0,q))`, se calcula el filtro causal (11.12), de longitud $m$ $\underline{\psi} = (1, \psi_1, \ldots, \psi_m)$, con la instrucción

```
psi = ImpulseVMA(phi=-m1.coef[(p+1):(p+q)],
theta = m1.coef[1:p],Trunc.Series = m).
```

Si se utiliza la función `poly.cal(r)`, donde `a=poly.cal(ra)`, `b=poly.cal(rb)` son los polinomios $\Phi_p(z)$ y $\Theta_q(z)$, respectivamente, entonces se escribe
`psi=ImpulseVMA(phi=-a[2:p],theta=b[2:q],Trunc.Series = m)`

**La función filter**. Un paso importante en la implementación de la metodología Box-Jenkins es calcular el filtro causal inverso dado en (11.8)

$$\nu(L) = \frac{\Phi_{p_x}(L)}{\Theta_{q_x}(L)} = 1 + \nu_1 L + \nu_2 L^2 + \ldots$$

con la instruccion

```
nu=ImpulseVMA(phi=m1.coef[(p+1):(p+q)],
ma=m1.coef[1:p],lag.max=m).
```

El filtro $\nu(L)$ se aplica a $X_n$ y es inmediato que $\nu(L)X_n = U_n$. Esta operación se hace mediante la función `filter()` de la librería `stats`. Se escribe

`u=filter(x,rev(nu),"conv",sides=1,circular=TRUE)`. Un detalle importante es el uso de `rev(nu)`, que especifica se invierte el orden de la sucesión `nu`.

**Los Pasos para la metodología Box-Jenkins en R**

**Paso 1**. Para implementarlo en R se procede a estimar el modelo para $(X_n) \sim ARMA(p_x, q_x)$, luego se calculan los primeros valores del filtro $\nu(L)$ con la función `ARMAtoMA()` .

```
mx = arima(x,order=c(px,qx))
nu = ARMAtoMA(ar = mx$coef[(px+1):(px+qx)],
     ma = mx$coef[1:px], lag.max = 30)
```

La operación de "blanqueo" consiste en aplicar el filtro inverso $\nu(L)$ a ambos procesos $X_n$ y $Y_n$. El filtro (**??**) se define a partir de los polinomios $\Phi_p(z)$ y $\Theta_q(z)$ del modelo para $X_n$ estimado con `arima()`. **Paso 2**. Definiendo $W_n = \nu(L)Y_n, U_n = \nu(L)X_n$ y $V_n = \nu(L)Z_n$ se obtiene

En esta ecuación $(V_n)$ y $(U_n)$ son procesos independientes por hipótesis, ya que $(Z_n)$ y $(X_n)$ lo son. Además $(U_n) \sim RB(0, \sigma_u^2)$. Para calcular los procesos $W_n, U_n$ en R se utiliza la función `filter()` de la librería stats

```
w = filter(y,filter = nu, "conv", sides = 1, circular = TRUE)
u = filter(x,filter = nu, "conv", sides = 1, circular = TRUE)
```

**Paso 3**.La instrucción con R consiste en calcular solamente $\widehat{\rho}_{w,u}(k)$ y graficarla

```
acf(ts.intersect(w,u),na.action = na.pass)
```

## 11.2. *Cointegración

# APÉNDICE A

# Datos de Series de Tiempo

## A.1. Series con Tendencia y Estacionalidad

### A.1.1. Ejemplo 1. Cinco Series

Para cargar los datos en R simplemente copiar cada bloque “structure” en el programa. Las series tiene componente estacional y/o tendencia.

```
"X1" <-
structure(c(13630, 14109, 15945, 16637, 17136, 17701, 18894,
21479, 18297, 23610, 18546, 14781, 13975, 13468, 13490, 17759,
17889, 17298, 19260, 20123, 18777, 23709, 16981, 17134, 14197,
14749, 16270, 18028, 17939, 19218, 19683, 19711, 20396, 20659,
15398, 18549, 15441, 14844, 16610, 17958, 17945, 20473, 19482,
21151, 20952, 21222, 19489, 17968, 15818, 14423, 18010, 18544,
20123, 21131, 19371, 23507, 22605, 21042, 21547, 17417, 19043,
16207, 17941, 18075, 19914, 21450, 18702, 22845, 21272, 21430,
22094, 15988, 17670, 16825, 18226, 19497, 21406, 20008, 22962,
23116, 20327, 25791, 19850, 18185, 19292, 16934, 18314, 21890,
21460, 22641, 23237, 23495, 23948, 27086, 19827, 19827),
```

```
.Tsp = c(1990, 1997.91666666667, 12), class = "ts")

"X2" <-
structure(c(2993, 4579, 4138, 4062, 3206, 4041, 3784, 2754, 4037,
4318, 5069, 5353, 3302, 4651, 3614, 3558, 3013, 3814, 3821, 3331,
3931, 4598, 5432, 5233, 3374, 4587, 3943, 4306, 3643, 3563, 3759,
3085, 3317, 5335, 6067, 5163, 3255, 4048, 4154, 4117, 3537, 3478,
4006, 3604, 3682, 5180, 5274, 5205, 3456, 3849, 4064, 3966, 3604,
3980, 4385, 3664, 4290, 5814, 6037, 5492, 3492, 5612, 4319, 4078,
3924, 4319, 4390, 3545, 4184, 5316, 6330, 6349, 3390, 4550, 4060,
3530, 3458, 4153, 3797, 4069, 4000, 5424, 6257, 5316, 3522, 4520,
3390, 3609, 4060, 4266, 4588, 4182, 3551, 6231, 6668, 5921),
.Tsp = c(1990, 1997.91666666667, 12), class = "ts")

"X3" <-
structure(c(1574, 1368, 1387, 1109, 1257, 1376, 2143, 1208, 2007,
1876, 1702, 1819, 1802, 1205, 1684, 1682, 1991, 2394, 1914, 2499,
2130, 2529, 2328, 2076, 2496, 1647, 2518, 2205, 2395, 2891, 2712,
2427, 2477, 2860, 2505, 3355, 1760, 2318, 3111, 2570, 2868, 3042,
2749, 2839, 3140, 2909, 2982, 3667, 2814, 2732, 3265, 3166, 2792,
3742, 3099, 3278, 4120, 3553, 3675, 3799, 3427, 3234, 3733, 3642,
3553, 3647, 3624, 2973, 3597, 3731, 4092, 4100, 2762, 3953, 4152,
4229, 4419, 4774, 4313, 4060, 4664, 4374, 4419, 4908, 4321, 4772,
4361, 4969, 5111, 5014, 4858, 5159, 5086, 5379, 5605, 5269),
.Tsp = c(1990, 1997.91666666667, 12), class = "ts")

"X4" <-
structure(c(718, 659, 831, 645, 777, 637, 983, 894, 994, 1079,
1604, 681, 559, 606, 967, 900, 940, 876, 724, 1044, 964, 975,
969, 839, 750, 700, 752, 638, 976, 660, 1169, 866, 940, 1084,
653, 1150, 502, 828, 750, 911, 918, 1125, 1123, 1075, 1132, 1058,
1083, 1093, 763, 905, 1066, 841, 1085, 1080, 1180, 1382, 1170,
1399, 1023, 1402, 523, 1224, 1259, 1009, 1340, 1524, 1170, 1751,
1094, 1421, 1664, 1973, 962, 898, 1016, 1025, 1428, 1145, 1431,
1202, 1903, 1206, 1433, 1805, 635, 836, 912, 1085, 938, 1360,
1396, 1340, 1715, 1591, 998, 921),
.Tsp = c(1990, 1997.91666666667, 12), class = "ts")
```

```
"X5" <-
structure(c(856, 1695, 1747, 1845, 1323, 1618, 1542, 1341, 1704,
1437, 1896, 1881, 647, 1683, 1253, 1308, 1383, 1538, 1373, 1409,
1475, 1274, 1617, 1719, 779, 1583, 1616, 1692, 1507, 1568, 1500,
1545, 1373, 1599, 1962, 1827, 1340, 1293, 1465, 1799, 1550, 1374,
1834, 1707, 1623, 1629, 1672, 2062, 1281, 1523, 1693, 2225, 1352,
1625, 1972, 1617, 2461, 2090, 1320, 1740, 1031, 2326, 2063, 1776,
1380, 1749, 1671, 2324, 2116, 1215, 1493, 1663, 1090, 1901, 1840,
1773, 1817, 1632, 1048, 1009, 878, 1103, 1290, 1129, 766, 1426,
1276, 861, 1281, 1200, 1058, 1486, 1107, 1276, 1319, 1253),
.Tsp = c(1990, 1997.91666666667, 12), class = "ts")
```

### A.1.2. Ejemplo 2. Consumo domiciliario agua potable en Medellín

Son siete series que corresponden al consumo domiciliario mensual de agua potable, en metros cúbicos, en Medellín, entre 01/2002 y 02/2009, de los seis estratos (E1,...,E6).

```
E1 <- ts(c(
830585,862065,819126,846259,797484,822275,817202,808691,849281,785705,
816179,794372,810886,842975,808247,779136,812041,823208,859626,791757,
874626,821428,839239,866664,929256,930349,922565,945067,876293,914848,
901043,883872,914414,863295,869320,883188,876428,885178,844189,865512,
868821,843532,841804,845474,848497,839140,829033,835413,808128,850091,
815426,828197,837568,849059,849659,884204,851915,828371,853098,849168,
868531,894938,894784,849512,869826,851201,883266,910082,902557,885969,
830532,884486,901322,912116,878968,846227,882920,887664,868967,878299,
884827,881470,846974,881508,911565,895798),
frequency = 12, start = c(2002,1),end=c(2009,2))
```

```
E2 <- ts(c(
3971013,4047717,3800956,3996145,3746531,3878994,3881793,3842148,4004402,
3731894,3952091,3831181,3955677,4104482,3808791,3695691,3955968,3976956,
3952781,3782935,4030056,3864257,3842991,3820538,4113630,4119284,4111950,
4118195,3806820,3968917,4000465,3939983,4083424,3844038,3918498,3879706,
4110543,4014215,3913236,4010316,3909789,3846572,3961508,3913592,4011225,
3922224,3841417,3999491,3795465,4009919,3854934,3929686,3939443,3982917,
4014318,4153911,4013923,3905882,3982978,4053938,4109499,4235207,4253578,
4031997,3995562,3967422,4150281,4194999,4157934,4063502,4005597,4119769,
4249717,4282180,4026789,4142479,4092584,4129862,4025220,4107953,4165228,
4098281,3919459,4131321,4268906,4201312),
```

```
frequency = 12, start = c(2002,1),end=c(2009,2))

E3 <- ts(c(
4289646,4360063,4122174,4277444,4098109,4219165,4140381,4237069,4332006,
4090702,4296672,4152125,4197431,4444280,4182874,4008303,4273428,4337354,
4250530,4115744,4402179,4152760,4141799,4154357,4275861,4348270,4366593,
4359502,4075882,4178149,4206317,4228685,4352285,4118340,4268701,4133591,
4371138,4281707,4238128,4326347,4167633,4099662,4284719,4259002,4364109,
4295135,4186947,4323053,4122255,4283617,4204119,4298895,4269430,4291015,
4328754,4469938,4386520,4281562,4311298,4356524,4411549,4510562,4588383,
4369930,4250003,4271059,4462517,4487729,4489106,4381471,4195577,4439998,
4482539,4516261,4329306,4380152,4323603,4371817,4261146,4355593,4451290,
4340448,4162281,4352946,4395039,4399908),
frequency = 12, start = c(2002,1),end=c(2009,2))

E4 <- ts(c(
1217466,1301264,1223000,1245300,1234224,1247983,1213676,1299168,1263516,
1231687,1297885,1222159,1209971,1269712,1301712,1219499,1252521,1280500,
1277585,1223212,1353537,1262290,1249965,1266496,1247729,1274681,1334440,
1288604,1249738,1276211,1250433,1291328,1359462,1248971,1326074,1264522,
1288042,1288262,1310493,1314526,1269072,1272218,1286302,1339012,1343318,
1339787,1304250,1309399,1265588,1273902,1297944,1332145,1316248,1307335,
1299378,1373785,1359523,1346393,1305493,1379201,1341553,1342878,1416894,
1377080,1310540,1314538,1379672,1370980,1404261,1419119,1314071,1397678,
1392864,1374966,1416715,1367503,1344382,1374954,1371417,1374139,1438310,
1403824,1321147,1391531,1380644,1377416),
frequency = 12, start = c(2002,1),end=c(2009,2))

E5 <- ts(c(
 919003,1072792, 953995, 969484,1001430, 987649, 943763,1071933, 979635,
1004341,1058158, 964956, 959495, 992008,1060873, 976456,1017342,1005237,
1019765, 987224,1103186,1024675,1012451,1038227,,975603,,972791,1101383,
1033757,1007687,1029451,1019049,1154071,1105857,1027462,1095586,1007931,
1019537,1007079,1088062,1076923,1040349,1029928,1066408,1084946,1096110,
1116210,1059769,1079636,1033898, 974538,1062797,1113183,1051377,1061349,
1082328,1090139,1134724,1115018,1080466,1116532,1108965,1013549,1163244,
1125324,1037362,1060212,1124409,1091552,1140620,1150021,1030509,1125640,
1100176,1030624,1140577,1063250,1063201,1113295,1081073,1070065,1173631,
1121170,1056010,1098898,1056032,1048162),
frequency = 12, start = c(2002,1),end=c(2009,2))

E6 <- ts(c(
```

```
411310,562540,492259,488079,526026,519228,482174,577138,507632,520075,
546454,493409,492322,456760,561440,513925,521828,518497,530498,504558,
582742,542139,532877,558820,522660,459051,593756,550561,520271,551585,
525128,533818,593406,539314,602251,527376,532884,491869,587439,581779,
556217,548896,573440,556283,595247,605525,581950,586437,554534,476104,
582210,610106,568132,583566,617242,588585,623685,625737,617741,621826,
616176,521329,688246,651002,569869,603937,641515,621708,652043,689640,
592125,654952,645666,540491,675843,623267,613993,639755,627214,600945,
687069,685458,596244,661623,609800,555597),
frequency = 12, start = c(2002,1),end=c(2009,2))
```

# Bibliografía

ABRAHAM, B. J. AND R. A. LEDOLTER (2000): *Introduction to Time Series and Forecasting*, New York: John Wiley and Sons, 2nd ed.

ARAGON, Y. (2008): *Séries temporelles avec R Méthodes et Cas*, France: Springer Science + Business Media LLC.

BOX, G. AND G. C. TIAO (1975): "Intervention Analysis with Applications ot Economic and Environmental Problems," *JASA*, 70, 70–79.

BOX, G. E. P., G. M. JENKINS, AND G. C. REINSEL (1976): *Time Series Analysis, Forecasting and Control*, New York: Holden-Day, third edition ed.

BROCKWELL, P. J. AND R. A. DAVIS (2002): *Introduction to Time Series and Forecasting*, New York: Springer-Verlag, Inc., 2nd ed.

BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Society, B*, 37, 149–163.

CANOVA, F. AND B. HANSEN (1995): "Are seasonal patterns constant over time?. A Test for seasonal stability," *Journal of Business and Economics Statistics*, 13, 237–252.

CHU, C.-S. J., K. HORNIK, AND C.-M. KAUN (1995): "MOSUM tests for parameter constancy," *Biometrika*, 82, 603–617.

CLEVELAND, R. B., W. S. CLEVELAND, J. E. MCRAE, AND I. TERPPENING (1990): "STL: A Seasonal Trend Decomposition Procedure Based on Loess," *Journal of Official Statistics*, 6, 3–73.

COWPERTWAIT, P. AND A. METCALFE (2009): *Introductory Time Series Analysis with R*, New York: Springer Science + Business Media LLC, third edition ed.

CRYER, J. D. AND K.-S. CHAN (2008): *Time Series Analysis with Applications in R*, New York: Springer Science + Business Media LLC.

DIEBOLD, F. (1999): *Elementos de Pronóstico*, México: International Thomson Editores.

ELDER, J. AND P. E. KENNEDY (2001): "Testing for unit roots: what should student be taught?" *Journal of Economic Education*, 32, 137–146.

GHYSELS, E. AND D. R. OSBORN (2001): *The econometric analysis of seasonal time series*, Cambridge University Press.

GRANDELL, J. (2014): "Time series analysis," Lecture Notes,SF2943 Time Series Analysis.

HARVEY, A. (2006): "Forecasting with unobserved components time series models," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Amsterdam, The Netherlands: Elsevier North-Holland, chap. 7, 330–358.

HARVEY, A. C. AND J. DURBIN (1986): "The Effect of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modeling," *Journal of the Royal Statistical Society, A*, 149, 187–227.

HAYWOOD, J. AND J. RANDAL (2006): "Seasonality and structural breaks: NZ visitor arrivals and 9/11," Tech. rep., MSCS Research Report 06.

HYLLEBERG, S., R. ENGLE, C. GRANGER, AND B. YOO (1990): "Seasonal integration and cointegration," *Journal of Econometrics*, 44, 215–238.

HYNDMAN, R. J. AND A. B. KOEHLER (2006): "Another look at measures of forecast accuracy," *International journal of forecasting*, 22, 679–688.

HYNDMAN, R. J., A. B. KOEHLER, J. K. ORD, AND R. D. SNYDER (2008): *Forecasting with Exponential Smoothing. The State Space Approach*, Berlín, Heildelberg: Springer Science + Business Media LLC.

KLEINBAUM, D., L. KUPPER, AND K. MULLER (1988): *Applied regression analysis and other multivariable methods*, PWS-KENT Publishers Co.

LÓPEZ-DE LACALLE, J. (2017): *uroot: Unit Root Tests for Seasonal Time Series*, r package version 2.0-9.

OSBORN, D. R., A. P. CHUI, J. P. SMITH, AND C. R. BIRCHENHALL (1988): "Seasonality and the order of integration for consumption," *Oxford Bulletin of Economics and Statistics*, 50, 361–377.

POVEDA, G. (1982): "Curva Logística y Mínimos Cuadrados," *Lecturas Matemáticas*, 3, 111–132.

PRADO, R. AND M. WEST (2010): *Time series: modeling, computation, and inference*, CRC Press.

QIU, D. (2015): *rmaf: Refined Moving Average Filter*, r package version 3.0.1.

RODRIGUES, P. M. AND D. R. OSBORN (1999): "Performance of seasonal unit root tests for monthly data," *Journal of Applied Statistics*, 26, 985–1004.

RYAN, J. A. AND J. M. ULRICH (2017): *xts: eXtensible Time Series*, r package version 0.10-0.

SHEATHER, S. (2009): *A Modern Approach to Regression with R*, Springer Texts in Statistics, Springer New York.

SHUMWAY, R. H. AND D. S. STOFFER (2005): *Time Series Analysis and Its Applications. With R examples*, New York: Springer Science + Business Media LLC.

SOARES, L. J. AND M. C. MEDEIROS (2008): "Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data," *International Journal of Forecasting*, 24, 630–644.

VENABLES, W. AND B. RIPLEY (2002): *Modern Applied Statistics with S*, Statistics and Computing, Springer.

VINOD, H. D. (1973): "Generalization of the Durbin-Watson Statistic for Higher Order Autoregressive Processes," *Communications in Statistics*, 2, 115–144.

WINTERS, P. (1960): "Forecasting sales by exponentially weighted moving averages," *Management Science*, 6, 324–342.

ZEILEIS, A., K. HORNIK, C. KLEIBER, AND W. KRAEMER (2003): "Testing and Dating of Structural Changes in Practice," *Computational Statistics & Data Analysis*, 44, 109–123.

ZEILEIS, A., F. LEISCH, K. HORNIK, AND C. KLEIBER (2002): "strucchange: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7, 232–274.

# Índice alfabético


$AR(1)$, 108
$CUSUM$, 74
$R^2$, 28

AIC, 29
Algoritmo de Brockwell y Davis, 90

Banda de Bartlett, 102
bIC, 29

Componente Estacional, 35, 51
Componente Estructural, 55
Consistencia, 30
Criterio de Información
    de Akaike, 29
    de Schwarz, 29

Datos de Series de Tiempo, 201

Ecuación
    Yule-Walker, 132
Ecuación Yule–Walker, 132
Eficiencia Asintótica, 30
El Modelo de Espacio de Estados, 69
Error cuadrático medio, 24
Estabilidad Estructural, 71
Estacionario en Covarianza, 97
Estadístico
    Ljung-Box, 106
Estadístico
    Durbin-Watson, 109
Estimación
    de la Autocorrelación, 100
    de la Autocovarianza, 100
Estimador de la Varianza del Error, 24
Estimadores Recursivos, 73
EWMA, 84

fac, 99
Filtro
    causal, 84
Función
    de Autocovarianza, 97
Función
    de Autocorrelación, 99
    de Autocorrelación Parcial, 122
función de R
    dwtext, 110

Hyndman, R., 148

librería
    dynlm, 114
    lmtest, 110

Método de Descomposición STL, 65

Método de Suavizamiento
Exponencial Simple, 84
Marcha Aleatoria, 157
Matriz Toeplitz, 144
Media Móvil, 83
Media Móvil
Bilateral, 84
Unilateral, 84
Modelo
aditivo de componentes, 35
log-Lineal, 40
Modelo de Suavizamiento
de Holt-Winters, 67
Modelo tendencia
cúbico, 37
cuadrático, 37
exponencial, 37
lineal, 37
logístico, 37

Operador de Rezago, 120

Proceso
$ARIMA(p, d, q)$, 156
SARIMA, 173
Proceso
$AR(1)$ , 108
AR(p), 129
ARMA(p,q), 140
integrado, 159
MA(q) , 120
Proceso MA Invertible, 123
Pronóstico
de densidad, 41
por intervalo, 41
Pronósticos
con base en la tendencia, 41
definición, 41
Propiedades
de procesos MA$(q)$, 121
Prueba
Dickey-Fuller, 166
Aumentada, 169
Prueba
Diebold-Mariano, 164
Durbin-Watson, 108
Ljung-Box, 105
Prueba MOSUM, 76
Pruebas
de Raíz Unitaria Estacional, 179
Pruebas de Incorrelación, 104

R cuadrado, 28
R cuadrado ajustado, 28
Raíz Unitaria, 156
Autoregresiva, 156
Recursivos
Estimadores, 72
Regresión Loess, 61
Representación causal
ARMA, 142
Residuales
Recursivos , 74
Residuales Recursivos Cuadrados CUSUM-SQ, 75
Residuos
estructurales, 55
Ruido Blanco, 95
Gaussiano, 95

Supercosistencia, 165

Tendencia, 36
Tendencia Estocástica, 159
Teorema
de Bartlett, 101

Validación cruzada, 41

validación cruzada, 42
Variable
    Indicadora Estacional, 53