
Técnicas de Pronósticos

Aplicaciones con R

TÉCNICAS DE PRONÓSTICOS

Aplicaciones con R

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

Universidad Nacional de Colombia
Medellín

Copyright ©2006 Norman Diego Giraldo Gómez

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

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

p. cm.—(Colección Notas de Clase)

“Universidad Nacional de Colombia.”

Incluye referencias bibliográficas e índice.

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

1. Probabilidades—Teoría. 2. Matemáticas

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

519.2

G887c

Diagramación en LaTeX.

Impresión

Editorial

Índice general

1. Introducción al Curso	1
1.1. Objetivos del Curso	2
1.2. Metodología, Requisitos y Evaluación	3
1.3. Por qué Pronósticos?	4
1.4. La estimación de la tendencia	5
1.5. Ejemplos de Pronósticos	6
1.6. Pronósticos y aprendizaje de máquinas	7
1.7. Programa del Curso: Preparación	8
1.8. Programa del Curso: Parte I	8
1.9. Programa del Curso: Parte II	8
1.10. Programa del Curso: Parte III	9
1.11. Instalar R	10
1.12. “Task View: Time Series Analysis”	11
1.13. Bibliografía en Moodle	11

2. Repaso del Lenguaje R	13
2.1. Características de R	13
2.1.1. Instalación de R	14
2.1.2. Ayudas de R	14
2.1.3. Instalación librerías de R	14
2.1.4. Algunos manuales de R	16
2.1.5. Algunos manuales Internet	16
2.2. Estructuras de datos	17
2.2.1. Objeto vectores	17
2.2.2. Objeto data.frame	18
2.2.3. Objeto list	18
2.3. Lectura de datos en archivos	19
2.3.1. La función read.table	19
2.3.2. La función ts	20
2.3.3. Funciones de lectura de archivos Excel	23
2.3.4. Series con datos faltantes	24
2.3.5. Funciones para escritura de datos	26
2.4. Repaso del Modelo de Regresión Lineal	27
2.4.1. Estimación de los Parámetros	30
2.4.2. Mínimos cuadrados ordinarios	30
2.4.3. Pronósticos	31
2.4.4. Error cuadrático medio, MSE	32
2.4.5. Residuos	32
2.4.6. Mínimos Cuadrados Nolineales	36
2.4.7. Regresión Local Polinómica	37

2.5. Pruebas de Ajuste del Modelo	38
2.5.1. Estadísticos de Ajuste y de Selección de Modelos	42
2.5.2. Pruebas F parciales múltiples	47
3. Modelo Aditivo de Componentes	49
3.1. El Modelo Aditivo de Componentes de series de tiempo	49
3.2. Modelos para la Tendencia	51
3.3. Estimación de la Tendencia	53
3.4. Pronósticos con base en la Tendencia	54
3.5. Medidas de precisión en los pronósticos	56
3.5.1. Caso de Estudio: Pronóstico de Ventas al Menudeo	58
3.6. Modelos para la Componente Estacional	63
3.7. Modelo de componentes aditivas	67
3.8. Problemas	73
4. Pruebas de Estabilidad Estructural	77
4.1. Pruebas de Estabilidad Estructural en Modelos de Componentes	77
4.2. Estimación recursiva de parámetros	79
4.3. Pruebas CUSUM	80
4.4. Implementación en R: la librería strucchange	83
4.5. Identificación del punto de quiebre	88
5. Suavizadores, Espacio de Estados y Redes neuronales	91
5.1. Regresión Local Loess	91
5.2. Método híbrido Loess más componente estacional	95
5.3. Método de descomposición STL	96

5.4. Modelo de Holt-Winters	99
5.5. Modelos de Espacio de Estados	102
5.6. Redes neuronales autoregresivas	105
6. Medias Móviles y Filtros	119
6.1. Medias Móviles	119
6.2. Método de descomposición con base en Medias móviles	127
6.3. Librerías para filtrado de series en R	129
6.4. Problemas	130
7. Pruebas de Incorrelación	131
7.1. Introducción	131
7.2. Series Estacionarias en Covarianza	133
7.3. Estimación de las funciones de Autocovarianza y Autocorrelación. .	136
7.4. Pruebas de Incorrelación	142
7.4.1. Prueba Ljung-Box (LB)	143
7.4.2. Prueba Durbin-Watson (DW)	147
7.4.3. Prueba Durbin-Watson Generalizada	152
7.5. Alternativas cuando los Residuos Estructurales muestran Autocorrelación	154
7.5.1. Incluir rezagos de la variable dependiente en el modelo . . .	154
7.5.2. Transformación Cochran-Orcutt	158
7.5.3. Pruebas de Heterocedasticidad	158
8. Modelos ARMA para la Componente Aleatoria	161
8.1. Introducción	161
8.2. Procesos de Medias Móviles de orden q , $MA(q)$	162

8.2.1.	Propiedades de los procesos MA(q)	163
8.2.2.	Condición de Invertibilidad del Proceso MA(q)	165
8.2.3.	Función fac parcial de un Proceso MA(q) invertible	166
8.2.4.	Estimación de procesos MA(q)	168
8.3.	Procesos Autorregresivos de Orden p, AR(p)	172
8.3.1.	Condición Suficiente para que un AR(p) sea Estacionario en Covarianza	173
8.3.2.	La Función de Autocovarianza de un Proceso AR(p)	175
8.3.3.	Función fac parcial de un Proceso AR(p)	178
8.3.4.	Estimación de procesos AR(p)	179
8.3.5.	Simulación de procesos AR(p)	181
8.3.6.	Análisis de un caso de aplicación de AR(p)	181
8.4.	Procesos Autoregresivos y de Medias Móviles ARMA(p,q)	185
8.4.1.	Procesos ARMA y la representación de Wold	187
8.4.2.	Simulación de modelos ARMA	189
8.4.3.	La función de autocovarianza de un proceso ARMA	190
8.4.4.	Identificación de modelos ARMA	191
8.4.5.	Estimación de modelos ARMA	193
8.5.	Modelos ARMA estacionales, SARMA	197
8.5.1.	La simulación de los modelos SARMA	201
8.5.2.	La identificación de los modelos SARMA	202
8.5.3.	La estimación de los modelos SARMA	204
9.	Raíces Unitarias y Tendencias Estocásticas (Modelos ARIMA)	211
9.1.	Introducción	211
9.2.	Modelos ARIMA	212

9.3. Modelos de Marcha Aleatoria	213
9.4. Ejemplos de modelos ARIMA	217
9.5. Pruebas de Hipótesis para Raíces Unitarias	223
9.6. Prueba Dickey-Fuller	225
9.7. Valores críticos de la prueba Dickey-Fuller	226
9.8. Prueba Dickey-Fuller Aumentada	228
9.9. La prueba KPSS	234
9.10. La prueba PP	235
9.11. Pruebas de No linealidad en Series de Tiempo	237
9.12. Modelos híbridos ARIMA - Red Neuronal	239
9.13. Combinación de pronósticos	241
10. Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)	243
10.1. Modelos SARIMA	243
10.2. Pruebas de Raíz Unitaria Estacional	252
10.3. La prueba Canova-Hansen	252
10.4. La prueba HEGY	257
10.5. Prueba OCSB	262
10.6. Ejemplo de resumen	264
11. Modelos de Rezagos Distribuidos y Cointegración	269
11.1. Modelo de rezagos distribuidos	269
11.2. *Cointegración	275
A. Datos de Series de Tiempo	277
A.1. Series con Tendencia y Estacionalidad	277

A.1.1. Ejemplo 1. Cinco Series	277
A.1.2. Ejemplo 2. Consumo domiciliario agua potable en Medellín	279

Índice de figuras

1.1. potencia de energía: entrada agua en metros cúbicos diarios en embalses	4
1.2. logprecio kwh bolsa energía colombia	4
2.1. Producción diaria	21
2.2. Gráfica de una serie mensual	23
2.3. Gráfica producción mensual de crudo para países del Golfo y para los no pertenecientes a la Opec (en rojo)	25
2.4. Serie con datos faltantes versus serie con imputación mediante splines	27
2.5. gráficas A: y versus x_1 , B: residuos r , C: y versus \hat{y} , D: histograma residuos r	34
3.1. Descomposición de la Serie Y_t	50
3.2. Tasa de Empleo, Australia, 02/1978 - 09/2014	52
3.3. Logaritmo del precio del kwh en la bolsa energía Colombia	53
3.4. Ventas al menudeo, del texto Diebold(1999)	59

3.5. Pronósticos (—) versus Observados (-o-o-)	62
3.6. Average Monthly Temperatures at Nottingham, 1920-1939	65
3.7. Ajuste con Variables Indicadoras y tendencia lineal. La componente estacional parece cambiar con el tiempo, es posiblemente un cambio estructural.	70
3.8. Ajuste Logístico a Población de Medellín	75
4.1. Estimadores Recursivos para el Modelo de Producción de Cemento	81
4.2. Región de Rechazo de la Prueba <i>CUSUM</i>	82
4.3. Resultados prueba <i>CUSUM</i> gráfica	85
4.4. Resultados prueba <i>MOSUM</i> gráfica	85
4.5. Punto de Quiebre 1974-Q4 en la Serie y en los Residuos	89
5.1. Suavizamiento Loess de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06	93
5.2. Descomposición <i>STL</i> número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06	97
5.3. Pronóstico modelo híbrido Loess + Indicadoreas, serie de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06	98
5.4. Estimación de la tendencia con los Métodos Loess, <i>STL</i> y Holt-Winters	102
5.5. Ajuste con <i>NNAR(3,1,2)[12]</i> , serie de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06	107
6.1. Precios de la Acción de Microsoft Suavizados con <i>SES</i>	122
7.1. Ejemplo Función de Autocovarianza	134
7.2. Ejemplo Función de Autocorrelación.	136

7.3. Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario	137
7.4. Ejemplos de Fac y Bandas de Bartlett	140
7.5. $\log(\text{igbc})$ y $\text{diff}(\log(\text{igbc}))$, a izquierda, con sus fac muestrales, a derecha	140
7.6. Fac de los Residuos Estructurales Serie Cementos	141
7.7. Densidad χ_m^2	144
7.8. (A): σ_t , (B): Y_t , (C): Z_t , (D): fac de Y_t	146
7.9. Compara fac de serie no estacionaria versus estacionaria.	147
7.10. Fac teóricas de un $AR(1)$	148
7.11. La fac de residuos $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$	151
7.12. Y estimada versus residuales estructurales	159
8.1. Fac teóricas de modelos $MA(q)$	164
8.2. fac y facp de un $MA(3)$ invertible.	167
8.3. Residuos estructurales de la serie de producción de cemento	170
8.4. (a) Pronosticos Estructurales versus Estructurales + $MA(6)$, (b) Raíces y círculo unitario	170
8.5. Círculo Unitario	174
8.6. fac de $ar(2)$	178
8.7. Aportes hidrológicos en Gw/h, 01/07/98 - 31/05/14	182
8.8. fac y facp de Aportes hidrológicos	184
8.9. fac y facp de residuos de $AR(30)$	185
8.10. Fac Muestral de $ARMA(p, q)$	191
8.11. Posibles submodelos $(1,1)=AR$, $(1,2)$, $(2,1)=ARMA$, $(2,2)=SARMA$	203
8.12. Serie de llamadas diarias a un Call Center.	205

8.13. Residuos estructurales para la serie de llamadas diarias a un Call Center. Posible modelo S-ARMA	206
8.14. Pronósticos estructurales (azul) versus pronósticos estructurales incorporando pronósticos con S-ARMA(5,3)(2,0)[7], para la serie de llamadas diarias a un Call Center.	208
8.15. Pronósticos con tres modelos para producción de cemento	209
9.1. Trayectorias de Marchas Aleatorias	214
9.2. Pronósticos de la Serie PNB-USA.	217
9.3. (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada	218
9.4. Pronósticos usd/yen con ARIMA(3,1,2) (continua) y Tendencia Lineal+AR(2) (punteada)	219
9.5. Serie USD por Libra	220
9.6. fac y facp muestrales de proceso con raíz unitaria	223
9.7. fac y facp muestrales de la serie USD/Pound	231
9.8. Pronósticos con modelo híbrido ARIMA-NNAR de la serie USD/-Pound	240
9.9. combinacion de pronósticos con 6 modelos para la serie USD/Pound	242
10.1. Aplicación de diferencia simple y luego diferencia estacional a la serie de producción de cemento	246
10.2. (A) Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12], del Ejemplo 10.1.2, con (B) diferencia ordinaria, (C) diferencia estacional, (D) ambas diferencias	248
10.3. Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12], serie doblemente diferenciada y su acf y pacf. Nótese que las autocorrelaciones en los rezagos 1,12,13 son significativas, con el signo correcto	251
10.4. Raices de la unidad, $z^{12} = 1$	259

10.5. Serie de Producción Trimestral de Cemento, Australia	265
10.6. Serie Diferenciada de Producción Trimestral de Cemento	266
10.7. Pronósticos a 8 trimestres de la Producción Trimestral de Cemento .	267
10.8. Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA	267

Índice de cuadros

2.1. Tabla de decisiones en una prueba de hipótesis	39
2.2. Resultados estimación MCO modelo m1	45
2.3. Resultados estimación MCO modelo m2	45
2.4. Resultados estimación MCO modelo m3	45
2.5. Resultados los estadísticos de ajuste y selección	46
3.1. Ajuste de los Modelos Lineal, Cuadrático, Cúbico y Exponencial . .	61
3.2. Estadísticos de Selección de Modelos	61
3.3. Estadísticos de calidad de pronósticos para tres modelos	63
3.4. Resultados del Ajuste del Modelo con Variables Estacionales	70
3.5. Coeficientes estimados en el Modelo Lineal con variables trigonométri- cas	72
3.6. Seis primeros valores ajustados con variables Indicadoras y con Variables Trigonómicas	72
3.7. Población de Medellín censos 1905 - 2005	74

5.1. Calidad de pronósticos para los modelos Loess+Estacional versus STL	98
5.2. Medidas de calidad de ajuste, tres modelos estructurales	104
5.3. Medidas de calidad de pronósticos, tres modelos estructurales	104
5.4. Modelos para Pronosticar	110
5.5. Asignación de modelos y serie por grupo	111
7.1. Resultados de la Prueba DW Generalizada	154
7.2. Resultados del Modelo estructural con dos rezagos 1,2 de Y_t	156
7.3. Resultados Prueba DWG para el modelo con rezagos 1,2	157
8.1. Estimadores OLS, Burg, Y-W, MaxVer en un AR(2)	180
8.2. Coeficientes estimados del modelo SARMA(1,0,2)(3,0,1)[4]	210
8.3. Comparación de medidas de calidad de pronósticos	210
9.1. Parámetros del modelo ARIMA(3,1,2)	219
9.2. Valores críticos de la prueba Dickey-Fuller	227
9.3. Ejemplo de resultado de la prueba Dickey-Fuller, con $n = 50$ obser- vaciones	227
9.4. Tabla de decisiones en una prueba de hipótesis	227
10.1. Parámetros estimados de SARIMA(3,1,2)(1,1,2)[4]	266
11.1. Modelos de Transferencia	272

Índice de códigos en R

2.1. Ejemplo de uso de read.table con datos en una dirección URL . . .	20
2.2. Ejemplo de uso de la función structure para lectura de datos	21
2.3. Ejemplo de lectura de datos desde archivos Excel	23
2.4. Ejemplo de imputacion de datos faltantes en una serie de tiempo . .	24
2.5. Ejemplo de regresión lineal con R	33

CAPÍTULO 1

Introducción al Curso

Índice

1.1. Objetivos del Curso	2
1.2. Metodología, Requisitos y Evaluación	3
1.3. Por qué Pronósticos?	4
1.4. La estimación de la tendencia	5
1.5. Ejemplos de Pronósticos	6
1.6. Pronósticos y aprendizaje de máquinas	7
1.7. Programa del Curso: Preparación	8
1.8. Programa del Curso: Parte I	8
1.9. Programa del Curso: Parte II	8
1.10. Programa del Curso: Parte III	9
1.11. Instalar R	10
1.12. “Task View: Time Series Analysis”	11
1.13. Bibliografía en Moodle	11

1.1. Objetivos del Curso

1. El objetivo es desarrollar modelos estadísticos de series de tiempo que permitan el cálculo de pronósticos.
2. Introducir el manejo de software R para implementarlos.
3. El curso se apoya en el texto de Notas de Clase, que desarrolla los temas siguientes, de manera articulada. Las Notas de Clase siguieron inicialmente el texto de Diebold [1999], complementándolo en el software R.
 - a) Modelo básico de componentes
 - b) Espacio de estados, redes neuronales
 - c) Modelos ARMA para los residuos del modelo de componentes
 - d) Modelos integrados ARIMA-SARIMA
 - e) Comparaciones de pronósticos
4. El cálculo de pronósticos con énfasis en ejemplos de series reales.

1.2. Metodología, Requisitos y Evaluación

1. Metodología: clases presenciales y/o remotas, de teoría y talleres de práctica con R. Material complementario en Moodle (temas de trabajos, datos, documentos de referencia).
2. Texto guía: Giraldo,N.(2006) Técnicas de Pronósticos con R.
<https://repositorio.unal.edu.co/handle/unal/76272> (no actualizado)
3. Las exposiciones se harán siguiendo las Notas de Clase, con posibles modificaciones de éstas en el transcurso del semestre. Y referencias a textos de consulta en Moodle (incluyendo Wikipedia)
4. Pre-requisitos:
 - a) Inferencia estadística: pruebas de hipótesis, estimación de parámetros.
 - b) Análisis de regresión múltiple.
 - c) Programación básica en lenguaje R.
5. Bibliografía: textos que desarrollan la teoría de series de tiempo utilizando el software R, por ejemplo, Cowpertwait and Metcalfe [2009], Shumway and Stoffer [2005], Cryer and Chan [2008], Aragon [2008], Hyndman et al. [2008].
6. Evaluación: 3 trabajos, de igual valor, en grupos, con base en las prácticas en R. Los temas son tipo proyecto, lo que significa que los trabajos 1,2,3 son tres fases de un análisis que tiene como objetivo comparar los pronósticos de las metodologías introducidas.

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

1.3. Por qué Pronósticos?

Los pronósticos se utilizan en forma constante en diversos campos: economía, finanzas, mercadeo, medio ambiente, ingeniería, salud. Su finalidad es proveer una guía para las decisiones que deban tomarse.

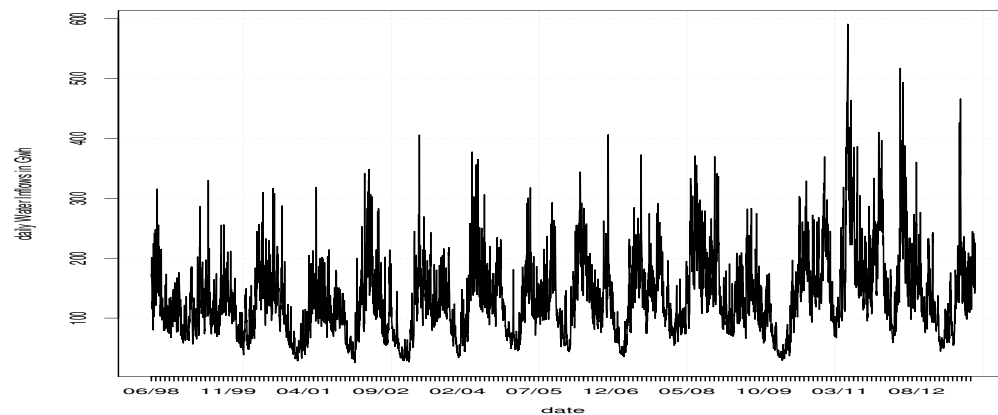


Figura 1.1: potencia de energía: entrada agua en metros cúbicos diarios en embalses

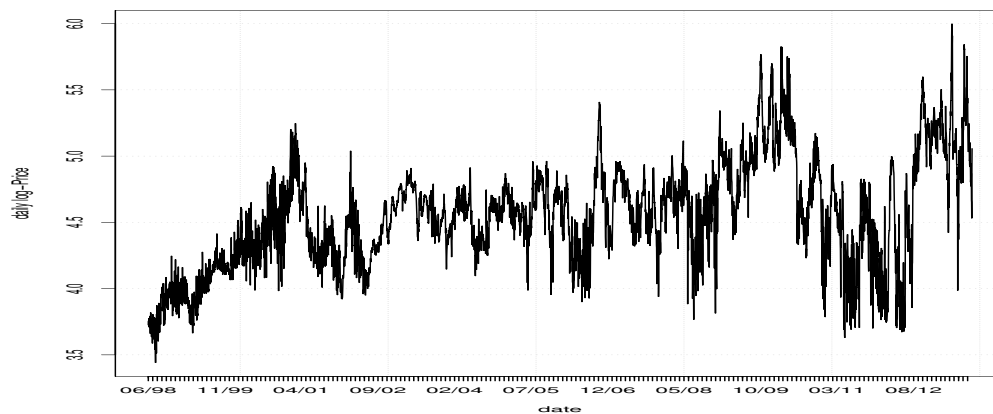


Figura 1.2: logprecio kwh bolsa energía colombia

1.4. La estimación de la tendencia

Por qué es útil estimar tendencias en el tiempo?. La estimación de la tendencia es básica para calcular pronósticos.

Pero también es útil

1. Para detectar cambios estructurales, p. ej. el efecto de una nueva ley ó regulación
2. Para establecer existencia de fenómenos que no se sospechan inicialmente ó son polémicos: Existe una burbuja especulativa en los precios de vivienda en Medellín?
3. Para establecer comparaciones sobre evolución de variables claves: desempleo, PIB,...
4. Para monitoreo de variables claves: prevalencia de enfermedades infecciosas,

1.5. Ejemplos de Pronósticos

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. Caso: Leonisa SA. Problema: pronósticos de ventas de diferentes modelos en el Catálogo.
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 como 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 en los diferentes POT.
6. 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.
7. En Planeación. El pronóstico de consumo de energía eléctrica domiciliaria es fundamental para las decisiones de generación a corto y mediano plazo.
8. En Turismo. El pronósticos de número de turistas mensuales para determinar la demanda hotelera.
9. 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.

1.6. Pronósticos y aprendizaje de máquinas

Por qué Pronósticos en series de tiempo aparece como un tema en Analítica?.

- AWS: “Amazon Forecast es un servicio de pronóstico de series temporales basado en el machine learning (ML) y creado para el análisis de las métricas.”

En el sitio web ⁽¹⁾: “Basado en la misma tecnología utilizada en Amazon.com, Forecast utiliza el aprendizaje automático para combinar datos de series temporales con variables adicionales para crear pronósticos.

Por ejemplo, la demanda de un color particular de una camisa puede cambiar con las estaciones y la ubicación de la tienda.

Una vez que proporcione sus datos, Forecast los examinará automáticamente, identificará lo que es significativo y producirá un modelo de pronóstico capaz de hacer predicciones que son hasta un 50 % más precisas que mirar solo datos de series temporales”.

- Prophet. En el sitio web ⁽²⁾: “Facebook Prophet es una herramienta desarrollada por Facebook para pronosticar objetos o datos de series temporales. Ayuda a las empresas a conocer el comportamiento de sus productos al pronosticar precios, ventas o el clima.

La herramienta Facebook Prophet se basa en un modelo de componentes, es decir, tendencia, estacionalidad y días festivos que ayuda a hacer modelos predictivos más precisos con estas restricciones. Es mucho mejor que el modelo ARIMA ya que ayuda a sintonizar y ajustar los parámetros de entrada.”

¹<https://docs.aws.amazon.com/es-es/managedservices/latest/userguide/forecast.html>

²<https://www.geeksforgeeks.org/time-series-analysis-using-facebook-prophet-in-r-programming/>

1.7. Programa del Curso: Preparación

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

1.8. Programa del Curso: Parte I

Se introduce el modelo de descomposición aditivo clásico, ó modelo de componentes latentes

$$Y_n = T_n + S_n + Z_n$$

Serie = Tendencia + Estacionalidad + Residuo aleatorio

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

1.9. Programa del Curso: Parte II

Se introduce el análisis del Residuo aleatorio Z_n

$$Y_n = T_n + S_n + Z_n$$

1. Cap 6: herramientas estadísticas para analizar Z_n

2. Cap 7: modelos para Z_n
3. Cap 8: ensamblar el modelo completo de componentes y pronosticar
4. Segundo Trabajo: (30 %)

1.10. Programa del Curso: Parte III

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

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

1.11. Instalar R

Qué es R? En Wikipedia (³) responden:

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

Cómo se obtiene R y se instala?.

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

³[https://es.wikipedia.org/wiki/R_\(lenguaje_de_programación\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programación))

1.12. “Task View: Time Series Analysis”

CRAN Task View: Time Series Analysis

Maintainer: Rob J Hyndman ⁽⁴⁾

- Forecasting and Univariate Modeling
- Frequency analysis
- Decomposition and Filtering
- Seasonality
- Stationarity, Unit Roots, and Cointegration
- Analysis of large groups of time series
- Non-linear time models (Neural Networks)
- Resampling

Sitios web

- <http://r-statistics.co/Time-Series-Analysis-With-R.html>
- <https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/>

1.13. Bibliografía en Moodle

- Giraldo, N. *Notas de Clase. Series de Tiempo con R*.
En <https://repositorio.unal.edu.co/handle/unal/76272>
- * Diebold, F. *Elementos de Pronósticos*. International Thomson Editores, 1999.
- Klaus Neusser *Time Series Analysis in Economics*.
www.neusser.ch/downloads/TimeSeriesBook.pdf

⁴<http://robjhyndman.com/hyndsight/>

- R. Adhikari and R. K. Agrawal *An Introductory Study on Time Series Modeling and Forecasting*.
<https://arxiv.org/abs/1302.6613>
- Ihaka, Ross *Time Series Analysis. Lecture Notes*.
<https://www.stat.auckland.ac.nz/~ihaka/726/notes.pdf>
- R.H. Shumway and D.S. Stoffer *Time Series Analysis and Applications Using the R Statistical Package*.
- R.H. Shumway and D.S. Stoffer: A gentle introduction to time series analysis is now available (for free). <http://www.stat.pitt.edu/stoffer/tsa4/tsaEZ.pdf>
- M. Dettling *Applied Time Series Analysis*
https://stat.ethz.ch/education/semesters/ss2013/atsa/ATSA-Scriptum-SS2013_130415.pdf
- Mahmood Arai *A Brief Guide to R for Beginners in Econometrics.*.
<homes.chass.utoronto.ca/~jmaheu/4050/Rintro.pdf>

CAPÍTULO 2

Repaso del Lenguaje R

2.1. Características de R

1. Es un software libre, de código abierto, para programación orientada a objetos tipo S3 y S4. Una ejemplo de la capacidad de la programación con objetos S4 está en las librerías `distr`, `distrEx` (mirar el documento de ayuda de éstas).
2. La información sobre R está en el sitio web <http://www.cran.r-project.org/>, incluyendo
3. Task View
4. R Journal
5. Manuals
6. Search (Google)
7. Packages
8. R Binaries

2.1.1. Instalación de R

1. Entrar a `http://www.cran.r-project.org/`.
2. Seleccionar: “R Binaries”
3. Seleccionar: “Windows”
4. Seleccionar: “base”
5. Seleccionar: “Download R 4.1.1 for Windows (86 megabytes, 32/64 bit)”.
6. Ejecutar el archivo .exe desde el directorio donde se descargó.

2.1.2. Ayudas de R

La ayuda (help) de R esta es una direccion URL interna que se abre con un navegador. Por ejemplo, para utilizar Chrome hay que colocar la dirección interna del ejecutable del navegador en la consola de R

```
options(browser="C:/Program Files (x86)/Google/Chrome/Application/chrome
```

luego, en la ventana de comandos se escribe, por ejemplo, `help("mean")`, ó el signo de interrogación y la palabra sobre la que se busca ayuda `?mean`.

En la ventana Windows de R en el vínculo “Ayuda”, abre una página html con un menú de ayudas. Es útil para explorar el contenido de las librerías.

2.1.3. Instalación librerías de R

R instala varias librerías por defecto, por ejemplo, la librería `stats`. El resto de las librerías se deben instalar desde la consola ó mediante comandos. Los comandos siguientes muestran el número de librerías disponibles a la fecha. Nota: entre 2020 y 2021 este número aumento un 10 %, aproximadamente.

```
nrow(available.packages())
--- Please select a CRAN mirror for use in this session
Sys.Date()
"2020-10-01"
```

```
16321
"2021-10-07"
17842
```

Instalación desde la consola de R.

- Dar Enter en Paquetes
- Luego Instalar paquete(s)
- Luego Secure Cran mirror: escoge una localización del cran, p.ej. Ecuador(Quito)
- Luego se escoge el **repositorio**. Por defecto es CRAN, pero existen alternos como R-Forge, BioC, Omegahat.
- A continuación aparece la lista de paquetes que se pueden instalar.

Algunas librerías para series de tiempo:

- `forecast`: ajuste, diagnósticos, pronósticos.
- `mFilter`: varios filtros para tendencia.
- `tseries`: varias funciones.
- `TSA`: funciones del texto
- `uroot`: pruebas de hipótesis
- `arima`: modelos arima-sarima.
- `nnet`: redes neuronales.

Instalación con comandos. En la librería `utils` (instalada por defecto) está el comando

```
install.packages("forecast", dependencies=TRUE)
```

En algunos casos una librería ya no está activa en el CRAN (deprecated), pero todavía se proporcionan los archivos de instalación en un archivo comprimido con extensión `.tar.gz`. En Google hay instrucciones para instalar una librería desde el archivo comprimido.

Algunas librerías pueden estar localizadas fuera del CRAN, por ejemplo en `github.com`.

Cuáles librerías utilizar para algún procedimiento se pueden encontrar ingresando palabras claves seguidas de la frase “with R”, en, por ejemplo, Google.

2.1.4. Algunos manuales de R

- Entrar a `http://www.cran.r-project.org/`.
- Seleccionar: “Contributed”,
- Seleccionar: “Manuals”

En Contributed:

- Seleccionar: “Non-English Documents”. Hay varias opciones.
- “R para Principiantes”, the Spanish version of “R for Beginners”, translated by Jorge A. Ahumada (PDF).
- A Spanish translation of “An Introduction to R” by Andrés González and Silvia González (PDF, Texinfo sources).
- “Metodos Estadisticos con R y R Commander” by Antonio Jose Saez Castillo (PDF, ZIP, 2010-07-08).

2.1.5. Algunos manuales Internet

dirección url: `https://rc2e.com/`: “Welcome to the R Cookbook 2nd Edition”, por James (JD) Long, and Paul Teetor, 2019-09-26.

dirección url: `https://stats.idre.ucla.edu/`: UCLA: Institute for digital research and education. Statistical consulting. (ver el vínculo para la página de R: “Class notes: Introduction to R”).

2.2. Estructuras de datos

2.2.1. Objeto vectores

Hay varios tipos de vectores: lógico, entero, doble (numérico), complejos y carácter. Todos se generan con el comando “c()”.

```
# Ejemplo de tipos de vectores
e = c(34, 35, 56, 78)
g = c("h", "m", "m", "m")
s = c(3.4, 4.5, 2.4, 12.3)
L = c(TRUE, FALSE, FALSE, NA)
r = complex(3)
r[1] = -0.8 - 1.3i
r[2] = Conj(r[1])
r[3] = 3
#-----
str(e)
num [1:4] 34 35 56 78
str(g)
chr [1:4] "h" "m" "m" "m"
str(s)
num [1:4] 3.4 4.5 2.4 12.3
str(L)
logi [1:4] TRUE FALSE FALSE NA
str(r)
cplx [1:3] -0.8-1.3i -0.8+1.3i 3+0i
#-----
e = as.integer(e)
str(e)
int [1:4] 34 35 56 78
#-----
a = as.numeric(L)
a
[1] 1 0 0 NA
```

```
str(a)
num [1:4] 1 0 0 NA
```

2.2.2. Objeto data.frame

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, pero deben tener el mismo número de filas. Las columnas pueden tener nombres. Un ejemplo simple es el siguiente.

Se define D como data.frame, y contiene las columnas edad, genero, salario. Entonces

```
e = c(34, 35, 56, 78)
g = c("h", "m", "m", "m")
s = c(3.4, 4.5, 2.4, 12.3)
D = data.frame(edad=e, genero=g, salario=s)
str(D)
'data.frame':  4 obs. of  3 variables:
 $ edad      : num  34 35 56 78
 $ genero    : Factor w/ 2 levels "h","m": 1 2 2 2
 $ salario   : num  3.4 4.5 2.4 12.3
es = D$edad/D$salario
es
[1] 10.000000  7.777778 23.333333  6.341463
```

2.2.3. Objeto 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 nuevamente. El siguiente es un ejemplo simple de objeto list

```
# ejemplo list
a = matrix(c(2, 3, 4, 5), 2, 2)
```



```

b = c("web", "real")
d = rnorm(120, 2, 3)
L = list(a=a, b=b, d=d)
L[[2]]
[1] "web" "real"
L[[2]][1]="complex"
L[[2]]
[1] "complex" "real"

```

2.3. Lectura de datos en archivos

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

2.3.1. 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 una alternativa 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.

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.

Código R 2.1: Ejemplo de uso de `read.table` con datos en una dirección URL

```
#-----
archivo = "http://www.medellin.unal.edu.co/~ndgiraldo/
Datos/Datos%20curso%20Series%20II/fechaydatos.prn"
G = read.table(archivo, header = TRUE,
stringsAsFactors=FALSE)
attach(G)
#-----grafica
np = length(x)
#-----convierte fecha a formato de R
fechas = as.Date(fecha, format="%d/%m/%y")
ejex.mes = seq(fechas[1], fechas[np], "months")
ejex.año = seq(fechas[1], fechas[np], "years")

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

2.3.2. La función `ts`

En esta sección consideramos una serie de tiempo T_t , $t = 1, 2, \dots, T$, como un vector columna. En R se ingresa inicialmente como un objeto tipo vector numérico.

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

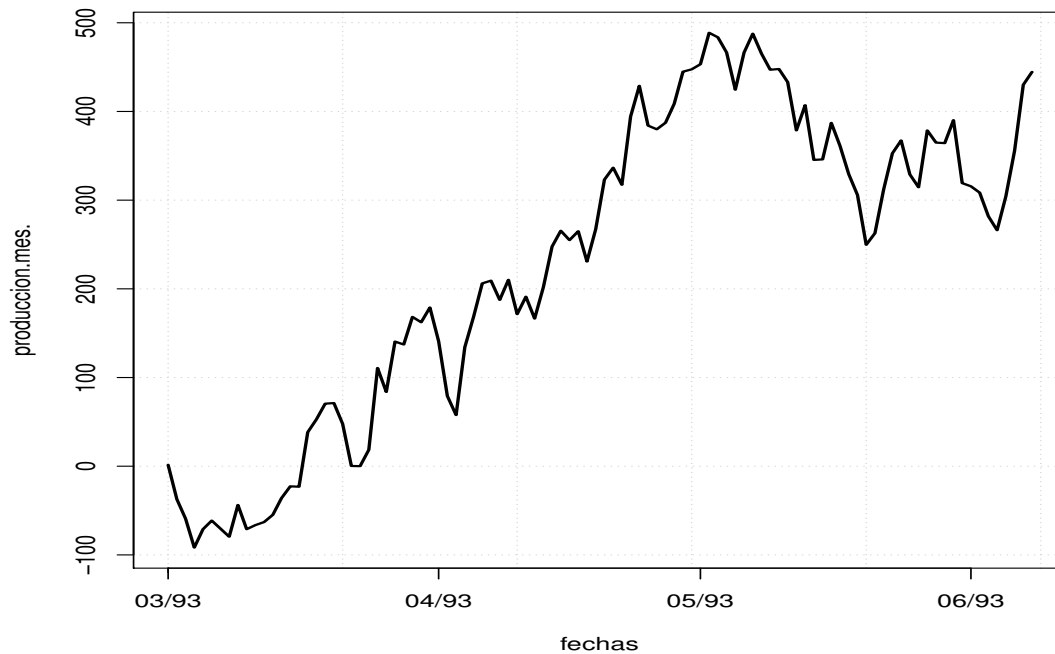


Figura 2.1: Producción diaria

Se puede ingresar los datos de la serie de tiempo directamente en el programa mediante la instrucción `structure()`.

Código R 2.2: Ejemplo de uso de la función `structure` para lectura de datos

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

# convertir a objeto ts
```

```

y=ts(y,frequency=12,start=c(1990,1))

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

# grafica con fechas
ts.plot(y,main="serie F")

# 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='b',
ylab='produccion.mes.')
axis.Date(1, at=ejex.mes, format="%m/%y")
axis.Date(1, at=ejex.año, labels = FALSE, tcl = -0.2)
# añadir una horizontal en el nivel 4000
abline(h = 4000, col='red')
# cual es la fecha en la cual se supera el nivel
# 4000 por primera vez?
j=1
while( y[j] < 4000 ){
j = j + 1}
# respuesta
fechas[j]
[1] "1994-11-01"
# añadir una vertical en esta fecha
points(fechas[j],y[j],type='h',col='blue')

```

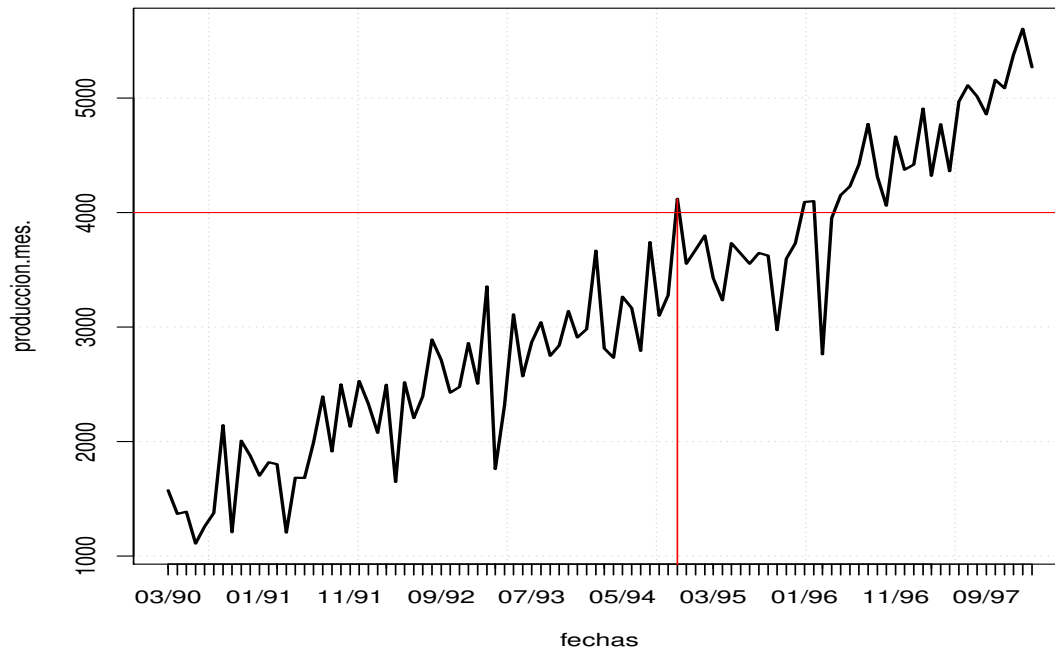


Figura 2.2: Gráfica de una serie mensual

2.3.3. Funciones de lectura de archivos Excel

Las funciones de lectura de archivos Excel son: `G = read.csv("nombre", header = TRUE)`, para archivos excel “delimitados por comas”, de extensión .csv.

Los archivos con extensión “.xls”, “.xlsx” se pueden leer con la función `read_excel()` de la librería `readxl`.

Código R 2.3: Ejemplo de lectura de datos desde archivos Excel

```
#-----
install.packages("readxl")
library(readxl)
res <- read_excel("international-petroleum-world-cr.xlsx")
attach(res)

# El comando plot.ts(D) grafica varias series
```

```

#(hasta 10 máximo), que deben estar en una matriz ó
# data.frame, por ejemplo, D.

str(res)
Classes 'tbl_df', 'tbl' and 'data.frame':      486 obs. of
5 variables:
 $ fecha : POSIXct, format: "1973-01-01" ...
 $ mexico: num  453 458 458 463 468 468 463 463 463 473 ...
 $ golfo : num  20051 20335 20332 20241 21369 ...
 $ noopec: num  25378 25627 25654 25812 25913 ...
 $ total : num  54389 54930 54995 55049 56323 ...
plot.ts(res[,c(2,3,4,5)])

# Pero se pueden comparar las series utilizando el
# comando lines.
#-----grafica
np = nrow(res)
#-----convierte fecha a formato de R
fechas = as.Date(fecha, format="%Y/%m/%d")
ejex.mes = seq(fechas[1], fechas[np], "months")
ejex.año = seq(fechas[1], fechas[np], "years")

plot(fechas, res$golfo, xaxt="n", panel.first = grid(),
type='l', lwd=2, ylab='produccion.mes.', ylim=c(9000,45000))
axis.Date(1, at=ejex.mes, format="%m/%y")
axis.Date(1, at=ejex.año, labels = FALSE, tcl = -0.2)
lines(fechas, res$noopec, col='red', lwd=2)

```

2.3.4. Series con datos faltantes

Varias librerías permiten imputar datos faltantes en series de tiempo. Por ejemplo `pastecs`.

Código R 2.4: Ejemplo de imputacion de datos faltantes en una serie de tiempo

```

#-----

```

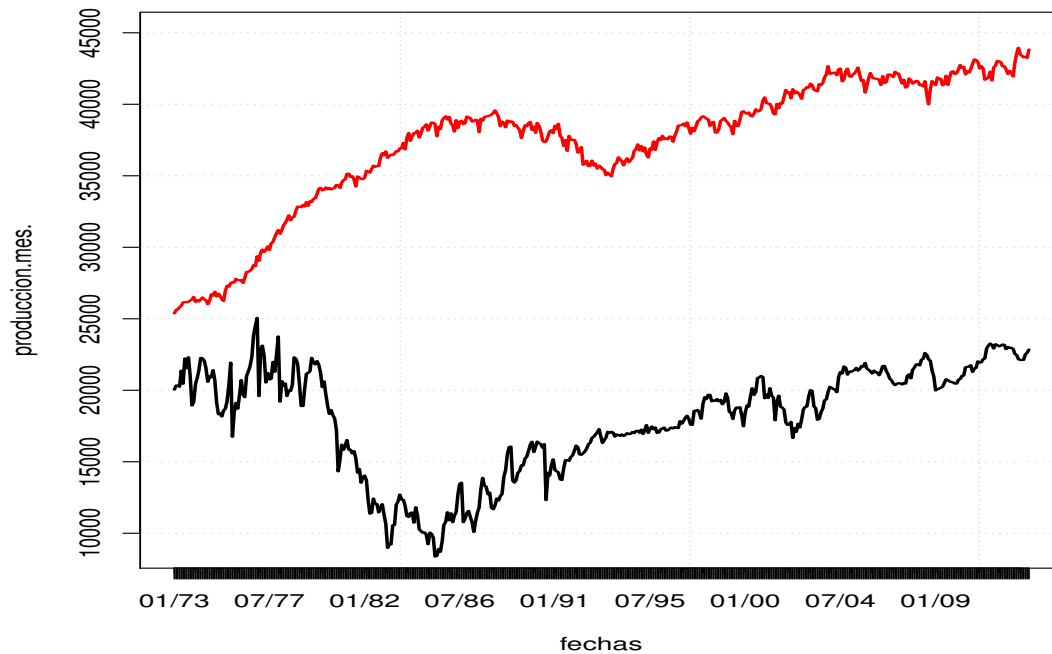


Figura 2.3: Gráfica producción mensual de crudo para países del Golfo y para los no pertenecientes a la Opec (en rojo)

```
install.packages("pastecs")
library(pastecs)
library(help="datasets")
attach(airquality)
help(airquality)
# Daily readings of the following air quality values for May 1,
# 1973 (a Tuesday) to September 30, 1973.
# Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at
# Roosevelt Island
# 'data.frame':  153 obs. of  6 variables:
# $ Ozone   : int  41 36 12 18 NA 28 23 19 8 NA ...
# $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
# $ Wind    : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
# $ Temp    : int  67 72 74 62 56 66 65 59 61 69 ...
```

```

# $ Month   : int   5 5 5 5 5 5 5 5 5 5 5 ...
# $ Day      : int   1 2 3 4 5 6 7 8 9 10 ...

# Nótese que la serie Ozone tiene datos faltantes , NA.
# El procedimiento de imputación es con base en
# interpolación mediante splines.
n = length(Ozone)
x = seq(1,n)
ozone.lleno = regul(x, y=Ozone,
xmin=min(x), n=length(x),
units="days", deltat=1,
methods="spline",
rule=2, f=0.5, periodic=FALSE,
window=(max(x) - min(x))/(n - 1),
split=100, specs=NULL)

str(ozone.lleno)
layout(1:2)
ts.plot(Ozone)
ts.plot(ozone.lleno$y$Series, ylab="imputada")

```

2.3.5. Funciones para escritura de datos

Hay funciones para escribir datos en un archivo, como la función `write.matrix` de la librería `MASS`

```

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

```

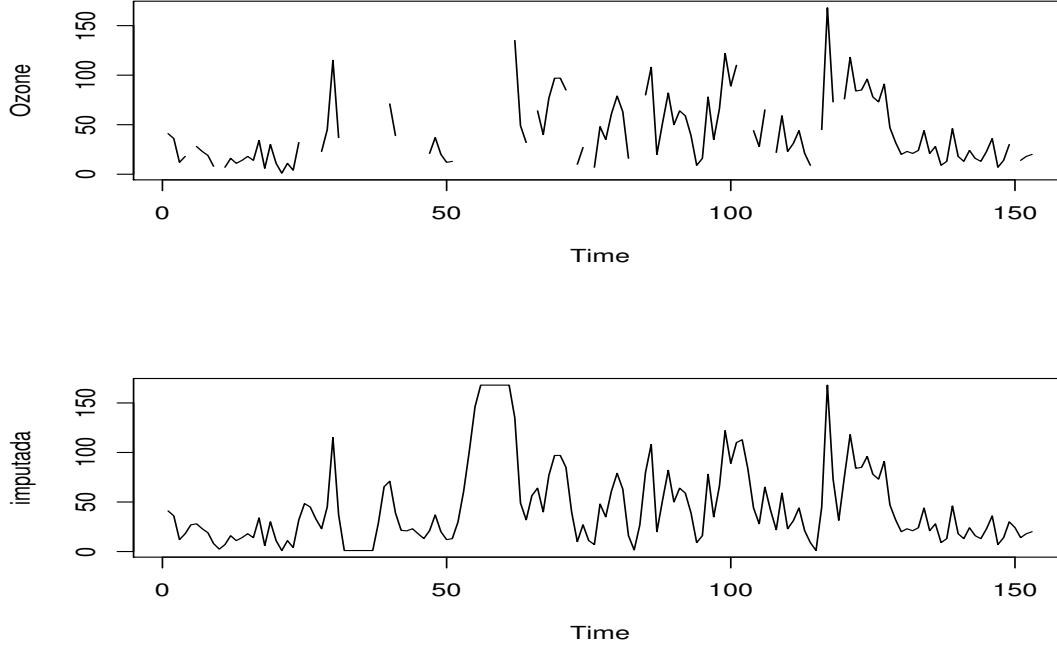



Figura 2.4: Serie con datos faltantes versus serie con imputación mediante splines

2.4. Repaso del Modelo de Regresión Lineal

El modelo de regresión lineal múltiple es básico para los modelos de componentes de series de tiempo. Una exposición más completa de este modelo está en, por ejemplo, en Kleinbaum et al. [1988] y Sheather [2009, caps. 5,6].

El modelo siguiente es lo suficientemente general para introducir los conceptos básicos. Suponga que se tiene una muestra de tamaño T de la variable aleatoria Y , y T mediciones de las variables X_1, X_2 , que no se consideran aleatorias, que es una manera de expresar que estos datos están prefijados

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

Además suponga que $\varepsilon_t \sim iid(0, \sigma^2)$ es una muestra de una variable aleatoria continua de media cero y varianza σ^2 . Por ejemplo, una $N(0, \sigma^2)$ ó una t-Student t_ν . Si se cumple que

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

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

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

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

entonces el modelo se puede expresar en forma matricial como

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}. \quad (2.2)$$

Los supuestos del modelo lineal son

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

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

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

El modelo se dice que es lineal porque Y_t se relaciona linealmente con $X_{1,t}, X_{2,t}$, es decir, Y_t es una combinación lineal de $X_{1,t}, X_{2,t}$. Relaciones de la forma siguiente son ejemplos de modelos no lineales:

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}}} + \varepsilon_t, \quad (2.3)$$

$$Y_t = \begin{cases} \alpha Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \leq r \\ \beta Y_{t-1} + \gamma \varepsilon_t & \text{if } Y_{t-1} > r \end{cases} \quad (2.4)$$

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-1} \varepsilon_{t-1} + \varepsilon_t \quad (2.5)$$

El modelo (2.3) es un ejemplo simple de una red neuronal. La forma general de una red neuronal con $p = 1, 2, \dots$ nodos de entrada y $q = 1, 2, \dots$ nodos ocultos es

$$Y_t = \alpha_0 + \sum_{j=1}^q \alpha_j f \left(\beta_{0,j} + \sum_{k=1}^p \beta_{j,k} X_{k,t} \right) + \varepsilon_t, \quad (2.6)$$

donde se escoje $f(x) = (1 + e^x)^{-1}$, entre otras opciones. Para el caso de (2.3) se tiene $p = 2, q = 1$. En la sección §5.6 se retoman las redes neuronales con más detalle.

Es conveniente señalar que, además de los conceptos de modelos lineal y no lineal, también existen los de serie lineal y no lineal, ver por ejemplo, Priestley [1988, Section 11.5.1, pag. 869]. Pero son diferentes.

En la sección §5.6 se definirá con mayor precisión serie no lineal. Muchas series conocidas se sabe que son de naturaleza no lineal, por ejemplo, las que muestran ciclos como las poblaciones de animales e insectos, la serie del número de manchas solares en la superficie del sol, entre otras.

Pero para decidir apriori si una serie de tiempo es no lineal se recurre a ciertas pruebas de hipótesis.

Una vez que se tiene evidencia de la no linealidad de la serie, se recomienda utilizar modelos no lineales, como las redes neuronales. En Venables and Ripley [2002], pag. 243:

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

El pronóstico de Y_t se define en general como una esperanza condicional

$$\begin{aligned} \hat{Y}_t &= \mathbb{E}(Y_t | Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}) \\ &= f(Y_{t-1}, \dots, Y_{t-p}), \end{aligned}$$

para cierta función $f(\cdot)$. Si la serie es lineal y cumple otras condiciones adicionales, entonces esta esperanza es una función lineal de las variables predictoras, por lo que un modelo lineal es adecuado. Aplicar un modelo no lineal sería innecesario. En

cambio, si la serie es no lineal la esperanza condicional se podría aproximar por un modelo no lineal.

En el blog de F. Diebold, Diebold [2018], hay discusiones y observaciones útiles sobre el uso de los métodos de Aprendizaje de Máquina aplicados a series de tiempo.

2.4.1. Estimación de los Parámetros

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

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

2.4.2. Mínimos cuadrados ordinarios

Los estimadores MCO de los coeficientes β_i en (2.1), se definen como el vector $\hat{\underline{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ que minimiza la función objetivo

$$G(\underline{\beta}) = \sum_{t=1}^T [Y_t - (\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t})]^2, \quad (2.7)$$

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

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

Y se escribe

$$\hat{\underline{\beta}} = \underset{\underline{\beta} \in \mathbb{R}^3}{\operatorname{argmin}} G(\underline{\beta}).$$

Minimizar la función objetivo $G(\underline{\beta})$ es resolver el problema de estimación por MCO. Sin embargo, $\hat{\underline{\beta}}$ satisface la fórmula matricial,

$$\hat{\underline{\beta}} = (X'X)^{-1}X'Y, \quad (2.8)$$

ver la comprobación en la sección al final. Hay dos propiedades básicas del estimador MCO $\hat{\underline{\beta}}$. Es un estimador lineal insesgado, es decir,

$$\mathbb{E}(\hat{\underline{\beta}}) = \underline{\beta}, \quad (2.9)$$

ver la comprobación en la sección al final. Se puede comprobar que el estimador MCO coincide, en este caso del modelo lineal, con el estimador de máxima verosimilitud. Lo cual implica que comparte la propiedad de mínima varianza, que significa,

$$Var(\hat{\beta}_i) \leq Var(\hat{\beta}_i^*), \quad i = 0, 1, 2,$$

para $\hat{\beta}_i^*$ cualquier estimador diferente de MCO.

Se comprueba que de (2.8) se obtiene la expresión para la matriz de varianzas y covarianzas del vector $\hat{\underline{\beta}}$ como

$$Var(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1}, \quad (2.10)$$

ver la comprobación en la sección al final. La diagonal de la matriz $\sigma^2(X'X)^{-1}$ contiene las varianzas de los $\hat{\underline{\beta}}$, indicadas por

$$Var(\hat{\beta}_i) = \sigma_{\hat{\beta}_i}^2 = diag(\sigma^2(X'X)^{-1})_i. \quad (2.11)$$

2.4.3. Pronósticos

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

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t} + \hat{\beta}_2 X_{2,t}, \quad t = 1, \dots, T. \quad (2.12)$$

Nótese que se cumple $\hat{Y} = X\hat{\underline{\beta}}$. Reemplazando (2.8) en esta expresión se obtiene

$$\hat{Y} = X(X'X)^{-1}X'Y = HY, \quad (2.13)$$

donde la matriz simétrica

$$H = X(X'X)^{-1}X' \quad (2.14)$$

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

$$\hat{Y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,T+h} + \hat{\beta}_2 X_{2,T+h}, \quad h = 1, \dots, m, \quad (2.15)$$

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

2.4.4. Error cuadrático medio, MSE

El error cuadrático medio, MSE (*mean square error*), se define como

$$MSE = \frac{1}{T-k} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 = \frac{1}{T-k} \sum_{t=1}^T \hat{\varepsilon}_t^2, \quad (2.16)$$

donde k es el número de coeficientes en el modelo, es decir, la dimensión de β . El MSE se define también como el estimador de la varianza del error

$$\hat{\sigma}^2 = MSE. \quad (2.17)$$

Y se define el estimador

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}. \quad (2.18)$$

MSE se calcula directamente con instrucción

```
MSE = anova(m2) ['Residuals', 'Mean Sq']
```

2.4.5. Residuos

Se define los residuos estimados como

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t, \quad t = 1, \dots, T. \quad (2.19)$$

En forma vectorial es

$$\underline{\hat{\varepsilon}} = \underline{Y} - X\underline{\hat{\beta}} = \underline{Y} - H\underline{Y} = (I - H)\underline{Y}. \quad (2.20)$$

La esperanza y la varianza de $\underline{\hat{\varepsilon}}$ son

$$\mathbb{E}(\underline{\hat{\varepsilon}}) = \mathbb{E}(\underline{Y} - X\underline{\hat{\beta}})$$

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

$$\text{Var}(\underline{\hat{\varepsilon}}) = \sigma^2(I - H), \quad (2.21)$$

ver la comprobación de esta identidad en la sección al final.

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

Los residuos estandarizados se definen como

$$r_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}\sqrt{1 - H_{t,t}}}. \quad (2.22)$$

Ejemplo 2.4.1. *Los datos de tres variables Y, X_1, X_2 , para ajustar el modelo de regresión (2.1) están en la página web*

<http://www.stat.umn.edu/geyer/5102/data/ex5-3.txt>

Código R 2.5: Ejemplo de regresión lineal con R

```
#-----
#-----leer datos en una url
archivo = "http://www.stat.umn.edu/geyer/5102/data/ex5-3.txt"
# archivo = "ex5-3.txt"
D = read.table(archivo, header=T)
str(D)
attach(D)
#-----estimar el modelo y=x1+x2+e, por MCO
m2 = lm(y ~ x1 + x2)
summary(m2)
#-----calcular residuos
ehat = residuals(m2)
#-----estimar mse y sigma
sigma=summary(m2)$sigma
MSE = sigma^2
#-----directamente
MSE = anova(m2)[ 'Residuals', 'Mean Sq']
#-----calcular residuos estandarizados r
```

```

require(MASS)
r = stdres(m2)
#-----valores ajustados
yhat = fitted(m2)
#-----graficas
par(mfrow=c(2,2))
plot(x1,y,type='p',main="A")
plot(r,main="B")
abline(h=0)
plot(y,yhat,type='p',main="C")
plot(ehat,type='p',main="D")
abline(h=0)

```

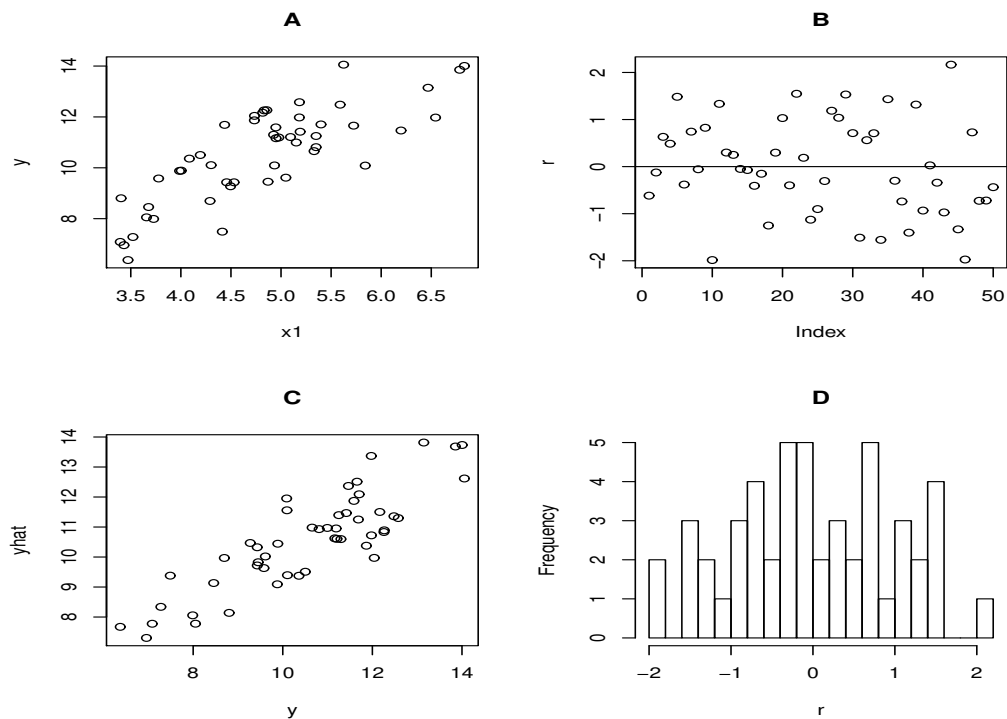


Figura 2.5: gráficas A: y versus x1, B: residuos r, C: y versus yhat, D: histograma residuos r

Demostraciones de identidades

- Demostración de la fórmula (2.8).

Demostración. Expresando $G(\underline{\beta})$ por

$$G(\underline{\beta}) = \|\underline{Y} - X\underline{\beta}\|^2 = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}).$$

donde \underline{x}' es el transpuesto de \underline{x} , también indicado por \underline{x}^\top . Utilizando fórmulas para derivadas matriciales se plantea obtener el mínimo como el valor $\underline{\beta}$ que resuelve la ecuación

$$\frac{\partial}{\partial \underline{\beta}} G(\underline{\beta}) = \underline{0}.$$

Se puede comprobar que (ver, por ejemplo, ⁽¹⁾, ó Wikipedia ⁽²⁾)

$$\frac{\partial}{\partial \underline{\beta}} G(\underline{\beta}) = 2X'(\underline{X}\underline{\beta} - \underline{Y})$$

por tanto, la solución es

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

□

- Comprobación de la propiedad de insesgamiento (2.9)

Demostración.

$$\begin{aligned} \mathbb{E}(\hat{\underline{\beta}}) &= \mathbb{E}((X'X)^{-1}X'\underline{Y}) = (X'X)^{-1}X'\mathbb{E}(\underline{Y}) \\ &= (X'X)^{-1}X'\mathbb{E}(X\underline{\beta} + \underline{\varepsilon}) = (X'X)^{-1}X'(X\underline{\beta} + \mathbb{E}(\underline{\varepsilon})) \\ &= (X'X)^{-1}X'X\underline{\beta} = \underline{\beta}. \end{aligned} \tag{2.23}$$

□

- Demostración de la identidad (2.10)

¹<http://www.matrixcalculus.org/>

²https://en.wikipedia.org/wiki/Matrix_calculus

Demostración. Con base en la expresión para $\hat{\underline{\beta}}$

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

de donde

$$\begin{aligned}Var(\hat{\underline{\beta}}) &= \mathbb{E}((\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})') = (X'X)^{-1}X'\mathbb{E}(\underline{\varepsilon}\underline{\varepsilon}')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\end{aligned}$$

□

- Demostración de la identidad (2.21).

Demostración. Por definición de Varianza de un vector aleatorio de media cero,

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

A partir de $\mathbb{E}(\underline{Y}\underline{Y}') = Var(\underline{Y}) = Var(X\underline{\beta} + \underline{\varepsilon}) = \mathbb{E}(\underline{\varepsilon}\underline{\varepsilon}')$, la hipótesis $\varepsilon_t \sim iid(0, \sigma^2)$ implica $\mathbb{E}(\underline{\varepsilon}\underline{\varepsilon}') = \sigma^2 I$, con $I \in \mathbb{R}^{T \times T}$ la matriz identidad. Luego $\mathbb{E}(\underline{Y}\underline{Y}') = \sigma^2 I$. Se obtiene entonces $Var(\hat{\underline{\varepsilon}}) = \sigma^2(I - H)(I - H)'$. Se puede comprobar que $H^2 = H$ y por tanto $(I - H)(I - H)' = I - H$, luego finalmente se obtiene $Var(\hat{\underline{\varepsilon}}) = \sigma^2(I - H)$. □

2.4.6. 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, por ejemplo, de la forma

$$Y_t = g(X_{1,t}, X_{2,t}; \underline{\beta}) + \varepsilon_t, t = 1, 2, \dots, T. \quad (2.24)$$

donde $g(x_1, x_2; \underline{\beta}) : \mathbb{R}^2 \rightarrow \mathbb{R}$ es una función dada que depende de los parámetros $\underline{\beta}$. Nótese que debe ser un modelo con errores aditivos. La función en R para mínimos

cuadrados no lineales es `nls()` y se aplica de la misma forma que `lm()`. Por ejemplo, para el modelo no lineal (2.3), pag. 28

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}}} + \varepsilon_t,$$

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

```
m = nls(Y ~ a0+a1/(1+exp(b0+b1*X1+b2*X2)),
start=list(a0=0.5,a1=1,b0=0.1, b1=1, b2=1))
```

Nótese que la función requiere unos valores iniciales para los parámetros. En el capítulo siguiente hay un ejemplo concreto de estimación de este modelo.

2.4.7. Regresión Local Polinómica

Es una alternativa a la regresión lineal, que puede aproximar casos no lineales. Se puede adaptar al caso de series de tiempo, en cuyo caso produce estimadores de la tendencia de una serie.

Suponga los datos: $(X_{1,i}, X_{2,i}, Y_i), i = 1, \dots, N$. El modelo de regresión local polinómica asume que existe una función $\mu(x_1, x_2)$ que cumple

$$Y_i = \mu(X_{1,i}, X_{2,i}) + \varepsilon_i \quad (2.25)$$

La función $\mu(x_1, x_2)$ se puede aproximar en una vecindad del punto $(X_{1,i}, X_{2,i})$ mediante un desarrollo de Taylor bivariado de orden 2

$$\begin{aligned} \mu(x_1, x_2) &\approx a_0 + a_1(x_1 - X_{1,i}) + a_2(x_2 - X_{2,i}) + \frac{1}{2}a_3(x_1 - X_{1,i})^2 + \\ &a_4(x_1 - X_{1,i})(x_2 - X_{2,i}) + \frac{1}{2}a_5(x_2 - X_{2,i})^2 \\ &= G(X_{1,i}, X_{2,i}; \underline{a}) \end{aligned}$$

con $\underline{a} = (a_1, \dots, a_5)'$.

Defina la función tricúbica

$$W(x) = (1 - |x|^3)^3, \quad -1 \leq x \leq 1. \quad (2.26)$$

Y defina para un punto (x_1, x_2) los pesos

$$W_i(x_1, x_2) = W\left(\frac{\|(X_{1,i}, X_{2,i}) - (x_1, x_2)\|}{h}\right) \quad (2.27)$$

donde $h > 0$ es un parámetro que se fija arbitrariamente, y los índices i se escogen de tal forma que

$$\frac{\|(X_{1,i}, X_{2,i}) - (x_1, x_2)\|}{h} \leq 1$$

Para cada punto (x_1, x_2) se calcula

$$\hat{\underline{a}}(x_1, x_2) = \underset{\underline{a} \in \mathbb{R}^5}{\operatorname{argmin}} \left\{ \sum_{i=1}^N W_i(x_1, x_2) (Y_i - G(X_{1,i}, X_{2,i}; \underline{a}))^2 \right\} \quad (2.28)$$

Es estimador de regresión múltiple local en el punto (x_1, x_2) se define como el primer componente del vector $\hat{\underline{a}}(x_1, x_2)$

$$\hat{\mu}(x_1, x_2) = \hat{a}_0(x_1, x_2) \quad (2.29)$$

2.5. Pruebas de Ajuste del Modelo

El problema de estimación por MCO no requiere el supuesto de Normalidad en los residuos: $\varepsilon_t \sim iid N(0, \sigma^2)$. Asumir este supuesto adicional implica que los estimadores MCO $\hat{\underline{\beta}}$ se distribuyen Normal multivariado, $\hat{\underline{\beta}} \sim N(\underline{\beta}, \sigma^2(X'X)^{-1})$. Por tanto, las distribuciones de las marginales $\hat{\beta}_i$ son Normales univariadas

$$\hat{\beta}_i \sim N(\beta_i, \sigma_{\hat{\beta}_i}^2), \quad (2.30)$$

donde las varianzas $\sigma_{\hat{\beta}_i}^2$ son los elementos de la diagonal de la matriz $\sigma^2(X'X)^{-1}$. Además, con $\hat{\sigma}_{\hat{\beta}_i} = \sqrt{\hat{\sigma}^2(X'X)^{-1}_{i,i}}$, denominado el error estándar, se definen los estadísticos t-Student, t_i

$$t_i = \hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i}. \quad (2.31)$$

Terminología básica

Un repaso de la terminología básica de las pruebas de hipótesis. Para realizar un prueba se consideran cuatro pasos:

1. Hipótesis de la Prueba.
2. Estadístico de la Prueba.
3. Distribución del Estadístico bajo H_0 .
4. Decisión.

		H_0 real	
		V	F
H_0 según prueba	V	OK	error II
	F	error I	OK

Tabla 2.1: Tabla de decisiones en una prueba de hipótesis

La prueba se define colocando expresiones para las hipótesis nula H_0 y la alterna H_1 .

El Error Tipo I es rechazar la Hipótesis nula H_0 siendo ésta cierta. Se trata de un falso positivo porque en ocasiones la alterna H_1 es el resultado de interés. El nivel de significación α se define como

$$\mathbb{P}(\text{Error Tipo I}) = \alpha.$$

El Error Tipo II es no rechazar la nula H_0 siendo la alterna H_1 la cierta. Se trata de un falso negativo. La probabilidad de cometer este error es

$$\mathbb{P}(\text{Error Tipo I}) = \beta.$$

Además, la probabilidad $1 - \beta$, denominada la potencia de la prueba, se interpreta como la probabilidad de rechazar la nula correctamente, es decir, la probabilidad de rechazar H_0 cuando H_1 es cierta. Fijado el nivel de significación α y el tamaño de la muestra, la potencia puede ser baja ó alta.

Las pruebas que tienen una baja potencia tienden a no rechazar la nula H_0 . No tienen capacidad para detectar H_1 . En ocasiones, la alterna H_1 puede estar muy cerca de la nula H_0 , y se requiere una prueba con alta potencia para detectarla.

Pruebas de Significación de los parámetros

La prueba se describe así. Para $i = 1, \dots, k$,

1. Hipótesis nula versus alterna:

$$H_0 : \beta_i = 0 \text{ versus } H_1 : \beta_i \neq 0.$$

Rechazar la nula se expresa diciendo: “el parámetro es significativamente diferente de cero”, ó simplemente “es significativo”.

2. Estadístico de prueba $t_i = \hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i}$.
3. Si H_0 es cierta entonces $t_i \sim t_{T-k}$, t-Student con $T - k$ grados de libertad, donde k es la dimensión del vector $\underline{\beta}$.
4. Decisión. Se plantean 3 maneras equivalentes para decidir acerca de la hipótesis nula H_0 .

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

- 2 Defina el valor p como

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

Si $\text{Valorp} < 0.05$ se rechaza H_0 a un nivel de significancia del 5 %, de lo contrario no se rechaza H_0 .

- 3 Con base en intervalos de confianza para el parámetro. Se rechaza la nula a un nivel de significancia de 5 % si

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

Note que, con R se pueden obtener los intervalos anteriores con la instrucción `confint(m2, level=...)`. Por ejemplo,

Ejemplo 2.5.1. Continuando con el Ejemplo 2.4.1, las pruebas t-Student se pueden programar, como se muestra a continuación. Para obtener los coeficientes estimados, los estadísticos t y los valores p se puede utilizar la función `coeftest` de la librería `lmtest`. También el comando `summary(modelo)`.

```
#-----pruebas de significacion
require(lmtest)
coeftest(m2)
t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.43421     0.90937   0.4775 0.6352321
x1           1.41790     0.17192   8.2473 1.091e-10 ***
x2           0.67427     0.16884   3.9934 0.0002275 ***
---
Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1
```

Prueba F

Se define la hipótesis

$$H_0 : \forall j = 1, 2, \dots, k, \beta_j = 0$$

$$H_1 : \exists j, \ni b_j \neq 0$$

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

$$\begin{aligned} SST &= \sum_{t=1}^T (Y_t - \bar{Y})^2 \\ &= \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^T (Y_t - \hat{Y}_t)^2 \\ &= SSR + SSE. \end{aligned} \tag{2.32}$$

El estadístico para esta prueba se define como

$$F = \frac{SSR/(k-1)}{SSE/(T-k)} = \frac{SSR/(k-1)}{MSE} \tag{2.33}$$

el cual se distribuye, bajo H_0 , con una distribución tipo F-Fisher

$$F \stackrel{H_0}{\sim} F_{k-1, T-k}. \tag{2.34}$$

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

Ejemplo 2.5.2. Continuando con el Ejemplo 2.4.1, con $k = 3$, $T = 50$, $\alpha = 0,05$, se tiene que $F_{k-1,T-k,\alpha} = F_{2,47,0.95} = 3.195056$, utilizando el comando: `qf(0.95, 2, 47)`. Luego, dado $F_{obs} = 66.48$, como $F_{obs} > 3.2$, se rechaza la nula H_0 , es decir, $X_{1,t}$ y $X_{2,t}$ tienen capacidad predictiva sobre Y_t .

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

2.5.1. Estadísticos de Ajuste y de Selección de Modelos

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

$$R^2 = 1 - \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} = 1 - \frac{SSE}{SST}. \quad (2.35)$$

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

$R^2 \geq 0.7$ buena

$R^2 = 0.6$ buena pero no mucho

$R^2 = 0.4$ regular-mala

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

Hay que anotar que en los textos sobre regresión se recomienda no usar el R^2 si el modelo no tiene intercepto. Carece de sentido en este caso.

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

$$\bar{R}^2 = 1 - \frac{(T-1)SSE}{(T-k)SST}. \quad (2.36)$$

Como $R^2 = 1 - SSE/SST$, entonces \bar{R}^2 se puede escribir en terminos de R^2 de la siguiente manera

$$\bar{R}^2 = 1 - \frac{T-1}{T-k}(1 - R^2). \quad (2.37)$$

En las expresiones siguientes el símbolo \nearrow indica aumento en una variable, y \searrow disminución. En (2.37), si $k \nearrow$, entonces $T - k \searrow$ y $\frac{T-1}{T-k} \nearrow$, luego $\bar{R}^2 \searrow$. Lo que se interpreta como que \bar{R}^2 penaliza aumentar el número de parámetros.

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

$$AIC = e^{2k/T} \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T} = \frac{(T-k)\hat{\sigma}^2}{T} e^{2k/T}. \quad (2.38)$$

El AIC también se define como el logaritmo de (2.38), y además se puede relacionar con la suma de cuadrados $G(\hat{\beta})$ en (2.7), usando las identidades en (2.16) y (2.17), para obtener

$$\begin{aligned} AIC &= \log \left(\frac{(T-k)\hat{\sigma}^2}{T} \right) + \frac{2k}{T} \\ &= \log \left(\frac{G(\hat{\beta})}{T} \right) + \frac{2k}{T} \end{aligned}$$

Se observa que si $k \nearrow$, entonces $AIC \nearrow$. Si $G(\hat{\beta}) \searrow$, entonces $AIC \searrow$.

Por lo que el AIC disminuye mientras menor sea $G(\hat{\beta})$ ó $\hat{\sigma}$, es decir, mientras mejor sea el ajuste, pero aumenta si el número de parámetros k aumenta. El modelo con menor AIC debe ser el modelo que mejor ajusta con el menor número de parámetros, lo que se relaciona con el principio de parsimonia (navaja de Ockam, ver ⁽³⁾)

En Diebold [1999, pág. 73]

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

Se define el criterio de información de Schwarz BIC

$$BIC = T^{\frac{k}{T}} \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T} = T^{\frac{k}{T}} \frac{(T-k)\hat{\sigma}^2}{T} \quad (2.39)$$

El BIC también se define como el logaritmo de (2.39).

$$BIC = \log \left(\frac{(T-k)\hat{\sigma}^2}{T} \right) + \frac{k}{T} \log(T). \quad (2.40)$$

³https://es.wikipedia.org/wiki/Navaja_de_Ockham

De la definición del criterio BIC se observa:

Si $k \nearrow$, entonces $BIC \nearrow$. Si $G(\hat{\beta}) \searrow$, entonces $BIC \searrow$.

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

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

Definición 2.5.1 (Consistencia). *Si se examinan varios modelos, y el modelo generador de los datos (MGD) está incluído, el criterio se dice consistente si a medida que T aumenta, la probabilidad de que el criterio seleccione el MGD tiende a 1.*

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

Definición 2.5.2 (Eficiencia Asintótica). *Un criterio es eficiente asintóticamente si a medida que T aumenta, elige un modelo cuya varianza de error de pronóstico a un paso*

$$Var(Y_{t+1} - \hat{Y}_{t+1} | Y_s, s \leq t)$$

tiende a la que se obtendría con el MGD.

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

“Muchos autores recomiendan usar el modelo más parsimonioso que selecciona el BIC en igualdad de circunstancias.”

Ejemplo 2.5.3. *Continuando con el Ejemplo 2.4.1, se examinan 3 modelos lineales:*

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \varepsilon_t, \quad (2.41a)$$

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t, \quad (2.41b)$$

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{2,t}^2 + \varepsilon_t \quad (2.41c)$$

Las instrucciones en R se dan a continuación. Los resultados aparecen en la tablas siguientes.

```
#----estimar 4 modelos
m1 = lm(y ~ x1)
summary(m1)
m2 = lm(y ~ x1 + x2)
summary(m2)
x3 = x1^2
m3 = lm(y ~ x1 + x2 + x3)
summary(m3)
```

Tabla 2.2: Resultados estimación MCO modelo m1

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.378	0.880	2.703	0.009
x1	1.698	0.180	9.446	0.000

Tabla 2.3: Resultados estimación MCO modelo m2

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.434	0.909	0.477	0.635
x1	1.418	0.172	8.247	0.000
x2	0.674	0.169	3.993	0.000

Tabla 2.4: Resultados estimación MCO modelo m3

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.650	3.375	-1.674	0.101
x1	4.044	1.416	2.857	0.006
x2	0.617	0.167	3.683	0.001
x3	-0.262	0.140	-1.869	0.068

El comando en R con resultados se muestra a continuación.

Escoger uno de los modelos comparando los estadísticos de ajuste y de selección. Para esto se diseñó la función (de autor) `medidas`, que se debe cargar con la función `source`, como aparece en el código siguiente.

```
medidas = function(m, y, k) {
```

```

T = length(y)
yest = fitted(m)
sse = sum((yest-y)^2)
ssr = sum((y-mean(y))^2)
mse = sse/(T-k)
R2 = 1 - sse/ssr
Ra2 = 1 - (T-1)*(1-R2)/(T-k)
aic = log((T-k)*exp(2*k/T)*mse/T)
bic = log(T^(k/T)*(T-k)*mse/T)
M = c(mse,Ra2,aic,bic)
return(M)
}

```

Y su aplicación con los resultados, permite escoger uno de los modelos mediante los estadísticos de ajuste y selección, como se muestra en el código siguiente.

```

source("medidas.r")
k1 = 2; k2 = 3; k3 = 4;
N1 = medidas(m1,y,k1)
N2 = medidas(m2,y,k2)
N3 = medidas(m3,y,k3)
N = cbind(N1,N2,N3)
rownames(N) = c("MSE", "R2-ajus", "AIC", "BIC")
colnames(N) = c("m1", "m2", "m3")
(N)

```

Los resultados están en la tabla 2.5 siguiente.

Tabla 2.5: Resultados los estadísticos de ajuste y selección

	m1	m2	m3
MSE	1.240	0.945	0.898
R2-ajus	0.643	0.728	0.741
AIC	0.254	0.002	-0.031
BIC	0.330	0.116	0.122

Cuál modelo se escoge?. La elección estaría entre m2 y m3. El AIC da el m3. Pero el BIC da el m2.

2.5.2. Pruebas F parciales múltiples

En casos en los que los modelos a comparar consisten de un modelo inicial con k variables predictoras y un modelo ampliado con $k + p$ variables predictoras, es posible medir el efecto de la inclusión de las nuevas p variables predictoras, mediante una prueba F parcial múltiple. Tomando como ejemplo $p = 2$, el modelo completo de puede escribir

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_1^* X_{1,t}^* + \beta_2^* X_{2,t}^* + \varepsilon_t, \quad (2.42)$$

La hipótesis nula es que las variables predictoras adicionales no contribuyen de manera significativa en la predicción de Y_t

$$H_0 : \beta_1^* = \beta_2^* = 0. \quad (2.43)$$

El estadístico de la prueba F parcial múltiple es

$$F = \frac{(SSR(ampliado) - SSR(reducido))/(k - 1)}{MSE(ampliado)} \quad (2.44)$$

donde ampliado corresponde al modelo en (2.42) y reducido es el modelo original. El estadístico se distribuye, bajo H_0 , según una distribución tipo F-Fisher

$$F \stackrel{H_0}{\sim} F_{k-1, T-k-p}. \quad (2.45)$$

De la ayuda de la función `anova`: `help('`anova`')`, cuando se le da una secuencia de objetos, `anova` prueba los modelos entre sí en el orden especificado, mediante pruebas F parciales múltiples.

Ejemplo 2.5.4. *Continuando con los 3 modelos del Ejemplo 2.5.3, se aplicó la función `anova` `anova(m1,m2,m3)`.*

```
Analysis of Variance Table
```

```
Model 1: y ~ x1
```

Model 2: $y \sim x_1 + x_2$

Model 3: $y \sim x_1 + x_2 + x_3$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	59.498				
2	47	44.424	1	15.074	16.7929	0.0001675 ***
3	46	41.290	1	3.134	3.4914	0.0680637 .

La conclusión es que al comparar m_1 y m_2 se rechaza la nula. Luego, añadir $X_{2,t}$ aumenta la capacidad predictiva. Sin embargo, incorporar $X_{2,t}^2$ resulta en no rechazo de la nula, aunque con un valor p cercano a 0.05. La conclusión es escoger el modelo m_2 , como lo hizo el estadístico BIC en el análisis de las medidas de selección.

CAPÍTULO 3

Modelo Aditivo de Componentes

Una serie de tiempo es una sucesión finita 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 SO_2 en ppm (ppm=partes por millón, $100ppm = 262mg/m^3$), medida semanalmente en una estación de monitoreo, es importante para vigilar la calidad del aire en una ciudad.

En este capítulo se introduce el Modelo Aditivo de Componentes y se analizan modelos específicos para las componentes de la tendencia y la estacionalidad, definidas a continuación.

3.1. El Modelo Aditivo de Componentes de series de tiempo

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

$$Y_t = T_t + S_t + \varepsilon_t, \quad (3.1)$$

donde T_t es la tendencia, S_t es la componente estacional y ε_t es la componente de errores. Las componentes T_t y S_t son funciones determinísticas de t . Su evolución es completamente predecible. El modelo (3.1) se denomina también Modelo Estructural Básico, ver Harvey [2006].

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

$$Y_t = T_t S_t e^{\varepsilon_t}. \quad (3.2)$$

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

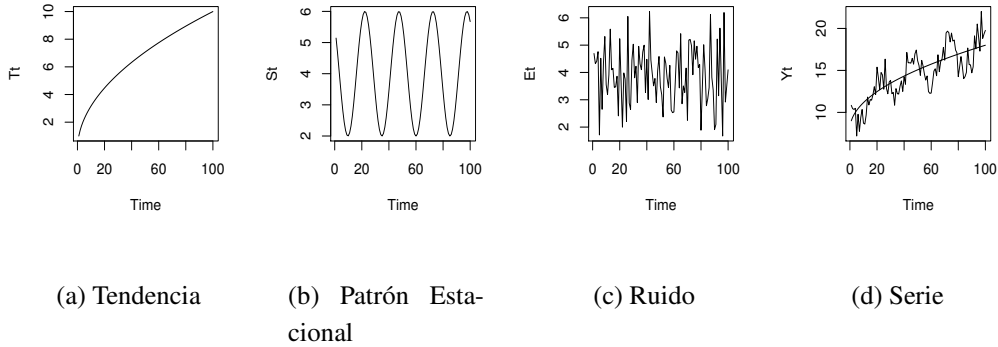


Figura 3.1: Descomposición de la Serie Y_t

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

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

La serie $\hat{\varepsilon}_t$ se modela y estima para finalmente 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 $h = 1, 2, \dots, m$ un horizonte de pronóstico. Sin embargo, puede suceder que la serie $\hat{\varepsilon}_t$ sea incorrelacionada, es decir, $Corr(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+s}) = 0$, para $s \neq 0$. En este caso $\hat{\varepsilon}_{T+h} = 0, \forall h > 0$.

3.2. Modelos para la Tendencia

Definición 3.2.1 (Tendencia). *Se define como una función T_t de t que describe la evolución lenta y a largo plazo del nivel medio de la serie.*

La función T_t puede depender de parámetros, que deben estimarse. Pero también es posible utilizar modelos no-paramétricos, como medias móviles y regresiones polinómicas locales. Los modelos paramétricos se pueden denominar “modelos globales”, en contraposición a los no-paramétricos, denominados “modelos locales”.

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

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

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

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

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

$$\text{Exponencial-cuadrático : } T_t = \exp(\beta_0 + \beta_1 t + \beta_2 t^2). \quad (3.3e)$$

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

La forma de la tendencia depende de los rangos de los parámetros, por ejemplo, en la tendencia cuadrática (3.3b) se cumple:

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

La componente T_t en algunos casos puede ser estacional, es decir, periódica, de período largo. Por ejemplo, en una serie diaria, S_t es una función periódica de período 30 días, y T_t también periódica de período 360 días.

Ejemplo 3.2.1. *Un análisis de tendencia puede ser útil para obtener conclusiones sobre el efecto de políticas generales.*

Por ejemplo, los datos mensuales de la tasa de empleo en Australia entre 02/1978 y 08/2104 se pueden ver en la Figura 3.2. Este indicador se ajusta estacionalmente y se mide en términos de miles de personas de 15 años y más como porcentaje de la población en edad laboral. Un análisis de la tendencia requiere considerar el cambio de tendencia que parece darse desde finales de 2008.

Se pueden consultar datos sobre tasa de empleo en varios países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), ver (¹).

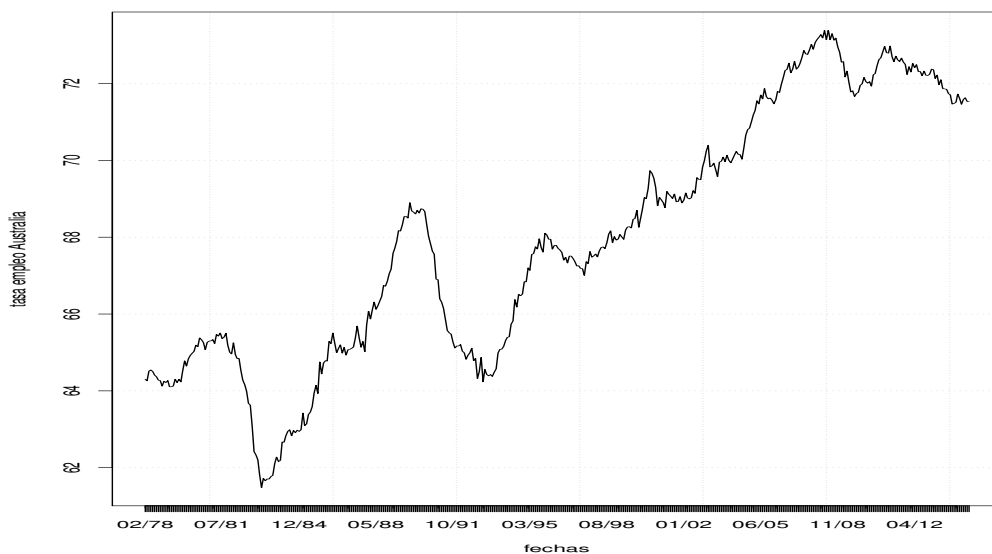


Figura 3.2: Tasa de Empleo, Australia, 02/1978 - 09/2014

Ejemplo 3.2.2. *Establecer la tendencia de una serie puede proporcionar pronósticos a corto y mediano plazo. Por ejemplo, en la Figura 3.3, donde se muestra la evolución del log-precio del kwh de energía eléctrica en el mercado MEN, un pronóstico puede ser la base para la negociación de contratos de futuros: compra a plazo a generadores.*

¹ <https://data.oecd.org/emp/employment-rate.htm>

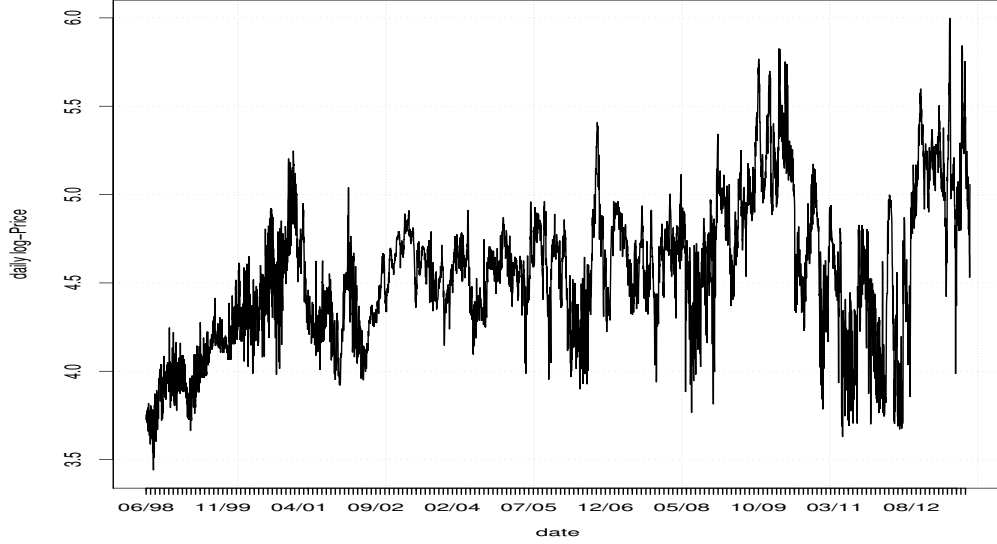


Figura 3.3: Logaritmo del precio del kwh en la bolsa energía Colombia

3.3. Estimación de la Tendencia

En este capítulo se introduce la estimación de la tendencia mediante modelos de regresión lineal y no lineal. En un capítulo posterior se introducen modelos no paramétricos. Son modelos que incluyen regresión polinómica local, filtros lineales y no lineales y medias móviles. Hay otros métodos que no se consideran en este curso, por ejemplo, wavelets (onditas). En ocasiones la estimación de la tendencia se denomina también “suavizar una serie” ó “extraer de la tendencia de una serie”.

Para la estimación de los parámetros $\underline{\beta} = (\beta_0, \beta_1, \beta_2)'$ en los modelos lineales (3.3a), (3.3b) y (3.3c) 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.

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

Para los modelos (3.3d) y (3.3f) 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 $\underline{\beta}$.

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 η_t es el término de error, de forma que $\eta_t = S_t + \varepsilon_t$. Por ejemplo.

1. En el caso lineal $T_t = \beta_0 + \beta_1 t$ con $\underline{\beta} = (\beta_0, \beta_1)'$, se ajusta el modelo de regresión lineal: $Y_t = \beta_0 + \beta_1 t + \eta_t$.
2. En el caso cuadrático $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ con $\underline{\beta} = (\beta_0, \beta_1, \beta_2)'$, 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 la variable explicativa adicional t^2 .
3. El modelo con tendencia exponencial-lineal (3.3d),

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

es no lineal y está relacionado, aunque no es equivalente, con un modelo con tendencia lineal para el logaritmo de Y_t , denominado modelo Logarítmico Lineal ó Log-Lineal, definido como

$$\ln Y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (3.5)$$

Es posible estimar por mínimos cuadrados ordinarios el modelo Log-Lineal (3.5) y utilizar los parámetros estimados $\hat{\beta}_0, \hat{\beta}_1$ como valores iniciales en la estimación del modelo con tendencia exponencial por mínimos cuadrados no lineales. Pero los parámetros estimados en ambos modelos no necesariamente coinciden. Y los valores estimados \hat{Y}_t en el modelo exponencial-lineal (3.3d) no necesariamente coinciden con los valores estimados $\exp(\widehat{\ln Y_t})$ con (3.5).

3.4. Pronósticos con base en la Tendencia

Definición 3.4.1. (Pronósticos). Suponga la serie con tendencia

$$Y_t = T_t + \eta_t, \quad t = 1, \dots, T,$$

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

$$\hat{Y}_{T+j} = \hat{T}_{T+j}, \quad j = 1, \dots, h. \quad (3.6)$$

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

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

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

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

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,

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

Otros tipos de pronósticos, en general, son

- El pronóstico por intervalo se obtiene si se asume que $\eta_t \sim iid N(0, \sigma^2)$. Entonces se cumple que $\widehat{Y}_{T+h} \pm 1.96\widehat{\sigma}$ es un intervalo de confianza del 95 %. Por ejemplo, en el modelo de tendencia lineal, el IC es

$$\widehat{\beta}_0 + \widehat{\beta}_1(T + h) \pm 1.96\widehat{\sigma}, \quad (3.8)$$

donde $\widehat{\sigma}^2 = MSE = \frac{1}{T-2} \sum_{t=1}^T e_t^2$

- El pronóstico de densidad es $Y_{T+h} \sim N(\widehat{Y}_{T+h}, \widehat{\sigma}^2)$.

Los pronósticos se pueden calcular en R mediante la función `predict(objeto, ...)` de la librería `forecast`. De la ayuda:

“predict(objeto,...) es una función genérica para pronósticos a partir de los resultados de varias funciones de ajuste de modelos. La función invoca métodos particulares que dependen de la clase del primer argumento”.

Por ejemplo, con un modelo de tendencia cuadrática $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \eta_t$, para $t = 1, 2, \dots, T$, si se genera un modelo de regresión lineal y se quiere calcular los pronósticos fuera de la muestra, en los tiempos $T+1, \dots, T+m$, se programa

```
m2 = lm(y ~ t + t2)
tt = seq(T+1, T+m)
tt2 = tp*tp
y.pron = predict(m2, data.frame(t=tt, t2=tt2))
```

3.5. Medidas de precisión en los pronósticos

Definición 3.5.1. *El procedimiento de validación cruzada consiste en reservar los últimos m datos de una muestra para contrastarlos con los pronósticos realizados a partir del modelo estimado con los $T - m$ primeros datos, utilizando ciertas medidas de calidad de pronósticos.*

Asuma $Y_t, t = T - m + 1, \dots, T$ y que se calcularon con estos datos el valor de los pronósticos \hat{Y}_t y los residuos estimados $\hat{\eta}_t = Y_t - \hat{Y}_t$. Se definen varias medidas para la precisión del pronóstico, por ejemplo

1. El error absoluto porcentual medio ó MAPE:

$$MAPE = \frac{1}{m} \sum_{t=T-m+1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|. \quad (3.9)$$

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

$$RMSE = \sqrt{\frac{1}{m} \sum_{t=T-m+1}^T (Y_t - \hat{Y}_t)^2}. \quad (3.10)$$

3. El estadístico U de Theil,

$$U = \sqrt{\frac{\sum_{t=T-m+1}^T (Y_t - \hat{Y}_t)^2}{\sum_{t=T-m+1}^T Y_t^2}} \quad (3.11)$$

En Bliemel [1973] se argumenta que si $U \geq 1$ entonces el modelo de pronósticos es equivalente a los pronósticos “ingenuos”, que consisten en pronosticar con el valor inmediatamente anterior en la serie. Sin embargo, en un blog reciente ⁽²⁾, R. Hyndman alega que no hay razón para que el U tenga que ser menor de uno porque es un cociente de valores medios cuadrados. Se excluyó el U de Theil de las versiones recientes de la función `accuracy`. Sin embargo, el U de Theil se puede calcular con la función `TheilU` de la librería `DescTools`.

Éstas y otras medidas se pueden calcular con la función `accuracy` de la librería `forecast`, con la instrucción `B=accuracy(y, yhat)`. Ver más detalles en Hyndman and Koehler [2006]. Devuelve un vector `B`, con componentes:

ME: Mean Error
 RMSE: Root Mean Squared Error
 MAE: Mean Absolute Error
 MPE: Mean Percentage Error
 MAPE: Mean Absolute Percentage Error
 MASE: Mean Absolute Scaled Error
 ACF1: Autocorrelation of errors at lag 1.

Ejemplo 3.5.1. *Suponga que se calcularon pronósticos para tres modelos: lineal, cuadrático y exponencial-lineal, que están como columnas en la matriz `pr`. El código siguiente calcula RMSE, MAE, MAPE, U-Theil. Note que se excluyeron ME y MPE porque producen valores negativos. En la Tabla 3.3 están los resultados.*

```
#-----Ejemplo código para calidad de pronósticos
R = rbind(accuracy(yf,pr[,1]),
accuracy(yf,pr[,2]),
accuracy(yf,pr[,3]))
rownames(R) = c("M.lin","M.cuad","M.exp")
require(DescTools)
Utheil=c(TheilU(yf,pr[,1], type=2),
TheilU(yf,pr[,2], type=2),
```

²<https://stats.stackexchange.com/questions/8396/how-to-produce-theils-u-with-package-forecast-2-16-in-r>

```

TheilU(yf,pr[,3], type=2))
R = cbind(R,Utheil)
R = R[,-c(1,4)]
(R)

```

3.5.1. 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.

Descripción de los Datos

Ventas al menudeo en USD a precios del año 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.

Validación cruzada. Usar el período 01/1955 - 12/1993 para estimar los modelos (período de entrenamiento) y el período 01/1994 - 12/1994 para examinar la eficiencia de los pronósticos fuera de la muestra.

Lectura de Datos, validación cruzada

```

D = read.table("ventas_al_menudeo.dat",header=T)
attach(D) # utiliza el nombre de las columnas como variables
# RTRR es el volumen de ventas en grandes almacenes en USD de 1999
# La variable RTRR del archivo tiene datos faltantes NA

y = na.omit(RTRR)/10000
# Convertir los datos en un objeto tipo ts
y = ts(y,frequency=12,start=c(1955,01))

```

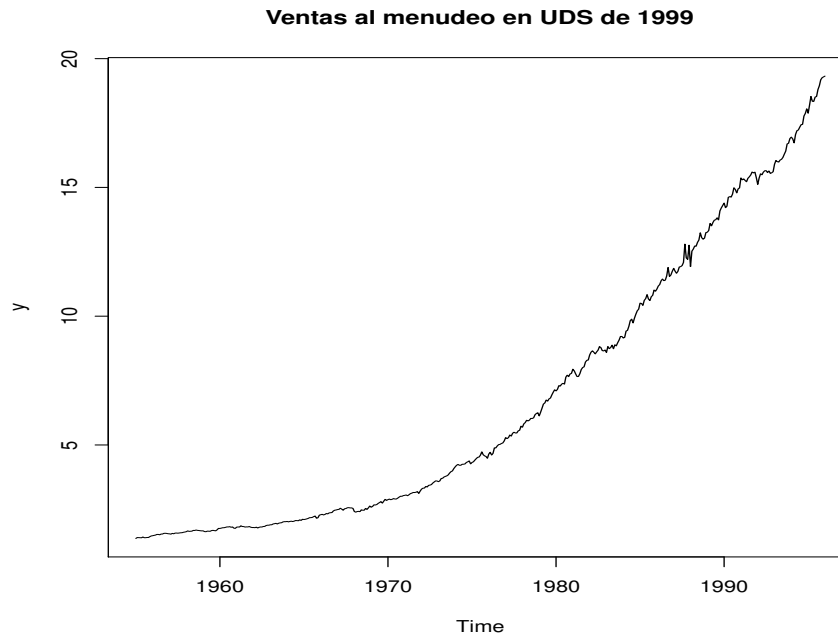



Figura 3.4: Ventas al menudeo, del texto Diebold(1999)

```
# Generar datos para validacion cruzada: dejar el
# ultimo año
T = length(y)
yi = y[1:(T-12)]
yf = y[(T-12+1):T]
```

Ajustar modelos:

```
# 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)
```

Resultados del ajuste de los Modelos

Nótese de los resultados en la Tabla 3.1 que el modelo cúbico no ajusta ya que el coeficiente de t^3 no da significativo al nivel de 5 %.

Calcular los Estadísticos de Selección del Modelo

Se procedió a calcular los estadísticos de calidad de ajuste, y los resultados está en la Tabla 3.2. Para calcular los estadísticos se usa la función `medidas()`, que los calcula para cada modelo.

```
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)
```

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

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

	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
AIC	1.001	-2.543	-2.544	-1.327
BIC	1.018	-2.517	-2.509	-1.309

Tabla 3.2: Estadísticos de Selección de Modelos

De la Tabla 3.2 se puede concluir que el modelo que mejor ajusta los datos es el modelo cuadrático (3.3b) ya que presenta el menor BIC.

Cálculo de Pronósticos con los Modelos

```
# Pronósticos sin incluir el cubico
tt=seq((T-12+1),T,1)
tt2 = tt*tt
pr = mat.or.vec(length(tt),3)
pr[,1] = predict(mod.lin,data.frame(t=tt))
pr[,2] = predict(mod.cuad,data.frame(t=tt,t2=tt2))
```

```

pr[,3] = predict(mod.exp,data.frame(t=tt))
# graficas de los pronosticos
par(mfrow=c(1,1))
plot(tt,yf,type='b',lwd = 2,
ylim = c(10,22))
lines(tt,pr[,2],col='blue',lwd=2)
lines(tt,pr[,1],col='red')
lines(tt,pr[,3],col='magenta')

```

Los pronósticos están en las columnas de la matriz `pr`, generada en el programa anterior. A partir del examen de la Figura (3.5) se puede concluir que el modelo cuadrático genera pronósticos confiables a corto plazo.

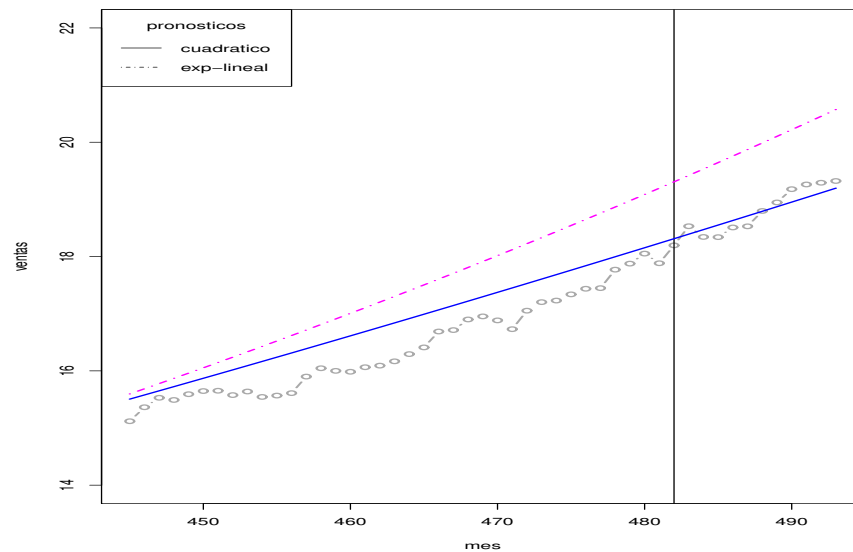


Figura 3.5: Pronósticos (—) versus Observados (-o-o-)

Cálculo de estadísticos de calidad de pronósticos

```

# Estadísticos de calidad
R = rbind(accuracy(yf,pr[,1]),
accuracy(yf,pr[,2]),

```

```

accuracy(yf,pr[,3]))
rownames(R) = c("M.lin", "M.cuad", "M.exp")
require(DescTools)
Utheil=c(TheilU(yf,pr[,1], type=2),
TheilU(yf,pr[,2], type=2),
TheilU(yf,pr[,3], type=2))
R = cbind(R,Utheil)
R = R[,-c(1,4)] # excluye ME y MPE
(R)

```

Tabla 3.3: Estadísticos de calidad de pronósticos para tres modelos

	RMSE	MAE	MAPE	U-Theil
M.lin	3.924	3.913	26.325	0.209
M.cuad	0.157	0.145	0.771	0.008
M.exp	1.169	1.163	5.836	0.062

3.6. Modelos para la Componente Estacional

En esta sección se introducen modelos para la componente estacional S_t , en el modelo de componentes de una serie: $Y_t = T_t + S_t + \varepsilon_t$, con T_t la tendencia y ε_t la componente aleatoria. Además se exponen procedimientos de estimación de S_t con base en regresión lineal múltiple.

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

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

Definición 3.6.1 (Componente Estacional). *La componente estacional S_t se define como una función periódica, de período $s = 2, 3, \dots$ que describe la estacionalidad*

de una serie. El conjunto de valores de S_t para $t = 1, \dots, s$ se denomina el patrón estacional. El período estacional s es el número mínimo de períodos que tarda el patrón estacional en repetirse.

La unidad de tiempo t de Y_t es día, semana, mes, trimestre, semestre, año. Definir el período estacional, por ejemplo, como $s = 12$, significa que el patrón estacional consiste de doce meses, para una serie mensual. De manera similar con $s = 4$, serie trimestral. Un período semanal es $s = 52$ y se muestrea 52 veces al año.

También se puede definir un modelo con período en días, cuando la serie se muestrea diariamente. Por ejemplo, si $s = 7$, entonces el patrón estacional que se repite cada 7 días. Podría ser $s = 30$, período mensual. También es posible un período de $s = 365$ días.

Ejemplo 3.6.1. *Una ejemplo de serie con componente estacional es la serie **nottem**, en la librería stats de R. Corresponde a la serie mensual “Average Monthly Temperatures at Nottingham, 1920-1939”, que contiene las temperaturas promedio de grados Fahrenheit tomadas mensualmente en el Castillo de Nottingham (UK), por 20 años, entre 1920 y 1939, ver ⁽³⁾. 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.*

³<http://www.astrostatistics.psu.edu/datasets/R/html/datasets/html/00Index.html>

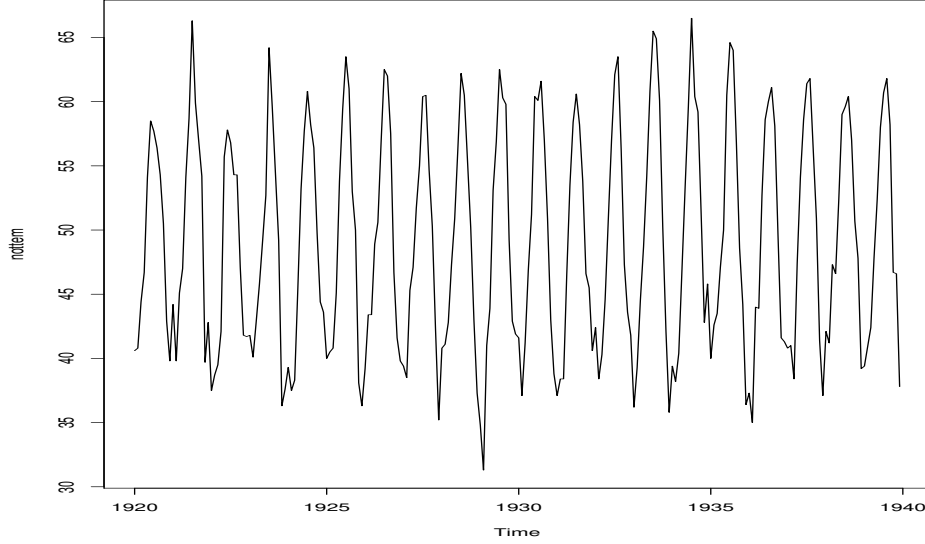


Figura 3.6: Average Monthly Temperatures at Nottingham, 1920-1939

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 periódica con periodo s , es decir cumple: $S_{t+s} = S_t$ 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 . Luego, el conjunto de funciones periódicas de período s es un espacio vectorial de dimensión s .
3. Una base para este espacio vectorial está conformada por las s variables indicadoras estacionales, $I_j(t) \in \{0, 1\}$, para $j = 1, \dots, s, t = 1, 2, \dots$, definidas por

$$I_j(t) = \begin{cases} 1, & t = j, j + s, j + 2s, \dots \\ 0, & \text{en otro caso} \end{cases} \quad (3.12)$$

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

$$S_t = \sum_{j=1}^s \delta_j I_j(t), \quad t = 1, 2, \dots \quad (3.13)$$

La condición para $I_j(t) = 1$ se puede dar también de la forma

$$I_j(t) = 1 \Leftrightarrow t \equiv j \pmod{s} \Leftrightarrow t - j = s \lfloor t/s \rfloor. \quad (3.14)$$

donde $\lfloor x \rfloor$ es la parte entera de x , p.ej. $\lfloor 2.34 \rfloor = 2$. La expresión en (3.14) se lee “ t congruente con j , módulo s ”, lo cual significa que $t - j$ es divisible por s , es decir que el cociente $\frac{t-j}{s}$ es un entero, que se expresa como $t - j = s \lfloor t/s \rfloor$.

Por ejemplo, con $s = 4$, es $I_3(11) = 1$?. Si $s = 4, j = 3, t = 11$, utilizando (3.14),

$$\begin{aligned} (t - j)/s &= \lfloor t/s \rfloor \\ (11 - 3)/4 &= \lfloor 11/4 \rfloor \Leftrightarrow 8/4 = \lfloor 2.75 \rfloor \Leftrightarrow 2 = 2. \end{aligned}$$

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

4. También se puede utilizar el conjunto

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

Para S_t periódica de período s , escogemos $m \geq 1$ y constantes $\beta_{1,j}, \beta_{2,j}, j = 1, \dots, m$ 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 en los rangos siguientes

$$\text{rango } m = \begin{cases} 1 \leq m \leq \frac{s}{2} - 1, & \text{si } s \text{ es par} \\ 1 \leq m \leq \frac{s-1}{2}, & \text{si } s \text{ es impar} \end{cases} \quad (3.15)$$

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

5. Nótese que en el modelo con variables indicadoras el número de parámetros aumenta con el período s . Para períodos grandes, por ejemplo $s = 360$, el modelo (3.13) no sería práctico. Una alternativa sería un modelo con variables trigonométricas

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

3.7. Modelo de componentes aditivas

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 (3.13), entonces el modelo de componentes consiste en el modelo de regresión lineal múltiple para Y_t está dado por:

$$Y_t = \beta_0 + \sum_{j=1}^k \beta_j t^j + \sum_{j=1}^s \delta_j I_j(t) + \varepsilon_t, \quad (3.16)$$

donde $k = 1, 2, 3$, β_i , $i = 1, \dots, k$ y δ_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.

A continuación se enumeran los pasos para estimar el modelo con tendencia y estacionalidad S_t con variables indicadoras, (3.16).

1. Identificar el período de la serie Y_t , s , si es que existe. Al declarar en R la serie como objeto “ts” con frecuencia, por ejemplo 12, se está identificado el posible período como $s = 12$. Sin embargo, el caso $s = 360$ no se incluye con indicadoras. En este caso sería preferible usar funciones trigonométricas, como se explica más adelante.
2. Generar las s variables indicadoras $I_j(t)$ para $t = 1, 2, \dots, T$. Es decir, generar una matriz con entrada (t, j) dada por $I_j(t)$, donde al variar $j = 1, \dots, s$ varían las columnas y al variar $t = 1, \dots, T$ varían las filas.
3. Estimar los parámetros en el modelo (3.16) mediante regresión lineal. Nótese que se trata de estimar los parámetros de las componentes de tendencia y estacional, conjuntamente.

La función en (3.16)

$$T_t + S_t = \beta_0 + \sum_{j=1}^k \beta_j t^j + \sum_{j=1}^s \delta_j I_j(t) \quad (3.17)$$

se denominará “componente estructural”, y los residuales ε_t en (3.16) 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 β_0 , pues, aunque no se hace explícito, esta constante está asociada con una variable explicativa que es constante e igual a 1 (ver Diebold [1999, pág. 90]). En consecuencia, el modelo (3.16) se modifica y queda

$$Y_t = \beta_0 + \sum_{j=1}^k \beta_j t^j + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t, \quad (3.18)$$

Implementación en R del modelo de componentes aditivas

Suponga que Y_t tiene una componente estacional de período s . Se declara como objeto `ts` en R, con frecuencia s con el comando `ts()` al incluir la opción, por ejemplo, `start=c(1995, 2)`, que indica que la serie mensual inicia en febrero de 1995. Ejemplo: `y = ts(y, frequency=12, start=c(1995, 2))`.

El primer paso consiste en generar las variables indicadoras $I_j(t)$, $j = 1, \dots, s$, $t = 1, \dots, T$. Luego calcular una matriz de ceros y unos, por ejemplo, It , de dimensión $T \times (s - 1)$, 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 1. De acuerdo a este valor inicial se genera la matriz It .

El siguiente código en R calcula la matriz $I_j(t)$ de dimensión $T \times s$.

```
#funcion para calcular la matriz It
```

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

En la librería `forecast` hay varios recursos y funciones para estimar y pronosticar con el modelo (3.18). Con la función `seasonaldummy` de la librería `forecast` se genera la matriz It . Adicionalmente, el comando `seasonaldummyf` genera otra matriz de ceros y unos, por ejemplo, It_p , con $s - 1$ columnas y con un número de filas igual al número de pronósticos, a partir del primer tiempo de pronósticos.

Adicionalmente, en `forecast` está implementada la función `tslm` que calcula el modelo de componentes pero con la tendencia lineal solamente.

```
fit <- tslm(y ~ trend + season)
plot(forecast(fit, h=20))
```

Ejemplo 3.7.1. *Considere la serie de producción de cemento Portland, trimestral, en toneladas por mil, entre Q1 1956 y Q3 1994, en Australia. El modelo de componentes que se asume para esta serie es de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{4-1} \delta_j I_t(t) + \varepsilon_t. \quad (3.19)$$

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

Los resultados del modelo lineal con variables estacionales (3.19) está en el Cuadro 3.4 siguiente. Una comparación entre la serie ajustada y la estimada se puede ver en la Figura (3.7). Con un R-cuadrado ajustado $\overline{R}^2 = 0.8966$, el ajuste resulta aceptable. También se observa que los coeficientes de las variables indicadoras $ItQ2$, $ItQ3$ no son significativos, sin embargo, $ItQ1$ sí, y por tanto, concluimos existe una componente estacional en la serie.

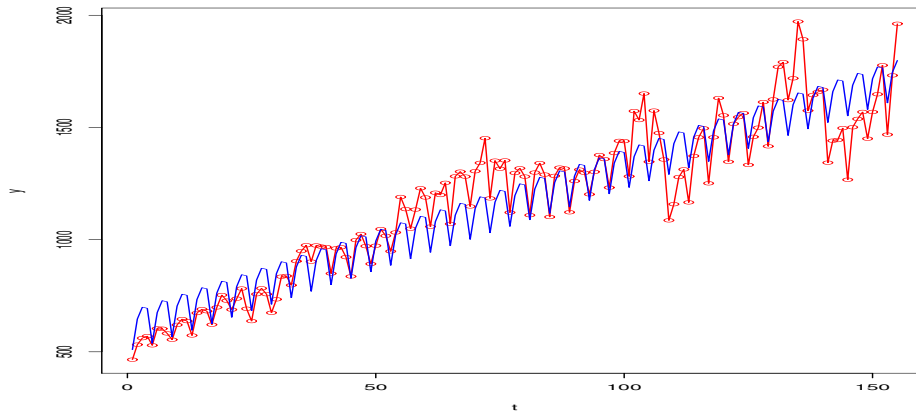


Figura 3.7: Ajuste con Variables Indicadoras y tendencia lineal. La componente estacional parece cambiar con el tiempo, es posiblemente un cambio estructural.

Tabla 3.4: 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

Código R 3.7.1.

```
# ----Ejemplo con producción de cemento
#-----cargar datos
# fpp2: Data for "Forecasting: Principles and Practice" (2nd Edition)
require(fpp2)
# Total quarterly production of Portland cement in Australia
# (in millions of tonnes) from 1956:Q1 to 2014:Q1.
y = qcement
# y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
frequency(y)
[1] 4
```

```
#----modelo con tendencia lineal y estacionalidad
#----con variables indicadoras estacionales
library(forecast)
t = seq(1,length(y))
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)
#----
library(xtable)
print(xtable(mod1), digits=4)
r1 = mod1$residuals
yhat1 = mod1$fitted.values
#-----
```

Ejercicio 3.7.1. *Cambiar el modelo (3.19) por uno con tendencia cuadrática. Cómo se compara con el modelo (3.19) utilizando las medidas de ajuste (AIC, BIC, R2 ajustado, MSE), y las pruebas F parciales?*

Ejemplo 3.7.2. *Continuando con el Ejemplo (3.7.1), ahora se estima el modelo (3.19), pero utilizando funciones trigonométricas. El procedimiento de estimación empieza con la generación de las variables explicativas $\sin(2\pi it/s)$, $\cos(2\pi it/s)$. Para calcular estas variables se utiliza la función `It.trig = fourier(y,k)`, la cual calcula los armónicos $\sin(2\pi jt/s)$, $\cos(2\pi jt/s)$, para cada $j = 1, 2, \dots, k$. Nótese que al ser $s = 4$ para, entonces el número de armónicos es como máximo $s/2 - 1 = 1$.*

$$Y_t = \beta_0 + \beta_1 t + \beta_1 \sin(2\pi t/4) + \beta_2 \cos(2\pi t/4) + \varepsilon_t. \quad (3.20)$$

Se incluye el cálculo de pronósticos a 8 trimestres con los modelos de variables indicadoras y de funciones trigonométricas. Los resultados están en la Tabla 3.5.

Código R 3.7.2.

```
#----modelo con tendencia lineal y estacionalidad
#----con variables trigonometricas
It.trig = fourier(y,1)
mod2 = lm(y ~ t + It.trig)
```

Tabla 3.5: Coeficientes estimados en el Modelo Lineal con variables trigonométricas

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.6104	0.0184	33.25	0.0000
ti	0.0074	0.0002	34.43	0.0000
It.trigS1-4	-0.0825	0.0129	-6.41	0.0000
It.trigC1-4	0.0156	0.0130	1.21	0.2294
It.trigC2-4	0.0303	0.0091	3.32	0.0012

```
summary(mod2)
yhat2 = mod2$fitted.values
```

En la Tabla 3.6 aparecen los valores estimados dentro de la muestra \hat{Y}_t , para los modelos lineal con indicadoras y lineal con trigonométricas. La conclusión que puede obtenerse es que en este caso es indiferente utilizar variables indicadoras ó funciones trigonométricas, dado que los pronósticos dentro de la muestra son muy similares.

Tabla 3.6: Seis primeros valores ajustados con variables Indicadoras y con Variables Trigonómicas

	lineal+indicadoras	lineal+trigonom
1	0.51	0.51
2	0.64	0.64
3	0.68	0.68
4	0.69	0.69
5	0.53	0.53
6	0.67	0.67

Ejemplo 3.7.3. *Un modelo para la demanda diaria de energía del mercado Nord Pool (países escandinavos) es el siguiente. La serie de precios es diaria para un período de 30/12/1996-26/04/2000.*

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

escribe el modelo como sigue.

$$Y_t = a + bt + s_t + S_t + e_t, \quad (3.21)$$

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

$$S_t = \alpha \cos\left(\frac{2\pi t}{365}\right) + \beta \sin\left(\frac{2\pi t}{365}\right).$$

En este modelo la componente con período anual se modela con funciones trigonométricas y la componente con período semanal con variables indicadoras. El siguiente código en R implementa el modelo (3.21).

Código R 3.7.3.

```
# 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)
```

3.8. Problemas

1. Con base en los datos de la tasa de empleo en Australia, en la Figura 3.2, ajuste varios modelos de tendencia, dejando como período de validación cruzada los 3 últimos años (36 meses). Los datos se pueden descargar de la página web mencionada en el Ejemplo 3.2.1.
2. En el Cuadro 3.7 se muestran los datos del total de habitantes en Medellín,

según los Censos desde 1905 hasta 2005, según el DANE (⁴). En Poveda

Tabla 3.7: 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

[1982] se planteó ajustar los datos de desde 1938 hasta 1973 mediante un modelo de componentes con tendencia logística, (3.3f).

$$Y_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)} + \varepsilon_t. \quad (3.22)$$

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

⁴http://es.wikipedia.org/wiki/Demografía_de_Medellín

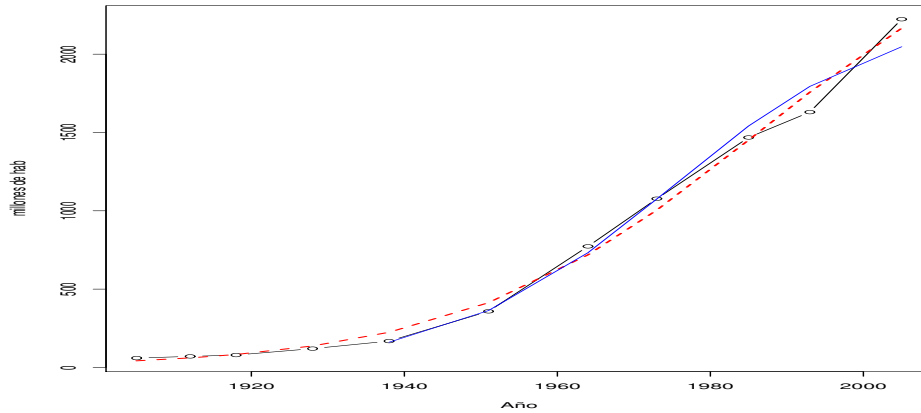


Figura 3.8: Ajuste Logístico a Población de Medellín

3. Consumo de agua mensual en el área metropolitana en metros cúbicos, por estrato, para los estratos 1,2,3,4,5,6, incluyendo el consumo total. Desde enero de 2002. En total son 7 series que se identificarán con los nombres : consumo 1, consumo 2,..., consumo 6, consumo total. Los datos están en el archivo consumo.dat, en la página del curso. Diseñe un plan inicial de análisis y descríbalos 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?.

CAPÍTULO 4

Pruebas de Estabilidad Estructural

4.1. Pruebas de Estabilidad Estructural en Modelos de Componentes

Para introducir la definición de Estabilidad Estructural en los modelos de Componentes, suponga el modelo de regresión lineal

$$Y_t = \beta_0 + \sum_{i=1}^k \beta_i X_{i,t} + \varepsilon_t, \quad t = 1, \dots, T, \quad (4.1)$$

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

$$Y_t = \beta_{0,t} + \sum_{i=1}^k \beta_{i,t} X_{i,t} + \varepsilon_t, \quad (4.2)$$

y quedaría definido un modelo con coeficientes variables. Nótese que (4.1) también se puede escribir como

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

con $\underline{X}_t = (1, X_{t,1}, X_{t,2}, \dots, X_{t,k})' \in \mathbb{R}^{1+k}$ y $\underline{\beta} \in \mathbb{R}^{1+k}$, vector de parámetros. En el caso que los parámetros varíen en el tiempo el modelo quedaría

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

Definición 4.1.1. *Estabilidad Estructural.* Si se asume el modelo de coeficientes variables (4.2), la estabilidad estructural del modelo se define como la hipótesis nula de que los coeficientes se mantengan constantes en el tiempo $\beta_{i,t} \equiv \beta_i$. La alterna es que algún $\beta_{i,t}$ cambie con t .

$$H_0 : \forall i = 1, \dots, k, \beta_{i,t} \equiv \beta_i \quad (4.3)$$

$$H_1 : \exists i, \beta_{i,t} \neq \beta_i.$$

Ejemplo 4.1.1. *Algunos ejemplos de cambio estructural en la literatura, debidos a intervenciones en fechas específicas, son*

- *EL nivel del rio Nilo, serie anual, con un cambio en 1898 con la primera represa de Ashwan.*
- *El número de muertes en accidentes automovilísticos mensuales en Gran Bretaña. La introducción del uso obligatorio del cinturón de seguridad en Enero 31 de 1983 cambió la serie mediante una intervención que disminuyó el nivel. En Harvey and Durbin [1986] se hace un análisis completo utilizando modelos estructurales.*
- *En Box and Tiao [1975, pag. 70] se exponen dos intervenciones en la ciudad de Los Angeles: la apertura 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 .*
- *El efecto de los atentados de 11 de Septiembre de 2001 en New York y Washington, en el número de turistas que ingresaron a varios países (ver Haywood and Randal [2006])*
- *El efecto de medidas de restricción de porte de armas de fuego en el número de homicidios por tal causa en varias ciudades.*

Dos técnicas para chequear el supuesto de estabilidad estructural son

1. La estimación recursiva de los parámetros del modelo. La idea básica de este tipo de pruebas es esta observación en Diebold [1999, pág. 92]

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

2. Las pruebas CUSUM, CUSUMSQ, MOSUM.

El objetivo al aplicar estas técnicas es asegurar que el modelo para pronósticos es estable. En caso de inestabilidad estructural, los pronósticos con un modelo de regresión global podrían no ser confiables. En este caso sería recomendable utilizar los procedimientos del Capítulo 5, como Holt-Winters, el modelo de espacio de estados en §5.5, pag.102. También un modelo de regresión no paramétrica como Loess.

4.2. Estimación recursiva de parámetros

El procedimiento de estimación recursiva de parámetros consiste en asumir el modelo de coeficientes variables (4.2)

$$Y_t = \beta_{0,t} + \sum_{i=1}^k \beta_{i,t} X_{i,t} + \varepsilon_t,$$

y, en lugar de utilizar todos los datos para la estimación, se utilizan los k primeros datos y se estiman los parámetros $\hat{\beta}_{i,k}$. Luego se utilizan los $k + 1$, y se estiman los $\hat{\beta}_{i,k+1}$, y así sucesivamente, hasta utilizar los T datos. Los parámetros obtenidos $\hat{\beta}_{i,t}$ $t = k, \dots, T$ se denominan Estimadores Recursivos.

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

Se toma un número inicial de datos k un poco mayor de $2 + s - 1$, por ejemplo, $k = 2 + s - 1 + 10$, para tratar de evitar fluctuaciones en los estimadores por poca información, y se corre el siguiente programa.

```
#-----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 4.2.1. *Con base en el Ejemplo 3.7.1, pag. 69, 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 (3.19), pag. 69, están en la Figura 4.1.*

Lo que puede concluirse de esta figura es que se aprecia un cambio en las trayectorias de los estimadores, por ejemplo en β_0 , en β_1 y en δ_3 . Por tanto, se puede concluir de este diagnóstico que existe evidencia de inestabilidad estructural en el modelo.

4.3. Pruebas CUSUM

Las pruebas CUSUM y CUSUM cuadrado fueron introducidas por Brown et al. [1975] para chequear la hipótesis nula de estabilidad estructural de un modelo de regresión lineal, dada en (4.3). Son herramientas estándar en el análisis de series económicas. El estadístico CUSUM de la prueba se define como sigue.

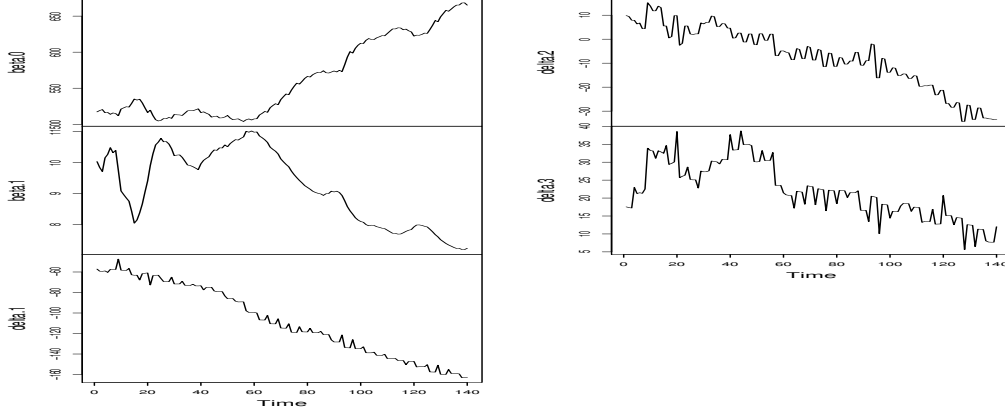


Figura 4.1: Estimadores Recursivos para el Modelo de Producción de Cemento

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

$$\hat{Y}_t = \hat{\beta}_{0,t} + \sum_{j=1}^k \hat{\beta}_{j,t} X_{j,t},$$

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

$$\hat{\varepsilon}_{t+1,t} = Y_{t+1} - \hat{Y}_t. \quad (4.4)$$

Si σ^2 es la varianza de error ε_t , se puede probar que

$$\hat{\varepsilon}_{t+1,t} \sim N(0, \sigma^2 r_t), \quad (4.5)$$

donde

$$r_t = 1 + \underline{X}'_{t+1} \left(\sum_{j=1}^t \underline{X}_j \underline{X}'_j \right)^{-1} \underline{X}_{t+1}, \quad (4.6)$$

con \underline{X}'_t vectores fila.

Definición 4.3.1 (Residuales Recursivos estandarizados). *Los residuales recursivos estandarizados se definen como*

$$W_{t+1,t} = \frac{\hat{\varepsilon}_{t+1,t}}{\hat{\sigma} \sqrt{r_t}}, \quad (4.7)$$

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

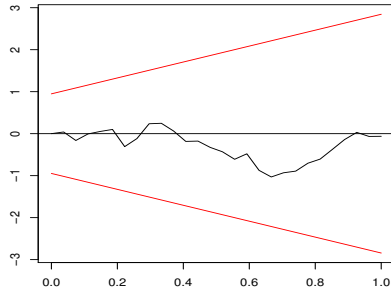
Definición 4.3.2 (CUSUM). La suma acumulada de los residuos recursivos estandarizados ó *CUSUM*, se define como

$$CUSUM_t = \frac{1}{\sigma_W} \sum_{i=1}^t W_{i+1,i}, \quad t = k, k+1, \dots, T-1, \quad (4.8)$$

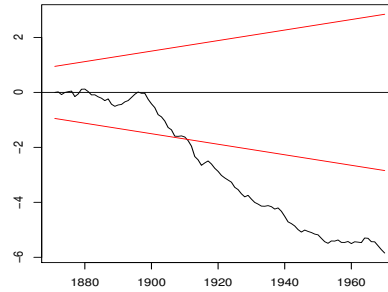
donde

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

La región de aceptación de la prueba CUSUM para la hipótesis (4.3) es el interior de una región (ó banda) limitada por 2 líneas rectas. Si la gráfica de los residuales CUSUM permanece dentro de esta banda entonces no se rechaza la hipótesis nula de estabilidad estructural. La expresión para las rectas es



(a) No se rechaza H_0



(b) Se rechaza H_0

Figura 4.2: Región de Rechazo de la Prueba *CUSUM*.

$$\pm a \left(\sqrt{T-k} + 2 \frac{t-k}{\sqrt{T-k}} \right), \quad t = k, k+1, \dots, T \quad (4.9)$$

con $a = 0.948$ para un nivel $\alpha = 0.05$. En la Figura 4.2 se muestran los dos casos posibles.

Definición 4.3.3. El estadístico *MOSUM*, basado en una media de residuales recursivos ó residuales MCO, sobre una ventana móvil, se define con las cantidades:

$m = T - k - 1$ y $0 < h < 1$, la proporción de datos en la ventana móvil, como

$$MOSUM_j = \frac{1}{\hat{\sigma}\sqrt{[hm]}} \left| \sum_{t=k+j+1}^{k+j+[hm]} W_{t+1,t} \right|, 0 \leq j \leq m - [hm] \quad (4.10)$$

La región de aceptación de la prueba MOSUM para la hipótesis (4.3) es el interior de una región (ó banda) limitada por 2 líneas horizontales. Si la gráfica de MOSUM permanece dentro de esta banda entonces no se rechaza la hipótesis nula de estabilidad estructural. Los detalles están en Chu et al. [1995, pag. 606]. En este artículo se hace la anotación de que la prueba MOSUM puede ser más sensible a los cambios estructurales por el hecho de estar basada en una ventana móvil y no en la totalidad de los datos.

Definición 4.3.4. *El estadístico CUSUMSQ, la suma acumulada de residuos recursivos cuadrados, se define como*

$$CUSUMSQ_t = \frac{\sum_{i=k}^t W_{i+1,i}^2}{\sum_{i=k}^T W_{i+1,i}^2}, \quad t = k, k+1, \dots, T \quad (4.11)$$

Si existe estabilidad estructural la gráfica de $CUSUMSQ_t$ versus t es una línea recta. Períodos en los cuales aumente la inestabilidad se notarán apartados de esta línea. Sin embargo, el siguiente comentario de A. Zeileis es importante ⁽¹⁾:

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

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

4.4. Implementación en R: la librería strucchange

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

Funciones de la librería

¹<https://stat.ethz.ch/pipermail/r-help/2005-January/064022.html>

1. **recresid**: Calcula los residuales recursivos estandarizados, $W_{t+1,t}$ en (4.7). La suma acumulada de éstos es el CUSUM. La instrucción en R para calcularlos es `rrc = recresid(y ~ t + It)`.

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

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

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

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

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

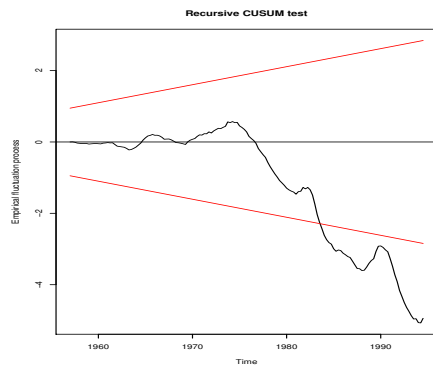
Ejemplo 4.4.1. Con base en el Ejemplo (3.7.1), pag. 69, de la serie de producción trimestral de cemento, la gráfica del estadístico CUSUM para el modelo con tendencia lineal y variables indicadoras estacionales (3.19), pag. 69, está en la Figura 4.3.

A partir de los resultados en la Figura 4.3 se rechazaría la hipótesis nula de estabilidad estructural en el modelo, ya que el estadístico CUSUM se sale de las bandas de confianza. Igual conclusión con la prueba MOSUM, que está en la Figura 4.4. La conclusión es que el modelo (3.19), pag. 69 detecta un cambio estructural en la serie.

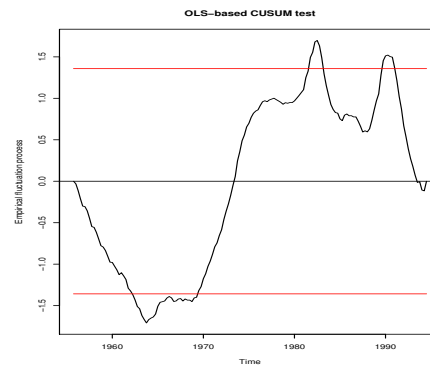
La prueba formal, con un nivel de significación de 5 %, confirma el rechazo de la hipótesis nula de estabilidad.

	estadistico	valor.p
Rec-CUSUM	1.703042	1.781954e-05

OLS-CUSUM	1.708886	5.814298e-03
Rec-MOSUM	2.246572	1.000000e-02
OLS-MOSUM	2.098502	1.000000e-02

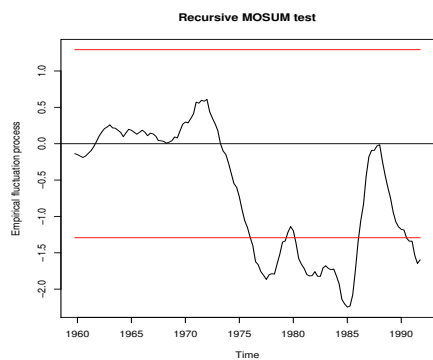


(a) prueba CUSUM residuales recursivos

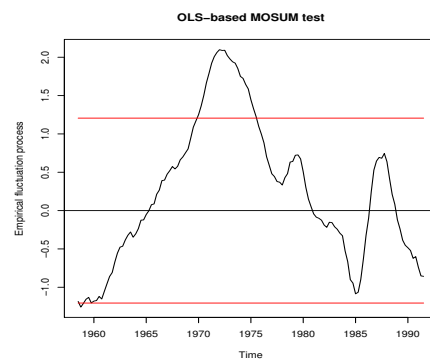


(b) prueba CUSUM residuales ols

Figura 4.3: Resultados prueba CUSUM gráfica



(a) prueba MOSUM residuales recursivos



(b) prueba MOSUM residuales mco

Figura 4.4: Resultados prueba MOSUM gráfica

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

Código R 4.4.1.

```

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

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

#----modelo con tendencia lineal y estacionalidad
T = length(y)
t = seq(1,T)
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)

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

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

```

```

"beta.1", "delta.1", "delta.2",
"delta.3")
plot.ts(parm, main="")

#----pruebas cusum graficas

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

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

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

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

#----pruebas cusum formales con valor p

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

```

Observación: La prueba CUSUM asume el supuesto de incorrelación de los errores, $Corr(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+k}) = 0$ para $k \neq 0$. Para decidir si se cumple este supuesto se utiliza el valor del estadístico Durbin-Watson (esta prueba se define en el capítulo 7).

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, $\hat{\sigma}_{\hat{\beta}_i}$, no son confiables. En el Ejemplo siguiente la prueba Durbin-Watson produce el resultado

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

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

4.5. Identificación del punto de quiebre

La función `breakpoints` permite identificar los puntos de quiebre ó cambio estructural. Suponga el modelo $Y_t = \underline{X}_t' \beta + \varepsilon_t$, $t = 1, \dots, T$. Se asumen m puntos de quiebre o de cambio estructural en los cuales los coeficientes cambian de un valor a otro, permaneciendo constantes entre estos puntos. Los puntos de quiebre son $t_j, j = 1, \dots, m$, y cumplen

$$Y_t = \underline{X}_t' \beta_j + \varepsilon_t, \quad t_{j-1} < t \leq t_j, \quad j = 1, \dots, m. \quad (4.12)$$

Los índices j de los puntos de quiebre se estima con la función `breakpoints`. Pero el número de puntos de quiebre, m , debe escogerse apriori. Por ejemplo, para que estime $m = 2$ puntos se utilizan las instrucciones:

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

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

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

```

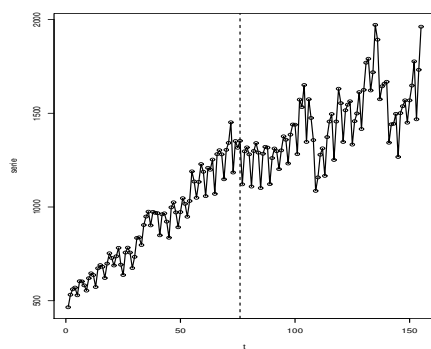
B= bp.n$extract.breaks(bp.n$RSS.table,breaks=2)

#---- 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)

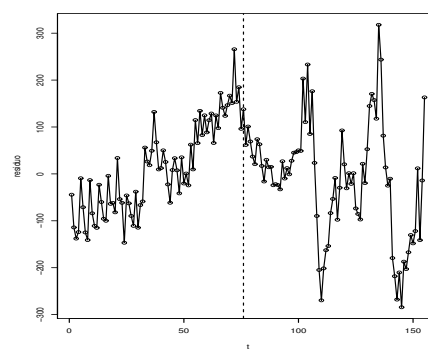
r1 = mod1$residuals
#----grafica de residuos OLS con punto de quiebre
plot(t,r1,type='o',ylab='residuo')
abline(v=B[1],lty=2,lwd=2)
abline(v=B[2],lty=2,lwd=2)

#----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)

```



(a) Con la Serie



(b) Con los residuos

Figura 4.5: Punto de Quiebre 1974-Q4 en la Serie y en los Residuos

El punto de quiebre se detecta en el cuarto trimestre de 1974 (la observación número 76). Ver el código en R anterior para la aplicación de la función `breakpoints`.

En este caso 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 (5.5). 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.

CAPÍTULO 5

Suavizadores, Espacio de Estados y Redes neuronales

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

5.1. Regresión Local Loess

La Regresión Loess, denominada en R inicialmente Lowess por sus siglas en inglés: *Locally weighted scatterplot smoothing*, es un modelo de regresión no paramétrica que regresa y_i versus x_i , pero no asume un modelo global fijo, es decir, no asume un intercepto y una pendiente fijas, sino variables, de manera local. Local significa una ventana móvil que contiene un número determinado de datos de la variable

explicativa. La motivación para utilizar Loess es que se puede estimar la tendencia de la serie T_t sin utilizar un modelo global paramétrico, que en muchos casos resulta muy restrictivo.

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

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \dots, n. \quad (5.1)$$

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

1. Se escoge q con $1 \leq q \leq n$.
2. Se escogen los q valores x_i más cercanos a x .
3. Defina $\lambda_q(x)$ la distancia máxima entre x y los q x_i escogidos. Luego

$$0 \leq \frac{|x_i - x|}{\lambda_q(x)} \leq 1.$$

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

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

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

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

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

```
T = length(y)
t = seq(1, T)
yw = loess(y ~ t, span = 0.75, degree = 1,
control = loess.control(surface = "direct"))
```

El parámetro `span = alfa` es un número $\alpha \in (0, 1)$, que determina el ancho de la ventana en términos de un porcentaje de datos, $q = \lfloor \alpha n \rfloor$. Mientras más cercano a 1 más suaviza porque se asemeja a una regresión lineal ó cuadrática global.

El parámetro `degree = 1, 2` determina si la regresión local es lineal ó cuadrática.

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

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

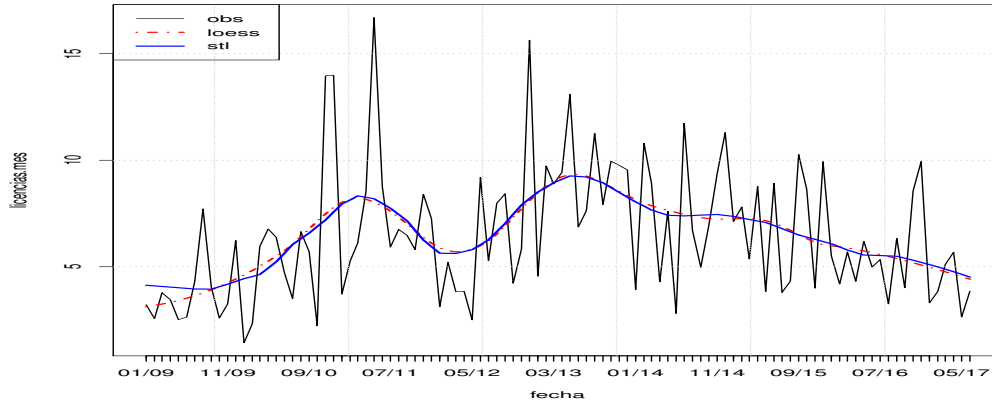


Figura 5.1: Suavizamiento Loess de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

Prueba Mann-Kendall para detectar tendencia monótona

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

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

donde $\text{signo}(x) = \pm 1$ si $x \gtrless 0$, y $\text{signo}(0) = 0$. Bajo H_0 , S se distribuye aproximadamente Normal, con media cero. Un valor positivo de S indica una tendencia creciente, y uno negativo una tendencia decreciente. La prueba está implementada en la librería `kendall`.

Este es el código en R para suavizamiento con Loess, ver Figura 5.3.

Código R 5.1.1.

```
# Analisis de número de unidades aprobadas para vivienda
# 88 municipios 2009/01 - 2017/06
# leer
G = read.table("series_histo_88_mpio_jun17.prn",
header = TRUE, stringsAsFactors = FALSE)
attach(G)

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

# Puede asumirse una tendencia global?
# detectar tendencia monotona con la prueba Mann-Kendall
# Ho: no hay tendencia, Ha: tendencia monotona
library(Kendall)
  MannKendall(y)
tau = 0.179, 2-sided pvalue =0.0078129
# como valor p menor de 0.05 rechaza la nula
#-----utilizar loess para estimar tendencia
m = 12
n = length(y)
yi = ts(y[1:(n-m)],frequency=12)
yf = ts(y[(n-m+1):n],frequency=12)

t = seq(1,(n-m))
yw = loess(yi ~ t, span=0.5,
control = loess.control(surface = "direct"))
yest.yw = yw$fitted
```

5.2. Método híbrido Loess más componente estacional

En el modelo de componentes con variables indicadoras (3.16), pág 68,

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

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

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

donde $g(t)$ es la estimación con Loess de la tendencia que reemplaza a $\beta_0 + \sum_{j=1}^k \beta_j t^j$. Nótese que se puede estimar el conjunto total de coeficientes estacionales δ_j , porque no hay, estrictamente, un problema de multicolinealidad.

Debido a que el modelo híbrido (5.3) es un modelo aditivo, el procedimiento de estimación se puede realizar en dos pasos. Estimar por Loess la tendencia, obteniendo $\hat{T}_t = \hat{g}(t)$, y luego estimar la componente estacional y el error estructural mediante regresión lineal con respecto a la variable $Y_t - \hat{T}_t$. El código R 5.2.1 siguiente muestra cómo realizar el procedimiento. En la Figura 5.3 se muestran los resultados.

Código R 5.2.1.

```
#-----pronostico modelo híbrido
#           pronostico con Loess
tt = seq((n-m+1), n, 1)
pr.yw = predict(yw, data.frame(t = tt))
#-----pronostico de la parte estacional
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) }

Itt = dummy(12,m,7)
pr.st = predict(mod1,data.frame(It=I(Itt)))
pr.y = pr.yw+pr.st

```

5.3. Método de descomposición STL

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

La estimación de la componente estacional \hat{S}_t por este método no es equivalente a la que se obtiene mediante los modelos paramétricos de componentes estacionales ó trigonométricas.

Además, este método permite calcular pronósticos con la función `forecast` de la librería del mismo nombre. Se aplicará para obtener las componentes de la serie Y_t y realizar un pronóstico.

Ejemplo 5.3.1. *Al aplicar el método STL a los datos del Ejemplo (5.1.1) se obtiene la descomposición en la Figura (5.2). El tercer panel es la tendencia estimada por Loess. El segundo la componente estacional. El tercero es la componente del error. La componente estacional tiene un rango estimado $[-2, 6] + 6.8 = [4.8, 12.8]$, (se añade $\min(Y_t) = 6.83$ para colocar la serie a nivel de la serie Y_t), y la Y_t un rango $[6.8, 31]$. La componente estacional tiene un rango que abarca una parte importante del rango de la serie Y_t .*

```

#-----descomposicion y pronostico con stl
y.stl = stl(yi,"per")

```

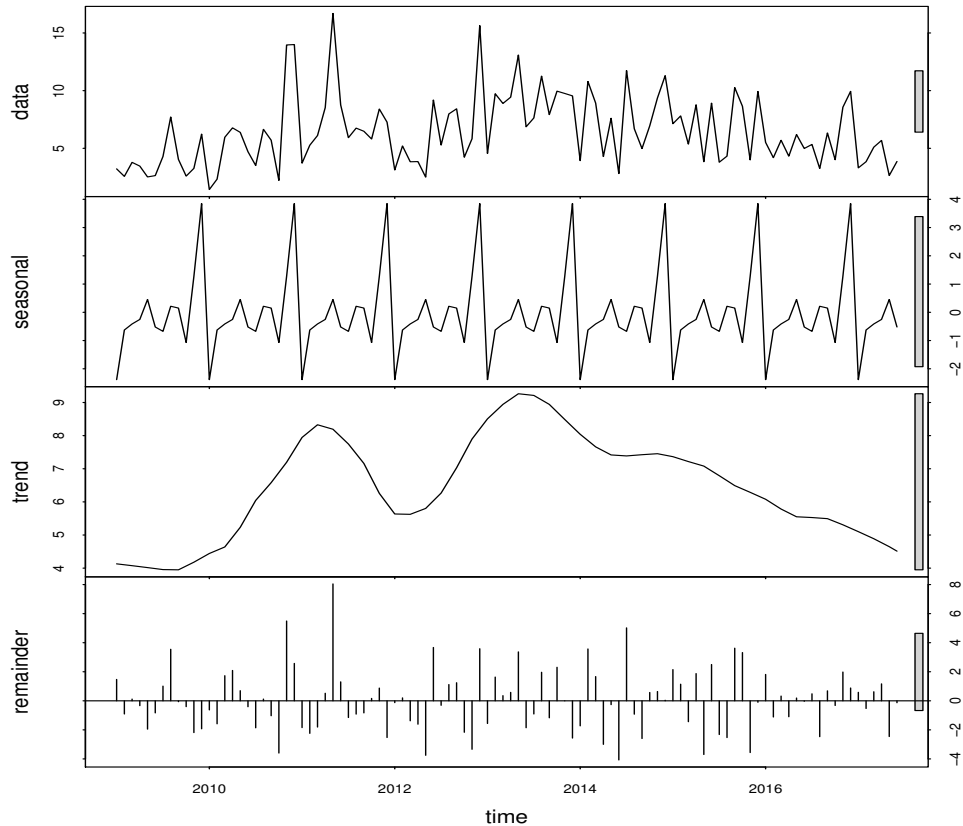


Figura 5.2: Descomposición STL número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

```
plot(y.stl)
St.stl = y.stl$time.series[,1]
Tt.stl = y.stl$time.series[,2]
#-----pronostico
pr.stl = forecast(y.stl,method="ets",h=m)$mean
#-----grafica pronosticos
plot(tt,yf,type='b',lwd=1,ylim=c(8,21))
lines(tt,pr.y,col='red',lwd=1)
lines(tt,pr.stl,col='blue',lwd=1)
legend("topleft", c("Obs","hibrido:Loess+Ind","STL"),
pch = c(1, 3, 5),col = c("black","red","blue"))
```

```
#-----compara calidad pronosticos
pr.y = ts(pr.y,frequency=12,start=c(2016,06))
pr.stl = ts(pr.stl,frequency=12,start=c(2016,06))
yf = ts(yf,frequency=12,start=c(2016,06))
M = rbind(accuracy(pr.y,yf),accuracy(pr.stl,yf))
rownames(M) = c("hibrido:Loess+Ind","STL")
(M)
```

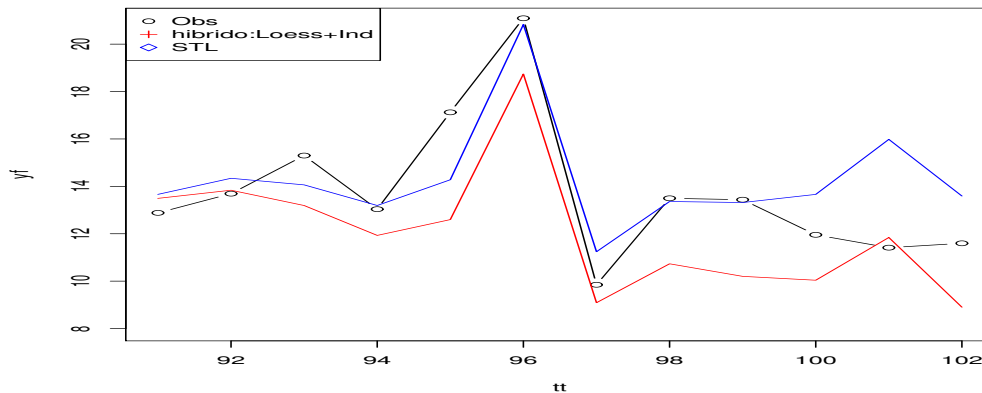


Figura 5.3: Pronóstico modelo híbrido Loess + Indicadoreas, serie de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

Tabla 5.1: Calidad de pronósticos para los modelos Loess+Estacional versus STL

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
hibrido:Loess+Ind	1.69	2.27	1.89	11.82	13.41	0.04	0.78
STL	-0.55	1.84	1.32	-5.82	10.46	0.37	0.64

Otras librerías ó funciones en R que también realizan la descomposición de T_t, S_t, ε_t , son, por ejemplo,

1. La función `decompose()`.
2. La librería `timesac` tiene la función `decomp()` que realiza la descomposición incluyendo una componente autoregresiva y otra para fechas de intervenciones, $Y_t = T_t + S_t + R_t + TA_t + \varepsilon_t$.

3. La función `ma.filter()` de la librería `rmaf` (Refined Moving Average Filter, Qiu [2015]).
4. La librería `tsR`
5. Página web: discusión extensa de descomposición
http://course1.winona.edu/bdeppa/FIN20335/Handouts/Time_Series_Decomposition.html

5.4. Modelo de Holt-Winters

En el modelo de componentes con variables indicadoras (3.16), pág 68,

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

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

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

El modelo Holt-winters asume que los parámetros varían de acuerdo a marchas aleatorias, definidas como

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

donde $\varepsilon_{\beta_{0,t}} \sim iidN(0, \sigma_{\beta_0}^2)$, $\varepsilon_{\beta_{1,t}} \sim iidN(0, \sigma_{\beta_1}^2)$, $\varepsilon_{\delta_{j,t}} \sim iidN(0, \sigma_{\delta_j}^2)$, son variables aleatorias distribuídas iid Normales. Se dice que los parámetros en (5.4) “varían lentamente”.

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

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

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

Definición 5.4.1. *El algoritmo Holt-Winters consiste en el sistema de ecuaciones recursivas para μ_t, β_t y S_t , dado por*

$$\begin{aligned}\mu_t &= \alpha(Y_t - S_{t-s}) + (1 - \alpha)(\mu_{t-1} + \beta_{t-1}), \\ \beta_t &= \beta(\mu_t - \mu_{t-1}) + (1 - \beta)\beta_{t-1}, \\ S_t &= \gamma(Y_t - \mu_t) + (1 - \gamma)S_{t-s},\end{aligned}\tag{5.5}$$

donde $t = 2, \dots, T$.

Las constantes α, β, γ , cada una en $(0, 1)$, son los parámetros de suavizamiento (*tunnig parameters*). Se escogen de manera que se minimize el MSE. O también, por tanteo. Según Abraham and Ledolter [2000, pag. 169],

“Si las componentes de tendencia y estacionalidad cambian rápidamente, las constantes α_j deben tomarse cercanas a 1.0. Si las componentes son estables y varían lentamente, se escogen cercanas a 0.0”.

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

$$Y_t = \sum_{j=1}^s \delta_j I_j(t) + \varepsilon_t,\tag{5.6}$$

$$\hat{\varepsilon}_t = \beta_0 + \beta_1 t + \eta_t, \quad t = 1, 2, \dots, T.\tag{5.7}$$

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

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

$$\hat{Y}_t = \mu_t + \beta_t + S_t.$$

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

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

En: Harvey [2006], pag. 358

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

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

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

Ejemplo 5.4.1. *Continuando con el Ejemplo 5.1.1, de los datos del DANE sobre estadísticas de edificación, correspondientes a número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06. Se puede observar una comparación entre Loess y Holt-Winters en la Figura (5.4)*

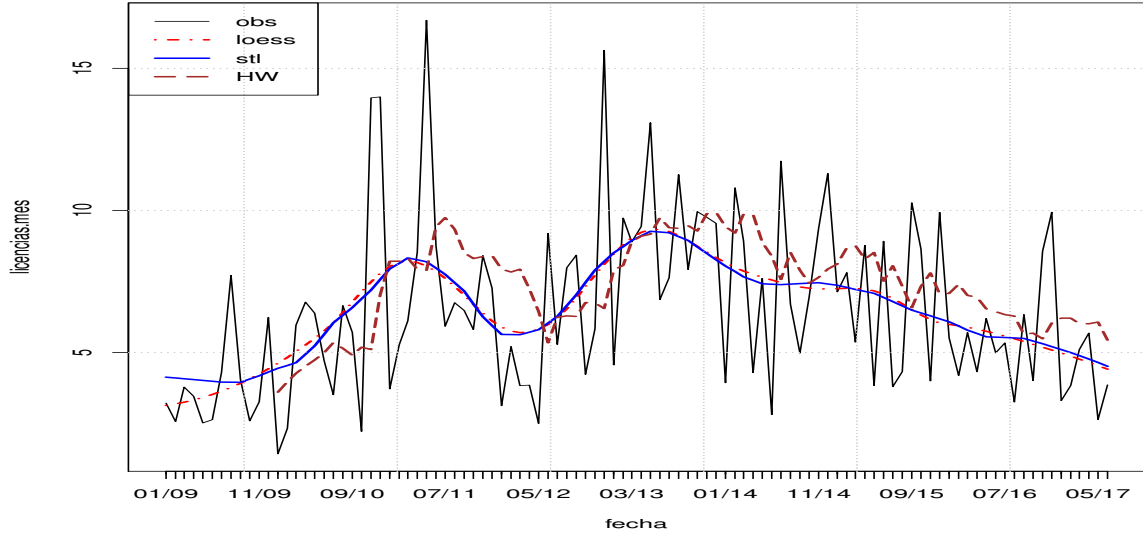


Figura 5.4: Estimación de la tendencia con los Métodos Loess, STL y Holt-Winters

5.5. Modelos de Espacio de Estados

Una serie $\{Y_t, t = 0, 1, \dots\}$ sigue un modelo lineal Gaussiano de Espacio de Estados si cumple el sistema de ecuaciones, ver [Helske, 2017, pag. 2]

$$\underline{Y}_t = Z_t \underline{\mu}_t + \underline{c}_t + \underline{\epsilon}_t, \quad (5.8)$$

$$\underline{\mu}_t = T_t \underline{\mu}_{t-1} + \underline{d}_t + R_t \underline{\eta}_t, \quad (5.9)$$

$$\underline{\mu}_0 \sim N_m(\underline{\mu}_0, P_0).$$

donde $\underline{\mu}_t$ es la variable latente.

La formulación es vectorial, pero aplica a casos escalares. Con

- $\underline{Y}_t \in \mathbb{R}^p, \underline{\mu}_t \in \mathbb{R}^m, \underline{\epsilon}_t \in \mathbb{R}^p, \underline{\eta}_t \in \mathbb{R}^q$
- Las matrices $Z_t \in \mathbb{R}^{p \times m}, T_t \in \mathbb{R}^{m \times m}, R_t \in \mathbb{R}^{m \times q}$ pueden ser independientes de t .
- $P_0 \in \mathbb{R}^{m \times m}$ es una matriz de varianza-covarianza.
- Los errores $\epsilon_t \sim iid N_p(0, H), \eta_t \sim iid N_q(0, Q)$, se asumen Normales multi-variados, independientes entre ellos.

- Las matrices $H \in \mathbb{R}^{p \times p}$, $Q \in \mathbb{R}^{q \times q}$ son matrices de varianzas-covarianzas.

El modelo anterior es general y se pueden definir varios casos particulares. Ejemplos de modelos de espacio de estados en ⁽¹⁾

Existen varias librerías en R para estimar modelos de Espacio de Estados. La función `StructTS` y la librería `d1m`. Entre muchas otras!.

El modelo de tendencia local (*local level model*), asume que se observa Y_t pero la serie tiene un nivel local μ_t que no puede observarse porque está distorsionado por un ruido. El nivel local es una marcha aleatoria.

$$Y_t = \mu_t + \epsilon_t, \quad (5.10)$$

$$\mu_t = \mu_{t-1} + \alpha_t, \quad t = 1, 2, \dots, T,$$

con $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$, $\alpha_t \sim iidN(0, \sigma_\alpha^2)$. Los parámetros son $\mu_0, \sigma_\epsilon^2, \sigma_\alpha^2$. Este modelo se puede estimar con la instrucción

```
mod.loc <- StructTS(y, type = "level")
print(mod.loc$coef)
```

El modelo de tendencia lineal local (*local linear level model*), asume que se observa Y_t pero la serie tiene un nivel local μ_t más una tendencia β_t . La tendencia es una marcha aleatoria.

$$Y_t = \mu_t + \epsilon_t, \quad (5.11)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \alpha_t,$$

$$\beta_t = \beta_{t-1} + \delta_t, \quad t = 1, 2, \dots, T,$$

con $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$, $\alpha_t \sim iidN(0, \sigma_\alpha^2)$, $\delta_t \sim iidN(0, \sigma_\delta^2)$. Los parámetros son $\mu_0, \sigma_\epsilon^2, \sigma_\alpha^2, \sigma_\delta^2$. Este modelo se puede estimar con la instrucción

```
mod.t <- StructTS(y, type = "trend")
print(mod.t$coef)
```

El modelo estructural Gaussiano con componente estacional, BSM (*Basic Structural Model*), se define como

$$Y_t = \mu_t + S_t + \epsilon_t, \quad (5.12)$$

¹<https://github.com/pteetor/StateSpaceModels>.

$$\begin{aligned}\mu_t &= \mu_{t-1} + \beta_{t-1} + \alpha_t, \\ \beta_t &= \beta_{t-1} + \delta_t, \\ S_t &= S_{t-1} + \dots + S_{t-s+1} + \gamma_t,\end{aligned}$$

con $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$, $\alpha_t \sim iidN(0, \sigma_\alpha^2)$, $\delta_t \sim iidN(0, \sigma_\delta^2)$, $\gamma_t \sim iidN(0, \sigma_\gamma^2)$. Los parámetros son $\mu_0, \sigma_\epsilon^2, \sigma_\alpha^2, \sigma_\delta^2, \sigma_\gamma^2$. Este modelo se puede estimar por máxima verosimilitud con la instrucción

```
mod.bsm <- StructTS(y, type = "BSM")
print(mod.bsm$coef)
```

Ejemplo 5.5.1. Con el Ejemplo 5.1.1, de los datos del DANE sobre estadísticas de edificación, correspondientes a número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06. Se ajustó los tres modelos estructurales a los datos. El modelo BSM resultó claramente superior tanto en calidad de ajuste como en calidad de pronósticos, como lo muestran las Tablas 5.2 y 5.3.

Tabla 5.2: Medidas de calidad de ajuste, tres modelos estructurales

	R2-ajus	MSE	logAIC	logBIC
local	0.492	11.768	2.498	2.581
tendencia	0.484	11.946	2.524	2.635
bsm	0.725	6.378	1.907	2.046

Tabla 5.3: Medidas de calidad de pronósticos, tres modelos estructurales

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
local	-0.898	2.991	2.499	-10.526	18.767	0.035	0.850
tendencia	-0.751	2.913	2.392	-9.346	17.808	0.012	0.813
bsm	-0.746	2.303	1.728	-7.473	13.194	0.243	0.775

El Modelo de Espacio de Estados, ETS (A, A, A) definido en Hyndman et al. [2008, pag. 45, sec. 3.4.3], es similar al modelo BSM, y también incorpora la hipótesis de variación lenta de acuerdo a marchas aleatorias. Los parámetros (α, δ, γ) controlan la varianza de un único error aleatorio ϵ_t , que afecta cada ecuación del sistema. Concretamente,

$$Y_t = \mu_t + \beta_t + S_{t-s} + \epsilon_t, \quad (5.13)$$

$$\begin{aligned}\mu_t &= \mu_{t-1} + \beta_{t-1} + \alpha\epsilon_t, \\ \beta_t &= \beta_{t-1} + \delta\epsilon_t, \\ S_t &= S_{t-s} + \gamma\epsilon_t,\end{aligned}$$

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

La estimación de los parámetros se realiza mediante la función `ets()` de la librería `forecast`.

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

Es un método que hace parte de la familia de modelos de Espacio de Estados, introducida en Hyndman et al. [2002] y Hyndman et al. [2008]. La convención es que cada modelo se describe con tres letras escogidas de N, A, M y Z. En todos los casos, N = ninguno, A = aditivo, M = multiplicativo y Z = seleccionado automáticamente.

- La primera letra indica el tipo de error (A, M, Z);
- la segunda letra indica el tipo de tendencia (N, A, M o Z);
- y la tercera letra indica el tipo de estacionalidad (N, A, M o Z).

Por ejemplo, ANN es un suavizado exponencial simple con errores aditivos.

5.6. Redes neuronales autoregresivas

Si se tiene la serie de tiempo $Y_t, t = 1, \dots, T$, los valores Y_{t-k} para $k = 1, 2, \dots$ se denominan rezagos de orden k .

La forma general de una red neuronal de una capa oculta (hidden layer), con q nodos en la capa oculta y p nodos de entrada se definió en (2.6). Pero la red neuronal

autoregresiva se define con rezagos Y_{t-k} de orden $k \leq p$.

$$Y_t = \alpha_0 + \sum_{j=1}^q \alpha_j f \left(\beta_{0,j} + \sum_{k=1}^p \beta_{j,k} Y_{t-k} \right) + \varepsilon_t, \quad (5.14)$$

donde $f(x) = (1 + e^x)^{-1}$. Para el caso de $p = 2, q = 1$ se tiene

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}}} + \varepsilon_t,$$

El modelo (5.14) se indica por $NNAR(p, q)$. Cuando se incluyen P rezagos estacionales de la forma $Y_{t-s}, Y_{t-2s}, \dots, Y_{t-Ps}$, donde s es el período de la componente estacional, se indica por $NNAR(p, P, q)[s]$. Ejemplos

■ $NNAR(2,1)$:

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}}} + \varepsilon_t,$$

■ $NNAR(3,2)$:

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_{1,0} + \beta_{1,1} Y_{t-1} + \beta_{1,2} Y_{t-2} + \beta_{1,3} Y_{t-3}}} + \frac{\alpha_2}{1 + e^{\beta_{2,0} + \beta_{2,1} Y_{t-1} + \beta_{2,2} Y_{t-2} + \beta_{2,3} Y_{t-3}}} + \varepsilon_t,$$

■ $NNAR(2,1,1)[12]$:

$$Y_t = \alpha_0 + \frac{\alpha_1}{1 + e^{\beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \delta_1 Y_{t-12}}} + \varepsilon_t,$$

Los modelos $NNAR(p, m, q)[s]$ están implementados en la librería `forecast` con la función `nnetar(y, p, P = 1, size)`, donde `size = q` es el número de nodos en la capa oculta. Es posible utilizarla de la forma `nnetar(y)`, en cuyo caso los órdenes $p, P, size$ se escogen con valores por defecto, que son $P = 1, p$ se escoge mediante un modelo autoregresivo de orden p , $AR(p)$, que mejor ajuste según el AIC, y $q = (P + p + 1)/2$.

La función incluye la transformación Box-Cox

$$Y_t^{(\lambda)} = \begin{cases} (Y_t^\lambda - 1)/\lambda, & \lambda \neq 0, \\ \log(Y_t), & \lambda = 0. \end{cases}$$

que se aplica antes de estimar el modelo. Colocando `nnetar(y, p, P = 1, size, lambda=0)`, se excluye su aplicación.

Ejemplo 5.6.1. Con el Ejemplo 5.1.1, de los datos del DANE sobre estadísticas de edificación, correspondientes a número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06. Se ajustó un modelo NNAR a los datos.

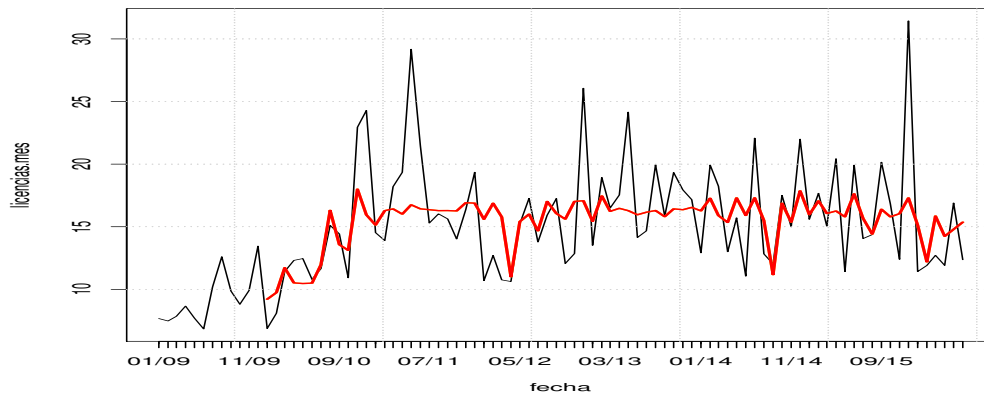


Figura 5.5: Ajuste con NNAR(3,1,2)[12], serie de número de unidades aprobadas para vivienda, mensuales, en 88 municipios en el período 2009/01 - 2017/06

```
require(forecast)
y.nnar = nnetar(yi, lambda=0)
print(y.nnar)
Series: yi
Model: NNAR(3,1,2)[12]
sigma^2 estimated as 0.04706
#-----valores ajustados
yhat.nnar= fitted(y.nnar)
```

Los pronósticos se calculan con la función `forecast`.

```
pr.nnar = forecast(y.nnar, h=m)$mean
pr.nnar = ts(pr.nnar, frequency=12, start=c(2016, 06))
```

	R2-ajus	MSE	logAIC	logBIC
local	0.492	11.768	2.498	2.581
tendencia	0.484	11.946	2.524	2.635
bsm	0.725	6.378	1.907	2.046
NNAR	0.280	14.578	2.831	3.223

Trabajo No 1 de Técnicas de Pronósticos

Pronósticos de Series de Tiempo con modelos para la Tendencia y la Estacionalidad

Octubre de 2021

Características del Trabajo

El trabajo para cada grupo consiste de 4 puntos enunciados en la sección siguiente. Los puntos se deben resolver con la serie asignada en Tabla 5.5. Las series se describen en la última sección y se indica en dónde se encuentran los datos. Tiene un valor de 33 % de la nota definitiva.

- **Presentación.** Elaborar el reporte en formato Word o Latex, con formato final pdf. Con el nombre de los integrantes del grupo y el número del grupo en el encabezado.
- **Detalles.** Desarrollar cada punto por separado. Las gráficas y las tablas deben contener una numeración, información resumida (caption), que permita entender rápidamente de qué se trata.
- **Entrega.** Cargar el informe pdf en el Moodle antes de la fecha de entrega. El archivo pdf debe identificarse con los apellidos de los autores, por ejemplo, “alvarez.correa.pdf”. Tiene un plazo de entrega de 1 semana (en días hábiles) a partir de la fecha de publicación en Moodle.

Puntos del Trabajo

Defina la estrategia de validacion cruzada escogiendo una parte de los datos para ajustar y otra para comparar los pronósticos. En series mensuales use 12 meses, en trimestrales use 8 trimestres. En frecuencia diaria use 15 días.

1. (25/25) Estime el modelo (C) asignado en la Tabla 5.4. Reporte las tablas de parametros estimados, estadísticos t y F, valores p. Reporte la gráfica de los valores ajustados \hat{Y}_t , versus los datos observados Y_t , para el período de entrenamiento.
2. (25/25) Estime el modelo (EE) asignado en la Tabla 5.4. Reporte la información que se obtenga del procedimiento de estimación aplicado: parámetros, tipo de red NNAR, varianzas del modelo. Reporte la gráfica de los pronósticos y de los datos observados, para el período de comparación.
3. (25/25) Reporte MSE, AIC, BIC y R-cuadrado ajustado para los 2 modelos C y EE. Concluya cuál modelo ajustó mejor.
4. (25/25) Calcule los pronósticos para la validación cruzada con los 2 modelos C y EE. Reporte MAPE, RMSE, U-Theil para éstos. Concluya cuál modelo pronostica mejor.

Opciones	Modelos
(C) Componentes	(1.1) cuadrático + indicadoras (1.2) cúbico + indicadoras (1.3) exponencial (lineal + indicadoras) (1.4) exponencial (cuadrático + indicadoras)
(EE) Espacio de Estados, Suavizadores	(2) Híbrido Loess + indicadoras (3) Holt-Winters (5.5), pag 100 (4) Modelo ETS-AAA (5.13), pag 104 (5) Espacio de Estados BSM (5.12), pag 103 (6) Red neuronal autoregresiva NNAR (5.14), pag 106

Tabla 5.4: Modelos para Pronosticar

Tabla 5.5: Asignación de modelos y serie por grupo

Grupo = Serie No	C	EE	Estudiantes
1 ✓	1.1	2	–
2 ✓	1.1	3	
3 ✓	1.1	4	–
4 ✓	1.1	5	–
5 ✓	1.1	6	–
6 ✓	1.2	2	–
7 ✓	1.2	3	–
8 ✓	1.2	4	–
9 ✓	1.2	5	
10 ✓	1.2	6	
11 ✓	1.3	2	
12 ✓	1.3	3	
13 ✓	1.3	4	–
14 ✓	1.3	5	
15 ✓	1.3	6	
16 ✓	1.4	2	
17 ✓	1.4	3	
18 ✓	1.4	4	
19 ✓	1.4	5	
20 ✓	1.4	6	

Series para el Trabajo

La serie asignada a cada integrante está en la lista siguiente. Los archivos están en Moodle en la carpeta del tema del trabajo ó se pueden cargar desde las librerías indicadas en cada serie.

(1) producción total vino

Instrucciones de lectura. Archivo en Moodle

```

Wine sales
#Australia, monthly
#Jan 1980-Jun 1994
#Total --- Total Wine sales (th. of litres)
#by wine makers in bottles <= 1L.
#ABS (Australian Bureau of Statistics)
# Time Series Data Library
Vino = read.table("Wine.dat", header = TRUE, stringsAsFactors = FALSE)
attach(Vino)
y = ts(Vino$Total,frequency=12,start=c(1980,01))

```

(2) Demanda diaria energía eléctrica en Bogotá,

```

Empresa Codensa S.A., en Gwh,
entre 1995-07-20 y 1997-10-23.
Archivo en Moodle
S = read.table("codensa.diaria.dat",
header = TRUE, stringsAsFactors = FALSE)

```

(3) precio de carne pavo

```

turkey.price.ts: Monthly Average Turkey Price, January 2001 to April
his time series shows the average retail price of turkey in the Unit
S = read.table("turkey.price.dat",
header = TRUE, stringsAsFactors = FALSE)
attach(S)

```

(4) La serie h02,

```

en la librería fpp,
se describe como: ``Monthly cortecosteroid
drug sales in Australia from 1992 to 2008''.
Se carga con las instrucciones
require(fpp)
y = h02
# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1992/1/1"),
length.out = length(y), by = "months")

```

(5) producción vino tinto

Instrucciones de lectura. Archivo en Moodle

Wine sales

#Australia, monthly

#Jan 1980-Jun 1994

#Total --- Total Wine sales (th. of litres)

#by wine makers in bottles <= 1L.

#ABS (Australian Bureau of Statistics)

Time Series Data Library

Vino = read.table("Wine.dat", header = TRUE, stringsAsFactors = FALSE)

attach(Vino)

y = ts(Vino\$Red, frequency=12, start=c(1980, 01))

(6) Empleo masculino Suiza.

Time series from <http://www.dataserie.org>,

a comprehensive and up-to-date

collection of open data from Switzerland

require(dataseries)

D=dataseries::ds("ch_comb_jobs.596.tot.1.0")

y = D\$ch_comb_jobs.596.tot.1.0

names(y) = "empleo total hombres trimestral Suiza"

fechas = D\$time

y=ts(y, frequency=4)

ts.plot(y)

(7) Turistas Suiza.

Time series from <http://www.dataserie.org>,

a comprehensive and up-to-date

collection of open data from Switzerland

require(dataseries)

D=dataseries::ds("ch_comb_jobs.596.tot.1.0")

y = D\$ch_comb_jobs.596.tot.1.0

names(y) = "empleo total hombres trimestral Suiza"

```

fechas = D$time
y=ts(y,frequency=4)
ts.plot(y)

```

(8) accidentes fatales auto en UK,

```

# leer datos de la serie Seatbelts:
UKDriverDeaths is a time series giving the monthly
totals of car drivers in Great Britain killed or
seriously injured Jan 1969 to Dec 1984.
Compulsory wearing of seat belts was
introduced on 31 Jan 1983.
data(Seatbelts)
# car drivers killed.
y = Seatbelts[,2]
ts.plot(y)

```

(9) La serie IPC,

```

Variación porcentual mensual del Indice de precios
al consumidor, 2000-2012.
library(readxl)
res <- read_excel("IPCporcentual.xlsx", 1) # lee el primer libro
attach(res)
y = ts(IPC, frequency=12,start=c(2000,01))
ts.plot(y)

```

(10) La serie PIB Colombia,

```

PIB Trimestral Colombia, Sin desestacionalizar
a precios constantes 2005.
G = read.table("PIBtrimestral.dat", header = TRUE,
stringsAsFactors=FALSE)
attach(G)
fechas = seq(as.Date("2000/1/1"),
length.out = length(pib), by = "quarter")
y = ts(pib,frequency=4,start=c(2000,01))

```


(11) precio carne hamburguesa

```

A time series objects consiting of average monthly
retail prices per pound of ham in the United States
between January 2001 and April 2008
S = read.table("ham.price.dat",
header = TRUE, stringsAsFactors = FALSE)
attach(S)

```

(12) turismo Chipre

```

# leer turismo chipre: 1 serie
tu.chipre = read.table("turismo.chipre.dat",
header = TRUE, stringsAsFactors = FALSE)
y = tu.chipre$y
# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1979/1/1"),
length.out = length(y), by = "month")

```

(13) indice vivienda USA Schiller San Diego

```

# leer indice SDXRSA indice Schiller
# sobre variación de los precios de vivienda
# en la ciudad de San Diego, EUA.
# ver https://fred.stlouisfed.org/series/SDXRSA
D = read.csv("SDXRSA.csv", header = TRUE,
stringsAsFactors = FALSE)
# usar la variacion mensual
y = diff(log(D$SDXRSA),1,1)

```

(14) generacion energía eléctrica

```

en la librería fpp, se describe como:
`` Electricity monthly total net
generation. January 1973–October 2010.``
require(fpp)
data(usmelec)

```

```

y = ts(usmelec,frequency=12, start=c(1973,01)
# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1973/1/1"),
length.out = length(y), by = "months")
plot(fechas,y,type='l')

```

(15) ventas antidiabético

```

en la librería fpp, se describe como
`Monthly anti-diabetic drug sales
in Australia from 1992 to 2008`.
require(fpp)
data(a10)
y = ts(a10,frequency=12,start=c(1992,01))
# generar un vector de fechas, clase 'Date'
fechas = seq(as.Date("1992/1/1"),
length.out = length(y), by = "months")
plot(fechas,y,type='l')

```

(16) Producción industrial España

```

# descripcion
Time-Series [1:61] from 1978 to 1983:
Brockwell and Davis (1991, Series E, p. 556) "
"Industrial production, Spain: monthly"
# instalacion
devtools::install_github("FinYang/tsdl")
require(tsdl)
# lectura
D <- subset(tsdl,12,"Production")
y = ts(D[[17]],frequency=12,start=c(1978,01))
par(mfrow=c(1,1))
t= seq(1,length(y))
ts.plot(y,type='b')

```

(17) serie turismo

```
# descripcion
The Tcomp package provides data from the tourism
forecasting competition described in George Athanasopolous,
Rob J. Hyndman, Haiyan Song, Doris C. Wu (2011)
``The tourism forecasting competition'',
International Journal of Forecasting 27 (2011) 822-844.
# instalacion
require(Tcomp)
data(package='Tcomp')
# lectura
y1 = tourism$M82$x
ts.plot(y,type='b')
```

(18) turismo Australia

```
# descripcion
Quarterly visitor nights (in millions) spent by
international tourists to Australia. 1999-2015
# instalacion
require(fpp2)
# lectura
data(austourists )
y = ts(austourists,frequency=4,start=c(1999,01))
ts.plot(y,type='b')
```

(19) producción cerveza Australia,

```
en la librería fpp, se describe como:
Producción total de cerveza en Australia: ``
Total quarterly beer production in Australia
(in megalitres) from 1956:Q1 to 2008:Q3.''.
Se carga con las instrucciones
require(fpp)
y = ts(ausbeer,frequency=4,start=c(1956,01))
ts.plot(y)
```

(20) poblacion insectos

```
# descripcion
poblacion de un tipo de mosca (blowfly)
en el estudio Brillinger, Guckenheimer, Guttorp and Oster (1980)
# instalacion
devtools::install_github("FinYang/tsdl")
require(tsdl)
D <- subset(tsdl,1,"Ecology")
# definir un periodo s=18 usando
require(sazedR)
S(D[[7]])
# lectura
y = ts(D[[7]],frequency=18)
ts.plot(y,type='b')
```

CAPÍTULO 6

Medias Móviles y Filtros

6.1. Medias Móviles

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

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

$$\bar{Y}_t = \sum_{j=-m}^m w_j Y_{t-j}, \quad t = m+1, \dots, T-m \quad (6.1)$$

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

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

Nótese que en la serie transformada \bar{Y}_t se pierden $2m$ datos. Las Medias Móviles se utilizan como estimadores de la tendencia, es decir, $\hat{T}_t = \bar{Y}_t$. Esto permite remover

la tendencia al formar la serie sin tendencia $X_t = Y_t - \bar{Y}_t$. Las Medias Móviles bilaterales se utilizan también para estimar la componente estacional. En este caso $\hat{S}_t = \bar{Y}_t$, y la serie $X_t = Y_t - \bar{Y}_t$ se denomina la serie des-estacionalizada. Algunos ejemplos de Medias Móviles son: (ver Diebold [1999, pág. 84])

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

$$\bar{Y}_t = \sum_{j=-m}^m w_j Y_{t-j}. \quad (6.2)$$

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

$$w_j = \frac{1}{2m+1}, \quad j = -m, \dots, m, \quad (6.3)$$

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

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

Ejemplo 6.1.2. *La Media Móvil Unilateral*

$$w_j = \frac{1}{m+1}, \quad j = 0, 1, \dots, m \text{ se define como} \quad (6.4)$$

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

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

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

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

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

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

para $t = 2, 3, \dots, T$. 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. Se comprueba a partir de la relación

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

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

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

Ejemplo 6.1.4. Código en R para SES (EWMA) con la función `emaTA()` de la librería `fTrading`. Los resultados están en la Figura 6.1. Nótese que en “lambda” se ingresa el valor de α .

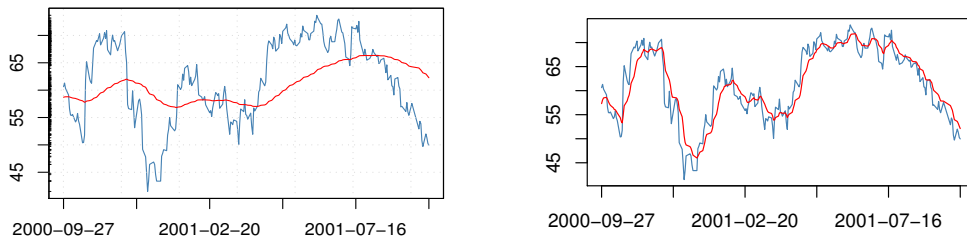
(a) $\alpha = 0.019$ (b) $\alpha = 0.189$

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

```
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")
```

La función filter de R

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

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


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

Los argumentos se definen así.

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

La opción: method = “convolution”. Se utiliza para calcular Medias Móviles.

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

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

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

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

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

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

Por ejemplo, para extraer la tendencia de la serie de temperaturas mensuales, “notem” (ver Ejemplo 3.6, pag. 65) se programa

```
#-----Ejemplo media móvil doble
Yt = nottem
f <- frequency(Yt)
w <- c(0.5, rep(1, f - 1), 0.5)/f
Tt = filter(Yt, w, "conv", 2, T, NULL)
Xt = Yt - Tt
t=seq(1, length(Yt))
par(mfrow = c(3, 1))
plot(t, Yt, type='l')
plot(t, Tt, type='l')
plot(t, Xt, type='l')
```

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

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

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

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

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

se puede generar X así:

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

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

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

La opción: `method = recursive`. Calcula una serie X definida por

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

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

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

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

```

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

```

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

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

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

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

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:

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

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

```

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

```

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

6.2. Método de descomposición con base en Medias móviles

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

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

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

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

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

$$\hat{T}_t = \frac{1}{s} \left(\frac{1}{2} Y_{t-q} + Y_{t-q+1} + \dots + Y_t + \dots + Y_{t+q-1} + \frac{1}{2} Y_{t+q} \right), \quad (6.9)$$

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

$$\hat{T}_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}, \quad (6.10)$$

La segunda parte del procedimiento consiste en estimar la componente estacional. Primero se calcula \hat{T}_t con el filtro (6.9) ó (6.10). Segundo, se estima \hat{S}_k para cada $k = 1, 2, \dots, s$ (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

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

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

Tercero, se coloca

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

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

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

Código R 6.2.1.

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

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

  w <- double(p)

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

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

```
return(D)
}
```

Ejemplo 6.2.1. *La programación en R del filtro (6.9) se puede hacer con la función `filter()`.*

$$T_t = \frac{1}{12} \left(\frac{1}{2} Y_{t-6} + \sum_{u=-5}^5 Y_{t-u} + \frac{1}{2} Y_{t+6} \right), \quad t = 7, \dots, 45, \quad (6.13)$$

```
#-----Programa R para estimar la tendencia
#-----en una serie con componente estacional
s = frequency(y)
w = c(0.5, rep(1, s - 1), 0.5)/s
Tt = filter(y,w,"conv",2,T,NULL)
# Notese que el filtro estima las componente de tendencia y estacionali
# exactamente como lo hace la función stl()
m = stl(y, "per")
Tt.stl = m$time.series[,2]
par(mfrow = c(1,1))
plot(t,Tt,type='l'); lines(t,Tt.stl,col='red');
```

6.3. Librerías para filtrado de series en R

Algunas librerías en R para filtrado de series son las siguientes.

La librería `robfilter` tiene siete filtros

1. `robreg.filter` Simple regression filters
2. `hybrid.filter` Median and repeated median hybrid filters
3. `dw.filter` Two-step location-regression-based filters
4. `wrm.filter` Weighted repeated median filters
5. `robust.filter` Regression filters with additional rules (outlier and level shift detection)

- 6. `adore.filter` Adaptive repeated median filters
- 7. `madore.filter` Multivariate adaptive repeated median filters

La librería `mFilter` (Decomposition of a time series into trend and cyclical components using various filters) de M. Barcilar, tiene los filtros clásicos

- 1. `bkfilter` Baxter-King filter of a time series
- 2. `bwfilter` Butterworth filter of a time series
- 3. `cffilter` Christiano-Fitzgerald filter of a time series
- 4. `hpfiler` Hodrick-Prescott filter of a time series
- 5. `trfilter` Trigonometric regression filter of a time series

Otras técnicas de ajuste estacional

Existen otras técnicas de ajuste estacional que son técnicas ad-hoc para calcular 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

6.4. 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.

CAPÍTULO 7

Pruebas de Incorrelación

7.1. Introducción

En el Modelo de Componentes Estructurales aditivo la serie se asume que se puede descomponer en las componentes de tendencia, estacionalidad y error:

$$Y_t = T_t + S_t + \varepsilon_t, \quad t \in \mathbb{Z}.$$

La componente que interesa analizar en este capítulo es ε_t . La serie estimada $\hat{\varepsilon}_t$, denominada “residuo estructural”, 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 autocorrelacionada ó, por el contrario, es solamente ruido aleatorio. En el primer caso es posible pronosticar los residuos estructurales e incorporar éstos a los pronósticos con la tendencia y la componente estacional. En el segundo caso el pronóstico de un ruido aleatorio es cero. Para clarificar la idea de ruido aleatorio se introduce la siguiente definición básica.

Definición 7.1.1 (Ruido Blanco). Una serie $(\varepsilon_t, t \in \mathbb{Z})$ se dice que es Ruido Blanco si cumple

$$\mathbb{E}(\varepsilon_t) \equiv 0, \quad (7.1a)$$

$$Var(\varepsilon_t) \equiv \sigma^2, \quad (7.1b)$$

$$\forall k \neq 0, Cov(\varepsilon_t, \varepsilon_{t+k}) = 0. \quad (7.1c)$$

Se denota $\varepsilon_t \sim RB(0, \sigma^2)$.

Si además cumple que $\varepsilon_t \sim N(0, \sigma^2)$ se dice que ε_t es Ruido Blanco Gaussiano y se denota $\varepsilon_t \sim RBG(0, \sigma^2)$. Nótese que en este caso la sucesión debe ser iid, $\varepsilon_t \sim iid N(0, \sigma^2)$, ya que una sucesión incorrelacionada que se distribuya Normal debe ser iid. En este caso también se cumple $\mathbb{E}(\varepsilon_{t+k}|\varepsilon_t) = \mathbb{E}(\varepsilon_{t+k}) = 0$, luego, en una sucesión $\varepsilon_t \sim iid N(0, \sigma^2)$ los pronósticos a k pasos son cero.

Obsérvese que si no se cumple $\forall k \neq 0, Cov(\varepsilon_t, \varepsilon_{t+k}) = 0$ entonces se tendría $\exists k \neq 0, Corr(\varepsilon_t, \varepsilon_{t+k}) \neq 0$. Lo cual implica que los errores estarán autocorrelacionados, y por tanto, podrían ser pronosticados. Por ejemplo, con $T_t = \beta_0 + \beta_1 t$,

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

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

Verificar las condiciones (7.1a), (7.1b) y (7.1c) se denomina: “Validación de los supuestos de los errores”. La condición (7.1a) se cumple automáticamente si ε_t es el residuo de la regresión $Y_t = T_t + S_t + \varepsilon_t$. Las otras dos condiciones se comprueban aplicando pruebas de hipótesis adecuadas. El resto del capítulo se dedicará a exponer pruebas para (7.1c), denominadas “pruebas de incorrelación”.

Para la condición de varianza constante, (7.1b), las pruebas disponibles en la literatura, por ejemplo, White, Breusch-Pagan y Goldfeld-Quandt, se denominan “pruebas de heterocedasticidad”. Asumen que la hipótesis alterna equivale a suponer que $Var(\varepsilon_t)$ es una función de una combinación de las variables explicativas, por ejemplo, de t, t^2 , las variables indicadoras en las columnas de la matriz It , ó de los valores estimados \hat{Y}_t , que es una combinación particular de las anteriores.

Esta forma de la hipótesis alterna no es útil si se asume el modelo de componentes, ya que lo que se busca es precisamente eliminar el efecto de tales variables en los residuos estructurales. Rechazar la hipótesis nula de varianza constante llevaría a

tratar de corregir el efecto de heterocedasticidad y a modificar el modelo de manera sustancial. Por lo que optamos por no realizar la validación de este supuesto en esta etapa del modelo y concentrarnos en las pruebas de incorrelación.

7.2. Series Estacionarias en Covarianza

Para determinar si $\widehat{\varepsilon}_t, t = 1, \dots, T$ presenta autocorrelación o no, se introducen dos pruebas de hipótesis y la función de autocovarianza. El concepto de serie estacionaria es básico en la teoría de series de tiempo. Es diferente del concepto de serie estacional.

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

Definición 7.2.1 (Serie de Tiempo Estacionaria en Covarianza). *Una serie $(Y_t, t \in \mathbb{Z})$ se dice estacionaria en covarianza si cumple dos condiciones.*

$$\mathbb{E}(Y_t) \equiv \mu, \quad (7.2a)$$

$$Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1). \quad (7.2b)$$

donde $Cov(Y_{t_1}, Y_{t_2}) = E((Y_{t_1} - \mu)(Y_{t_2} - \mu))$ y $R(t)$ es una función par, $R(-t) = R(t), \forall t$.

Es decir, la covarianza entre Y_{t_1} y Y_{t_2} depende únicamente de la distancia entre los tiempo t_2 y t_1 , y no de si $t_1 < t_2$ ó $t_1 > t_2$. Es decir,

$$Cov(Y_{t_1}, Y_{t_2}) = R(t_2 - t_1) = R(t_1 - t_2) = Cov(Y_{t_2}, Y_{t_1}).$$

Por ejemplo,

$$Cov(Y_{-4}, Y_3) = R(3 - (-4)) = R(7) = Cov(Y_0, Y_7),$$

$$Cov(Y_3, Y_4) = R(4 - 3) = R(1) = Cov(Y_0, Y_1).$$

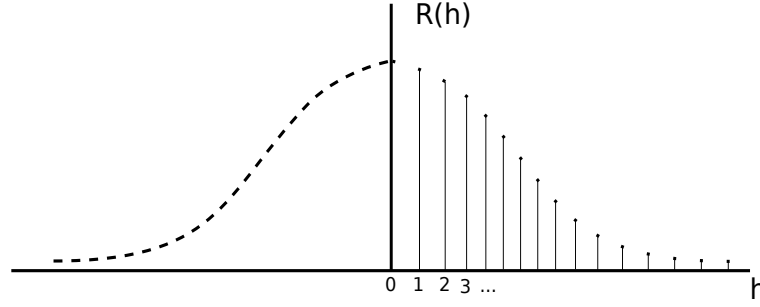


Figura 7.1: Ejemplo Función de Autocovarianza

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

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

Algunos ejemplos de funciones de autocovarianza para series estacionarias en covarianza se muestran a continuación.

$$R(h) = \sigma^2 e^{-\alpha|h|}, \quad \alpha > 0, \quad (7.3a)$$

$$R(h) = \sigma^2 e^{-\alpha|h|} \cos(\beta h), \quad \alpha > 0, \beta \in \mathbb{R}, \quad (7.3b)$$

$$R(h) = \sigma^2 e^{-\alpha|h|} \left(\cos(\beta h) + \frac{\alpha}{\beta} \sin(\beta|h|) \right), \quad \alpha > 0, \beta \in \mathbb{R}, \quad (7.3c)$$

$$R(h) = 2\alpha\sigma^2(1 - 2\alpha h^2)e^{-\alpha h^2}, \quad \alpha > 0, \quad (7.3d)$$

$$R(h) = \sigma^2 e^{-\alpha^2 h^2}, \quad \alpha > 0, \quad (7.3e)$$

$$R(h) = \sigma^2 e^{-\alpha|h|} \left(1 + \alpha|h| + \frac{1}{3}\alpha^2 h^2 \right), \quad \alpha > 0, \quad (7.3f)$$

Algunas propiedades de las series estacionarias en covarianza

Definición 7.2.2. *Un componente determinístico es una serie Z_t que tiene la propiedad de que si se conoce su historia hasta un tiempo t_0 , dada por $Z_s, s \leq t_0$, entonces se puede conocer Z_t para $t > t_0$.*

Por ejemplo, si $A \sim N(\mu, \sigma^2)$ entonces $Z_t = A \cos(\lambda t), t \geq 0$ es un componente determinístico.

Algunas propiedades permiten formarse una idea más clara de como es una serie estacionaria en covarianza.

1. Una serie estacionaria en covarianza con trayectorias monótonas (crecientes ó decrecientes) solamente puede ser una constante. Es decir,

$$\begin{aligned} Y_t \leq Y_{t+1}, \forall t \in \mathbb{Z} &\Rightarrow Y_t \equiv Y_0, \\ Y_t \geq Y_{t+1}, \forall t \in \mathbb{Z} &\Rightarrow Y_t \equiv Y_0. \end{aligned}$$

2. Una serie estacionaria en covarianza toma valores en un intervalo $[a, b]$ infinitas veces ó nunca. Es decir, si

$$\begin{aligned} A &= (\#\{t : Y_t \in [a, b]\} = 0), \\ B &= (\#\{t : Y_t \in [a, b]\} = \infty), \end{aligned}$$

entonces $\mathbb{P}(A \cup B) = 1$ y $\mathbb{P}(A \cap B) = 0$.

3. En vista de las dos propiedades anteriores, una serie estacionaria en covarianza fluctúa alrededor de la media $\mathbb{E}(Y_t) = \mu$. Y se mantiene dentro de un intervalo finito. Se dice entonces que presenta reversión en la media.
4. La función de autocovarianza de una serie estacionaria en covarianza sin componentes determinísticas, $R(h)$, tiende a cero cuando h tiende a infinito:

$$\lim_{h \rightarrow \infty} R(h) = 0.$$

5. Como este resultado es de la forma $p \Rightarrow q$, se cumple la contra-recíproca, es decir, $no(q) \Rightarrow no(p)$. Por lo tanto, si la función de autocorrelación $R(h)$ no tiende a cero la serie no puede ser estacionaria en covarianza.

En el modelo de componentes $Y_t = T_t + S_t + \epsilon_t$, lo que se espera es que los residuos ϵ_t sean una serie estacionaria en covarianza, incluyendo un ruido blanco. Un instrumento útil para examinar la posibilidad de que una serie sea estacionaria es la función de autocorrelación.

Definición 7.2.3 (Función de Autocorrelación). *La función de autocorrelación teórica de una serie estacionaria en covarianza (Y_t , $t \in \mathbb{Z}$) se define como*

$$\rho(h) := \text{Corr}(Y_t, Y_{t+h}) = \frac{R(h)}{R(0)}, \quad h = 0, 1, 2, \dots$$

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 $R(0) = \text{Var}(Y_t) = \sigma^2 > 0$, entonces $|R(h)| \leq \sigma^2 < \infty$, $\forall h$.

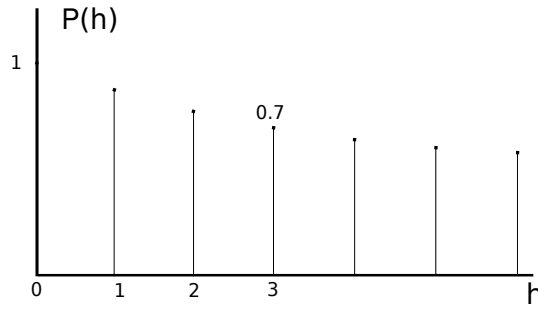


Figura 7.2: Ejemplo Función de Autocorrelación.

Los valores de $\rho(h)$ indican qué grado de asociación lineal existe entre dos valores de la serie Y_t separados una distancia en el tiempo de h . Por ejemplo, $\rho(3) = 0.7$ se interpreta como que Y_t y Y_{t+3} tienen una correlación de 0.7.

Si se observan las gráficas de $\rho(k)$ en las figuras (7.2), éstas parecen tender a cero, luego se puede aceptar que hay evidencia de que las series correspondientes son estacionarias en covarianza.

7.3. Estimación de las funciones de Autocovarianza y Autocorrelación.

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

$$\rho(k) = \frac{E((Y_t - \mu)(Y_{t+k} - \mu))}{E(Y_t - \mu)^2} \quad (7.4)$$

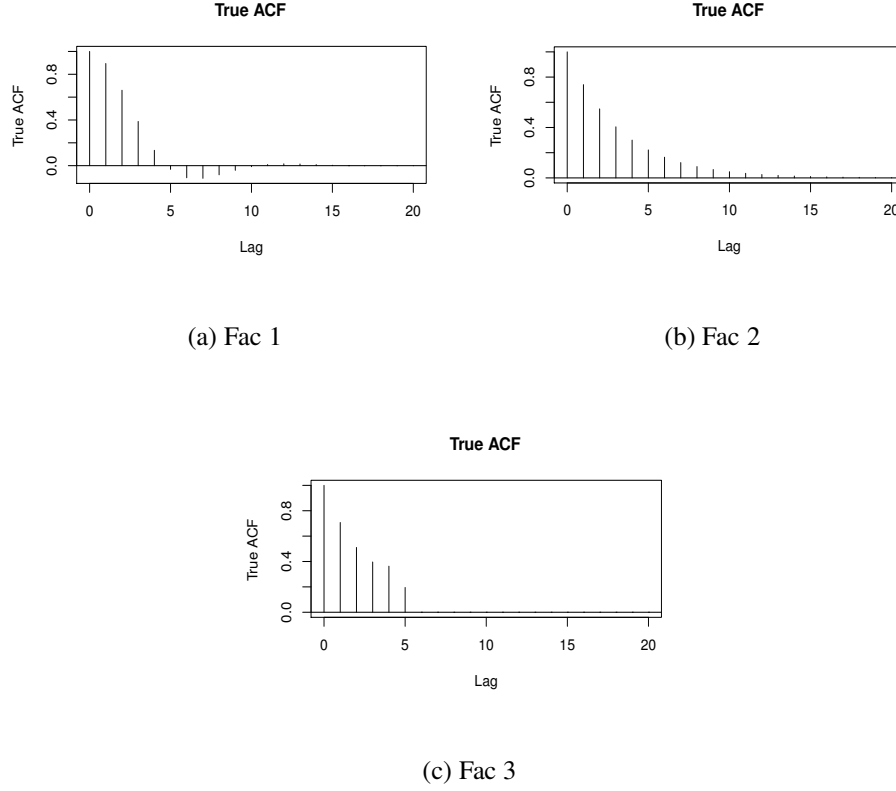


Figura 7.3: Ejemplos de Funciones de Autocorrelación de un Proceso Estacionario

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

$$\hat{\rho}(k) = \frac{\sum_{j=k+1}^T (Y_j - \bar{Y})(Y_{j-k} - \bar{Y})}{\sum_{j=1}^T (Y_j - \bar{Y})^2}, \quad (7.5)$$

donde $\bar{Y} = \frac{1}{T} \sum_{j=1}^T Y_j$ es la media de la trayectoria. Una cuestión práctica es decidir el rango m de los valores de $k = 1, 2, \dots, m$. Una posibilidad es $m = T/4$.

También se utiliza

$$\hat{\rho}(k) = \frac{\frac{1}{T-k+1} \sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\frac{1}{T-1} \sum_{t=1}^T (Y_t - \bar{Y})^2}. \quad (7.6)$$

El cálculo de la fac estimada se realiza en el lenguaje R mediante la función `acf(x)`. La opción `ci.type = "ma"` calcula las “bandas de Bartlett”, que se definen más adelante.

```
acf(x, lag.max = m, ci.type = "ma",
type = c("correlation", "covariance", "partial"))
#---opcion para eliminar la autocorrelacion en cero
require(TSA)
TSA::acf(x, lag.max = m, ci.type = "ma",
drop.lag.0 = TRUE)
```

Pregunta: Cómo debería ser la Función de Autocorrelación para un Proceso Ruido Blanco?.

Si $(\varepsilon_t, t \in \mathbb{Z})$ es ruido blanco, como $E(\varepsilon_t) = 0$, se cumple que la media de ε_t es constante y como $Cov(\varepsilon_t, \varepsilon_{t+k}) = \sigma^2$, para $k = 0$, entonces la función de autocovarianza cumple:

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

La función de autocorrelación teórica $\rho(k)$ (fac) de un ruido blanco es cero excepto en $k = 0$. Sin embargo, las autocorrelaciones muestrales $\hat{\rho}(k)$ de un ruido blanco tienen la propiedad siguiente.

Proposición 7.3.1 (Teorema de Bartlett). *Si $(\varepsilon_t, t \in \mathbb{Z})$ es ruido blanco entonces las autocorrelaciones muestrales $\hat{\rho}(k) = \widehat{Cov}(\varepsilon_t, \varepsilon_{t+k})$, $k = 1, 2, \dots, m$, donde $1 < m < T$ es un entero arbitrario, con base en una muestra $\{\varepsilon_1, \dots, \varepsilon_T\}$, cumplen*

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

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

Es decir, si T es grande y ε_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

$$P(-2\sigma \leq X \leq 2\sigma) \approx 0.95,$$

las autocorrelaciones $\hat{\rho}(k)$ deben estar en el intervalo $[-2/\sqrt{T}, 2/\sqrt{T}]$, con una probabilidad de 0.95 aproximadamente. El valor de m se puede tomar, como se mencionó, como $T/4$.

Definición 7.3.1. La región en \mathbb{R}^2 dada por $\{(x, y) \in [0, m] \times [-2/\sqrt{T}, 2/\sqrt{T}]\}$ se denomina “bandas de Bartlett”.

Se puede realizar un diagnóstico gráfico, con base en $\hat{\rho}(k)$, la fac estimada de los residuos estructurales estimados $\hat{\varepsilon}_t$, con base en la banda de Bartlett, para decidir si los residuos son ruido blanco. Si se observa

$$\forall k \leq m, \hat{\rho}(k) \in [-2/\sqrt{T}, 2/\sqrt{T}] \quad (7.7)$$

entonces hay evidencia de que los residuos estructurales son ruido blanco y por tanto se puede pasar a realizar pronósticos de la serie con la parte estructural. Si se observan algunos k para los cuales $\hat{\rho}(k) \notin [-2/\sqrt{T}, 2/\sqrt{T}]$, entonces se tiene evidencia de que el residuo estructural no es ruido blanco y tiene una dinámica autocorrelacionada que se puede modelar, por ejemplo mediante un proceso ARMA, con el fin de mejorar los pronósticos estructurales.

Como ilustración, al observar la Figura (7.4(a)), se puede concluir que hay evidencia de que Y_t es ruido blanco. En la Figura (7.4(b)) no se cumple que $\forall k, \hat{\rho}(k) \in [-2/\sqrt{T}, 2/\sqrt{T}]$, por tanto, hay evidencia de que Y_t no es ruido blanco.

Ejemplo 7.3.1. En ⁽¹⁾: “El Índice General de la Bolsa de Valores de Colombia o IG-BC era el índice bursátil más representativo de Colombia creado por la Bolsa de Valores de Colombia (BVC). Este índice fue reemplazado por el nuevo índice COLCAP (en nov-2013)”. El código siguiente calcula la fac de la serie $y = \log(\text{igbc})$ entre 2009-02-02 y 2010-12-30, y de su primera diferencia $r = \text{diff}(\log(y), 1, 1)$. El resultado se muestra en la Figura (7.5).

¹https://es.wikipedia.org/wiki/Indice_General_de_la_Bolsa_de_Valores_de_Colombia

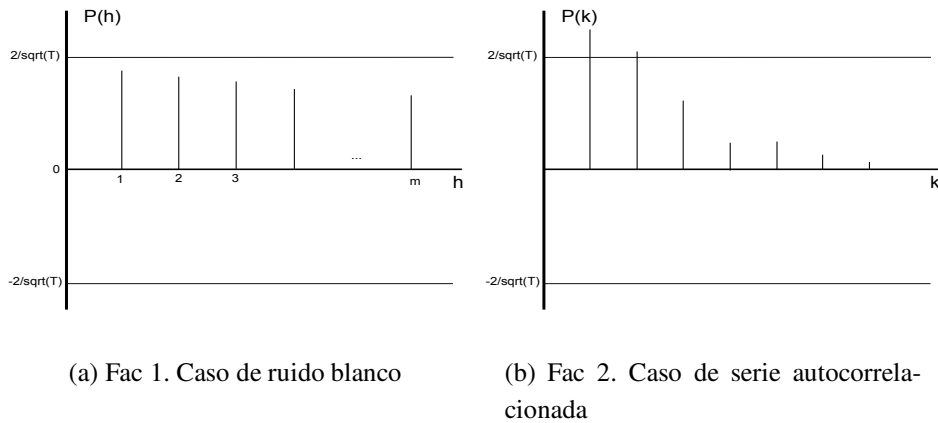


Figura 7.4: Ejemplos de Fac y Bandas de Bartlett

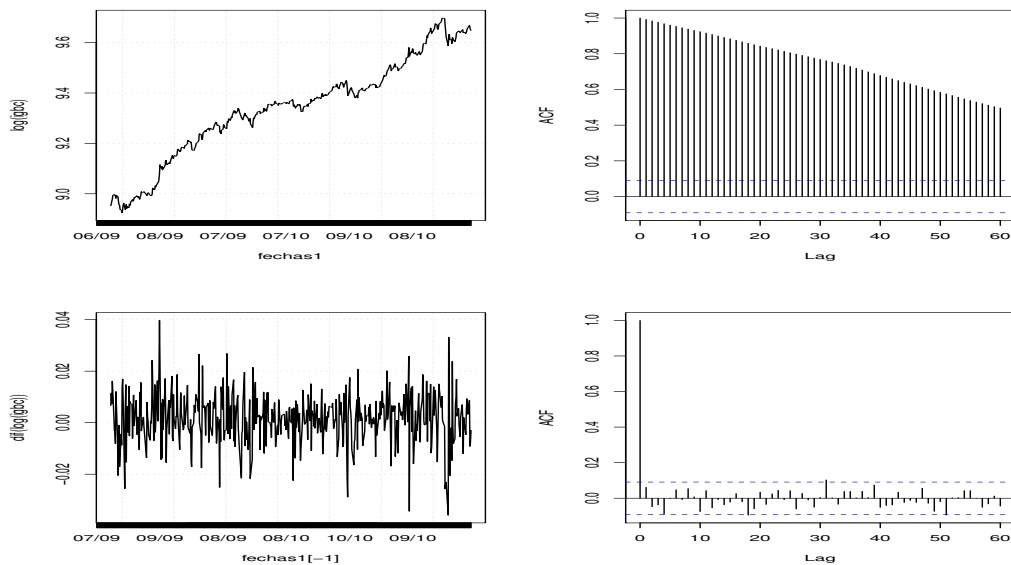


Figura 7.5: $\log(\text{igbc})$ y $\text{diff}(\log(\text{igbc}))$, a izquierda, con sus fac muestrales, a derecha

Puede apreciarse que la fac de $\log(\text{igbc})$ aparece muy persistente, tendiendo a cero lentamente, lo cual se puede interpretar como que no es estacionaria en covarianza. En el capítulo 7 se analizará este hecho mediante pruebas de hipótesis. En cambio, la fac de la serie de las diferencias aparece con todos sus valores dentro de las bandas de Bartlett, por lo que puede concluirse que puede tratarse de un ruido blanco. El

valor $m = 60$ corresponde a 60 días, y se escogió arbitrariamente.

```
# cálculo de la fac
require(TSA)
t = seq(1,length(y))
r = diff(y,1,1)
par(mfrow=c(2,2))
plot(t,y,type='l')
TSA::acf(y, lag.max = 60, ci.type = "ma",
drop.lag.0 = TRUE)
plot(t[-1],r,type='l')
acf(r,60)
```

Ejemplo 7.3.2. Retomando el Ejemplo (3.7.1) en la pag. 69, para modelación de la serie de producción de cemento Portland, trimestral, en toneladas por mil, entre Q1 1956 y Q3 1994, en Australia. Para la serie se asumió un modelo de la forma

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

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

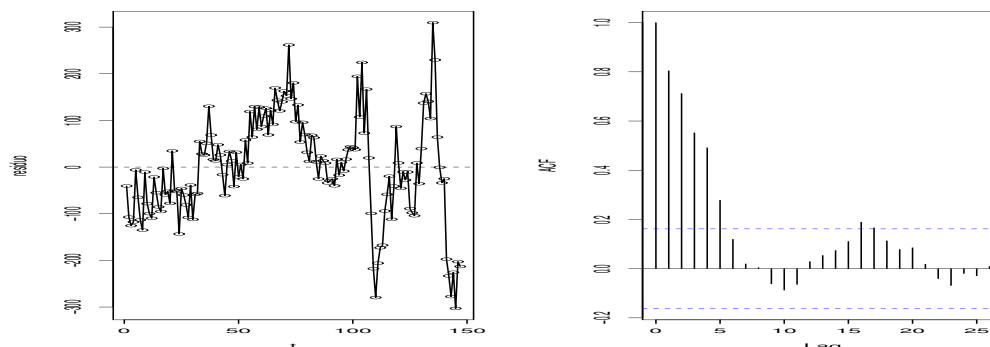


Figura 7.6: Fac de los Residuos Estructurales Serie Cementos

Es evidente que los residuos no son ruido blanco. En la gráfica de la fac muestral, en el panel derecho de la Figura 7.6, se puede apreciar cómo las primeras cinco

autocorrelaciones se salen de las Bandas de Bartlett. Además, en la gráfica del panel izquierdo, la gráfica de $\hat{\varepsilon}_t$ muestra períodos de incremento y luego de decrecimiento, que puede tomarse como evidencia de autocorrelación.

7.4. Pruebas de Incorrelación

Decidir si una serie es ruido blanco usando la función de autocorrelación estimada $\hat{\rho}(k)$ es equivalente a decidir si todas son estadísticamente cero, para $k \geq 1$. La gráfica de $\hat{\rho}(k)$ con las bandas de Bartlett dá un diagnóstico. Dos pruebas de hipótesis para chequear la incorrelación son:

1. Prueba Ljung-Box.
2. Prueba Durbin-Watson.

En las pruebas de hipótesis consideramos cuatro pasos:

1. Enunciado de las Hipótesis de la Prueba.
2. Estadístico de la Prueba.
3. Distribución del Estadístico bajo H_0 .
4. Decisión.

El Error Tipo I es rechazar la Hipótesis nula H_0 siendo ésta cierta. Se trata de un falso positivo porque en ocasiones la alterna H_1 es el resultado de interés. El nivel de significación α se define como

$$\mathbb{P}(\text{Error Tipo I}) = \alpha.$$

El Error Tipo II es no rechazar la nula H_0 siendo la alterna H_1 la cierta. Se trata de un falso negativo. La probabilidad de cometer este error es

$$\mathbb{P}(\text{Error Tipo I}) = \beta.$$

Además, la probabilidad $1 - \beta$, denominada la potencia de la prueba, se interpreta como la probabilidad de rechazar la nula correctamente, es decir, la probabilidad de rechazar H_0 cuando H_1 es cierta.

Las pruebas que tienen una baja potencia tienden a no rechazar la nula H_0 . No tienen capacidad para detectar H_1 . En ocasiones, la alterna H_1 puede estar muy cerca de la nula H_0 , y se requiere una prueba con alta potencia para detectarla.

7.4.1. Prueba Ljung-Box (LB)

A partir del Teorema de Bartlett, (7.3.1), si $\varepsilon_t \sim RB(0, \sigma^2)$, $t = 1, \dots, T$ y $\hat{\rho}(k) = \widehat{Corr}(\varepsilon_t, \varepsilon_{t+k})$, $k = 1, 2, \dots, m$, 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 $(\sqrt{T}\hat{\rho}(k))^2 \stackrel{a}{\sim} \chi_1^2$, independientes, por lo tanto, por la propiedad reproductiva de la distribución χ_ν^2 ,

$$Q = T \sum_{k=1}^m \hat{\rho}^2(k) \stackrel{a}{\sim} \chi_m^2. \quad (7.8)$$

El estadístico de la prueba Ljung-Box se define a partir de este resultado.

1. Hipótesis de la Prueba

$$\begin{aligned} H_0 : \varepsilon_t &\sim RB(0, \sigma^2) \\ H_1 : no(H_0) \end{aligned}$$

2. Estadístico de Prueba

Ljung y Box modificaron el estadístico Q en (7.8) para mejorar la aproximación de Q a la χ_m^2 en muestras pequeñas, usando

$$Q_{LB} = T(T+2) \sum_{k=1}^m \frac{1}{T-k} \hat{\rho}^2(k). \quad (7.9)$$

3. Distribución del Estadístico de la Prueba bajo H_0

Si $H_0 : \varepsilon_t \sim RB(0, \sigma^2)$ es cierta se cumple $Q_{LB} \stackrel{a}{\sim} \chi_m^2$.

4. Decisión

- a) Colocando Q_{obs} el estadístico observado y *Valor p* $p = P(\chi_m^2 \geq Q_{\text{obs}} | H_0 \text{ cierto})$, si *Valor p* $p < 0.05$, se rechaza H_0 . Si *Valor p* $p > 0.05$, no se rechaza H_0 .

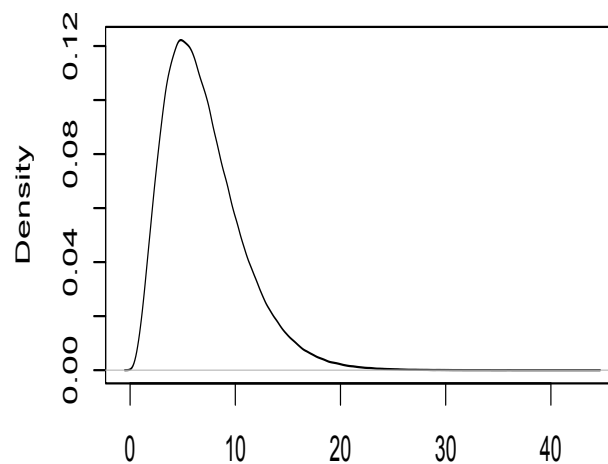


Figura 7.7: Densidad χ_m^2 .

Un Valor muy pequeño de Q_{obs} significa que *todas* las $\hat{\rho}^2(k)$ son pequeñas, luego no se rechaza H_0 .

- b) **Usando valores críticos:** Si $\chi_{m,\alpha}^2$ es valor crítico de χ_m^2 correspondiente a la probabilidad α , α nivel de significación, 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 . Es una prueba de cola derecha.

La prueba Ljung-Box se programa con la función `Box.test`. Por ejemplo

```
Box.test(rn, lag = 35 , type = "Ljung-Box")
X-squared = 379.43, df = 35, p-value < 2.2e-16,
```

por tanto, rechaza la hipótesis nula de que los residuos son Ruido Blanco.

Ejemplo 7.4.1. *Este es un ejemplo sobre la importancia de la hipótesis de Ruido Blanco en la prueba Ljung-Box. Suponga una sucesión periódica de período $p = 126$, dada por*

$$\sigma_t = 10(\sin(2\pi t/p) + 2), t = 1, 2, \dots$$

Y suponga una sucesión aleatoria $Z_t \sim iidN(0, 1)$. Se forma la sucesión $Y_t = \sigma_t Z_t$. Se tiene inmediatamente

$$\begin{aligned}\mathbb{E}(Y_t) &\equiv 0, \\ \text{Var}(Y_t) &= \text{Var}(\sigma_t Z_t) = \sigma_t^2, \\ \text{Corr}(Y_t, Y_{t+s}) &= \frac{\mathbb{E}(Y_t Y_{t+s})}{\sigma_t \sigma_{t+s}} = \frac{\mathbb{E}(Y_t) \mathbb{E}(Y_{t+s})}{\sigma_t \sigma_{t+s}} = 0, \quad t, s = 1, 2, \dots\end{aligned}$$

Entonces Y_t no es Ruido Blanco, es una serie incorrelacionada heterocedástica. En la Figura 7.8 se muestran trayectorias simuladas de Z_t y Y_t , y la fac de ésta. Qué resultado da la prueba Ljung-Box?.

```
#-----ruido  periódico
n = 350; p = 126; t = seq(1,n);
sn = 10*(sin(2*pi*t/p) + 2)
en = rnorm(n, 0, 1)
rn = sn*en
Box.test(rn, lag = 16 , type = "Ljung-Box")
data:  rn
X-squared = 28.407, df = 16, p-value = 0.02825
Box.test(rn, lag = 32 , type = "Ljung-Box")
Box-Ljung test
data:  rn
X-squared = 51.864, df = 32, p-value = 0.01462
```

Ejemplo 7.4.2. *Ejemplo con datos de la accion de Siemens AG. Ver la Figura 7.9. Calcular la prueba Ljung-Box para los rendimientos logarítmicos y calcular la fac para la serie nivel (precio) y la serie retorno. Se utiliza la función de R, `Box.test`.*

Código R 7.4.1.

```
#-----ejemplo accion Siemens bolsa Frankfurt
```

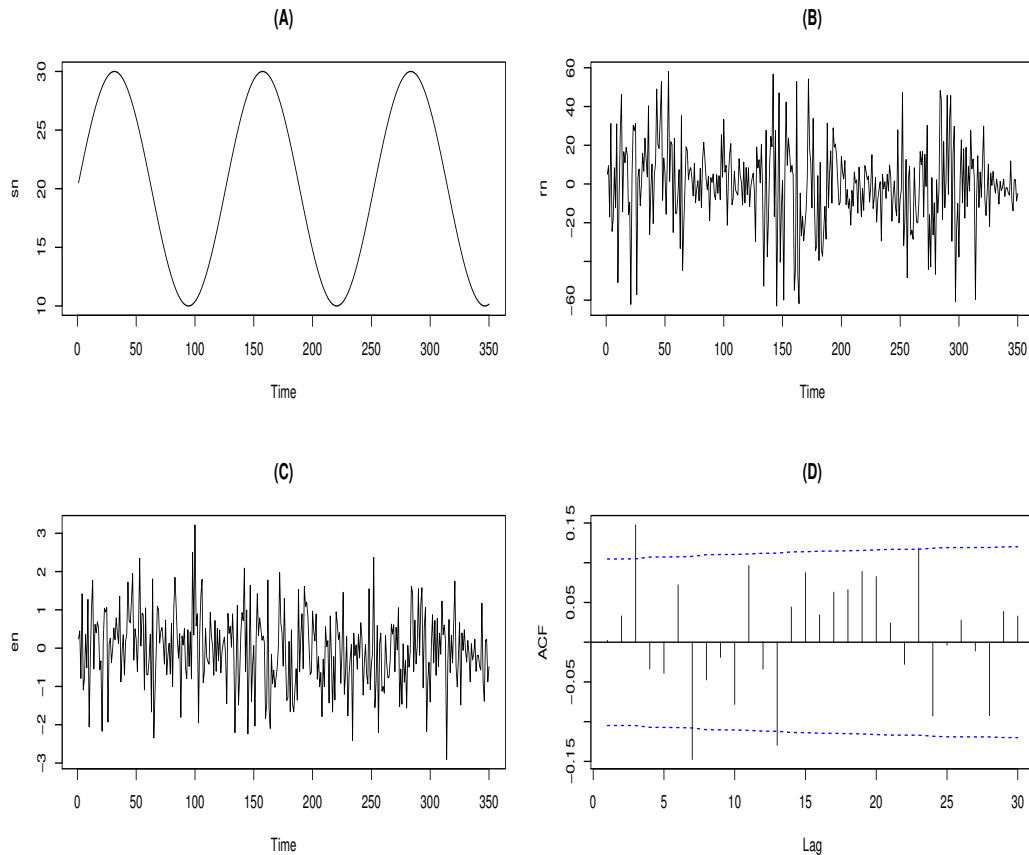


Figura 7.8: (A): σ_t , (B): Y_t , (C): Z_t , (D): fac de Y_t

```

https://es-us.finanzas.yahoo.com/q/hp?s=SIE.F+Precios+históricos
D = read.csv("siemens.SIE.F.csv", header=TRUE, stringsAsFactors=FALSE)
Xn = D$Close; Zn = diff(log(Xn),1,1);
require(TSA)
par(mfrow=c(2,2))
ts.plot(Xn)
TSA::acf(Xn,90,ci.type="ma",drop.lag.0 = TRUE)
ts.plot(Zn)
TSA::acf(Zn,30,ci.type="ma",drop.lag.0 = TRUE)
#-----pruebas Ljung-Box
Box.test(Zn, lag = 15 , type = "Ljung-Box")
X-squared = 30.983, df = 15, p-value = 0.008831

```

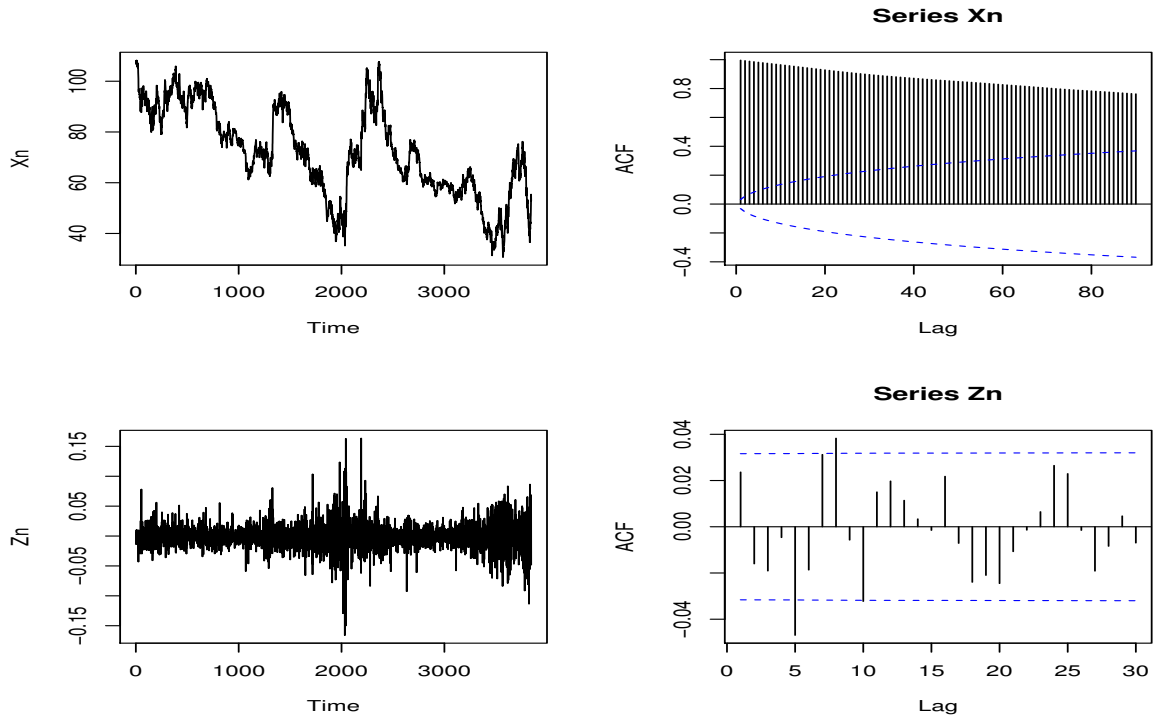



Figura 7.9: Compara fac de serie no estacionaria versus estacionaria.

7.4.2. Prueba Durbin-Watson (DW)

La prueba DW se basa en la definición del modelo Auto regresivo de orden uno, $AR(1)$.

Definición 7.4.1. Una serie $(Y_t, t \in \mathbb{Z})$ se dice que sigue un modelo $AR(1)$ con media μ si cumple

$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (7.10)$$

donde $|\phi| < 1$, $\mu \in \mathbb{R}$ y $\varepsilon_t \sim RB(0, \sigma^2)$. Se escribirá $Y_t \sim AR(1)$.

Propiedades. Si $Y_t \sim AR(1)$ entonces es un proceso estacionario en covarianza, tal que

- $E(Y_t) \equiv \mu$

- $Var(Y_t) \equiv \frac{\sigma^2}{1 - \phi}$
- $Cov(Y_t, Y_{t+k}) = R(k) = \frac{\sigma^2 \phi^k}{1 - \phi^2}, k = 0, 1, \dots$
- $\rho(k) = \phi^k, k = 0, 1, 2, \dots$

luego la fac teórica de un $AR(1)$ tiene las formas generales

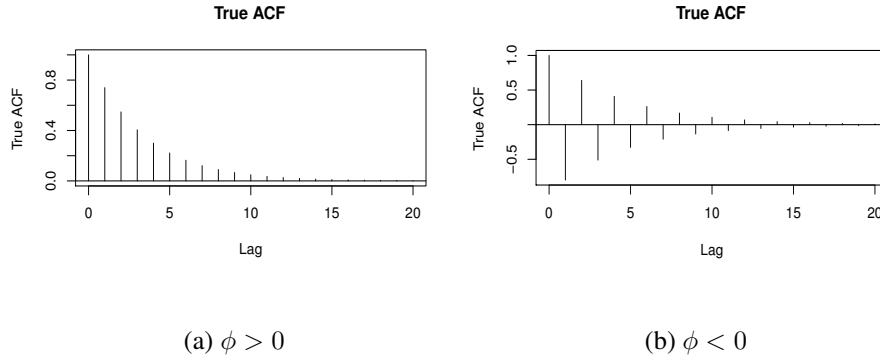


Figura 7.10: Fac teóricas de un $AR(1)$.

Ahora, si se tiene una serie $(Y_t, t = 1, \dots, T)$ y se quiere probar que es incorrelacionada, asumimos que $\mathbb{E}(Y_t) = 0$; en caso contrario, se transforma $Y_t^* = Y_t - \bar{Y}_t$.

La prueba DW se basa en asumir primero que $Y_t \sim AR(1)$, es decir, Y_t sigue el modelo (7.10), con $\mu = 0$.

1. Hipótesis de la Prueba

$$H_0 : \phi = 0$$

$$H_1 : \phi \neq 0.$$

Nótese que si H_0 es cierta, entonces $Y_t = \varepsilon_t$ y como $\varepsilon_t \sim RB(0, \sigma^2)$ y $Y_t \sim RB(0, \sigma^2)$.

2. Estadístico de Prueba

Definición 7.4.2. El estadístico DW se define como

$$d = \frac{\sum_{t=2}^T (Y_t - Y_{t-1})^2}{\sum_{t=1}^T Y_t^2}. \quad (7.11)$$

3. Distribución del estadístico bajo H_0 . La distribución de d bajo H_0 depende de la matriz de diseño X , para todos los modelos de regresión $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$. Por lo que debe ser calculada para cada X .

4. Decisión.

De (7.11), como $\sum_{t=2}^T (Y_t - Y_{t-1})^2 = \sum_{t=2}^T Y_t^2 + \sum_{t=2}^T Y_{t-1}^2 - 2 \sum_{t=2}^T Y_t Y_{t-1}$, se aproxima

$$\begin{aligned} d &= \frac{\sum_{t=2}^T Y_t^2 + \sum_{t=1}^{T-1} Y_t^2 - 2 \sum_{t=2}^T Y_t Y_{t-1}}{\sum_{t=1}^T Y_t^2} \\ &\approx \frac{\sum_{t=2}^T Y_t^2 + \sum_{t=2}^T Y_t^2 - 2 \sum_{t=2}^T Y_t Y_{t-1}}{\sum_{t=1}^T Y_t^2} \\ &= 2 - 2\hat{\rho}(1). \end{aligned}$$

es decir,

$$d \approx 2(1 - \hat{\rho}(1)). \quad (7.12)$$

A partir de (7.12) se puede concluir que, como $-1 < \hat{\rho}(1) < 1$, entonces $0 < d < 4$. Y si $\hat{\rho}(1) \approx 0$ entonces $d \approx 2$ y no rechaza la nula.

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

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

Colocando d_{obs} el estadístico observado y $\text{Valor } p = P(d \geq d_{\text{obs}} | H_0 \text{ cierto})$, si $\text{Valor } p < 0.05$, se rechaza H_0 . Si $\text{Valor } p > 0.05$, no se rechaza H_0 .

4. Prueba DW en R

- a) La función `dwtest` en la librería `lmtest`. Permite probar hipótesis alternas $H_1 : \phi > 0, \phi \neq 0, \phi < 0$, colocando `dwtest(yt ~ t + It)`. Calcula d y el Valor p .
- b) La función `durbin.watson` en la librería `car`. Utiliza bootstrap para calcular la distribución del estadístico d . Se programa `durbinWatsonTest(yt ~ t + It)`.

Ejemplo 7.4.3. *Ejemplos de resultados con las dos funciones anteriores para residuos de un modelo $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$.*

```
require(car)
durbinWatsonTest(lm(yt ~ t), simulate=TRUE, reps=1000)
#-----resultado
lag Autocorrelation D-W Statistic p-value
  1          0.1353341          1.729167          0
Alternative hypothesis: rho != 0
#-----
require(lmtest)
dwtest(yt ~ t)
#-----resultado
data: yt ~ t
DW = 1.7292, p-value = 4.473e-05
alternative hypothesis: true autocorrelation
is greater than 0
#-----fac residuos estimados
require(TSA)
acf(et, 90, ci.type='ma', drop.lag.0=TRUE)
t = seq(1, length(et))
```

Ejemplo 7.4.4. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

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

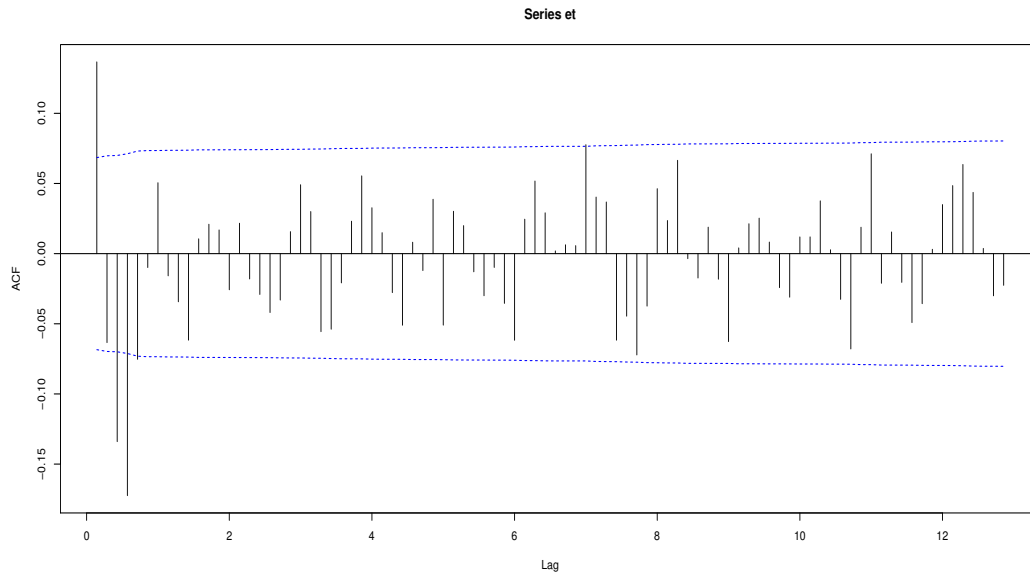


Figura 7.11: La fac de residuos $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$

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 (con las dos funciones de R mencionadas). Los resultados de R se muestran a continuación del código siguiente.

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

Código R 7.4.2.

```
# ejemplo de pruebas LB, DW
# para la serie de produccion de cemento
library(car)
library(forecast)
library(lmtest)
#----
E = read.table(cementq.dat, header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
#----modelo con tendencia lineal y estacionalidad
```

```

#----con variables indicadoras estacionales
t = seq(1,length(y))
It = seasonaldummy(y)
mod1 = lm(y ~ t + It)
summary(mod1)
et = residuals(mod1)
# pruebas Ljung-Box
Box.test(et, lag = 25, type = "Box-Pierce")
# resultado -----
Box-Ljung test
X-squared = 293.9028, df = 25, p-value < 2.2e-16

# pruebas Durbin-Watson
#-----con dwtest
dwtest(mod1) # de la libreria lmtest
Durbin-Watson test
DW = 0.3646, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0

durbinWatsonTest(mod1) # de la libreria car
#-----con durbin.watson
lag Autocorrelation D-W Statistic p-value
1 0.8043803 0.3645806 0
Alternative hypothesis: rho != 0
-----

```

7.4.3. Prueba Durbin-Watson Generalizada

Las pruebas Ljung-Box y Durbin-Watson coincidirían en su decisión siempre que la auto correlación de orden 1, $\hat{\rho}(1)$, sea dominante, es decir, sea la mayor de todas.

Pero la prueba DW puede no rechazar la hipótesis nula de incorrelación, siendo falsa, en el caso en que las primeras autocorrelaciones no sean significativas pero las de rezagos mayores sí.

Para incorporar esta posibilidad, Vinod [1973] introdujo la prueba Durbin-Watson Generalizada (DWG). Consiste en una serie de pruebas en secuencia, para $j = 1, 2, \dots$, asumiendo en cada j un modelo de la forma $Y_t = \phi_j Y_{t-j} + \eta_t$. La prueba para el paso j es

$$H_0 : \phi_j = 0$$

$$H_1 : \phi_j \neq 0.$$

dado que en las pruebas anteriores $j - 1, \dots, 1$, no se rechazó la hipótesis nula. El estadístico Durbin-Watson generalizado para el paso j se define como

$$d_j = \frac{\sum_{t=j+1}^T (Y_t - Y_{t-j})^2}{\sum_{t=1}^T Y_t^2}. \quad (7.13)$$

La prueba DWG se encuentra implementada en la librería `car`. Por ejemplo, colocando

```
durbinWatsonTest(mod,max.lag=10)
```

se obtienen los valores de los estadísticos d_j , para $j = 1, \dots, 10$, los correspondientes valores de $\hat{\rho}(j)$ y los valores p . Esta secuencia de pruebas permite determinar el orden de la menor auto correlación significativa.

Ejemplo 7.4.5. *Un caso de aplicación de la prueba GDW, muestra que no rechaza la nula de incorrelación en los rezagos $k = 1, 2, 3, 4$, pero a partir del rezago $k = 5$ rechaza la nula. Luego, se concluye que los residuos no son ruido blanco.*

Código R 7.4.3.

```
# ejemplo de prueba DWG
library(car)
library(forecast)
library(lmtest)
#---modelo con tendencia lineal y estacionalidad
#---con variables indicadoras estacionales
t = seq(1,length(y))
It = seasonaldummy(y)
```

```

mod1 = lm(y ~ t + It)
summary(mod1)
r = residuals(mod1)
# prueba Durbin-Watson Generalizada
# de la libreria car
durbinWatsonTest(mod1,max.lag=10)

```

Tabla 7.1: Resultados de la Prueba DW Generalizada

	Lag	rho.hat	dw	Valor.p
1	1.00	0.10	1.71	0.21
2	2.00	0.07	1.74	0.19
3	3.00	-0.01	1.89	0.93
4	4.00	-0.06	1.96	0.53
5	5.00	0.80	0.39	0.00
6	6.00	0.71	0.55	0.00
7	7.00	0.60	0.76	0.00
8	8.00	0.53	0.89	0.00
9	9.00	0.34	1.26	0.00
10	10.00	0.20	1.53	0.01

7.5. Alternativas cuando los Residuos Estructurales muestran Autocorrelación

En el caso de detectarse autocorrelación en los residuos estructurales el procedimiento consistirá en buscar un modelo tipo ARMA, autorregresivo de media móvil, que se tratará en el Capítulo 8. Pero es posible utilizar otras alternativas.

7.5.1. Incluir rezagos de la variable dependiente en el modelo

Una alternativa consiste en incluir en el modelo una ó varias componentes autorregresivas, es decir, aumentar el número de variables explicativas incluyendo rezagos

de la variables dependiente. Supongamos que el modelo inicial es un modelo con tendencia lineal y componente estacional dado por

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \quad (7.14)$$

Y se incluyen en el modelo, por ejemplo, dos rezagos de la variable dependiente

$$Y_t = \beta_0 + \beta_1 t + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \sum_{j=1}^{s-1} \delta_j I_t(t) + \varepsilon_t. \quad (7.15)$$

Con este cambio 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 (7.15) se coloca

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

Sin embargo, se sabe que cuando se agregan rezagos de la variable dependiente, el estadístico Durbin-Watson se sesga hacia 2 y no rechaza la nula de ruido blanco, rebaja la potencia.

Ejemplo 7.5.1. *Retomando el Ejemplo (7.4.4), para modelación de la serie de producción de cemento Portland, trimestral. Se puede preguntar si el modelo (7.15) logra capturar la estructura autoregresiva, detectada anteriormente. El programa siguiente muestra los pasos en la estimación de los modelos (7.14) y (7.15).*

Código R 7.5.1.

```
# ejemplo de pruebas LB, DW y DWG
# para la serie de produccion de cemento
# ademas ejemplo con rezagos con libreria dynlm
library(car)
library(forecast)
library(lmtest)
# modelo incorporando terminos rezagados de la serie
```

```
# utilizando la libreria dynlm
require(dynlm)
mod2 = dynlm(y ~ t + It + L(y,1)+ L(y,2))
summary(mod2)
```

Los resultados de la estimación con el modelo (7.15) se muestran en la Tabla 7.2 siguiente.

Tabla 7.2: Resultados del Modelo estructural con dos rezagos 1,2 de Y_t

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	100.1147	35.5231	2.82	0.0055
t	0.8695	0.4035	2.15	0.0329
ItQ1	-159.4054	15.5466	-10.25	0.0000
ItQ2	74.9556	21.9855	3.41	0.0009
ItQ3	65.1675	17.0202	3.83	0.0002
L(yi, 1)	0.6505	0.0828	7.86	0.0000
L(yi, 2)	0.2180	0.0839	2.60	0.0103

La pruebas LB rechaza la nula de ruido blanco, pero la DW no.

```
r2 = residuals(mod2)
# pruebas Ljung-Box
Box.test(r2, lag = 15, type = "Ljung-Box")
Box-Ljung test
data: r2
X-squared = 50.508, df = 15, p-value = 9.944e-06

# pruebas Durbin-Watson

dwtest(mod2)
Durbin-Watson test
data: mod2
DW = 1.9078, p-value = 0.2669
alternative hypothesis: true autocorrelation is greater than 0
```

```

durbinWatsonTest(mod2)
lag Autocorrelation D-W Statistic p-value
  1      0.01205702      1.907794    0.546
Alternative hypothesis: rho != 0
durbinWatsonTest(mod2,max.lag=10)

```

La prueba DWG con 10 rezagos no rechazan la nula para las auto correlaciones de órdenes 1,2,3. Pero la de orden 4 sí es significativa, y se rechaza la hipótesis de ruido blanco para este modelo. Ver la Tabla 7.3. La estrategia de incluir variables rezagadas no produce un resultado concluyente en este caso particular.

Tabla 7.3: Resultados Prueba DWG para el modelo con rezagos 1,2

	r	dw	p
1	0.0231	1.9484	0.6980
2	0.0793	1.8313	0.2720
3	-0.0762	2.1134	0.4100
4	0.3807	1.1955	0.0000
5	-0.0747	2.0854	0.4140
6	-0.2402	2.3958	0.0040
7	-0.1780	2.2050	0.0840
8	0.0815	1.6810	0.0920
9	-0.1215	2.0790	0.2360
10	-0.1959	2.2123	0.0480
11	-0.1247	2.0237	0.2900
12	0.1753	1.4218	0.0000
13	-0.0313	1.7453	0.5680
14	-0.0868	1.8542	0.7960
15	-0.0424	1.7558	0.6980

Como se anotó, la prueba DW se afecta por el termino $\phi_1 Y_{t-1}$, y el estadístico se sesga hacia 2, es decir, la prueba tiende a aceptar H_0 . La prueba tiene baja potencia para detectar la hipótesis alterna H_a .

7.5.2. Transformación Cochran-Orcutt

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

Si la prueba DW detecta la alterna $\phi = \rho(1) \neq 0$, entonces se procede a transformar el modelo para la serie Y_t . Escribiendo el modelo estructural como un modelo de regresión lineal en (4.1), $Y_t = \underline{X}_t' \underline{\beta} + \varepsilon_t$, donde la prueba DW asume $\varepsilon_t = \phi \varepsilon_{t-1} + \eta_t$ y η_t es ruido blanco, la transformación Cochran-Orcutt se define como

$$Y_t - \phi Y_{t-1} = \alpha(1 - \hat{\rho}(1)) + \beta(\underline{X}_t - \hat{\rho}(1)\underline{X}_{t-1}) + \eta_t. \quad (7.16)$$

De esta manera, los residuos η_t son ruido blanco. Esta transformación depende de una autocorrelación en los residuos tipo AR(1). No sería de esperar que funcionara en casos de estructuras de autocorrelación más complejas, como será el caso de la mayoría de las series de tiempo a considerar.

Ejemplo 7.5.2. *Retomando el Ejemplo (7.4.4), para modelación de la serie de producción de cemento Portland, trimestral, al aplicar la transformación Cochran-Orcutt mediante la función `cochrane.orcutt` de la librería `orcutt`, se obtiene que se rechaza la nula de ruido blanco. La función `cochrane.orcutt`, desafortunadamente, no es compatible con las funciones para Durbin-Watson y DWG.*

```
# transformacion Cochran-Orcutt
require(orcutt)
mod2 = cochrane.orcutt(mod1)
r = mod2$Cochrane.Orcutt$residuals
Box.test(r, lag = 25, type = "Ljung-Box")
# pruebas Durbin-Watson
dwtest(mod2) # no funciona!
durbinWatsonTest(mod2) # no funciona!
```

7.5.3. Pruebas de Heterocedasticidad

Ejemplo 7.5.3. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral.*

Al estimar los residuos estructurales $\hat{\varepsilon}_t$ es evidente que muestran heterocedasticidad, como se aprecia a partir de la Figura 7.12.

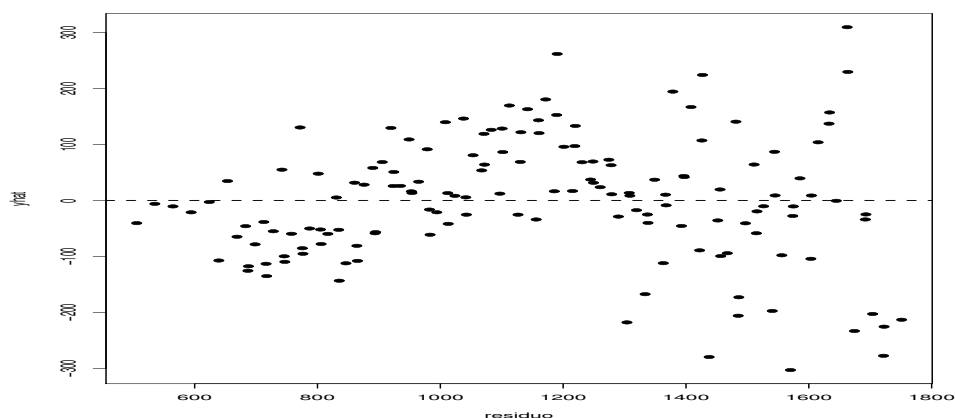


Figura 7.12: Y estimada versus residuales estructurales

Se pueden aplicar las pruebas de heterocedasticidad, Breusch-Pagan, Prueba de Scores para Varianza no constante y Goldfeld-Quandt. Los resultados se muestran a continuación.

```
-----
studentized Breusch-Pagan test
BP = 19.9866, df = 4, p-value = 0.0005025

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 23.03938    Df = 1    p = 1.587171e-06

Goldfeld-Quandt test
GQ = 6.8213, df1 = 69, df2 = 68, p-value = 7.005e-14
-----
```

Sin embargo, la estrategia consiste en buscar un modelo tipo ARMA(p,q) para los residuos estructurales y no una transformación como la que se sugeriría a partir de los resultados de estas pruebas. Por lo que la verificación de la homocedasticidad de la varianza de los residuos estructurales no se incluirá en las pruebas de validación de los residuos.

CAPÍTULO 8

Modelos ARMA para la Componente Aleatoria

8.1. Introducción

En los modelos de descomposición $Y_t = T_t + S_t + \varepsilon_t$, $t = 1, 2, \dots$ 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 cuatro posibles modelos.

1. Medias Móviles de orden q , $MA(q)$.
2. Autoregresivos de orden q , $AR(p)$.
3. Autoregresivos de Medias Móviles, $ARMA(p, q)$.
4. Autoregresivos de Medias Móviles Estacionales, $SARMA(p, q)(p_s, q_s)$.

Con relación a los tres primeros modelos Diebold [1999](ver pag. 129) anota lo siguiente : “Los tres modelos varían en su capacidad de capturar distintos tipos de comportamiento de autoregresión. Comenzaremos dando las características de las

funciones de autocorrelación y las cantidades relacionadas con cada modelo; éstas no tiene nada que ver con los datos ni con la estimación, pero son fundamentales para desarrollar una comprensión básica de las propiedades de los modelos, necesarias para llevar a cabo pronósticos inteligentes”.

Anotamos que Diebold [1999] no incluye el caso de Autoregresivos de Medias Móviles Estacionales, $SARMA(p, q)(p_s, q_s)$.

Definición 8.1.1. *El Operador Rezago se denota L (lag, en inglés) y se define por*

$$L(Y_t) = Y_{t-1}, \quad t \in \mathbb{Z}.$$

Es decir, L opera sobre una serie rezagándola un período hacia atrás. De igual manera $L(Y_{t-1}) = Y_{t-2}$, luego $L(L(Y_t)) = L^2(Y_t) = Y_{t-2}$ y en general $L^p(Y_t) = Y_{t-p}$, $p \geq 1$. 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 β_j , $j = 0, 1, \dots, p$ si se considera el polinomio en L

$$B_p(L) = \beta_0 I + \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p = \sum_{j=0}^p \beta_j L^j, \quad (8.1)$$

entonces, al aplicarlo a Y_t se obtiene

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

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

Definición 8.2.1 (Proceso $MA(q)$). *Se dice que una serie $(Y_t, t \in \mathbb{Z})$ sigue un proceso $MA(q)$, $q = 1, 2, \dots$ de media móvil de orden q , si se cumple que*

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}, \quad (8.2)$$

para constantes $\theta_1, \dots, \theta_q$ y $\varepsilon_t \sim RB(0, \sigma^2)$. La expresión para (8.2) con el operador L es $Y_t = \Theta_q(L)(\varepsilon_t)$, donde

$$\Theta_q(z) = 1 + \theta_1 z + \dots + \theta_q z^q \quad (8.3)$$

se denomina el polinomio de media móvil.

Ejemplo 8.2.1. Defina $Y_t \sim MA(2)$ dado por

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \\ \theta_1 &= -0.4, \quad \theta_2 = 0.4, \\ \varepsilon_t &\sim iid N(0, 9), \quad t \in \mathbb{Z}. \end{aligned}$$

8.2.1. Propiedades de los procesos MA(q)

Suponga que $Y_t \sim MA(q)$ según (8.2). Entonces

1. $\mathbb{E}(Y_t) \equiv 0$. Se cumple inmediatamente pues $\mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t) + \theta_1 \mathbb{E}(\varepsilon_{t-1}) + \dots + \mathbb{E}(\theta_q \varepsilon_{t-q}) = 0$, ya que ε_t es ruido blanco. Se puede definir procesos MA(q) con media μ diferente de cero colocando $Y_t = \mu + \Theta_q(L)(\varepsilon_t)$.
2. $Var(Y_t) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2$. Debido a que ε_t es incorrelacionado se tiene que

$$\begin{aligned} Var(Y_t) &= Var(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}) \\ &= (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2. \end{aligned}$$

Nótese que $Var(Y_t) > Var(\varepsilon_t) = \sigma^2$.

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

$$R(k) = \sigma^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k} = \sigma^2 (\theta_k + \theta_1 \theta_{1+k} + \dots + \theta_q), \quad 0 \leq k \leq q, \quad (8.4)$$

Y $R(k) = 0$, $k \geq q+1$, con $\theta_0 = 1$. La ecuación (8.4) se puede interpretar como una evidencia de que un $MA(q)$ es un proceso débilmente correlacionado, ya que su autocovarianza es cero a partir de un cierto rezago. Por esta razón se pueden ver los procesos MA(q) como alternativas al Ruido Blanco, que es completamente incorrelacionado.

Ejemplo 8.2.2. *Evaluar $R(k)$ con el proceso $MA(2)$ del Ejemplo 8.2.1*

El modelo es $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$, $t \in \mathbb{Z}$, con $\varepsilon_t \sim iid N(0, 9)$, y los parámetros $\theta_1 = -0.4$, $\theta_2 = 0.4$, $\sigma_2 = 9$. Entonces hay que evaluar

$$R(k) = 9 \sum_{j=0}^{2-k} \theta_j \theta_{j+k}, \quad 0 \leq k \leq 2,$$

con $\theta_0 = 1$,

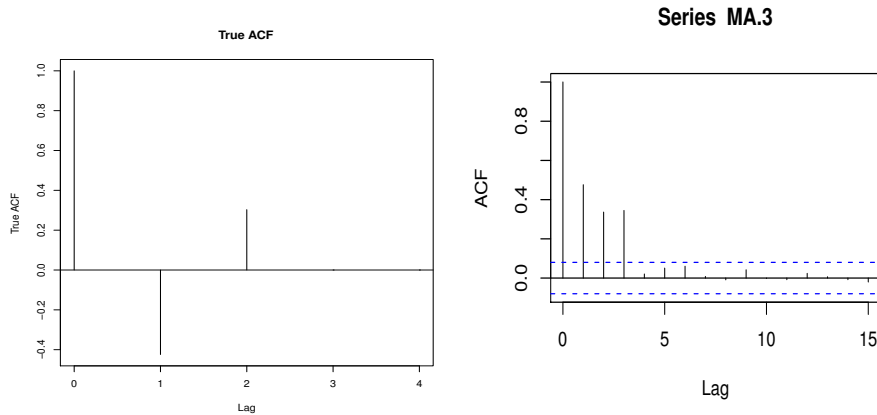
$$R(0) = 9(1 + 0.4^2 + 0.4^2) = 11.88$$

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

$$R(2) = 9 \sum_{j=0}^{2-2} \theta_j \theta_{j+2} = 9(\theta_0 \theta_2) = 9(0.4) = 3.6.$$

Por tanto, la fac es

$$\begin{aligned} \rho(0) &= 1, \quad \rho(1) = -\frac{5.04}{11.88} = -0.42, \quad \rho(2) = \frac{3.6}{11.88} = 0.30 \\ \rho(3) &= \rho(4) = \dots = 0 \end{aligned}$$



(a) Función de Autocorrelación del Ejemplo 8.2.1

(b) fac muestral de un $MA(3)$

Figura 8.1: Fac teóricas de modelos $MA(q)$

Conclusión. De acuerdo con (8.4), si la fac muestral de una serie Y_t termina abruptamente puede tratarse de un $MA(q)$. Por ejemplo, en la Figura 8.1(b) sería factible un modelo $MA(3)$.

4. Un $MA(q)$ siempre es un proceso estacionario con fac, $\rho(k) = \frac{R(k)}{R(0)}$, $k \in \mathbb{Z}$.
5. Los $MA(q)$ son aproximadores universales de procesos estacionarios en covarianza (lineales). La posibilidad de encontrar un $MA(q)$ óptimo (aunque el orden q puede ser alto) está garantizada por un resultado teórico que dice que dado un proceso estacionario en covarianza que cumple cierta condición técnica (su densidad espectral deber ser continua en $[-\pi, \pi]$), entonces siempre existirá un $MA(q)$ que lo aproxime con un grado deseado de error (ver Brockwell and Davis [2002], pag. 132, Corollary 4.4.1).
6. La simulación de un proceso $MA(q)$ se puede hacer con la función `filter()`, descrita en la sección 6.1, pag.122, ya que el proceso es una media móvil unilateral de un ruido blanco. Si se denota por `et` el vector de ruido blanco con varianza σ^2 y se indica por `theta` un vector con los parámetros $\theta_1, \dots, \theta_q$, se simulan n valores de $Y_t \sim MA(q)$ ejecutando el comando

```
et = rnorm(n, 0, sigma)
y = filter(et, theta, "conv", sides=1, circular=TRUE) .
#Produce el mismo resultado que
y=arima.sim(list=(order(0, 0, q), ma=theta), n=n, sd=sigma) .
```

8.2.2. Condición de Invertibilidad del Proceso $MA(q)$

Los procesos $MA(q)$ pueden presentar un problema de falta de identificabilidad. Puede ocurrir que existan dos conjuntos de parámetros $\underline{\theta}^{(i)} = (\theta_1^{(i)}, \dots, \theta_q^{(i)})$, para $i = 1, 2$, con $\underline{\theta}^{(1)} \neq \underline{\theta}^{(2)}$ y sin embargo se cumple que las respectivas autocovarianzas (8.4) son iguales, $R(k, \underline{\theta}^{(1)}) \equiv R(k, \underline{\theta}^{(2)})$. Si se cumple la condición siguiente, “de invertibilidad”, no ocurre este problema.

Definición 8.2.2. Dado un proceso $MA(q)$, $Y_t = \theta_q(L)(\varepsilon_t)$ donde

$$\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q,$$

Si las q raíces de la ecuación

$$\Theta_q(z) = 0,$$

$z_1, z_2, \dots, z_q \in \mathbb{C}$, cumplen $|z_j| > 1$, se dice que el proceso Y_t es invertible.

Es decir, si $\Theta_q(z)$ cumple que “sus raíces están fuera del círculo unitario”, lo cual se expresa también como

$$\Theta_q(z) \neq 0, \forall z \in \mathbb{C}, |z| \leq 1. \quad (8.5)$$

Ejemplo 8.2.3. Consideremos el proceso $Y_t \sim MA(2)$ en el Ejemplo 8.2.1,

$$Y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.4\varepsilon_{t-2},$$

y veamos si es invertible. Hallamos las raíces del polinomio $\Theta_2(z)$

$$\begin{aligned} \Theta_2(z) &= 1 - 0.4z + 0.4z^2 = 0, \\ \therefore z &= \frac{0.4 \pm \sqrt{0.4^2 - 4(0.4)(1)}}{2(0.4)} = \frac{1}{2} \pm \frac{3}{2}i \end{aligned}$$

por tanto

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

luego Y_t es invertible.

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

Definición 8.2.3 (Función de Autocorrelación Parcial (facp)). Suponga que $(Y_t, t \in \mathbb{Z})$ es serie estacionaria. La facp es una función de k , $\alpha(k)$, $k = 1, 2, \dots$ definida por

$$1. \alpha(1) = \rho(1)$$

$$2. \alpha(k) = \text{Corr}(U_1, U_k) \text{ donde}$$

$$U_1 = Y_1 - E(Y_1|Y_2, \dots, Y_k)$$

$$U_k = Y_{k+1} - E(Y_{k+1}|Y_2, \dots, Y_k), \quad k = 2, \dots$$

Lo que la facp hace es correlacionar Y_1 y Y_{k+1} pero eliminando primero el efecto de las variables intermedias Y_2, \dots, Y_k . 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 sólo si se asume la condición de invertibilidad para un $MA(q)$. En este caso se tiene el siguiente resultado:

$$\lim_{k \rightarrow \infty} \alpha(k) = 0, \quad (8.6)$$

y se concluye que la facp de un $MA(q)$ siempre decrece a cero.

Ejemplo 8.2.4. En las Figuras siguientes 8.2 se observa la fac y la facp de un $MA(3)$ invertible.

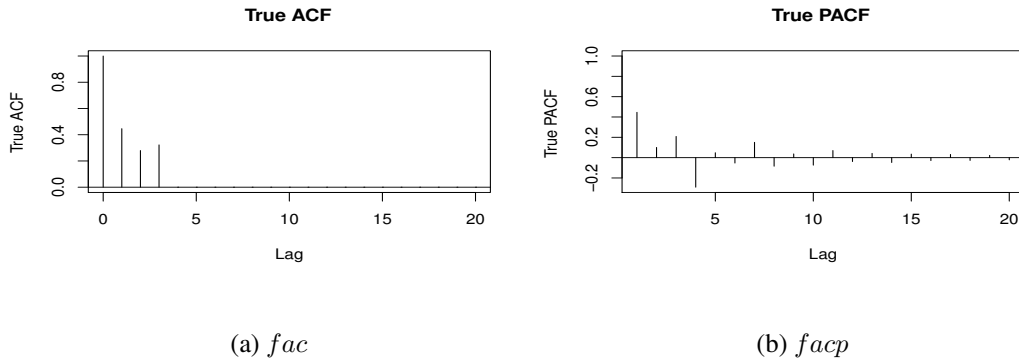


Figura 8.2: fac y facp de un $MA(3)$ invertible.

La prueba de (8.6) es como sigue.

Demostración. Suponga que se cumple la condición de invertibilidad (8.5) para el polinomio

$$\Theta_q(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$$

Entonces la función $\Theta_q(z)^{-1} = 1/\Theta_q(z)$ se puede desarrollar en una serie de Taylor alrededor de $z = 0$, de la forma

$$\Theta_q(z)^{-1} = \sum_{j=0}^{\infty} \psi_j z^j, \quad (8.7)$$

para $|z| < 1$ y tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\psi_0 = 1$.

Si se reemplaza z por L se tiene un operador en el rezago L de la forma $\Theta_q(L)^{-1} = \sum_{j=0}^{\infty} \psi_j L^j$, el cual, aplicado a Y_t produce

$$\begin{aligned} \Theta_q(L)^{-1} Y_t &= \sum_{j=0}^{\infty} \psi_j L^j (Y_t) = \sum_{j=0}^{\infty} \psi_j Y_{t-j} \\ &= Y_t + \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \dots \end{aligned}$$

Ahora, multiplicando ambos miembros de por $\Theta_q(L)^{-1}$ se obtiene

$$\begin{aligned} Y_t &= \Theta_q(L) \varepsilon_t \\ \Leftrightarrow \Theta_q(L)^{-1} Y_t &= \Theta_q(L)^{-1} \Theta_q(L) \varepsilon_t = \varepsilon_t \\ \therefore \Theta_q(L)^{-1} Y_t &= \varepsilon_t \\ \Leftrightarrow Y_t + \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \dots &= \varepsilon_t \\ \therefore Y_t &= -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \dots + \varepsilon_t, \end{aligned}$$

Por tanto, si se hace la regresión de Y_t sobre los primeros k rezagos Y_{t-j} , $j = 1, \dots, k$,

$$Y_t = -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \dots - \psi_k Y_{t-k} + \varepsilon_t \quad (8.8)$$

entonces el k -ésimo coeficiente es $\alpha(k) = \psi_k$. Pero la serie $\sum_{k=0}^{\infty} \psi_k$ es absolutamente convergente, entonces $\psi_k \rightarrow 0$. Por tanto, $\alpha(k) \rightarrow 0$ cuando $k \rightarrow \infty$ y se concluye que la facp de un MA(q) siempre decrece a cero. \square

8.2.4. Estimación de procesos MA(q)

La estimación de procesos $Y_t \sim MA(q)$ no es directa. Una de las razones es que en $Y_t = \theta_q(L) \varepsilon_t$ los valores del ruido blanco ε_t no son observables. Supongamos una muestra de tamaño T de la serie, Y_1, \dots, Y_T .

En el caso de estimadores de mínimos cuadrados ordinarios ver (2.7), pag.30, el estimador de $\underline{\theta} = (\theta_1, \dots, \theta_q)'$ se define como el vector $\hat{\underline{\theta}}$ que minimiza la función

$$G(\underline{\theta}) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left(Y_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \right)^2, \quad (8.9)$$

Pero los ε_{t-j} no son observables. Se puede recurrir a la invertibilidad, y utilizar la serie en (8.8) pero truncada hasta un cierto n . Permitiendo seguir con el signo igual, se escribe

$$Y_t = -\psi_1 Y_{t-1} - \psi_2 Y_{t-2} - \dots - \psi_n Y_{t-n} + \varepsilon_t. \quad (8.10)$$

Considerando la identidad

$$\Theta_q(z)^{-1} \Theta_q(z) = (1 + \psi_1 z + \psi_2 z^2 + \dots)(1 + \theta_1 z + \dots + \theta_q z^q) = 1$$

se pueden despejar los ψ_j en función del vector $\underline{\theta} = (\theta_1, \dots, \theta_q)'$, escribiendo $\psi_j = \psi_j(\underline{\theta})$. La minimización se hace con base en (8.10) con la función objetivo

$$G(\underline{\theta}) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T \left(Y_t + \sum_{j=1}^n \psi_j(\underline{\theta}) Y_{t-j} \right)^2. \quad (8.11)$$

El correspondiente estimador $\hat{\underline{\theta}}$ se denomina estimador de mínimos cuadrados condicionales. Es asintóticamente eficiente y tiene las mismas propiedades asintóticas del estimador de máxima verosimilitud.

En R hay varias funciones para estimar procesos MA(q). La función `arma` de la librería `tseries` y la función `arima`. La primera calcula los valores p de los coeficientes, y la segunda permite calcular pronósticos. Ambas utilizan estimadores de mínimos cuadrados condicionales. El siguiente código muestra cómo estimar y pronosticar un proceso MA(q).

Ejemplo 8.2.5. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

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

Al estimar los residuos estructurales $\hat{\varepsilon}_t$ se observó que posiblemente muestran una dinámica autocorrelacionada. Se confirmó este hecho con las pruebas Ljung-Box y Durbin-Watson. En este Ejemplo se investiga si los residuos podrían seguir una dinámica MA(q).

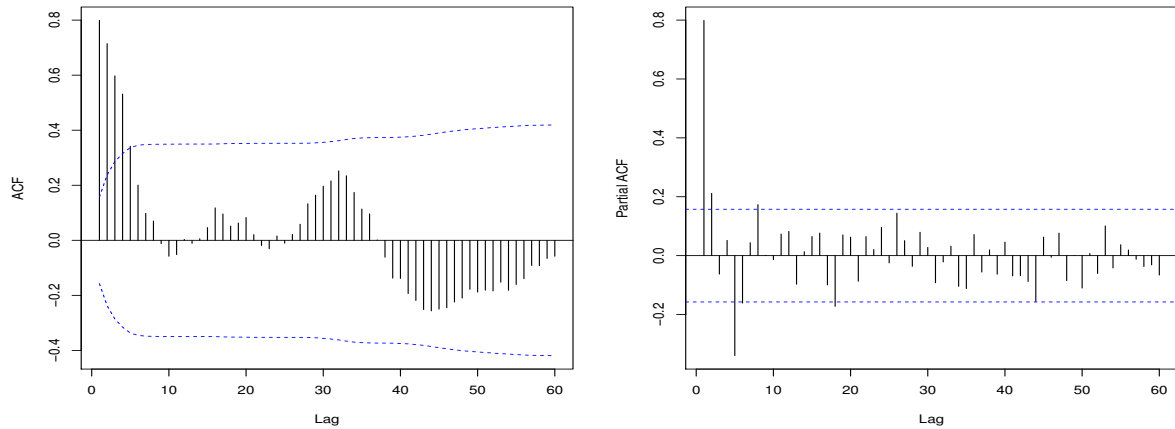
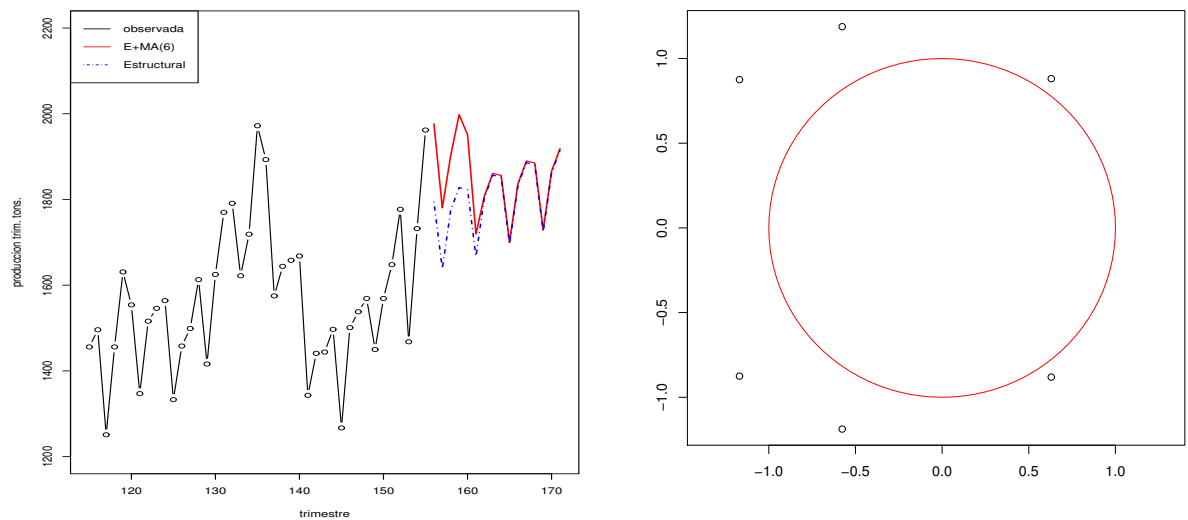


Figura 8.3: Residuos estructurales de la serie de producción de cemento



(a) Pronósticos

(b) Raíces círculo unitario

Figura 8.4: (a) Pronosticos Estructurales versus Estructurales + MA(6), (b) Raíces y círculo unitario

En la Figura 8.3 se muestran la fac y la fac parcial de los residuos estructurales, que muestran que un MA(q) es factible, con $q \geq 4$. El código R siguiente muestra que un

MA(6) puede presentar mejor ajuste. En la Figura 8.4, panel izquierdo, se muestran pronósticos con el modelo estructural y el modelo MA(6) para los residuos. Las raíces del polinomio del modelo MA(6) se grafican en la Figura 8.4, panel derecho. El código R siguiente muestra la estimación y análisis del MA(6).

Estima MA(6) para los residuos con dos funciones de estimación: arma y arima.

```
#-----
require(tseries)
mod4.1 = arma(r,order = c(0,6))
summary(mod4.1)
mod4.2 = arima(r,order = c(0,0,6))
summary(mod4.2)
```

Examinar residuos del MA(6)

```
#-----
r4.2 = residuals(mod4.2)

par(mfrow=c(2,2))
plot(t,r4.2,type='o',ylab='residuo')
abline(h=0,lty=2)
plot(density(r4.2),xlab='x',main= '')
acf(r4.2,60,ci.type="ma",drop.lag.0=TRUE,main="")
pacf(r4.2,60,main="")
cpgram(r4.2)
#----- prueba Ljung-Box
Box.test(r4.2, lag = 12, type = "Ljung-Box")
X-squared = 9.5687, df = 12, p-value = 0.6537
```

Calcular 16 pronósticos estructurales añadiendo 16 pronósticos MA(6).

```
m = 16
tp = seq(T+1,T+m)
ltp = seasonaldummyf(y,m)
# el objeto mod1 es el modelo estructural
```

```

y.p = predict(mod1,data.frame(t=tp,It=I(Itp)))
r.p = predict(mod4.2,n.ahead=m)$pred
#----- pronosticos conjuntos
yr.p = y.p+r.p

#-----
require(polynom)
theta = coef(mod4.2); theta = c(1,theta[-7]);
(Mod(polyroot(theta)))
require(signal)
zplane(filt=rev(theta),a=c(1))

#----- graficar
m0 = 40; N = seq(T-m0,T);
par(mfrow=c(1,1))
plot(seq(T-m0,T),y[N],type='b',
ylim=c(1200,2200),xlim=c(T-m0,T+m),
ylab='produccion trim. tons.',xlab='trimestre')
lines(seq(T+1,T+m),yr.p,lty=7,lwd=2,col='red')
lines(seq(T+1,T+m),y.p,lty=4,lwd=2,col='blue')
legend("topleft",
legend=c("observada","E+MA(6)","Estructural"),
col=c('black','red','blue'),lty=c(1,7,4))

```

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

Definición 8.3.1 (Proceso AR(p)). *Un proceso $(Y_t, t \in \mathbb{Z})$ sigue un modelo AR(p) de media cero si es solución de la ecuación recursiva siguiente.*

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \quad (8.12)$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$, $\varphi_j, j = 1, 2, \dots, p$ son los coeficientes (parámetros) y $p = 1, 2, \dots$ es el orden. El caso $p = 0$ se interpreta como $Y_t = \varepsilon_t$.

Se define

$$\Phi_p(z) = 1 - \sum_{j=1}^p \varphi_j z^j \quad (8.13)$$

el polinomio autorregresivo. Usando el operador de rezago L se puede escribir (8.12) como

$$\Phi_p(L)(Y_t) = \varepsilon_t.$$

Un proceso $Y_t \sim AR(p)$ con media diferente de cero, $\mathbb{E}(Y_t) = \mu \neq 0$, se define mediante la ecuación

$$\Phi_p(L)(Y_t - \mu) = \varepsilon_t. \quad (8.14)$$

Obsérvese que

$$\begin{aligned} \Phi_p(L)(Y_t - \mu) &= \Phi_p(L)(Y_t) - \Phi_p(L)(\mu) \\ &= \Phi_p(L)(Y_t) - \mu(1 - \sum_{j=1}^p \varphi_j) = \varepsilon_t. \end{aligned}$$

Luego,

$$\begin{aligned} \Phi_p(L)(Y_t) &= \varphi_0 + \varepsilon_t \\ \Leftrightarrow Y_t &= \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \end{aligned}$$

con $\varphi_0 = \mu(1 - \sum_{j=1}^p \varphi_j)$. La constante φ_0 se denomina intercepto.

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

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

$$|z_i| > 1. \quad (8.15)$$

Es decir, si $\Phi_p(z)$ cumple que “sus raíces están fuera del círculo unitario”, expresado como

$$\forall z \in \mathbb{C}, |z| \leq 1, \Phi_p(z) \neq 0. \quad (8.16)$$

En la Figura 8.5 se observa el círculo unitario junto con una raíz z_j y su conjugado \bar{z}_j , ambas fuera del círculo.

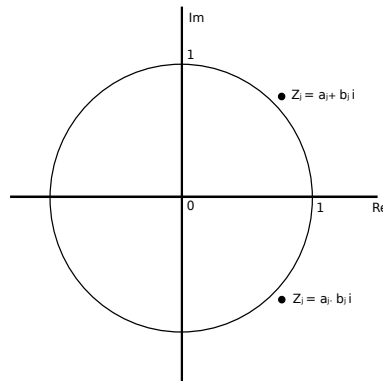


Figura 8.5: Círculo Unitario

Sin embargo, la condición (8.16) no es necesaria. Puede suceder que algunas de las raíces de $\Phi_p(z) = 0$ estén adentro y otras afuera del círculo unitario, pero si se cumple que $\Phi_p(z) \neq 0, \forall z \in \mathbb{C}, |z| = 1$, en este caso Y_t es estacionario en covarianza.

Ejemplo 8.3.1. El proceso $Y_t \sim AR(2)$,

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t, \quad (8.17)$$

cumple que es estacionario en covarianza ya que las raíces características de $\Phi_2(z) = 1 - 1.5z + 0.9z^2 = 0$, que son $\frac{5}{6} \pm \frac{\sqrt{15}i}{6}$, están fuera del círculo unitario. Se puede hacer la comprobación con las funciones `InvertibleQ()` de la librería `FitAR`, y `armaRoots()` de la librería `fArma`, pero que se puede cargar como programa independiente, como sigue.

```
phi = c(1.5, -0.9)
require(FitAR)
InvertibleQ(phi)
[1] TRUE
source("armaRoots.r")
armaRoots(phi)
      re      im      dist
```

1 0.8333 0.6455 1.0541
 2 0.8333 -0.6455 1.0541

Para un proceso $Y_t \sim AR(p)$ de media cero, que cumple la condición para ser estacionario en covarianza (8.16), se cumple:

$$i) \quad \mathbb{E}(Y_t) = 0, \quad (8.18)$$

$$ii) \quad \sum_{j=1}^p \varphi_j < 1. \quad (8.19)$$

Demostración. i). Para comprobar (8.18), si Y_t es estacionario en covarianza entonces $\mathbb{E}(Y_t) \equiv \mu$. Además,

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

pero todas las esperanzas son iguales a μ , luego

$$\mu = \varphi_1 \mu + \varphi_2 \mu + \cdots + \varphi_p \mu.$$

Si $\mu \neq 0$ entonces $1 = \varphi_1 + \cdots + \varphi_p$, y por tanto el polinomio autorregresivo evaluado en $z = 1$ se anula, $\Phi_p(1) = 0$, lo cual es una contradicción ya que $\forall z \in \mathbb{C}, |z| \leq 1$ $\Phi_p(z) \neq 0$. Luego debe tenerse que $\mu = 0$, es decir, el proceso definido en (8.12) es de media cero. \square

Demostración. ii). Usando la técnica de reducción al absurdo, para comprobar (8.19), se asume lo contrario, es decir $\sum_{j=1}^p \varphi_j \geq 1$. Si ocurre el caso $\sum_{j=1}^p \varphi_j = 1$, equivale a $\Phi_p(1) = 0$, lo cual es una contradicción con la condición (8.16). Si se asume $\sum_{j=1}^p \varphi_j > 1$ entonces $\Phi_p(1) < 0$. Pero $\Phi_p(0) = 1$. Como la función $\Phi_p(z)$ es continua en $[0, 1]$ debe existir un punto $0 < z < 1$ en el cual $\Phi_p(z) = 0$. Pero eso es nuevamente una contradicción con la condición (8.16), porque este z sería una raíz de $\Phi_p(z) = 0$ que cumple $|z| < 1$. \square

8.3.2. La Función de Autocovarianza de un Proceso AR(p)

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

de Yule–Walker, dada por

$$R(k) = \sum_{j=1}^p \varphi_j R(k-j), \quad k = 1, 2, \dots, \quad (8.20)$$

donde el valor inicial está dado por $R(0) = \sum_{j=1}^p \varphi_j R(j) + \sigma^2$.

A partir de (8.20) es inmediato que la fac $\rho(k) = R(k)/R(0)$ cumple una ecuación similar, al dividir estas ecuaciones por $R(0)$.

$$\rho(k) = \sum_{j=1}^p \varphi_j \rho(k-j), \quad k = 1, 2, \dots \quad (8.21)$$

La ecuación (8.21) es una ecuación recursiva lineal de orden p , homogénea. Requiere p valores iniciales

$$\rho(0) = 1, \rho(1) = q_1, \dots, \rho(p-1) = q_{p-1}.$$

Se puede resolver por fórmula ó utilizando un software de cálculo simbólico como por ejemplo, Maple ó Maxima ⁽¹⁾.

Demostración. Con $\mu = \mathbb{E}(Y_t)$, la ecuación $Y_t = \varphi_0 + \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t$ equivale a $Y_t - \mu = \sum_{j=1}^p \varphi_j (Y_{t-j} - \mu) + \varepsilon_t$. Multiplicando ambos miembros por $Y_{t-k} - \mu$, con $1 \leq k \leq t$, y tomando valor esperado $\mathbb{E}(\cdot)$ se obtiene

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

Ya que se cumple $\mathbb{E}(\varepsilon_t(Y_{t-k} - \mu)) = 0$ porque, a partir de la definición del proceso Y_t en (8.12), Y_{t-k} depende de ε_s con $s \leq t-k$, que son variables incorrelacionadas con ε_t . El caso $k = 0$ es similar, con $\mathbb{E}(\varepsilon_t(Y_t - \mu)) = \text{Var}(\varepsilon) = \sigma^2$. \square

Ejemplo 8.3.2. Retomando el Ejemplo 8.3.1, para el proceso $Y_t \sim AR(2)$, $Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t$, estacionario en covarianza, con $\varepsilon_t \sim iidN(0, \sigma^2)$, $\sigma^2 = 9$. Encontrar su autocorrelación.

¹<http://maxima.sourceforge.net/es/index.html>

La ecuación de Yule-Walker para $\rho(k)$

$$\rho(k) = 1.5\rho(k-1) - 0.9\rho(k-2), \quad k \geq 2. \quad (8.22)$$

Hay que encontrar las condiciones iniciales, $\rho(0), \rho(1)$. Primero se encuentra la varianza del proceso, $R(0)$.

$$\begin{aligned} R(0) &= \sum_{j=1}^2 \varphi_j R(j) + \sigma^2 \\ \Leftrightarrow 1 &= 1.5\rho(1) - 0.9\rho(2) + \sigma^2/R(0) \\ \Leftrightarrow R(0) &= \frac{\sigma^2}{1 - 1.5\rho(1) + 0.9\rho(2)} \end{aligned}$$

Reemplazando $k = 1$ en (8.22) y con $\rho(0) = 1$ se obtiene

$$\begin{aligned} \rho(1) &= 1.5\rho(0) - 0.9\rho(1) \\ \therefore \rho(1) &= \frac{15}{19}. \end{aligned}$$

Se puede encontrar la solución de (8.22) usando una teoría básica de ecuaciones lineales recurrentes homogéneas, ver ⁽²⁾. Con Maxima se obtiene

$$\rho(n) = \frac{(2\sqrt{10})^n \sqrt{15} 3^{n-1} \sin\left(\operatorname{atan}\left(\frac{\sqrt{15}}{5}\right) n\right)}{19 \times 20^n} + \frac{(2\sqrt{10})^n 3^n \cos\left(\operatorname{atan}\left(\frac{\sqrt{15}}{5}\right) n\right)}{20^n}$$

Nótese que en este caso la fac tiene una semi-periodicidad, de período

$$\frac{2\pi}{\arctan\left(\frac{\sqrt{15}}{5}\right)} = 9.5,$$

y además es amortiguada exponencialmente, tendiendo a cero.

La función de autocovarianza, $R_X(k)$, se pueden calcular en el lenguaje R con la función `ARMAacf()` como sigue. La gráfica está en la Figura 8.6,

```
phi = c(1.5, -0.9)
acf.ar2 = ARMAacf(ar=phi, ma=numeric(0), lag.max=30)
str(acf.ar2)
Named num [1:31] 1 0.789 0.284 -0.284 -0.682 ...
- attr(*, "names")= chr [1:31] "0" "1" "2" "3" ...
```

²https://en.wikipedia.org/wiki/Recurrence_relation#Roots_of_the_characteristic_polynomial

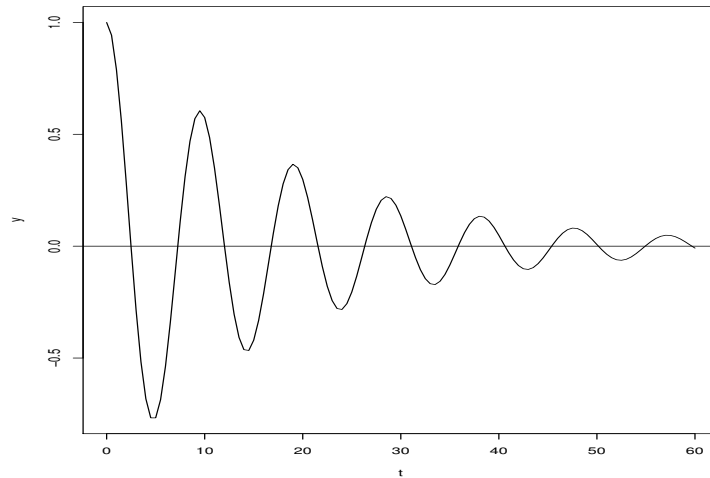


Figura 8.6: fac de ar(2)

8.3.3. Función fac parcial de un Proceso AR(p)

El razonamiento para determinar la facp de un AR(p) es directo. La facp de un proceso $AR(p)$ es $\alpha(k)$ tal que $\hat{\alpha}(k)$ es el coeficiente $\hat{\beta}_{k,k}$ en la regresión

$$Y_t = \beta_0 + \beta_{k,1}Y_{t-1} + \cdots + \beta_{k,k}Y_{t-k} + \eta_t, \quad k = 2 \quad (8.23)$$

pero como $\beta_{k,k} = 0$ si $k \geq p + 1$ entonces $\hat{\alpha}(k) = 0$ si $k \geq p + 1$. Se concluye que la fac parcial de un AR(p) es cero para rezagos $k \geq p + 1$.

Ejemplo 8.3.3. Retomando el Ejemplo 8.3.1, para el proceso $Y_t \sim AR(2)$,

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t,$$

estacionario en covarianza, la función de autocovarianza parcial, $\alpha_X(k)$, se pueden calcular en R con la función siguiente

```
phi = c(1.5, -0.9)
PAcf = ARMAacf(ar=phi,
ma=numeric(0), lag.max=30, pacf=TRUE)
str(PAcf)
num [1:30] 7.89e-01 -9.00e-01 3.10e-15 -8.67e-16 -9.40e-16 ...
```


8.3.4. Estimación de procesos AR(p)

La estimación de un proceso $Y_t \sim AR(p)$ es directa y existen varios métodos para realizarla. Si se tiene una muestra del proceso, Y_1, \dots, Y_T , se pueden utilizar:

1. Los estimadores Yule-Walker con base en la fac muestral. Se basan en la siguiente identidad. La fac de un $AR(p)$, $\rho(k) = \text{Corr}(Y_t, Y_{t+k})$, $k = 1, 2, \dots, p, p+1, \dots$, cumple lo siguiente.

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

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

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

$$A\underline{\varphi} = \underline{\rho}. \quad (8.24)$$

Luego, dados los estimadores de la fac $\hat{\rho}(1), \dots, \hat{\rho}(p)$ se pueden reemplazar en la expresiones para A y $\underline{\rho}$, y resolver (8.24) para obtener

$$\hat{\underline{\varphi}} = \hat{A}^{-1}\hat{\underline{\rho}}. \quad (8.25)$$

Estos estadísticos $\hat{\underline{\varphi}}$ así calculados se denominan los estimadores de Yule-Walker de $\underline{\varphi}$.

2. Estimadores por mínimos cuadrados ordinarios. En el caso de estimadores de mínimos cuadrados ordinarios, ver (2.7), pag.30, el estimador de $\underline{\varphi} = (\varphi_1, \dots, \varphi_p)'$ se define como el vector $\hat{\underline{\varphi}}$ que minimiza la función

$$G(\underline{\varphi}) = \sum_{t=1+p}^T \varepsilon_t^2 = \sum_{t=1+p}^T \left(Y_t - \sum_{j=1}^p \varphi_j Y_{t-j} \right)^2. \quad (8.26)$$

Un problema que surge con estos estimadores es que el vector estimado $\hat{\underline{\varphi}}$ puede no cumplir la condición (8.16), para que la solución Y_t sea estacionaria en covarianza. El método de Burg permite corregir esta posibilidad.

3. estimadores de máxima verosimilitud, asumiendo que los errores cumple $\varepsilon_t \sim iidN(0, \sigma^2)$,

Los métodos de máxima verosimilitud y Burg se pueden consultar en Brockwell and Davis [2002] y Grandell [2014]

Estimación de un proceso AR(p) en R

La estimación de un proceso AR(p) en R se puede hacer mediante varias funciones en R. La función `ar()` de la librería `stats` provee la estimación por los métodos: mínimos cuadrados, Yule-Walker, máxima verosimilitud y Burg.

```
ar(x, aic = TRUE, order.max = NULL,
    method = c("yule-walker", "burg", "ols", "mle", "yw")
#-----
aic : Logical flag. If TRUE then the Akaike Information
Criterion is used to choose the order of the
autoregressive model. If FALSE, the model of order
order.max is fitted.
```

Ejemplo 8.3.4. Retomando el Ejemplo 8.3.1, para el proceso $Y_t \sim AR(2)$,

$$Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t,$$

estacionario en covarianza, los resultados de una estimación con base en una muestra simulada, utilizando `ar()` con las opciones “ols”, “burg”, “yw” y “mle” se muestran a continuación.

	ols	burg	yw	mle
1	1.48	1.48	1.48	1.50
2	-0.88	-0.88	-0.88	-0.90

Tabla 8.1: Estimadores OLS, Burg, Y-W, MaxVer en un AR(2)

La función `FitAR()` de la librería `FitAR` utiliza un algoritmo para máxima verosimilitud. Por ejemplo, las instrucciones siguientes determinan un orden p óptimo mediante el mínimo AIC, considerando todas las posibilidades hasta un máximo, en el ejemplo de 30.

```
p = SelectModel(y, lag.max=32, Best=1)
m2 = FitAR(y, p)
coef(m2)
```

La función `arima()` de `stats` utiliza máxima verosimilitud. Un inconveniente con esta función es que reporta la media con el nombre de intercepto. Así, podría generar confusión ya que intercepto se refiere a φ_0 .

8.3.5. Simulación de procesos AR(p)

La simulación de un proceso AR(p) se puede hacer mediante la función `filter()`, escribiendo `y=filter(et, phi, "recursive", init=y0)`, donde `phi=c(f1, ..., fp)` es el vector de coeficientes φ_j en la ecuación $Y_t = \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t$, y $(\varepsilon_t) \sim RB(0, \sigma^2)$. El vector `y0=c(y1, ..., yp)` consiste de los primeros p valores necesarios para iniciar la recursión AR(p).

Ejemplo 8.3.5. *En el caso del Ejemplo 8.3.1 la instrucción es*

```
et = rnorm(1000, 0, 2)
y=filter(et, c(1.5, -0.9), "recursive", init=y0)
```

donde `y0=c(0, 0)`, si se toman los valores iniciales como la media del proceso, que es cero.

La simulación de un proceso AR(p) de media μ diferente de cero, se realiza colocando `y=mu+filter(et, phi, "recursive", init=y0)`

8.3.6. Análisis de un caso de aplicación de AR(p)

La serie de aportes hidrológicos diarios en Colombia muestra la cantidad de agua convertida en energía, en todas las represas para generación de electricidad en el país. La unidad es Gw/hora. Para los datos entre Julio 1 de 1998 y 31 de Mayo de 2014, la gráfica de los aportes, Y_t está en la Figura 8.7. Inicialmente se propone un

modelo de componentes lineal y trigonométricas, de la forma

$$Y_t = b_0 + b_1 t + \sum_{j=1}^k \alpha_j \cos(2\pi j t / 365) + \beta_j \sin(2\pi j t / 365) + \varepsilon_t, \quad (8.27)$$

donde (ε_t) es una serie con autocorrelacionada, que posiblemente puede ser AR(p). Los análisis en el programa R siguiente muestran que este supuesto es correcto.

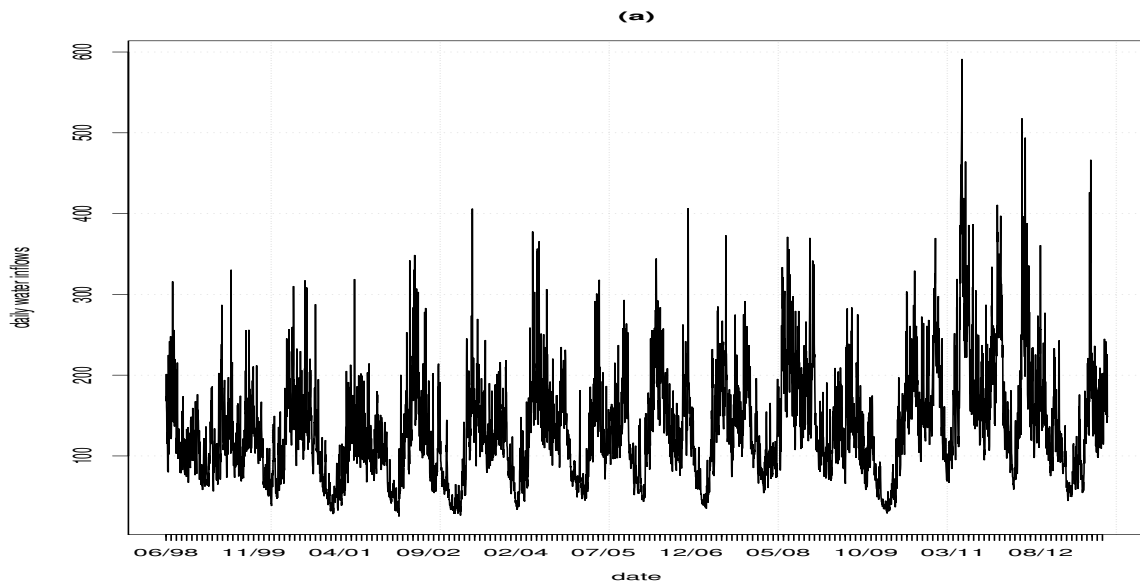


Figura 8.7: Aportes hidrológicos en Gw/h, 01/07/98 - 31/05/14

```
# Ejemplo de analisis AR(p)
# Serie de aportes hidrológicos diarios
# en Gw/h
#-----modela estacionalidad con sen, cos
#----- serie con frecuencia diaria
Ap = ts(Ap,frequency = 365)
require(forecast)
It.trig = fourier(Ap,4)
t = seq(1,length(Ap))/100
mod2 = lm(Ap ~ t + It.trig)
summary(mod2)
```

Los resultados son aceptables, aunque el R^2 está bajo.

```
Call: lm(formula = Ap ~ t + It.trig)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	104.80070	1.34523	77.906	< 2e-16	***
t	1.00769	0.04182	24.094	< 2e-16	***
It.trigS1-365	32.45044	0.95059	34.137	< 2e-16	***
It.trigC1-365	24.35062	0.95159	25.589	< 2e-16	***
It.trigS2-365	-11.17645	0.95123	-11.749	< 2e-16	***
It.trigC2-365	31.80005	0.95038	33.460	< 2e-16	***
It.trigS3-365	0.43385	0.95037	0.457	0.648	
It.trigC3-365	-3.70841	0.95124	-3.899	9.79e-05	***
It.trigS4-365	-7.41607	0.95115	-7.797	7.52e-15	***
It.trigC4-365	4.54505	0.95022	4.783	1.77e-06	***

```
---
```

```
Multiple R-squared: 0.4021, Adjusted R-squared: 0.4012
```

```
F-statistic: 415.6 on 9 and 5561 DF, p-value: < 2.2e-16
```

Identificar los residuos con fac y fac parcial en la Figura 8.8.

```
#-----examen fac y fac parcial residuos
r = residuals(mod2)
require(TSA)
par(mfrow=c(1,2))
TSA::acf(r,60, drop.lag.0 = TRUE, ci.type="ma")
pacf(r,60)
```

Se puede concluir que los residuos estructurales pueden ser modelados mediante un AR(p), con $p \geq 9$. Para identificar p se puede usar la función `SelectModel`

```
#-----identifica AR(p)
#install.packages("FitAR")
require(FitAR)
n = length(Ap)
```

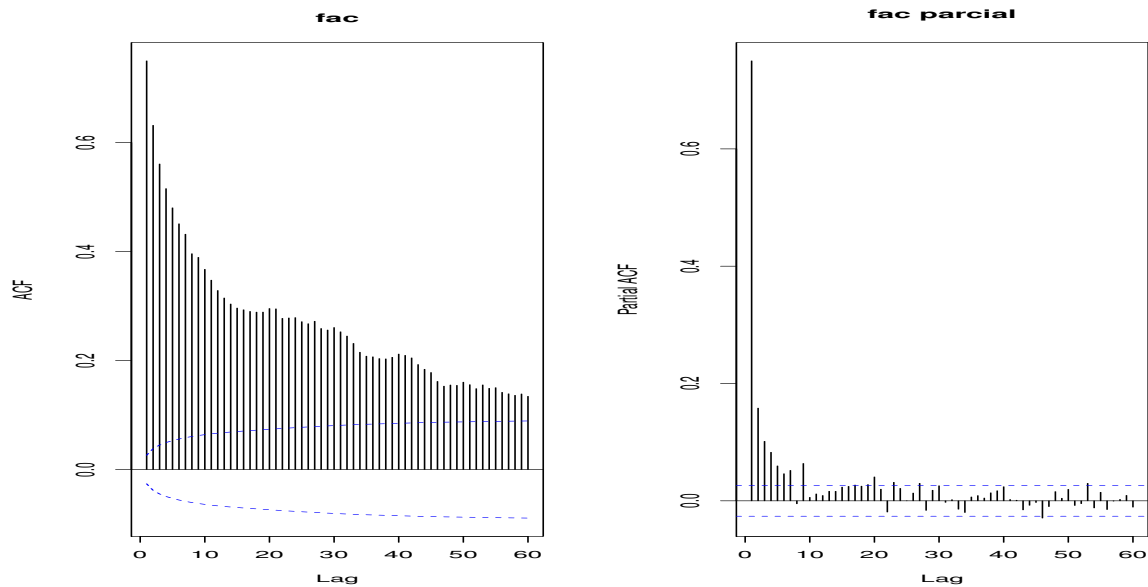


Figura 8.8: fac y facp de Aportes hidrológicos

```
pvec = SelectModel(r, ARModel="AR",
Criterion="BIC",
lag.max=floor(10*log10(n)), Best=1)
(p=pvec)
9
```

ó también la función ar

```
# aic Logical flag. If TRUE then the Akaike Information
# Criterion is used
mod1 = ar(r,aic = TRUE, method=c("burg"))
print(mod1)
Order selected 30
```

Los residuos del modelo AR(30) deben ser ruido blanco. La Figura 8.9 muestra que la fac está dentro de las bandas de Bartlett. Finalmente, la prueba Ljung-Box muestra que no se rechaza la nula de ruido blanco.

```
Box.test(et,50,type= "Ljung-Box")
```

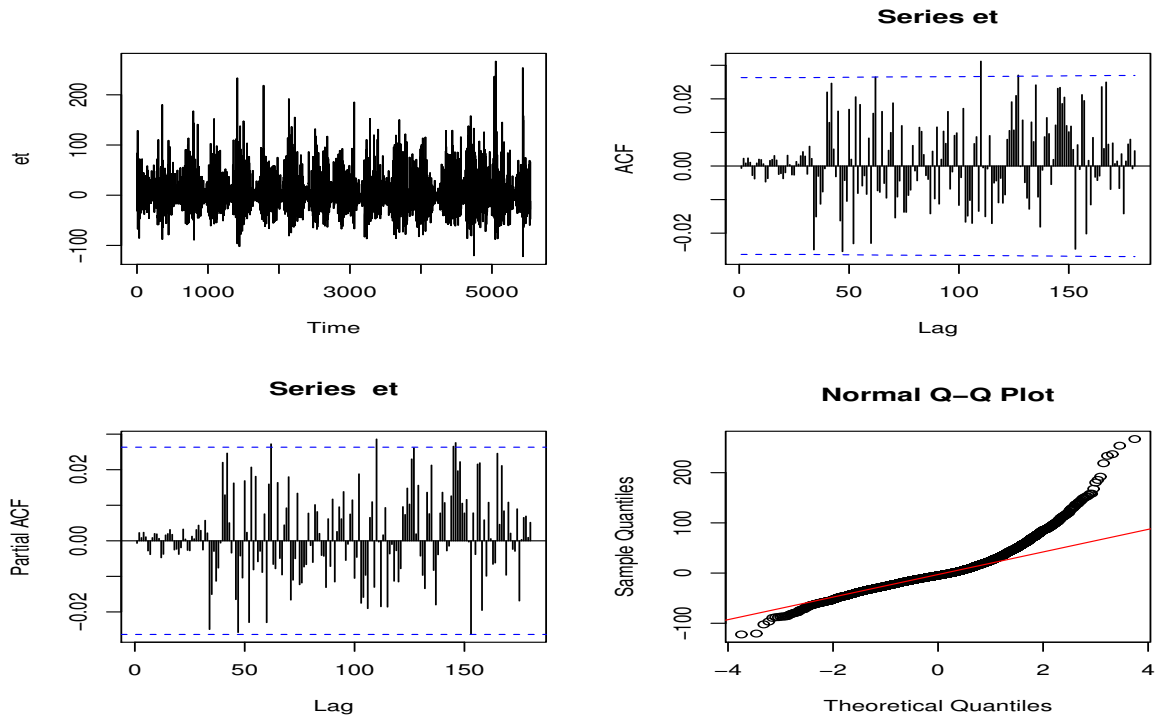


Figura 8.9: fac y facp de residuos de AR(30)

Box-Ljung test

X-squared = 23.374, df = 50, p-value = 0.9995

8.4. Procesos Autoregresivos y de Medias Móviles AR-MA(p,q)

Una manera de introducir los procesos ARMA(p,q) es la siguiente. Si en un proceso $Y_t \sim AR(p)$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}$$

donde $\varepsilon_t \sim RB(0, \sigma^2)$, se reemplaza ε_t por un proceso $MA(q)$,

$$Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

entonces se obtiene un proceso ARMA(p,q),

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + Z_t \\ &= \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \end{aligned}$$

Por tanto, un proceso ARMA(p,q) es un modelo que combina las propiedades de memoria larga de los AR(p) con las propiedades de ruido débilmente autocorrelacionado en los MA(q), y que tiene suficiente flexibilidad y parsimonia para representar una variedad grande de procesos estacionarios en covarianza.

Usando la notación del operador de rezago L se puede escribir un proceso $ARMA(p, q)$ como

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ \Leftrightarrow Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \cdots - \varphi_p Y_{t-p} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ \Leftrightarrow \Phi_p(L)(Y_t) &= \Theta_q(L)(\varepsilon_t), t \in \mathbb{Z}, \end{aligned} \quad (8.28)$$

donde

$$\Phi_p(z) = 1 - \sum_{j=1}^p \varphi_j z^j, \quad p = 1, 2, \dots \quad (8.29)$$

$$\Theta_q(z) = 1 + \sum_{j=1}^q \theta_j z^j, \quad q = 1, 2, \dots \quad (8.30)$$

son los polinomios autoregresivo y de media móvil, respectivamente.

Un proceso $Y_t \sim ARMA(p, q)$ de media cero, se define como una solución de la ecuación recursiva (8.28), dados los coeficientes φ_j, θ_j , y dado el ruido blanco $(\varepsilon_t, t \in \mathbb{Z})$. En la práctica, sin embargo, se trata es de un problema inverso: dada la serie (Y_1, \dots, Y_T) , encontrar los coeficientes (parámetros) y estimar el ruido blanco.

Las condiciones de estacionariedad de la parte AR(p) y de invertibilidad de la parte MA(q) se asumen en el modelo ARMA(p,q), (8.28). Es decir, se asume que las raíces de las ecuaciones $\Phi_p(z) = 0$ y $\Theta_q(z) = 0$ están fuera del círculo unitario. Además se asume que estos polinomios no tienen raíces en común. Si se cumplen estas condiciones el proceso $Y_t \sim ARMA(p, q)$ es estacionario e identificable. Un ARMA(p,q) de media diferente de cero se define por la ecuación

$$\Phi_p(L)(Y_t) = \varphi_0 + \Theta_q(L)(\varepsilon_t)$$

donde $\varphi_0 = \mu(1 - \sum_{j=1}^p \varphi_j)$.

Ejemplo 8.4.1. Suponga un proceso $Y_t \sim ARMA(1, 1)$ de media cero, dado por

$$\Phi_1(L)(Y_t) = \Theta_1(L)(\varepsilon_t) \quad (8.31)$$

donde $\Phi_1(L) = 1 - \varphi L$ y $\Theta_1(L) = 1 + \theta L$. Es decir $Y_t = \varphi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$. Si $|\varphi| < 1$ y $|\theta| < 1$ es estacionario e invertible. Por ejemplo

$$Y_t = 0.9Y_{t-1} + \varepsilon_t - 0.4\varepsilon_{t-1},$$

con $\varepsilon_t \sim RB(0, \sigma^2)$. Si se asume $E(Y_t) = \mu = 2$ entonces el proceso queda

$$\begin{aligned} Y_t &= (1 - \varphi)\mu + \varphi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \\ &= 2(1 - 0.9) + 0.9Y_{t-1} + \varepsilon_t - 0.4\varepsilon_{t-1}. \end{aligned}$$

Ejemplo 8.4.2. Considere el proceso $Y_t \sim ARMA(2, 1)$ dado por

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

que equivale a

$$\begin{aligned} Y_t - 1.5Y_{t-1} + 0.9Y_{t-2} - 2(1 - 1.5 + 0.9) &= \varepsilon_t - 0.4\varepsilon_{t-1} \\ (1 - 1.5L + 0.9L^2)(Y_t - 2) &= (1 - 0.4L)\varepsilon_t, \end{aligned} \quad (8.32)$$

con ecuación característica $1 - 1.5z + 0.9z^2 = 0$, y sus raíces dadas por

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

con módulo $|z| = 1.05 > 1$. Por tanto Y_t es un proceso estacionario en covarianza. También es invertible. En este caso, es válido dividir ambos miembros de la ecuación (8.32) por $1 - 1.5L + 0.9L^2$, para obtener una expresión equivalente

$$Y_t = 2 + \frac{1 - 0.4L}{1 - 1.5L + 0.9L^2} \varepsilon_t.$$

8.4.1. Procesos ARMA y la representación de Wold

Suponga que $Y_t \sim ARMA(p, q)$ y que las raíces de las ecuaciones $\Phi_p(z) = 0$ y $\Theta_q(z) = 0$ están fuera del círculo unitario.

La capacidad de los procesos $ARMA(p, q)$ de representar una amplia variedad de procesos estacionarios en covarianza se basa el siguiente resultado conocido como Teorema de Wold.

Proposición 8.4.1. *Todo proceso estacionario en covarianza X_t , sin componentes determinísticos (ver 7.2.2, pag. 134), satisface una identidad de la forma*

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots, \quad (8.33)$$

con $\psi_0 = 1$ y $\sum_{j=0}^{\infty} |\psi_j| < \infty$, y $\varepsilon_t \sim RB(0, \sigma^2)$. La identidad (8.33) se denomina representación $MA(\infty)$ ó de Wold, de X_t .

La relación entre la representación $MA(\infty)$ y los procesos ARMA(p,q) consiste en una aproximación. Si se define la función de variable $z \in \mathbb{C}$

$$\Psi(z) := \sum_{j=0}^{\infty} \psi_j z^j, |z| < 1$$

se puede aplicar un resultado que se conoce como aproximación de Padé, ver ⁽³⁾.

Aproximación de Padé: siempre es posible encontrar una función racional, es decir, un cociente de polinomios, de la forma $\Theta_q(z)/\Phi_p(z)$, que aproxime bien a $\Psi(z)$, en el sentido de que existe una constante $C > 0$ y enteros $p, q \geq 0$, para los cuales se cumple

$$\forall z \in \mathbb{C}, \quad \left| \Psi(z) - \frac{\Theta_q(z)}{\Phi_p(z)} \right| \leq C|z|^{p+q+1}. \quad (8.34)$$

La aplicación es: si se toma un Y_t , estacionario en covarianza, entonces aplica la representación $MA(\infty)$ y debe existir la función $\Psi(z)$. Pero por la aproximación de Padé se pueden encontrar p, q tales que

$$\Psi(z) \approx \frac{\Theta_q(z)}{\Phi_p(z)} \quad (8.35)$$

Entonces Y_t puede aproximarse con un modelo ARMA(p,q), tal que $\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t$.

Fué la idea base de George E. P. Box y Gwilym Jenkins, implementada en su libro Box et al. [1976], en donde se introdujo la metodología Box-Jenkins, que resultó muy exitosa y se difundió extensamente, ver ⁽⁴⁾

En la práctica no se conoce la función $\Psi(z)$, pero se espera que, al encontrar un posible modelo ARMA(p,q), se cumpla la aproximación (8.34). Entonces el

³https://es.wikipedia.org/wiki/Aproximaci3n_de_Pad3

⁴https://en.wikipedia.org/wiki/Box-Jenkins_method

problema básico inicial es encontrar los órdenes p y q adecuados, que se conoce como el problema de identificación de un modelo ARMA(p, q).

Dados los coeficientes $a = (\varphi_1, \dots, \varphi_p)$ y $b = (\theta_1, \dots, \theta_q)$ se pueden calcular los coeficientes $(\psi_j, j = 0, 1, \dots)$ en (8.33), mediante la función `ARMAtoMA(a, b, m)`.

Ejemplo 8.4.3. Considere el proceso ARMA(2,1) dado por $(1 - L + \frac{1}{4}L^2)Y_t = (1 + L)\varepsilon_t$. La representación $MA(\infty)$ está dada por: (el primero es ψ_1)

```
ARMAtoMA(c(1.0, -0.25), c(1.0), 10)
[1] 2.00000000 1.75000000 1.25000000 0.81250000 0.50000000
[6] 0.29687500 0.17187500 0.09765625 0.05468750 0.03027344
```

El plan de análisis con procesos ARMA consiste en los siguientes pasos, con algunas librerías R para implementarlos.

1. Identificación
2. Estimación
3. Chequeo del ajuste.
4. Pronósticos con el modelo. Utilizar la función de R `predict` para pronósticos.

8.4.2. Simulación de modelos ARMA

La simulación de un proceso ARMA(p, q) se puede hacer mediante la función `arima.sim`. El ejemplo siguiente muestra los polinomios autoregresivo y de media móvil de un ARMA(3,2)

```
#----- Ejemplo la funcion arima.sim
# con los coeficientes a y b.
Phi3 = 1 - 0.1466381*x - 0.1430615*x^2 - 0.3576538*x^3
Theta2 = 1 + 1.420118*x + 0.591716*x^2
a = c(1, - 0.1466, - 0.1430, - 0.3576)
b = c(1, 1.4201, 0.5917)
```

```
y = arima.sim(list(order=c(3,0,2), ar=-a[2:4],
ma=b[2:3]), n = 300, sd=sqrt(0.3))
```

Utilizando la función `filter()`. Requiere dos aplicaciones de esta función. En la primera calcula la componente MA(q), y en la segunda calcular la componente AR(p) con errores MA(q). La instrucción es

```
#----- usar la funcion filter.
et = rnorm(n,0,sqrt(0.3))
x = stats::filter(et,b,"convolution",sides=1,circular=TRUE)
y = stats::filter(x,-a[2:4],"recursive",init=rep(0,3))
```

donde se toman los 3 valores iniciales como la media del proceso, que es cero.

La simulación de un proceso AR(p) de media μ diferente de cero, se realiza añadiendo esta cantidad en la segunda ecuación de la simulación anterior.

8.4.3. La función de autocovarianza de un proceso ARMA

La función de autocovarianza de un proceso $Y_t \sim \text{ARMA}(p,q)$ estacionario de media cero se indica por $R(k) = \text{Cov}(Y_t, Y_{t+k})$, para $k = 0, 1, \dots$. Multiplicando ambos miembros por Y_{t-k} y tomando esperanza $\mathbb{E}(\cdot)$ en la representación $MA(\infty)$ en (8.33) se obtienen las ecuaciones recursivas siguientes, similares a las ecuaciones Yule-Walker para AR(p), en (8.20). Defina $n = \max(p, q + 1)$,

$$R(k) = \varphi_1 R(k-1) + \dots + \varphi_p R(k-p) + \sigma^2 \sum_{j=k}^q \theta_j \psi_{j-k}, \quad (8.36)$$

$$k = 0, 1, \dots, n-1,$$

$$R(k) = \varphi_1 R(k-1) + \dots + \varphi_p R(k-p), \quad k = n, n+1, \dots \quad (8.37)$$

Ejemplo 8.4.4. (tomado de Brockwell and Davis [2002], pag. 93). Considere el proceso ARMA(2,1) dado por $(1 - L + \frac{1}{4}L^2)Y_t = (1 + L)\varepsilon_t$. Entonces $n = \max(p, q + 1) = 2$, por tanto, para $k = 0, 1$ se tiene el sistema lineal

$$R(0) - R(1) + \frac{1}{4}R(2) = \sigma^2(\psi_0\theta_0 + \psi_1\theta_1) = \sigma^2(1 + \psi_1),$$

$$R(1) - R(0) + \frac{1}{4}R(1) = \sigma^2(\theta_1\psi_0) = \sigma^2\psi_0 = \sigma^2. \quad (8.38)$$

Se tiene $\psi_1 = 2$ del Ejemplo anterior. Usando la segunda ecuación (8.37), se obtiene $R(2) = R(1) - \frac{1}{4}R(0)$, luego, reemplazando en el sistema (8.38), y resolviendo, se obtienen $R(0) = 32\sigma^2/3$, $R(1) = 28\sigma^2/3$. Se requiere especificar σ^2 . Utilizando la segunda ecuación (8.37),

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

se puede calcular la autocovarianza recursivamente.

Las funciones `fac` y `facp` teóricas se pueden calcular con la función `ARMAacf`, como se muestra en el código siguiente.

```
#----- fac facp teórica
Phi3 = 1 - 0.1466381*x - 0.1430615*x^2 - 0.3576538*x^3
Theta2 = 1 + 1.420118*x + 0.591716*x^2
a = c(1, - 0.1466, - 0.1430, - 0.3576)
b = c(1, 1.4201, 0.5917)
fac.k = ARMAacf(ar = -a[2:4], ma=b[2:3], lag.max = 20)
facp.k = ARMAacf(ar = -a[2:4], ma=b[2:3], lag.max = 20, pacf=TRUE)
```

8.4.4. Identificación de modelos ARMA

No es inmediato cómo identificar los modelos $ARMA(p, q)$ mediante la `fac` y `facp`. Si $q \geq p \geq 1$ entonces para $k \geq q+1$ se cumple (8.36) y para $1 \leq k \leq p-1$, $\rho(k)$ no tiene un patrón general, luego la `fac` muestral presenta un patrón definido solamente para $k \geq q+1$.

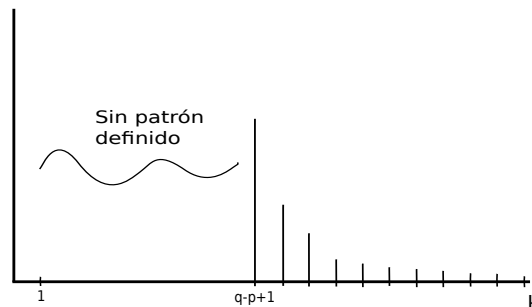


Figura 8.10: Fac Muestral de $ARMA(p, q)$.

Un método directo (“de barrido”) consiste en la identificación de p y q con el AIC. Se basa en comparar varios modelos $ARMA(p, q)$ para valores de, por ejemplo, $p = 0, 1, 2, 3$ y $q = 0, 1, 2, 3$, buscando el que produzca el menor AIC , definido por

$$AIC = -2L(\hat{\underline{\beta}}) + 2k, \quad k = p + q \quad (8.39)$$

donde $\hat{\underline{\beta}}$ está dada más adelante en (8.43) y $L()$ es la función de log-verosimilitud asumiendo distribución Normal para el ruido blanco ε_t . Se identifica un posible modelo parsimonioso con este mínimo AIC . El código siguiente en R muestra una programación básica con base en la función `arima()` para estimación por máxima verosimilitud.

Código R 8.4.1.

```
#---metodo de barrido, para escoger un posible
# modelo ARMA(p,q), con p,q=0,1,2,3
nll = matrix(nrow=4,ncol=4,
dimnames=list(paste("p=", 0:3, sep=""),
paste("q=", 0:3, sep="")))
aic = nll
lb = nll
for (p in (0:3)) {
for (q in (0:3)) {
ARMAlabel = sprintf("ARMA(%d,%d)",p,q)
armamodel = arima(y,order=c(p, 0, q),include.mean = TRUE)
lbtest = Box.test(na.omit(armamodel$resid), lag = 12,type="Ljung")
nll[p+1,q+1] = armamodel$loglik;
aic[p+1,q+1] = armamodel$aic;
lb[p+1,q+1] = lbtest$p.value}}
cat("LogLik:\n")
print(nll)
cat("AIC:\n")
print(aic)
cat("Box-Ljung-Test:\n")
print(lb)
min(aic)
```

En el lenguaje R existen varias funciones para identificar un posible modelo ARMA.

1. La función `auto.arima()` de la librería `forecast`. Para ampliar ver el sitio ⁽⁵⁾. De este sitio se toma esta nota inicial:

La función `auto.arima()` en R usa una variación del algoritmo de Hyndman-Khandakar (Hyndman & Khandakar, 2008), que combina pruebas de raíz unitaria, minimización del AICc y MLE para obtener un modelo ARIMA. Los argumentos de `auto.arima()` proporcionan muchas variaciones en el algoritmo.

2. la función `autoarmafit()` de la librería `timsac`; el programa asume que el modelo $\text{ARMA}(p,q)$ se define con $\Theta_q(z) = 1 - \sum_{j=1}^q \theta_j z^j$, es decir, los coeficientes θ_j cambian de signo en esta librería con respecto a la convención utilizada en R y en estas Notas. Sin embargo, es una función muy inestable y genera en ocasiones desbordamiento de memoria.

3. La función `armasubsets()` con la librería `library(TSA)`.

8.4.5. Estimación de modelos ARMA

La estimación de procesos $Y_t \sim \text{ARMA}(p,q)$ se basa en un supuesto de que el vector $\underline{Y} = (Y_1, \dots, Y_n)'$ se distribuye Normal multivariado con media $\underline{\mu}$, y matriz de covarianzas $\Sigma = [\text{Cov}(Y_i, Y_j)]_{n \times n}$. Como el proceso se asume estacionario, se cumple $\text{Cov}(Y_i, Y_j) = R(j-i)$, donde $R(k)$ es la función de autocovarianza de Y_t . La forma de Σ es la de una matriz tipo Toeplitz: simétrica con las sub-diagonales idénticas:

$$\Sigma = \begin{bmatrix} R(0) & R(1) & R(2) & \cdots & R(n-1) \\ R(1) & R(0) & R(1) & \cdots & R(n-2) \\ R(2) & R(1) & R(0) & \cdots & R(n-3) \\ \vdots & \vdots & & \ddots & \\ R(n-1) & R(n-2) & R(n-3) & \cdots & R(0) \end{bmatrix}. \quad (8.40)$$

$R(k)$ se calcula mediante las ecuaciones Yule-Walker, (8.36) y (8.37). Por tanto, colocando $\underline{\beta} = (\mu, \sigma^2, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)'$, la matriz Σ depende del vector $\underline{\beta}$,

⁵<https://otexts.com/fpp2/arima-r.html>

y se escribe $\Sigma(\underline{\beta})$. Este supuesto permite implementar la estimación por máxima verosimilitud. Se escribe la densidad Normal Multivariada como

$$f(\underline{y}, \underline{\beta}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma(\underline{\beta}))}} \exp \left(-\frac{1}{2} (\underline{y} - \underline{\mu})' \Sigma(\underline{\beta})^{-1} (\underline{y} - \underline{\mu}) \right) \quad (8.41)$$

donde $\underline{\mu} = (\mu, \dots, \mu)' \in \mathbb{R}^n$. La función de log-verosimilitud se define a partir del logaritmo de la densidad (8.41), $\log(f(\underline{y}, \underline{\beta}))$, y está dada por

$$\begin{aligned} L(\underline{\beta}) &:= \sum_{j=1}^n \log f(\underline{y}_j, \underline{\beta}) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det(\Sigma(\underline{\beta})) - \frac{1}{2} (\underline{y} - \underline{\mu})' \Sigma(\underline{\beta})^{-1} (\underline{y} - \underline{\mu}). \end{aligned} \quad (8.42)$$

El estimador ML de máxima verosimilitud se define como

La función de verosimilitud $L(\underline{\beta})$ se maximiza con respecto a los parámetros del modelo (coeficientes AR y MA) en $\underline{\beta}$.

$$\hat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}} (-L(\underline{\beta})). \quad (8.43)$$

Para hacerlo, se requiere un algoritmo de optimización y el filtro de Kalman ⁽⁶⁾. El filtro de Kalman se usa para calcular el valor de la función de verosimilitud para un conjunto dado de valores de parámetros. No es un tema directo, requiere elaboración, ver un ejemplo en ⁽⁷⁾.

O se puede usar el algoritmo de Durbin-Levinson, en la función `DLLoglikelihood` de la librería `ltsa`.

Este valor se pasa como argumento a un algoritmo de optimización, por ejemplo, la función `R optim`.

La librería `ltsa`: Linear time series analysis, de McLeod, Yu y Krougly, tiene funciones relacionadas con la estimación. De la descripción: “Methods of developing linear time series modelling. Methods are given for loglikelihood computation, forecasting and simulation”.

Existen varias funciones del lenguaje R para estimación de modelos ARMA. Por ejemplo,

⁶https://es.wikipedia.org/wiki/Filtro_de_Kalman

⁷<https://cran.r-project.org/web/packages/FKF/vignettes/FKF.html>

1. La función `arima()` estima primero por mínimos cuadrados condicionales y luego por máxima verosimilitud.
2. La función `coefstest()` de la librería `lmtest` calcula los valores p de los coeficientes con base en el modelo estimado con `arima`. Además, con `arima()` se provee el AIC.
3. La función `arma()`, de `tseries`, estima mediante mínimos cuadrados condicionales.
4. La función `FitARMA()`, de `FitARMA` para estimación de ARIMA(p,d,q) con el algoritmo Fast Maximum Likelihood de McLeod y Zhang(2007).
5. La función `armafit()`, de `timsac`, para ajuste de modelos ARMA con el algoritmo Davidon, ⁽⁸⁾. La librería es un producto del IMS, ver ⁽⁹⁾

Ejemplo 8.4.5. *Se simula un ARMA(3,2) tal que las raíces de $\Phi_3(z) = 0$ y $\Theta_2(z) = 0$, tienen módulos fuera del círculo unitario. Las tres raíces de $\Phi_3(z) = 0$ se indican por $z1[1]$, $z1[2]$, $z1[3]$ y los coeficientes del polinomio autorregresivo de grado 3, mónico, $\Phi_3(z)$,*

$$\begin{aligned}\Phi_3(z) &= (1 - z/z1[1])(1 - z/z1[2])(1 - z/z1[3]) \\ &= 1 - \varphi_1 z - \varphi_2 z^2 - \varphi_3 z^3 \\ &= 1 - 0.1466381z - 0.1430615z^2 - 0.3576538z^3\end{aligned}$$

Estos coeficientes se calculan en el código siguiente en el vector `a`. El código incluye la identificación mediante la función `auto.arima`, la estimación con `arima`.

Código R 8.4.2.

```
# simulación, identificación, estimación y
# predicción de un ARMA(3,2)
#-----raices autoregresivas 3
library(polynom)
z1 = complex(3)
z1[1] = -0.8 - 1.3i
```

⁸https://en.wikipedia.org/wiki/Davidon-Fletcher-Powell_formula

⁹<https://www.ism.ac.jp/editsec/csm/index.html>

```

z1[2] = Conj(z1[1])
z1[3] = 1.2
#-----polinomio AR
a = poly.calc(z1)
a = a/a[1]
#-----raices media movil  2
z2 = complex(2)
z2[1] = -1.2 -0.5i
z2[2] = Conj(z2[1])
#-----polinomio MA
b = poly.calc(z2)
(b = b/b[1])
#----- usar la funcion arima.sim
#           con los coeficientes a y b.
n = 300
y = arima.sim(list(order=c(3,0,2), ar=-a[2:4],
ma=b[2:3]), n=n,sd=sqrt(0.3))
#-----graficar
plot.ts(y)
#----- identificar el modelo
y = ts(y,frequency=250)
auto.arima(y)
#----- para su estimacion se usa la instruccion
mod1 = arima(y1, c(3,0,2))

```

Ejemplo 8.4.6. En el artículo Nowicka-Zagrajek and Weron [2002], los autores presentan un modelo de series de tiempo ARMA(p,q) con ruido hiperbólico para la demanda diaria de electricidad en el mercado eléctrico de California. Primero eliminan la componente estacional semanal y anual.

“Los estimadores de ML utilizados aquí se basan en la suposición de ruido blanco Gaussiano. Sin embargo, esto no excluye modelos con ruido no Gaussiano dado que la distribución muestral asintótica de los estimadores es la misma para $Z_t \text{ iid}(0, \sigma^2)$, independientemente de si Z_t es o no Gaussiano, ver Brockwell and Davis [1987, Sección 10.8]”.

El modelo identificado con el menor AIC modificado es

$$X_t = 0.332776X_{t-1} + Z_t - 0.383245Z_{t-1} \\ - 0.12908Z_{t-2} - 0.149307Z_{t-3} - 0.0531862Z_{t-6}$$

donde $Z_t \sim iidHyp(\mu, \alpha, \beta, \delta)$, es una distribución Hiperbólica. La distribución Hiperbólica es una distribución de 4 parámetros, α , β (asimetría), $\delta > 0$ (escala), $\mu \in \mathbb{R}$ (localización), con $0 \leq |\beta| < \alpha$, y $\gamma = \sqrt{\alpha^2 - \beta^2}$,

$$f(x) = \frac{\gamma}{2\alpha\delta K_1(\delta\gamma)} e^{\beta(x-\mu) - \alpha\sqrt{\delta^2 + (x-\mu)^2}},$$

donde $K_\lambda(x)$ es la función de Bessel modificada de segunda clase. En el artículo se encuentra que

$$\hat{\alpha} = 1.671304, \hat{\beta} = -0.098790, \hat{\delta} = 0.298285, \hat{\mu} = 0.076975.$$

Ejemplo 8.4.7. La serie de los precios anuales del cobre desde 1800 hasta 1997 que se puede descargar en R a través de la librería `tsdl`, tomada de Time Series Data Library. El modelo identificado con el menor AIC modificado es

$$X_t = 0.332776X_{t-1} + Z_t - 0.383245Z_{t-1} \\ - 0.12908Z_{t-2} - 0.149307Z_{t-3} - 0.0531862Z_{t-6}$$

donde $Z_t \sim iidHyp(\mu, \alpha, \beta, \delta)$, es una distribución Hiperbólica. Mediante la librería de R `GeneralizedHyperbolic` se pueden estimar los parámetros y además, se puede realizar una prueba de ajuste tipo Cramer-vonMises, con la función `hyperbCvMTest`.

8.5. Modelos ARMA estacionales, SARMA

Los modelos ARMA estacionales estacionarios se introducen con el fin de modelar estacionalidades remanentes en los residuos estructurales. Por tanto, dependen del parámetro s definido como el período de la serie. Pero también son modelos útiles por sí solos.

Hasta el momento no se ha introducido un procedimiento para identificar y estimar el período en una serie dada. Es útil contar con algunos. Se pueden mencionar, por ejemplo

1. La función `periodicity` de la librería `xts`, ver Ryan and Ulrich [2017]. De la página de ayuda: “Estimate the periodicity of a time-series-like object by calculating the median time between observations in days”.

```
et = resid(modelo)
et = ts(et, frequency=1)
periodicity(et)
```

2. Un programa de aplicación por R. Hyndman ⁽¹⁰⁾

Si realmente no tiene idea de cuál es la periodicidad, probablemente el mejor enfoque sea encontrar la frecuencia correspondiente al máximo de la densidad espectral. Sin embargo, el espectro a bajas frecuencias se verá afectado por la tendencia, por lo que primero debe eliminar la tendencia de la serie. La función R siguiente debería hacer el trabajo para la mayoría de las series ... devuelve 1 para los datos que no tienen una periodicidad fuerte y, de lo contrario, la duración del período

```
find.freq <- function(x)
{
  n <- length(x)
  spec <- spec.ar(c(x), plot=FALSE)
  if(max(spec$spec)>10)
# Arbitrary threshold chosen by trial and error.
  {
    period <- round(1/spec$freq[which.max(spec$spec)])
    if(period==Inf) # Find next local maximum
    {
      j <- which(diff(spec$spec)>0)
      if(length(j)>0)
      {
        nextmax <- j[1] + which.max(spec$spec[j[1]:500])
        period <- round(1/spec$freq[nextmax])
      }
    }
  }
}
```

¹⁰<https://stats.stackexchange.com/questions/1207/period-detection-of-a-generic-time-series>

```

        else
            period <- 1
        }
    }
else
    period <- 1
return(period) }

```

Ejemplo 8.5.1. Con los datos de demanda de energía eléctrica domiciliaria a la hora 6:00 pm, diaria, muestran una periodicidad semanal, período igual a 7.

Los modelos ARMA estacionales, S-ARMA modelan la estacionalidad incorporando, a un modelo ARMA, correlaciones entre valores de la serie separadas por un período s .

Definición 8.5.1. Un proceso $(Y_t, t \in \mathbb{Z})$ sigue un modelo ARMA Estacional de período $s > 1$, ó S-ARMA, denotado por $Y_t \sim SARMA(p, q)(p_s, q_s)[s]$, donde p, q, p_s, q_s son enteros no negativos, si para $\epsilon_t \sim RB(0, \sigma^2)$, cumple

$$\Phi_{p_s}(L^s)\Phi_p(L)Y_t = \Theta_{q_s}(L^s)\Theta_q(L)\epsilon_t, \quad (8.44)$$

donde

$$\Phi_p(L) = 1 - \sum_{j=1}^p \varphi_j L^j, \quad (8.45a)$$

$$\Theta_q(L) = 1 + \sum_{j=1}^q \theta_j L^j, \quad (8.45b)$$

$$\Phi_{p_s}(L^s) = 1 - \sum_{j=1}^{p_s} \varphi_{s,j} L^{js}, \quad (8.45c)$$

$$\Theta_{q_s}(L^s) = 1 + \sum_{j=1}^{q_s} \theta_{s,j} L^{js}. \quad (8.45d)$$

En (8.44) p, q son los órdenes AR y MA del modelo ARMA ordinario y p_s, q_s los órdenes AR y MA estacionales. Nótese que en (8.45c) y (8.45d) las potencias de L son de la forma L^s, L^{2s}, \dots

Se asume que las raíces de los cuatro polinomios anteriores están fuera del círculo unitario y además, no hay raíces comunes. De este supuesto se puede escribir el modelo de la forma siguiente

$$Y_t = \frac{\Theta_{q_s}(L^s)}{\Phi_{p_s}(L^s)} \frac{\Theta_q(L)}{\Phi_p(L)} e_t,$$

y se define $X_t = \frac{\Theta_q(L)}{\Phi_p(L)} e_t$, entonces X_t tiene autocorrelaciones tipo ARMA(p,q), y queda

$$Y_t = \frac{\Theta_{q_s}(L^s)}{\Phi_{p_s}(L^s)} X_t,$$

y por Teorema de Wold se tiene la identidad:

$$Y_t = X_t + \psi_1^{(s)} X_{t-s} + \psi_2^{(2s)} X_{t-2s} + \dots, \quad (8.46)$$

Una interpretación de (8.46) es que Y_t tiene efectos adicionales cada s períodos, incorporados a la autocorrelación entre períodos consecutivos, que se asemeja a una dinámica con autocorrelación ordinaria más estacionalidad de período s .

Ejemplo 8.5.2. El modelo $Y_t \sim ARMA(1, 1)(1, 2)[4]$ corresponde a

$$(1 - \varphi_1 L)(1 - \varphi_{4,1} L^4) Y_t = (1 + \theta_1 L)(1 + \theta_{4,1} L^4 + \theta_{4,2} L^8) \epsilon_t, \quad (8.47)$$

Y el modelo $Y_t \sim ARMA(1, 2)(1, 1)[12]$ corresponde a

$$(1 - \varphi_1 L)(1 - \varphi_{12,1} L^{12}) Y_t = (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{12,1} L^{12}) \epsilon_t, \quad (8.48)$$

Nótese que en un S-ARMA, por ejemplo el (8.48), al multiplicar los polinomios autorregresivos y de media móvil se obtiene

$$(1 - \varphi_1 L)(1 - \varphi_{12,1} L^{12}) = 1 - \varphi_1 L - \varphi_{12,1} L^{12} + \varphi_1 \varphi_{12,1} L^{13} \quad (8.49)$$

$$(1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{12,1} L^{12}) = 1 + \theta_1 L + \theta_2 L^2 + \theta_{12,1} L^{12} + \theta_1 \theta_{12,1} L^{13} + \theta_2 \theta_{12,1} L^{14}. \quad (8.50)$$

luego (8.48) es un ARMA en el cual hay un coeficiente autorregresivo $\varphi_1 \varphi_{12}$ en el rezago $1 + 12 = 13$, y también hay un coeficiente de media móvil $\theta_1 \theta_{12}$ en el rezago $1 + 12 = 13$ y otro $\theta_2 \theta_{12,1}$ en el rezago $2 + 12 = 14$. Estos podrían verse en la fac y en la fac parcial, además del rezago 12 (en ocasiones se denominan rezagos satélites).

8.5.1. La simulación de los modelos SARMA

Una manera de realizar la simulación de un modelo S-ARMA es generar los coeficientes auto-regresivos y de media móvil de las partes ordinaria y estacional y colocarlos en vectores, como se muestra en (8.49) y (8.50), de tal forma que se puedan ingresar a la función `arima.sim`.

La librería `CombMSC` tiene la función `sarima.Sim` que implementa la simulación de estos modelos.

Ejemplo 8.5.3. *Un ejemplo con base en el modelo en (8.49) y (8.50)*

```
library(polynom)
# Ejemplo coeficientes s-arma(1,2)(1,1)[12]
#-----
ar = polynomial(c(1,-0.3))
ma = polynomial(c(1,0.072,- 0.629))
ars = polynomial(c(1,-0.12))
mas = polynomial(c(1,- 0.021))
#----genera coeficientes ar y ma completos
period = 12
bs = polynomial(c(rep(0,period),1))
ars = predict(ars,bs)
mas = predict(mas,bs)

fullarpoly <- ar*ars
fullmapoly <- ma*mas

(Mod(polyroot(fullarpoly)))
(Mod(polyroot(fullmapoly)))

mo <- list()
mo$ar <- coef(fullarpoly)
mo$ma <- coef(fullmapoly)

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

```

zplane(filt=rev(mo$ma), a=rev(mo$ar))
#-----
n = 300
y = arima.sim(list(order=c(13,0,14), ar=-mo$ar[-1],
  ma=mo$ma[-1]), n=n, sd=sqrt(0.3))

```

8.5.2. La identificación de los modelos SARMA

Esta observación es útil con referencia al siguiente procedimiento de identificación gráfico, en la librería `library(TSA)`, con la función `armasubsets()`. Calcula un panel para identificación para modelos ARMA y SARMA, que detecta estos rezagos. Sin embargo, tiene limitaciones gráficas para rezagos de órdenes altos. A continuación se muestra una aplicación e interpretación de los resultados de la función.

Para identificar un posible modelo SARMA en una serie con período s , el procedimiento de identificación gráfico con `armasubsets()` permite escoger los órdenes p, q, p_s, q_s con base en un panel que muestra cuáles son los modelos más opcionados con respecto al BIC.

En el panel 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. La instrucción que grafica el panel es:

```

mod2 = lm(yi ~ t + It)
summary(mod2)
r = mod2$residuals
library(TSA)
res=armasubsets(y=r, nar=14, nma=14, y.name='r',
ar.method='ols')
plot(res)

```

En la Figura 8.11 aparecen cuatro paneles de ejemplo, generados con las instrucciones anteriores. Corresponden a un AR en la celda (1,1). En las celdas (1,2) y (2,1)

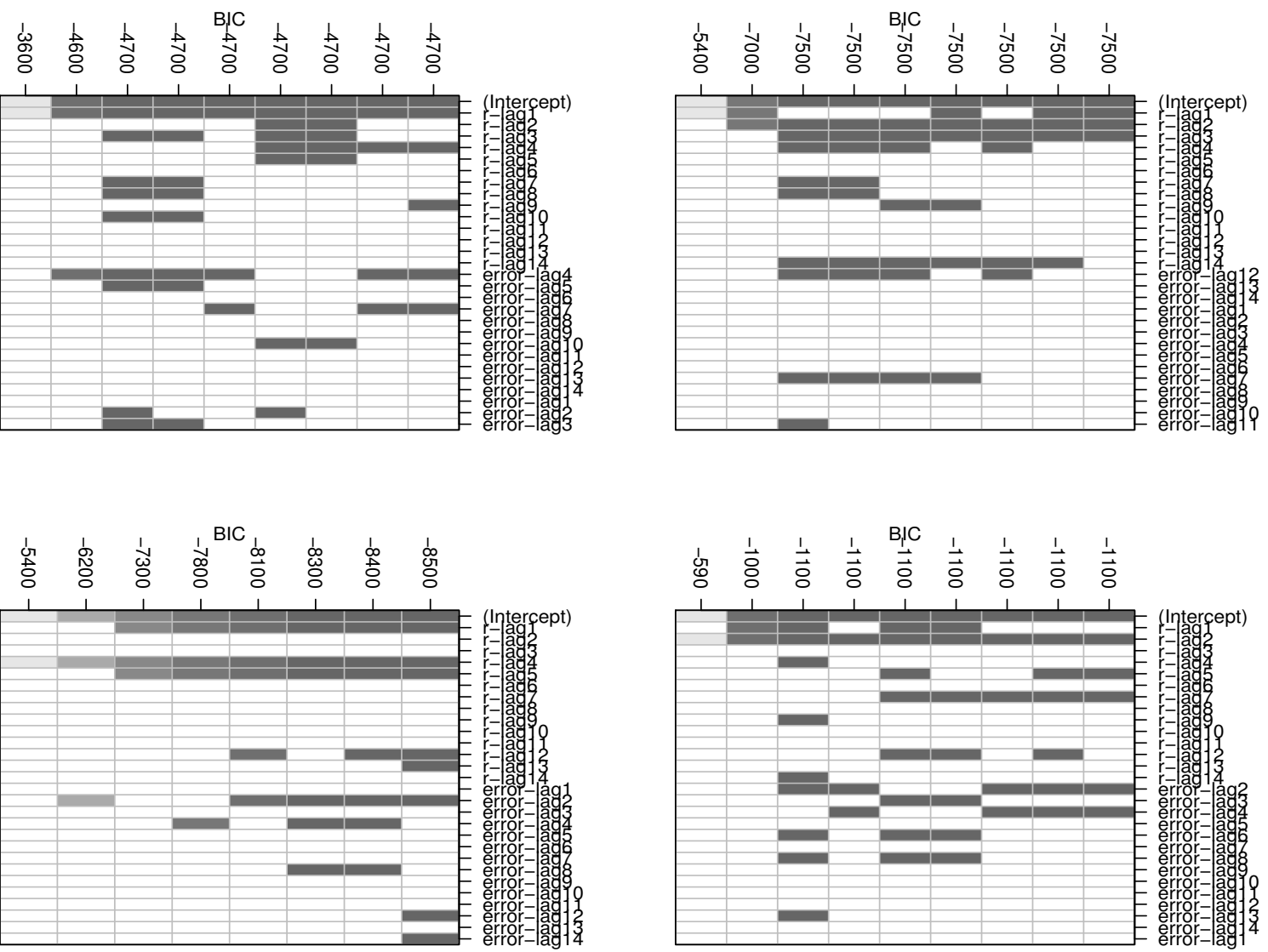


Figura 8.11: Posibles submodelos (1,1)=AR, (1,2),(2,1)=ARMA, (2,2)=SARMA

modelos ARMA. Y en la (2,2) un SARMA.

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.

$$(1 - \varphi_1 L)(1 - \varphi_{12,1} L^{12}) Y_t = (1 + \sum_{j=1}^2 \theta_j L^j)(1 + \theta_{12,1} L^{12}) \epsilon_t, \quad (8.51)$$

8.5.3. La estimación de los modelos SARMA

Los modelos SARMA se estiman con la función `arima()`. Por ejemplo, la estimación del modelo (8.51), $Y_t \sim ARMA(1, 2)(1, 1)[12]$, 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 inco-rrelación de éstos se hace por ejemplo con el examen de la *fac* y la prueba Ljung-Box.

Ejemplo 8.5.4. *La serie de llamadas diarias a un Call Center de cierta entidad bancaria muestra una componente estacional de período 7 días y una tendencia*

lineal, en la Figura 8.12. Se procedió a estimar estas componentes mediante un modelo de componentes con una función lineal y funciones trigonométricas. En la Figura 8.13 se pueden observar las características de los residuos del modelo estructural. Los comandos R siguientes implementan los pasos para identificar, estimar y pronosticar un modelo S-ARMA para estos residuos.

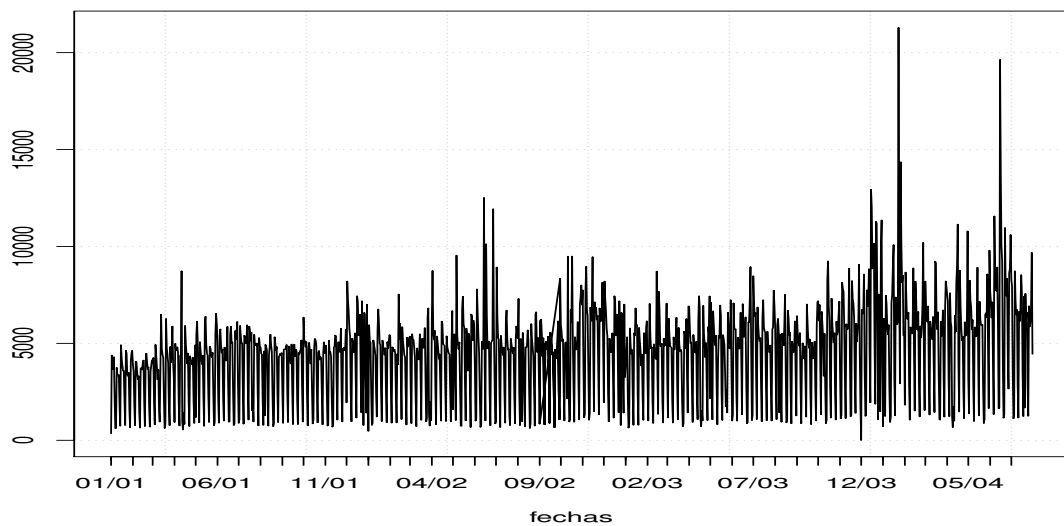


Figura 8.12: Serie de llamadas diarias a un Call Center.

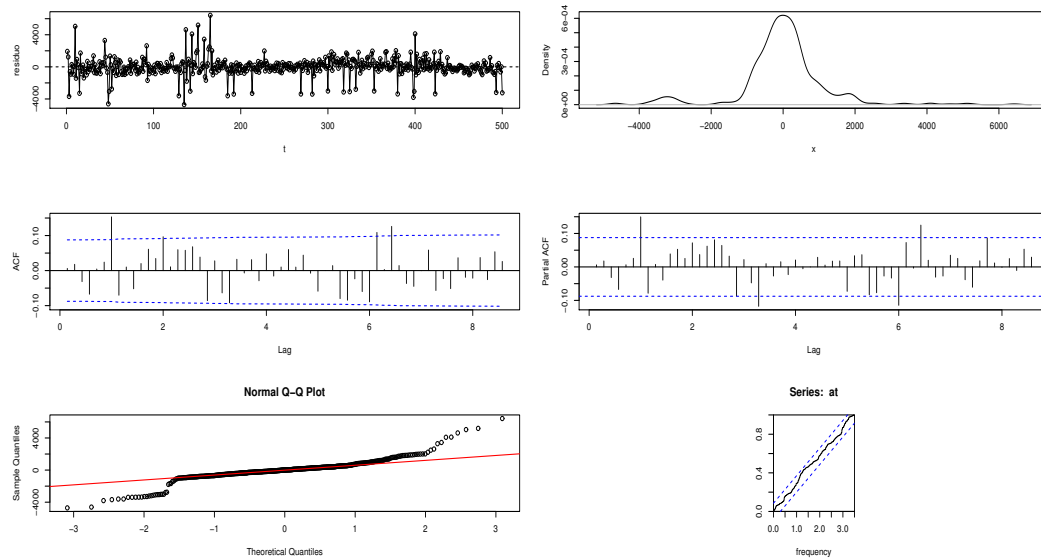


Figura 8.13: Residuos estructurales para la serie de llamadas diarias a un Call Center.
Posible modelo S-ARMA

```
#-----modelo de componentes serie llamadas call center
require(forecast)
y = ts(y,frequency=7)
It = fourier(y,2)
t = seq(1,length(y))
m.l = lm(y ~ t + It)
summary(m.l)
#-----residuos estructurales
yr = resid(m.l)
yr = ts(yr,frequency=7)

par(mfrow=c(3,2))
require(TSA)
plot(t,yr,type='o',ylab='residuo')
abline(h=0,lty=2)
plot(density(yr),xlab='x',main='')
```

```

acf(yr,60,drop.lag.0 = TRUE,main="")
pacf(yr,60,drop.lag.0 = TRUE,main="")
qqnorm(yr)
qqline(yr,col=2)
#-----identificador armasubsets
require(TSA)
res=armasubsets(y=yr,
nar=14,nma=14,y.name='y',ar.method='ols')
par(mfrow=c(1,1))
plot(res)
#-----identificador auto.arima
auto.arima(yr)
ARIMA(3,0,0)(2,0,0)[7] with zero mean

```

Observamos que el modelo identificado es S-ARMA, luego el residuo estructural contiene todavía una componente estacional. La estimación con la función `arima` muestra los siguientes resultados. Nótese que se aumentó en una unidad el orden autorregresivo estacional, para obtener residuos ruido blanco.

```

m.2 = arima(yr,order=c(3,0,0),seasonal=
list(order=c(3,0,0),period=7))
require(lmtest)
coeftest(m.2)
z test of coefficients:

```

ar1	0.227720	0.028435	8.0084	1.162e-15	***
ar2	-0.060973	0.028743	-2.1213	0.03389	*
ar3	0.064545	0.027813	2.3207	0.02030	*
sar1	0.245015	0.027231	8.9976	< 2.2e-16	***
sar2	0.158905	0.028121	5.6508	1.597e-08	***
sar3	0.198965	0.028062	7.0903	1.338e-12	***
intercept	-11.869552	144.912236	-0.0819	0.93472	

Los residuos de este modelo S-ARMA(5,3)(2,0)[7] deberían ser incorrelacionados. La prueba Ljung-Box arroja el siguiente resultado

```
at = resid(m.2)
```

```
Box.test(x = at, lag = 14, type="Ljung-Box")
X-squared = 19.926, df = 14, p-value = 0.1325
```

Por lo que no rechaza la nula de incorrelación. Una conclusión es que la serie de llamadas tiene una estacionalidad de período 7 tanto en la componente estructural como en los residuos.

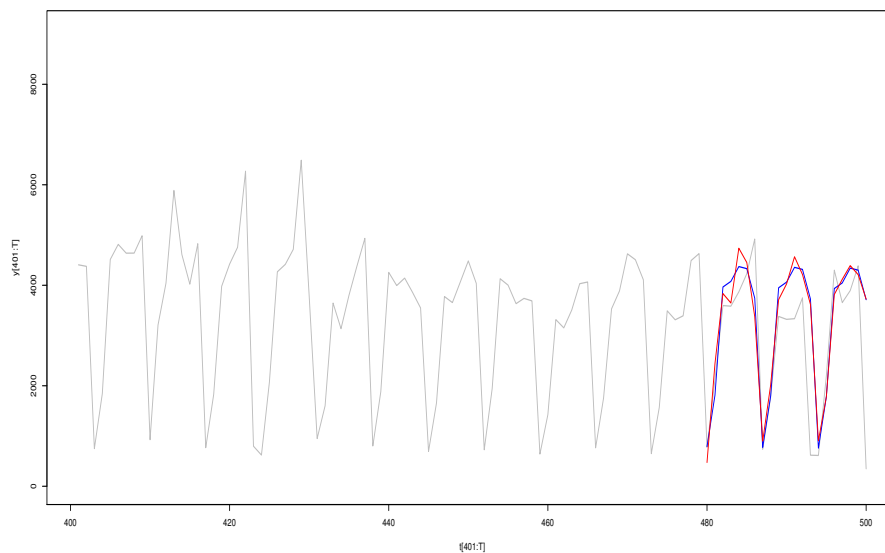


Figura 8.14: Pronósticos estructurales (azul) versus pronósticos estructurales incorporando pronósticos con S-ARMA(5,3)(2,0)[7], para la serie de llamadas diarias a un Call Center.

Los pronósticos estructurales a 21 días y los que se obtienen añadiéndoles los pronósticos del modelo S-ARMA(5,3)(2,0)[7] estimado están en la Figura 8.14

```
#-----calcular pronosticos a 21 dias
T = length(y)
tt = seq((T-21+1),T,1); Itt = fourier(yr,3,21);
py.l = predict(m.l,data.frame(t=tt,It=I(Itt)))
p.arma = predict(m.2,n.ahead=21)$pred
py.tot = py.l+p.arma
```

Ejemplo 8.5.5. Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma

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

Un esquema de análisis es el siguiente. Tomando en cuenta que anteriormente en el Ejemplo 8.2.5, en la pag. 169, se ajustó un MA(6) a los residuos estimados $\hat{\varepsilon}_t$. Ahora interesa examinar si se puede tener también $\hat{\varepsilon}_t \sim S\text{-ARMA}$, y cuál de éstos modelos produce mejores pronósticos.

Un modelo posible es SARMA(1,0,2)(1,0,2)[4], con media cero. En la Tabla 8.2 se pueden ver los resultados de la estimación, utilizando la función `arima()`. El resultado de la prueba Ljung-Box para incorrelación de los residuos es $X\text{-squared} = 23.9558$, $df = 26$, $p\text{-value} = 0.5785$.

En la Tabla 8.3 se pueden comparar las medidas de calidad de pronósticos para los modelos: Estructural (3.19), Estructural con residuos MA(6) y Estructural con residuos SARMA(1,0,2)(1,0,2)[4]. Este último modelo tiene los pronósticos de menor MAPE.

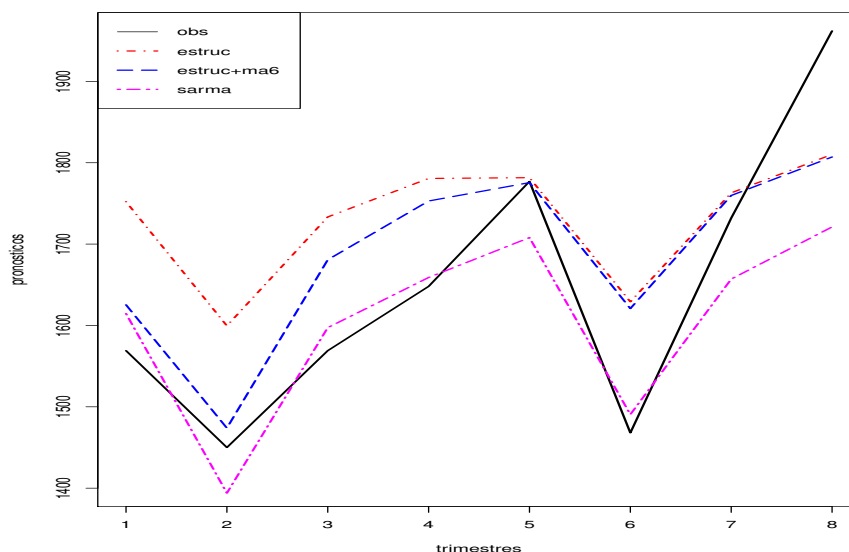


Figura 8.15: Pronósticos con tres modelos para producción de cemento

Tabla 8.2: Coeficientes estimados del modelo SARMA(1,0,2)(3,0,1)[4]

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.6852	0.0991	6.9143	0.0000
ma1	0.0234	0.1144	0.2047	0.8378
ma2	0.4114	0.1097	3.7503	0.0002
sar1	1.1032	0.1511	7.3023	0.0000
sar2	-0.4470	0.1400	-3.1927	0.0014
sar3	0.2762	0.0896	3.0833	0.0020
sma1	-0.5645	0.1461	-3.8638	0.0001
intercept	-44.0386	93.4644	-0.4712	0.6375

Tabla 8.3: Comparación de medidas de calidad de pronósticos

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
estructural	-84.3254	137.1416	122.2615	-5.7245	7.6581	0.1503	0.6695
E+ma6	-40.1263	97.0538	79.2159	-2.8445	4.8387	-0.0503	0.5103
E+sarma	41.6997	96.8725	68.5528	2.1812	3.9090	0.0445	0.5023

CAPÍTULO 9

Raíces Unitarias y Tendencias Estocásticas (Modelos ARIMA)

9.1. Introducción

En este capítulo se presenta un modelo alternativo al de componentes con errores ARMA, conocido como modelo ARIMA-SARIMA ó modelo de Box-Jenkins. El modelo ARIMA es una alternativa al modelo con tendencia, $Y_t = T_t + \varepsilon_t$, y el modelo SARIMA una alternativa al 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. En el capítulo siguiente se introducen los modelos SARIMA y las pruebas de raíz unitaria estacional para detectarlos.

9.2. Modelos ARIMA

Observación. Si en $Y_t = a + bt + \varepsilon_t$, $\varepsilon_t \sim RB(0, \sigma^2)$ se toma $\Delta Y_t = (1 - L)Y_t = Y_t - Y_{t-1}$, se obtiene

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

luego, llamando $W_t = \Delta Y_t$ y $\eta_t = \varepsilon_t - \varepsilon_{t-1}$, se obtiene $W_t = b + \eta_t$. Nótese que η_t es MA(1) y por tanto W_t es un MA(1) con media diferente de cero, estacionario en covarianza. Luego, diferenciar una serie no estacionaria, con tendencia lineal, puede producir una serie estacionaria en covarianza. Esta idea es básica en los modelos ARIMA. Pero requiere algunos conceptos para formalizarla.

Definición 9.2.1 (Raíz Unitaria Autoregresiva). Suponga un modelo $ARMA(p, q)$

$$\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2),$$

tal que una de las p raíces de la ecuación $\Phi_p(z) = 0$ es $z = 1$. Entonces se dice que Y_t tiene una raíz unitaria autorregresiva.

En este caso el polinomio $\Phi_p(z)$ factoriza como $\Phi_p(z) = (1 - z)\Phi_{p-1}(z)$ donde $\Phi_{p-1}(z)$ es un polinomio de grado $p - 1$ y

$$\Phi_{p-1}(L)(1 - L)Y_t = \Theta_q(L)\varepsilon_t,$$

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

$$(1 - L)Y_t = \Delta Y_t \sim ARMA(p - 1, q).$$

Ejemplo 9.2.1. Si $\Phi(L) = (1 - 0.2L)(1 - L)$, $\Theta(L) = 1 - 0.3L$, entonces

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

Entonces $Y_t \sim ARMA(2, 1)$, y no es estacionario en covarianza pues $1 - 1.2z + 0.2z^2 = 0$ tiene raíces $z_1 = 1$, $z_2 = 5$ y sólo una cumple $|z| > 1$. Pero $W_t = \Delta Y_t = Y_t - Y_{t-1}$ cumple $(1 - 0.2L)W_t = (1 - 0.3)\varepsilon_t \sim ARMA(1, 1)$ es estacionario en covarianza e invertible.

Definición 9.2.2 (Proceso ARIMA(p,1,q)). Una serie de tiempo $(Y_t, t \in \mathbb{Z})$ sigue un modelo $ARIMA(p, 1, q)$, con tendencia, si $(1-L)Y_t$ sigue un proceso $ARMA(p, q)$ estacionario en covarianza, con media diferente de cero. Es decir, si

$$\Phi_p(L)(1-L)Y_t = \mu + \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (9.1)$$

Nótese que si $Y_t \sim ARIMA(p, 1, q)$ entonces $\Delta Y_t = W_t$, con W_t estacionario. Si Y_0 se asume dada, entonces $Y_t = Y_0 + \sum_{j=1}^t \Delta Y_j$. Por lo que la operación de suma acumulada puede considerarse el inverso de la operación de diferencia.

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].

9.3. Modelos de Marcha Aleatoria

Las Marchas Aleatorias (Random Walks) son modelos estocásticos muy estudiados. Una lista de al menos 15 áreas de aplicaciones de las marchas aleatorias, incluyendo el arte (escultura), aparece en Wikipedia (¹).

La marcha aleatoria sin tendencia se define como

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (9.2)$$

Según como se tome el conjunto de valores de Y_t , E , y se defina la distribución de ε_t , se obtienen muchos tipos de marchas aleatorias unidimensionales. Las marchas aleatorias no acotadas tienen $E = \mathbb{R}, \mathbb{Z}$. Las acotadas tienen $E = \mathbb{R}^+, \mathbb{Z}^+$, $E = \{d_1, \dots, d_k\}$, $E = [a, b]$, etc.

También se definen marchas aleatorias multi-dimensionales, $E = \mathbb{R}^d$, $E = \mathbb{Z}^d$, $d = 2, 3$. Un resultado de teoría establece que para $d \geq 3$ una marcha aleatoria no puede tener distribución invariante.

¹https://en.wikipedia.org/wiki/Random_walk

Nótese que la marcha aleatoria en (9.2) es un proceso $Y_t \sim AR(1)$, de media cero, con coeficiente $\varphi = 1$. Las marchas aleatorias son ejemplos básicos de modelos ARIMA por lo que se denotará $Y_t \sim ARIMA(0, 1, 0)$.

La marcha aleatoria sin tendencia cumple, para $t \geq 0$, y Y_0 dado, no aleatorio,

$$Y_t = Y_0 + \sum_{i=0}^t \varepsilon_i, \quad (9.3a)$$

$$E(Y_t) = Y_0, \quad (9.3b)$$

$$Var(Y_t) = t\sigma^2. \quad (9.3c)$$

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

$$Y_t = \mu + Y_{t-1} + \varepsilon_t, \quad \mu \in \mathbb{R}. \quad (9.4)$$

La marcha aleatoria con tendencia cumple

$$Y_t = t\mu + Y_0 + \sum_{i=0}^t \varepsilon_i, \quad (9.5a)$$

$$E(Y_t) = t\mu + Y_0, \quad (9.5b)$$

$$Var(Y_t) = t\sigma^2. \quad (9.5c)$$

Un ejemplo de trayectorias de marchas aleatorias con y sin tendencia se muestran a continuación en la Figura 9.1.

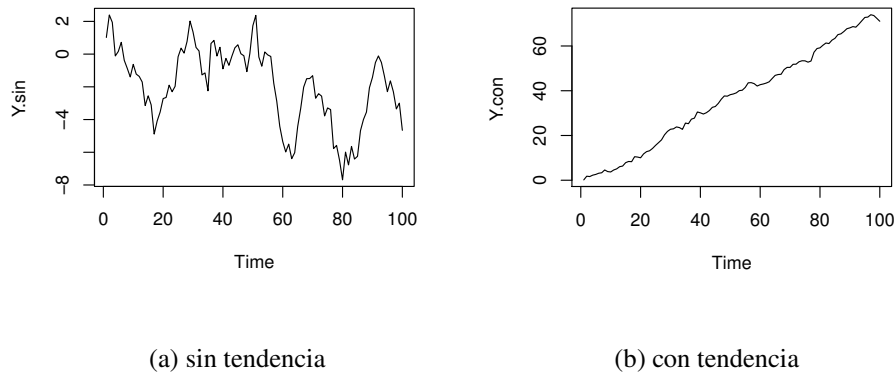


Figura 9.1: Trayectorias de Marchas Aleatorias

Obsérvese que la marcha aleatoria con tendencia Y_t , tiene una característica similar a una serie con tendencia lineal de la forma $X_t = a + bt + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, 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

$$\hat{Y}_T(j) = \mu + \varphi^j Y_T, \quad j = 1, 2, \dots, h.$$

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, a partir de $Y_t = Y_0 + \mu t + \sum_{j=1}^t \varepsilon_j$, para $k = 1, 2, \dots$,

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

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

$$Corr(Y_t, Y_{t+k}) = \frac{\sigma^2 t}{\sqrt{\sigma^2 t \sigma^2 (t+k)}} = \frac{1}{\sqrt{1 + k/t}}. \quad (9.6)$$

Para t grande tal que k/t sea muy pequeño, se puede aproximar a cero y obtener que las autocorrelaciones en marchas aleatorias con tendencia son casi uno, $\rho(k) \approx 1$. Como la tendencia no determina el valor de la autocorrelación, la conclusión también es válida en el caso de marchas aleatorias sin tendencia. Las marchas aleatorias son series con fuertes autocorrelaciones.

Ejemplo 9.3.1. Si $Y_t \sim ARMA(1, 1)$ con $E(Y_t) = 0$ entonces

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

y se puede comprobar que las autocorrelaciones de Y_t satisfacen:

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

$$\rho(k) = \rho(k-1), \quad k \geq 2.$$

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

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

Por lo tanto $\rho(k) \equiv 1, \forall k \geq 1$.

Los procesos $ARIMA(p, 1, q)$ se comportan como marchas aleatorias en ciertos aspectos. Las marchas aleatorias son muy autocorrelacionadas. Los procesos ARIMA también muestran una fuerte autocorrelación. La fac muestral es muy persistente y cercana a uno.

Definición 9.3.1. Si la serie Y_t no estacionaria en covarianza es tal que ΔY_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 similar.

Luego un $ARIMA(p, 1, q)$ es un proceso $I(1)$ y un $ARMA(p, q)$ es un $I(0)$.

Definición 9.3.2. Una serie $(Y_t, t \in \mathbb{Z})$ se dice que tiene tendencia estocástica si es integrada de orden $d = 1$.

La marcha aleatoria con tendencia (9.4) 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. De igual manera, un proceso $ARIMA(p, 1, q)$ es un proceso con tendencia estocástica.

Una serie con tendencia estocástica presenta períodos de duración aleatoria durante los cuales crece linealmente.

La implementación de los modelos ARIMA asume series con tendencia estocástica. La metodología consiste en diferenciar la series $d = 1, 2$ veces, hasta obtener otra serie estacionaria en covarianza, identificable por un modelo ARMA, y estimar los parámetros de este modelo.

9.4. Ejemplos de modelos ARIMA

A continuación se muestran algunos ejemplos de estimación de modelos ARIMA con la función `arima()`. En la sección siguiente se trata las pruebas de hipótesis para detectar la presencia de raíces unitarias.

Ejemplo 9.4.1. 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 $AR(2)$

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \eta_t, \quad \eta_t \sim RB(0, \sigma^2).$$

2. Modelo ARIMA

$$Y_t \sim ARIMA(1, 1, 0) \text{ con tendencia}$$

$$(1 - \varphi L)(1 - L)Y_t = \delta + \varepsilon_t.$$

Entre 1932 y 1933 hubo una grave recesión en EUA y los pronósticos con base en 1933 se hacen en una posición por debajo de la tendencia

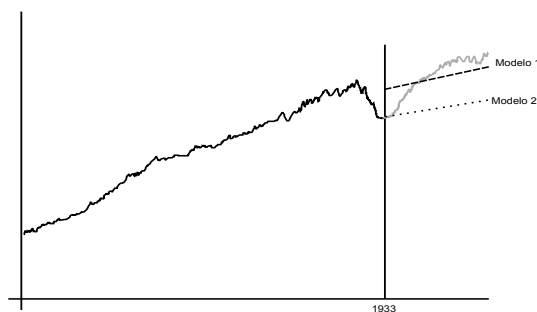


Figura 9.2: Pronósticos de la Serie PNB-USA.

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

Ejemplo 9.4.2. *Análisis de la Tasa de cambio: USD/YEN, el precio de 1 US Dólar en Yen, ver Diebold [1999, pág. 221]. La serie es mensual desde 1973-01 hasta 1996-07. Por ejemplo, $Y_t = 107$ significa que en el mes t se pagaban 107 Yen por 1 USD. El análisis siguiente es con el logaritmo de esta tasa. En lo que sigue y denota el logaritmo de la tasa y dy la serie diferenciada $dy = \text{diff}(y, 1, 1)$.*

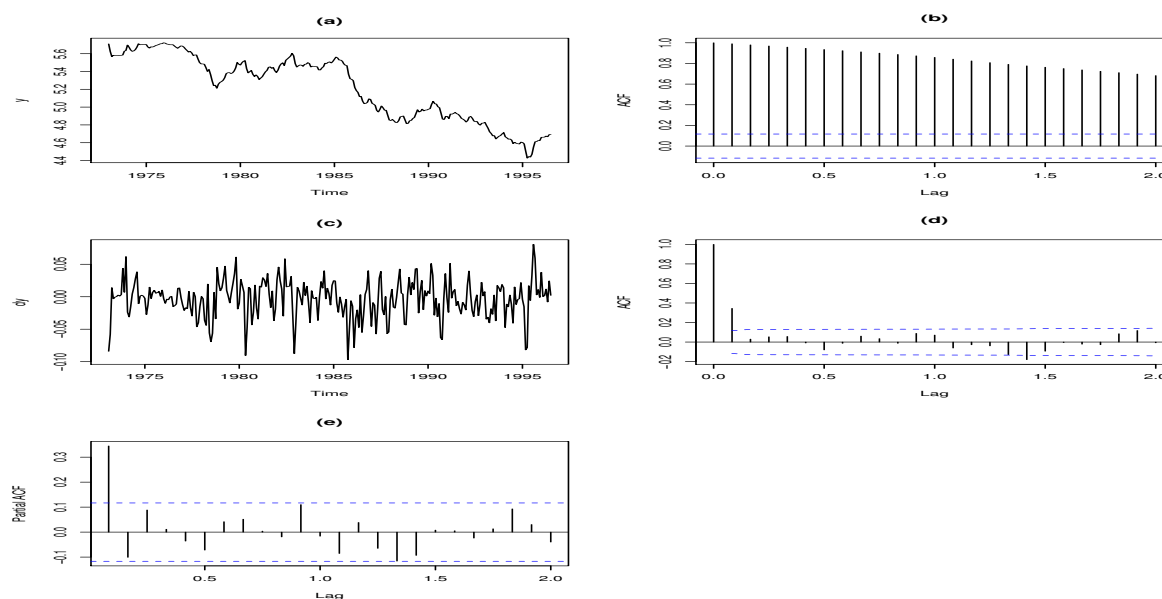


Figura 9.3: (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada

La Figura 9.3 (b) sugiere un proceso integrado $I(1)$ debido a que la fac es muy persistente; y las Figuras (d) y (e) sugieren un proceso estacionario para la serie diferenciada, por lo que un modelo ARMA puede ser apropiado.

Utilizando la función $z = \text{autoarmafit}(dy)$, de la librería `timsac`, con la serie diferenciada se obtiene un modelo ARMA(3,2).

Utilizando la función `arima(y, order=c(3,1,2))` se obtienen los resultados de la Tabla 9.1. Finalmente, utilizando la prueba Ljung-Box, se comprueba que los residuos de este modelo son Ruido Blanco. El resultado de la prueba es

X-squared = X-squared = 15.7889, df = 26, p-value = 0.9411,

Tabla 9.1: Parámetros del modelo ARIMA(3,1,2)

	parametros	sd.dev	est t
ar1	0.18	0.06	2.91
ar2	-0.93	0.01	-77.71
ar3	0.37	0.06	6.17
ma1	0.20	0.02	11.32
ma2	0.99	0.03	32.43

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 no es lo mismo estimar el modelo ARIMA(3,1,2) con la variable $\log(Y_t)$, que estimar el modelo ARMA(3,2) con la variable $\Delta \log(Y_t)$. Los pronósticos a 19 períodos se muestran en la Figura 9.4, comparados con los pronósticos de un modelo con tendencia lineal y error estructural AR(2).

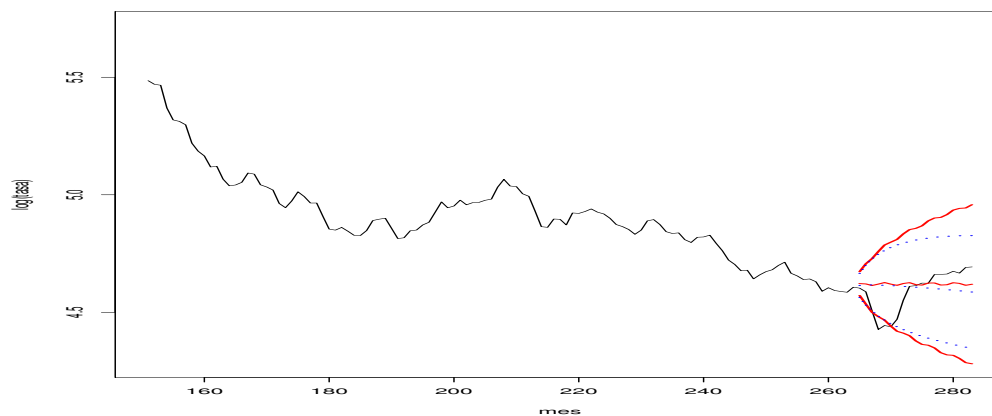
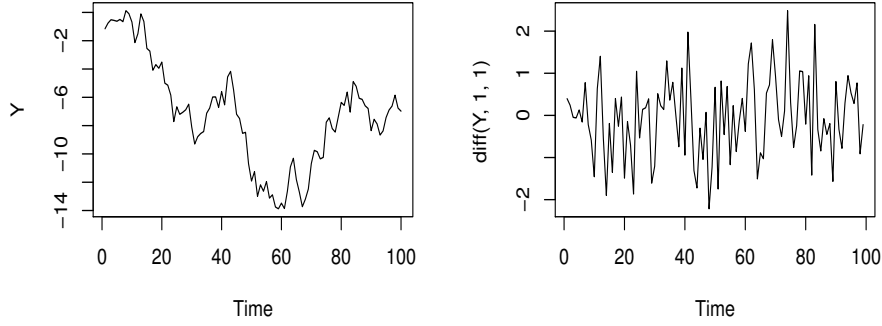


Figura 9.4: Pronósticos usd/yen con ARIMA(3,1,2) (continua) y Tendencia Lineal+AR(2) (punteada)

Ejemplo 9.4.3. Considerando Y_t la serie del logaritmo de la tasa de cambio US-D/Libra, es decir, el log del precio de 1 libra esterlina en usd, mensual, en el período 01/80-12/88, con 478 observaciones.

La trayectoria parece una marcha aleatoria sin tendencia, como se observa en la

Figura 9.6.



(a) Serie USD/Libra

(b) Primera Diferencia de la Serie USD/Libra

Figura 9.5: Serie USD por Libra

La primera diferencia parece estacionaria, casi ruido blanco. La fac de la primera diferencia confirma que es ruido blanco, obteniendo un modelo de marcha aleatoria con tendencia, dado por

$$(1 - L)Y_t = \mu + \varepsilon_t, \quad \mu = -0.0009.$$

Este modelo se denomina Modelo de Black-Scholes ó Modelo Log-Normal para el precio de un activo.

Ejemplo 9.4.4. Considerando Y_t la serie del empleo en Canadá de periodicidad trimestral en el período 01/1962-04/1993, desestacionalizada.

Se propuso un modelo $AR(2)$. Pero usando la función `auto.arima`, se obtiene $Y_t \sim ARIMA(1, 1, 0)$ sin intercepto, es decir,

$$(1 - \varphi L)(1 - L)Y_t = \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (9.7)$$

con $\hat{\varphi} = 0.4598$, $s.e. = 0.0758$, $\hat{\sigma} = 2.068$, $BIC = 491.22$, $AIC = 485.41$, $MAPE = 1.0453$.

Si se ajusta un modelo $AR(2)$ a la serie, es decir,

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad (9.8)$$

se obtiene $\hat{\varphi}_1 = 1.4505$, $\hat{\varphi}_2 = -0.4763$, $\hat{\sigma} = 2.022$, $\hat{\mu} = 97.498$, $AIC = 493.57$, $BIC = 505.22$, $MAPE = 1.0648$.

```
# programa en R
library(forecast) #
# 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))
```

La conclusión es que sería preferible utilizar el modelo (9.7) por tener un mejor ajuste. Pero puede utilizarse otro recurso.

La prueba Diebold-Mariano permite discriminar entre dos modelos de acuerdo a su capacidad predictiva dentro de la muestra.

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.

En el Ejemplo anterior, 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
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
```

Definición 9.4.1. *La hipótesis de mercados eficientes asume que los retornos del precio de las acciones son ARIMA(0,1,0).*

En ⁽²⁾: “En finanzas, la hipótesis del mercado eficiente (EFM) afirma que los precios de los activos que se negocian en un mercado financiero eficiente reflejan toda la información existente y se ajustan total y rápidamente a los nuevos datos que puedan surgir...Es por esta razón, que los precios de las acciones siguen una trayectoria incierta, es decir, el mejor pronóstico que existe para el precio de mañana es el precio de hoy. Esto es lo que se denomina Teoría de la marcha aleatoria...

Crash, burbujas y depresiones son compatibles con la hipótesis siempre que esta conducta no sea predecible...

Hay economistas, matemáticos y especuladores que no creen en la hipótesis del mercado eficiente. Piensan que hay razones poderosas para que existan ineficiencias: la lenta difusión de la información, el poder desigual de los distintos participantes en los mercados y la existencia de inversores profesionales que, al menos aparentemente, logran altos rendimientos...

Joseph Stiglitz, George Akerlof y Michael Spence ganaron el Premio Nobel de Economía por sus investigaciones sobre los mercados con información asimétrica y las inconsistencias, ineficiencias y fallos de los mercados financieros, derivados de la existencia de información asimétricas.”

En ⁽³⁾ : “Quienes está a favor de la EFM argumentan que no tiene sentido buscar acciones infravaloradas o tratar de predecir tendencias en el mercado a través de análisis fundamentales o técnicos”.

En un período de tiempo el precio puede aumentar por efecto de la demanda. Pero llega un momento en el cual la demanda se satura y los precios empiezan a oscilar,

²https://es.wikipedia.org/wiki/Hipótesis_del_mercado_eficiente

³<https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp>

o a decrecer. Hay períodos de duración aleatoria en los que aparecen tendencias lineales, que se asumen tendencias estocásticas, y por tanto, se pueden describir por un modelo $ARIMA(0,1,0)$.

Sin embargo, se dan casos en los que se encuentra un $ARIMA(0,1,1)$, denominado IMA(1), ver, por ejemplo, Box et al. [1976], pag. 103.

9.5. Pruebas de Hipótesis para Raíces Unitarias

La fac y facp muestrales en presencia de Raíces Unitarias

En un proceso Y_t con raíz unitaria, por ejemplo, $ARIMA(p, 1, q)$, la fac muestral converge a cero con mucha lentitud, la facp muestral presenta un valor cercano a 1 en el rezago $k = 1$ y los demás valores son aproximadamente cero. Por ejemplo la Figura 9.6 siguiente se muestra la fac y la facp típicas de un proceso con raíz unitaria.

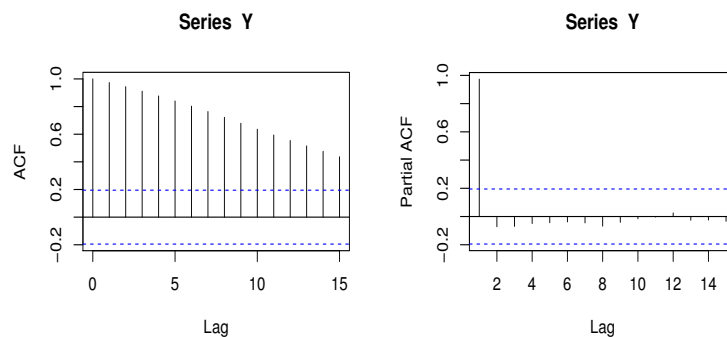


Figura 9.6: fac y facp muestrales de proceso con raíz unitaria

Necesidad de pruebas para detectar Raíces Unitarias

Hay que tener precaución con relación a tomar la decisión de diferenciar la serie para buscar un ARMA cuando se observan las características anteriores. Y es que es necesario aplicar una prueba de hipótesis para determinar si existe una raíz unitaria en lugar de una tendencia lineal determinística ó, incluso, un $AR(p)$ muy persistente. Las siguientes observaciones son pertinentes.

1. Diebold [1999, pág. 221]: “Si no hay raíz unitaria es conveniente utilizar modelos de niveles (componentes determinísticas) y es 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 un componente MA no invertible el cual causa problemas serios de estimación.”

Pruebas de raíz unitaria en R

Existen en la literatura de series de tiempo varias pruebas para detectar raíz unitaria en una serie de tiempo. Y varias librerías en R que las implementan.

1. Prueba Dickey-Fuller aumentada. En las librerías:
 - `forecast`, función: `ndiffs`
 - `fUnitRoots`, función: `adfTest`
 - `tseries`, función: `adf.test`
 - `aTSA`, función: `adf.test`
 - `urca`, función: `ur.df`
2. Prueba KPSS, en la librería `aTSA`, función: `kpss.test`
3. Prueba PP, en la librería `aTSA`, función: `pp.test`

Estimadores MCO en caso de raíz unitaria en modelos AR(1)

La discusión en esta sección es para un $Y_t \sim ARIMA(p, 1, q)$. Suponga un proceso $AR(1)$, $Y_t = \varphi Y_{t-1} + \varepsilon_t$, en el cual $\varphi = 1$ pero φ se estima usando MCO. Este estimador $\hat{\varphi}$, tiene dos propiedades, ver Diebold [1999]

1. Superconsistencia

a) En el caso $\varphi = 1$ el estimador de MCO $\hat{\varphi}_T$ tiene la propiedad de que

$$T(\hat{\varphi}_T - 1) \xrightarrow{d} Z, \quad T \rightarrow \infty, \quad (9.9)$$

donde Z es una variable aleatoria no degenerada, y el símbolo \xrightarrow{d} denota la convergencia en distribución.

b) En el caso $|\varphi| < 1$, el estimador de MCO, $\hat{\varphi}_T$ cumple

$$\sqrt{T}(\hat{\varphi}_T - \varphi) \xrightarrow{d} Z \quad T \rightarrow \infty, \quad (9.10)$$

donde Z es variable aleatoria no degenerada.

como $\sqrt{T} < T$, se dice que la convergencia en el caso (9.9) es más rápida que en el caso (9.10), y a esto se denomina superconsistencia, es decir, el estimador de MCO, $\hat{\varphi}_T$ de una raíz unitaria es superconsistente.

2. Sesgo de $\hat{\varphi}$

Si $\hat{\varphi}_T$ es el estimador de MCO de φ 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 en la serie. Aunque $\hat{\varphi}_T$ converge a φ , cuando $T \rightarrow \infty$, el sesgo puede ser apreciable en muestras no muy grandes.

9.6. Prueba Dickey-Fuller

La prueba Dickey-Fuller se basa en asumir que la serie se puede aproximar por un proceso AR(1) con tres variantes: media cero, media diferente de cero y tendencia lineal. Inicialmente se asume que Y_t sigue un modelo AR(1) de media cero y se procede a transformar este modelo de la siguiente manera.

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varepsilon_t, \\ \Leftrightarrow Y_t - Y_{t-1} &= (\varphi_1 - 1)Y_{t-1} + \varepsilon_t, \\ \Leftrightarrow \Delta Y_t &= \rho Y_{t-1} + \varepsilon_t. \end{aligned}$$

con $\rho = \varphi_1 - 1$. La existencia de una raíz unitaria $\varphi_1 = 1$ equivale a $\rho = 0$.

1. Prueba DF para el caso 1: $Y_t \sim AR(1)$ con media cero. Entonces

$$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t. \quad (9.11)$$

La hipótesis nula es $H_0 : \rho = 0$ versus la alterna $H_a : \rho < 0$. El estadístico de la prueba se denota por τ y su distribución bajo H_0 permite calcular los valores críticos, por ejemplo, $\tau_{0.05}$ para un nivel de significación de 5 %.

El criterio de rechazo es $\hat{\tau} < \tau_{0.05}$, con $\hat{\tau}$ el valor calculado del estadístico. El valor p es $\mathbb{P}(\tau < \hat{\tau} | H_0)$. Se rechaza la nula a un nivel de significación dado si el valor p es menor que tal nivel. En R la prueba DF se implementa, para este caso, con la librería `urca`, por medio de la función `ur.df(y, type="none", lags=0)`.

2. Prueba DF para el caso 2: $Y_t \sim AR(1)$ con media diferente de cero. Entonces se modifica (9.11) así:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \quad (9.12)$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0, H_a : \rho < 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y, type="drift", lags=0)`.

3. Prueba DF para el caso 3: suponiendo que $Y_t \sim AR(1)$ con media diferente de cero y con tendencia lineal, entonces

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t, \quad (9.13)$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0, H_a : \rho < 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y, type="trend", lags=0)`.

9.7. Valores críticos de la prueba Dickey-Fuller

Con base en la Tabla siguiente, tomada de Fuller [2009]

Como $\hat{\tau}$ es mayor que cualquiera de los valores críticos, no rechaza la nula de raíz unitaria. Note que el valor p es en este caso 0.2924.

Recordamos la definición de error en las pruebas de hipótesis. El error tipo I se denomina “falso positivo”. Es el evento de rechazar H_0 siendo cierta. La probabilidad del error tipo I es α , el nivel de significación.

Tabla 9.2: Valores críticos de la prueba Dickey-Fuller

Sample size	Without trend		With trend	
	1 %	5 %	1 %	5 %
T = 25	-3.75	-3.00	-4.38	-3.60
T = 50	-3.58	-2.93	-4.15	-3.50
T = 100	-3.51	-2.89	-4.04	-3.45
T = 250	-3.46	-2.88	-3.99	-3.43
T = 500	-3.44	-2.87	-3.98	-3.42
T = ∞	-3.43	-2.86	-3.96	-3.41

Tabla 9.3: Ejemplo de resultado de la prueba Dickey-Fuller, con $n = 50$ observaciones

Estadístico Prueba	1 %	5 %	10 %
$\hat{\tau} = -1.987$	-3.580	-2.930	-2.600

		H_0 real	
		V	F
H_0 según prueba	V	OK	error II
	F	error I	OK

Tabla 9.4: Tabla de decisiones en una prueba de hipótesis

Recordamos la definición de potencia de una prueba, $1 - \beta$ donde $\beta = \text{Prob}(\text{error II})$. El error tipo II se denomina “falso negativo”. Es el evento de no rechazar H_0 siendo falsa. Luego la potencia es la probabilidad de rechazar H_0 siendo falsa.

Una prueba con baja potencia significa que es poco capaz de detectar la hipótesis alterna cuando esta es cierta.

Fijado el nivel α y el tamaño de la muestra, la potencia puede ser baja ó alta.

Las pruebas DF se sabe que tienen baja potencia. Por tanto, no rechazar la hipótesis nula de raíz unitaria no necesariamente significa que se deba asumir que existe. Como señala Diebold [1999, pág 220]: “Las pruebas de raíz unitaria tienen problemas de potencia y tamaño de muestra. De potencia porque las hipótesis alternas son muy cercanas a la hipótesis nula”. La presencia de cambios estructurales en la serie se sabe que disminuye la potencia de las pruebas DF, pero no se desarrollará este tema.

9.8. Prueba Dickey-Fuller Aumentada

La prueba aumentada de Dickey-Fuller es preferible a la prueba básica, pero 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 asumir que $Y_t \sim AR(p)$, y determinar cuál de los tres casos considerados anteriormente, con $p = 1$, determina una mejor aproximación a la serie original. Nótese que los casos a examinar dependen del orden autorregresivo p , por lo que la búsqueda para, por ejemplo, $p = 1, 2, 3$, requiere examinar $3 \times 3 = 9$ casos. La decisión se toma con base en el menor AIC.

Caso 1. Suponiendo que $Y_t \sim AR(p)$ con media cero, caso indicado por “none”, entonces se realizan las transformaciones siguientes.

$$Y_t = \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t = \varphi_1 Y_{t-1} + \sum_{j=2}^p \varphi_j Y_{t-j} + \varepsilon_t. \quad (9.14)$$

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 (9.14) se transforma en

$$\Delta Y_t = \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (9.15)$$

Si hay una raíz unitaria se cumple

$$1 - \sum_{j=1}^p \varphi_j = 1 - \rho_1 = 0,$$

por tanto, $\rho = 0$. En este caso el modelo (9.15) equivale a

$$\Delta Y_t = \sum_{j=1}^{p-1} \rho_{j+1} \Delta Y_{t-j} + \varepsilon_t, \quad (9.16)$$

es decir un $AR(p-1)$ en la variable $Z_t = \Delta Y_t$. La hipótesis nula es $H_0 : \rho = 0$ y la alterna $H_a : \rho < 0$. No rechazar la nula significa que se detecta la raíz unitaria. El estadístico DF, τ tiene la misma distribución asintótica que el estadístico DF en el caso $AR(1)$ dado en (9.11). La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

“Así, los resultados del proceso $AR(1)$ se generalizan asintóticamente en forma directa a procesos de orden superior”. Diebold [1999, pág. 128]

Caso 2 Suponiendo $Y_t \sim AR(p)$ con media diferente de cero, caso indicado por “drift”. Con la misma notación del Caso 1 anterior, se realizan las transformaciones siguientes.

$$\begin{aligned} Y_t - \mu &= \sum_{j=1}^p \varphi_j (Y_{t-j} - \mu) + \varepsilon_t \\ Y_t &= \alpha + \varphi_1 Y_{t-1} + \sum_{j=2}^p \rho_j (Y_{t-j+1} - Y_{t-j}) + \varepsilon_t \\ \Delta Y_t &= \alpha + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \end{aligned} \quad (9.17)$$

donde $\alpha = \mu(1 - \sum_{j=1}^p \varphi_j)$, con los ρ_j definidos como en el Caso 1 anterior.

La hipótesis nula es $H_0 : \rho = 0$ (existe raíz unitaria) y la alterna $H_a : \rho < 0$. Bajo $H_0 : \rho = 0$ el término α se anula. La distribución asintótica del estadístico DF es igual a la del caso $AR(1)$ con media, (9.12). La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

Caso 3 Suponiendo $Y_t \sim AR(p)$ con tendencia lineal, caso indicado por “trend”. En este caso se define

$$Y_t = a + bt + \sum_{j=1}^p \varphi_j (Y_{t-j} - a - b(t-j)) + \varepsilon_t \quad (9.18)$$

que se puede reordenar de la siguiente forma

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (9.19)$$

donde

$$\begin{aligned} k_1 &= a \left(1 - \sum_{i=1}^p \varphi_i \right) + b \sum_{i=1}^p i \varphi_i, \\ k_2 &= b \left(1 - \sum_{i=1}^p \varphi_i \right), \\ \rho_1 &= \sum_{i=1}^p \varphi_i, \quad \rho = \rho_1 - 1. \end{aligned}$$

La hipótesis nula es $H_0 : \rho = 0$ (existe raíz unitaria) y la alterna $H_a : \rho < 0$. Bajo $H_0 : \rho = 0$ se tiene que $k_2 = 0$, $k_1 = b \sum_{i=1}^p i \varphi_i$ y el estadístico DF tiene la misma distribución asintótica del estadístico en el caso AR(1) con tendencia lineal, (9.13). La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

Ejemplo 9.8.1. Consideramos la serie de la tasa de cambio USD-Pound, libra inglesa, mensual, entre 1970-01 y 1997-08, como se muestra en la Figura 9.7. Se aplican los tres casos de la prueba ADF para existencia de raíz unitaria, con la librería `aTSA`. Los resultados conducen a no rechazar la nula, por lo que se detecta raíz unitaria, y es factible ajustar un modelo ARIMA.

```
require(aTSA)
aTSA::adf.test(y)
#-----resultados
Augmented Dickey-Fuller Test
alternative: stationary
```

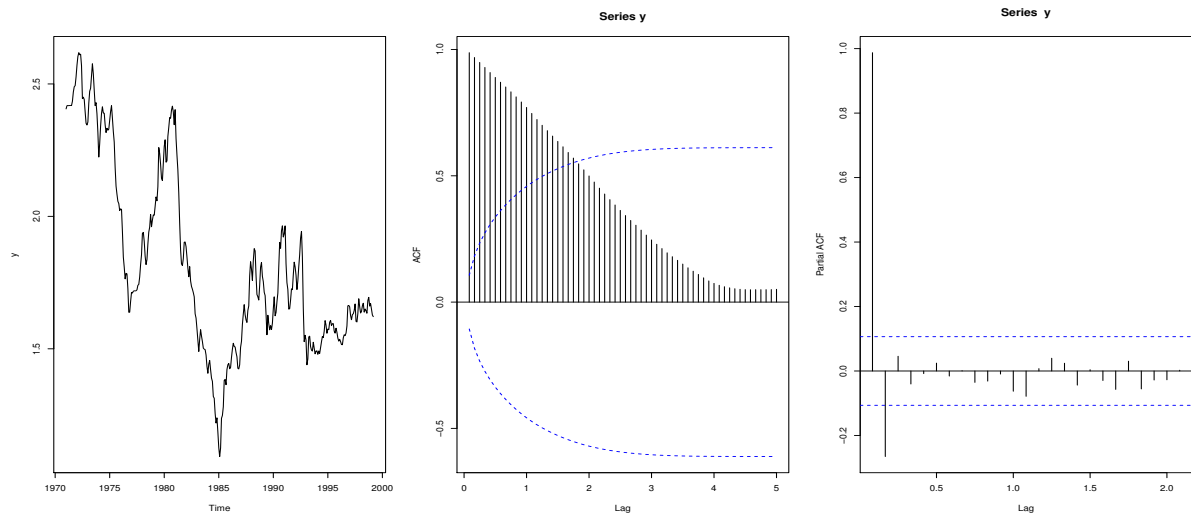


Figura 9.7: fac y facp muestrales de la serie USD/Pound

Type 1: no drift no trend

	lag	ADF	p.value
[1,]	0	-1.26	0.229
[2,]	1	-1.04	0.306
[3,]	2	-1.10	0.287
[4,]	3	-1.06	0.302
[5,]	4	-1.07	0.295
[6,]	5	-1.08	0.291

Type 2: with drift no trend

	lag	ADF	p.value
[1,]	0	-1.69	0.451
[2,]	1	-2.15	0.268
[3,]	2	-1.96	0.345
[4,]	3	-2.09	0.293
[5,]	4	-2.04	0.313
[6,]	5	-2.02	0.321

Type 3: with drift and trend

	lag	ADF	p.value
[1,]	0	-1.57	0.758

```
[2,] 1 -2.46 0.382
[3,] 2 -2.10 0.533
[4,] 3 -2.35 0.429
[5,] 4 -2.25 0.472
[6,] 5 -2.20 0.494
-----
```

Note: in fact, p.value = 0.01 means p.value <= 0.01

Ejemplo 9.8.2. Análisis detallado de la prueba ADF. 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()` al caso de series de tiempo, permitiendo incluir valores rezagados Y_{t-j} , con el comando `L(y, j)` y tendencia lineal $a + bt$ con el comando `trend(y)`, como variables predictoras. Por ejemplo, para el modelo de la ecuación (9.19), en la pag. 230, con $p = 3$

$$\begin{aligned}\Delta Y_t &= k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^3 \rho_j \Delta Y_{t-j+1} + \varepsilon_t, \\ &= k_1 + k_2 t + \rho Y_{t-1} + \rho_2 \Delta Y_{t-1} + \rho_3 \Delta Y_{t-2} + \varepsilon_t\end{aligned}$$

se programa, con `dy` la variable ΔY_t , y observando que

$$\begin{aligned}\Delta Y_{t-1} &= \Delta(L(Y_t)) = Y_{t-1} - Y_{t-2} = \\ L(Y_t - Y_{t-1}) &= L(\Delta(Y_t)).\end{aligned}$$

```
dy = diff(y, 1, 1)
reg.trend = dynlm(dy ~ trend(y) + L(y, 1) + L(dy, 1) + L(dy, 2)).
```

A continuación se programan las regresiones para $p = 1, 2$, excluyendo el caso “none” de media cero.

```
## estrategia de regresiones para la df aumentada
```

```
require(dynlm)

reg.drift1 = dynlm(dy ~ L(y,1))
reg.drift2 = dynlm(dy ~ L(y,1) + L(dy, 1))

reg.trend1 = dynlm(dy ~ trend(y) + L(y,1))
reg.trend2 = dynlm(dy ~ trend(y) + L(y,1) + L(dy, 1))

(c(AIC(reg.drift1),AIC(reg.drift2)))
(c(AIC(reg.trend1),AIC(reg.trend2)))

-----resultados de los valores AIC para los 4 modelos
-1202.376 -1240.019
-1203.215 -1245.147
```

Se detecta el caso `reg.trend1`, que corresponde a $p = 2$ como el de menor AIC.

Y la correspondiente prueba DF aumentada se programa con la función `ur.df()` de la librería `urca`

```
#-----programa la prueba ADF con p = 2 y trend

df.trend = ur.df(y = y, lags = 2, type = 'trend' )
summary(df.trend)
#-----resultados de la prueba DF aumentada
Test regression trend
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Value of test-statistic is: -2.58 2.9804 3.448
Critical values for test statistics:
      1pct   5pct 10pct
tau3  -3.98  -3.42 -3.13
phi2   6.15   4.71  4.05
phi3   8.34   6.30  5.36
```

A partir de los resultados se concluye que no se rechaza la hipótesis de raíz unitaria ya que $\hat{\tau} > \tau_{0.05}$, 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 9.4.2 se ajustó el modelo ARIMA(3,1,2) que produce buenos pronosticos del logaritmo de la tasa.

9.9. La prueba KPSS

La prueba KPSS (Kwiatkowski-Phillips-Schmidt-Shin), Kwiatkowski et al. [1992], aproxima la serie $Y_t, t = 1, 2, \dots, T$ con tres tipos de marchas aleatorias, en lugar de las tres transformaciones Dickey-Fuller de procesos AR(p) en las pruebas DF aumentadas.

$$Y_t = U_t + \epsilon_t, \text{ (none)} \quad (9.20a)$$

$$Y_t = \mu + U_t + \epsilon_t, \text{ (drift)} \quad (9.20b)$$

$$Y_t = \mu + \alpha t + U_t + \epsilon_t, \text{ (trend)} \quad (9.20c)$$

donde U_t es una marcha aleatoria sin tendencia de la forma $U_t = U_{t-1} + a_t$, con $a_t \sim iid N(0, \sigma^2)$. Y $\epsilon_t \sim iid N(0, \sigma_\epsilon^2)$, independiente de $\{a_t\}$.

La hipótesis nula es $H_0 : \sigma^2 = 0$, versus $H_a : \sigma^2 > 0$. No rechazar la nula equivale a asumir que la serie Y_t es estacionaria. Excepto por una tendencia lineal, en el caso “trend”. Entonces, la existencia de raíz unitaria es la hipótesis alterna.

El estadístico de la prueba es

$$K = \frac{1}{T^2} \frac{\sum_{j=1}^T S_j^2}{\hat{\sigma}^2} \quad (9.21)$$

donde $S_j^2 = \sum_{t=1}^j \hat{\epsilon}_t$, y $\hat{\sigma}^2$ es el estimador Newey-West de la varianza, con base en l rezagos, ver ⁽⁴⁾.

La prueba KPSS se puede programar en la librería `aTSA` con la función `kpss.test`. El parámetro `lag.short` es una variable lógica para el número de rezagos l en la fórmula del estimador Newey-West. Asumimos el valor por defecto `lag.short = TRUE`.

Ejemplo 9.9.1. *Considerando la serie de la tasa de cambio USD-Pound, libra inglesa, mensual, entre 1970-01 y 1997-08, como se muestra en la Figura 9.7. Se*

⁴https://en.wikipedia.org/wiki/Newey-West_estimator

aplican los tres casos de la prueba KPSS para existencia de raíz unitaria. Los resultados conducen a no rechazar la nula, por lo que no se detecta raíz unitaria, y no sería conveniente utilizar un modelo ARIMA.

```
#-----prueba KPSS
require(aTSA)
kpss.test(y, lag.short = TRUE, output = TRUE)
#-----resultados
KPSS Unit Root Test
alternative: nonstationary
```

```
Type 1: no drift no trend
```

```
lag    stat p.value
  4 0.0542    0.1
```

```
-----
```

```
Type 2: with drift no trend
```

```
lag    stat p.value
  4 0.0543    0.1
```

```
-----
```

```
Type 1: with drift and trend
```

```
lag    stat p.value
  4 0.0563    0.1
```

```
-----
```

```
Note: p.value = 0.01 means p.value <= 0.01
```

```
      : p.value = 0.10 means p.value >= 0.10
```

9.10. La prueba PP

La prueba PP, Phillips and Perron [1988], aproxima la serie Y_t con tres tipos de regresiones AR(1), similar a las pruebas DF básicas, pero sin realizar las transformaciones Dickey-Fuller.

$$Y_t = \rho Y_{t-1} + \epsilon_t, \text{ (none)} \quad (9.22a)$$

$$Y_t = \mu + \rho Y_{t-1} + \epsilon_t, \text{ (drift)} \quad (9.22b)$$

$$Y_t = \mu + \alpha t + \rho Y_{t-1} + \epsilon_t, \text{ (trend)} \quad (9.22c)$$

La hipótesis nula es $H_0 : \rho = 1$, versus $H_0 : \rho < 1$. Luego, no rechazar la nula indica que existe raíz unitaria. El estadístico de la prueba es similar al estadístico Dickey-Fuller, $\tau = (\rho - 1)/\sigma$, pero la desviación σ se calcula con el estimador estimador Newey-West de la varianza, con base en l rezagos.

La prueba PP se puede programar en la librería `aTSA` con la función `pp.test`. El parámetro `lag.short` es una variable lógica, como en la prueba KPSS, y se usará el valor por defecto `TRUE`. Los resultados muestran evidencia de existencia de raíz unitaria en los tres casos.

```
require(aTSA)
#-----prueba Phillips-Perron para la serie USD/Pound
aTSA::pp.test(y, lag.short = TRUE, output = TRUE)
#-----resultado
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
lag  Z_rho p.value
    5 -0.624   0.553
-----
Type 2: with drift no trend
lag  Z_rho p.value
    5 -5.88   0.41
-----
Type 3: with drift and trend
lag  Z_rho p.value
    5 -8.99   0.505
-----
Note: p-value = 0.01 means p.value <= 0.01
```

9.11. Pruebas de No linealidad en Series de Tiempo

Las definiciones siguientes están en Priestley [1988, Chap. 3]. A partir de la representación de Wold, 8.4.1, pag. 188, todo proceso estacionario en covarianza X_t , sin componentes determinísticos satisface una identidad, en media cuadrática, de la forma (8.33)

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

donde $\varepsilon_t \sim iid(0, \sigma^2)$. Esta es la expresión general para una serie lineal causal.

Una definición general de serie de tiempo Y_t no lineal, causal, es que existe una función $g(\cdot)$ tal que

$$Y_t = g(\varepsilon_t, \varepsilon_{t-2}, \dots). \quad (9.23)$$

Se trata de definir una expresión para este supuesto. Por medio de un argumento con base en la expansión de Taylor de la función $g(\cdot)$ alrededor del punto $(0, 0, \dots)$ se puede extender la representación en el caso lineal a una más general, conocida como expansión de Volterra

$$Y_t = \psi_0 + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{j,k} \varepsilon_{t-j} \varepsilon_{t-k-j} + \dots + \varepsilon_t. \quad (9.24)$$

La expansión anterior muestra que en el conjunto de series que tengan la forma (9.24) se tienen efectos aleatorios mucho más complejos que en el caso lineal. Ejemplos simples de series no lineales son

$$Y_t = \alpha + \beta Y_{t-1} + \gamma \varepsilon_t \varepsilon_{t-1}. \quad (9.25)$$

$$Y_t = \alpha + \beta Y_{t-1} + \gamma Y_{t-1} \varepsilon_{t-1} + \varepsilon_t. \quad (9.26)$$

$$Y_t = 2.0 + 0.2 \varepsilon_{t-1} + 0.5 \varepsilon_{t-1} \varepsilon_{t-2} + 1.3 \varepsilon_{t-1} \varepsilon_{t-3} + \varepsilon_t. \quad (9.27)$$

En todos los casos se asume $\varepsilon_t \sim iid N(0, \sigma^2)$. Interesa implementar algunas pruebas para detectar no linealidad.

La prueba Teräsvirta asume que la serie Y_t sigue un modelo no lineal tipo red neuronal autorregresiva $Y_t \sim NNAR(p, 0, q)$, con p nodos de entrada y q nodos en la capa oculta, definida en (5.14), más una componente AR(p) estacionaria en covarianza

$$\Phi_p(L)Y_t = \sum_{j=1}^q \alpha_j f \left(\beta_{0,j} + \sum_{k=1}^p \beta_{j,k} Y_{t-k} \right) + \varepsilon_t,$$

donde $f(x) = (1 + e^x)^{-1}$. La hipótesis nula es

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0. \quad (9.28)$$

$$H_a : \text{no } (H_0) \quad (9.29)$$

El estadístico es del tipo Multiplicador de Lagrange, ML, ⁽⁵⁾, pero es difícil de describir, ver Teräsvirta et al. [1993]. Se puede realizar con la librería `tseries`, con la función `terasvirta.test(x)`. El ejemplo siguiente está tomado de las ayudas de esta función.

```
#-----
require(tseries)
n= 300;x=double(n);
#---genera una serie no lineal: H0 es falsa
x[1] <- 0.0
for(i in (2:n)) {
x[i] <- 0.4*x[i-1] + tanh(x[i-1]) + rnorm(1, sd=0.5)}
x <- as.ts(x)
plot(x)
terasvirta.test(x)
#---resultado
Teraesvirta Neural Network Test
data:  x
X-squared = 26.269, df = 2, p-value = 1.976e-06
```

Una prueba similar es la prueba White, Lee et al. [1993], en la librería `tseries`.

```
#-----
require(tseries)
White Neural Network Test
data:  x
X-squared = 29.2, df = 2, p-value = 4.565e-07
```

La librería `nonlinearAnalysis` implementa en la función `nonlinearityTest` seis pruebas de no linealidad, incluyendo las pruebas Teraesvirta y White.

⁵https://en.wikipedia.org/wiki/Score_test

9.12. Modelos híbridos ARIMA - Red Neuronal

Una aplicación de las prueba de no linealidad es la propuesta de modelos híbridos ARIMA-NNAR, introducida por Zhang [2003].

En este trabajo Zhang [2003], se propone que si $Y_t \sim ARIMA(p, 1, q)$, definido en (9.1)

$$\begin{aligned}\Phi_p(L)(1-L)Y_t &= \mu + \Theta_q(L)\varepsilon_t, \\ \varepsilon_t &\sim RB(0, \sigma^2), \quad t = 1, 2, \dots, T,\end{aligned}$$

entonces la serie de residuos ARIMA $\hat{\varepsilon}_t$, podría tener una dinámica no lineal, que no se revela con las pruebas Ljung-Box de incorrelación. Por lo que es necesario utilizar pruebas de no linealidad previamente. En caso de no linealidad, se podría modelar con una red neuronal, por ejemplo una NNAR(p,0,q).

$$\hat{\varepsilon}_t \sim NNAR(p, 0, q),$$

de tal forma que los pronósticos se pueden combinar como pronósticos ARIMA más pronósticos NNAR,

$$\hat{Y}_{T+j} + \hat{\varepsilon}_{T+j}, j = 1, 2, \dots, h.$$

Ejemplo 9.12.1. *El código siguiente implementa un modelo ARIMA para la serie de la tasa de cambio USD/Pound, y los residuos los ajusta con una red neuronal auto regresiva NNAR. Los pronosticos a 30 meses se muestran en la Figura 9.8.*

```
#-----
auto.arima(yi)
ARIMA(0,1,1)(0,0,1)[12]
mod2 = stats::arima(yi,order=c(0,1,1),
seasonal=list(order=c(0,0,1),period=12))
require(lmtest)
coeftest(mod2)
#-----residuos
at=ts(mod2$residuals,frequency=12)
#-----pruebas de no linealidad
```

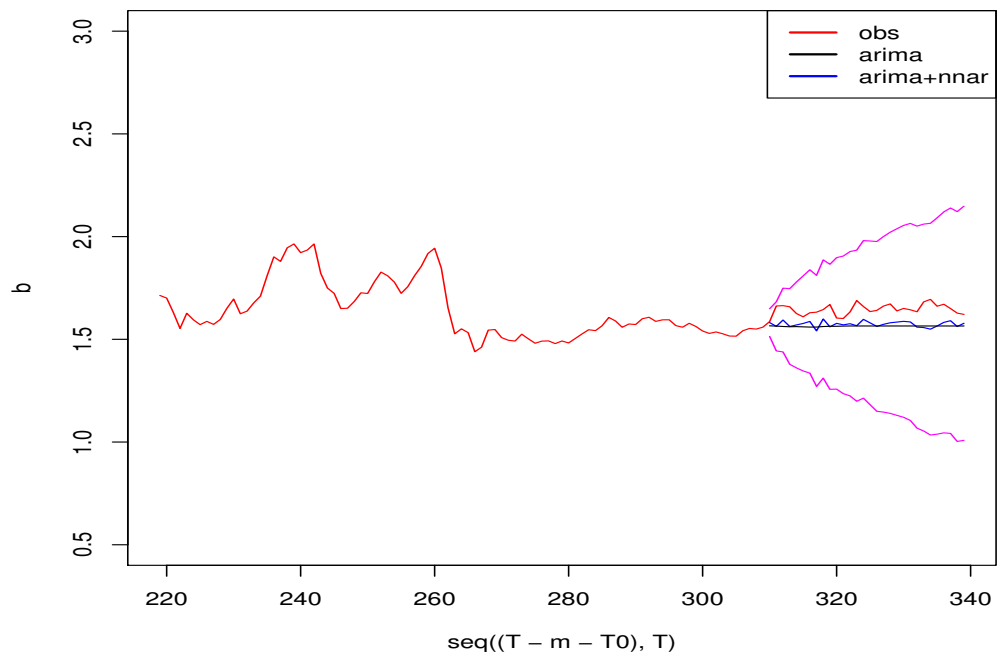


Figura 9.8: Pronósticos con modelo híbrido ARIMA-NNAR de la serie USD/Pound

```

terasvirta.test(at)
white.test(at)
#-----grafica
ts.plot(at)
#-----ajuste NNAR
mod3=nnetar(at)
print(mod3)
#-----pronosticos
pr.nnar = forecast(mod3,h=m)$mean
pr.nnar = ts(pr.nnar,frequency=12,start=c(1996,10))
pr.arima = predict(mod2,m)$pred
sde = predict(mod2,m)$se
pr.arima = ts(pr.arima,frequency=12,start=c(1996,10))
pr = pr.nnar+pr.arima

```

9.13. Combinación de pronósticos

De la página de Rob Hyndman (⁶): “It has been well-known since at least 1969, when Bates and Granger wrote their famous paper on “The Combination of Forecasts”, that combining forecasts often leads to better forecast accuracy. ”

La librería `forecastHybrid` realiza pronósticos con base en los modelos `auto.arima`, `ets`, `nnetar`, `stlm`, `tbats`, y los combina como una suma ponderada para producir un solo pronóstico.

Ejemplo 9.13.1. *El código siguiente implementa una combinación de pronósticos a 30 meses para la serie de la tasa de cambio USD/Pound. Los pronósticos a 30 meses se muestran en la Figura 9.9.*

```
require(forecastHybrid)
mod1 = hybridModel(yi, weights="equal")
#-----pronosticos
B=forecast(mod1,h=m)
Bf = B$pointForecasts
colnames(Bf)
[1] "auto.arima" "ets"          "thetam"          "nnetar"
[5] "stlm"       "tbats"
p1 = apply(Bf,1,mean)
#-----graficar
T0 = 90
b = c(yi[(T-m-T0):(T-m)],yf)
T1 = length(yi)
tt = seq((T-m+1),T)
par(mfrow=c(1,1))
plot(seq((T-m-T0),T),b,
type='l', col=2,ylim=c(0.5,3.0))
lines(tt,p1, col='blue')
lines(tt, p1+1.64*sde, col='magenta')
lines(tt, p1-1.64*sde, col='magenta')
```

⁶<https://robjhyndman.com/hyndsight/forecast-combinations/>

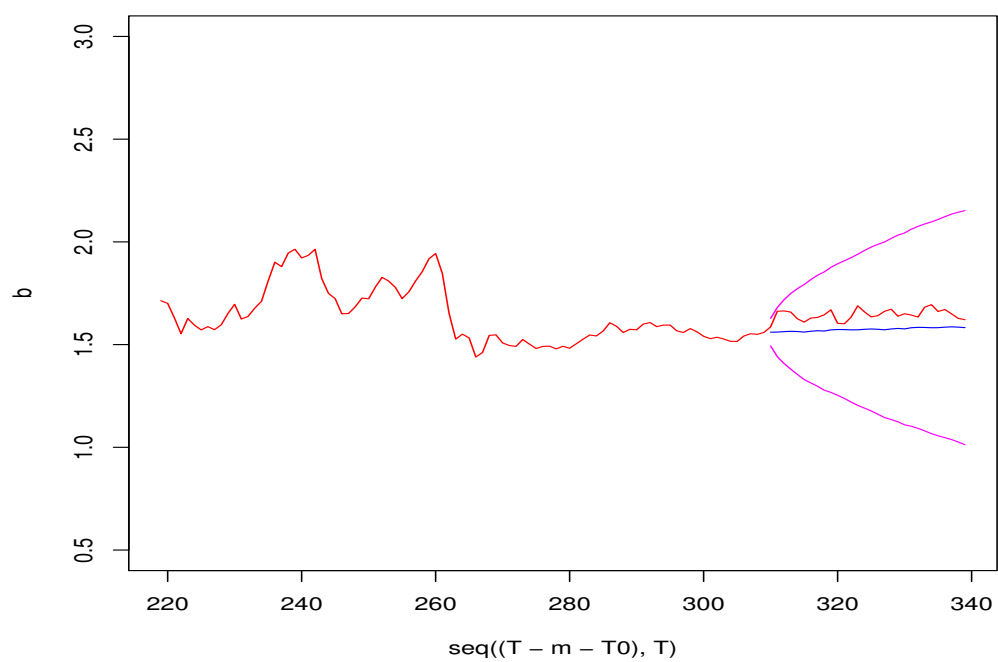


Figura 9.9: combinacion de pronósticos con 6 modelos para la serie USD/Pound

CAPÍTULO 10

Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)

10.1. Modelos SARIMA

En esta sección se utilizará el operador rezago estacional, L^s , donde $s > 1$, entero, es el período de una dinámica estacional, que se asume está presente en la serie Y_t . Se define mediante las operaciones

$$L^s(Y_t) = Y_{t-s}, \quad (10.1a)$$

$$(I - L^s)Y_t = \Delta_s Y_t = Y_t - Y_{t-s}, \quad (10.1b)$$

$$(I - L^s)^2 Y_t = \Delta_s^2 Y_t = (I - 2L^s + L^{2s})Y_t. \quad (10.1c)$$

Si la serie Y_t tiene una componente estacional con período s es posible eliminarla diferenciando una ó dos veces, es decir, transformando Y_t en W_t :

$$W_t = \Delta_s^D Y_t, \quad D = 1, 2. \quad (10.2)$$

En R la instrucción para calcular (10.2) es `diff(y, s, D)`. Se espera que la serie W_t sea estacionaria en covarianza, con una posible estructura *SARMA*. En caso

de existir una componente de tendencia y otra de estacionalidad en la serie Y_t , es posible eliminar ambas mediante diferenciaciones ordinaria y estacional

$$W_t = (1 - L^s)^D (1 - L)^d Y_t = \Delta_s^D \Delta^d Y_t \quad (10.3)$$

En R la transformación (10.3) se puede realizar con la instrucción siguiente

```
Wt = diff(diff(Yt, 1, d), s, D)
```

La transformación (filtro lineal) $X_t = \Delta^d \Delta_s^D Y_t$ elimina la tendencia y la estacionalidad de Y_t , dejando una posible estructura $SARMA(p, q)(p_s, q_s)[s]$, definida en (8.44), pag. 199.

Ejemplo 10.1.1. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

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

En el Ejemplo 8.2.5, en la pag. 169, se ajustó un MA(6) a los residuos estimados $\hat{\varepsilon}_t$. En el Ejemplo 8.5.5, en la pag. 208, se estimó un modelo alterno SARMA(2,1)(2,1)[4] para los residuos. En la Figura 10.1 se muestra el efecto de la diferenciación ordinaria, $U_t = \text{diff}(Y_t, 1, d)$, en el panel superior derecho. Luego el de la diferenciación estacional $W_t = \text{diff}(U_t, s, D)$, en el panel inferior derecho. El resultado es que la serie W_t muestra una dinámica que puede corresponder a un $ARIMA(3, 0, 2)(0, 0, 1)[4]$ with zero mean.

Código R 10.1.1.

```
#-----Ejemplo diferencias serie cemento
E = read.table("cementq.dat", header = TRUE)
attach(E)
y = ts(y, frequency=4, start=c(1956, 1), end=c(1994, 3))

#----- con stl()
m1 = stl(y, s.window = 'per')
plot(m1)
```

```

dyos = diff(diff(y, 4, 1), 1, 1)
dys = diff(y, 4, 1)
dyo = diff(y, 1, 1)

par(mfrow = c(2, 2))
ts.plot(y, main=" (A) ")
ts.plot(dyo, main=" (B) ")
ts.plot(dys, main=" (C) ")
ts.plot(dyos, main=" (D) ")

require(TSA)
par(mfrow = c(1, 2))
acf(dyos, 16, drop.lag.0=TRUE, ci.type="ma")
pacf(dyos, 16)

dyos = ts(dyos, frequency=4)
require(forecast)
auto.arima(dyos)
ARIMA(3, 0, 2)(0, 0, 1)[4] with zero mean

```

Definición 10.1.1. *Se dice que la serie $(Y_t, t \in \mathbb{Z})$, sigue un modelo SARIMA(p, d, q)(p_s, D, q_s)[s] si cumple que*

$$\Phi_{p_s}(L^s)\Phi_p(L)\Delta_s^D\Delta^d Y_t = \Theta_{q_s}(L^s)\Theta_q(L)\varepsilon_t, \quad (10.4)$$

donde $d, D = 1, 2$, y $\varepsilon_t \sim RB(0, \sigma^2)$. Y

$$\begin{aligned} \Phi_p(L) &= 1 - \sum_{j=1}^p \varphi_j L^j, \\ \Theta_q(L) &= 1 + \sum_{j=1}^q \theta_j L^j, \\ \Phi_{p_s}(L^s) &= 1 - \sum_{j=1}^{p_s} \varphi_{s,j} L^{js}, \\ \Theta_{q_s}(L^s) &= 1 + \sum_{j=1}^{q_s} \theta_{s,j} L^{js}. \end{aligned}$$

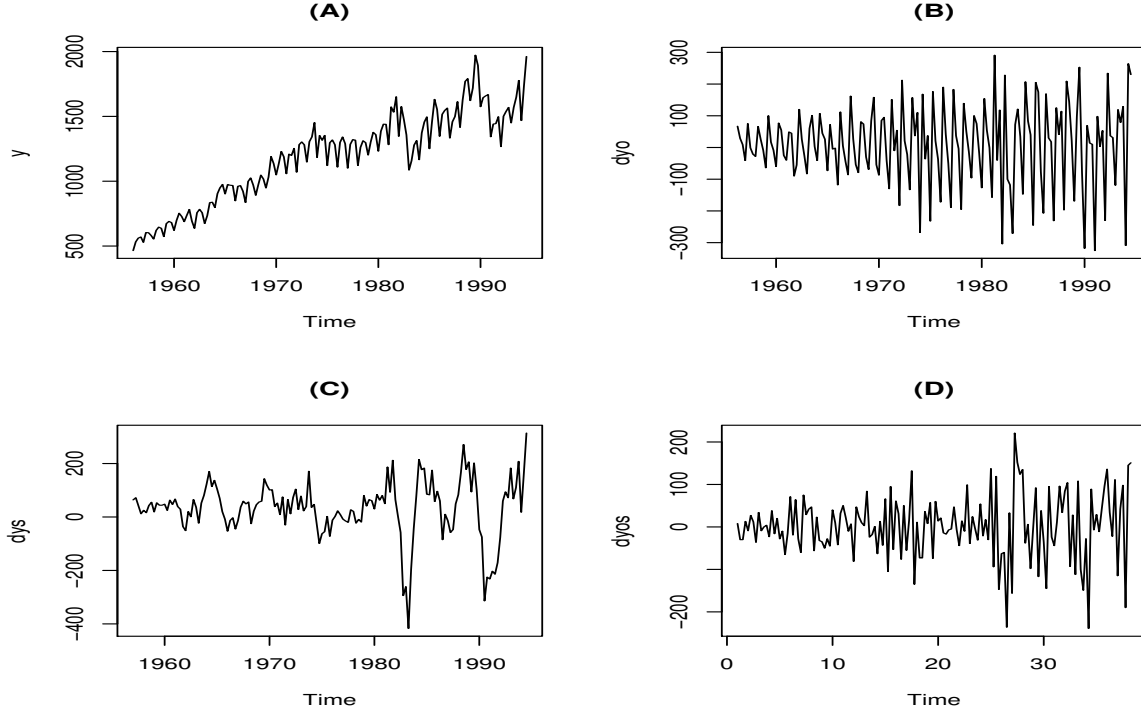


Figura 10.1: Aplicación de diferencia simple y luego diferencia estacional a la serie de producción de cemento

Las notaciones en la Definición anterior no son las utilizadas en la mayoría de los textos, están tomadas del libro Prado and West [2010]. La idea del modelo SARIMA es que la serie diferenciada W_t que satisface

$$\varphi_p(L)\Phi_P(L^s)W_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2).$$

tiene un papel equivalente al residuo estructural ε_t en el modelo de Componentes

$$Y_t = \sum_{j=0}^k \beta_j t^j + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t. \quad (10.5)$$

En (10.4), al diferenciar la serie Y_t , se eliminan la tendencia y estacionalidad aleatorias. Pero diferenciar en (10.5) para eliminar las respectivas componentes determinísticas puede ser 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 10.1.2. Suponga un modelo $SARIMA(1, 1, 0)(1, 1, 0)[12]$, es decir, $s = 12$, $p = 1, q = 0, p_s = 1, q_s = 0, d = D = 1$, dado por

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})(1 - L^{12})(1 - L)Y_t = \varepsilon_t,$$

Colocando $X_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ se obtiene, de manera equivalente, $(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})X_t = \varepsilon_t$. Cuando se desarrollan los polinomios autoregresivos del modelo se obtiene

$$(1 - \varphi_1 L - \varphi_{12} L^{12} + \varphi_1 \varphi_{12} L^{13})X_t = \varepsilon_t,$$

ó también, $X_t = \varphi_1 X_{t-1} - \varphi_{12} X_{t-12} + \varphi_1 \varphi_{12} X_{t-13} + \varepsilon_t$.

Note que este modelo es un $X_n \sim AR(13)$ con restricciones sobre ciertos coeficientes que se asumen cero. Se podría indicar por $AR([1, 12, 13])$. El parámetro $\varphi_1 \varphi_{12}$ se estima como un solo valor φ_{13} y puede ser no significativo. En el caso que sea no significativo el modelo se dice sin interacción, y se expresa como

$$(1 - \varphi_1 L - \varphi_{12} L^{12})X_t = \varepsilon_t.$$

Considere los 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*, ver Smith [2019], para simular trayectorias de modelos SARIMA y así visualizar este modelo.

```
# simulacion sarima(1,1,0)(1,1,0)[12] con la libreria CombMSC
library(CombMSC)
library(forecast)
library(TSA)

fi=0.8; fi12=0.737;
sigma= sqrt(2)
y = sarima.Sim(n = 20, period =12,
model=list(order=c(1,1,0),ar = fi,
ma = NULL, sd = sigma),
seasonal=list(order=c(1,1,0),ar=fi12,ma = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)

#-----note: longitud de la serie: 20*12=240
```

```

dyls = diff(diff(y,12,1),1,1)
dys = diff(y,12,1)
dyl = diff(y,1,1)

```

```

par(mfrow = c(2,2))
ts.plot(y,main="(A) ")
ts.plot(dyl,main="(B) ")
ts.plot(dys,main="(C) ")
ts.plot(dyls,main="(D) ")

```

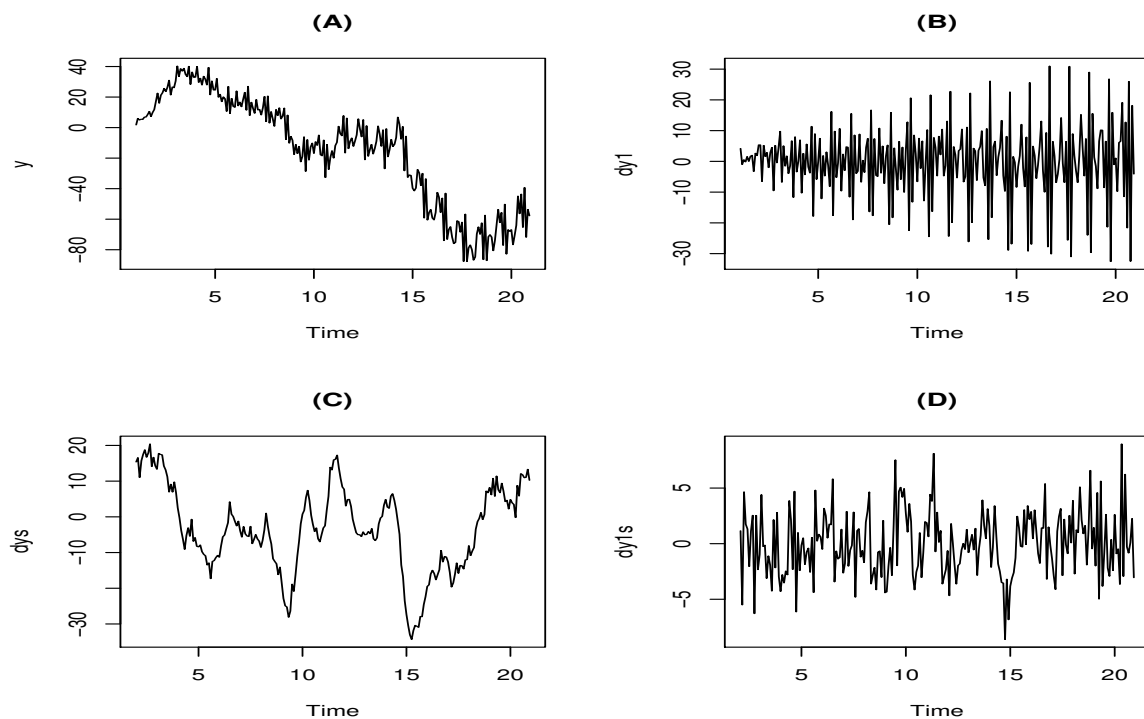


Figura 10.2: (A) Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12], del Ejemplo 10.1.2, con (B) diferencia ordinaria, (C) diferencia estacional, (D) ambas diferencias

Ejemplo 10.1.3. Modelo “Air-passengers” Se define como un modelo multiplicativo de la forma SARIMA(0,1,1)(0,1,1)[12] dado por

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1)(1 + \theta_{12}L^{12})\varepsilon_t, \quad |\theta_1| < 1, |\theta_{12}| < 1. \quad (10.6)$$

Es un modelo utilizado con frecuencia para modelar series con tendencia lineal y componente estacional. Por ejemplo la serie “airline”, del texto de Box y Jenkins: “Monthly totals of international airline passengers, 1949 to 1960”. En R se puede visualizar con

```
data("AirPassengers")
AP = AirPassengers
ts.plot(AP)
```

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 1,12,13. La función de autocorrelación de W_t es

$$\begin{aligned}\sigma^2 &= (1 + \theta_1^2)(1 + \theta_{12}^2) \\ \rho_1 &= \theta_1(1 + \theta_{12}^2) \\ \rho_{12} &= \theta_{12}(1 + \theta_1^2) \\ \rho_{13} &= \theta_1\theta_{12}\end{aligned}$$

Suponga el modelo air-passengers con parámetros $\theta_1 = -0.377$, $\theta_{12} = -0.572$, $\sigma^2 = 0.0014$, el siguiente código simula una trayectoria de este modelo. El resultado está en la Figura 10.3. Nótese que $\rho_1, \rho_{12} < 0$ y $\rho_{13} > 0$.

Código R 10.1.2.

```
#----- simular el modelo airline
ti=-0.377; ti12=-0.572;
sigma= sqrt(0.0014)
library(CombMSC)
y = sarima.Sim(n = 20, period =12,
model=list(order=c(0,1,1),ma = ti,
ar = NULL, sd = sigma),
seasonal=list(order=c(0,1,1),ma=ti12,ar = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)
```

```

Wt = diff(diff(y,12,1),1,1)
par(mfrow=c(2,2))
ts.plot(y)
ts.plot(Wt)
acf(Wt,48,drop.lag.0=TRUE,ci.type="ma")
pacf(Wt,48)

# identificar
auto.arima(dy)
ARIMA(0,0,1)(0,0,1)[12] with zero mean

# identificador armasubset() libreria TSA
res=armasubsets(y=dy,
nar=14,nma=14,
y.name='dy',
ar.method='ols')
layout(1:1)
plot(res)

# estimar
m1 = arima(y,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=12))
summary(m1)

# pronosticar
pr = predict(m1,n.ahead=12)$pred

```

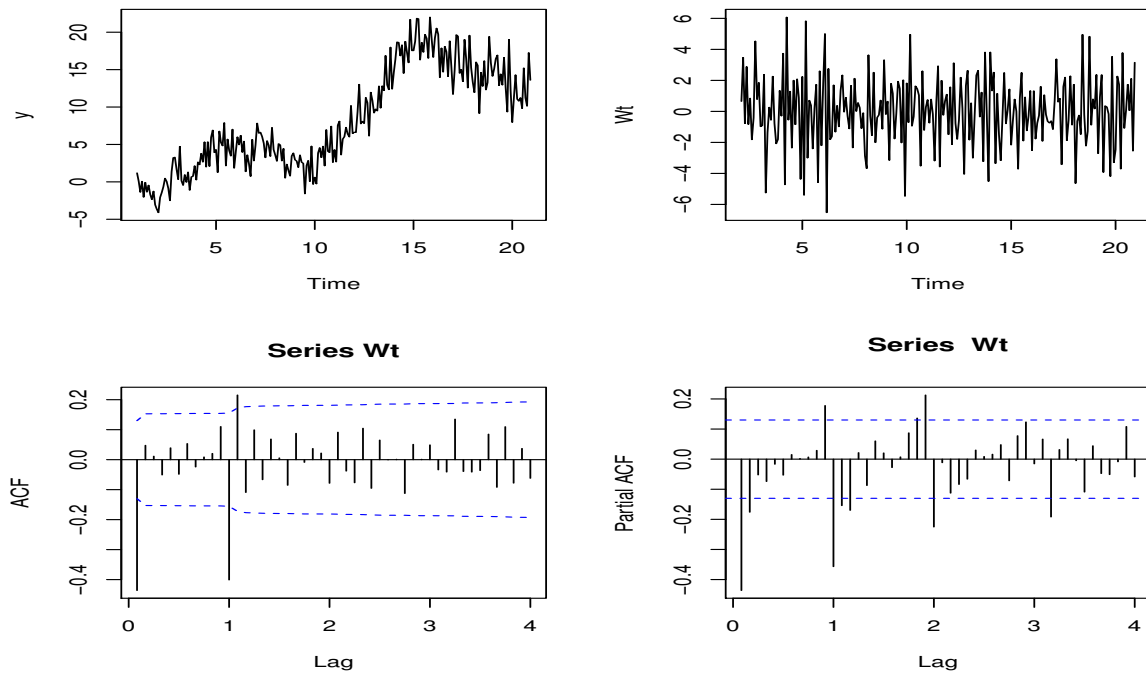



Figura 10.3: Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12], serie doblemente diferenciada y su acf y pacf. Nótese que las autocorrelaciones en los rezagos 1,12,13 son significativas, con el signo correcto

Ejercicio 10.1.1. Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA $SARIMA(1, 1, 1)(1, 1, 1)[4]$ definido por

$$(1 - \varphi_1 L)(1 - \varphi_4 L^4)(1 - L)(1 - L^4)Y_t = (1 - \theta_1 L)(1 - \theta_4 L^4)\varepsilon_t.$$

Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA con la serie Y_t . Es equivalente a estimar un modelo ARMA(1,1)(1,1) con la serie diferenciada, $W_t = (1 - L)(1 - L^4)Y_t$? Use los siguientes valores para los parámetros.

```
fi=-0.8; fi4=-0.37; ti=-0.64; ti4=-0.513; sigma= sqrt(0.014)
```

10.2. Pruebas de Raíz Unitaria Estacional

Para introducir la idea de las pruebas de raíz unitaria estacional hay que recordar que una serie $Y_t \sim ARMA(p, q)$ se dice integrada, ó con una raíz unitaria si se cumple que el polinomio autorregresivo $\Phi_p(z)$ tiene una raíz igual a 1. Es decir, si se cumple que $\Phi_p(1) = 0$.

Definición 10.2.1. Una serie de tiempo $Y_t \sim SARMA(p, q)(p_s, q_s)[s]$ se dice integrada estacional, de período $s > 1$, si se cumple que algunas ó todas las raíces de $z^s = 1$ son raíces de la ecuación del polinomio autorregresivo conjunto $\Phi_p(z)\Phi_{p_s}(z) = 0$.

Las raíces de $z^s = 1$ se denominan las raíces de la unidad, y son s números complejos. Con $s = 4, 12$, por ejemplo.

Para el caso $s = 4$ se tiene que las cuatro raíces de la unidad son $1, -1, i, -i$, donde $i = \sqrt{-1}$ es la unidad imaginaria.

Para el caso $s = 12$, 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 $z = re^{i\theta}$ y la fórmula de Euler $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, donde $r = |z|$ y θ es el ángulo que forma $|z|$ con el eje X, se pueden identificar las raíces unitarias con los ángulos $\theta = 2\pi j/s, j = 0, 1, \dots, s-1$, ya que se tiene $r = 1$ en todos los casos, pues las raíces unitarias tienen módulo 1. Estos ángulos se denominan las “frecuencias estacionales”. Por ejemplo, para el caso $s = 12$ se identifican con los ángulos $\theta_j = 2\pi j/12, j = 0, 1, \dots, 11$

Se han desarrollado pruebas para detectar raíces unitarias estacionales, como las pruebas HEGY, ver Hylleberg et al. [1990], y la prueba Canova-Hansen, indicada por CH, ver Canova and Hansen [1995], que se exponen a continuación,

10.3. La prueba Canova-Hansen

Los modelos de componentes estructurales con componente estacional y (posibles) errores SARMA y los SARIMA pueden considerarse, 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”.

En Ghysels and Osborn [2001, pag. 29] se anota:

"[...] for the sample sizes often observed in practice, it may be difficult to discriminate between deterministic seasonality and a seasonal unit root process."

La prueba CH se basa en esta observación. La inestabilidad estructural de la componente estacional se toma como una característica equivalente a la existencia de una raíz unitaria estacional, opuesta a Componente estacional más residuo estacionario estacional (SARMA).

Ghysels and Osborn [2001, pag. 42]:

"It is now well known that a series generated by a unit root process can wander widely and smoothly over time without any inherent tendency to return to its underlying mean value [...]. In the seasonal context, there are S unit root processes, none of which has an inherent tendency to return to a deterministic pattern. As a result, the values for the seasons can wander widely and smoothly in relation to each other [...]"

Descripción de la Prueba

El modelo que asume la prueba es un modelo sin tendencia y con componente estacional, descrita mediante funciones trigonométricas, 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

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \sin\left(\frac{2\pi jt}{s}\right), \quad (10.7)$$

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 (10.7) es

$s - 1$, el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores

$$\underline{\gamma} = (\gamma_1, \dots, \gamma_{s/2})',$$

$$\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'.$$

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 modelos de marcha aleatoria:

$$\underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \quad (10.8)$$

$$\underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t, \quad (10.9)$$

donde $\underline{u}_t, \underline{v}_t$ son vectores iid, independientes de media cero y varianzas no nulas σ^2, σ_*^2 . 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.

$$H_0 : \quad \underline{\gamma}_t \equiv \underline{\gamma}, \underline{\gamma}_t^* \equiv \underline{\gamma}^*,$$

$$H_a : \quad \underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t.$$

En Canova and Hansen [1995, pag. 240, sec. 2.2], se re-escribe la prueba con base en cierto parámetro τ^2 , colocando $H_0 : \tau^2 = 0$, versus $H_a : \tau^2 > 0$. El estadístico de la prueba es del tipo Multiplicador de Lagrange, de la forma

$$LM = \frac{1}{T^2} tr \left((A' \Omega A)^{-1} A' \left(\sum_{j=1}^T \underline{F}_j \underline{F}_j' \right) A \right), \quad (10.10)$$

donde A, Ω, F_j son matrices, $tr()$ es la traza de una matriz, T es la longitud de la serie, pero las definiciones específicas no se considerarán aquí y se remite al artículo original. Además, en el artículo se proveen los valores críticos para varios niveles de significación. La nula se rechaza si el valor observado del estadístico es mayor que el valor crítico

$$LM_{obs} > LM_{\alpha}.$$

Implementación en R

La prueba Canova-Hansen está implementada en la librería `uroot`, con la función `CH.test`, que tiene los parámetros siguientes

```
ch.test(x, type = c("dummy", "trigonometric"),
lag1 = FALSE, NW.order = NULL, sid = NULL,
xreg = NULL, pvalue = c("RS", "raw"),
rs.nobsreg = 13),
```

la serie debe estar en la variable x , y debe haberse declarado como objeto “ts”, con `frequency=4, 12`, solamente.

El parámetro `type = c("dummy", "trigonometric")` indica si se utilizan variables trigonométricas ó indicadoras para modelar la estacionalidad.

El parámetro `lag1` es variable lógica. Si `TRUE` se incluye el primer rezago de la serie como variable explicativa en la regresión. Por defecto se coloca `FALSE`.

El parámetro `NW.order` es un entero que determina cuántos rezagos se utilizan para la estimación de la matriz de covarianza en la fórmula de Newey y West.

El parámetro `sid` es un vector numérico optional que especifica cuáles frecuencias deberían incluirse en la prueba. Por ejemplo, para $s = 4$ hay 4 frecuencias que corresponden a los ángulos $0, \pi/2, \pi, 3\pi/2$. Por defecto se escogen todas las frecuencias.

El parámetro `xreg` es un vector ó matrix opcional de variables exógenas con el mismo número de filas que la serie x .

Nótese, sin embargo, que para aplicar la prueba CH es necesario eliminar la tendencia lineal de la serie Y_t , en caso de existir. Para aplicar la prueba se prefirió filtrar previamente esta tendencia, colocando $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$, donde \hat{T}_t se estimó mediante el filtro `stl()` ó Loess.

Ejemplo 10.3.1. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelar de la serie de producción de cemento Portland, trimestral, y decidir entre los modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA, se aplica la prueba Canova-Hansen para determinar la posibilidad de una raíz unitaria estacional que apoye un posible modelo SARIMA.*

La prueba se implementa con los comandos siguientes. Nótese que se utilizó el procedimiento `stl()` para extraer la tendencia de la serie.

Código R 10.3.1.

```

#-----Ejemplo prueba Canova-Hansen
E = read.table("cementq.dat", header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
ts.plot(y)
#-----eliminar la tendencia con stl()
m1 = stl(y, s.window = 'per')
s1 = m1$time.series[,1]; e1 = m1$time.series[,3];
y1 = s1+e1
#-----implementacion de la prueba Canova-Hansen
require(uroot)
res = ch.test(y1,
type = "trigonometric", pvalue='raw')
res
#-----resultado
data:  y1
      statistic pvalue
pi/2      2.4867   0.01 **
pi         1.0541   0.01 **
joint      2.761   0.01 **
Test type: seasonal cycles
NW covariance matrix lag order: 4
First order lag: no
Other regressors: no
P-values: interpolation in original tables

```

Como se detecta una raíz unitaria estacional al nivel de 5 % en dos de las frecuencias de un total de dos (período 4), se rechaza la nula y existe evidencia de inestabilidad estructural en la componente estacional. Por tanto, el modelo SARIMA sería recomendable para modelar la serie.

10.4. La prueba HEGY

La prueba HEGY (Hylleberg, Engle, Granger, Yoo), en Hylleberg et al. [1990], es una generalización de la prueba aumentada de Dickey-Fuller. Permite decidir cuáles de las frecuencias estacionales corresponden a raíces estacionales significativas. En caso de detectarse al menos una, exceptuando la frecuencia que corresponde a cero, se concluirá que la serie en cuestión está integrada estacionalmente, que es la hipótesis nula.

Esta sección describe la idea básica de la prueba, que utiliza una regresión lineal múltiple. Los estadísticos para detectar las raíces unitarias estacionales son tipo t-Student y F, correspondientes a los estadísticos para pruebas de hipótesis sobre los parámetros de la regresión. Esta sección es con base en el artículo de Hylleberg et al. [1990].

Inicialmente se asume que la serie Y_t tiene una componente estacional de período s par, y se puede aproximar por un proceso AR(s), tal que

$$\Phi_s(L)Y_t = \varepsilon_t,$$

para ε_t ruido blanco. Se asume las s raíces estacionales como los números complejos $\theta_j \in \mathbb{C}$, $j = 1, \dots, s$, en los cuales $\Phi_s(\theta_j)$ es un número finito. Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - z/\theta_k, \\ \Delta(z) &= \prod_{k=1}^s \delta_k(z), \\ \lambda_k &= \Phi_s(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k). \end{aligned} \quad (10.11)$$

Entonces se cumple, por una identidad debida a Lagrange, que

$$\Phi_s(z) = \sum_{k=1}^s \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\Phi_s^*(z), \quad (10.12)$$

donde $\Phi_s^*(z)$ es un polinomio posiblemente infinito ó racional. Nótese que, por la definición (10.11) se cumple

$$\Phi_s(\theta_k) = 0 \Leftrightarrow \lambda_k = 0 \quad (10.13)$$

por lo tanto, para determinar si θ_k es raíz estacional se debe comprobar $\lambda_k = 0$. Pero estas λ_k se pueden asimilar a los coeficientes en una regresión lineal múltiple que se define a partir de la identidad (10.12). Y por tanto, probar $\lambda_k = 0$ se puede hacer, en principio, mediante una prueba t-Student de significación de parámetros, ó con una prueba F.

Por ejemplo, para el caso $\Phi_s(z) = 1 - z^4$, donde $\theta_1 = 1, \theta_2 = -1, \theta_3 = i, \theta_4 = -i$, la identidad (10.12) se transforma en

$$\begin{aligned}\Phi_s(z) &= \lambda_1 z(1+z)(1+z^2) + \lambda_2 (-z)(1-z)(1+z^2) \\ &+ \lambda_3 (-iz)(1-z)(1+z)(1-iz) + \lambda_4 (iz)(1-z)(1+z)(1+iz) \\ &+ \Phi_s^*(z)(1-z^4).\end{aligned}\quad (10.14)$$

Si se definen nuevas constantes $\pi_1, \pi_2, \pi_3, \pi_4$ mediante las relaciones:

$$\pi_1 = -\lambda_1, \pi_2 = -\lambda_2, 2\lambda_3 = -\pi_3 + i\pi_4, 2\lambda_4 = -\pi_3 - i\pi_4$$

entonces la identidad (10.14) queda

$$\begin{aligned}\Phi_s(z) &= -\pi_1 z(1+z+z^2+z^3) - \pi_2 (-z)(1-z+z^2-z^3) \\ &- (\pi_4 + \pi_3 z)(-z)(1-z^2) + \Phi_s^*(z)(1-z^4).\end{aligned}\quad (10.15)$$

Finalmente, en $\Phi_s(L)Y_t = \varepsilon_t$ se reemplaza $\Phi_s(L)$ por la correspondiente expresión obtenida en (10.15). Si se definen las variables

$$\begin{aligned}X_{1,t} &= (1 + L + L^2 + L^3)Y_t, \\ X_{2,t} &= -(1 - L + L^2 - L^3)Y_t, \\ X_{3,t} &= -(1 - L^2)Y_t, \\ X_{4,t} &= (1 - L^4)Y_t,\end{aligned}$$

y se reemplaza $\Phi_s^*(L) = 1$, se obtiene la ecuación

$$X_{4,t} = \pi_1 X_{1,t} + \pi_2 X_{2,t} + \pi_3 X_{3,t} + \varepsilon_t. \quad (10.16)$$

La ecuación (10.16) es una regresión lineal múltiple. A partir de esta ecuación se plantea la prueba de hipótesis HEGY para el caso de período $s = 4$. La hipótesis en la prueba HEGY, para el caso $s = 4$, se compone de tres hipótesis, para las raíces unitarias estacionales: $\pm 1, \pm i$, definidas de la manera siguiente.

$$H_0 : \Phi_s(1) = 0 \Leftrightarrow \pi_1 = 0, \quad H_a : \Phi_s(1) > 0 \Leftrightarrow \pi_1 < 0,$$

$$H_0 : \Phi_s(-1) = 0 \Leftrightarrow \pi_2 = 0, \quad H_a : \Phi_s(2) > 0 \Leftrightarrow \pi_2 < 0,$$

$$H_0 : |\Phi_s(i)| = 0 \Leftrightarrow \pi_3 = \pi_4 = 0, \quad H_a : no(H_0).$$

Si alguna hipótesis nula no se rechaza se asume que existe raíz unitaria estacional, excepto la raíz $z = 1$ que corresponde a una raíz unitaria para tendencia aleatoria y es equivalente a la prueba Dickey-Fuller.

Los estadísticos de las pruebas son t-Student para $z = 1, -1$, y F para $\pm i$. Las decisiones se toman con base en los valores p correspondientes.

En el caso $s = 12$, hay 12 raíces unitarias, $z_j = \cos(2\pi j/12) + i \sin(2\pi j/12)$, $j = 0, 1, \dots, 11$.

Ver un ejemplo de código en R para calcular las raíces y graficarlas en: ⁽¹⁾.

Hay 7 hipótesis, dos para $z = 1, -1$, y una para cada una de las cinco parejas de raíces unitarias estacionales conjugadas.

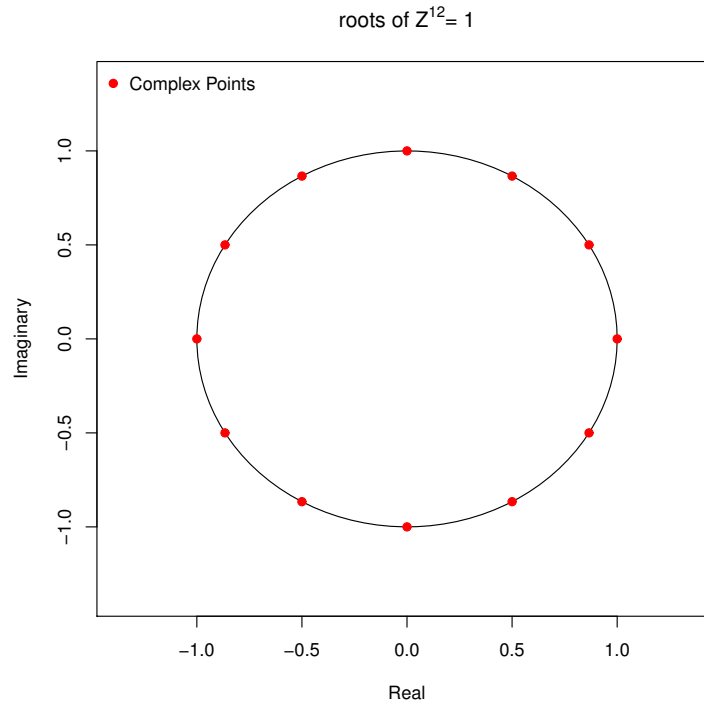


Figura 10.4: Raíces de la unidad, $z^{12} = 1$

¹<https://codereview.stackexchange.com/questions/132959/roots-of-unity-in-r>

No rechazar alguna de las hipótesis nulas equivale a aceptar que existe una raíz unitaria estacional.

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 modelo de tendencia lineal más componentes estacionales con residuos posiblemente SARMA.

Hay que tener en cuenta la potencia de la prueba: $\text{Prob}(\text{rechazar } H_0 \mid H_1 \text{ es cierta})$, 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.

Implementación en R

La prueba HEGY está implementada en la función `hegy.test`, de la librería `uroot`, ver López-de LaCalle [2017]. Ejemplo:

```
hegy.test(x, deterministic = c(1, 0, 0),
lag.method = "fixed", maxlag = 1,
pvalue = "raw")
```

- la serie debe estar en la variable x , y debe haberse declarado como objeto “ts”.
- El parámetro `deterministic=c(1, 1, 1)` se refiere a si se desea incluir en el modelo un intercepto, tendencia lineal o componente estacional con variables indicadoras. El modelo se define en este caso como

$$\Phi_s(L)[Y_t - (\beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t))] = \varepsilon_t. \quad (10.17)$$

En caso contrario se coloca `deterministic=c(1, 0, 0)`

- El parámetro `pvalue`, según la ayuda:

a character specifying the method employed to compute

p-values: "RS", the default, interpolation based on response surface regressions; "bootstrap", bootstrap; "raw" interpolation in the tables provided in the reference papers.

Ejemplo 10.4.1. *Retomando el Ejemplo (10.3.1) en la pag. 255, para modelación de la serie de producción de cemento Portland trimestral, donde se aplicó la prueba Canova-Hansen, se va a aplicar ahora la prueba HEGY con el mismo objetivo de determinar si el modelo SARIMA es adecuado. Los comandos en R son los siguientes.*

```
# prueba hegy con intercepto, tendencia y variables indicadoras.
require(uroot)
hegy.out1 = hegy.test(x=y,
deterministic = c(1,1,1),
lag.method = "fixed", maxlag = 1)
#----resultado
statistic p-value
t_1      -4.153  0.0046 **
t_2      -3.1308 0.0206 *
F_3:4    14.9083      0 ***
F_2:4    12.533      0 ***
F_1:4    15.6967      0 ***
Deterministic terms: constant + trend + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions
```

El resultado anterior muestra que se rechaza la hipótesis nula de raíces unitarias en $1, -1, \pm i$, indicadas por t_{pi_1} , t_{pi_2} , $F_{pi_3:4}$. La última corresponde a las raíces en las frecuencias $\pi/2, 3\pi/2$, la primera corresponde a la frecuencia cero, y es una raíz unitaria para la tendencia, similar a la prueba Dickey-Fuller. Luego, no sería apropiado ajustar un modelo SARIMA a la serie.

Si se repite la prueba con la serie sin tendencia, rechaza nuevamente la hipótesis de raíces unitarias estacionales.

```
# prueba hegy sin tendencia.
```

```

require(uroot)
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
t1 = m1$time.series[,2]
y1 = y-t1
hegy.out1 = hegy.test(x=y1,
deterministic = c(0,0,1),
lag.method = "fixed", maxlag = 1)
#-----resultado
      statistic p-value
t_1      -5.4194      0 ***
t_2      -3.064    0.0257 *
F_3:4     14.9748      0 ***
F_2:4     12.475      0 ***
F_1:4     19.6446      0 ***
Deterministic terms: constant + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions

```

10.5. Prueba OCSB

La prueba Osborn, Chui, Smith, Birchenhall para raíz unitaria estacional, ver Osborn et al. [1988] y Ghysels and Osborn [2001] pag. 66. La prueba se basa en la regresión

$$\Delta \Delta_s Y_t = \beta_1 \Delta_s Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t, \quad (10.18)$$

con $\Delta = I - L$, $\Delta_s = I - L^s$. Utilizando la identidad

$$I - L^s = (I - L)(I + L + \dots + L^{s-1}),$$

aplicando este operador a Y_{t-1} se tiene

$$\Delta_s Y_{t-1} = (I + L + \dots + L^{s-1}) \Delta Y_{t-1}$$

luego en (10.18)

$$\begin{aligned} \Delta_s \Delta Y_t &= \beta_1 \Delta_s Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t \\ \Leftrightarrow \Delta_s \Delta Y_t &= \beta_1 (I + L + \dots + L^{s-1}) \Delta Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t \end{aligned}$$

y todas las variables en la regresión (10.18) incluyen la primera diferencia Δ . Por tanto, como afirman Ghysels and Osborn [2001] pag. 66:

An interpretation of the OCSB test, therefore, is that it tests seasonal integration in a process that has been first differenced.

En Rodrigues and Osborn [1999] se propone la regresión sin la primera diferencia Δ

$$\Delta_s Y_t = \beta_1(I + L + \dots + L^{s-1})Y_{t-1} + \beta_2 Y_{t-s} + \epsilon_t. \quad (10.19)$$

como modelo para la prueba de hipótesis nula de raíz unitaria estacional, incluyendo la raíz unitaria ordinaria. La expresión de la hipótesis nula y la alterna en Osborn et al. [1988], pag. 364 es

$$H_0 : Y_t \sim I(1, 1) \quad (10.20)$$

$$H_1 : Y_t \sim I(0, 0) \vee I(1, 0) \vee I(0, 1) \quad (10.21)$$

donde el símbolo $I(1, 1)$ significa integrado ordinario y estacional, y los otros $I(1, 0)$, $I(0, 1)$, $I(0, 0)$ se interpretan de manera similar. El símbolo \vee es la “o” lógica.

El estadístico de la prueba se define como el estadístico de la prueba ADF, $t_{\hat{\beta}_2}$, con respecto a la regresión (10.19). La hipótesis equivale en este caso a $H_0 : \beta_2 = 0$, vs. $H_1 : \beta_2 < 0$. Se rechaza la nula cuando $t_{\hat{\beta}_2} < t_\alpha$.

Además, el estadístico de la prueba ADF, $t_{\hat{\beta}_1}$ se puede utilizar para la hipótesis de raíz unitaria ordinaria $H_0 : Y_t \sim I(1)$. La hipótesis equivale en este caso a $H_0 : \beta_1 = 0$, vs. $H_1 : \beta_1 < 0$.

La distribución asintótica de estos estadísticos bajo H_0 no se incluyen aquí.

Pero ambas pruebas no pueden equipararse a la prueba Dickey-Fuller, porque, como Ghysels and Osborn [2001] pag. 66 observan, los valores críticos ADF no son estrictamente válidos para éstas, en consecuencia para la prueba OCSB se requieren valores críticos calculados por simulación.

La prueba OCSB está implementada en la librería `forecast`, con la función `ocsb.test`. Los valores críticos en esta implementación se calculan mediante simulación. Esta función calcula solamente $t_{\hat{\beta}_2}$ y el valor crítico para el nivel de significación de 5 %.

Ejemplo 10.5.1. *Retomando los Ejemplos (10.3.1) y (10.4.1) en las pags. 255 y 261, para la modelación de la serie de producción de cemento Portland trimestral, donde se aplicaron las pruebas Canova-Hansen y HEGY. Se va a aplicar ahora la prueba OCSB con el mismo objetivo de determinar si el modelo SARIMA es adecuado.*

Los comandos en R son los siguientes.

```
#-----serie con tendencia y estacionalidad
require(forecast)
ocsb.test(y)
Test statistic: -2.092, 5% critical value: -1.8927

#-----decision: -2.092 < -1.8927, rechaza Ho
alternative hypothesis: stationary
#-----eliminar tendencia
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
s1 = m1$time.series[,1]
t1 = m1$time.series[,2]
e1 = m1$time.series[,3]
y1 = y-t1
ocsb.test(y1)
Test statistic: -2.1177, 5% critical value: -1.8927
```

En el primer caso el estadístico observado es menor que el valor crítico por lo que rechaza la nula $H_0 : Y_t \sim I(1, 1)$. Y por tanto acepta una de las alternas $H_1 : Y_t \sim I(0, 0) \vee I(1, 0) \vee I(0, 1)$. Además, como se eliminó la tendencia previamente, esto excluye la alternativa $I(1, 0)$, y como se rechaza la nula que contiene la integración estacional, la alterna debe ser $Y_t \sim I(0, 0)$. Es decir, el modelo SARIMA no es adecuado y se debería utilizar el de Componentes.

10.6. Ejemplo de resumen

Ejemplo 10.6.1. *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral, para decidir entre los*

modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA.

$$1) Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t, \varepsilon_t \sim MA(6), \quad (10.22)$$

$$2) Y_t = SARIMA(3, 1, 2)(1, 1, 2)[4]. \quad (10.23)$$

En la Figura 10.5 se puede ver la serie con y sin tendencia. Se observa que la componente estacional no es constante, por lo que es posible que se tenga un modelo SARIMA (10.23) en lugar del modelo de componentes (10.22). Al examinar la fac de

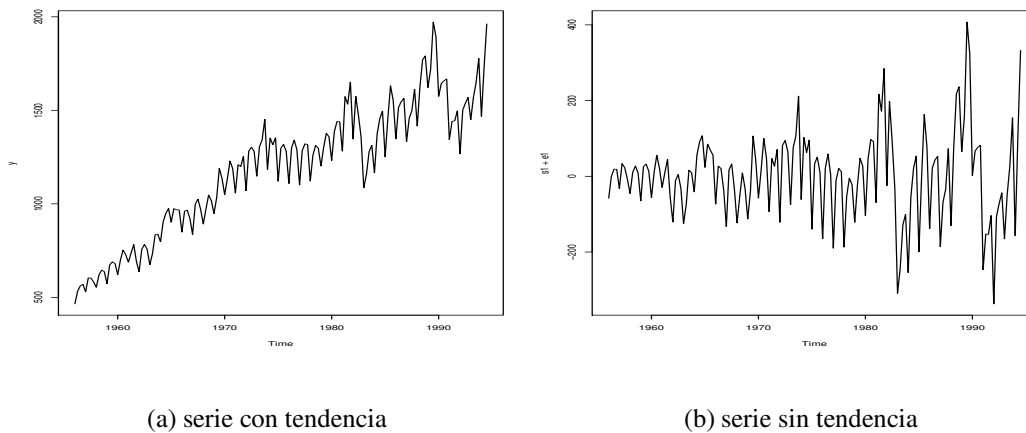


Figura 10.5: Serie de Producción Trimestral de Cemento, Australia

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 10.6.

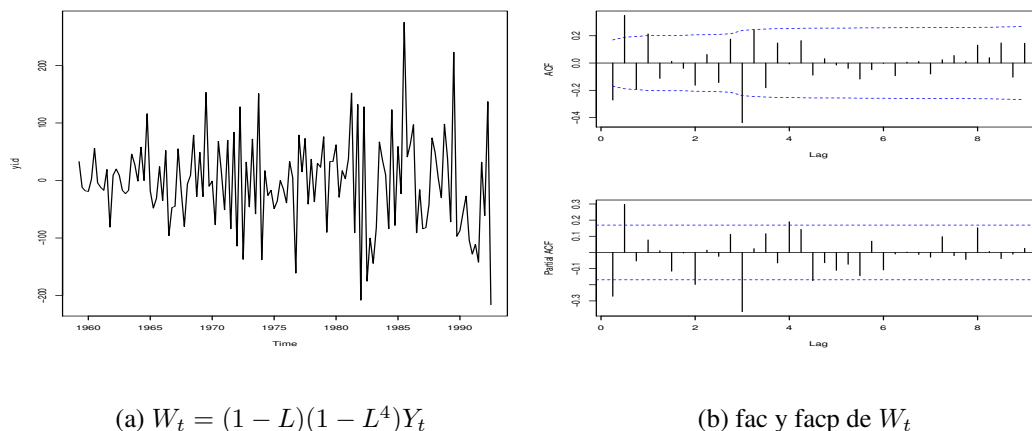


Figura 10.6: Serie Diferenciada de Producción Trimestral de Cemento

Mediante el uso de la función `armasubSets()` de la librería TSA, aplicada a la serie W_t se llega al modelo $SARIMA(3, 1, 2)(1, 1, 2)[4]$. Los parámetros estimados están en la Tabla 10.1.

Tabla 10.1: Parámetros estimados de $SARIMA(3,1,2)(1,1,2)[4]$

	parametros	sd.dev	est t
ar1	0.55	0.30	1.85
ar2	0.57	0.20	2.82
ar3	-0.31	0.12	-2.60
ma1	-0.73	0.32	-2.28
ma2	-0.23	0.31	-0.73
sar1	0.90	0.25	3.56
sma1	-1.56	0.32	-4.83
sma2	0.67	0.18	3.80

La prueba Ljung-Box para los residuos de este modelo arroja el resultado $X\text{-squared} = 34.672$, $df = 26$, $p\text{-value} = 0.1189$, por lo que puede aceptarse este modelo para pronosticar. Los pronósticos con el modelo SARIMA versus los obtenidos con el modelo componentes (10.22) se pueden

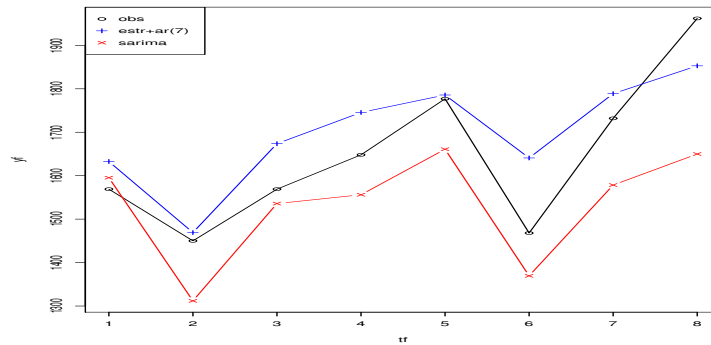


Figura 10.7: Pronósticos a 8 trimestres de la Producción Trimestral de Cemento

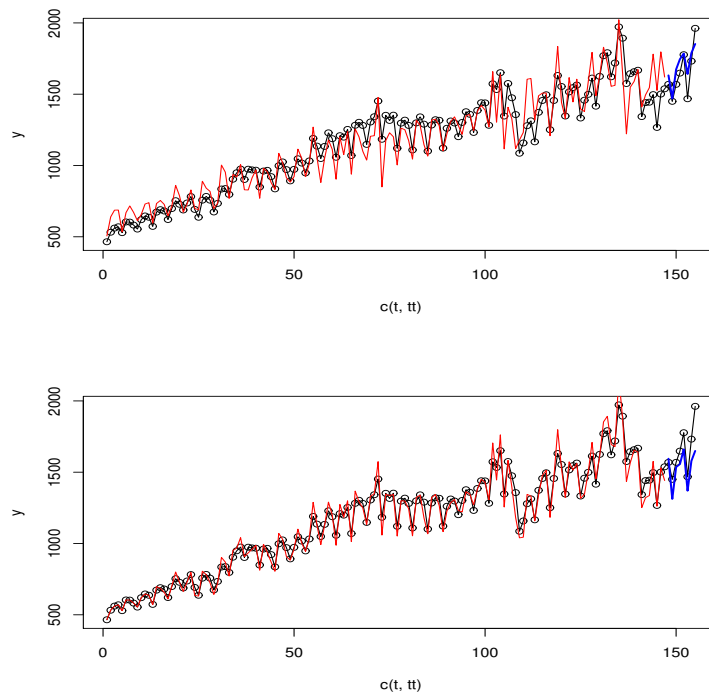


Figura 10.8: Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA

observar en la Figura 10.7. Se llegó a que el SARIMA no se detecta como modelo factible por las prueba CH, HEGY y OCSB. El MAPE de sus pronósticos, a 8 trimestres, es 7.89, mayor que el del modelo de Descomposición con errores MA(6),

que es 4.64. Sin embargo, el modelo SARIMA tiene mejor ajuste dentro de la muestra. En la Figura 10.8 se puede comparar tanto los ajustes dentro de la muestra como los pronósticos.

Cuál modelo escoger?. Desde el punto de vista de la validación cruzada a 8 trimestres, el modelo de Descomposición. Desde el punto de vista del ajuste dentro de la muestra, el modelo SARIMA.

CAPÍTULO 11

Modelos de Rezagos Distribuidos y Cointegración

11.1. Modelo de rezagos distribuidos

El objetivo de este capítulo es desarrollar la idea de la regresión lineal entre dos ó más procesos estacionarios. Si se consideran $(X_n, Y_n, Z_n, n \in \mathbb{Z})$ tres procesos estacionarios en covarianza, de media cero, y $(\beta_j, j = 0, 1, \dots) \in \ell_1$ es una sucesión de reales con $\beta_0 = 1$ entonces se puede definir el proceso estacionario en covarianza $D_n = \sum_{j=0}^{\infty} \beta_j X_{n-j}$, y definir un modelo de regresión lineal de la forma

$$Y_n = \alpha_0 + \alpha_1 D_n + Z_n, \quad (11.1)$$

para α_0, α_1 parámetros. Reemplazando D_n en función de $(X_n, n \in \mathbb{Z})$ se obtiene

$$Y_n = \alpha_0 + \alpha_1 \sum_{j=0}^{\infty} \beta_j X_{n-j} + Z_n, \quad (11.2)$$

donde los (β_j) son también parámetros. Si se tienen muestras finitas de los procesos (Y_n, X_n) entonces no es posible estimar el conjunto infinito de parámetros en (11.2) por lo que se hace necesario introducir algunas restricciones. Una de éstas consiste

en asumir que la función de transferencia $\beta(z) = \sum_{j=0}^{\infty} \beta_j z^j$ se puede reemplazar (aproximar) por un filtro causal tipo ARMA(p,q) de la forma

$$\beta(z) = \frac{\Theta_q(z)}{\Phi_p(z)} = \frac{1 + \sum_{j=1}^q \theta_j z^j}{1 - \sum_{j=1}^p \varphi_j z^j}, \quad (11.3)$$

para $|z| < 1$, donde $\Phi_p(z) \neq 0$, $\Theta_q(z) \neq 0 \forall z, |z| \leq 1$. También en (11.2) se asume que $(Z_n) \sim ARMA(p_z, q_z)$ y $(X_n) \sim ARMA(p_x, q_x)$, incluyendo la posibilidad de S-ARMA estacional, son independientes. Entonces (11.2) se escribe

$$Y_n = \alpha_0 + \alpha_1 \frac{\Theta_q(L)}{\Phi_p(L)} X_n + Z_n. \quad (11.4)$$

Si se redefinen los parámetros θ_j de tal forma que $\alpha_1(1 + \sum_{j=1}^q \theta_j) \equiv \sum_{j=0}^q \theta_j$ y se reemplaza X_n por X_{n-b} , donde $b = 0, 1, \dots$, entonces (11.4) queda

$$Y_n = \alpha_0 + \frac{\Theta_q(L)}{\Phi_p(L)} X_{n-b} + Z_n. \quad (11.5)$$

El modelo (11.5) se define como “modelo de función racional con rezagos distribuidos, con intercepto”. El rezago X_{n-b} indica la demora que tarda en surtir efecto un cambio de las X_n en las Y_n .

El principal problema del modelo (11.5) es la identificación de los órdenes p, q, p_z, q_z . Dos posibilidades consisten en: 1) el método de Box-Jenkins con base en preblanqueado y 2) búsqueda directa ensayando varios valores de p, q, p_z, q_z y escogiendo el modelo de mínimo AIC. Esta última requiere un método de estimación. Los métodos de estimación se presentan más adelante.

Método de identificación de Box-Jenkins

El método se describe en cuatro pasos. A partir del modelo (11.5) se coloca $\alpha_0 = 0$ y se re-definen los coeficientes β_j mediante la identidad

$$\frac{\Theta_q(z)z^b}{\Phi_p(z)} = \sum_{j=0}^{\infty} \beta_j z^j = \beta(z). \quad (11.6)$$

Los pasos son los siguientes; más adelante se dan las correspondientes instrucciones en R.

Paso 1. Se estima un modelo para $(X_n) \sim ARMA(p_x, q_x)$ tal que

$$X_n = \frac{\Theta_{q_x}(L)}{\Phi_{p_x}(L)} U_n, \quad (11.7)$$

donde $(U_n) \sim RB(0, \sigma_u^2)$. Se calcula el filtro causal inverso

$$\nu(L) = \frac{\Phi_{p_x}(L)}{\Theta_{q_x}(L)} = 1 + \nu_1 L + \nu_2 L^2 + \dots \quad (11.8)$$

Paso 2. Se aplica el filtro inverso (11.8) a izquierda en la ecuación (11.5). Se calculan $W_n = \nu(L)Y_n$ y $U_n = \nu(L)X_n$. Definiendo $V_n = \nu(L)Z_n$ se obtiene

$$W_n = \beta(L)U_n + V_n. \quad (11.9)$$

En esta ecuación (V_n) y (U_n) son procesos independientes por hipótesis, ya que (Z_n) y (X_n) lo son; y son observables (calculables), en cambio V_n no lo es, pero esto no causa inconveniente debido a que se elimina al tomar esperanzas en el paso siguiente.

Paso 3. Se multiplica a derecha ambos miembros de la ecuación (11.9) por U_{n-k} , para $k = 0, 1, 2, \dots$, y luego se toma esperanza. En el miembro de la izquierda se obtiene la covarianza cruzada $R_{w,u}(k) = Cov(W_n, U_{n+k}) = \mathbb{E}(W_n U_{n+k})$. Se tiene además

$$\begin{aligned} R_{w,u}(k) &= \mathbb{E}(\beta(L)U_n U_{n-k}) + \mathbb{E}(V_n U_{n-k}) \\ &= \sum_{j=0}^{\infty} \beta_j \mathbb{E}(U_n U_{n-k}) \\ &= \beta_k \sigma_u^2. \end{aligned}$$

Se estiman los β_k mediante la relación

$$\hat{\beta}_k = \frac{\hat{R}_{w,u}(k)}{\hat{\sigma}_u^2} = \frac{\hat{\rho}_{w,u}(k)\hat{\sigma}_w}{\hat{\sigma}_u}. \quad (11.10)$$

Paso 4. Según (11.6), a cada par de órdenes (p, q) y polinomios $\Phi_p(z)$, $\Theta_q(z)$, le corresponde una sucesión (β_k) . E inversamente, dados los $(\hat{\beta}_k)$ les debe corresponder ciertos (p, q) y polinomios $\Phi_p(z)$, $\Theta_q(z)$. En la Figura 11.1 a la derecha de cada patrón (β_k) se muestra el correspondiente modelo. Comparando (β_k) con $(\hat{\beta}_k)$, se elige el (p, q) del caso con mejor correspondencia, y se puede iniciar el proceso de estimación.

Implementación en R

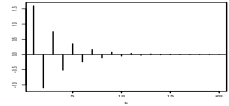
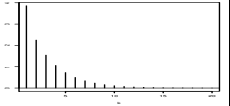
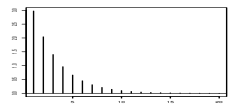
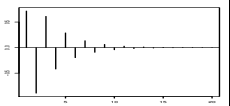
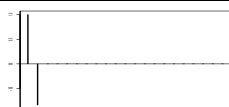
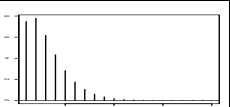

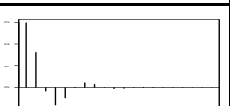
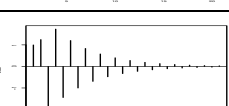
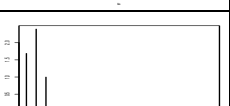
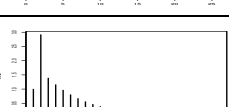
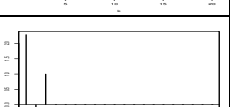
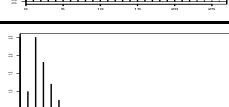
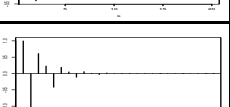
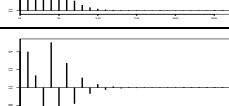
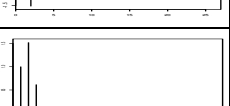
(p, q)	$\beta_j, j = 0, 1, \dots$	Modelo	(p, q)	$\beta_j, j = 0, 1, \dots$	Modelo
(1,0)		$\frac{\theta_0}{1-\varphi_1 L}$	(1,1)		$\frac{\theta_0+\theta_1 L}{1-\varphi_1 L}$
(1,0)		$\frac{\theta_0}{1-\varphi_1 L}$	(1,1)		$\frac{\theta_0+\theta_1 L}{1-\varphi_1 L}$
(0,1)		$\frac{\theta_0+\theta_1 L}{1}$	(2,0)		$\frac{\theta_0}{1+\varphi_1 L+\varphi_2 L^2}$
(0,1)		$\frac{\theta_0+\theta_1 L}{1}$	(2,0)		$\frac{\theta_0}{1-\varphi_1 L-\varphi_2 L^2}$
(1,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L}$	(0,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1}$
(1,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L}$	(0,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1}$
(2,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L-\varphi_2 L^2}$	(2,1)		$\frac{\theta_0+\theta_1 L}{1-\varphi_1 L-\varphi_2 L^2}$
(2,2)		$\frac{\theta_0+\theta_1 L+\theta_2 L^2}{1-\varphi_1 L-\varphi_2 L^2}$	(2,1)		$\frac{\theta_0+\theta_1 L}{1-\varphi_1 L-\varphi_2 L^2}$

Tabla 11.1: Modelos de Transferencia

Para implementar el método de identificación Box-Jenkins en R es conveniente primero explicar las funciones: `arima()`, `arima.sim()`, `poly.calc()`, `ARMAtoMA()` e `ImpulseVMA`.

La función `arima`. Estima modelos SARIMA. La especificación del modelo a estimar depende de cómo R representa los procesos ARMA(p,q), que es como sigue.

$$Y_n = \varphi_0 + \sum_{j=1}^p \varphi_j Y_{n-j} + Z_n + \sum_{j=1}^q \theta_j Z_{n-j}. \quad (11.11)$$

Al aplicar `m1 = arima(y, order=c(p, 0, q))` el objeto `m1` contiene un nodo (slot) denominado `coef` tal que `m1$coef` es el vector $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \varphi_0)$.

La función poly.calc. Para definir un modelo ARMA estacionario en covarianza e invertible se puede comenzar definiendo las raíces de los polinomios autorregresivo, $\Phi_p(z)$ y de media móvil, $\Theta_q(z)$, de tal forma que se garantice que están por fuera del círculo unitario.

Si el vector r contiene p raíces, la función $a = \text{poly.calc}(r)$ de la librería `polynom`, calcula los coeficientes de un polinomio a , con las raíces r . El objeto a es un vector de coeficientes correspondientes a las potencias de z en orden ascendente. Por ejemplo, $-2.796 + 0.41*z + 0.4*z^2 + z^3$. Sin embargo, los polinomios $\Phi_p(z)$ y $\Theta_q(z)$ deben ser mónicos, es decir, el término constante debe ser la unidad. Si se normaliza escribiendo $a = a/a[1]$, el resultado es un polinomio mónico con raíces r , por ejemplo, $1 - 0.146*z - 0.143*z^2 - 0.357*z^3$. Si este polinomio fuera $\Phi_3(z) = 1 - \sum_{j=1}^3 \varphi_j z^j$, el vector de los coeficientes φ_j se obtienen como $-a[2:4]$. Si el polinomio fuera $\Theta_q(z) = 1 + \sum_{j=1}^q \theta_j z^j$, el vector de coeficientes θ_j se obtiene como $a[2:4]$; el detalle es el signo menos en los coeficientes autorregresivos.

La función arima.sim. Si los polinomios $a = \text{poly.calc}(ra)$, $b = \text{poly.calc}(rb)$ son los polinomios $\Phi_p(z)$ y $\Theta_q(z)$, respectivamente, se pueden simular n valores de un proceso ARMA utilizando `arima.sim()`,
 $X_n = \text{arima.sim}(\text{list}(\text{order} = c(p, 0, q), \text{ar} = -a[2:p], \text{ma} = b[2:q]), n = n, \text{sd} = \text{sigma})$.

Las funciones ARMAtoMA, ImpulseVMA. Dado el proceso $(X_n) \sim \text{ARMA}(p_x, q_x)$ en (11.7) para obtener su representación causal, dada por la sucesión $(\psi_j, j = 0, 1, 2, \dots)$,

$$\frac{\Theta_{q_x}(z)}{\Phi_{p_x}(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \quad (11.12)$$

con $\psi_0 = 1$, se utilizan las funciones `ARMAtoMA` de la librería `stats` e `ImpulseVMA` de la librería `portes`. En adelante solamente se utilizará esta última. Si los vectores `phi`, `theta` corresponden a $(\varphi_1, \dots, \varphi_p)$ y $(\theta_1, \dots, \theta_1)$, respectivamente, se escribe `ImpulseVMA(phi, theta, m)`, y el parámetro `m` es el número de términos en el desarrollo (11.12).

Si se utiliza la función `arima()` para estimar un modelo ARMA, con $m1 = \text{arima}(x, \text{order} = c(p, 0, q))$, se calcula el filtro causal (11.12), de longitud m $\underline{\psi} = (1, \psi_1, \dots, \psi_m)$, con la instrucción

```
psi = ImpulseVMA(phi=-m1.coef[(p+1):(p+q)],
theta = m1.coef[1:p], Trunc.Series = m) .
```

Si se utiliza la función `poly.cal(r)`, donde $a = \text{poly.cal}(ra)$, $b = \text{poly.cal}(rb)$ son los polinomios $\Phi_p(z)$ y $\Theta_q(z)$, respectivamente, entonces se escribe

```
psi=ImpulseVMA(phi=-a[2:p],theta=b[2:q], Trunc.Series = m)
```

La función filter. Un paso importante en la implementación de la metodología Box-Jenkins es calcular el filtro causal inverso dado en (11.8)

$$\nu(L) = \frac{\Phi_{p_x}(L)}{\Theta_{q_x}(L)} = 1 + \nu_1 L + \nu_2 L^2 + \dots$$

con la instrucción

```
nu=ImpulseVMA(phi=m1.coef[(p+1):(p+q)],
ma=m1.coef[1:p], lag.max=m) .
```

El filtro $\nu(L)$ se aplica a X_n y es inmediato que $\nu(L)X_n = U_n$. Esta operación se hace mediante la función `filter()` de la librería `stats`. Se escribe `u=filter(x, rev(nu), "conv", sides=1, circular=TRUE)`. Un detalle importante es el uso de `rev(nu)`, que especifica se invierte el orden de la sucesión `nu`.

Los Pasos para la metodología Box-Jenkins en R

Paso 1. Para implementarlo en R se procede a estimar el modelo para $(X_n) \sim ARMA(p_x, q_x)$, luego se calculan los primeros valores del filtro $\nu(L)$ con la función `ARMAtoMA()`.

```
mx = arima(x, order=c(px, qx))
nu = ARMAtoMA(ar = mx$coef[(px+1):(px+qx)],
ma = mx$coef[1:px], lag.max = 30)
```

La operación de “blanqueo” consiste en aplicar el filtro inverso $\nu(L)$ a ambos procesos X_n y Y_n . El filtro (??) se define a partir de los polinomios $\Phi_p(z)$ y $\Theta_q(z)$ del modelo para X_n estimado con `arima()`. **Paso 2.** Definiendo $W_n = \nu(L)Y_n$, $U_n = \nu(L)X_n$ y $V_n = \nu(L)Z_n$ se obtiene

En esta ecuación (V_n) y (U_n) son procesos independientes por hipótesis, ya que (Z_n) y (X_n) lo son. Además $(U_n) \sim RB(0, \sigma_u^2)$. Para calcular los procesos W_n, U_n en R se utiliza la función `filter()` de la librería `stats`

```
w = filter(y, filter = nu, "conv", sides = 1, circular = TRUE)
u = filter(x, filter = nu, "conv", sides = 1, circular = TRUE)
```

Paso 3. La instrucción con R consiste en calcular solamente $\hat{\rho}_{w,u}(k)$ y graficarla

```
acf(ts.intersect(w,u), na.action = na.pass)
```

11.2. *Cointegración

APÉNDICE A

Datos de Series de Tiempo

A.1. Series con Tendencia y Estacionalidad

A.1.1. Ejemplo 1. Cinco Series

Para cargar los datos en R simplemente copiar cada bloque “structure” en el programa. Las series tiene componente estacional y/o tendencia.

```
"X1" <-  
structure(c(13630, 14109, 15945, 16637, 17136, 17701, 18894,  
21479, 18297, 23610, 18546, 14781, 13975, 13468, 13490, 17759,  
17889, 17298, 19260, 20123, 18777, 23709, 16981, 17134, 14197,  
14749, 16270, 18028, 17939, 19218, 19683, 19711, 20396, 20659,  
15398, 18549, 15441, 14844, 16610, 17958, 17945, 20473, 19482,  
21151, 20952, 21222, 19489, 17968, 15818, 14423, 18010, 18544,  
20123, 21131, 19371, 23507, 22605, 21042, 21547, 17417, 19043,  
16207, 17941, 18075, 19914, 21450, 18702, 22845, 21272, 21430,  
22094, 15988, 17670, 16825, 18226, 19497, 21406, 20008, 22962,
```

```
23116, 20327, 25791, 19850, 18185, 19292, 16934, 18314, 21890,
21460, 22641, 23237, 23495, 23948, 27086, 19827, 19827),
.Tsp = c(1990, 1997.916666666667, 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.
- BLIEMEL, F. (1973): “Theil’s forecast accuracy coefficient: A clarification,” .
- BOX, G. AND G. C. TIAO (1975): “Intervention Analysis with Applications of Economic and Environmental Problems,” *JASA*, 70, 70–79.
- BOX, G. E. P., G. M. JENKINS, AND G. C. REINSEL (1976): *Time Series Analysis, Forecasting and Control*, New York: Holden-Day, third edition ed.
- BROCKWELL, P. J. AND R. A. DAVIS (1987): *Time Series: Theory and Method*, New York: Springer-Verlag, Inc., 1nd ed.
- (2002): *Introduction to Time Series and Forecasting*, New York: Springer-Verlag, Inc., 2nd ed.
- BROWN, R. L., J. DURBIN, AND J. M. EVANS (1975): “Techniques for testing the constancy of regression relationships over time,” *Journal of the Royal Statistical Society, B*, 37, 149–163.
- CANOVA, F. AND B. HANSEN (1995): “Are seasonal patterns constant over time?. A Test for seasonal stability,” *Journal of Business and Economics Statistics*, 13, 237–252.
- CHU, C.-S. J., K. HORNIK, AND C.-M. KAUN (1995): “MOSUM tests for parameter constancy,” *Biometrika*, 82, 603–617.
- CLEVELAND, R. B., W. S. CLEVELAND, J. E. MCRAE, AND I. TERPPENING (1990): “STL: A Seasonal Trend Decomposition Procedure Based on Loess,” *Journal of Official Statistics*, 6, 3–73.
- COWPERTWAIT, P. AND A. METCALFE (2009): *Introductory Time Series Analysis with R*, New York: Springer Science + Business Media LLC, third edition ed.

- CRYER, J. D. AND K.-S. CHAN (2008): *Time Series Analysis with Applications in R*, New York: Springer Science + Business Media LLC.
- DIEBOLD, F. (1999): *Elementos de Pronóstico*, México: International Thomson Editores.
- DIEBOLD, F. X. (2018): “No Hesitations,” <https://fxdiebold.blogspot.com/2017/03/machine-learning-and-econometrics-vi.html>.
- ELDER, J. AND P. E. KENNEDY (2001): “Testing for unit roots: what should student be taught?” *Journal of Economic Education*, 32, 137–146.
- FULLER, W. A. (2009): *Introduction to statistical time series*, vol. 428, John Wiley & Sons.
- GHYSELS, E. AND D. R. OSBORN (2001): *The econometric analysis of seasonal time series*, Cambridge University Press.
- GRANDELL, J. (2014): “Time series analysis,” Lecture Notes, SF2943 Time Series Analysis.
- HARVEY, A. (2006): “Forecasting with unobserved components time series models,” in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Amsterdam, The Netherlands: Elsevier North-Holland, chap. 7, 330–358.
- HARVEY, A. C. AND J. DURBIN (1986): “The Effect of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modeling,” *Journal of the Royal Statistical Society, A*, 149, 187–227.
- HAYWOOD, J. AND J. RANDAL (2006): “Seasonality and structural breaks: NZ visitor arrivals and 9/11,” Tech. rep., MSCS Research Report 06.
- HELSKE, J. (2017): “KFAS: Exponential Family State Space Models in R,” *Journal of Statistical Software*, 78, 1–39.
- HYLLEBERG, S., R. ENGLE, C. GRANGER, AND B. YOO (1990): “Seasonal integration and cointegration,” *Journal of Econometrics*, 44, 215–238.

- HYNDMAN, R. J. AND A. B. KOEHLER (2006): “Another look at measures of forecast accuracy,” *International journal of forecasting*, 22, 679–688.
- HYNDMAN, R. J., A. B. KOEHLER, J. K. ORD, AND R. D. SNYDER (2008): *Forecasting with Exponential Smoothing. The State Space Approach*, Berlín, Heidelberg: Springer Science + Business Media LLC.
- HYNDMAN, R. J., A. B. KOEHLER, R. D. SNYDER, AND S. GROSE (2002): “A state space framework for automatic forecasting using exponential smoothing methods,” *International Journal of forecasting*, 18, 439–454.
- KLEINBAUM, D., L. KUPPER, AND K. MULLER (1988): *Applied regression analysis and other multivariable methods*, PWS-KENT Publishers Co.
- KWIATKOWSKI, D., P. C. PHILLIPS, P. SCHMIDT, Y. SHIN, ET AL. (1992): “Testing the null hypothesis of stationarity against the alternative of a unit root,” *Journal of econometrics*, 54, 159–178.
- LEE, T.-H., H. WHITE, AND C. W. GRANGER (1993): “Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests,” *Journal of Econometrics*, 56, 269–290.
- LÓPEZ-DE LACALLE, J. (2017): *uroot: Unit Root Tests for Seasonal Time Series*, r package version 2.0-9.
- NOWICKA-ZAGRAJEK, J. AND R. WERON (2002): “Modeling electricity loads in California: ARMA models with hyperbolic noise,” *Signal Processing*, 82, 1903–1915.
- OSBORN, D. R., A. P. CHUI, J. P. SMITH, AND C. R. BIRCHENHALL (1988): “Seasonality and the order of integration for consumption,” *Oxford Bulletin of Economics and Statistics*, 50, 361–377.
- PHILLIPS, P. C. AND P. PERRON (1988): “Testing for a unit root in time series regression,” *Biometrika*, 75, 335–346.
- POVEDA, G. (1982): “Curva Logística y Mínimos Cuadrados,” *Lecturas Matemáticas*, 3, 111–132.

- PRADO, R. AND M. WEST (2010): *Time series: modeling, computation, and inference*, CRC Press.
- PRIESTLEY, M. B. (1988): *Non-linear and Non-stationary Time Serie Analysis*, London, UK: Academic Press Limited.
- QIU, D. (2015): *rmaf: Refined Moving Average Filter*, r package version 3.0.1.
- RODRIGUES, P. M. AND D. R. OSBORN (1999): “Performance of seasonal unit root tests for monthly data,” *Journal of Applied Statistics*, 26, 985–1004.
- RYAN, J. A. AND J. M. ULRICH (2017): *xts: eXtensible Time Series*, r package version 0.10-0.
- SHEATHER, S. (2009): *A Modern Approach to Regression with R*, Springer Texts in Statistics, Springer New York.
- SHUMWAY, R. H. AND D. S. STOFFER (2005): *Time Series Analysis and Its Applications. With R examples*, New York: Springer Science + Business Media LLC.
- SMITH, A. K. (2019): *CombMSC: Combined Model Selection Criteria*, r package version 1.4.2.1.
- 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.
- TERÄSVIRTA, T., C.-F. LIN, AND C. W. GRANGER (1993): “Power of the neural network linearity test,” *Journal of time series analysis*, 14, 209–220.
- VENABLES, W. AND B. RIPLEY (2002): *Modern Applied Statistics with S*, Statistics and Computing, Springer.
- VINOD, H. D. (1973): “Generalization of the Durbin-Watson Statistic for Higher Order Autoregressive Processes,” *Communications in Statistics*, 2, 115–144.
- WINTERS, P. (1960): “Forecasting sales by exponentially weighted moving averages,” *Management Science*, 6, 324–342.

- ZEILEIS, A., K. HORNIK, C. KLEIBER, AND W. KRAEMER (2003): “Testing and Dating of Structural Changes in Practice,” *Computational Statistics & Data Analysis*, 44, 109–123.
- ZEILEIS, A., F. LEISCH, K. HORNIK, AND C. KLEIBER (2002): “strucchange: An R Package for Testing for Structural Change in Linear Regression Models,” *Journal of Statistical Software*, 7, 232–274.
- ZHANG, G. P. (2003): “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, 50, 159–175.

Índice alfabético

- AR*(1), 147
- CUSUM*, 82
- R^2 , 42
- AIC, 43
- Algoritmo de Brockwell y Davis, 127
- Bandas de Bartlett, 139
- bIC, 43
- Componente determinístico, 134
- Componente Estacional, 50, 63
- Componente Estructural, 68
- Consistencia, 44
- Criterio de Información
 - de Akaike, 43
 - de Schwarz, 43
- Datos de Series de Tiempo, 277
- Desarrollo
 - de Volterra, 237
 - de Wold, 188
- Ecuación
 - Yule-Walker, 175
- Ecuación Yule–Walker, 175
- Eficiencia Asintótica, 44
- Error cuadrático medio, 32
- Estabilidad Estructural, 78
- Estacionario en Covarianza, 133
- Estadístico
 - Ljung-Box, 143
- Estadístico
 - Durbin-Watson, 149
- Estimación
 - de la Autocorrelación, 136
 - de la Autocovarianza, 136
- Estimador de la Varianza del Error, 32
- Estimadores Recursivos, 79
- EWMA, 120
- fac, 136
- Filtro
 - causal, 120
- Función
 - de Autocovarianza, 133
- Función
 - de Autocorrelación, 136
 - de Autocorrelación Parcial, 166
- función de R
 - dwtext, 149

- Hyndman, R., 198
- librería
- dynlm, 155
 - lmtest, 149
- Método de Descomposición STL, 96
- Método de Suavizamiento
- Exponencial Simple, 120
- MAPE: Error absoluto porcentual promedio, 57
- Marcha Aleatoria, 213
- Matriz hat, 31
- Matriz Toeplitz, 193
- Media Móvil, 119
- Media Móvil
- Bilateral, 120
 - Unilateral, 120
- Medidas de precisión de pronósticos
- accuracy, 57
 - MAPE, 57
 - RMSE, 57
 - U Theil, 57
- Modelo
- $AR(1)$, 147
 - aditivo de componentes, 49
 - ARIMA, 213
 - Espacio de Estados, 102
 - híbrido ARIMA - Red Neuronal, 239
 - log-Lineal, 54
- Modelo de Suavizamiento
- de Holt-Winters, 99
- Modelo tendencia
- cúbico, 51
 - cuadrático, 51
 - exponencial, 51
 - lineal, 51
 - logístico, 51
- Modelos de Espacio de Estados, 102
- MSE, 32
- Navaja de Ockam, 43
- Operador Rezago, 162
- Principio de parsimonia, 43
- Proceso
- ARIMA, 213
 - SARIMA, 243
- Proceso
- $AR(p)$, 172
 - $ARMA(p,q)$, 186
 - integrado, 216
 - $MA(q)$, 162
- Proceso MA Invertible, 166
- Pronóstico
- de densidad, 55
 - por intervalo, 55
- Pronósticos
- combinación, 241
 - con base en la tendencia, 54
 - definición, 55
- Propiedades
- de procesos $MA(q)$, 163
- Prueba
- Dickey-Fuller, 225
 - Aumentada, 228
- Prueba
- Diebold-Mariano, 221
 - Durbin-Watson, 147
 - KPSS, 234
 - Ljung-Box, 143
 - no linealidad Teräsvirta, 237

- White, 238
- Prueba MOSUM, 82
- Pruebas
 - de no linealidad, 236
 - de Raíz Unitaria Estacional, 252
- Pruebas de Incorrelación, 142
- R cuadrado, 42
- R cuadrado ajustado, 42
- Raíz Unitaria, 212
 - Autoregresiva, 212
- Raíz unitaria estacional, 252
- Redes neuronales Autoregresivas NNAR, 105
- Regresión Loess, 91
- Representación
 - $MA(\infty)$, 187
 - causal, 187
- Residuales
 - Recursivos , 81
- Residuales Recursivos Cuadrados CUSUM-SQ, 83
- Residuos
 - estructurales, 68
- Residuos estandarizados, 33
- Ruido Blanco, 132
 - Gaussiano, 132
- Supercosistencia, 224
- Tendencia, 51
- Tendencia Estocástica, 216
- Teorema
 - de Bartlett, 138
 - de Wold, 187
- Transformación Box-Cox, 106
- Validación cruzada, 56
- Variable
 - Indicadora Estacional, 65