

---

Notas de Clase

Series de Tiempo con R



# NOTAS DE CLASE

## Series de Tiempo con R

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

**Universidad Nacional de Colombia**  
Medellín

Copyright ©2006 Universidad Nacional de Colombia.  
Profesor Norman Diego Giraldo Gómez.

Publicado por Editorial  
ISBN 000-000-000-0

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 de la Editorial.

**Centro de Documentación Rafael Botero, UN-Medellín:**

Series de Tiempo 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

---

<b>1. Introducción al Lenguaje R</b>	<b>1</b>
1.1. Introducción al Curso . . . . .	1
1.1.1. Notas sobre Pronósticos . . . . .	1
1.2. Introducción al R . . . . .	2
1.3. Uso de R de Manera Interactiva versus Ejecución desde un Programa . . . .	4
1.3.1. Interactiva . . . . .	4
1.3.2. Ejecución desde un Programa . . . . .	5
1.3.3. Lectura de Datos . . . . .	8
1.4. Repaso del Modelo de Regresión Lineal . . . . .	9
1.4.1. Estimación de los Parámetros . . . . .	10
1.4.2. Pruebas de Ajuste del Modelo . . . . .	12
1.4.3. Estadísticos de Ajuste y de Selección de Modelos . . . . .	13
1.4.4. Mínimos Cuadrados Nolineales . . . . .	15
1.5. Ejemplo de Regresión con R . . . . .	16
1.6. Manejo de Series de Tiempo con R . . . . .	18

<b>2. Modelado y Pronóstico de la Tendencia</b>	<b>19</b>
2.1. El Modelo Aditivo de Componentes de Series de Tiempo . . . . .	19
2.2. Estimación de la Tendencia . . . . .	23
2.3. Pronósticos con base en la Tendencia . . . . .	24
2.4. Caso de Estudio: Pronóstico de Ventas al Menudeo . . . . .	24
2.4.1. Descripción de los Datos . . . . .	25
2.4.2. Programa en R . . . . .	25
2.5. Problemas . . . . .	30
<b>3. Recursos en R Para Análisis de Tendencia</b>	<b>33</b>
3.1. Introducción . . . . .	33
3.2. Metodologías de Descomposición con Base en Suavizadores . . . . .	34
3.2.1. Regresión Local Loess . . . . .	34
3.2.2. STL: Método de Descomposición Basada en Regresión Loess . . .	36
3.3. Filtros Lineales, Medias Móviles y Suavizadores . . . . .	36
3.4. La Función Filter . . . . .	39
3.5. Ejemplo de Análisis de Tendencia con Suavizadores . . . . .	42
3.6. Notas . . . . .	46
<b>4. Modelado y Pronóstico incluyendo la Componente Estacional</b>	<b>49</b>
4.1. Definiciones . . . . .	49
4.2. Estimación de la Componente Estacional . . . . .	51
4.2.1. Modelos con Variables Indicadoras . . . . .	51
4.2.2. Modelos con Funciones Trigonómicas . . . . .	54
4.3. Procedimientos Recursivos de Estimación para Diagnosticar Estabilidad Es- tructural en Modelos de Pronósticos . . . . .	57
4.4. Métodos de Filtrado para la Componente Estacional . . . . .	66

4.4.1. Algoritmo de Brockwell y Davis . . . . .	66
4.4.2. Método de Holt-Winters . . . . .	69
4.5. Notas . . . . .	70
4.6. Problemas . . . . .	71
<b>5. Validación de los Supuestos sobre los Errores</b>	<b>73</b>
5.1. Series Estacionarias en Covarianza . . . . .	75
5.1.1. Estimación de las funciones de Autocovarianza y Autocorrelación. .	78
5.2. Pruebas de Incorrelación . . . . .	81
5.2.1. Prueba Ljung-Box (LB) . . . . .	81
5.2.2. Prueba Durbin-Watson (DW) . . . . .	83
5.3. Alternativas cuando los Residuos Estructurales muestran Autocorrelación .	88
<b>6. Modelos ARMA para la Componente Aleatoria</b>	<b>93</b>
6.1. Introducción . . . . .	93
6.2. Procesos de Medias Móviles de orden $q$ , $MA(q)$ . . . . .	94
6.2.1. Propiedades . . . . .	94
6.2.2. Condición de Invertibilidad del Proceso $MA(q)$ . . . . .	97
6.2.3. Función fac parcial de un Proceso $MA(q)$ invertible . . . . .	98
6.2.4. Implementación en R . . . . .	99
6.3. Procesos Autorregresivos de Orden $p$ , $AR(p)$ . . . . .	100
6.3.1. Condición Suficiente para que un $AR(p)$ sea Estacionario en Covarianza	100
6.3.2. Algunas Propiedades de los Procesos $AR(p)$ Estacionarios . . . . .	101
6.3.3. Proceso $AR(1)$ . . . . .	105
6.4. Procesos Autoregresivos y de Medias Móviles $ARMA(p,q)$ . . . . .	107
6.4.1. Propiedades de los Modelos ARMA . . . . .	109

6.4.2. Librerías para identificación, estimación y pronósticos de modelos ARMA . . . . .	111
6.4.3. Identificación de modelos ARMA . . . . .	111
6.4.4. Estimación de modelos ARMA . . . . .	112
6.5. Modelos ARMA Estacionales, SARMA . . . . .	114
<b>7. Raíces Unitarias y Tendencias Estocásticas (ARIMA)</b>	<b>119</b>
7.1. Introducción . . . . .	119
7.2. Modelos ARIMA . . . . .	119
7.3. Raíces Unitarias, Estimación y Prueba de Hipótesis . . . . .	127
7.3.1. Prueba Dickey-Fuller (DF) . . . . .	129
7.3.2. Prueba Dickey-Fuller Aumentada . . . . .	130
<b>8. Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)</b>	<b>137</b>
8.1. Modelos SARIMA . . . . .	137
8.1.1. Modelo SARIMA . . . . .	137
8.2. Pruebas de Raíz Unitaria Estacional . . . . .	142
8.2.1. La prueba HEGY . . . . .	143
8.2.2. La prueba Canova-Hansen . . . . .	150
<b>A. Datos de Series de Tiempo</b>	<b>155</b>
A.1. Series con Tendencia y Estacionalidad . . . . .	155
A.1.1. Ejemplo 1. Cinco Series . . . . .	155
A.1.2. Ejemplo 2. Consumo domiciliario agua potable en Medellín . . . . .	157



---

## Índice de figuras

---

2.1. Descomposición de la Serie $Y_t$ . . . . .	20
2.2. Ajuste Logístico a Población de Medellín . . . . .	22
2.3. Pronósticos (—) versus Observados (-o-o-) . . . . .	30
3.1. Suavizamiento Loess de Número de Licencias de Construcción en Medellín, 1986-2003 . . . . .	35
3.2. Suavizamiento STL de Número de Licencias de Construcción en Medellín, 1986-2003 . . . . .	36
3.3. Precios de la Acción de Microsoft Suavizados con SES . . . . .	38
3.4. Suavizamientos SES, Holt, Media Movil unilateral 30 del IGBC, 2009-2010	43
4.1. Average Monthly Temperatures at Nottingham, 1920-1939 . . . . .	50
4.2. Ajuste con Variables Indicadoras . . . . .	54
4.3. Estimadores Recursivos para el Modelo de Producción de Cemento . . . . .	60
4.4. Región de Rechazo de la Prueba <i>CUSUM</i> . . . . .	61
4.5. Resultados prueba CUSUM gráfica . . . . .	63
4.6. Punto de Quiebre 1974-Q4 en la Serie y en los Residuos . . . . .	64

4.7. Ajuste con el Método de Holt-Winters . . . . .	70
4.8. Pronósticos Varios Métodos . . . . .	71
5.1. Ejemplo Función de Autocovarianza . . . . .	76
5.2. Ejemplo Función de Autocorrelación. . . . .	76
5.4. Fac Teórica de un Ruido Blanco. . . . .	77
5.3. Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario . . .	78
5.5. Ejemplos de Fac y Bandas de Bartlett . . . . .	80
5.6. Fac de los Residuos Estructurales Serie Cementos . . . . .	80
5.7. Densidad $\chi_m^2$ . . . . .	82
5.8. Fac Muestral de un $AR(1)$ . . . . .	84
5.9. Residuales estructurales versus Y estimada . . . . .	87
6.1. Función de Autocorrelación. . . . .	96
6.2. FAC muestral de un $MA(3)$ . . . . .	96
6.3. $FAC$ y $FACP$ de un $MA(3)$ . . . . .	99
6.4. $FAC$ y $FACP$ del Ejemplo. . . . .	100
6.5. Círculo Unitario . . . . .	101
6.6. FAC de $Y_t = \varphi Y_{t-1} + \varepsilon_t$ . . . . .	104
6.7. FACP Muestral de $AR(p)$ . . . . .	105
6.8. Fac Muestral de $ARMA(p, q)$ . . . . .	112
6.9. Posibles submodelos (1,1)=AR, (1,2),(2,1)=ARMA, (2,2)=SARMA . . . .	116
7.1. Trayectorias de Marchas Aleatorias . . . . .	121
7.2. Pronósticos de la Serie PNB-USA. . . . .	124
7.3. (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada . . . . .	124

7.4. Pronósticos usd/yen con ARIMA(3,1,2)(continua) y Tendencia Lineal+AR(2) (punteada) . . . . .	125
7.5. Serie USD por Libra . . . . .	126
7.6. FAC y FACP Muestral de un ARIMA(p,1,q). . . . .	128
8.1. Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12] . . . . .	139
8.2. Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12] . . . . .	141
8.3. Serie de Producción Trimestral de Cemento, Australia . . . . .	146
8.4. Serie Diferenciada de Producción Trimestral de Cemento . . . . .	148
8.5. Pronósticos a 8 trimestres de la Producción Trimestral de Cemento . . . . .	149
8.6. Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descom- posicion + AR, Panel inferior: SARIMA . . . . .	150



# CAPÍTULO 1

---

## Introducción al Lenguaje R

---

### 1.1. 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].

#### 1.1.1. Notas sobre Pronósticos

Los pronósticos se utilizan en forma constante en diversos campos: economía, finanzas, mercadeo, medio ambiente, ingeniería. Una finalidad es proveer una guía para las decisiones que deben tomarse. Algunos ejemplos en estos campos pueden ser los siguientes.

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. 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.
4. En Economía los pronósticos sobre ciclos económicos de recesión y expansión para guía de aumento ó disminución de las tasas interbancarias.
5. En Planeación. Los pronósticos de demanda de viviendas, en los diferentes rangos, para orientar políticas de uso del suelo.
6. En Planeación. Los pronósticos de gastos sirven para definir los presupuestos futuros.
7. 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.
8. En Planeación. El pronóstico de consumo de energía eléctrica domiciliaria es fundamental para las decisiones de generación.
9. En Turismo. El pronósticos de número de turistas mensuales para determinar la demanda hotelera.
10. 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.

## **1.2. Introducción al R**

1. Qué es R? <http://www.r-project.org/>. Es un software libre, de código abierto, para programación orientada a objetos, dedicado a cómputos estadísticos y financieros.
2. Qué es R-metrics? <https://www.rmetrics.org/>. Es un sistema de librerías (toolboxes) de R para realizar finanzas computacionales e ingeniería financiera. R posee aproximadamente 1600 librerías sobre diversas aplicaciones. R-metrics posee 10 librerías especializadas 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 2.13.1 for Windows).
- d) Ejecutar el archivo .exe desde el escritorio.

4. Detalles de instalación: R instala unas librerías por defecto. El resto de las 1600 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, Colombia 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`: estimación por 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).

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

### 1.3.1. Interactiva

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

- `> a = 7.5` #equivale a los dos siguientes

- `a <- 7.5`

- `7.5 -> a` Crea un objeto numérico “a” con el valor 7.5.

- `> 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
```

### 1.3.2. Ejecución desde un Programa

En lugar de ó al mismo tiempo del manejo interactivo se pueden usar programas. Los programas en R tienen extensión *.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 dividir 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(y2)
  s = ((n1-1)*s1 + (n2-1)*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 1.3.1.** # 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 1.3.2.** Si *D* se ha definido como data.frame, y contiene las columnas edad, genero, salario, entonces 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 matriz, o un data.frame, o un list.

### Ejemplo 1.3.3.

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

### 1.3.3. 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 1.3.4.** *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/~ndgiraldd/
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)
```

### Otras funciones de lectura y escritura

Hay otras funciones de lectura, como `G = read.csv("nombre", header = TRUE)`, que lee archivos excel “delimitados por comas”, de extensión .csv. Y funciones para escribir datos como la función `G = write.matrix(data.frame, archivo)`.

```
#-----
D = data.frame(fechas=fecha,x=x)
require(MASS)
write.matrix(D,"h.dat")
```

## 1.4. Repaso del Modelo de Regresión Lineal

El modelo de regresión lineal múltiple es básico para algunos de los modelos de series de tiempo. Suponga que se tienen unas muestras de las variables aleatorias  $Y, X_1, X_2$ ,

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

Además suponga que  $\varepsilon_t \sim iid N(0, \sigma^2)$  es una muestra de una Normal. Si se cumple que

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \quad (1.1)$$

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. Se asume que las variables  $X_{1,t}$ ,  $X_{2,t}$  están dadas. Los supuestos del modelo lineal son

1. Las  $\varepsilon_t$  no están correlacionadas.
2. Las  $\varepsilon_t$  tienen varianzas constante.
3.  $\varepsilon_t$  es independiente de las  $X_{i,t}$ . Lo cual implica

- a)  $\mathbb{E}(\varepsilon_t|X_{1,t}, X_{2,t}) = \mathbb{E}(\varepsilon_t) = 0$ , luego  $E(Y_t|X_{1,t}, X_{2,t}) = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}$ .
- b)  $Var(Y_t|X_{1,t}, X_{2,t}) = Var(\varepsilon_t|X_{1,t}, X_{2,t}) = Var(\varepsilon_t) = \sigma^2$ .

donde  $\sigma^2$  se denomina la varianza del error, y es constante.

4. El rango de la matriz  $X$  es 3, es decir, no existen constantes  $a, b$  tal que  $aX_{1,t} + bX_{2,t} = 1$ , es decir, 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, por ejemplo

$$Y_t = \frac{1}{1 + e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}}} + \varepsilon_t, \quad (1.2)$$

es no lineal.

#### 1.4.1. Estimación de los Parámetros

Los parámetros de los modelos (1.1) y (1.2), son:  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$  y se busca estimarlos a partir de la muestra de  $Y_t$ ,  $X_{1,t}$ ,  $X_{2,t}$ . 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

$$G(\underline{\beta}) = \sum_{t=1}^T (Y_t - (\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}))^2, \quad (1.3)$$

es decir,  $\underline{\hat{\beta}}$  cumple que  $\forall \underline{\beta}, G(\underline{\hat{\beta}}) \leq G(\underline{\beta})$ . Y se puede escribir

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

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. Si se define la matriz de diseño  $X_{T \times 3}$  como

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

entonces  $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$  es el modelo en forma matricial y se puede comprobar que  $\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y}$  es el estimador MCO.

Este estimador  $\underline{\hat{\beta}}$  es un estimador lineal insesgado, es decir,  $E(\underline{\hat{\beta}}) = \underline{\beta}$ , de mínima varianza, osea,  $Var(\hat{\beta}_i) \leq Var(\hat{\beta}_i^*)$  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  $\hat{\beta}_i \sim N(\cdot, \cdot)$  y se pueden definir los estadísticos de ajuste como la t y la F.

Note que una vez definidos los  $\hat{\beta}_i$ , se define los residuos estimados como

$$\hat{\varepsilon}_t = Y_t - (\hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \hat{\beta}_2 X_{2,t}), \quad t = 1, \dots, T. \quad (1.4)$$

Y se definen los pronósticos dentro de la muestra (interpolar) como

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \hat{\beta}_2 X_{2,t}, \quad t = 1, \dots, T, \quad (1.5)$$

y fuera de la muestra (extrapolar) como

$$\hat{Y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,T+h} + \hat{\beta}_2 X_{2,T+h}, \quad h = 1, \dots, m, \quad (1.6)$$

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

### 1.4.2. Pruebas de Ajuste del Modelo

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

Si  $\varepsilon_t \sim iidN(0, \sigma^2)$  se cumple que las  $\hat{\beta}_i \sim N(\beta_i, \sigma_{\hat{\beta}_i}^2)$ , donde  $\hat{\sigma}_{\hat{\beta}_i} = S_{\hat{\beta}_i}$  = error estándar. Además se cumple que  $t = \hat{\beta}_i / S_{\hat{\beta}_i}$  se distribuye  $t_{T-k}$ , donde  $k$  es el número de parámetros  $\beta_i$ , y es el estadístico de la prueba de significancia de los parámetros  $\beta_i$ . Esta prueba se describe así. Para  $i = 1, \dots, k$ ,

1.  $H_0 : \beta_i = 0$  versus  $H_1 : \beta_i \neq 0$ .
2. Estadístico de prueba  $t_i = \hat{\beta}_i / S_{\hat{\beta}_i}$ .
3. Si  $H_0$  es cierta entonces  $t_i \sim t_{T-k}$ .
4. Decisión. Se plantean 3 formas equivalentes.
  - 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 Si  $0 \notin [\hat{\beta}_i - 1.96S_{\hat{\beta}_i}, \hat{\beta}_i + 1.96S_{\hat{\beta}_i}]$  se rechaza a un nivel de significancia de 5 %.
  - 3 Defina  $Valorp = P(|t_{T-k}| > |t_{obs}|) = P(t_{T-k} > |t_{obs}|) + P(t_{T-k} < -|t_{obs}|)$ . Si  $Valorp < 0.05$  se rechaza  $H_0$  a un nivel de significancia del 5 %, de lo contrario no se rechaza  $H_0$ .

#### Prueba F

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

$$SST = \sum_{t=1}^T (Y_t - \bar{Y})^2 \quad (1.7)$$

$$= \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 \quad (1.8)$$

$$= SSR + SSE. \quad (1.9)$$

Se define el estadístico  $F$  para probar la hipótesis

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{k-1} = 0 \text{ versus } H_1 : \text{no}(H_0)$$



donde el estadístico está dado por

$$F = \frac{\frac{SSR}{k-1}}{\frac{SSE}{T-k}} \stackrel{H_0}{\sim} F_{k-1, T-k}. \quad (1.10)$$

Si se rechaza  $H_0$ , las variables  $X_{1,t}, \dots, X_{k-1,t}$  tienen capacidad para predecir los valores de  $Y_t$ . Si  $F > F_{k-1, T-k, \alpha}$  se rechaza  $H_0$ .

**Ejemplo 1.4.1.** Con  $k = 3$ ,  $T = 48$ ,  $\alpha = 0,05$ , se tiene que  $F_{k-1, T-k, \alpha} = F_{2, 45, 0.05} = 3.2$ , luego si  $F = 30.89$ . Como  $F > 3.2$  se rechaza  $H_0$ , es decir,  $X_{1,t}$  y  $X_{2,t}$  tienen capacidad predictiva sobre  $Y_t$ .

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

Se define el estimador de la varianza del error como

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 = \frac{1}{T-k} \sum_{t=1}^T \hat{\varepsilon}_t^2, \quad (1.11)$$

donde  $k$  es el número de  $\beta$ 's en el modelo. Se denomina “el error cuadrático medio estimado” ó MSE estimado.

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

$$R^2 = 1 - \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (1.12)$$

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

$$R^2 \geq 0.7 \text{ buena}$$

$$R^2 = 0.6 \text{ buena pero no mucho}$$

$$R^2 = 0.4 \text{ regular-mala}$$

$$R^2 = 0.2 \text{ desechar, seguramente la prueba } F \text{ y las pruebas } t \text{ no rechazan}$$

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

$$\bar{R}^2 = 1 - \frac{(T-1)SSE}{(T-k)SST}. \quad (1.13)$$

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

$$\bar{R}^2 = 1 - \frac{T-1}{T-k}(1 - R^2). \quad (1.14)$$

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

$$AIC = e^{2k/T} \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T} = \frac{(T-k)\hat{\sigma}^2}{T} e^{2k/T}. \quad (1.15)$$

donde  $\sum_{t=1}^T \hat{\varepsilon}_t^2 = (T-k)\hat{\sigma}^2$ . 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 (1.15).

De la definición del criterio AIC se observa:

- Si  $k \nearrow$ , entonces  $e^{\frac{2k}{T}} \nearrow$  luego  $AIC \nearrow$ .
- Si  $MSE \searrow$ , entonces  $AIC \searrow$ .

*«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» Diebold [1999, pág. 73]*

Se define el criterio de información de Schwarz BIC

$$BIC = T^{k/T} \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T} \quad (1.16)$$

El BIC también se define como el logaritmo de (1.16).

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

**Resultado:**  $S^2$ ,  $AIC$  no son consistentes pero  $BIC$  sí es.

**Definición 1.4.2 (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  $Y_{T+1} - \hat{Y}_{T+1}$

**Resultado:**  $AIC$  es eficiente asintóticamente pero  $BIC$  no.

«Muchos autores recomiendan usar el modelo más parsimonioso que selecciona el  $BIC$  en igualdad de circunstancias» Diebold [1999, pág. 75].

#### 1.4.4. 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

$$Y = g(X; \beta) + \epsilon \quad (1.17)$$

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 (1.2)

$$Y_t = \frac{1}{1 + e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}}} + \varepsilon_t, \quad (1.18)$$

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

```
m = nls(Y ~ 1/(1+exp(w+a*X1+b*X2)),
start=list(w=0.1, a=1, b=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.

## 1.5. Ejemplo de Regresión con R

En este ejemplo se muestran algunas de las capacidades de R. En particular: 1) La lectura de datos en un sitio de la web a partir de la dirección url. 2) Los estadísticos para escogencia del mejor modelo. 3) Gráficas de los modelos ajustados.

Los datos son de tres variables  $Y, X_1, X_2$ , para un modelo de regresión (1.1). Utilizamos datos de un ejemplo de la página web

<http://www.stat.umn.edu/geyer/5102/examp/reg.html>

Se examinan tres modelos lineales:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \varepsilon_t, \quad (1.19)$$

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \quad (1.20)$$

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \quad (1.21)$$

Los resultados aparecen en el Cuadro 1.1

```
#-----
# programa en R
library(xtable) # para reportar tablas en LaTeX para informes
library(fRegression) # para estadísticos especiales de regresion

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

m1 = lm(y ~ x1)
summary(m1)
print(xtable(m1),digits=4)

m2 = lm(y ~ x1 + x2)
summary(m2)
print(xtable(m2),digits=4)

m3 = lm(y ~ -1 + x1 + x2)
summary(m3)
print(xtable(m3),digits=4)
```

Tabla 1.1: 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

Las Tablas Anova para los modelos m2, m3 son las siguientes.

```
#-----
anova(m2)

Residual standard error: 0.9722 on 47 degrees of freedom
Multiple R-squared: 0.7388,
Adjusted R-squared: 0.7277
F-statistic: 66.48 on 2 and 47 DF,
p-value: 1.987e-14
#-----
anova(m3)

Residual standard error: 0.9644 on 48 degrees of freedom
Multiple R-squared: 0.9922,
Adjusted R-squared: 0.9919
F-statistic: 3061 on 2 and 48 DF,
p-value: < 2.2e-16
```

Y los valores de AIC aparecen a continuación.

```
#-----
(c(AIC(m1), AIC(m2), AIC(m3)))
156.5899 143.9822 142.2242
```

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.

## 1.6. Manejo de Series de Tiempo con R

En esta sección consideramos una serie de tiempo  $T_t$ ,  $t = 1, 2, \dots, T$ , 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 `ts`, mediante la función `ts()`.

---

### Modelado y Pronóstico de la Tendencia

---

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

Pueden mencionarse varios motivos por los cuales es útil analizar las series de tiempo. Por ejemplo, para obtener pronósticos de valores futuros de la serie, con el fin de ayudar a tomar decisiones que tienen consecuencias importantes. O para entender mejor el mecanismo de generación de los datos, que puede no ser claro inicialmente en una investigación.

#### 2.1. El Modelo Aditivo de Componentes de Series de Tiempo

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

$$Y_t = T_t + S_t + \varepsilon_t, \quad (2.1)$$

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 de  $t$  determinísticas. Su evolución es perfectamente predecible.

La componente  $T_t$  en algunos casos también puede ser una componente estacional, pero de baja frecuencia, o, equivalentemente, una componente con período muy 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:

$$Y_t = T_t S_t \exp \varepsilon_t. \quad (2.2)$$

En la Figura (2.1) 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).

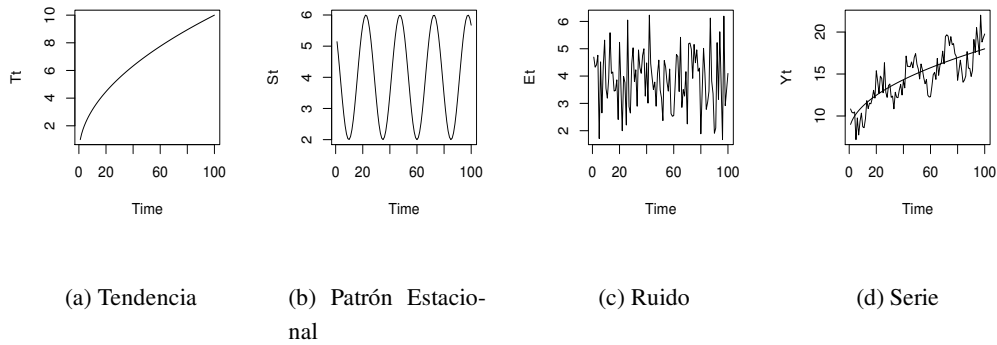


Figura 2.1: Descomposición de la Serie  $Y_t$

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 reconstruir  $Y_t$ ,  $\hat{Y}_t = \hat{T}_t + \hat{S}_t + \hat{\varepsilon}_t$ , 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  $Y_1, \dots, Y_T$  con  $h = 1, 2, \dots, m$ . Sin embargo, puede suceder que la serie  $\hat{\varepsilon}_t$  sea incorrelacionada, es decir,  $\text{Corr}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+s}) = 0$ , para  $s \neq 0$ . En este caso  $\hat{\varepsilon}_{T+h} = 0, \forall h > 0$ .

**Definición 2.1.1 (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  $T_t$  depende de parámetros, que deben estimarse.

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

Lineal

$$T_t = \beta_0 + \beta_1 t, \quad (2.3)$$



Cuadrático

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad (2.4)$$

Cúbico

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3, \quad (2.5)$$

Exponencial

$$T_t = \exp(\beta_0 + \beta_1 t). \quad (2.6)$$

Logístico

$$T_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)}, \quad (2.7)$$

En la tendencia cuadrática se puede observar:

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).

**Definición 2.1.2.** *El modelo Logarítmico Lineal ó Log-Lineal se define como*

$$\ln Y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (2.8)$$

Corresponde a un modelo con tendencia lineal para el logaritmo de  $Y_t$ . En (2.8) al tomar exponencial se tiene  $Y_t = \exp(\beta_0 + \beta_1 t + \varepsilon_t)$ , que es similar al modelo con tendencia exponencial (2.6),  $Y_t = \exp(\beta_0 + \beta_1 t) + \varepsilon_t$ . Sin embargo, son modelos diferentes y se estiman por métodos diferentes.

**Ejemplo 2.1.1.** *En el Cuadro 2.1 se muestran los datos del total de habitantes en Medellín, según los Censos desde 1905 hasta 2005, según el DANE (<sup>1</sup>). En Poveda [1982] se plan-*

---

<sup>1</sup>[http://es.wikipedia.org/wiki/Demografía\\_de\\_Medellín](http://es.wikipedia.org/wiki/Demografía_de_Medellín)

Tabla 2.1: 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

teó ajustar los datos de desde 1938 hasta 1973 mediante un modelo de componentes con tendencia logística, (2.7).

$$Y_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)} + \varepsilon_t. \quad (2.9)$$

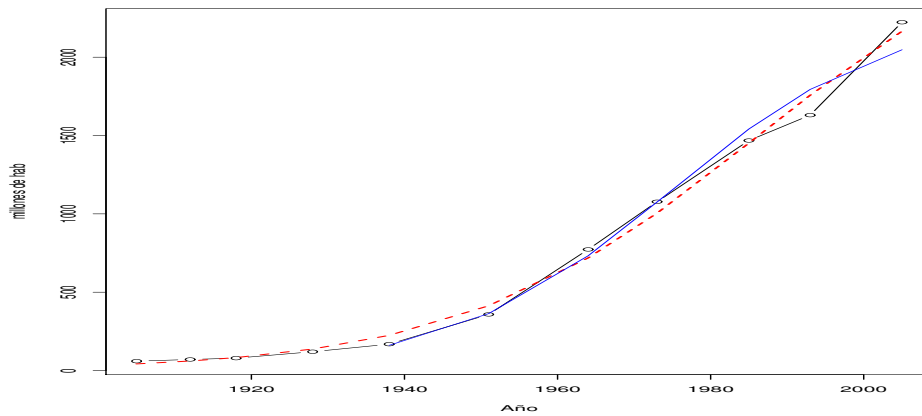


Figura 2.2: Ajuste Logístico a Población de Medellín

## 2.2. Estimación de la Tendencia

En este capítulo se introduce la estimación de la tendencia mediante modelos de regresión lineal y no lineal. Son modelos paramétricos. En el capítulo siguiente se introducen modelos no paramétricos para estimar la tendencia, como los suavizadores, los filtros lineales y no lineales y las 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 (2.3), (2.4), (2.5) y (2.8) 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  $G(\underline{\beta}) = \sum_{t=1}^T (Y_t - T_t(\underline{\beta}))^2$  toma el valor mínimo.

$$\hat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}} G(\underline{\beta}). \quad (2.10)$$

Para los modelos (2.6) y (2.7) se usa el método de mínimos cuadrados no lineales, que también minimiza la suma de errores cuadrados  $G(\underline{\beta}) = \sum_{t=1}^T (Y_t - T_t(\underline{\beta}))^2$ , pero  $T_t(\underline{\beta})$  es una función no lineal de  $\underline{\beta}$ .

El modelo LogLineal (2.8) es equivalente, algebraicamente, a

$$Y_t = \exp(\beta_0 + \beta_1 t + \varepsilon_t). \quad (2.11)$$

Sin embargo, este último modelo es no lineal y no coincide con el modelo exponencial, (2.6),  $Y_t = \exp(\beta_0 + \beta_1 t) + \varepsilon_t$ . Es posible estimar por mínimos cuadrados ordinarios el modelo LogLineal y utilizar los parámetros estimados  $\hat{\beta}_0, \hat{\beta}_1$  como valores iniciales en la estimación del modelo exponencial por mínimos cuadrados no lineales. Pero los parámetros estimados en ambos modelos no necesariamente coinciden.

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  $t^2$ .

### 2.3. Pronósticos con base en la Tendencia

Suponga la serie con tendencia  $Y_t = T_t + \eta_t$ ,  $t = 1, \dots, T$ , con  $(\eta_t)$  una sucesión  $iid(0, \sigma^2)$ . Los pronósticos de  $Y_t$  en los tiempos  $T + 1, T + 2, \dots, T + h$ ,  $h \geq 1$  se definen como

**Definición 2.3.1.**

$$\hat{Y}_{T+j} = \hat{T}_{T+j}, j = 1, \dots, h. \quad (2.12)$$

donde  $\hat{T}_t$  es la función estimada de la tendencia. Por ejemplo, en el modelo lineal  $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$ , al reemplazar  $t$  por  $T + h$  se obtiene  $Y_{T+h} = \hat{\beta}_0 + \hat{\beta}_1(T + h) + \hat{\varepsilon}_{T+h}$ . Pero el pronóstico  $\hat{Y}_{T+h}$ , puede o no ser cero. Es cero si los residuos  $\hat{\varepsilon}_t$ ,  $t = 1, \dots, T$  son independientes. Para decidir esto se realiza una prueba de incorrelación y una prueba de normalidad. Puede suceder que los residuos estimados  $\hat{\varepsilon}_t$  sean autocorrelacionados. En este caso el pronóstico es  $\hat{Y}_{T+j} = \hat{T}_{T+j} + \hat{\varepsilon}_{T+j}$ .

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

**Definición 2.3.2.**

$$\hat{Y}_{T+j} = \mathbb{E}(Y_{T+j} | Y_1, \dots, Y_T), j = 1, \dots, h. \quad (2.13)$$

Otros tipos de pronósticos son

- El pronóstico por intervalo se obtiene si se asume que  $\varepsilon_t \sim iid N(0, \sigma^2)$ . 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  $Y_{T+h} \sim N(\hat{Y}_{T+h}, \hat{\sigma}^2)$ .

### 2.4. 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.

### 2.4.1. 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. Esta técnica para verificar la eficiencia del modelo estimado se llama “validación cruzada”.

### 2.4.2. 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
# RTRR 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(l yi ~ 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**

Nótese de los resultados del cuadro 2.2 que el modelo cúbico no ajusta ya que el coeficiente de  $t^3$  no da 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:  $R^2$  ajustado, MSE,  $\log(\text{AIC})$ ,  $\log(\text{BIC})$  se usará la función `medidas()`, que los calcula para cada modelo.

```
medidas = function(m,y,k){
# y = serie, m = modelo, k = numero parametros
```

Tabla 2.2: 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

```

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, mse, aic, bic)
names(M) = c("R2-ajus", "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 2.3: Estadísticos de Selección de Modelos

Del cuadro 2.3 se puede concluir que el modelo que mejor ajusta los datos es el modelo cuadrático (2.4) 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 (2.4). 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
```

Year	1950	1960	1970	1975	1980	1985
Unemployed	1869	271	149	1074	889	2304
Year	1988	1989	1990	1991	1992	1993
Unemployed	2242	1869	1883	1689	1808	2270

Tabla 2.4: Unemployed2 Data.

```
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 (2.3) se puede concluir que el modelo cuadrático genera pronósticos confiables a corto plazo.

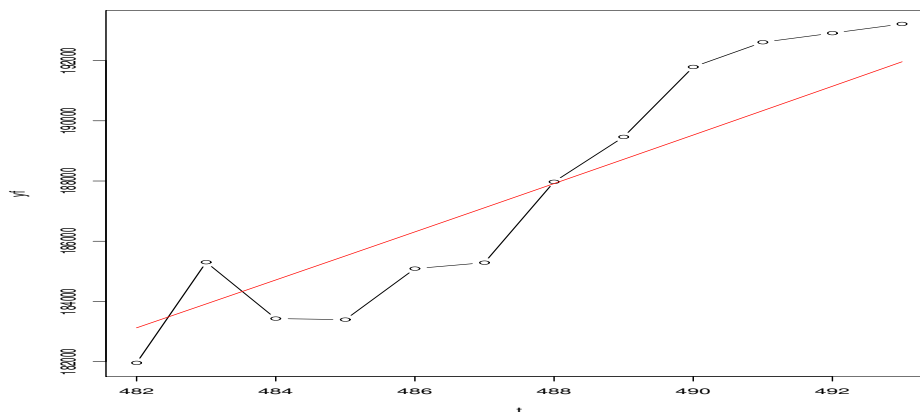


Figura 2.3: Pronósticos (—) versus Observados (-o-o-)

## 2.5. Problemas

1. (Unemployed2 Data) Table 2.4 lists total numbers of unemployed (in thousands) in West Germany between 1950 and 1993. Compare a logistic trend function with an allometric one. Which one gives the better fit?
2. 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 descríbalos en 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 definió 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  $\hat{\delta}_i$  versus  $i = 1, 2, \dots, s$ .
- 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?.



---

### Recursos en R Para Análisis de Tendencia

---

#### 3.1. Introducción

En este capítulo se introducen algunas de las herramientas de R para la implementación de la descomposición  $Y_t = T_t + S_t + \varepsilon_t$ , y para la estimación de la tendencia  $T_t$ . Hay que hacer énfasis en que estos métodos no asumen un modelo global para la tendencia  $T_t$  sino local, es decir, no son modelos con parámetros fijos, de la forma, por ejemplo,  $T_t = \beta_0 + \beta_1 t$ , sino que  $\beta_0$  y  $\beta_1$  cambian en el tiempo para permitir mayor flexibilidad.

1. La función **loess()** realiza un suavizamiento con base en una regresión local Loess.
2. Las funciones **decompose()** y **stl()** ambas realizan una descomposición de la serie  $Y_t$  en las tres componentes.
3. Hay librerías con varios tipos de filtros para suavizamiento y extracción de componentes de tendencia y estacionales como mFilter, que tiene implementado el filtro Hodrick-Prescott.
4. También la librería timsac tiene la función **decomp()** que realiza una descomposición incluyendo una componente autoregresiva y otra para fechas de intervenciones,  $Y_t = T_t + S_t + R_t + TA_t + \varepsilon_t$ .

5. Mediante la función **filter()** de la librería stats se pueden implementar varios tipos de medias móviles y filtros lineales, por ejemplo, filtros tipo Henderson y filtros recursivos, que son herramientas útiles para la estimación de la tendencia  $T_t$ .

En la sección 2.1 se introducen los métodos de descomposición con base en suavizadores. En la sección 2.2 se definen las medias móviles. EL uso de las funciones filter (stats) y filter (signal) se describe en la sección 2.3.

## 3.2. Metodologías de Descomposición con Base en Suavizadores

### 3.2.1. Regresión Local Loess

La Regresión Loess (inicialmente Lowess: Locally wighted 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 descripción del algoritmo Loess es la siguiente. Suponga  $(x_i, Y_i)$ ,  $i = 1, \dots, n$ . El objetivo de Loess es calcular una función de regresión local  $g(x)$  que es el análogo de, por ejemplo,  $a + bx$ , es decir,  $Y_i = g(x_i) + \varepsilon_i$ ,  $i = 1, \dots, n$ . Suponga  $x_{(1)} \leq x \leq x_{(n)}$ , entonces  $g(x)$  se calcula así:

1. Se escoge  $q \in \mathbb{Z}^+$  tal que  $q \leq n$ .
2. Se escogen los  $q$  valores  $x_i$  más cercanos a  $x$
3. Defina  $w(x) = \begin{cases} (1 - x^3)^3 & , 0 \leq x \leq 1 \\ 0 & , x \geq 1 \end{cases}$
4. Defina  $\lambda_q(x)$  la distancia de  $x$  al  $x_i$  más alejado entre los  $q$  escogidos.
5. Defina  $v_i(x) = w\left(\frac{|x_i - x|}{\lambda_q(x)}\right)$ , para cada  $x_i$  escogido.
6. Ajuste  $Y_i = a + bx_i$  ó  $Y_i = a + bx_i + cx_i^2$  con MCP ponderando cada  $x_i$  con  $v_i(x)$ .
7. Defina  $g(x)$  como el valor estimado de la regresión en  $x$ .

En series de tiempo se toma, por ejemplo,  $(Y_i, t_i)$ , es decir,  $x_i = t_i$ . En cada  $t$  se reemplaza  $Y_t$  por  $\hat{Y}_t = \hat{g}(t)$ .

**Ejemplo 3.2.1.** Si se toman los datos del número de licencias para vivienda nueva, mensuales, en Medellín entre 01/1986 y 06/2003, se observa un fuerte aumento durante los años 1986 - 1993. Los datos se tomaron de la página web de Camacol <sup>(1)</sup>. A partir de 1993 empieza un decrecimiento. Sin embargo, a partir de 1996 se observa un aparente cambio de concavidad y se sugiere una recuperación.

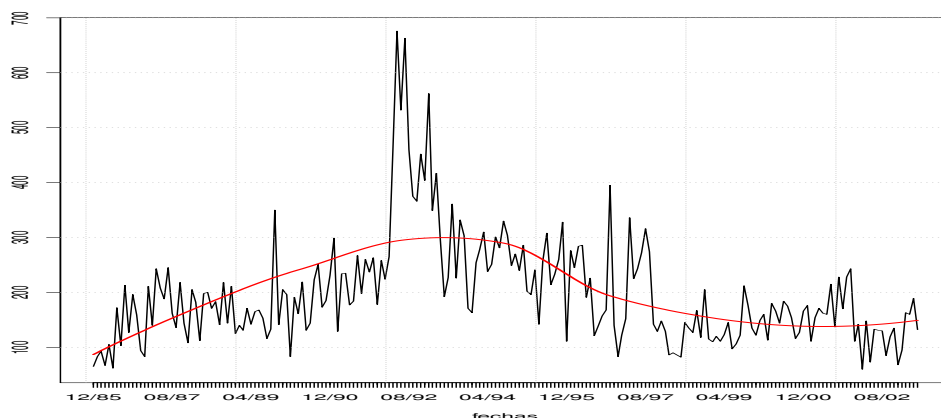


Figura 3.1: Suavizamiento Loess de Número de Licencias de Construcción en Medellín, 1986-2003

Este es el código en R para suavizamiento con Loess, ver Figura 3.1.

```
G = read.table("licenciasmedellin.dat",
header = TRUE, stringsAsFactors = FALSE)
attach(G)
yw=loess(y ~ time(y))
np = length(y)
fecha = seq(as.Date("1986/01/01"), as.Date("2003/06/03"),
by="months")
fechas = strptime(as.character(fecha), "%Y-%m-%d")

plot(fechas,y, 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)
```

<sup>1</sup>[www.camacol.com.co](http://www.camacol.com.co)

```
lines(fechas,yw$fitted, xaxt="n", panel.first = grid(),
type='l',col='red',lwd=2)
```

### 3.2.2. STL: Método de Descomposición Basada en Regresión Loess

STL es un método para estimar las componentes  $T_t$  y  $S_t$  con base en Regresión Loess, desarrollado por Cleveland et al. [1990]. STL consiste de una secuencia de dos aplicaciones iteradas de Regresión Loess. Para aplicar este método se debe especificar una frecuencia de muestreo relacionada con el período de la componente estacional. La forma de especificar esta frecuencia es declarando la variable en donde se encuentran los datos como un objeto `ts` con frecuencia (52, 12, 4, 1), es decir, semanal, mensual, trimestral o anual, respectivamente.

**Ejemplo 3.2.2.** Al aplicar el método *Stl* a los datos del Ejemplo (3.2.1) y graficar la serie suavizada se obtiene el resultado de la Figura (3.2).

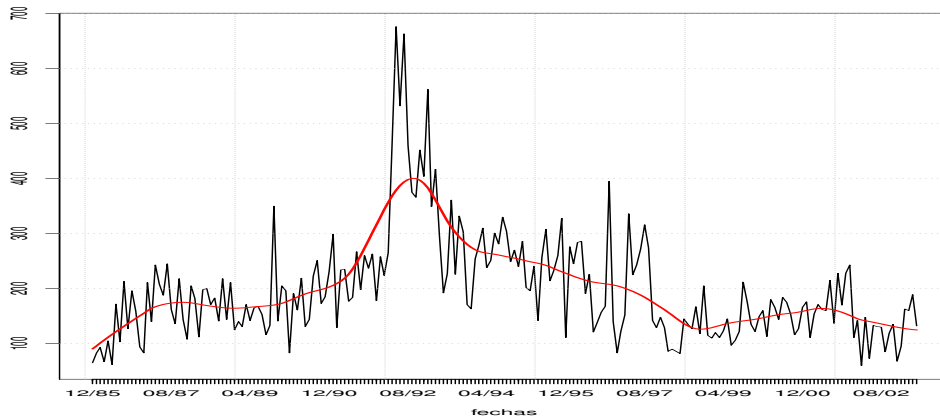


Figura 3.2: Suavizado STL de Número de Licencias de Construcción en Medellín, 1986-2003

### 3.3. Filtros Lineales, Medias Móviles y Suavizadores

**Definición 3.3.1 (Filtro Lineal).** Un filtro lineal es un operador que convierte una serie  $Y_t$  en otra  $\bar{Y}_t$  tal que

$$\bar{Y}_t = \sum_{j=-m}^m w_j Y_{t-j}, \quad t = -m, \dots, T-m \quad (3.1)$$



donde  $w_{-m}, w_{-m+1}, \dots, w_0, \dots, w_m$  son pesos predeterminados. En algunos casos para determinar la tendencia y reducir las fluctuaciones locales (suavizar) se restringen a  $\sum_{r=-m}^m w_r = 1$ . Este filtro lineal se denomina *Media Móvil*. Usualmente se escogen  $w_{-j} = w_j$ , pesos simétricos.

Los filtros lineales se utilizarán como estimadores de la tendencia, es decir,  $\hat{T}_t = \bar{Y}_t$ , cuando se observe que la tendencia no parece factible que sea descrita por una única función paramétrica, por ejemplo de la forma polinómica  $T_t = \beta_0 + \sum_{j=1}^k \beta_j t^j$ .

Una aplicación del filtro lineal es remover la tendencia, es decir, si  $\hat{T}_t = \bar{Y}_t$ , la serie  $X_t = Y_t - \hat{T}_t$  es la serie sin tendencia.

Si se aplica (3.1) a  $Y_t = T_t + S_t + \varepsilon_t$  el filtro lineal altera la amplitud de las oscilaciones de  $S_t$ . Ese efecto se llama ganancia del filtro.

**Ejemplo 3.3.1.** En Diebold [1999, pág. 84], se define:

- La Media Móvil General Bilateral con  $w_j = w_{-j}$ ,  $j = 0, 1, \dots, m$ .

$$\bar{Y}_t = \sum_{j=-m}^m w_j Y_{t-j}. \quad (3.2)$$

El término  $Y_{t-j}$  se llama “rezago de orden  $j$ ”. Si  $j < 0$  el rezago  $Y_{t-j}$  es un valor “futuro”.

- La Media Móvil Bilateral

$$w_j = \frac{1}{2m+1}, \quad j = -m, \dots, m. \quad (3.3)$$

- La Media Móvil Unilateral

$$w_j = \frac{1}{m+1}, \quad j = 0, 1, \dots, m. \quad (3.4)$$

El entero  $m \geq 1$  es el “ancho de ventana” de la media móvil.

La Media Móvil Unilateral (3.4) utiliza valores presentes y pasados de la serie. Es un ejemplo de filtro lineal “causal”. La Media Móvil Bilateral (3.3) no es un filtro lineal causal.

**Ejemplo 3.3.2.** Un ejemplo de filtro lineal es el Suavizamiento Exponencial Simple (SES), definido por

$$\bar{Y}_t = \sum_{j=0}^m \alpha(1-\alpha)^j Y_{t-j}, \quad \alpha \in (0, 1), m > 1. \quad (3.5)$$

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

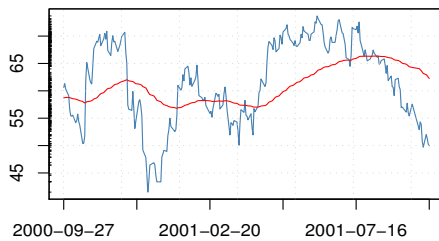
La ecuación (3.5) se puede obtener de la relación recursiva  $\bar{Y}_t = \alpha Y_t + (1 - \alpha)\bar{Y}_{t-1}$  para  $t = 2, 3, \dots, T$ . Se comprueba que  $\bar{Y}_t = (1 - \alpha)^{m+1}\bar{Y}_{t-m-1} + \alpha \sum_{j=0}^m (1 - \alpha)^j Y_{t-j}$ . Si  $m$  es grande se puede aproximar  $(1 - \alpha)^{m+1} \approx 0$  y por tanto  $\bar{Y}_t = \sum_{j=0}^m \alpha(1 - \alpha)^j Y_{t-j}$ .

- Si  $\alpha \approx 0$ , entonces el filtro reduce la serie casi a una constante. De la relación  $\bar{Y}_t = (1 - \alpha)^{m+1}\bar{Y}_{t-m-1} + \alpha \sum_{j=0}^m (1 - \alpha)^j Y_{t-j}$ , si  $\alpha = 0$ , entonces  $\bar{Y}_t = \bar{Y}_{t-m-1}$  para todo  $m > 1$ . Con lo cual  $\bar{Y}_t = \bar{Y}_0$ .
- Si  $\alpha \approx 1$ , entonces  $\bar{Y}_t = Y_t$ , es decir, se reproduce la serie y no la suaviza. Con  $\alpha = 0.2$  ya se observa esta característica.

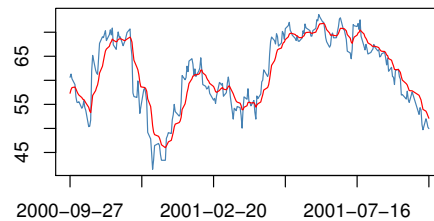
En R la función `emaTA(x, lambda = 0.019, startup = 0)` de la librería `fTrading` implementa EWMA  $\bar{Y}_t = \sum_{j=0}^m \alpha(1 - \alpha)^j Y_{t-j}$ . Se usa la fórmula  $\alpha = \frac{2}{m+1}$ , y a la función se le puede ingresar  $m$  en lugar de  $\alpha$ . El siguiente es un ejemplo de código en R para EWMA.

**Ejemplo 3.3.3.** Código en R para SES (EWMA) con `emaTA`. Los resultados están en la Figura 3.3. 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 3.3: Precios de la Acción de Microsoft Suavizados con SES

### 3.4. La Función Filter

Una función útil para programa medias móviles y suavizadores es la función `filter`. 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`. La exposición siguiente es sobre la función `filter` de la librería `stats`. Cuando se utilice la de la librería `signal` se escribe `signal::filter()`. Esta última se describe en la sección de Notas.

La forma de utilizar esta función es la siguiente.

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

#### Argumentos

- y: Es la serie que se quiere suavizar. Puede ser multivariada.
- w: Es el vector con los pesos  $\underline{w} = (w_0, w_1, \dots, w_m)$ , con  $w_j$  real ó matriz.
- 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"

La opción `method = "convolution"` se utiliza para calcular medias móviles.

**Ejemplo 3.4.1.** Para calcular  $\bar{Y}_t = \frac{1}{m+1} \sum_{j=0}^m Y_{t-j}$  se programa:

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

**Ejemplo 3.4.2.** Para calcular  $\bar{Y}_t = \sum_{j=0}^m w_j Y_{t-j}$

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

La opción "circular=TRUE" requiere más explicación. En la media móvil unilateral se calcula  $\bar{Y}_t = \sum_{j=0}^m w_j Y_{t-j}$ . Si se escribe la serie en la forma  $Y_{t-T+1}, \dots, Y_{t-1}, Y_t$ , que es como la lee filter, y se tienen los  $m+1$  pesos  $w_0, w_1, \dots, w_m$ , al aplicar  $X = \text{filter}(Y, w, \text{"conv"}, 1, T, \text{NULL})$  se obtiene una serie  $X = (X_{t-T+1}, \dots, X_{t-1}, X_t)$ , calculada de la siguiente manera.

$$\begin{aligned} X_t &= w_0 Y_t + w_1 Y_{t-1} + \dots + w_m Y_{t-m} \\ X_{t-1} &= w_0 Y_{t-1} + w_1 Y_{t-2} + w_2 Y_{t-3} + \dots + w_m Y_{t-m-1} \\ &\vdots \\ X_{t-T+1} &= w_0 Y_{t-T+1} + w_1 Y_{t-T} + w_2 Y_{t-T-1} + \dots + w_m Y_{t-T-m+1} \end{aligned}$$

Pero los datos  $Y_{t-T}, \dots, Y_{t-T-m+1}$  no están disponibles. El efecto de la opción TRUE para "circular" es reemplazarlos por los valores  $Y_t, Y_{t-1}, \dots, Y_{t-m+1}$ , respectivamente. De manera similar se reemplazan en las medias bilaterales.

**Ejemplo 3.4.3.** *Ejemplo simple de convolution con reemplazo de valores.*

```
y = c(1,1,3,1); w = c(1,2,3)
x = filter(y,w,"conv",1,T,NULL)
# x tiene cuatro elementos calculados así
x[4] = 1(1) + 2(3) + 3(1) = 10
x[3] = 1(3) + 2(1) + 3(1) = 8
x[2] = 1(1) + 2(1) + 3(1) = 6
x[1] = 1(1) + 2(1) + 3(3) = 12
```

**Ejemplo 3.4.4.** *Para calcular  $\bar{Y}_t = \sum_{j=-m}^m w_{|j|} Y_{t-j}$*

```
# k = 2m+1
w = c(w0,w1, w2, ..., wk)
x = filter(y,w,"conv",2,T,NULL)
```

**Ejemplo 3.4.5.** *Cálculo de EWMA con la función filter. Como en EWMA (3.5) se tiene*

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

*se puede generar X así:*

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

**Ejemplo 3.4.6.** La siguiente media móvil de orden 12 estima la componente  $T_t$  en una serie con componente estacional de período  $p = 12$ .

$$T_t = \frac{1}{12} \left( \frac{1}{2} Y_{t-6} + \sum_{u=-5}^5 Y_{t-u} + \frac{1}{2} Y_{t+6} \right), \quad t = 7, \dots, 45, \quad (3.6)$$

Aplicar este filtro a la serie de tiempo “nottem”, que está por defecto en R, y corresponde a las temperaturas promedio mensuales, en grados Fahrenheit, del castillo de Nottingham, en Inglaterra.

La programación en R del filtro (3.6) se puede hacer de la manera siguiente.

```
#-----
Yt = nottem
ts.plot(Yt)

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');

# Notese que el filtro estima la componente de tendencia
# exactamente como lo hace la función stl()
m = stl(Yt, "per")
Tmt = m$time.series[,2]

par(mfrow = c(1,1))
plot(t, Tt, type='l'); lines(t, Tmt, col='red');
```

### La opción method= “recursive”

La opción method = "recursive" calcula una serie  $X$  definida por

$$X_t = \sum_{j=1}^m w_j X_{t-j} + Y_t, \quad (3.7)$$

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

**Ejemplo 3.4.7.** 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 3.4.8.** 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)
```

### 3.5. Ejemplo de Análisis de Tendencia con Suavizadores

Este programa calcula varios suavizamientos de la serie del índice de la Bolsa de Colombia, IGBC, durante el período 19/01/2009 a 19/08/2010, que fué un período al alza. Los resultados de muestran a en la Figura 3.4 continuación

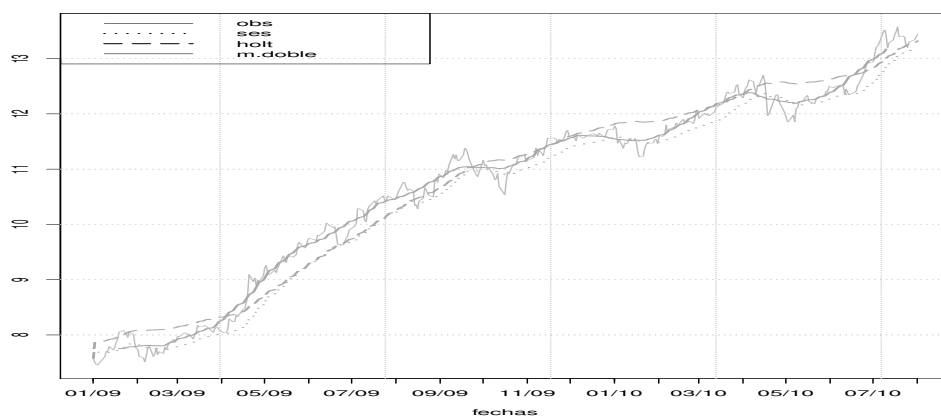


Figura 3.4: Suavizamientos SES, Holt, Media Movil unilateral 30 del IGBC, 2009-2010

```
# analisis del igbc mediante ewma, medias moviles
# bi-unilaterales y filter
```

```
library(forecast)
library(fTrading)
```

```
source("holt.r")
```

```
D = read.table("igbc.dat",
header = TRUE, stringsAsFactors = FALSE)
attach(D)
```

```
n1 = which(fecha=="19/01/2009")
n2 = which(fecha=="19/08/2010")
x = igbc[n1:n2]/1000
f = fecha[n1:n2]
```

```
# suavizamiento exponencial EWMA
```

```
f22 = emaTA(x, lambda = 0.072, startup = 30)
```

```
# suavizamiento Holt
```

```
f.obj = function(theta,yt){
```

```

a = theta[1]
b = theta[2]
n = length(yt)
yet = holt(yt,a,b,n)$yet
err = yt[2:n]-yet[1:(n-1)]
mse = mean(err^2)
return(mse)
}

# parametros que minimizan el MSE
(alpha = optim(
c(0.5,0.5),
f.obj,
lower = rep(1.0e-06,2),
upper = rep(0.9999999,2),
method = "L-BFGS-B",
yt=x)$par)

# otro par de parametros
alfa1 = 0.032
alfa2 = 0.0003
np = length(x)

f33 = holt(x,alfa1,alfa2,np)$yet

# medias bilaterales (filtro henderson)

w13 = c(-0.019, -0.028, 0.0, 0.066, 0.147, 0.214,
0.240, 0.214, 0.147, 0.066, 0.0, -0.028, -0.019)

w23 = c(-0.004, -0.011, -0.016, -0.015, -0.005,
0.013, 0.039, 0.068, 0.097, 0.122, 0.138, 0.148,
0.138, 0.122, 0.097, 0.068, 0.039, 0.013, -0.005,
-0.015, -0.016, -0.011, -0.004)

fw13 = filter(x,w13,"conv",2,F,NULL)
fw23 = filter(x,w23,"conv",2,F,NULL)
f4 = filter(x,rep(1/30,30),"conv",2,F,NULL)

```



```

np = length(x)
fecha = as.Date(f, "%d/%m/%Y")
fechas = strptime(as.character(fecha), "%Y-%m-%d")

plot(fechas,x, xaxt="n",panel.first = grid(),
type='l',ylab='',col='grey')
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)

lines(fechas,f22, xaxt="n", panel.first = grid(),
type='l',lty=3,lwd=2,col='darkgrey')
lines(fechas,f33, xaxt="n", panel.first = grid(),
type='l',lty=5,lwd=2,col='darkgrey')
lines(fechas,f4, xaxt="n", panel.first = grid(),
type='l',lty=7,lwd=2,col='darkgrey')

legend("topleft",
c("obs","ses","holt", "m.doble"),
lty = c(1,3,5,7),
lwd=c(1,2,2,2) )

plot(fechas,x, xaxt="n",panel.first = grid(),
type='l',ylab='',col='grey')
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)

lines(fechas,f13, xaxt="n", panel.first = grid(),
type='l',lty=3,lwd=2,col='darkgrey')
lines(fechas,f23, xaxt="n", panel.first = grid(),
type='l',lty=5,lwd=2,col='darkgrey')
lines(fechas,f4, xaxt="n", panel.first = grid(),

```

```

type='l',lty=7,lwd=2,col='darkgrey')

legend("topleft",
c("obs","hender13","hender23","m.doble"),
lty = c(1,3,5,7),
lwd=c(1,2,2,2) )

```

### 3.6. Notas

#### Sobre 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:

$$a = c(1, a_1, \dots, a_p)$$

$$b = c(b_0, b_1, \dots, b_q)$$

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  $Y = (Y_{t-T+1}, Y_{t-T}, \dots, Y_{t-1}, Y_t)$  se obtiene una serie filtrada  $X = (X_{t-T+1}, \dots, X_{t-1}, X_t)$  definida por:

$$X_t + \sum_{j=1}^p a_j X_{t-j} = \sum_{j=0}^q b_j Y_{t-j}. \quad (3.8)$$

**Ejemplo 3.6.1.** 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 (3.4.8). 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 (3.8) 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.

## Sobre EWMA

Nótese que también a partir de  $X_t = \alpha Y_t + (1 - \alpha)X_{t-1}$ , se puede programar

```
w = 1-a
x = filter(a*y,w,"rec")
n = length(x)
m = floor((2-a)/a)
x = x[(m+1):n]
```

## Método de Suavizamiento de Holt

El método de suavizamiento de Holt consiste en un par de ecuaciones recursivas, con un par de parámetros,  $(\alpha, \beta) \in (0, 1) \times (0, 1)$ .

$$a_t = \alpha Y_t + (1 - \alpha)(a_{t-1} + b_{t-1}), \quad (3.9)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}. \quad (3.10)$$

con  $\hat{Y}_t = a_t + b_t$ , para  $t = 1, \dots, T$ . Los valores iniciales  $a_0, b_0$  se toman como el intercepto y la pendientes estimadas por regresión en el modelo  $Y_t = a + bt + \varepsilon_t$ . Un programa en R para una función en R que retorna  $(\hat{Y}_t, a_t, b_t)$ , se muestra a continuación.

```
#suavizamiento de Holt -----
holt = function(yt,a,b,n){
bst = double(n)
ast = double(n)
yet = double(n)

rw = lm(yt ~ time(yt))
ast[1] = rw$coef[1]
bst[1] = rw$coef[2]

yet[1] =yt[1]
for (i in 2:n)
{
ast[i] = a*yt[i]+(1-a)*(ast[i-1]+bst[i-1])
bst[i] = b*(ast[i]-ast[i-1])+(1-b)*bst[i-1]
yet[i] = ast[i]+bst[i]
```

48

```
}  
D = data.frame(yet,ast,bst)  
return(D) }
```

---

### Modelado y Pronóstico de Series Estacionales

---

#### 4.1. Definiciones

En este capítulo se introducen modelos para la componente estacional  $S_t$  que forma 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 y filtros.

**Definición 4.1.1 (Componente Estacional).** *La componente estacional  $S_t$  se define como una función no aleatoria, periódica de período  $s$ . Los valores de  $S_t$  para  $t = 1, \dots, s$  se denominan “el patrón estacional”. El período estacional  $s$  es el número mínimo de períodos que tarda el patrón estacional en repertirse.*

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 la unidad de tiempo es mes, y se muestrea doce veces al año. De manera similar con  $s = 4$ , trimestral. Un período semanal es  $s = 52$  y se muestrea 52 veces al año. También se puede definir un modelo con  $t$  en días, y  $s = 7$ , en el cual se muestrea  $360 = 52(7)$ , veces al año y se define un patrón estacional que se repite cada 7 días.

**Ejemplo 4.1.1.** *Una ejemplo de series con componente estacional es la serie **nottem**, en la librería stats de R. Corresponden 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. <sup>(1)</sup>

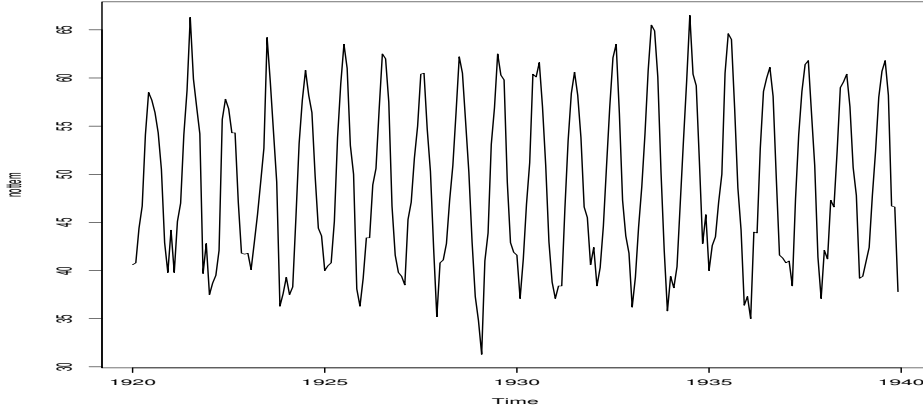


Figura 4.1: Average Monthly Temperatures at Nottingham, 1920-1939

### Propiedades de $S_t$

En todo lo que sigue  $S_t$ ,  $t = 1, 2, \dots$  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  $s$ ,  $S_{t+s} = S_t$  para  $t = 1, 2, \dots$ . Por tanto, sólo se requiere definir  $S_t$  en los primeros  $s$  valores,  $S_t$ ,  $t = 1, 2, \dots, s$ . 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  $s$ . 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  $j = 1, \dots, s$ ,  $t = 1, 2, \dots$ , definidas por

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

<sup>1</sup><http://www.astrostatistics.psu.edu/datasets/R/html/datasets/html/00Index.html>

O, de manera equivalente,  $I_j(t) = 1$ , si  $t = s\lfloor t/s \rfloor + j$ , donde  $\lfloor x \rfloor$  es la parte entera de  $x$ ,  $I_j(t) = 0$  en otro caso. Por ejemplo, si  $s = 4, j = 3, t = 11$ , entonces como  $11 = 4\lfloor 11/4 \rfloor + 3$ , se tiene  $I_3(11) = 1$ . También se puede usar el concepto de “congruencia módulo  $s$ ”,  $I_j(t) = 1$  si  $t \equiv j \pmod{s}$ , es decir, si  $t - j$  es divisible por  $s$ . Por tanto, si  $S_t$  es periódica de período  $s$ , existen constantes  $\delta_j, j = 1, \dots, s$  tales que

$$S_t = \sum_{j=1}^s \delta_j I_j(t), \quad t = 1, 2, \dots \quad (4.2)$$

4. También se puede utilizar el conjunto  $\{\sin(2\pi jt/s), \cos(2\pi jt/s), j = 1, \dots, s\}$ . Para  $S_t$  periódica de período  $s$ , escogemos  $m \geq 1$  y constantes  $b_{1,j}, b_{2,j}, j = 1, \dots, m$  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  $1 \leq m \leq \frac{s}{2} - 1$ , si  $s$  es par, o por  $1 \leq m \leq \frac{s-1}{2}$  si  $s$  es impar. Valores  $m = 1, 2$  son recomendables por parcimonia. Note que con  $m = 2$  habrían cuatro parámetros  $\beta_{1,j}, \beta_{2,j}$  a estimar mediante regresión lineal.
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 y sería preferible el modelo con variables trigonométricas.

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 para  $Y_t$  está dado por:

$$Y_t = \beta_0 + \sum_{j=1}^k \beta_j t^j + \sum_{j=1}^s \delta_j I_j(t) + \varepsilon_t, \quad (4.3)$$

donde  $k = 1, 2, 3$  y  $\beta_j, j = 1, \dots, k$  y  $\delta_j, j = 1, \dots, s$  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.

## 4.2. Estimación de la Componente Estacional

### 4.2.1. Modelos con Variables Indicadoras

A continuación se enumeran los pasos para estimar la componente  $S_t$ , asumiendo primero el modelo con variables indicadoras (4.2).

1. Identificar el período de la serie  $Y_t$ ,  $s$ , si es que existe. Con R al declarar la serie como objeto “ts” con frecuencia, por ejemplo 12, ya está identificado el posible período como  $s = 12$ .
2. Generar las  $s$  variables indicadoras  $I_j(t)$  para  $t = 1, 2, \dots, T$ . Es decir, generar la matriz  $I_j(t)$ , donde al variar  $j = 1, \dots, s$  variamos las columnas y al variar  $t = 1, \dots, T$  varían las filas.
3. Estimar los parámetros en el modelo  $Y_t = T_t + \sum_{j=1}^s \delta_j I_j(t) + \varepsilon_t$  mediante regresión lineal. Nótese que se trata de estimar los parámetros de las componentes de tendencia y estacional, conjuntamente.

La función  $T_t + S_t$  se denominará “componente estructural”, y los residuales  $\varepsilon_t$  se denominarán “residuos estructurales”.

### El Problema de la Multicolinealidad con Variables Indicadoras

Se deben incluir 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.

El modelo(4.3) se modifica y queda

$$Y_t = \beta_0 + \sum_{j=1}^k \beta_j t^j + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t, \quad (4.4)$$

ver Diebold [1999, pág. 90].

### Estimación en R del modelo con Variables Indicadoras

En la librería `forecast` hay varios recursos y funciones para estimar y pronosticar con el modelo (4.3). Suponga que  $Y_t$  tiene una componente de período  $s$  que se declara con frecuencia  $s$ .

El primer paso consiste en generar las variables indicadoras  $I_j(t)$ ,  $j = 1, \dots, s$ ,  $t = 1, \dots, T$ . 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  $Y_1$  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 uno. De acuerdo a este valor inicial se genera la matriz  $It$ . 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  $T$ . 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` es interesante analizar el siguiente código en R que calcula la matriz  $I_j(t)$  de dimensión  $T \times s$ .<sup>(2)</sup>

**Ejemplo 4.2.1.** #-----

```
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 4.2.2.** *Ejemplo con R para modelación de la serie de producción de cemento Portland, trimestral, en toneladas por mil, entre Q1 1956 y Q3 1994, en Australia. La serie se asume que es de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t. \quad (4.5)$$

*Es decir, el modelo a estimar es lineal para la tendencia, más una componente estacional.*

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

---

<sup>2</sup>autor : Ledwing Osorio Cárdenas

```

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  $Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t) + \varepsilon_t$ , 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  $\bar{R}^2 = 0.8966$ , el ajuste resulta aceptable. Sin embargo, la componente estacional  $S_t$  parece no ser constante.

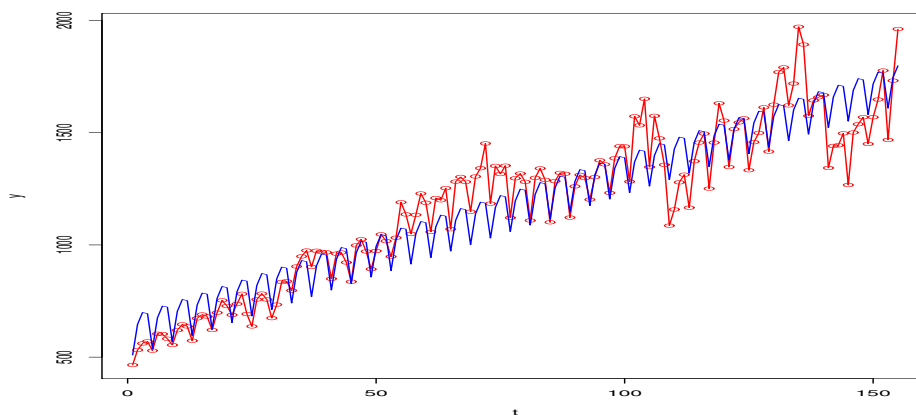


Figura 4.2: Ajuste con Variables Indicadoras

#### 4.2.2. Modelos con Funciones Trigonómicas

En este caso el modelo a estimar es:

$$Y_t = \beta_0 + \sum_{j=1}^m \beta_j t^j + \sum_{j=1}^k \beta_{1,j} \sin(2\pi j t / s) + \beta_{2,j} \cos(2\pi j t / s) + \varepsilon_t, \quad (4.6)$$

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

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  $\text{sen}(2\pi it/s)$ ,  $\text{cos}(2\pi it/s)$ . Para calcular estas variables se utiliza la función `It.trig = fourier(y, j)`, donde  $j = 1, 2, \dots$ , la cual calcula los armónicos  $\text{sen}(2\pi jt/s)$ ,  $\text{cos}(2\pi jt/s)$ , para cada  $t = 1, 2, \dots, T$ .

**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.*

*Se asume que la serie de precios es la suma de cuatro componente: componente lineal ( $a + bt$ ), componente estacional semanal ( $s_t$ ), componente estacional anual ( $S_t$ ), y componente aleatoria ( $e_t$ ):*

$$Y_t = a + bt + s_t + S_t + e_t, \quad (4.7)$$

$$s_t = \sum_{j=1}^6 \delta_j I_j(t),$$

$$S_t = \alpha \cos\left(\frac{2\pi t}{365}\right) + \beta \text{sen}\left(\frac{2\pi t}{365}\right),$$

*En este modelo la componente con período anual se modela con funciones trigonométricas y la componente con período semanal se modela con variables indicadoras. Cuando se aplica la descomposición con la función `stl()` ésta estima la componente de tendencia como  $a + bt + S_t$ . El siguiente código en R implementa el modelo (4.7).*

```
# 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 = lm(y ~ t + St + x1 + x2)
summary(modelo1)

```

**Ejemplo 4.2.4.** *Continuando con el Ejemplo (4.2.2), ahora se estima el modelo (5.13), pero utilizando funciones trigonométricas.*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^2 \beta_{1,j} \sin(2\pi j t / 4) + \beta_{2,j} \cos(2\pi j t / 4) + \varepsilon_t. \quad (4.8)$$

*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.*

```

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

It.trig = fourier(y,2)
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,2,8)
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 Trigonómicas

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

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.

### 4.3. Procedimientos Recursivos de Estimación para Diagnosticar Estabilidad Estructural en Modelos de Pronósticos

«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.» Diebold [1999, pág. 92]

Suponga el modelo de regresión lineal

$$Y_t = \sum_{i=1}^k \beta_i X_{i,t} + \varepsilon_t, \quad t = 1, \dots, T, \quad (4.9)$$

donde  $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ .

Si los parámetros  $\beta_i$  no fueran constantes sino que cambiarían con el tiempo  $t$ , de manera aleatoria o no, se tendría  $Y_t = \sum_{i=1}^k \beta_{i,t} X_{i,t} + \varepsilon_t$ , un modelo con coeficientes variables.

Note que (4.9) también se puede escribir como  $Y_t = \underline{X}_t' \underline{\beta} + \varepsilon_t$  con  $\underline{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{k,t})$  vector fila en  $\mathbb{R}^k$  y  $\underline{\beta} \in \mathbb{R}^k$  vector de parámetros. Pero en caso de que los parámetros varíen en el tiempo el modelo quedaría  $Y_t = \underline{X}_t' \underline{\beta}_t + \varepsilon_t$ .

**Definición 4.3.1 (Estabilidad Estructural).** Si  $Y_t = \sum_{i=1}^k \beta_i X_{i,t} + \varepsilon_t$ ,  $t = 1, \dots, T$ , la estabilidad estructural del modelo se define como el supuesto de que los coeficientes  $\beta_i$  se mantengan constantes en cada tiempo  $t$ . Lo contrario es que algún  $\beta_i$  cambie con  $t$ . La primera afirmación es la hipótesis nula y la segunda es la alterna.

**Ejemplo 4.3.1.** Algunos ejemplos de cambio estructural en la literatura, debidos a intervenciones en fechas específicas, son

- *EL nivel del río 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 medida 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 de 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  $O_3$ .*

Dos técnicas para chequear el supuesto de estabilidad estructural son la estimación recursiva de los parámetros del modelo y las pruebas tipo CUSUM. El objetivo al aplicar estas técnicas es asegurar que el modelo para pronósticos es estable. En caso de no tener estabilidad estructural los pronósticos con un modelo de regresión global podrían no ser confiables. En este caso sería recomendable utilizar un procedimiento adaptativo como Holt, Holt-Winters, Loess, ó un modelo de variables estructurales.

### Estimación Recursiva de los Parámetros

**Definición 4.3.2 (Estimadores Recursivos).** *En lugar de utilizar todos los datos para la estimación ( $t = 1, \dots, T$ ), si el modelo tiene  $k$  parámetros se utilizan los  $k$  primeros datos y se estiman los parámetros  $\hat{\beta}_{j,k}$ . Luego se utilizan los primeros  $k + 1$ , y se estiman los  $\hat{\beta}_{j,k+1}$ , y así sucesivamente hasta utilizar los  $T$  datos. Los parámetros obtenidos  $\hat{\beta}_{j,t}$   $t = k, \dots, T$  se denominan Estimadores Recursivos.*

Para implementar en R la estimación recursiva se utilizan las siguientes instrucciones. Suponga que el número de parámetros es  $k$  y el total de datos es  $T$  y que se trata de un modelo con tendencia lineal y estacional, dado por  $Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_j(t) + \epsilon_t$ ,  $t = 1, \dots, T$ . El total de parámetros es  $k = 2 + s - 1$ . Se toman 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.

```
#-----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 4.3.2.** *Con base en el Ejemplo (4.2.2), 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 (5.13), pag. 86, están en la Figura 4.5.*

*Lo que puede concluirse 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 concluir de este diagnóstico que existe evidencia de inestabilidad estructural en el modelo.*

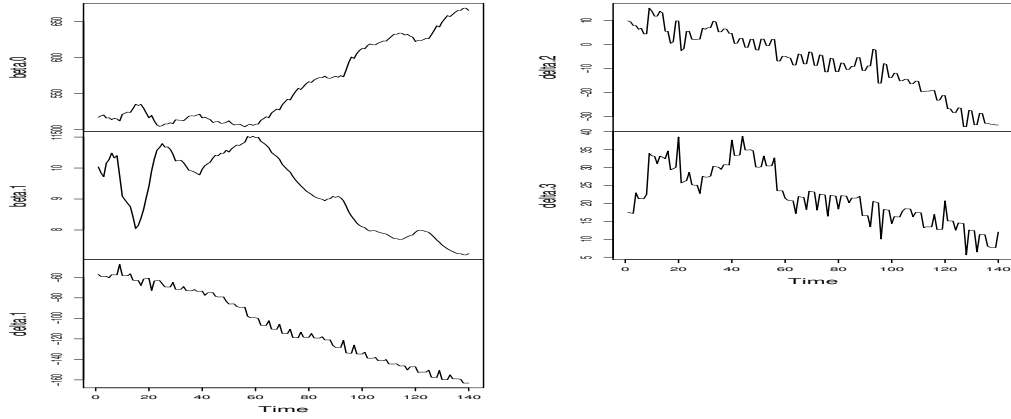


Figura 4.3: Estimadores Recursivos para el Modelo de Producción de Cemento

### Pruebas CUSUM

**Definición 4.3.3 (Residuales Recursivos.).** Para cada  $t = k, k + 1, \dots, T - 1$  se calcula el pronóstico a un paso,  $\hat{Y}_t = \sum_{j=1}^k \hat{\beta}_{j,t} X_{j,t}$  donde  $\hat{\beta}_{j,t}$  son los estimadores recursivos en  $t$ . y se forma el Residual Recursivo

$$\hat{\varepsilon}_{t+1,t} = Y_{t+1} - \hat{Y}_t, \quad (4.10)$$

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

$$\hat{\varepsilon}_{t+1,t} \sim N(0, \sigma^2 r_t), \quad (4.11)$$

con  $r_t = 1 + \underline{X}'_{t+1} \left( \sum_{j=1}^t \underline{X}_j \underline{X}'_j \right)^{-1} \underline{X}_{t+1}$ .

Los residuales recursivos estandarizados se definen como

$$W_{t+1,t} = \frac{\hat{\varepsilon}_{t+1,t}}{\hat{\sigma} \sqrt{r_t}}, \quad \hat{\sigma}^2 = MSE, \quad (4.12)$$

y se tiene que  $W_{t+1,t} \stackrel{i.i.d.}{\sim} N(0, 1)$ .

**Definición 4.3.4 (CUSUM).** La suma acumulada de los residuos recursivos ó CUSUM, se define como

$$CUSUM_t = \sum_{j=1}^t \frac{W_{i+1,i}}{\sigma_W}, \quad t = k, k + 1, \dots, T - 1, \quad (4.13)$$

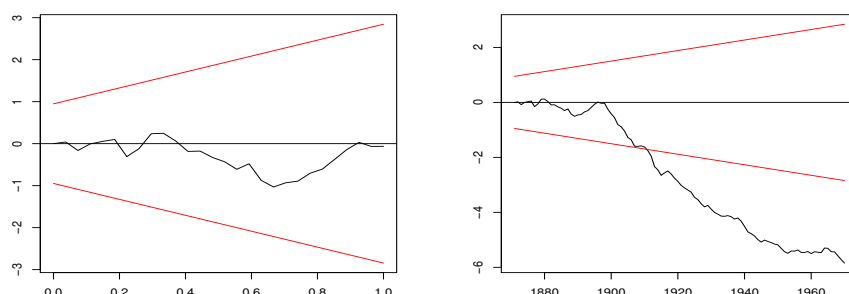
donde

$$\sigma_W = \sqrt{\frac{\sum_{i=k}^{T-1} (W_{i+1,i} - \bar{W})^2}{T - k}}, \quad \bar{W} = \frac{1}{T - k} \sum_{i=k}^{T-1} W_{i+1,i}.$$



**Uso:** El estadístico  $CUSUM_t$  se utiliza para chequear la *estabilidad estructural del modelo*.

La región de aceptación de la prueba es una banda formada por 2 líneas rectas.



(a) No se rechaza  $H_0$

(b) Se rechaza  $H_0$

Figura 4.4: Región de Rechazo de la Prueba  $CUSUM$ .

La expresión para las rectas es

$$\pm a \left( \sqrt{T-k} + 2 \frac{t-k}{\sqrt{T-k}} \right) \quad (4.14)$$

con  $a = 0.948$  para un nivel  $\alpha = 0.05$ . En las Figuras 4.4 se muestra gráficamente los dos casos posibles.

**Definición 4.3.5.** La suma acumulada de residuos recursivos cuadrados ó  $CUSUMSQ$  se define como

$$CUSUMSQ_t = \frac{\sum_{i=k}^t W_{i+1,i}^2}{\sum_{i=k}^T W_{i+1,i}^2}, \quad t \in [k, T] \quad (4.15)$$

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. Introducido por Brown et al. [1975], es una herramienta estándar en el análisis de series económicas.

### Implementación en R: la librería strucchange

El objeto de esta librería son las pruebas de cambios estructurales en modelos de regresión.

### Algunas 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.

```
rrc = recresid(y ~ t + It)
```

2. **efp:** Calcula los residuales recursivos junto con la región de aceptación (Bandas). No es propiamente una prueba de hipótesis sino un diagnóstico gráfico. Hay dos versiones. Una que utiliza los residuales recursivos y otra con base en los residuales de la regresión por Mínimos Cuadrados.

```
prueba.cusum = 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 utilice para calcular.

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

4. **breakpoints:** Suponga el modelo  $Y_t = \underline{X}_t' \beta + \varepsilon_t$ ,  $t = 1, \dots, T$  se asume  $m$  puntos de quiebre o de cambio en los cuales los coeficientes cambian de un valor a otro, permaneciendo constantes entre los puntos. Los puntos de quiebre son tales que

$$Y_i = \underline{X}_i' \beta_j + \mu_i, \quad i_{j-1} + 1 \leq i \leq i_j, \quad j = 1, \dots, m + 1. \quad (4.16)$$

Los puntos de quiebre (el índice)  $i_j$  se estiman con la función `breakpoints`.

```
bp.n = breakpoints(y2 ~ t + It, breaks=2)
summary(bp.n)
# en el vector B estan las observaciones que
# corresponden a cambios estructurales
B= bp.n$extract.breaks(bp.n$RSS.table,breaks=2)
```

**Ejemplo 4.3.3.** Con base en el Ejemplo (4.2.2), de la serie de producción trimestral de cemento, la gráfica del estadístico CUSUM para el modelo con tendencia lineal y variables indicadoras estacionales (5.13), pag. 86, está en la Figura 4.5.

A partir de los resultados en la Figura 4.5 se rechazaría la hipótesis nula de estabilidad estructural en el modelo, ya que el estadístico CUSUM se sale de las bandas de confianza. La prueba formal confirma el rechazo de la hipótesis nula de estabilidad.

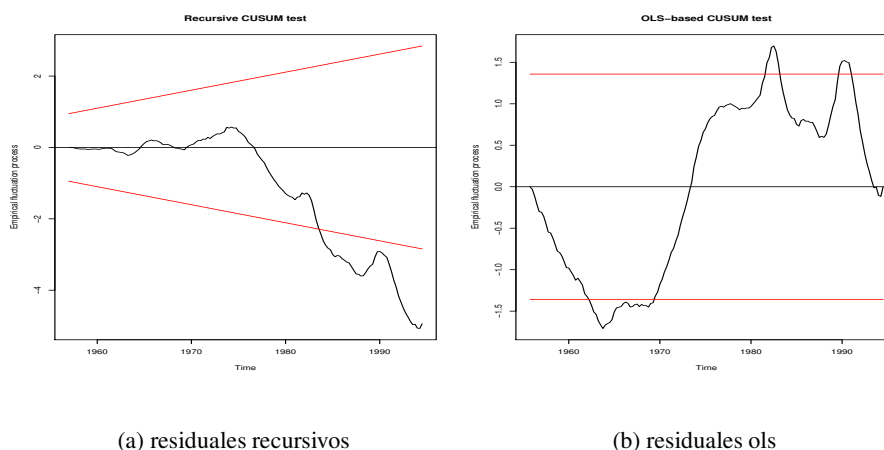


Figura 4.5: Resultados prueba CUSUM gráfica

```
data: prueba.cusum
S = 1.703, p-value = 1.782e-05
```

*La conclusión es que el modelo (5.13), pag. 86 detecta un cambio estructural en la serie. El punto de quiebre se detecta en el cuarto trimestre de 1974 (la observación número 76). Ver el código en R a continuación para la aplicación de la función `breakpoints`. En este caso concluimos 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 (4.27). También es posible ampliar el modelo de descomposición para incluir una intervención que modele el cambio en la observación 76. Pero el análisis de intervenciones es tema de otro capítulo.*

*El código R siguiente muestra la programación de los comandos con los cuales se generaron los análisis y las gráficas de esta parte*

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

E = read.table("F:/Curso pronósticos/Datos/cementq.dat",
header = TRUE)
attach(E)
```

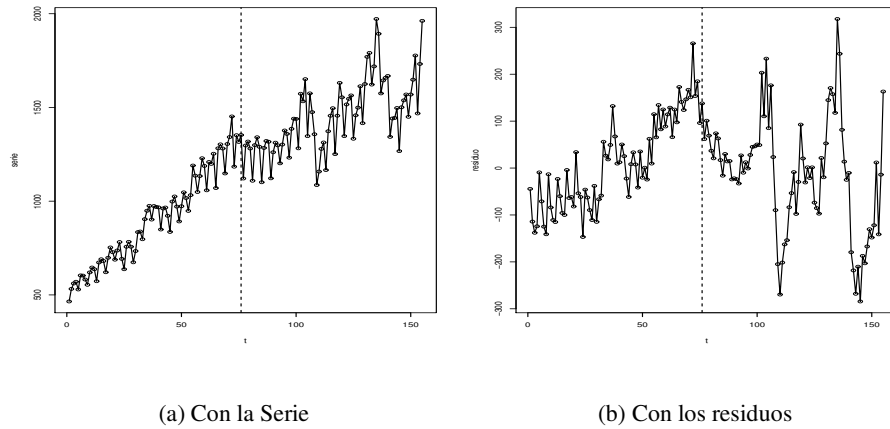


Figura 4.6: Punto de Quiebre 1974-Q4 en la Serie y en los Residuos

```
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)

#----pruebas cusum formales con valor p
sctest(prueba.cusum1)
sctest(prueba.cusum2)

#----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)
#-----
```

#### 4.4. Métodos de Filtrado para la Componente Estacional

En esta sección se exponen dos métodos para estimar la componente  $S_t$  de la serie  $Y_t = T_t + S_t + \varepsilon_t$ ,  $t = 1, \dots, T$ . Una con base en filtros lineales, de la forma

$$\hat{S}_t = \sum_{j=-r}^r u_j Y_{t+j}, \quad t = r+1, \dots, T-r, \quad (4.17)$$

Y otra con base en un algoritmo recursivo, que suaviza la serie sin eliminar la componente estacional.

Existen otras técnicas de ajuste estacional que son técnicas ad-hoc para calcular índices 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](http://www.census.gov/srd/www/x12a)

##### 4.4.1. Algoritmo de Brockwell y Davis

Este Algoritmo está en en Brockwell and Davis [2002, pág. 24] y se describe a continuación. Con base en la serie  $Y_1, \dots, Y_T$ , primero se estima la tendencia con una media móvil diseñada para eliminar la componente estacional y disminuir el ruido aleatorio, definida por:

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

$$\hat{T}_t = s^{-1} \left( \frac{1}{2} Y_{t-q} + Y_{t-q+1} + \dots + Y_t + \dots + Y_{t+q-1} + \frac{1}{2} Y_{t+q} \right), \quad (4.18)$$

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

$$\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}, \quad . \quad (4.19)$$

Segundo, se estima  $\hat{S}_k$  para cada  $k = 1, 2, \dots, s$ . Es decir, se estima el patrón estacional, promediando los valores de  $Y_t - \hat{T}_t$  para todos los  $t$  que corresponden a la estación  $k$  (son los  $t = s[t/s] + k$ ). Con  $q = \lfloor s/2 \rfloor$  se define

$$\hat{S}_k^{(1)} = \frac{\sum_{j=a}^b Y_{k+sj} - \hat{T}_{k+sj}}{b-a+2}, \quad (4.20)$$

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

$$\hat{S}_k = \hat{S}_k^{(1)} - \frac{1}{q} \sum_{k=1}^s \hat{S}_k^{(1)}, \quad k = 1, 2, \dots, s, \quad (4.21)$$

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  $\hat{S}_t$ ,  $t = 1, \dots, T$ .

**Ejemplo 4.4.1.** Tomando por ejemplo  $T = 250$ ,  $s = 12$  y  $q = 6$  entonces

$$\hat{T}_t = \frac{1}{12} \left( \frac{1}{2} Y_{t-6} + Y_{t-5} + \dots + Y_{t+5} + \frac{1}{2} Y_{t+6} \right),$$

con  $7 \leq t \leq 244$ , entonces se tiene que

$$\hat{S}_k^{(1)} = \frac{\sum_{j=a}^b Y_{k+12j} - \hat{T}_{k+12j}}{b-a+2},$$

donde

$$a = \left\lfloor \frac{(6-k)_+}{12} \right\rfloor + 1,$$

$$b = \left\lfloor \frac{(250-6-k)}{12} \right\rfloor.$$

Para  $k = 1$

$$\begin{aligned} \hat{S}_1^{(1)} &= \frac{1}{21} \sum_{j=1}^{20} Y_{1+12j} - \hat{T}_{1+12j} \\ &= \frac{1}{21} \left( (Y_{1+12} - \hat{T}_{1+12}) + (Y_{1+24} - \hat{T}_{1+24}) + \dots + (Y_{1+240} - \hat{T}_{1+240}) \right). \end{aligned}$$

Note que  $\hat{T}_{244}$  es el último valor de  $\hat{T}_t$ , usamos  $\hat{T}_{241}$  con  $Y_{241}$ . Para  $k = 12$

$$\begin{aligned}\hat{S}_{12}^{(1)} &= \frac{1}{20} \sum_{j=0}^{19} Y_{12+12j} - \hat{T}_{12+12j} = \frac{1}{20} \sum_{j=0}^{19} Y_{12(1+j)} - \hat{T}_{12(1+j)} \\ &= \frac{1}{20} \left( (Y_{12} - \hat{T}_{12}) + (Y_{24} - \hat{T}_{24}) + \cdots + (Y_{240} - \hat{T}_{240}) \right).\end{aligned}$$

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

```
seascomp = function(x,p) {
#-----algoritmo Brockwell-Davis
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)
}
```

#### 4.4.2. Método de Holt-Winters

Suponga una serie de la forma  $Y_t = \beta_{0,t} + \beta_{1,t}t + S_t + \varepsilon_t$ , para  $t = 1, \dots, T$ , donde  $\beta_{0,t}, \beta_{1,t}$  varían lentamente, y  $S_t$  es la componente estacional de período  $s$ , que se asume varía lentamente también. El método de Holt-Winters es un algoritmo que permite estimar recursivamente  $\mu_t = \beta_{0,t}$ ,  $\beta_t = \beta_{1,t}$  y  $S_t$  a partir de unos valores iniciales. Es una generalización del método de Holt (3.9) pag. 47, y fué introducido por Winters, Winters [1960]. El algoritmo es el siguiente.

$$\mu_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(\mu_{t-1} + \beta_{t-1}), \quad (4.22)$$

$$\beta_t = \beta(\mu_t - \mu_{t-1}) + (1 - \beta)\beta_{t-1}, \quad (4.23)$$

$$S_t = \gamma(Y_t - \mu_t) + (1 - \gamma)S_{t-s}, \quad (4.24)$$

donde  $t = 2, \dots, T$ , y  $\alpha, \beta, \gamma \in (0, 1)$  son las constantes de suavizamiento y  $\mu_0, \beta_0$  y  $S_1, \dots, S_s$  son valores iniciales dados. El valor suavizado de la serie en  $t$  es  $\hat{Y}_t = \mu_t + \beta_t + S_t$ .

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

$$Y_t = \sum_{j=1}^s \delta_j I_j(t) + \varepsilon_t, \quad (4.25)$$

$$\hat{\varepsilon}_t = \beta_0 + \beta_1 t + \eta_t, \quad (4.26)$$

es decir,  $S_j = \hat{\delta}_j$ ,  $\beta_0 = \hat{\beta}_1$ ,  $\mu_0 = \hat{\beta}_0 + \hat{\beta}_1$ .

Las constantes  $\alpha, \beta, \gamma$  se escogen de manera que se minimice 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, varían lentamente, se escogen cercanas a cero”.

**Ejemplo 4.4.2.** *Continuando con el Ejemplo (4.2.2), de la serie de producción trimestral de cemento, se incluye el suavizamiento y 8 pronósticos con el método de Holt-Winters. Los pronósticos se mostraron en el Cuadro 4.2 anterior. Se puede observar el suavizamiento en la Figura (4.7) Adicionalmente, la gráfica de los ocho pronósticos con Holt-Winters se compara con los de los modelos con variables indicadoras y trigonométricas en la Figura (4.8). En el Cuadro 4.8 se pueden ver el MAPE para cuatro métodos. Entonces se puede concluir que los modelos con variables indicadoras y con trigonométricas producen pronósticos muy similares, pero aún superiores a los generados con Holt-Winters. En este caso se puede concluir que el modelo con variables indicadoras sería preferible para pronosticar.*

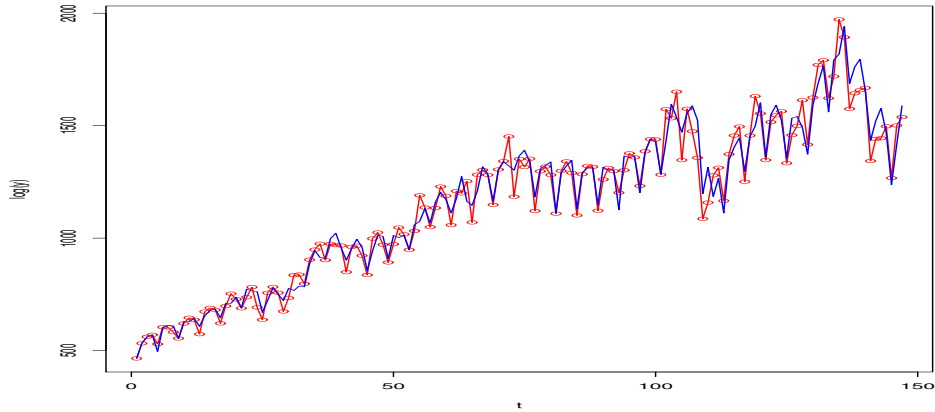


Figura 4.7: Ajuste con el Método de Holt-Winters

	Holt-Winters	Ind	Trig	Exp-Est
ME	-143.84	84.33	84.77	-113.08
RMSE	176.71	137.14	137.68	163.16
MAE	144.74	122.26	123.99	121.55
MPE	-9.52	5.03	5.06	-7.31
MAPE	9.58	7.13	7.23	7.84

## 4.5. Notas

1. **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 ( $S_{\hat{\beta}_i}$ ) no son confiables. Pero en el ejemplo la prueba Durbin-Watson conduce al 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.

2. El Modelo de Componentes Estructurales de Harvey, Durbin y Koopmans se define

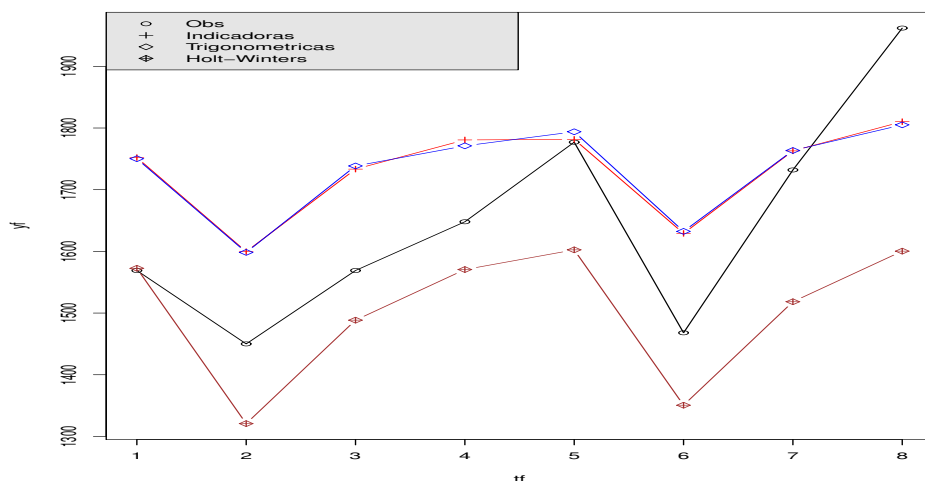


Figura 4.8: Pronósticos Varios Métodos

como.

$$x(t) = m(t) + s(t) + \epsilon(t), \quad (4.27)$$

$$m(t+1) = m(t) + n(t) + \xi(t), \quad (4.28)$$

$$n(t+1) = n(t) + \zeta(t), \quad (4.29)$$

$$s(t+1) = -s(t) - \dots - s(t-s+2) + w(t), \quad (4.30)$$

$$\epsilon(t) \sim N(0, s_\epsilon^2), \xi(t) \sim N(0, s_\xi^2), \zeta(t) \sim N(0, s_\zeta^2), w(t) \sim N(0, s_w^2).$$

Este modelo es una alternativa a Holt-Winters. Se estima mediante el Filtro de Kalman, y está implementado en la función `StructTS(y, type="BSM")`.

## 4.6. Problemas

1. (ver Brockwell and Davis [2002, pag.43, ex. 1.14]) Muestre que el filtro con coeficientes  $[a_{-2}, a_{-1}, a_0, a_1, a_2] = [-1, 4, 3, 4, -1]$  pasa polinomios de grado 3 y elimina componentes estacionales de período 3.



---

## Validación de los Supuestos sobre los Errores

---

Al considerar una serie de tiempo  $Y_t$  se asumirá que el índice  $t$  toma valores enteros negativos y positivos,  $t = 0, \pm 1, \pm 2, \dots$ , y se escribe  $(Y_t, t \in \mathbb{Z})$ . Los valores observados de  $Y_t$  forman un conjunto finito que se denomina una *trayectoria muestral*, y se escribe  $\{Y_1, Y_2, \dots, Y_T\}$ .

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  $\hat{\varepsilon}_t$ , denominada “residuos estructurales”, resulta de sustraer de  $Y_t$  las estimaciones de las dos primeras componentes, es decir

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

Para calcular los pronósticos  $\hat{Y}_{T+j}$ ,  $j = 1, \dots, h$ , se debe analizar la serie  $\hat{\varepsilon}_t$  para detectar si tiene una dinámica auto-correlacionada ó, por el contrario, es solamente ruido aleatorio. En el primer caso es posible pronosticar los residuos estructurales e incorporarlos a los pronósticos estructurales, es decir, 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 5.0.1 (Ruido Blanco).** Una serie  $(\varepsilon_t, t \in \mathbb{Z})$  se dice que es Ruido Blanco si cumple

1.  $E(\varepsilon_t) \equiv 0$ , (Media cero)
2.  $Var(\varepsilon_t) \equiv \sigma^2$ , (Varianza constante)
3.  $\forall k \neq 0, Cov(\varepsilon_t, \varepsilon_{t+k}) = 0$ , (Incorrelación)

Si además cumple que  $\varepsilon_t \sim N(0, \sigma^2)$  se dice que  $\varepsilon_t$  es Ruido Blanco Gaussiano (RBG).

Se denota  $\varepsilon_t \sim RB(0, \sigma^2)$  ó  $\varepsilon_t \sim RBG(0, \sigma^2)$ . Nótese que es este último 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. Por tanto también se cumple  $E(\varepsilon_{t+k}|\varepsilon_t) = E(\varepsilon_{t+k}) = 0$ , es decir, 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. Mientras que la condición que una serie sea ruido blanco implica que su pronóstico es cero. Por ejemplo, con  $T_t = \beta_0 + \beta_1 t$

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, \dots, T, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

$$\hat{Y}_{T+k} = \hat{\beta}_0 + \hat{\beta}_1(T+k) + \hat{\varepsilon}_{T+k}, \quad k = 1, 2, \dots, h,$$

pero  $\hat{\varepsilon}_{T+k} = E(\varepsilon_{T+k}|\varepsilon_T) = 0$ , luego

$$\hat{Y}_{T+k} = \hat{\beta}_0 + \hat{\beta}_1(T+k), \quad k = 1, 2, \dots, h,$$

que es la Ecuación de Pronósticos final.

La condición  $E(\varepsilon_t) \equiv 0$  se cumple automáticamente si  $\varepsilon_t$  es el residuo de la regresión  $Y_t = T_t + \varepsilon_t$ . Las otras dos condiciones se comprueban aplicando pruebas de hipótesis adecuadas. A estas comprobaciones se las denomina: “Validación de los supuestos de los errores”. El resto del capítulo se dedicará a exponer pruebas para detectar la condición de incorrelación.

Para la condición de varianza constante, es decir para la condición de homocedasticidad, las pruebas disponibles en la literatura, por ejemplo, White, Breusch-Pagan y Goldfeld-Quandt, asumen que la hipótesis nula es  $H_o : Var(\varepsilon_t) \equiv \sigma^2$  versus la hipótesis alterna  $H_a$  de que  $Var(\varepsilon_t)$  es una función de una combinación de las variables explicativas, por ejemplo,  $t$ ,

$t^2$ , las variables indicadoras en las columnas de la matriz  $It$ , ó de los valores estimados  $\hat{Y}_t$ , que es una combinación particular. Esta hipótesis alterna no es útil ya que lo que se busca es precisamente eliminar el efecto de tales variables en los residuos estructurales. Rechazar esa hipótesis nula 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 y concentrarnos en las pruebas de incorrelación.

## 5.1. Series Estacionarias en Covarianza

Para determinar si  $\hat{\varepsilon}_t, t = 1, \dots, T$  presenta autocorrelación o no, se requieren algunas pruebas de hipótesis y dos estadísticos conocidos como la función de autocovarianza y la función de autocorrelación. Para introducir estas dos funciones se requieren algunas definiciones. La definición de serie estacionaria es básica. Es diferente del concepto de serie estacional. Se darán con base en una serie  $Y_t, t \in \mathbb{Z}$  en general pero pensando en aplicarlas al caso concreto  $Y_t = \hat{\varepsilon}_t$ .

**Definición 5.1.1 (Serie de Tiempo Estacionaria en Covarianza).** *Una serie  $(Y_t, t \in \mathbb{Z})$  se dice estacionaria en covarianza o simplemente estacionaria si cumple dos condiciones:*

1.  $E(Y_t) \equiv \mu$ ,
2.  $Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1)$ , con  $R(t)$  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  $t_1$ . Como puede darse que  $t_1 < t_2$  ó  $t_1 > t_2$ , esta distacia es el valor absoluto  $|t_2 - t_1|$ .*

La condición 2. establece que debe existir una función  $R(h)$ , que sea par, es decir,  $R(h) = R(-h)$ , tal que  $Cov(Y_t, Y_{t+h}) = R(h)$ . Por tanto se debe cumplir una relación de simetría simple:  $Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1) = R(t_1 - t_2) = Cov(Y_{t_2}, Y_{t_1})$ . Por ejemplo,

$$\begin{aligned} Cov(Y_{-4}, Y_3) &= R(3 - (-4)) = R(7) = Cov(Y_0, Y_7), \\ Cov(Y_3, Y_4) &= R(4 - 3) = R(1) = Cov(Y_0, Y_1). \end{aligned}$$

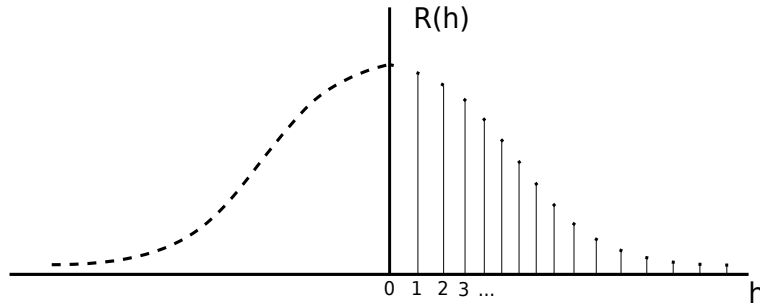


Figura 5.1: Ejemplo Función de Autocovarianza

Nótese que por ser funciones pares las autocovarianzas solamente se grafican en el semieje positivo. Las siguientes identidades son inmediatas.

$$\begin{aligned}
 Cov(Y_t, Y_{t+h}) &= E((Y_t - \mu)(Y_{t+h} - \mu)) \\
 &= E(Y_t Y_{t+h}) - \mu E(Y_t) - \mu E(Y_{t+h}) + \mu^2 \\
 &= E(Y_t Y_{t+h}) - \mu^2 - \mu^2 + \mu^2 \\
 &= E(Y_t Y_{t+h}) - \mu^2,
 \end{aligned}$$

Por lo tanto  $R(h) = E(Y_t Y_{t+h}) - \mu^2$ , ó también,  $R(t_2 - t_1) = E(Y_{t_2} Y_{t_1}) - \mu^2$ .

**Definición 5.1.2 (Función de Autocorrelación).** La función de autocorrelación teórica de una serie estacionaria en covarianza ( $Y_t$ ,  $t \in \mathbb{Z}$ ) se define como

$$\rho(h) = \frac{R(h)}{R(0)}, \quad h = 0, 1, 2, \dots$$

La función de autocorrelación se indicará con las siglas FAC (ACF en inglés).

Como  $|\rho(h)| \leq 1$  se cumple entonces 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  $\sigma^2 < \infty$ .

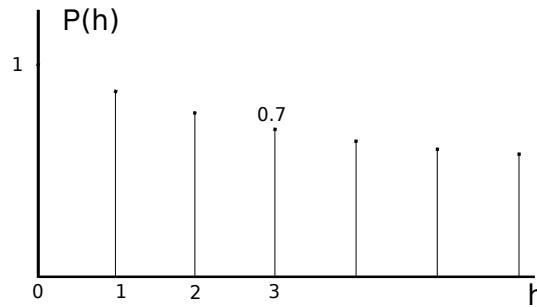


Figura 5.2: Ejemplo Función de Autocorrelación.



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  $Y_t$  fuera, por ejemplo, un índice de producción trimestral, este resultado permite concluir que el índice del primer y el cuarto trimestre tienen una asociación lineal alta.

Según Diebold [1999, pág. 110] si la serie  $(Y_t, t \in \mathbb{Z})$  es estacionaria en covarianza y no tiene componentes determinísticos entonces  $\rho(k) \rightarrow 0, k \rightarrow \infty$ . Un componente determinístico es otra serie  $Z_t$  que tiene la propiedad de que si se conoce su historia hasta un tiempo  $t_0$ , por ejemplo,  $Z_s, s \leq t_0$ , 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.

Como este resultado es de la forma  $p \Rightarrow q$ , se cumple la contra-recíproca, es decir,  $\neg q \Rightarrow \neg p$ , que se lee: no  $q$  implica no  $p$ . Por lo tanto, si la función de autocorrelación  $\rho(k)$  no tiende a cero la serie no puede ser estacionaria en covarianza. Si se observan las gráficas de  $\rho(k)$  en las figuras (5.1) ésta parece tender a cero, luego se puede aceptar que hay evidencia de que las series son estacionarias.

**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  $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0$  para  $h \neq 0$  entonces la función de autocovarianza cumple:

$$R(k) = \begin{cases} \sigma^2, & \text{si } k = 0 \\ 0, & \text{si } k \geq 1. \end{cases}$$

Adicionalmente

$$\rho(k) = \begin{cases} 1, & \text{si } k = 0 \\ 0, & \text{si } k \geq 1. \end{cases}$$

La función de autocorrelación teórica de un ruido blanco  $R(k)$  es cero excepto en  $k = 0$ .

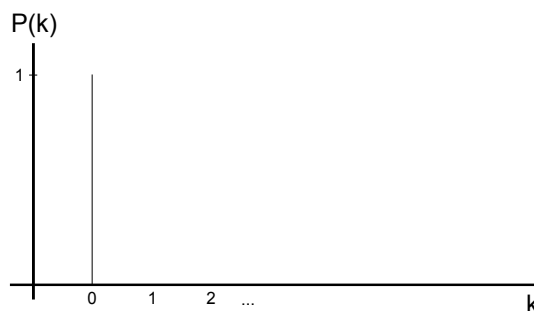


Figura 5.4: Fac Teórica de un Ruido Blanco.

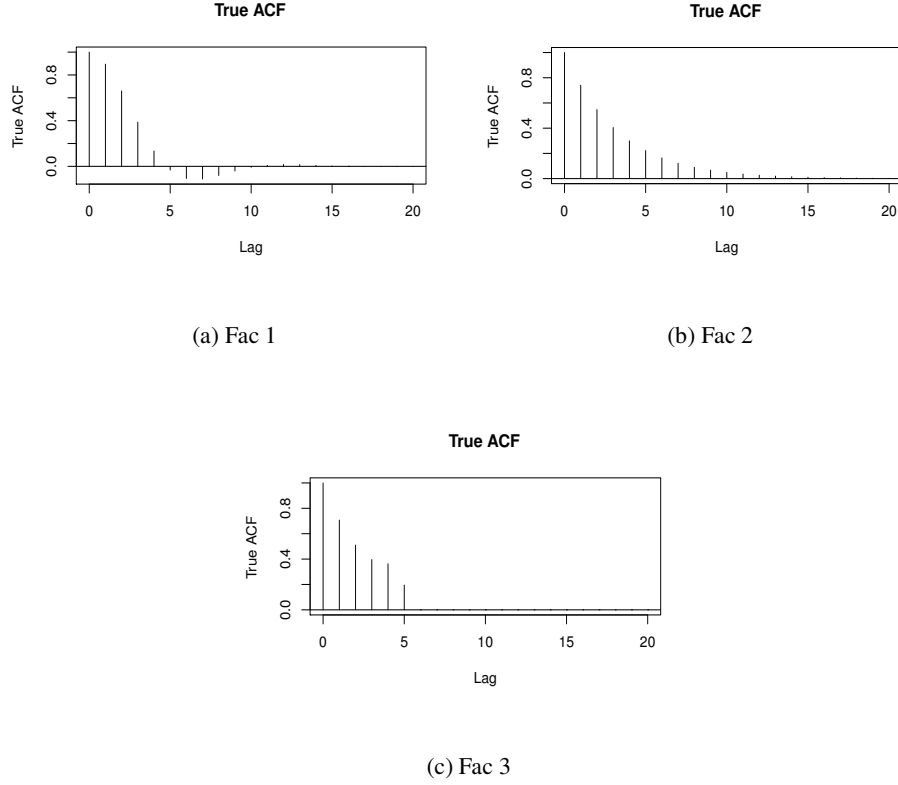


Figura 5.3: Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario

### 5.1.1. Estimación de las funciones de Autocovarianza y Autocorrelación.

A partir de la expresión teórica para  $\rho(k)$

$$\rho(k) = \frac{E((Y_t - \mu)(Y_{t+k} - \mu))}{E(Y_t - \mu)^2} \quad (5.1)$$

se puede obtener el estimador de  $\rho(k)$ ,  $\hat{\rho}(k)$  basado en la trayectoria  $\{Y_1, Y_2, \dots, Y_T\}$

$$\hat{\rho}(k) = \frac{\sum_{j=k+1}^T (Y_j - \bar{Y})(Y_{j-k} - \bar{Y})}{\sum_{j=1}^T (Y_j - \bar{Y})^2}, \quad (5.2)$$

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 de los valores de  $k = 1, 2, \dots, m$ . Una posibilidad es  $m = T/4$ .

También se utiliza

$$\hat{\rho}(k) = \frac{\frac{1}{T-k+1} \sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\frac{1}{T-1} \sum_{t=1}^T (Y_t - \bar{Y})^2}. \quad (5.3)$$

El cálculo de la Fac estimada se realiza en el lenguaje R mediante el la función `acf(x)`

**Proposición 5.1.1 (Bartlett (1946)).** Si  $(\varepsilon_t, t \in \mathbb{Z})$  es ruido blanco entonces las autocorrelaciones muestrales  $\hat{\rho}(k) = \widehat{Cov}(\varepsilon_t, \varepsilon_{t+k})$ ,  $k = 1, 2, \dots, m$ , donde  $m > 1$  es un entero arbitrario, con base en una muestra  $\{\varepsilon_1, \dots, \varepsilon_T\}$ , cumplen

1.  $\hat{\rho}(k)$  son independientes.
2.  $\hat{\rho}(k) \stackrel{a}{\sim} N(0, 1/T)$  para  $T$  grande.

donde  $(\stackrel{a}{\sim})$  significa: se distribuye aproximadamente.

Es decir, si  $T$  es grande y  $\varepsilon_t$  es ruido blanco entonces todas las  $\hat{\rho}(k)$  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  $P(-2\sigma < X < 2\sigma) \approx 0.95$ , por lo tanto, las autocorrelaciones  $\hat{\rho}(k)$  deben estar todas 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  $T/4$ . En SAS el valor por defecto es  $m = 26$ .

**Definición 5.1.3 (Bandas de Bartlett).** El intervalo  $[-2/T, 2/T]$  se denomina “Bandas de Bartlett”.

Al observar la figura (5.5(a)), se puede concluir que hay evidencia de que  $Y_t$  es ruido blanco ( $m = T/4$ ). En la figura (5.5(b)) no se cumple que  $\forall k, \hat{\rho}(k) \in [-2/T, 2/T]$  por tanto no se acepta que  $Y_t$  sea ruido blanco.

### Aplicación al Caso de los Residuos Estructurales

Se puede realizar un diagnóstico gráfico con base en  $\hat{\rho}(k)$  y las bandas de Bartlett. Si se observa  $\forall k \leq m, \hat{\rho}(k) \in [-2/T, 2/T]$  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  $\hat{\rho}(k) \notin [-2/T, 2/T]$

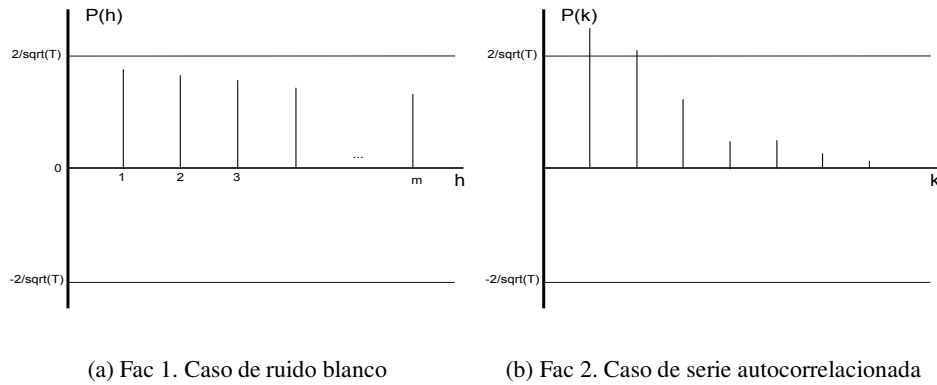


Figura 5.5: Ejemplos de Fac y Bandas de Bartlett

entonces se tiene evidencia de que el residuo estructural no puede ser ruido blanco y tiene una dinámica autorregresiva que se debe modelar, por ejemplo mediante un proceso ARMA, para aprovechar tal dinámica con el fin de mejorar los pronósticos estructurales.

**Ejemplo 5.1.1.** *Retomando el Ejemplo (4.2.2) en la pag. 53, 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*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t. \quad (5.4)$$

*Es decir, un modelo lineal para la tendencia, más una componente estacional, de período  $s = 4$ . Al estimar los residuos estructurales  $\hat{\varepsilon}_t$  se obtienen las gráficas siguientes.*

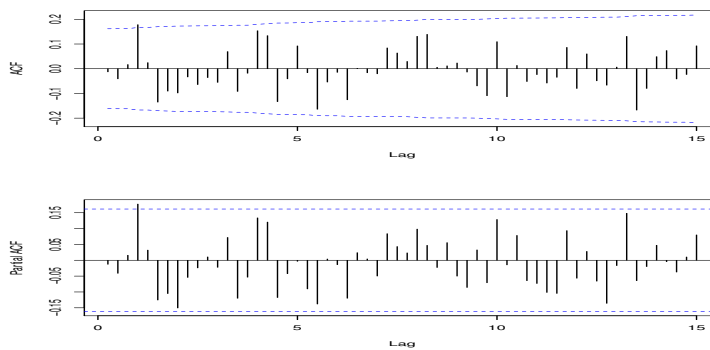


Figura 5.6: Fac de los Residuos Estructurales Serie Cementos

*Es evidente que los residuos no son ruido blanco. En la gráfica de la fac muestral, en el panel inferior de la Figura 5.6, se puede apreciar cómo las primeras cinco autocorrelaciones se*

salen de las Bandas de Bartlett. Además, en la gráfica del panel superior, la gráfica de  $\widehat{\varepsilon}_t$  muestra períodos de incremento y luego de decrecimiento, que puede tomarse como evidencia de autocorrelación. Este ejemplo se continúa más adelante en este Capítulo.

## 5.2. 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.

### 5.2.1. Prueba Ljung-Box (LB)

A partir del Teorema de Bartlett, (5.1.1), si  $\varepsilon_t \sim RB(0, \sigma^2)$ ,  $t = 1, \dots, T$  y  $\hat{\rho}(k) = \widehat{Cov}(\varepsilon_t, \varepsilon_{t+k})$ ,  $k = 1, 2, \dots, m$ , entonces, para  $T$  grande, se tiene

1.  $\hat{\rho}(k) \stackrel{a}{\sim} N(0, 1/T)$ .
2.  $\hat{\rho}(k)$  son idenpendientes.

Luego  $\sqrt{T}\hat{\rho}(k) \stackrel{a}{\sim} N(0, 1)$ , y  $(\sqrt{T}\hat{\rho}(k))^2 \stackrel{a}{\sim} \chi_1^2$ , por lo tanto

$$Q = T \sum_{k=1}^m \hat{\rho}^2(k) \stackrel{a}{\sim} \chi_m^2. \quad (5.5)$$

#### 1. Hipótesis de la Prueba

$$H_0 : \varepsilon_t \sim RB(0, \sigma^2)$$

$$H_1 : \varepsilon_t \text{ no son ruido blanco}$$

otra forma equivalente de escribir la prueba es

$$H_0 : \forall k, \hat{\rho}(k) = 0$$

$$H_1 : \exists k, \hat{\rho}(k) \neq 0$$

## 2. Estadístico de Prueba

Ljung y Box modificaron el estadístico  $Q$  en (5.5) para mejorar la aproximación de  $Q$  a la  $\chi_m^2$  en muestras pequeñas, usando

$$Q_{LB} = T(T+2) \sum_{k=1}^m \frac{1}{T-k} \hat{\rho}^2(k) \stackrel{a}{\sim} \chi_m^2. \quad (5.6)$$

Entonces, si  $H_0 : \varepsilon_t \sim RB(0, \sigma^2)$  es cierta se cumple  $\hat{\rho}(k) \stackrel{a}{\sim} N(0, 1/T)$ ,  $k = 1, \dots, m$  y por tanto  $Q_{LB} \stackrel{H_0}{\sim} \chi_m^2$  donde  $(\stackrel{H_0}{\sim})$  significa: distribuye asumiendo  $H_0$  cierta.

## 3. Decisión

a) Llame  $Q_{\text{obs}}$  el estadístico observado y *Valor  $p$*   $= P(\chi_m^2 \geq Q_{\text{obs}} | H_0 \text{ cierto})$

- 1) Si *Valor  $p$*   $< 0.05$ , se rechaza  $H_0$ .
- 2) Si *Valor  $p$*   $> 0.05$ , no se rechaza  $H_0$ .

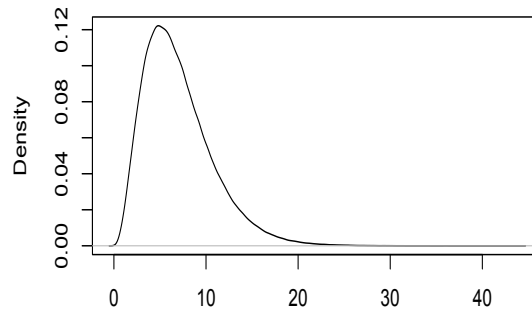


Figura 5.7: Densidad  $\chi_m^2$ .

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  $\chi_{m,\alpha}^2$  es valor crítico de  $\chi_m^2$  correspondiente a la probabilidad  $\alpha$ ,  $\alpha$  nivel de significancia, entonces si  $Q_{\text{obs}} < \chi_{m,\alpha}^2$ , no se rechaza  $H_0$  y si  $Q_{\text{obs}} > \chi_{m,\alpha}^2$  se rechaza  $H_0$ .

**Ejemplo 5.2.1.**  $m = 6$ ,  $\alpha = 0.05$ ,  $\chi_{6,0.05}^2 = 12.54$  Si el estadístico observado fuera  $Q_{\text{obs}} = 66.93$  entonces  $Q_{\text{obs}} > \chi_{6,0.05}^2$ , por lo tanto, se rechaza  $H_0$

#### 4. Prueba LB en R

Se usa la función `Box.test` de `stats`.

```
# Ejemplo de aplicacion de la prueba Ljung-Box
# Nótese que la funcion recibe un vector r
# que no necesariamente tiene que ser un vector de residuos
modelo = lm(y ~ t + It)
r = modelo$residuals
Box.test(r, lag = 25, type = "Ljung-Box")
Box.test(r, lag = 35, type = "Ljung-Box")
```

#### 5.2.2. Prueba Durbin-Watson (DW)

La prueba DW se basa en la definición del modelo  $AR(1)$ .

**Definición 5.2.1 (Proceso  $AR(1)$ ).** Una serie  $(Y_t, t \in \mathbb{Z})$  se dice  $AR(1)$  con media  $\mu$  si cumple que

$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (5.7)$$

donde  $\phi$  es un parámetro con  $|\phi| < 1$  y  $\varepsilon_t \sim RB(0, \sigma^2)$ .

**Propiedades:** Si  $Y_t \sim AR(1)$

1. Si  $\mu = 0$  entonces  $Y_t = \phi Y_{t-1} + \varepsilon_t$ .  
Si  $Y_t$  tiene media  $\mu$ ,  $Y_t^* = Y_t - \mu$  tiene media cero.
2.
  - $E(Y_t) = \mu$
  - $Var(Y_t) = \frac{\sigma^2}{1 - \phi^2}$
  - $Cov(Y_t, Y_{t+k}) = R(k) = \frac{\sigma^2 \phi^k}{1 - \phi^2}, k = 0, 1, \dots$
  - $\rho(k) = \phi^k, k = 0, 1, 2, \dots$

luego la FAC muestral de un  $AR(1)$  es de la forma

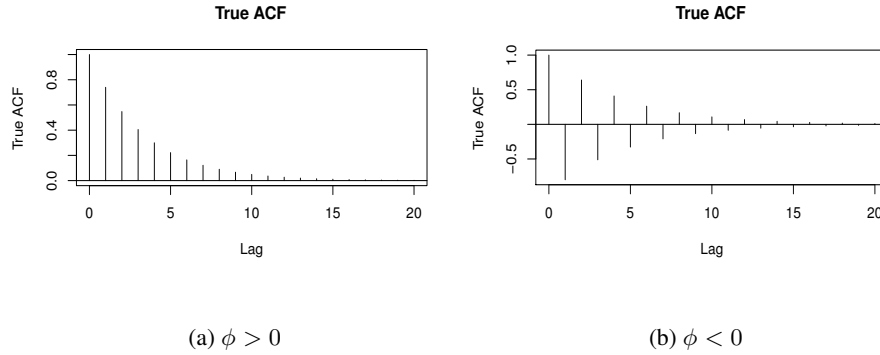


Figura 5.8: Fac Muestral de un  $AR(1)$ .

3. Además  $Y_t$  es estacionario en covarianza.

Ahora suponga una serie  $(Y_t, t = 1, \dots, T)$  y se quiere probar que es incorrelacionada. Asumimos que  $\bar{Y}_t = 0$ , 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 la ecuación (5.7).

### 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

El estadístico DW se define como

**Definición 5.2.2.**

$$d = \frac{\sum_{t=2}^T (Y_t - Y_{t-1})^2}{\sum_{t=1}^T Y_t^2}. \quad (5.8)$$



Entonces como  $\sum_{t=2}^T (Y_t - Y_{t-1})^2 = 2 \sum_{t=2}^T Y_t^2 - 2 \sum_{t=2}^T Y_t Y_{t-1}$ , se puede colocar

$$d \approx 2 - 2 \frac{\sum_{t=2}^T Y_t Y_{t-1}}{\sum_{t=1}^T Y_t^2} \approx 2 - 2\hat{\rho}(1),$$

es decir,

$$d \approx 2(1 - \hat{\rho}(1)). \quad (5.9)$$

A partir de (5.9) se puede concluir que, como  $-1 < \hat{\rho}(1) < 1$  entonces  $0 < d < 4$ . Y si  $\hat{\rho}(1) \approx 0$  entonces  $d \approx 2$ . Además,

Si

- $d < 2 \Leftrightarrow 1 - \hat{\rho}(1) < 1 \Leftrightarrow \hat{\rho}(1) > 0$
- $d < 2 \Leftrightarrow \hat{\rho}(1) < 0$

luego un  $d > 2$  indica posible autocorrelación positiva en la serie y  $d < 2$  posible autocorrelación negativa.

### 3. Decisión

La decisión se toma con base en el *Valor p*.

### 4. Prueba DW en R

a) **Librería lmtest:** Tiene la función `dwtest`. Permite probar hipótesis alternas

$H_1 : \phi > 0, H_1 : \phi \neq 0, H_1 : \phi < 0$ .

```
dwtest(formula, alternative = c("greater", "two.sided",
                                "less"))
```

`dwtest(y ~ t + It)` calcula  $d$  y el *Valor p*.

b) **Librería car:** Tiene la función `durbin.watson`, que es diferente de `dwtest`.

```
durbinWatsonTest(y ~ t + It)
```

### 5. Notas sobre la Prueba DW

Suponga que el modelo

$$Y_t = \sum_{j=1}^k \beta_j X_{j,t} + \varepsilon_t, \quad (5.10)$$

contiene un término autorregresivo  $Y_{t-1}$ ,

$$Y_t = \phi_1 Y_{t-1} + \sum_{j=1}^k \beta_j X_{j,t} + \varepsilon_t, \quad (5.11)$$

es decir, se incluye un término autorregresivo de la serie.

- a) Entonces la prueba DW se afecta por el término  $\phi_1 Y_{t-1}$ , y el estadístico se sesga hacia 2, es decir, la prueba tiende a aceptar  $H_0$ .

Se dice que la prueba tiene baja potencia para detectar la hipótesis alternativa  $H_1$  cuando es cierta. En este caso se modifica el estadístico DW a

$$h = \left( 1 - \frac{d}{2} \sqrt{\frac{T}{1 - TS_{\phi_1}^2}} \right) \quad (5.12)$$

pero la prueba `durbin.watson` la calcula automáticamente cuando se incluye la variable rezagada  $Y_{t-1}$ .

- b) Puede que en el modelo  $Y_t = \sum_{j=1}^k \beta_j X_{j,t} + \varepsilon_t$  se detecte autocorrelación en  $\hat{\varepsilon}_t$ . Pero al considerar  $Y_t = \phi_1 Y_{t-1} + \sum_{j=1}^k \beta_j X_{j,t} + \varepsilon_t$  la prueba modificada no detecte autocorrelación y en cambio  $\hat{\phi}_1$  sea significativo.

**Ejemplo 5.2.2.** Retomando el Ejemplo (5.1.1) en la pag. 80, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t. \quad (5.13)$$

Al estimar los residuos estructurales  $\hat{\varepsilon}_t$  se observó que posiblemente muestran una dinámica autocorrelacionada. Se puede confirmar este hecho con las pruebas Ljung-Box y Durbin-Watson ( en sus dos versiones).

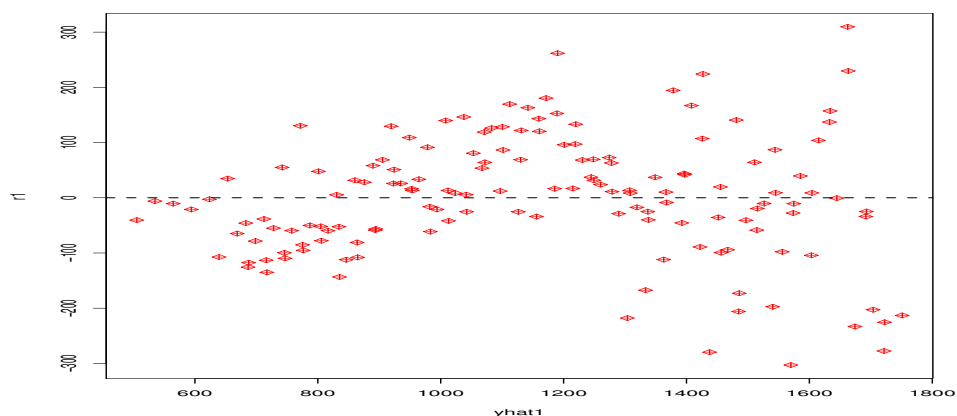


Figura 5.9: Residuales estructurales versus Y estimada

Box-Ljung test

X-squared = 293.9028, df = 25, p-value < 2.2e-16

Durbin-Watson test

DW = 0.3646, p-value < 2.2e-16

alternative hypothesis: true autocorrelation is greater than 0

lag	Autocorrelation	D-W	Statistic	p-value
1	0.8043803	0.3645806	0	

Alternative hypothesis:  $\rho \neq 0$

*Es claro que los residuos estructurales presentan autocorrelación. Es decir, es supuesto de incorrelación no es válido.*

*También se pueden correr las pruebas de heterocedasticidad, Breusch-Pagan, Prueba de Scores para Varianza no constante y Goldfeld-Quandt.*

studentized Breusch-Pagan test

BP = 19.9866, df = 4, p-value = 0.0005025

Non-constant Variance Score Test

Variance formula:  $\sim$  fitted.values

Chisquare = 23.03938 Df = 1 p = 1.587171e-06

Goldfeld-Quandt test

$GQ = 6.8213$ ,  $df1 = 69$ ,  $df2 = 68$ ,  $p\text{-value} = 7.005e-14$

*Es evidente que los residuos estructurales tiene heterocedasticidad. Se confirma con la Figura 5.9.*

### 5.3. Alternativas cuando los Residuos Estructurales muestran Autocorrelación

En el caso de detectarse autocorrelación en los residuos estructurales el procedimiento de consistirá en buscar un modelo tipo ARMA, autorregresivo de media móvil, que se tratarán en el Capítulo 6. Pero es posible utilizar otras alternativas.

Una es la transformación Cochran-Orcutt, que permite re-estimar los parámetros incluyendo una autocorrelación de tipo AR(1) en los residuos estructurales.

Otra alternativa consiste en incluir en el modelo una componente autorregresiva, 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

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \quad (5.14)$$

Que se transforma, por ejemplo, en el modelo con dos rezagos de la variable dependiente.

$$Y_t = \beta_0 + \beta_1 t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \quad (5.15)$$

Con esta transformación se trata de “capturar” la autocorrelación en los residuos estructurales estimados  $\hat{\varepsilon}_t$ . 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 (5.15) se coloca

```
require(dynlm)
mod2 = dynlm(yi ~ t + It + L(yi,1) + L(yi,2))
summary(mod2)
```

Este modelo logra capturar la estructura tanto como para que la prueba Durbin-Watson no rechace la hipótesis nula. Además, las pruebas de heterocedasticidad no rechazan la hipótesis nula de homocedasticidad. El programa siguiente muestra los pasos en la estimación de los modelos (5.14) y (5.15).

```

#--- programa en R para estimacion con rezagos de Yt
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))

yi = window(y,start=c(1956,1),end=c(1992,3))
yf = window(y,start=c(1992,4),end=c(1994,3))

#----modelo con tendencia lineal y estacionalidad
#----con variables indicadoras estacionales

t = seq(1,length(yi))
It = seasonaldummy(yi)
mod1 = lm(yi ~ t + It)
summary(mod1)
print(xtable(mod1),digits=6)

yhat1 = mod1$fitted.values
r1 = mod1$residuals

# analisis de los residuos estructurales
par(mfrow = c(1,1))
plot(t,r1,type='o',ylab='residuo')
abline(h=0,lty=2)
acf(r1,26,main=" ")

# grafica residuos versus yhat
plot(yhat1,r1,type='p',pch=9, col='red')
abline(h=0,lty=2,lwd=2)

# pruebas de heterocedasticidad
bptest(mod1,studentize = TRUE)

```

```

ncvTest(mod1)
gqtest(mod1)

# prueba Ljung-Box
Box.test(r1, lag = 25, type = "Ljung-Box")

# prueba Durbin-Watson
dwtest(mod1)
durbinWatsonTest(mod1)

# modelo incorporando terminos rezagados de la serie
# utilizando la libreria dynlm

require(dynlm)

mod2 = dynlm(yi ~ t + It + L(yi,1)+ L(yi,2))
summary(mod2)
yhat2 = mod2$fitted.values
r2 = mod2$residuals

# pruebas de heterocedasticidad

bptest(mod2, studentize = TRUE)
ncvTest(mod2)
gqtest(mod2)

# grafica residuos versus yhat
plot(as.numeric(yhat2), as.numeric(r2), type='p')
abline(h=0, lty=2, lwd=2)

layout(1:2)
plot(t[-c(1,2)], r2, type='o', ylab='residuo')
abline(h=0, lty=2)
acf(r2, 26, main=" ")

# prueba Durbin-Watson
dwtest(mod2, alternative="two.sided")
durbinWatsonTest(mod2)

```

```

# prueba Ljung-Box
Box.test(r2, lag = 25, type = "Ljung-Box")

# pronosticos con el modelo de regresion
T = length(yi)
Itp = seasonaldummyf(yi,8)
tp = seq(T+1,T+8,1)
prons.vi = predict(mod1,
data.frame(t = tp, It=I(Itp)))

# pronosticos con el modelo con rezagos
n = length(y)
lyf1=y[(n-8):(n-1)]
lyf2=y[(n-9):(n-2)]
Xt = cbind(rep(1,8),tp,Itp,lyf1,lyf2)
beta = dfm$coefficient
prons.din = Xt%*%beta

# grafica de pronosticos
plot(tp,yf, type = 'o',
ylim = c(1400,2000))
lines(tp,prons.vi, type = 'b', pch = 5,col='red' )
lines(tp,prons.din, type = 'b', pch = 3,col='blue' )

(cbind(accuracy(yf,prons.vi),
accuracy(yf,prons.din)))

```

	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





---

### Modelos ARMA para la Componente Aleatoria

---

#### 6.1. Introducción

En los modelos de descomposición  $Y_t = T_t + S_t + \varepsilon_t$ ,  $t = 1, 2, \dots$  se estima  $\hat{\varepsilon}_t$  y se determina si es o no ruido blanco mediante las pruebas Ljung-Box y Durbin-Watson. En caso de encontrar que  $\hat{\varepsilon}_t$  no es ruido blanco, el siguiente paso es modelar esta componente mediante tres posibles modelos

1. Medias Móviles de orden  $q$ ,  $MA(q)$ .
2. Autoregresivos de orden  $q$ ,  $AR(p)$ .
3. Medias Móviles Autoregresivos,  $ARMA(p, q)$ .

*“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 cantidades relacionadas con cada modelo, éstas no tienen nada que ver con datos ni estimación pero son fundamentales para desarrollar una comprensión básica de las propiedades de los modelos necesarios para llevar a cabo pronósticos inteligentes.” Diebold [1999, pág. 129]*

## 6.2. Procesos de Medias Móviles de orden $q$ , $MA(q)$

**Definición 6.2.1.** El Operador de Rezago se denota por  $L$  (lag, en inglés) y es tal que  $L(Y_t) = Y_{t-1}$ . 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  $\beta_j, j = 0, 1, \dots, p$

$$B_p(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p, \quad (6.1)$$

tal que

$$\begin{aligned} B_p(L)(Y_t) &= (\beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p)Y_t, \\ &= \sum_{j=0}^p \beta_j L^j Y_t, \\ &= \sum_{j=0}^p \beta_j Y_{t-j}, \\ &= \beta_0 Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p}. \end{aligned}$$

**Definición 6.2.2 (Proceso  $MA(q)$ ).** Se dice que una serie  $Y_t$  sigue un proceso  $MA(q)$ ,  $q = 1, 2, \dots$  de media móvil de orden  $q$ , si se cumple que

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}, \quad (6.2)$$

para constantes  $\theta_1, \dots, \theta_q$  y  $\varepsilon_t \sim RB(0, \sigma^2)$ . La expresión con el operador  $L$  es, si se define el polinomio

$$\theta_q(L) = 1 + \theta_1 L + \dots + \theta_q L^q, \quad (6.3)$$

entonces la ecuación (6.2) queda  $Y_t = \theta_q(L)(\varepsilon_t)$ .

### 6.2.1. Propiedades

1.  $E(Y_t) = 0$
  2.  $Var(Y_t) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2$
- luego  $Var(Y_t) > Var(\varepsilon_t)$ , en general.

3.  $Cov(Y_t, Y_{t+k}) = R(k)$ , donde

$$R(K) = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & k < q+1 \\ 0, & k \geq q+1 \end{cases} \quad (6.4)$$

con  $\theta_0 = 1$ .

4. Un  $MA(q)$  siempre es un proceso estacionario con  $\rho(k) = \frac{R(k)}{R(0)}$ .

Interpretación de (6.4). La ecuación (6.4) 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 valor. Por esta razón se puede ver los procesos  $MA(q)$  como alternativas al Ruido Blanco completamente incorrelacionado.

**Ejemplo 6.2.1.** Sea  $Y_t \sim MA(2)$  dado por

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 9), \quad t \in \mathbb{Z},$$

con

$$\theta_1 = -0.4, \theta_2 = 0.4, \sigma_2 = 9,$$

entonces

$$\begin{aligned} R(0) &= (1 + 0.4^2 + 0.4^2)9 = 11.88 \\ R(1) &= 9 \sum_{j=0}^{2-1} \theta_j \theta_{j+1} = 9(\theta_0 \theta_1 + \theta_1 \theta_2) \\ &= 9(-0.4 + (-0.4)(0.4)) = -5.04 \\ R(2) &= 9 \sum_{j=0}^{2-2} \theta_j \theta_{j+2} = 9(\theta_0 \theta_2) = 9(0.4) = 3.6. \end{aligned}$$

Entonces la FAC es

$$\begin{aligned} \rho(0) &= 1, \quad \rho(1) = -\frac{5.04}{11.88} = -0.42, \quad \rho(2) = \frac{3.6}{11.88} \\ \rho(3) &= \rho(4) = \dots = 0 \end{aligned}$$

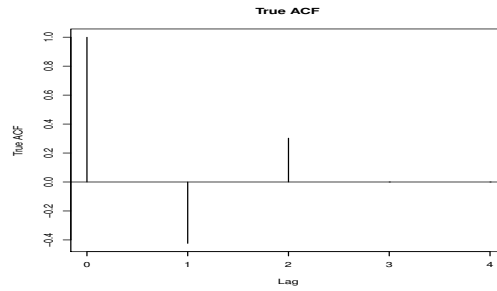
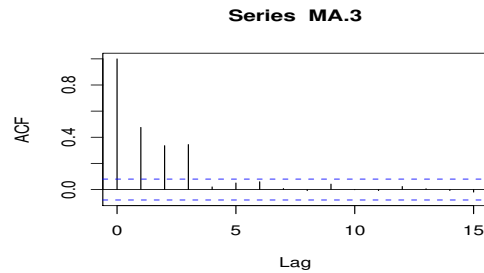


Figura 6.1: Función de Autocorrelación.

**Ejercicio 6.2.1.** Encuentre la FAC de

1.  $Y_t = \varepsilon_t - 0.5\varepsilon_{t-1} - 0.5\varepsilon_{t-2}$ .
2.  $Y_t = \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2} - 0.1\varepsilon_{t-3}$ .

**Conclusión** De acuerdo con (6.4), si la fac muestral de una serie  $Y_t$  termina abruptamente puede tratarse de un  $MA(q)$ . Por ejemplo, en la siguiente gráfica 6.2 sería factible un modelo  $MA(3)$ .

Figura 6.2: FAC muestral de un  $MA(3)$ .

**Definición 6.2.3 (Función de Autocorrelación Parcial (facp)).** Suponga que  $(Y_t, t \in \mathbb{Z})$  es estacionaria. La facp es una función de  $k$ ,  $\alpha(k)$ ,  $k = 1, 2, \dots$  definida por

1.  $\alpha(1) = \rho(1)$
2.  $\alpha(k) = \text{Corr}(\varepsilon_1, \varepsilon_k)$  donde

$$\begin{aligned}\varepsilon_1 &= Y_1 - E(Y_1|Y_2, \dots, Y_{k-1}) \\ \varepsilon_k &= Y_k - E(Y_k|Y_2, \dots, Y_{k-1}), \quad k = 2, \dots\end{aligned}$$

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  $Y_{t-1}, \dots, Y_{t-k}$  tal que  $Y_t = \phi_{k1}Y_{t-1} + \dots + \phi_{kk}Y_{t-k} + \varepsilon_t$  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  $MA(q)$

### 6.2.2. Condición de Invertibilidad del Proceso MA(q)

**Definición 6.2.4.** Dado un proceso  $MA(q)$ ,  $Y_t = \theta_q(L)(\varepsilon_t)$  donde  $\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$ , entonces considerando el polinomio en  $z \in \mathbb{C}$ ,  $\theta_q(z) = 1 + \theta_1 z + \dots + \theta_q z^q$  y sus  $q$  raíces  $(z_1, z_2, \dots, z_q) \in \mathbb{C}$ , es decir, valores  $z \in \mathbb{C}$  tales que  $\theta_q(z) = 0$ , se dice que el proceso  $Y_t$  es invertible si se cumple

$$|z_j| > 1, \quad \forall j = 1, \dots, q, \quad (6.5)$$

o también, si  $\theta_q(z) \neq 0$ ,  $\forall z$ ,  $|z| \leq 1$ . Note que (6.5) es equivalente a

$$\frac{1}{|z_j|} < 1, \quad \forall j = 1, \dots, q$$

es decir, los inversos de las raíces deben caer dentro del círculo unitario complejo.

**Ejemplo 6.2.2.** Sea  $Y_t \sim MA(a)$  tal que

$$Y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.4\varepsilon_{t-2}, \quad (6.6)$$

veamos si  $Y_t$  es invertible. Hallamos las raíces del polinomio  $\theta_q(z)$

$$\begin{aligned} \theta_2(z) &= 1 - 0.4z + 0.4z^2 = 0, \\ z &= \frac{0.4 \pm \sqrt{0.4^2 - 4(0.4)(1)}}{2(0.4)} \\ &= \frac{1}{2} \pm \frac{1}{2} \frac{10}{4} \sqrt{\frac{4}{10}} \sqrt{\frac{4}{10} - 4} = \frac{1}{2} \pm \frac{1}{2} \frac{\sqrt{10}}{2} \sqrt{-\frac{36}{10}} \\ &= \frac{1}{2} \pm \frac{1}{2} \frac{\sqrt{10}}{2} \sqrt{\frac{36}{10}} i = \frac{1}{2} \pm \frac{3}{2} i \end{aligned}$$

por tanto

$$|z| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\pm\frac{3}{2}\right)^2} = \sqrt{\frac{1}{4} + 9} > 1,$$

luego  $Y_t$  es invertible.

### 6.2.3. Función fac parcial de un Proceso MA(q) invertible

Suponga un proceso  $Y_t \sim MA(q)$  invertible,

$$Y_t = \theta_q(L)(\varepsilon_t). \quad (6.7)$$

Considere  $\theta_q(z) = 1 + \theta_1 z + \dots + \theta_q z^q$  entonces  $\theta_q(z) \neq 0$ ,  $|z| \leq 1$ , luego la función  $\frac{1}{\theta_q(z)}$  tiene desarrollo en serie de Taylor alrededor de  $z = 0$ , dado por

$$\frac{1}{\theta_q(z)} = 1 + \psi_1 z + \psi_2 z^2 + \dots = \sum_{j=0}^{\infty} \psi_j z^j, \quad \psi_0 = 1, \quad (6.8)$$

con  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ , donde  $\psi_j \rightarrow 0$  si  $j \rightarrow \infty$ . Multiplicando ambos miembros de (6.7) por  $\frac{1}{\theta_q(L)}$  se obtiene

$$\varepsilon_t = \frac{1}{\theta_q(L)} Y_t = \psi(L)(Y_t) = Y_t + \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \dots \quad (6.9)$$

Y despejando  $Y_t$

$$Y_t = -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \dots + \varepsilon_t, \quad (6.10)$$

de donde concluimos que si hace la regresión de  $Y_t$  sobre los primeros  $k$  rezagos  $Y_{t-j}$ ,  $j = 1, \dots, k$ , entonces el  $k$ -ésimo coeficiente es  $\alpha(k) = \psi(k) \neq 0$ ,  $\forall k$  y como  $\psi(k) \rightarrow 0$  entonces  $\alpha(k) \rightarrow 0$  cuando  $k \rightarrow \infty$ . Por tanto, la facp de un MA(q) decrece a cero. En las Figuras siguientes 6.3 se observa la fac y la facp de un MA(3).

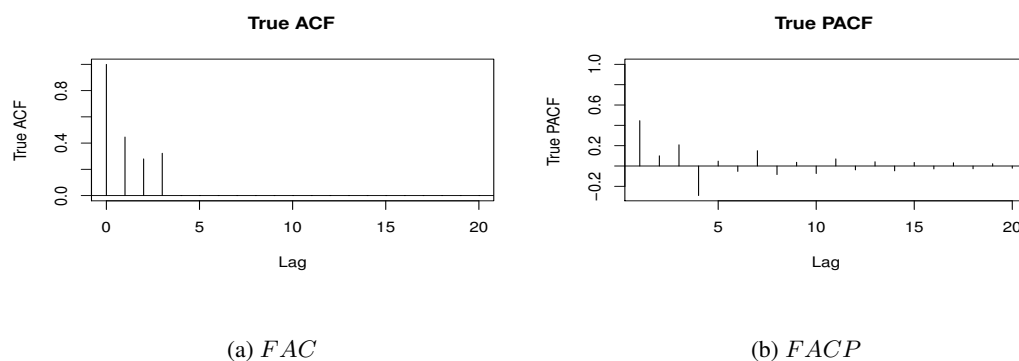


Figura 6.3:  $FAC$  y  $FACP$  de un  $MA(3)$ .

#### 6.2.4. Implementación en R

En R para identificar se usan las funciones `acf` y `pacf` y para estimar una de las funciones usadas es `arma` de la librería `tseries`.

**Ejemplo 6.2.3.** `library(forecast,tseries)`  
`n = 300`  
`theta = c(-1,-0.4,-0.4)`  
`(Mod(polyroot(theta)))`  
`y = arima.sim(list(order=c(0,0,2), ma=theta[2:3]), n=n, sd=sqrt(2.3))`  
  
`layout(1:3)`  
`ts.plot(y)`  
`acf(y, 30)`  
`pacf(y, 30)`  
  
`# Estimación: Función arma librería tseries`  
  
`modelo.y = arma(x=y, order=c(0,2))`  
`summary(modelo.y)`

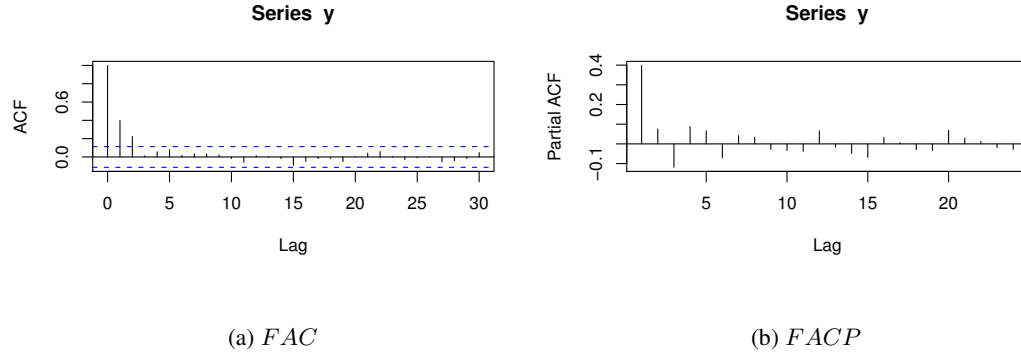


Figura 6.4: *FAC* y *FACP* del Ejemplo.

```
pred.y = predict(modelo.y, n.ahead=2)
plot(seq(1, 9, 1), c(tail(y), pred.y$pred), type='b')
points(seq(7, 9, 1), pred.y$pred, type='b', col='red')
```

### 6.3. Procesos Autorregresivos de Orden $p$ , $AR(p)$

**Definición 6.3.1 (Proceso  $AR(p)$ ).** Se dice que  $Y_n, n \in \mathbb{Z}$  sigue un proceso  $AR(p)$  de media cero si

$$Y_n = \varphi_1 Y_{n-1} + \varphi_2 Y_{n-2} + \cdots + \varphi_p Y_{n-p} + \varepsilon_n, \quad (6.11)$$

donde  $\varepsilon_n \sim RB(0, \sigma^2)$  y  $p = 1, 2, \dots$ . Usando el operador de rezago  $L$  se puede escribir (6.11) como

$$\varphi_p(L)(Y_n) = \varepsilon_n, \quad (6.12)$$

con  $\varphi_p(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p$ ,  $z \in \mathbb{C}$ , el polinomio autorregresivo.

#### 6.3.1. Condición Suficiente para que un $AR(p)$ sea Estacionario en Covarianza

La condición suficiente para que  $Y_t \sim AR(p)$  sea estacionario en covarianza es que las  $p$  raíces de la ecuación  $\varphi_p(z) = 0$ ,  $z_i$ ,  $i = 1, 2, \dots, p$  cumplan

$$|z_i| > 1. \quad (6.13)$$

Nótese que si  $z_j = a_j \pm ib_j$  entonces  $|z_j| = \sqrt{a_j^2 + b_j^2}$ . La condición (6.13) no es, sin embargo, necesaria. En palabras, la condición (6.13) se describe como “para que un proceso



autorregresivo de orden  $p$  sea estacionario en covarianza, es suficiente que las raíces del polinomio autorregresivo estén por fuera del círculo unitario”. El círculo unitario aparece en la Figura 6.5. En esta figura se observa la posición de la raíz  $z_j$  y su conjugado  $\bar{z}_j$ .

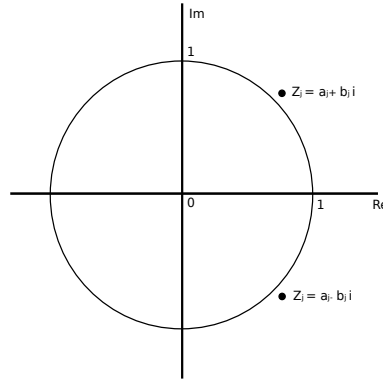


Figura 6.5: Círculo Unitario

Si el proceso  $Y_t$  es estacionario en covarianza se cumple que su media es constante,  $Y_t \equiv \mu$ , y tiene función de autocovarianza,  $Cov(Y_t, Y_{t+k}) = R(k)$ ,  $k \in \mathbb{Z}$ . De ahora en adelante siempre se asumirá que los procesos AR(p) cumplen la condición (6.13).

### 6.3.2. Algunas Propiedades de los Procesos AR(p) Estacionarios

**Proposición 6.3.1.** Para un proceso  $Y_t \sim AR(p)$ , (6.11) se cumple: i)  $E(Y_t) = 0$  y ii)  $\sum_{j=1}^p \varphi_j < 1$ .

*Demostración.* i). Si  $Y_t$  es estacionario en covarianza entonces  $\mathbb{E}(Y_t) = \mu$ . Además,

$$E(Y_t) = \varphi_1 E(Y_{t-1}) + \varphi_2 E(Y_{t-2}) + \cdots + \varphi_p E(Y_{t-p}) + 0,$$

pero todas las esperanzas son  $\mu$  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 (6.11) es de media cero. Para ii), si se asume que  $\sum_{j=1}^p \varphi_j \geq 1$ , el caso  $= 1$  implica nuevamente que  $\varphi_p(1) = 0$ , lo cual es una contradicción. 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.  $\square$

Un proceso  $Y_t \sim AR(p)$  con  $\mathbb{E}(Y_t) = \mu \neq 0$  se define mediante la relación siguiente.

$$\varphi_p(L)(Y_t) = \varphi_0 + \varepsilon_t, \quad (6.14)$$

donde

$$\begin{aligned} \varphi_0 &= \varphi_p(L)(\mu) \\ &= (1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p)\mu. \end{aligned}$$

Nótese que también se puede escribir  $Y_t = (1 - \varphi_1 - \cdots - \varphi_p)\mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$ , de donde  $Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \cdots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t$ . Es decir, el proceso  $(Y_t - \mu)$  es  $AR(p)$  de media cero.

### La Función de Autocovarianza de los Procesos AR(p)

La función de autocovarianza de un proceso  $Y_t \sim AR(p)$ ,  $R(k) = \text{Cov}(Y_t, Y_{t+k})$ , se puede calcular resolviendo una ecuación recursiva lineal denominada la ecuación de Yule–Walker.

**Proposición 6.3.2.** *Suponga un proceso  $AR(p)$ ,  $Y_n = \sum_{j=1}^p \varphi_j Y_{n-j} + \varepsilon_n$ , que satisface la condición ((6.13)). Su función de autocovarianza  $R(k)$  satisface la ecuación recursiva*

$$R(k) = \sum_{j=1}^p \varphi_j R(k-j), \quad k = 1, 2, \dots \quad (6.15)$$

denominada ecuación de Yule–Walker.

*Demostración.* Colocando  $\mu = \mathbb{E}(Y_n)$ , como  $Y_n = \sum_{j=1}^p \varphi_j Y_{n-j} + \varepsilon_n$ , al tomar esperanza en ambos miembros de esta identidad se obtiene  $\mu = \sum_{j=1}^p \varphi_j \mu + 0$ . Restando las expresiones anteriores se obtiene  $Y_n - \mu = \sum_{j=1}^p \varphi_j (Y_{n-j} - \mu) + \varepsilon_n$ . Multiplicando ambos miembros de esta identidad por  $Y_{n-k} - \mu$ , con  $k \leq n$ , y tomando valor esperado  $\mathbb{E}(\cdot)$  se obtiene

$$\begin{aligned} R(k) &= \mathbb{E}((Y_n - \mu)(Y_{n-k} - \mu)) \\ &= \sum_{j=1}^p \varphi_j \mathbb{E}((Y_{n-j} - \mu)(Y_{n-k} - \mu)) + \mathbb{E}(\varepsilon_n(Y_{n-k} - \mu)) \\ &= \sum_{j=1}^p \varphi_j R(k-j). \end{aligned}$$

En el resultado anterior se cumple  $\mathbb{E}(\varepsilon_n(Y_{n-k} - \mu)) = 0$  porque, a partir de la definición del proceso  $Y_n$  en (6.11),  $Y_{n-k}$  depende de  $\varepsilon_s$  con  $s \leq n-k$ , que son variables incorrelacionadas con  $\varepsilon_n$ .  $\square$

### La Varianza de los Procesos AR(p)

Si  $Y_t \sim AR(p)$  de media cero,  $\varphi_p(L)(Y_n) = \varepsilon_n$ , para  $\varepsilon_n \sim RB(0, \sigma^2)$ . Además, se asumió que se cumple que  $\forall z, |z| \leq 1, \varphi_p(z) \neq 0$ . Entonces el cociente  $\frac{1}{\varphi_p(z)}$  se puede desarrollar en serie de potencias de  $z$ , y colocar

$$\frac{1}{\varphi_p(z)} = \sum_{j=0}^{\infty} \psi_j z^j,$$

para ciertos coeficientes  $(\psi_j, j = 0, 1, \dots)$ , con  $\psi_0 = 1$ . Por tanto, se puede colocar

$$Y_n = \frac{1}{\varphi_p(L)}(\varepsilon_n) = \varepsilon_n + \psi_1 \varepsilon_{n-1} + \psi_2 \varepsilon_{n-2} + \dots \quad (6.16)$$

Tomando varianza a ambos miembros de (6.16), se obtiene  $Var(Y_n) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$ . La identidad (6.16) se puede interpretar como que todo proceso  $AR(p)$  que cumple la condición (6.13) es equivalente a un  $MA(\infty)$ . También se denomina (6.16) la representación causal de  $Y_t$ . Los coeficientes  $\psi_j$  se pueden calcular con la función en R, `ARMAtoMA(ar = numeric(0), ma = numeric(0), lag.max)`. En el argumento “ar=” se ingresa el vector de coeficientes autorregresivos  $(\varphi_1, \varphi_2, \dots, \varphi_p)$ . El argumento “lag.max” es el número de coeficientes  $\psi_j$  deseados.

### Estimación de la FAC de un AR(p)

La fac de un  $AR(p)$ ,  $\rho(k) = Corr(Y_t, Y_{t+k})$ ,  $k = 1, 2, \dots, p, p+1, \dots$ , cumple lo siguiente.

1. Dadas la matriz  $A$  y los vectores  $\underline{\varphi}, \underline{\rho}$

$$A = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(2) & \cdots & \rho(p-2) \\ \rho(2) & \rho(3) & \cdots & \rho(p-3) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{bmatrix}, \quad \underline{\varphi} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}, \quad \underline{\rho} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix},$$

entonces es válido el sistema lineal  $p \times p$

$$A\underline{\varphi} = \underline{\rho}. \quad (6.17)$$

Luego, dados los estimadores de la fac  $\hat{\rho}(1), \dots, \hat{\rho}(p)$  se puede resolver (6.17) y colocar  $\hat{\underline{\varphi}} = \hat{A}^{-1}\hat{\underline{\rho}}$ . Estos estadísticos  $\hat{\underline{\varphi}}$  así calculados se denominan los estimadores de Yule-Walker de  $\underline{\varphi}$

2. Se puede comprobar que la fac de un  $AR(p)$  cumple la ecuación siguiente.

$$\rho(k) = \varphi_1\rho(k-1) + \varphi_2\rho(k-2) + \dots + \varphi_p\rho(k-p), \quad k = p, p+1, \dots \quad (6.18)$$

Entonces (6.18) forma una ecuación en diferencias finitas con condiciones iniciales  $\rho(1), \dots, \rho(p)$ , para  $\rho(k)$ ,  $k \geq p+1$ , con solución

$$\rho(k) = s_1g_1^k + s_2g_2^2 + \dots + s_pg_p^p, \quad (6.19)$$

donde  $g_i = 1/z_i$  y  $z_i$  es la  $i$ -ésima raíz de la ecuación característica

$$1 - \varphi_1z - \varphi_2z^2 - \dots - \varphi_pz^p = 0 \quad (6.20)$$

con  $|z_i| > 1 \Leftrightarrow |g_i| < 1$ , luego se debe cumplir que  $\rho(k) \rightarrow 0$ ,  $k \rightarrow \infty$

**Nota 6.3.1.** Si  $g_i \approx 1$ , por ejemplo  $g_i = 1 - \varepsilon$  se tendrá  $s_ig_i^k = s_i(1 - \varepsilon)^k$  y por tanto  $\rho(k)$  decae a cero más lento que si  $g_i = \varepsilon$ .

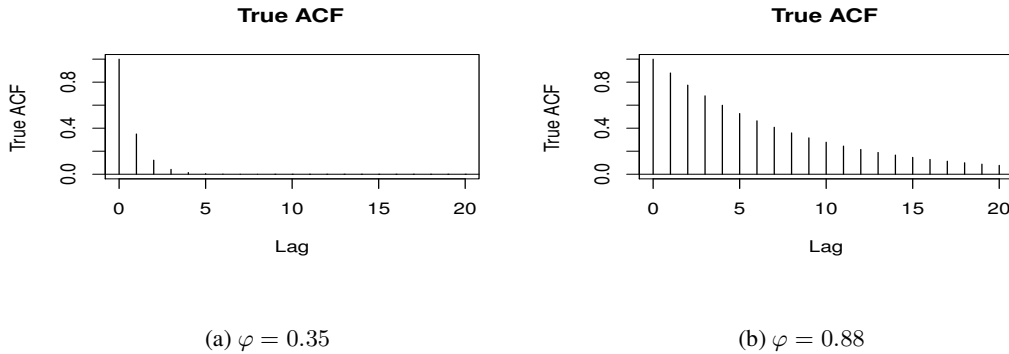


Figura 6.6: FAC de  $Y_t = \varphi Y_{t-1} + \varepsilon_t$ .

### FACP de los Procesos AR(p)

La FACP de un procesos  $AR(p)$  es  $\alpha(k)$  tal que  $\hat{\alpha}(k)$  es el coeficiente  $\hat{\beta}_{k,k}$  en la regresión

$$Y_t = \beta_0 + \beta_{k,1}Y_{t-1} + \cdots + \beta_{k,k}Y_{t-k} + a_t, \quad k = 2 \quad (6.21)$$

pero como  $\beta_{k,k} = 0$  si  $k \geq p + 1$  entonces  $\hat{\alpha}(k) = 0$  si  $k \geq p + 1$

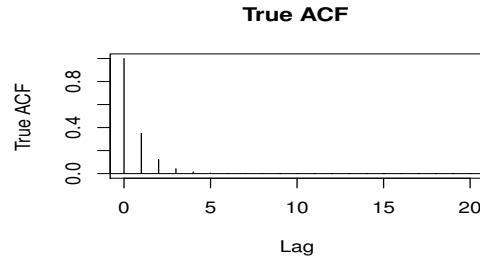


Figura 6.7: FACP Muestral de  $AR(p)$ .

**Ejemplo 6.3.1.** Sea  $Y_t \sim AR(2)$  con

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2) \\ Y_t &= 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t, \\ \varphi_2(z) &= 1 - 1.5z + 0.9z^2 = 0 && \text{ecuación característica} \\ z &= 0.83 \pm 0.64i, \quad |z| = 1.054 > 1 \end{aligned}$$

luego  $Y_t$  es estacionario en covarianza, además

$$\begin{aligned} \rho(k) &= 1 - 1.5\rho(k-1) + 0.9\rho(k-2), \quad k \geq 2 \\ \rho(0) &= 1, \quad \rho(1) = \frac{\varphi_1}{1 - \varphi_2} = \frac{1.5}{1.9} = 0.789. \end{aligned}$$

### 6.3.3. Proceso AR(1)

El proceso AR(1) se ha utilizado anteriormente por ejemplo, en la prueba Durbin-Watson. A continuación se desarrollan algunas de sus propiedades.

Si  $Y_t$  es un AR(1) de media  $\mu$ , entonces está definido por

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t, \quad \varphi_0 = (1 - \varphi_1)\mu, \quad (6.22)$$

donde el proceso  $Y_t$  es estacionario si todas las raíces  $z$  de la ecuación  $\varphi_1(z) = 0$  caen fuera del círculo unitario  $|z| > 1$ . Luego el  $AR(1)$  es estacionario en covarianza si y solo si  $|\varphi_1| < 1$ .

**Definición 6.3.2 (Marcha Aleatoria).** Se dice que  $Y_t$  es una marcha aleatoria (Random Walk) si cumple

$$Y_t = \mu + Y_{t-1} + \varepsilon_t, \quad (6.23)$$

nótese que es un  $AR(1)$  con  $\varphi_1 = 1$ .

### Propiedades del $AR(1)$

1.  $E(Y_t) = \mu = \frac{\varphi_0}{1 - \varphi_1}$
2.  $Cov(Y_t, Y_{t+k}) = \frac{\sigma^2 \varphi_1^k}{1 - \varphi_1^2}, \quad k = 0, 1, \dots$
3.  $\rho(k) = \varphi_1^k, \quad -1 < \varphi_1 < 1$ .

**Nota 6.3.2.** Diebold [1999, pág. 138], Si  $Y_t \sim AR(1)$  de media cero estacionario en covarianza entonces

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim R.B.(0, \sigma^2)$$

es decir

$$(1 - \varphi_1 L)Y_t = \varepsilon_t$$

y se puede escribir como

$$Y_t = \frac{1}{1 - \varphi_1 L} \varepsilon_t,$$

si se cumple que

$$f(z) = \frac{1}{1 - \varphi_1 z} = \sum_{j=0}^{\infty} \varphi_1^j z^j = 1 + \varphi_1 z + \varphi_1^2 z^2 + \dots$$

por que  $|\varphi_1 z| < 1$  ya que  $|\varphi_1| < 1$  y  $|z| < 1$ . Entonces

$$Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \dots,$$

y como los  $\varepsilon_t$  son incorrelacionados

$$\begin{aligned} \text{Var}(Y_t) &= \sigma^2 + \varphi_1 \sigma^2 + \varphi_1^2 \sigma^2 + \dots \\ &= \sigma^2(1 + \varphi_1 + \varphi_1^2 + \dots) = \sigma^2 \left( \frac{1}{1 - \varphi_1} \right) \end{aligned} \quad \text{Varianza Incondicional}$$

**Nota 6.3.3.** Si  $Y_t \sim AR(1)$  de media cero estacionario en covarianza entonces

$$E(Y_t | Y_{t-1}) = E(\varphi_1 Y_{t-1} + \varepsilon_t | Y_{t-1}) = \varphi_1 Y_{t-1},$$

y si se asume que  $\varepsilon_t$  son independientes de  $y_{t-1}, Y_{t-2}, \dots$

$$\begin{aligned} \text{Var}(Y_t | Y_{t-1}) &= \text{Var}(\varphi_1 Y_{t-1} + \varepsilon_t | Y_{t-1}) \\ &= \varphi_1^2 \text{Var}(Y_{t-1} | Y_{t-1}) + \text{Var}(\varepsilon_t) = \sigma^2 \end{aligned} \quad \text{Varianza Condicional}$$

## 6.4. Procesos Autoregresivos y de Medias Móviles ARMA(p,q)

Si en un proceso  $Y_t \sim AR(p)$

$$Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \dots - \varphi_p Y_{t-p} = \varepsilon_t, \quad \varepsilon_t \sim R.B.(0, \sigma^2), \quad (6.24)$$

se cambia  $\varepsilon_t$  por un modelo  $MA(q)$ ,  $Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$  entonces (6.24) queda

$$Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (6.25)$$

donde  $\varepsilon_t \sim RB(0, \sigma^2)$ . El efecto de este cambio es que los errores no se toman incorrelacionados sino con autocorrelación débil. 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. Usando la notación del operador de rezago (6.25) se puede definir el proceso  $ARMA(p, q)$  por

**Definición 6.4.1.** Un proceso  $Y_t \sim ARMA(p, q)$  se define mediante la ecuación (6.25), o también por

$$\varphi_p(L)(Y_t) = \theta_q(L)(\varepsilon_t), \quad t \in \mathbb{Z}, \quad (6.26)$$

donde  $\varepsilon_t \sim RB(0, \sigma^2)$ , y  $\varphi_p(z) = 1 - \sum_{j=1}^p \varphi_j z^j$ ,  $\theta_q(z) = 1 + \sum_{j=1}^q \theta_j z^j$  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), (6.26). Por lo tanto, se asume que las raíces de las ecuaciones  $\varphi_p(z) = 0$  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.

**Ejemplo 6.4.1.** Sea  $Y_t \sim ARMA(1, 1)$  dado por

$$\varphi_1(L)(Y_t) = \theta_1(L)(\varepsilon_t) \quad (6.27)$$

donde  $\varphi_1(L) = 1 - \varphi L$  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  $\varepsilon_t \sim RB(0, \sigma^2)$

**Ejemplo 6.4.2.** 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.

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{11} \delta_j I_j(t) + \varepsilon_t,$$

$$\varepsilon_t = \varphi \varepsilon_{t-1} + a_t + \theta a_{t-1},$$

con  $a_t \sim RB(0, \sigma^2)$ , el residuo del proceso arma. Nótese que el número de parámetros de este modelo es 16, incluyendo la varianza  $\sigma^2$ .

**Ejemplo 6.4.3.** Considere el proceso  $Y_t \sim ARMA(2, 1)$  dado por

$$Y_t = 2 + \frac{1 - 0.4L}{1 - 1.5L + 0.9L^2} \varepsilon_t, \quad \varepsilon_t \sim R.B.(0, \sigma^2)$$

osea

$$Y_t = 2(1 - 1.5 + 0.9) + 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t - 0.4\varepsilon_{t-1}$$

con ecuación característica

$$1 - 1.5z + 0.9z^2 = 0$$



y sus raíces dadas por

$$z = \frac{1.5 \pm \sqrt{1.5^2 - 4(0.9)}}{2(0.9)} = 0.83 \pm 0.645i$$

por tanto

$$|z| = 1.05 > 1$$

es un proceso estacionario en covarianza e invertible.

#### 6.4.1. Propiedades de los Modelos ARMA

1. Suponga  $Y_t \sim ARMA(p, q)$  entonces  $E(Y_t) = 0$ . Si el proceso es estacionario entonces se puede expresar  $\theta_q(z)/\varphi_p(z) = \sum_{j=0}^{\infty} \psi_j z^j$  con  $\psi_0 = 1$  y  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . A partir de  $\varphi_p(L)Y_t = \theta_q(L)\varepsilon_t$ , se puede escribir

$$Y_t = \frac{\theta_q(L)}{\varphi_p(L)}\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots,$$

por tanto

$$E(Y_t) = E(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots) = 0.$$

2. En caso de ser  $E(Y_t) = \mu \neq 0$  se coloca

$$Y_t = \mu + \frac{\theta_q(L)}{\varphi_p(L)}\varepsilon_t \quad (6.28)$$

de donde  $\varphi_p(L)Y_t = \varphi_p(1)\mu + \theta_q(L)\varepsilon_t$ . Por ejemplo, sea  $Y_t \sim ARMA(1, 1)$  con  $E(Y_t) = \mu$  entonces

$$Y_t = \mu + \frac{1 + \theta L}{1 - \varphi L}\varepsilon_t$$

luego

$$(1 - \varphi L)Y_t = (1 - \varphi L)\mu + (1 - \theta L)\varepsilon_t$$

pero  $(1 - \varphi L)\mu = \mu - \varphi\mu = (1 - \varphi)\mu$ , luego

$$Y_t = (1 - \varphi)\mu + \varphi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

3. La función de autocovarianza de un proceso  $Y_t \sim \text{ARMA}(p, q)$  estacionario de media cero. Si se indica por  $R(k) = \text{Cov}(Y_t, Y_{t+k})$  su función de autocovarianza, para  $k = 0, 1, \dots$  un método para calcular esta función se basa en la representación  $\varphi_p(L)Y_t = \theta_q(L)\varepsilon_t$ , con  $\theta_q(z)/\varphi_p(z) = \sum_{j=0}^{\infty} \psi_j z^j$ . Multiplicando ambos miembros por  $Y_{t-k}$  y tomando esperanza  $\mathbb{E}(\cdot)$  se obtienen las ecuaciones recursivas siguientes, similares a las ecuaciones Yule-Walker para AR(p), (6.15). Defina  $n = \max(p, q+1)$ ,

$$R(k) - \varphi_1 R(k-1) - \dots - \varphi_p R(k-p) = \begin{cases} \sigma^2 \sum_{j=k}^q \theta_j \psi_{j-k}, & \text{si } k = 0, 1, \dots, n-1 \\ 0, & \text{si } k = n, n+1, \dots \end{cases} \quad (6.29)$$

**Ejemplo 6.4.4.** (tomado de Brockwell and Davis [2002], pag. 93). Considere el proceso  $\text{ARMA}(2, 1)$  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

$$\begin{aligned} R(0) - R(1) + \frac{1}{4}R(2) &= \sigma^2(\psi_0\theta_0 + \psi_1\theta_1) = \sigma^2(1 + \psi_1), \\ R(1) - R(0) + \frac{1}{4}R(1) &= \sigma^2(\theta_1\psi_0) = \sigma^2\psi_0 = \sigma^2. \end{aligned} \quad (6.30)$$

Para calcular los coeficientes  $(\psi_j, j = 0, 1, \dots)$  se puede utilizar la función de `ARMAtoMA(a, b, m)`, la cual calcula los coeficientes  $\psi_j, j = 1, 2, \dots, m$ , dados los vectores  $a = (\varphi_1, \dots, \varphi_p)$  y  $b = (\theta_1, \dots, \theta_q)$ . 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 (6.29), se obtiene  $R(2) = R(1) - \frac{1}{4}R(0)$ , luego, reemplazando en el sistema (6.30), y resolviendo, se obtienen  $R(0) = 32\sigma^2/3$ ,  $R(1) = 28\sigma^2/3$ . Utilizando la segunda ecuación de (6.29),

$$R(k) = R(k-1) - \frac{1}{4}R(k-2), \quad k = 2, 3, \dots$$

se puede calcular la autocovarianza recursivamente.

4. La varianza de un proceso  $Y_t \sim \text{ARMA}(p, q)$  estacionario de media cero. Es evidente que a partir de la primera ecuación en (6.29) se puede definir un sistema lineal una de cuyas incógnitas es  $R(0)$ , la cual se resuelve en función de  $\sigma^2$ .

### 6.4.2. Librerías para identificación, estimación y pronósticos de modelos ARMA

El plan de análisis con procesos ARMA consiste en

1. Identificar el modelo ARMA(p,q).
2. Estimar el modelo.
3. Chequeo del ajuste del modelo a los datos.
4. Pronósticos con el modelo. O simulación del modelo.

Algunas de las librerías y funciones a utilizar para este análisis son

1. stat, con la función arima(), estima primero por mínimos cuadrados condicionales y luego por máxima verosimilitud.
2. tseries, con la función arma(), estima mediante mínimos cuadrados condicionales.
3. forecast, con la función auto.arima(), para identificación de modelos ARIMA.
4. FitAR, FitARMA, para estimación de AR(p) y ARMA(p,q).
5. timsac, con la función autoarmafit(), para identificación del modelo ARMA(p,q) con menor AIC. 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  $\theta_j$  se cambian de signo en esta librería. También tiene la función armafit() para ajuste de modelos ARMA.

### 6.4.3. 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 (6.18) 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  $k \geq q + 1$ .

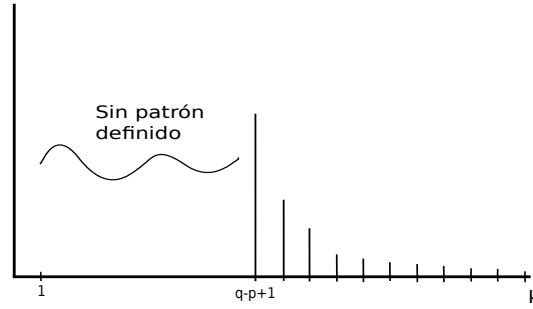


Figura 6.8: Fac Muestral de  $ARMA(p, q)$ .

Una alternativa consiste en buscar una pareja de órdenes  $(p, q)$  dentro de un rango inicial, por ejemplo  $p, q = 0, 1, \dots, 10$ , que minimice alguna función de  $\hat{\varepsilon}^2$ , como por ejemplo el AIC ó el BIC.

#### 6.4.4. Estimación de modelos ARMA

La estimación de procesos  $Y_t \sim ARMA(p, q)$  se basa en un supuesto: que el vector  $\underline{Y} = (Y_1, \dots, Y_n)'$  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:

$$\Sigma = \begin{bmatrix} R(0) & R(1) & \cdots & R(n-1) \\ R(1) & R(0) & \cdots & R(n-2) \\ R(2) & R(1) & \cdots & R(n-3) \\ \vdots & \vdots & & \vdots \\ R(n-1) & R(n-2) & \cdots & R(0) \end{bmatrix}. \quad (6.31)$$

Por ejemplo, para  $Y_t$  un  $AR(p)$ ,  $R(k)$  se calcula mediante las ecuaciones Yule-Walker, (6.15),  $R(k) = \mu + \sum_{j=1}^p \varphi_j R(k-j)$ . Por tanto, colocando  $\underline{\beta} = (\mu, \sigma^2, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)'$ , 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

$$f(\underline{y}, \underline{\beta}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma(\underline{\beta}))}} \exp \left( -\frac{1}{2} (\underline{y} - \underline{\mu}) \Sigma(\underline{\beta})^{-1} (\underline{y} - \underline{\mu})' \right) \quad (6.32)$$

donde  $\underline{\mu} = (\mu, \dots, \mu)' \in \mathbb{R}^n$ . La función de log-verosimilitud se define a partir del logaritmo de la densidad (6.32),  $\log(f(\underline{y}, \underline{\beta}))$ , y está dada por

$$\begin{aligned} L(\underline{\beta}) &:= \sum_{j=1}^n \log f(\underline{y}_j, \underline{\beta}) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det(\Sigma(\underline{\beta})) - \frac{1}{2} (\underline{y} - \underline{\mu})' \Sigma(\underline{\beta})^{-1} (\underline{y} - \underline{\mu}). \end{aligned} \quad (6.33)$$

El estimador ML de máxima verosimilitud de  $\underline{\beta}$  se define como

$$\hat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}}(-L(\underline{\beta})). \quad (6.34)$$

La identificación con base en el AIC se basa en comparar varios modelos  $ARMA(p, q)$  para valores de, por ejemplo,  $p = 0, 1, \dots, 20$  y  $q = 0, 1, \dots, 20$  con base en el criterio  $AIC$  el cual está definido por

$$AIC = -2L(\hat{\underline{\beta}}) + 2k, \quad k = p + q \quad (6.35)$$

Entonces identifica un posible modelo parsimonioso escogiendo el modelo con el mínimo  $AIC$ .

**Ejemplo 6.4.5.** Suponga que se requiere 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.

```
#----- simulacion de un ARMA(3,2)
#----- defina las raices del polinomio autorregresivo grado 3
z1 = complex(3)
z1[1] = -0.8 - 1.3i
z1[2] = conj(z1[1])
z1[3] = 1.2
```

Entonces las tres raíces son  $z1[1]$ ,  $z1[2]$ ,  $z1[3]$  y los coeficientes del polinomio autorregresivo de grado 3, mónico,  $\phi_3(z)$ ,

$$\begin{aligned} \phi_3(z) &= (z - z1[1])(z - z1[2])(z - z1[3]) \\ &= 1 - \varphi_1 z - \varphi_2 z^2 - \varphi_3 z^3 \\ &= 1 - 0.1466381 * z - 0.1430615 * z^2 - 0.3576538 * z^3 \end{aligned}$$

se calculan a continuación en el vector  $a$ .

```

a = poly.calc(z1)
a = a/a[1]
#----- defina las 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))
#----- 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)

plot(1:70, c(y[(3200-49):3200],py1$pred),
      ylim=c(min(py1$pred-1.64*py1$se),max()), type='b', col=2)
points(51:70,py1$pred, type='b', col='blue')
points(51:70, py1$pred+1.64*py1$se, type='l', col='blue')
points(51:70, py1$pred-1.64*py1$se, type='l', col='blue')

```

## 6.5. Modelos ARMA Estacionales, SARMA

Se define un modelo ARMA Estacional (<sup>1</sup>) como un modelo ARMA que incluye un modelo ARMA para los tiempos dentro de cada período estacional. Y otro modelo ARMA entre las estaciones, por ejemplo, entre los eneros, febreros, etc.. Se escribe  $Y_t \sim$

---

<sup>1</sup><http://robjhyndman.com/researchtips/cyclclcts/>

$ARMA(p, q)(p_s, q_s)_{[s]}$ .

$$Y_t \sim ARMA(p, q)(p_s, q_s)_{[s]} \Leftrightarrow \Phi_p(L)\Phi_{p_s}(L)Y_t = \Theta_q(L)\Theta_{q_s}(L)\epsilon_t. \quad (6.36)$$

donde  $(p, q)$  son los órdenes AR y MA del modelo ARMA ordinario y  $(p_s, q_s)$  los órdenes AR y MA estacionales, **entre** estaciones. Por ejemplo, el modelo  $Y_t \sim ARMA(1, 1)(1, 2)_{[s]}$  se indica por

$$(1 - \varphi_1 L)(1 - \varphi_s L^s)Y_t = (1 + \theta_1 L)(1 + \theta_s L^s + \theta_{2s} L^{2s})\epsilon_t, \quad (6.37)$$

donde, por ejemplo,  $s = 4, 12$ . Nótese que (6.37) es un tipo particular de ARMA. Otro ejemplo es  $Y_t \sim ARMA(1, 2)(1, 1)_{[s]}$  que se indica por

$$(1 - \varphi_1 L)(1 - \varphi_s L^s)Y_t = (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_s L^s)\epsilon_t, \quad (6.38)$$

**Una clave** para la identificación de modelos SARMA es que si se observa por ejemplo (6.38), con  $s = 12$ , 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  $1 + 12 = 13$ . Esto podría verse en la fac y en la fac parcial. La librería `library(TSA)` tiene la función `armasubsets` que calcula un panel para identificación para

## Identificación y Estimación de los SARMA

Los pasos siguientes estiman el modelo de tendencia lineal inicial. Luego muestra un procedimiento de identificación gráfico para el SARMA. De lo que se trata es de escoger los órdenes  $p, q, p_s, q_s$ .

Por ejemplo, si se supone que la serie tiene tendencia lineal y componente estacional de período  $s = 12$ . Para identificar un posible modelo SARMA en los residuos se programan las instrucciones siguientes.

```
#-----
mod2 = lm(yi ~ t )
summary(mod2)
r = mod2$residuals

# identificador grafico

library(TSA)
res=armasubsets(y=r, nar=14, nma=14, y.name='r',
ar.method='ols')
plot(res)
```

La última instrucción gráfica el panel (2,2) de la Figura 6.9. En la Figura 6.9 aparecen las

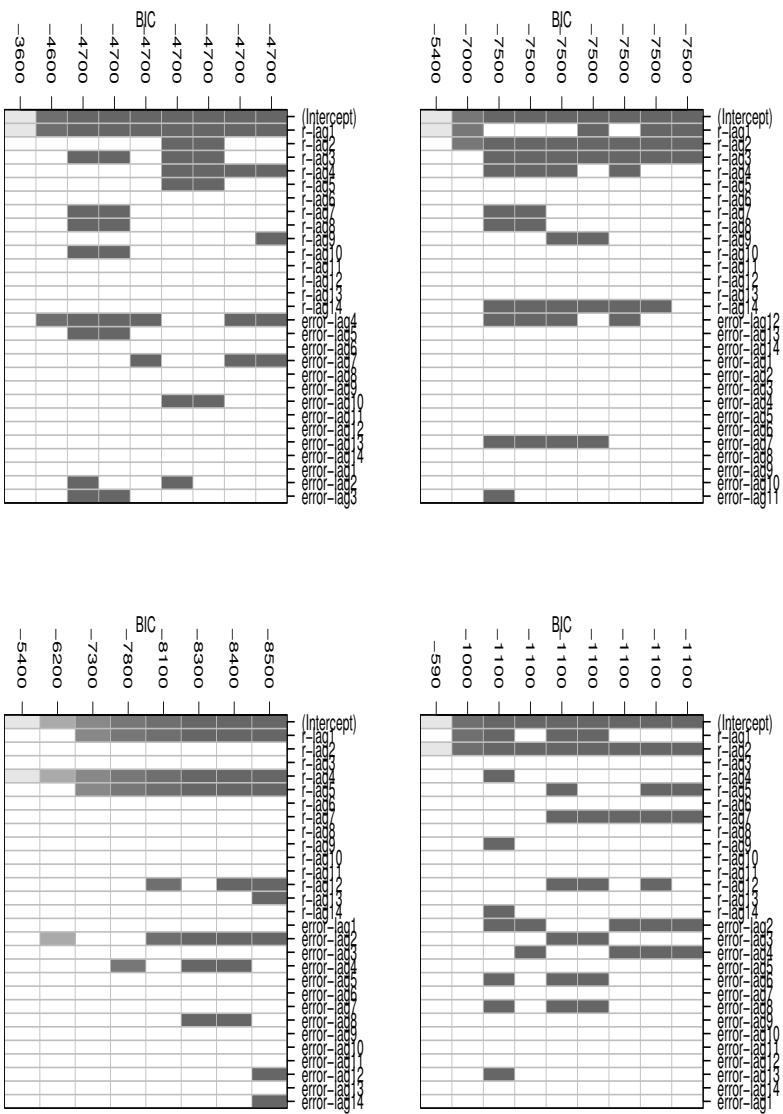


Figura 6.9: Posibles submodelos (1,1)=AR, (1,2),(2,1)=ARMA, (2,2)=SARMA identificaciones de un AR en la celda (1,1). En las celdas (1,2) y (2,1) aparecen modelos ARMA. Y en la (2,2) un SARMA.

En cada gráfica la información sobre los rezagos a incluir en el modelo está organizada 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 da 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.

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12}) Y_t = (1 + \sum_{j=1}^2 \theta_j L^j)(1 + \theta_{12} L^{12}) \epsilon_t, \quad (6.39)$$

### La estimación de los modelos SARMA

La estimación del modelo (6.39),  $Y_t \sim ARMA(1, 2)(1, 1)_{[12]}$ , 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.

### Una aplicación de los modelos ARMA estacionales, SARMA

Estos modelos son una alternativa a los modelos con componentes estacionales y de tendencia. Por ejemplo, un modelo con tendencia lineal y posible componente estacional,  $S_t$ , según el modelo de descomposición está dado por

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + S_t + \epsilon_t, \\ \epsilon_t &\sim ARMA(p, q). \end{aligned} \quad (6.40)$$

La alternativa es un modelo de la forma

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

$$\epsilon_t \sim SARMA(p, q)(p_s, q_s). \quad (6.41)$$

La idea es que en el modelo (6.41) el ARMA estacional modela la parte  $S_t + \epsilon_t$  del modelo (6.40).

---

## Raíces Unitarias y Tendencias Estocásticas (ARIMA)

---

### 7.1. Introducción

En este capítulo se presenta un modelo alternativo al de Descomposición 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.

### 7.2. 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

$$\begin{aligned} Y_t - Y_{t-1} &= a + bt + \varepsilon_t - (a + b(t-1) + \varepsilon_{t-1}) \\ &= b + \varepsilon_t - \varepsilon_{t-1} = b + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \theta_1 = -1, \end{aligned}$$

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.

**Definición 7.2.1 (Raíz Unitaria Autoregresiva).** Suponga un modelo  $ARMA(p, q)$

$$\varphi(L)_p Y_t = \theta(L)_q \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2),$$

tal que una de las  $p$  raíces de la ecuación  $\varphi_p(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p = 0$  es 1, entonces se dice que  $Y_t$  tiene una raíz unitaria autorregresiva.

En este caso el polinomio  $\varphi_p(L)$  factoriza como  $\varphi_p(L) = (1 - L)\varphi_{p-1}(L)$  donde  $\varphi_{p-1}$  es un polinomio de grado  $p - 1$  y

$$\varphi_{p-1}(1 - L)Y_t = \theta(L)\varepsilon_t,$$

luego si  $Y_t \sim ARMA(p, q)$  con una raíz unitaria entonces  $\Delta Y_t \sim ARMA(p - 1, q)$ .

**Ejemplo 7.2.1.** Si  $\varphi(L) = (1 - 0.2L)(1 - L)$ ,  $\theta(L) = 1 - 0.3L$

$$\begin{aligned} (1 - 0.2L)(1 - L)Y_t &= (1 - 0.2L)\varepsilon_t \\ (1 - 1.2L + 0.2L^2)Y_t &= \varepsilon_t - 0.3\varepsilon_{t-1} \\ Y_t - 1.2Y_{t-1} + 0.2Y_{t-2} &= \varepsilon_t - 0.3\varepsilon_{t-1}. \end{aligned}$$

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 7.2.2.** La marcha aleatoria sin tendencia se define como un proceso  $Y_t \sim AR(1)$ , de media cero, con coeficiente  $\varphi = 1$ , es decir

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (7.1)$$

La marcha aleatoria con tendencia se define como un proceso  $AR(1)$  de media diferente de cero, con coeficiente  $\varphi = 1$ , es decir

$$Y_t = \mu + Y_{t-1} + \varepsilon_t, \quad \mu \in \mathbb{R}. \quad (7.2)$$

La marcha aleatoria sin tendencia cumple, para  $t \geq 0$ ,  $Y_t = Y_0 + \sum_{i=0}^t \varepsilon_i$ . Luego  $E(Y_t|Y_0) = Y_0$  y  $Var(Y_t) = t\sigma^2$ .

Para la marcha aleatoria con tendencia

$$Y_t = t\delta + Y_0 + \sum_{i=0}^t \varepsilon_i, \quad (7.3)$$

$$E(Y_t|Y_0) = t\delta + Y_0, \quad (7.4)$$

$$Var(Y_t) = t\sigma^2. \quad (7.5)$$

Un ejemplo de trayectorias de marchas aleatorias con y sin tendencia se muestran a continuación en la Figura 7.1.

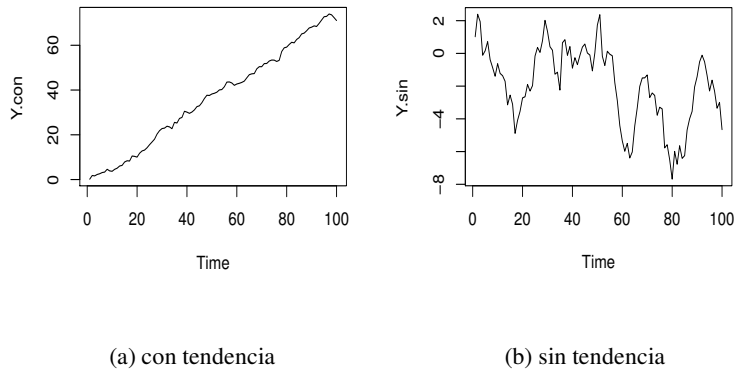


Figura 7.1: Trayectorias de Marchas Aleatorias

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,  $E(Y_t - Y_{t-1}) = \mu$  y  $E(X_t - X_{t-1}) = b$ .

Suponga  $Y_t = \mu + \varphi Y_{t-1} + \varepsilon_t$ ,  $t = 1, 2, \dots, T$ , 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  $j = 1, 2, \dots, h$ . 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 pronóstico es el último valor conocido.

Finalmente, en una marcha aleatoria con tendencia, se tiene  $Y_t = Y_0 + \mu t + \sum_{j=1}^t \varepsilon_j$ , y, por tanto, para  $k = 1, 2, \dots$ ,

$$Cov(Y_t, Y_{t+k}) = \mathbb{E}((Y_t - \mathbb{E}(Y_t))(Y_{t+k} - \mathbb{E}(Y_{t+k})))$$

$$\begin{aligned}
&= \mathbb{E}\left(\sum_{j=1}^t \varepsilon_j \sum_{j=1}^{t+k} \varepsilon_j\right) = \sum_{j=1}^t \sum_{i=1}^t \mathbb{E}(\varepsilon_j \varepsilon_i) \\
&= \sigma^2 t,
\end{aligned}$$

Como  $\text{Var}(Y_t) = \sigma^2 t$  entonces

$$\begin{aligned}
\text{Corr}(Y_t, Y_{t+k}) &= \sigma^2 t / \sqrt{\sigma^2 t (\sigma^2 (t+k))}. \\
&= 1 / \sqrt{1 + k/t}
\end{aligned}$$

Para  $t$  grande tal que  $k/t$  sea muy pequeño se puede aproximar 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.

**Definición 7.2.3.** 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.

**Definición 7.2.4.** Una serie  $Y_t, t \in \mathbb{Z}$  se dice que tiene tendencia estocástica si es integrada de orden  $d = 1, 2, I(d)$ .

La marcha aleatoria con tendencia (7.2) es un ejemplo de tendencia estocástica, en contraposición a una serie con tendencia lineal  $T_t = a + bt$ , la cual se dice que tiene tendencia determinística.

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 períodos 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.

La implementación de los modelos ARIMA empieza con series con tendencia estocástica. La metodología consiste en diferenciar la serie  $d$  veces hasta obtener otra serie estacionaria en covarianza, identificable por un modelo ARMA, y estimar los parámetros de este modelo.

**Definición 7.2.5 (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

$$\varphi(L)(1-L)Y_t = \delta + \theta(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (7.6)$$

y  $\varphi(z) = 0$  tiene las raíces fuera del círculo unitario.

Una serie de tiempo sigue un modelo  $ARIMA(p, d, q)$  si

$$\Delta^d Y_t = (1 - L)^d Y_t, \quad d = 1, 2, \dots$$

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]. De acuerdo con la definición (7.2.3) de serie integrada si  $Y_t \sim ARIMA(p, 1, q)$  entonces es integrada.

**Ejemplo 7.2.2.** 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:

$$\begin{aligned} \rho(1) &= \frac{(1 + \varphi\theta)(\varphi + \theta)}{1 + \theta^2 + 2\varphi\theta} \\ \rho(k) &= \rho(k - 1), \quad k \geq 2. \end{aligned}$$

Si  $\varphi \rightarrow 1$  obtenemos un  $ARIMA(0, 1, 1) = IMA(1, 1)$  y

$$\rho(1) \rightarrow \frac{(1 + \theta)(1 + \theta)}{1 + \theta^2 + 2\theta} = 1.$$

Por lo tanto  $\rho(k) \equiv 1, \forall k \geq 1$ . Nótese que el  $IMA(1, 1)$  es similar a la marcha aleatoria sin tendencia. Esto da lugar a afirmar que “los procesos  $ARIMA(p, 1, q)$  se comportan como marchas aleatorias en ciertos aspectos”.

**Ejemplo 7.2.3.** 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 $AR(2)$

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \varepsilon_t \\ \varepsilon_t &= \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \eta_t, \quad \eta_t \sim RB(0, \sigma^2). \end{aligned}$$

#### 2. Modelo con tendencia aleatoria

$$\begin{aligned} Y_t &\sim ARIMA(1, 1, 0) \text{ con tendencia} \\ (1 - \varphi L)(1 - L)Y_t &= \delta + \varepsilon_t \\ Z_t &= \delta + \varphi Z_{t-1} + \varepsilon_t, \quad Z_t = (1 - L)Y_t \end{aligned}$$

La primera diferencia de  $Y_t$  sigue un modelo  $AR(1)$  con media diferente de cero.

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

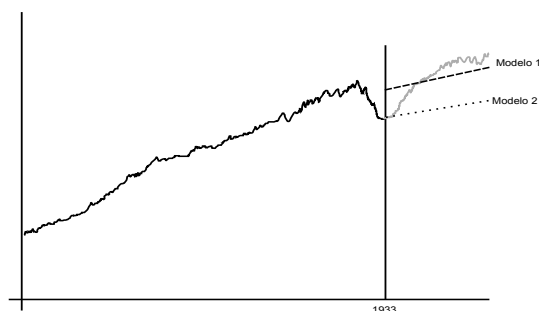


Figura 7.2: Pronósticos de la Serie PNB-USA.

El modelo 1 produce mejores pronósticos que siguen la tendencia, en cambio, el modelo 2 no, pues subestima los valores.

**Ejemplo 7.2.4.** 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.

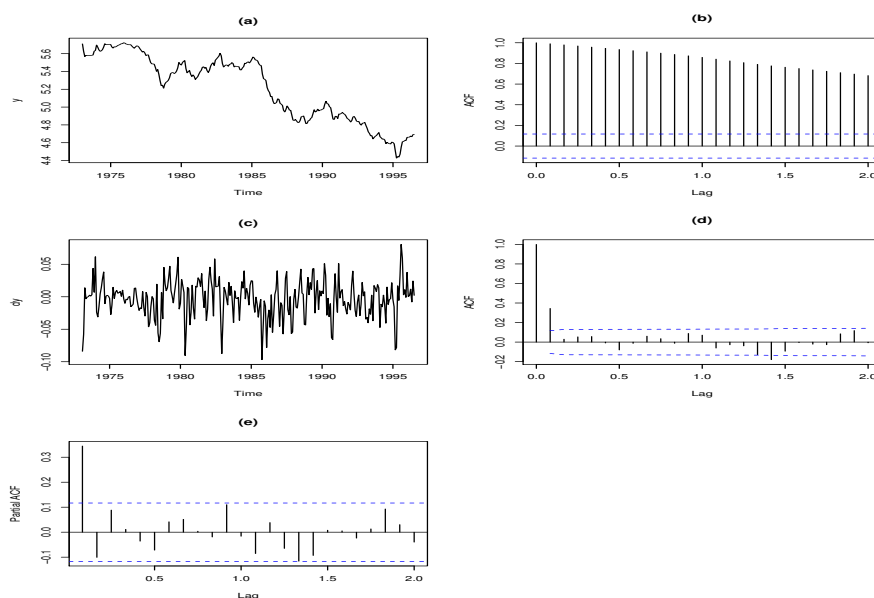


Figura 7.3: (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada



La Figura 7.3 (b) sugiere un proceso integrado y las Figuras (d) y (e) sugieren un ARMA para la serie diferenciada. Utilizando la función  $z = \text{autoarmafit}(dy)$ , de la librería `timsac`, con la serie diferenciada,  $dy = \text{diff}(y, 1, 1)$ , se obtiene un modelo ARMA(3,2). Utilizando la función  $\text{mod} = \text{arima}(y, \text{order}=c(3, 1, 2))$  se obtienen los resultados de la Tabla 7.1. Finalmente, utilizando la prueba Ljung-Box, se comprueba

Tabla 7.1: 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 resultado de la prueba es

$$X\text{-squared} = X\text{-squared} = 15.7889, \quad df = 26, \quad p\text{-value} = 0.9411,$$

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 7.4, comparados con los pronósticos de un modelo con tendencia lineal y error estructural AR(2).

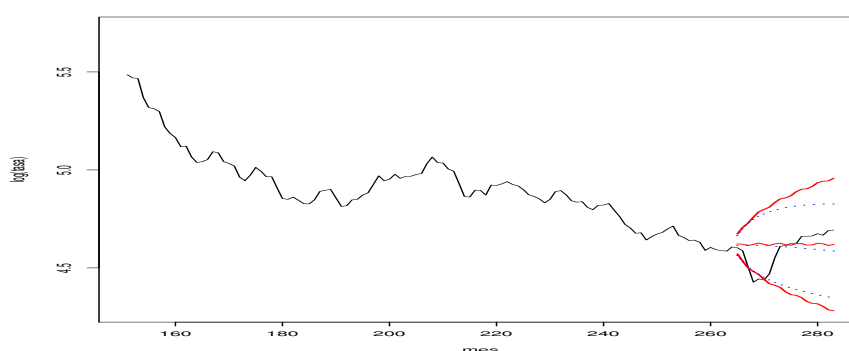


Figura 7.4: Pronósticos  $\text{usd/yen}$  con ARIMA(3,1,2)(continua) y Tendencia Lineal+AR(2) (punteada)

**Ejemplo 7.2.5.** 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 7.6.

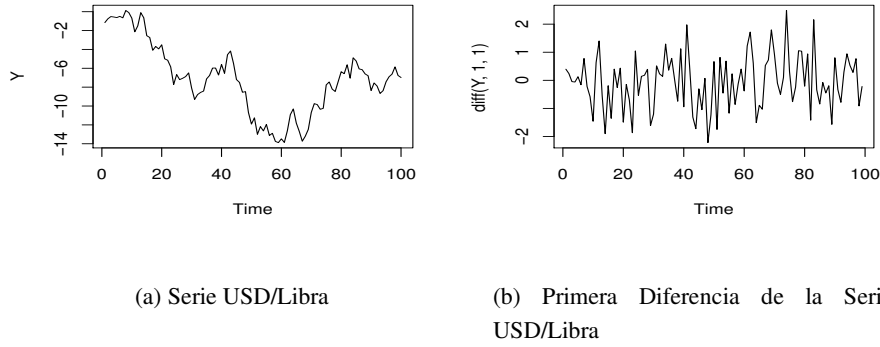


Figura 7.5: Serie USD por Libra

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 7.2.6.** Sea  $Y_t$  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,

$$(1 - \varphi L)(1 - L)Y_t = \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (7.7)$$

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,

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad (7.8)$$

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 (7.7) 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:  $p\text{-value} = 0.8412$ . 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
```

## 7.3. Raíces Unitarias, Estimación y Prueba de Hipótesis

### Minimos Cuadrados Ordinarios (MCO) y Raíces Unitarias

La discusión en general es para un  $Y_t \sim ARIMA(p, 1, q)$  pero se puede utilizar una marcha aleatoria sin tendencia. Suponga un proceso  $AR(1)$ ,  $Y_t = \varphi Y_{t-1} + \varepsilon_t$ , en el cual  $\varphi = 1$  pero  $\varphi$  se estima usando MCO. El estimador de MCO de  $\varphi$ ,  $\hat{\varphi}$ , tiene dos propiedades

#### 1. Superconsistencia

a) En el caso  $\varphi = 1$  el estimador de MCO  $\hat{\varphi}_T$  tiene la propiedad de que

$$T(\hat{\varphi}_T - 1) \xrightarrow{d} Z, \quad T \rightarrow \infty, \quad (7.9)$$

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

$$\sqrt{T}(\hat{\varphi}_T - \varphi) \xrightarrow{d} Z \quad T \rightarrow \infty, \quad (7.10)$$

donde  $Z$  es variable aleatoria no degenerada.

como  $\sqrt{T} < T$ , se dice que la convergencia en el caso (7.9) es más rápida que en el caso (7.10), y a esto se denomina superconsistencia, es decir, el estimador de MCO,  $\hat{\varphi}_T$  de una raíz unitaria es superconsistente.

## 2. Sesgo de $\hat{\varphi}$

Si  $\hat{\varphi}_T$  es el estimador de MCO de  $\varphi$  entonces  $E(\hat{\varphi}_T) < \varphi$  y el sesgo es mayor cuando  $\varphi = 1$ . El sesgo es  $\varphi - \hat{\varphi}_T$  y crece si se considera tendencia. Aunque  $\hat{\varphi}_T$  converge a  $\varphi$ , cuando  $T \rightarrow \infty$ , 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 y la facp muestral presenta un valor cercano a 1 en el rezago  $k = 1$  y el resto son aproximadamente cero. Por ejemplo la Figura 7.6 siguiente de la fac y la facp de la tasa USD/Libra.

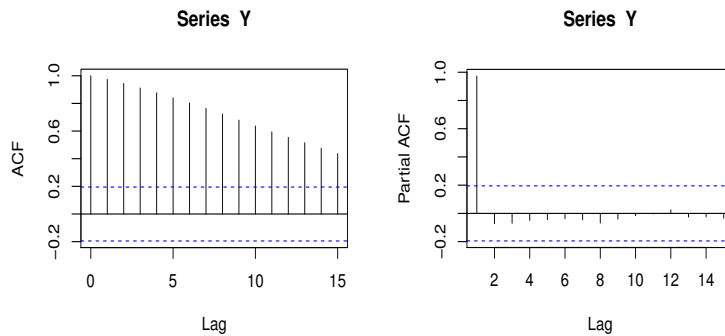


Figura 7.6: FAC y FACP Muestral de un  $ARIMA(p,1,q)$ .

Hay que tener una 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.

1. Diebold [1999, pág. 221]: “Si no hay raíz unitaria es conveniente utilizar modelos de niveles (componentes determinísticas) y es adecuado 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 determinística, tomar diferencias introduce una componente MA no invertible el cual causa problemas serios de estimación.”

### 7.3.1. 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) y se procede a transformar el 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 equivale a  $\varphi_1 = 1$ , es decir, a  $\rho = 0$ .

1. Prueba DF para el caso 1: suponiendo que  $Y_t \sim AR(1)$  con media cero, entonces

$$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t. \quad (7.11)$$

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, de tal forma que el criterio de rechazo es  $\hat{\tau} < \tau_{0.05}$ , con  $\hat{\tau}$  es valor calculado del estadístico. En R la prueba DF se implementa mediante la librería `urca` por medio de la función `ur.df(y, type="none", lags=0)`.

2. Prueba DF para el caso 2: suponiendo que  $Y_t \sim AR(1)$  con media diferente de cero, entonces

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \quad (7.12)$$

con la misma hipótesis. En R es `ur.df(y, type="drift", lags=0)`.

3. Prueba DF para el caso 3: suponiendo que  $Y_t \sim AR(1)$  con tendencia lineal, entonces

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t, \quad (7.13)$$

con la misma hipótesis. En R es `ur.df(y, type="trend", lags=0)`.

### 7.3.2. 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 incluirse al correr la regresión para la prueba de raíz unitaria. Incluir demasiados regresores puede generar una pérdida de potencia, mientras que no incluir suficientes puede generar resultados sesgados hacia el no rechazo de la hipótesis nula...incluir 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 determinar cuál de los siguientes tres casos determina una mejor aproximación a la serie original,  $Y_t$ . Nótese que los casos a examinar dependen del orden autorregresivo  $p$ , por lo que la búsqueda para por ejemplo,  $p = 1, 2, 3$ , requiere  $3 \times 3$  casos. La decisión se toma con base en el menor AIC.

**Caso 1** Suponiendo que  $Y_t \sim AR(p)$  con media cero, (“none”) entonces

$$Y_t = \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t = \varphi_1 Y_{t-1} + \sum_{j=2}^p \varphi_j Y_{t-j} + \varepsilon_t. \quad (7.14)$$

Defina

$$\rho_1 = \sum_{j=1}^p \varphi_j, \quad \rho_i = -\sum_{j=i}^p \varphi_j, \quad i = 2, \dots, p,$$

entonces con  $\rho = \rho_1 - 1$ , la ecuación (7.14) se transforma en

$$\Delta Y_t = \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (7.15)$$

Si hay una raíz unitaria se cumple  $1 - \varphi_1 - \varphi_2 - \dots - \varphi_p = 0$ , es decir,  $\rho_1 = 1$ . En este caso se tiene que el modelo (7.15) equivale a

$$\begin{aligned} Y_t - Y_{t-1} &= \sum_{j=2}^p \rho_j (Y_{t-j+1} - Y_{t-j}) + \varepsilon_t \\ \Delta Y_t &= \sum_{j=1}^{p-1} \rho_{j+1} \Delta Y_{t-j} + \varepsilon_t \end{aligned} \quad (7.16)$$

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,  $\hat{\tau}$  tiene la misma distribución asintótica que el estadístico DF para  $\rho = 0$ , ver caso 1 (7.11).

«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 (“drift”). Con la misma notación del caso 1,

$$\begin{aligned} Y_t - \mu &= \sum_{j=1}^p \varphi_j (Y_{t-j} - \mu) + \varepsilon_t \\ Y_t &= \alpha + \varphi_1 Y_{t-1} + \sum_{j=2}^p \rho_j (Y_{t-j+1} - Y_{t-j}) + \varepsilon_t \\ \Delta Y_t &= \alpha + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \end{aligned} \quad (7.17)$$

donde  $\alpha = \mu(1 - \sum_{j=1}^p \varphi_j)$ , con los  $\rho_j$  definidos como en el caso 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 porque  $\sum_{j=1}^p \varphi_j = 1$ . La distribución asintótica del estadístico DF es igual a la del caso  $AR(1)$  con media, ver caso 2 (7.12).

**Caso 3** Modelo  $AR(p)$  con tendencia lineal (“trend”). En este caso se define

$$Y_t = a + bt + \sum_{j=1}^p \varphi_j (Y_{t-j} - a - b(t-j)) + \varepsilon_t \quad (7.18)$$

que se puede reordenar de la siguiente forma

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (7.19)$$

donde

$$\begin{aligned} k_1 &= a \left( 1 - \sum_{i=1}^p \varphi_i \right) + b \sum_{i=1}^p i \varphi_i, \\ k_2 &= b \left( 1 - \sum_{i=1}^p \varphi_i \right), \\ \rho_1 &= \sum_{i=1}^p \varphi_i, \quad \rho = \rho_1 - 1. \end{aligned}$$

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), ver caso 3 (7.13).

No rechazar la hipótesis nula de raíz unitaria no necesariamente significa que se deba asumir que existe, debido a la baja potencia en estas pruebas. 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 ADF, pero no se desarrollará este tema.

**Ejemplo 7.3.1.** *Continuando con el Ejemplo 7.2.4, del logaritmo de la tasa de cambio USD/Yen, 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.*

**La Prueba Dickey-Fuller.** 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 yen/usd
library(forecast)
library(urca)
library(fUnitRoots)
uno = read.table("yen_usd.dat", header=T, stringsAsFactors=F)
```



```

attach(unos)
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)

----- 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

```

La prueba DF 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, luego no rechaza la hipótesis nula de raíz 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 yen-usd.

**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 incluir valores rezagados  $Y_{t-j}$ , con el comando `L(y, j)`, tendencia lineal  $a + bt$  con el comando `trend(y)`.

Por ejemplo, para programar el modelo de la ecuación (7.19), en la pag. 132,

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t, \quad (7.20)$$

con  $p = 3$ , queda  $\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \rho_2 \Delta Y_{t-1} + \varepsilon_t$ , se programa

```
reg.trend1 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1)).
```

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')`.

```
## estrategia de regresiones para la df aumentada
require(dynlm)

reg.drift0 = dynlm(dly ~ L(ly,1))
reg.drift1 = dynlm(dly ~ L(ly,1) + L(dly, 1))
reg.drift2 = dynlm(dly ~ L(ly,1) + L(dly, 1) + L(dly, 2))
reg.drift3 = dynlm(dly ~ L(ly,1) + L(dly, 1) + L(dly, 2) + L(dly, 3))

(c(AIC(reg.drift0),AIC(reg.drift1),AIC(reg.drift2),AIC(reg.drift3)))

reg.trend0 = dynlm(dly ~ trend(ly) + L(ly,1))
reg.trend1 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1))
reg.trend2 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1) + L(dly, 2))
reg.trend3 = dynlm(dly ~ trend(ly) + L(ly,1) + L(dly, 1) + L(dly, 2)
+ L(dly, 3))

(c(AIC(reg.trend0),AIC(reg.trend1),AIC(reg.trend2),AIC(reg.trend3)))

## se detecta el caso reg.trend1 como el de menor aic

df.trend = ur.df(y = ly, lags = 2, type = 'trend' )
summary(df.trend)

-----resultados de los valores AIC para los 8 modelos
-1202.376 -1240.019 -1236.218 -1232.318
-1203.215 -1245.147 -1240.917 -1237.595

-----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

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. Luego, es válido diferenciar la serie. En el Ejemplo 7.2.4 se ajustó el modelo ARIMA(3,1,2) que produce buenos pronósticos del logaritmo de la tasa.



---

## Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)

---

### 8.1. Modelos SARIMA

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, transformando  $Y_t$  a

$$W_t = (1 - L^s)^D Y_t = \Delta_s^D Y_t, \quad D = 1, 2, \quad (8.1)$$

esperando que la serie  $W_t$  sea estacionaria en covarianza y así proceder a buscar una estructura *ARMA*.

#### 8.1.1. Modelo SARIMA

Sin embargo, si hay estacionalidad, pueden existir estructuras *ARIMA intra* y *entre* los  $s$  períodos. Las *ARIMA intra* se refieren a modelos de la forma

$$\varphi_p(L) \Delta^d Y_t = \theta_q(L) \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (8.2)$$

con  $d = 0, 1, 2$ . Los *ARIMA entre* se refieren a modelos de la forma

$$\Phi_{p_s}(L^s) \Delta_s^D Y_t = \Theta_{q_s}(L^s) \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (8.3)$$

con  $D = 0, 1, 2$ . Y se define el modelo  $SARIMA(p, d, q)(p_s, D, q_s)_s$  como

$$\varphi_p(L)\Phi_{p_s}(L^s)\Delta^d\Delta_s^D Y_t = \theta_q(L)\Theta_{q_s}(L^s)\varepsilon_t, \quad (8.4)$$

La transformación (filtro lineal)  $X_t = \Delta^d\Delta_s^D Y_t = (1-L)^d(1-L^s)^D Y_t$  es la que elimina la tendencia y la estacionalidad de  $Y_t$ , dejando una estructura  $ARMA(p + sp_s, q + sq_s)$ . En el caso  $d = D = 0$  se obtiene un modelo  $SARMA(p, q)(p_s, q_s)$ , (6.36), pag. 115.

La idea es que la serie filtrada  $X_t$  en el modelo

$$\varphi_p(L)\Phi_P(L^s)X_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2).$$

tiene un papel similar o equivalente al residuo estructural  $\varepsilon_t$  en el modelo

$$Y_t = a + bt + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t. \quad (8.5)$$

En (8.4) al diferenciar la serie  $Y_t$  se eliminan la tendencia y estacionalidad aleatorias. Pero diferenciar en (8.5) 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 permitan decidir entre ambos modelos cuando hay evidencia de estacionalidad.

**Ejemplo 8.1.1.** 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

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})(1 - L)(1 - L^{12})Y_t = \varepsilon_t,$$

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. 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 8.1.2.** 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.

```
library(CombMSC)
library(forecast)

ti=0.8; ti4=-0.137;
sigma= sqrt(2)

Y = sarima.Sim(n = 22, period =12,
model=list(order=c(1,1,0),ar = ti,
ma = NULL, sd = sigma),
seasonal=list(order=c(1,1,0),ar=ti4,ma = NULL))

ts.plot(Y)
auto.arima(Y)
```

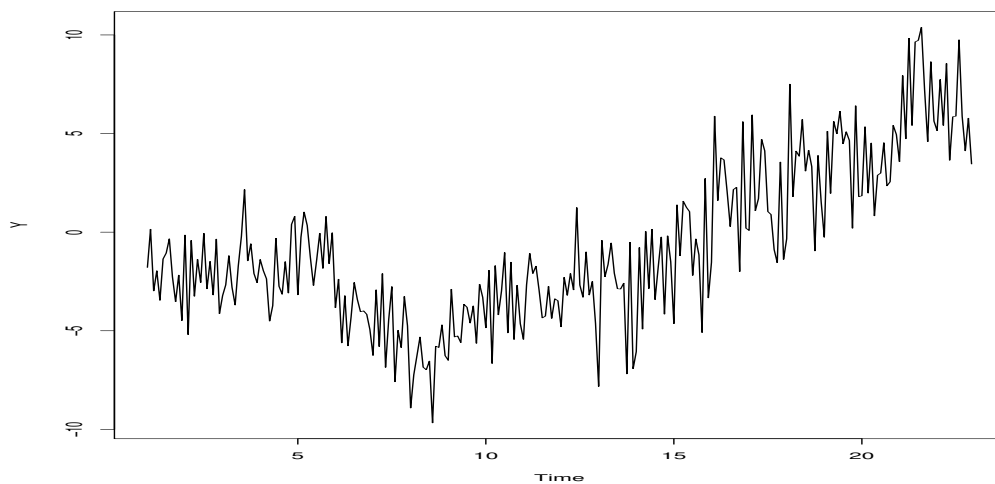


Figura 8.1: Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12]

**Ejemplo 8.1.3. Modelo “Air-passengers”** Se define como un modelo multiplicativo de la forma  $SARIMA(0, 1, 1)(0, 1, 1)_{12}$  dado por

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1)(1 + \theta_{12}L^{12})\varepsilon_t, \quad |\theta_1| < 1, |\theta_{12}| < 1. \quad (8.6)$$

*Es un modelo utilizado con frecuencia para modelar series con tendencia lineal y componente estacional. Por ejemplo la serie The classic Box Jenkins airline data. 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*

$$\rho_0 = (1 + \theta_1^2)(1 + \theta_{12}^2)$$

$$\rho_1 = \theta_1(1 + \theta_{12}^2)$$

$$\rho_{11} = \theta_1\theta_{12}$$

$$\rho_{12} = \theta_{12}(1 + \theta_1^2)$$

*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
ti=-0.377; ti4=-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=ti4,ar = NULL))
# graficar
ts.plot(Y)
# identificar
modelo = auto.arima(Y)
summary(modelo)
# estimar
arima(Y,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))
```



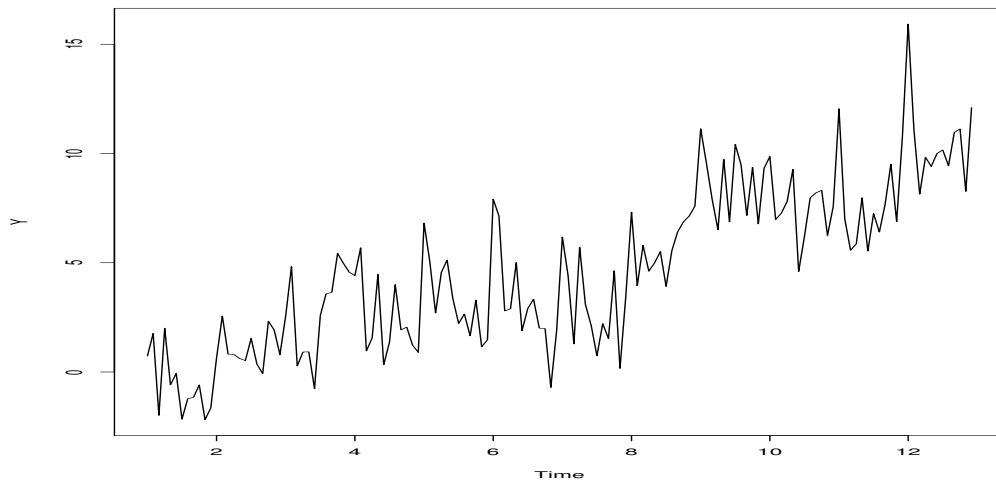


Figura 8.2: Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12]

**Ejemplo 8.1.4.** Sea un proceso SARIMA(1, 1, 1)(1, 1, 1)<sub>4</sub> definido por

$$(1 - \varphi_1 L)(1 - \varphi_4 L^4)(1 - L)(1 - L^4)Y_t = (1 - \theta_1 L)(1 - \theta_4 L^4)\varepsilon_t$$

se simula con el código siguiente. Es interesante examinar si en R es equivalente estimar el modelo SARIMA con la serie  $Y_t$ , y el modelo ARMA con la serie diferenciada,  $W_t = (1 - L)(1 - L^4)Y_t$ .

```
library(CombMSC)
library(forecast)

fi=-0.8; fi4=-0.37; ti=-0.64; ti4=-0.513; sigma= sqrt(0.014)

Y = sarima.Sim(n = 60, period =4,
model=list(order=c(1,0,1),
ar = fi, ma = ti, sd = sigma),
seasonal=list(order=c(1,1,1),ar=fi4,ma = ti4))

ts.plot(Y)

auto.arima(Y)
arima(Y,order=c(1,1,1),seasonal=list(order=c(1,1,1),period=4))
```

```

W = diff(diff(Y, 1, 1), 4, 1)
acf(W)
pacf(W)

arima(W, order=c(1, 0, 1), seasonal=list(order=c(1, 0, 1), period=4))

```

## 8.2. Pruebas de Raíz Unitaria Estacional

A partir de la definición de los modelos SARIMA hay dos posibles modelos para una serie con componente estacional:

1. Modelos con componente determinística, es decir, el modelo de descomposición.
2. Modelos no estacionarios, integrados estacionalmente, es decir, SARIMA(p,d,q)(P,D,Q)[s], con  $d \neq 0$  y  $D \neq 0$ .

Sin embargo, un modelo de la forma SARIMA(p,0,q)(P,0,Q)[s], equivalente a un ARMA(p+P,q+Q), podría ser tanto estacionario como mostrar características estacionales. Por tanto, un tercer modelo es el anterior, que se denomina un modelo “estacionario estacional”. El modelo con componente determinística no es estacionario, pero si se elimina la componente de tendencia, el proceso filtrado resulta la suma de un proceso determinístico periódico y otro estacionario en covarianza. Concretamente,  $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$ . Este modelo también se denominará “estacionario estacional”.

Se necesitan técnicas para discriminar cuál modelo aplicar. Aunque en el curso se han implementado criterios tales como la validación cruzada y la comparación del MAPE para escoger el mejor modelo para pronosticar, en la teoría de series de tiempo se ha dado mucho desarrollo en las pruebas de raíces unitarias, por ejemplo, las pruebas para detectar raíces unitarias estacionales, como las pruebas de Hylleberg y otros, indicada por HEGY, y la prueba Canova-Hansen, indicada por CH, que se exponen a continuación, con el propósito de utilizarlas como herramientas para escoger el modelo más adecuado.

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  $\varphi_p(z)$  tiene una raíz igual a 1. Es decir, si se cumple que  $\varphi_p(1) = 0$ . En este caso el polinomio se puede factorizar como  $\varphi_p(z) = (1 - z)\varphi_{p-1}(z)$ .

**Definición 8.2.1.** 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  $\varphi_p(z)$ .

Las raíces de  $z^s = 1$ , para por ejemplo  $s = 4, 12$ , se denominan las raíces de la unidad, y son  $s$  números complejos. 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. Para  $s = 12$  las raíces son  $\pm 1, \pm i, -(1 \pm i\sqrt{3})/2, (1 \pm i\sqrt{3})/2, -(\sqrt{3} \pm i)/2, (\sqrt{3} \pm i)/2$ .

Con la representación  $z = re^{i\theta}$  donde  $r = |z|$  y  $\theta$  es el ángulo que forma  $|z|$  con el eje X, se pueden identificar las cuatro raíces unitarias con los ángulos  $\theta = 2\pi j/4, j = 0, 1, 2, 3$ , ya que  $r = 1$  en todos los casos, pues las raíces unitarias tienen módulo 1. Estos ángulos se denominan las “frecuencias estacionales”.

### 8.2.1. La prueba HEGY

La prueba HEGY 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 se concluirá que la serie en cuestión está integrada estacionalmente integrada.

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. Lo que sigue está tomado del artículo de Hylleberg et al. [1990].

Inicialmente se asume que la serie  $Y_t$  está generada por un proceso AR(p), es decir, se cumple

$$\varphi(L)Y_t = \varepsilon_t,$$

para  $\varepsilon_t$  ruido blanco. Se asume que están dados  $p$  números complejos  $\theta_j \in \mathbb{C}, j = 1, \dots, p$ , en los cuales  $\varphi(\theta_j)$  es un número finito. Estos  $p$  números son las  $p$  raíces estacionales. El valor  $p$  es el período, indicado anteriormente por  $s$ . Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - z/\theta_k, \\ \Delta(z) &= \prod_{k=1}^p \delta_k(z), \\ \lambda_k &= \varphi(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k). \end{aligned} \tag{8.7}$$

Entonces se cumple, por una identidad (debida a Lagrange), que

$$\varphi(z) = \sum_{k=1}^p \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\varphi^*(z), \tag{8.8}$$

donde  $\varphi^*(z)$  es un polinomio posiblemente infinito ó racional. Nótese que, por la definición (8.7) se cumple

$$\varphi(\theta_k) = 0 \Leftrightarrow \lambda_k = 0 \quad (8.9)$$

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 (8.8). 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  $\varphi(z) = 1 - z^4$ , donde  $\theta_1 = 1, \theta_2 = -1, \theta_3 = i, \theta_4 = -i$ , la identidad (8.8) se transforma en

$$\begin{aligned} \varphi(z) = & \lambda_1 z(1+z)(1+z^2) + \lambda_2 (-z)(1-z)(1+z^2) \\ & + \lambda_3 (-iz)(1-z)(1+z)(1-iz) + \lambda_4 (iz)(1-z)(1+z)(1+iz) \\ & + \varphi^*(z)(1-z^4). \end{aligned} \quad (8.10)$$

Pero se tiene que  $\lambda_3, \lambda_4 \in \mathbb{C}$ , y además,  $\lambda_4 = \bar{\lambda}_3$ . 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 (8.10) queda

$$\begin{aligned} \varphi(z) = & -\pi_1 z(1+z+z^2+z^3) - \pi_2 (-z)(1-z+z^2-z^3) \\ & -(\pi_4 + \pi_3 z)(-z)(1-z^2) + \varphi^*(z)(1-z^4). \end{aligned} \quad (8.11)$$

Finalmente, a partir de  $\varphi(L)Y_t = \varepsilon_t$ , se reemplaza  $\varphi(L)$  por la correspondiente expresión obtenida al reemplazar la expresión para  $\varphi(z)$  en (8.11). Si se definen las variables

$$\begin{aligned} X_{1,t} &= (1 + L + L^2 + L^3)Y_t, \\ X_{2,t} &= -(1 - L + L^2 - L^3)Y_t, \\ X_{3,t} &= -(1 - L^2)Y_t, \\ X_{4,t} &= (1 - L^4)Y_t, \end{aligned}$$

se obtiene la ecuación

$$\varphi^*(L)X_{4,t} = \pi_1 X_{1,t} + \pi_2 X_{2,t} + \pi_3 X_{3,t} + \varepsilon_t. \quad (8.12)$$

Colocando  $\varphi^*(L) = 1$ , la ecuación (8.12) 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, una para cada raíz unitaria estacional:  $-1, \pm i$ . Nótese que la raíz  $z = 1$  corresponde a una raíz unitaria para tendencia aleatoria.

$$\begin{aligned} H_0 : \varphi(1) = 0 &\Leftrightarrow \pi_1 = 0, & H_a : \varphi(1) > 0 &\Leftrightarrow \pi_1 < 0, \\ H_0 : \varphi(-1) = 0 &\Leftrightarrow \pi_2 = 0, & H_a : \varphi(2) > 0 &\Leftrightarrow \pi_2 < 0, \\ H_0 : |\varphi(i)| = 0 &\Leftrightarrow \pi_3 = \pi_4 = 0, & H_a : no(H_0). \end{aligned}$$

En el caso  $s = 12$ , hay 7 hipótesis, dos para  $z = 1, -1$ , y una para cada pareja de raíces unitarias estacionales conjugadas.

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.

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 se refiere a la probabilidad:  $\text{Prob}(\text{rechazar } H_0 \mid H_a \text{ es cierta})$ , 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 serie 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 cierto ARMA con órdenes altos, ó también como un modelo de descomposición, como se vió al comienzo del curso.

## Implementación en R

La prueba HEGY está implementada en la función `HEGY.test`, de la librería `uroot`, (<sup>1</sup>). Al colocar `res = HEGY.test(y, itsd, regvar, selectlags)`, la serie debe estar en la variable `y`, y debe haberse declarado como objeto “ts”. El parámetro `itsd` se refiere a si se desea incluir en el modelo un intercepto, tendencia lineal y componente estacional con

---

<sup>1</sup>La librería `uroot` está en <http://r-forge.r-project.org/bin/windows/contrib/latest/>, en un archivo zip con el cual se instala directamente

variables indicadoras. El modelo se define en el caso afirmativo como

$$(1 - L^4)(Y_t - (\beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t))) = \varepsilon_t. \quad (8.13)$$

Colocando `itsd=c(1,1,c(1,2))` se especifica este modelo. En caso de que no se desee incluir estas variables se coloca `itsd=c(0,0,c(0))`. El parámetro `regvar` se coloca `regvar=0` si no se incluyen variables exógenas en el modelo. En caso contrario se incluye una matriz con los valores de las variables. Por ejemplo, se podría incluir el primer rezago de  $y$ ,  $Y_{t-1}$ . El parámetro `selectlags` se coloca por defecto con los valores `selectlags = list(mode = "bic", Pmax = 12)`.

**Ejemplo 8.2.1.** Retomando el Ejemplo (5.1.1) en la pag. 80, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t) + \varepsilon_t, \quad (8.14)$$

$$\varepsilon_t \sim AR(7). \quad (8.15)$$

En la Figura 8.3 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 en lugar del modelo de componentes (8.14).

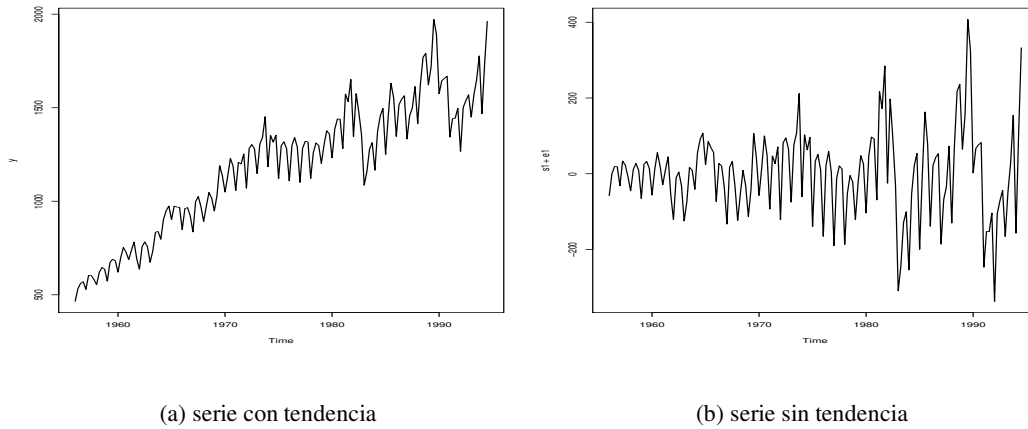


Figura 8.3: Serie de Producción Trimestral de Cemento, Australia

Con el fin establecer cuál modelo es más adecuado se procede a realizar la prueba HEGY. Los comandos en R son

```
# prueba hegy
res = HEGY.test(wts=y,
itsd=c(0,0,c(0)),
regvar=0, selectlags=list(mode="signf", Pmax=NULL))
res
-----resultados
Null hypothesis: Unit root.
Alternative hypothesis: Stationarity.
-----
      HEGY statistics:
              Stat. p-value
tpi_1      3.245      0.1
tpi_2     -1.020      0.1
Fpi_3:4     2.254      0.1
Fpi_2:4     1.854      NA
Fpi_1:4     4.101      NA
```

*El resultado anterior muestra que no 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, es apropiado ajustar un modelo SARIMA a la serie. Al examinar la fac de la serie diferenciada  $w_t = (1 - L)(1 - L^4)Y_t$ , 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 8.4.*

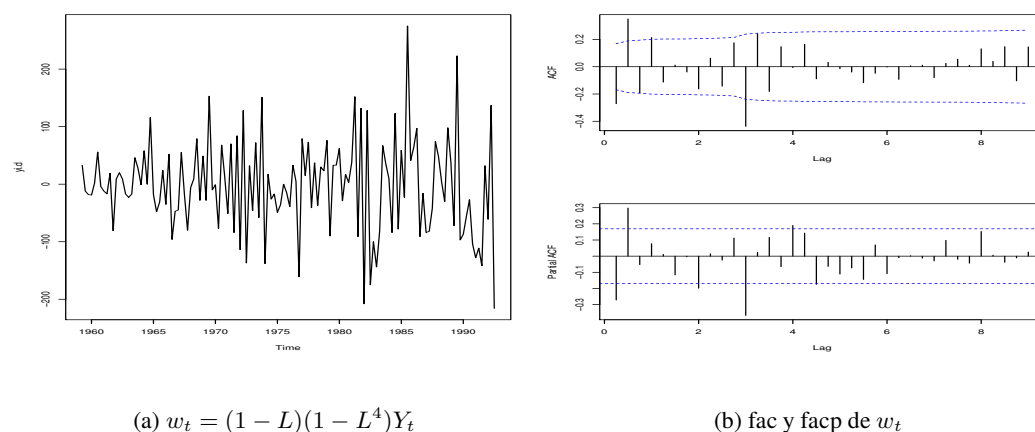


Figura 8.4: Serie Diferenciada de Producción Trimestral de Cemento

*Mediante ensayo y error, empezando con un modelo SARIMA(1, 1, 1)(1, 1, 1)[4], hasta conseguir residuos ruido blanco se llega al modelo SARIMA(3, 1, 2)(1, 1, 2)[4].*

Tabla 8.1: 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  $X^2 = 34.672$ ,  $df = 26$ ,  $p\text{-value} = 0.1189$ , por lo que puede aceptarse este modelo para pronosticar. Los pronósticos son el modelo SARIMA versus los obtenidos con el modelo componentes (8.14) se pueden observar en la Figura 8.5.*



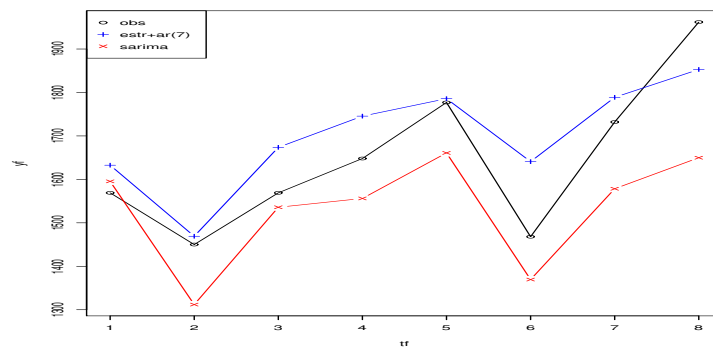


Figura 8.5: Pronósticos a 8 trimestres de la Producción Trimestral de Cemento

Como conclusión, se llega a que el SARIMA se detecta como modelo correcto por la prueba HEGY. Pero el MAPE de sus pronósticos, a 8 trimestres, es 7.89, mayor que el del modelo de Descomposición con errores AR(7), que es 4.64. Sin embargo, el modelo SARIMA tiene mejor ajuste dentro de la muestra. En la Figura 8.6 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.

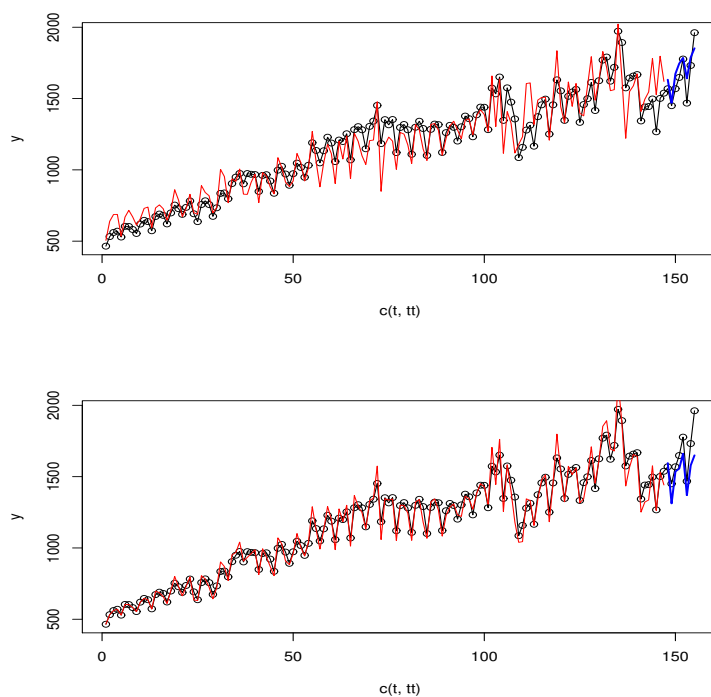


Figura 8.6: Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA

### 8.2.2. 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 idea. 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.

## Descripción de la Prueba

El modelo que asume la prueba es un modelo sin tendencia y con componente estacional, descrita mediante funciones trigonométricas, 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

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \sin\left(\frac{2\pi jt}{s}\right), \quad (8.16)$$

donde se asume  $s$  es par,  $t = 1, 2, \dots$ . Nótese que cuando  $j = s/2$  entonces  $\sin\left(\frac{2\pi jt}{s}\right) = \sin(\pi t) \equiv 0$ . Luego, el número de parámetros en el modelo (8.16) es  $s - 1$ , el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores  $\underline{\gamma} = (\gamma_1, \dots, \gamma_{s/2})'$ ,  $\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'$ .  $\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'$

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 un modelo de marcha aleatoria,  $\underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t$ ,  $\underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t$ , 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 un parámetro  $\tau^2$ , tal que  $H_0 : \tau^2 = 0$ , versus  $H_a : \tau^2 > 0$ . El estadístico de la prueba  $L$ , requiere resultados avanzados para su definición, por lo que se remite al artículo original. Además en el artículo se proveen los valores críticos para varios niveles de significación.

## Implementación en R

La prueba Canova-Hansen está implementada en en la función `CH.test`, que tiene los parámetros siguientes. Al colocar `res = CH.test(y, freq, f0, DetTr)`, la serie debe estar en la variable `y`, y debe haberse declarado como objeto “ts”. El parámetro `freq` se refiere cuáles frecuencias se prueban para detectar raíces unitarias. Nótese que si  $s = 4$  hay 2 frecuencias:  $-1, \pm i$ , por lo que se coloca `freq=c(1, 1)`. En el caso  $s = 12$  hay 6 frecuencias y se coloca `freq=c(1, 1, 1, 1, 1, 1)`. El parámetro `f0 = 0, 1` se coloca igual a 1 si la variable  $Y_{t-1}$  se incluye como variable explicativa en el modelo, y se coloca igual a

0 si no se incluye. El parámetro DetTr es una variable lógica con valores TRUE, FALSE. La opción TRUE indica que se incluye en el modelo una tendencia lineal. Nótese, sin embargo, que para aplicar la prueba Canova and Hansen [1995] 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  $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$ , donde  $\hat{T}_t$  se estimó mediante el filtro `stl()`. En consecuencia colocamos DetTr=FALSE.

**Ejemplo 8.2.2.** *Continuando el Ejemplo 8.2.1, para decidir entre los modelos de Descomposición con errores AR(7) y el SARIMA*

$$1) Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t, \varepsilon_t \sim AR(7),$$

$$2) Y_t = SARIMA(3, 1, 2)(1, 1, 2)[4].$$

*La prueba se implementa con los comandos*

```
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
ch.out1 = CH.test(wts=y1, freq=c(1,1,1), f0=0, DetTr=FALSE)
ch.out1
-----
Y los resultados son

Canova & Hansen test
-----

Null hypothesis: Stationarity.
Alternative hypothesis: Unit root.
Frequency of the tested cycles: pi/2 , pi , NA ,

L-statistic: 2.274

Critical values:
0.10 0.05 0.025 0.01
1.28 1.47 1.63 1.88
```

*Como el estadístico observado es 2.274, mayor que el valor crítico al nivel de 5 %, se rechaza la hipótesis nula y se detecta una raíz unitaria estacional, que corresponde a inestabilidad estructural en la componente estacional. Por tanto, el modelo SARIMA es preferible para modelar la serie.*



## 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, 1997.916666666667, 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, 1997.916666666667, 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, 1997.916666666667, 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, 1997.916666666667, 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, 1997.916666666667, 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 of Economic and Environmental Problems,” *JASA*, 70, 70–79.
- 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.
- 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.
- 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.
- HYLLEBERG, S., R. ENGLE, C. GRANGER, AND B. YOO (1990): “Seasonal integration and cointegration,” *Journal of Econometrics*, 44, 215–238.

- HYNDMAN, R. J., A. B. KOEHLER, J. K. ORD, AND R. D. SNYDER (2008): *Forecasting with Exponential Smoothing. The State Space Approach*, Berlín, Heidelberg: Springer Science + Business Media LLC.
- POVEDA, G. (1982): “Curva Logística y Mínimos Cuadrados,” *Lecturas Matemáticas*, 3, 111–132.
- 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.
- WINTERS, P. (1960): “Forecasting sales by exponentially weighted moving averages,” *Management Science*, 6, 324–342.

---

## Índice alfabético

---

$AR(1)$ , 81  
 $CUSUM$ , 58  
 $R^2$ , 13  
  
AIC, 14  
Algoritmo de Brockwell y Davis, 64  
  
Banda de Bartlett, 77  
bIC, 14  
  
Componente Estacional, 19, 47  
Componente Estructural, 50  
Consistencia, 15  
Criterio de Información  
    de Akaike, 14  
    de Schwarz, 14  
  
Data  
    Unemployed2, 29  
Datos de Series de Tiempo, 153  
  
Ecuación Yule–Walker, 100  
Eficiencia Asintótica, 15  
Error cuadrático medio, 13  
Estabilidad Estructural, 56  
Estacionario en Covarianza, 73  
Estadístico  
    Ljung-Box, 80

Estadístico  
    Durbin-Watson, 82  
Estimación  
    de la Autocorrelación, 76  
    de la Autocovarianza, 76  
Estimador de la Varianza del Error, 13  
Estimadores Recursivos, 57  
EWMA, 35  
  
FAC, 74  
Filtro  
    causal, 35  
Filtro Lineal, 34  
Función  
    de Autocovarianza, 73  
Función  
    de Autocorrelación, 74  
    de Autocorrelación Parcial, 94  
Función de R  
    Box.test, 81  
función de R  
    dwtext, 83  
  
librería  
    dynlm, 86  
    lmtest, 83  
  
Método de Descomposición STL, 34

- Método de Descomposición Stl, 34
- Método de Suavizamiento
  - de Holt, 45
  - de Holt-Winters, 67
  - Exponencial Simple, 35
- Marcha Aleatoria, 104
- Matriz Toeplitz, 110
- Media Móvil, 35
  - Bilateral, 35
  - Unilateral, 35
- Modelo
  - aditivo de componentes, 19
  - cúbico, 21
  - cuadrático, 21
  - exponencial, 21
  - lineal, 21
  - log-Lineal, 21
  - logístico, 21
- Modelos SARIMA, 135
- Operador de Rezago, 92
- Proceso
  - $ARIMA(p, d, q)$ , 120
  - SARIMA, 135
- Proceso
  - $AR(1)$ , 81
  - $AR(p)$ , 98
  - $ARMA(p, q)$ , 105
  - estacionario estacional, 140
  - integrado, 120
  - $MA(q)$ , 92
- Proceso Invertible, 95
- Pronóstico
  - de densidad, 24
  - por intervalo, 24
- Pronósticos
  - con base en la tendencia, 24
  - definición, 24
- Propiedades
  - de procesos  $MA(q)$ , 92
- Prueba
  - Dickey-Fuller, 127
  - Aumentada, 128
- Prueba
  - Diebold-Mariano, 125
  - Durbin-Watson, 81
  - Ljung-Box, 79
- Pruebas
  - de Raíz Unitaria Estacional, 140
- Pruebas de Incorrelación, 79
- R cuadrado, 13
- R cuadrado ajustado, 13
- Raíz Unitaria, 118
  - Autoregresiva, 118
- Recursivos
  - Estimadores, 57
- Regresión Loess, 32
- Residuales
  - Recursivos, 58
- Residuales Recursivos Cuadrados CUSUMSQ, 59
- Residuos
  - estructurales, 50
- Ruido Blanco, 72
  - Gaussiano, 72
- Supercosistencia, 126
- Tendencia, 20
- Tendencia Estocástica, 120
- Teorema
  - de Bartlett, 77
- validación cruzada, 25
- Variable
  - Indicadora Estacional, 48