



Daniel Peña

# ANÁLISIS DE SERIES TEMPORALES

Alianza Editorial



Daniel Peña

# Análisis de series temporales

Alianza Editorial

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.

© Daniel Peña Sánchez de Rivera, 2005  
© Alianza Editorial, S. A., Madrid, 2005  
Calle Juan Ignacio Luca de Tena, 15; 28027 Madrid; teléf. 91 393 88 88  
[www.alianzaeditorial.es](http://www.alianzaeditorial.es)  
ISBN: 84-206-9128-3  
Depósito legal: M. 32.494-2005  
Impreso en Lavel  
Madrid  
Printed in Spain

*A mis maestros, que me enseñaron el camino;  
a mis doctorandos y coautores, con los que ha sido un placer recorrerlo;  
a Mely, Jorge y Alvaro, con amor,  
con los que siempre cuento.*

# Índice

Prefacio . . . . .	15
1. Introducción a las series temporales . . . . .	18
1.1. Ejemplos de series temporales univariantes . . . . .	18
1.2. Ejemplos de series multivariantes . . . . .	25
1.3. Contenido del libro . . . . .	27
1.4. Un poco de historia . . . . .	30
1.5. Programas de ordenador . . . . .	34
1.6. Lecturas complementarias . . . . .	35
Apéndice 1.1: Descripción de la entrada de datos con algunos programas de ordenador . . . . .	37
2. Análisis descriptivo de una serie temporal . . . . .	43
2.1. Introducción . . . . .	43
2.2. El análisis de tendencias deterministas . . . . .	44
2.2.1. Tipos de modelos . . . . .	44
2.2.2. Estimación . . . . .	46
2.2.3. Limitaciones del ajuste de tendencias deterministas . . . . .	49
2.3. Métodos de alisado . . . . .	52
2.3.1. El modelo de alisado simple . . . . .	52
2.3.2. El método de alisado doble de Holt . . . . .	56
2.4. Métodos de descomposición para series estacionales . . . . .	58
2.5. Estacionalidad y el ajuste de ciclos . . . . .	61
2.5.1. Definiciones básicas . . . . .	61
2.5.2. Representación de la estacionalidad por un ciclo . . . . .	65
2.6. Exploración de múltiples ciclos: el periodograma . . . . .	68

## Índice

---

2.7. Conclusiones y lecturas complementarias . . . . .	72
Apéndice 2.1: Regresión múltiple y la estimación del periodograma . . . . .	74
 3. Series temporales y procesos estocásticos . . . . .	78
3.1. Introducción . . . . .	78
3.2. El concepto de proceso estocástico . . . . .	79
3.2.1. Propiedades de las distribuciones marginales . . . . .	82
3.2.2. Propiedades de las distribuciones condicionadas . . . . .	85
3.3. Procesos estacionarios . . . . .	86
3.3.1. Definición . . . . .	86
3.3.2. Combinaciones de procesos estacionarios (*) . . . . .	89
3.4. Proceso de ruido blanco . . . . .	90
3.5. Estimación de los momentos de procesos estacionarios . . . . .	92
3.5.1. Estimación de la media . . . . .	92
3.5.2. Estimación de las autocovarianzas y autocorrelaciones . . . . .	95
3.6. El espectro de un proceso estacionario (*) . . . . .	98
Apéndice 3.1: Condición de convergencia . . . . .	103
Apéndice 3.2: Diferencia de Martingalas . . . . .	103
Apéndice 3.3: Las fórmulas de Bartlett para las autocovarianzas estimadas . . . . .	104
Apéndice 3.4: El espectro y las autocovarianzas . . . . .	104
Apéndice 3.5: Generación de procesos con el ordenador . . . . .	106
 4. Procesos autorregresivos . . . . .	109
4.1. Introducción . . . . .	109
4.2. El proceso autorregresivo de primer orden (AR(1)) . . . . .	110
4.2.1. Esperanza y varianza . . . . .	111
4.2.2. Función de autocovarianzas y autocorrelación . . . . .	113
4.2.2.1. Función de autocorrelación simple (fas) . . . . .	113
4.2.2.3. Representación del proceso AR(1) como suma de innovaciones	114
4.3. El proceso AR(2) . . . . .	115
4.3.1. Función de autocovarianzas . . . . .	117
4.3.2. Función de autocorrelación simple . . . . .	118
4.3.3. Representación del AR(2) como suma de innovaciones . . . . .	119
4.4. El proceso autorregresivo general (AR( $p$ )) . . . . .	123
4.4.1. Función de autocorrelación simple . . . . .	124
4.4.2. Ecuaciones de Yule-Walker . . . . .	124
4.4.3. Representación del AR( $p$ ) suma de innovaciones . . . . .	126
4.5. La función de autocorrelación parcial . . . . .	126
Apéndice 4.1: Notación de operadores . . . . .	131
Apéndice 4.2: Ecuaciones en diferencias . . . . .	133
Apéndice 4.3: Cálculo de los coeficientes de autocorrelación parcial . . . . .	139
Apéndice 4.4: Generación de procesos AR . . . . .	140
 5. Procesos de media móvil y ARMA . . . . .	142
5.1. Introducción . . . . .	142
5.2. El proceso de media móvil de orden uno (MA(1)) . . . . .	143
5.2.1. Función de autocorrelación simple y parcial . . . . .	144
5.3. El proceso MA( $q$ ) . . . . .	146
5.4. El proceso MA( $\infty$ ). La descomposición de Wold . . . . .	149

## Análisis de series temporales

---

5.4.1. Los procesos AR y MA y el proceso general . . . . .	152
5.5. El proceso ARMA(1, 1) . . . . .	154
5.6. Procesos ARMA( $p, q$ ) . . . . .	156
5.7. Los procesos ARMA y la suma de procesos estacionarios . . . . .	158
Apéndice 5.1: Función de correlación inversa . . . . .	162
6. Procesos integrados y de memoria larga . . . . .	166
6.1. Introducción . . . . .	166
6.2. Procesos integrados . . . . .	167
6.3. El paseo aleatorio . . . . .	172
6.4. El proceso de alisado exponencial simple . . . . .	176
6.5. Procesos integrados de orden dos . . . . .	178
6.6. Procesos integrados ARIMA . . . . .	179
6.7. Procesos integrados y tendencias . . . . .	181
6.8. Procesos de memoria larga (*) . . . . .	184
6.9. Lecturas complementarias . . . . .	189
Apéndice 6.1: La función Gamma y los procesos de memoria larga . . . . .	191
Apéndice 6.2: Generación de series no estacionarias . . . . .	192
7. Procesos ARIMA estacionales . . . . .	194
7.1. Introducción . . . . .	194
7.2. El concepto de estacionalidad y sus tipos . . . . .	195
7.3. El modelo ARIMA estacional . . . . .	199
7.4. Función de autocorrelación simple de un proceso estacional . . . . .	204
7.5. Función de autocorrelación parcial . . . . .	210
7.6. Generalizaciones y lecturas complementarias . . . . .	211
Apéndice 7.1: Un contraste simple de igualdad de la estructura estacional . . . . .	213
Apéndice 7.2: Generación de series estacionales . . . . .	213
8. Predicción con modelos ARIMA . . . . .	215
8.1. Introducción . . . . .	215
8.2. La ecuación de predicción de un modelo ARIMA . . . . .	216
8.2.1. La esperanza condicionada como predictor óptimo . . . . .	216
8.2.2. Cálculo de las predicciones . . . . .	219
8.3. Interpretación de las predicciones . . . . .	224
8.3.1. Procesos no estacionales . . . . .	224
8.3.2. Procesos estacionales . . . . .	230
8.3.2.1. El modelo de pasajeros de avión . . . . .	232
8.4. Varianza de las predicciones . . . . .	236
8.5. Adaptación de las predicciones . . . . .	239
8.6. Medidas de predecibilidad . . . . .	240
8.7. Lecturas complementarias . . . . .	242
Apéndice 8.1: Predicción y esperanza condicionada . . . . .	244
9. La identificación de los posibles modelos ARIMA . . . . .	247
9.1. Introducción . . . . .	247
9.2. Determinación de la transformación para estabilizar la varianza . . . . .	248
9.3. Determinación de la transformación para estabilizar la media . . . . .	253
9.3.1. Determinar el orden de diferenciación regular . . . . .	253
9.3.2. Determinar el orden de diferenciación estacional . . . . .	255

---

9.3.3. Contrastes de raíces unitarias . . . . .	256
9.3.3.1. Contraste de Dickey-Fuller . . . . .	257
9.3.3.2. El contraste de Dickey-Fuller aumentado . . . . .	261
9.4. La identificación de la estructura ARMA . . . . .	265
9.5. Lecturas complementarias . . . . .	272
Apéndice 9.1: La transformación Box-Cox para estabilizar la varianza . . . . .	273
Apéndice 9.2: Programas para la identificación . . . . .	274
 10. Estimación y selección de modelos ARIMA . . . . .	277
10.1. Introducción . . . . .	277
10.2. La función de verosimilitud de un proceso ARMA . . . . .	278
10.3. Procesos AR . . . . .	281
10.3.1. El proceso AR(1) . . . . .	281
10.3.2. Procesos AR(p) . . . . .	283
10.4. Estimación de modelos MA y ARMA . . . . .	285
10.4.1. Estimación MV condicional . . . . .	285
10.4.1.1. El algoritmo de Hannan y Rissanen . . . . .	286
10.4.2. Estimación MV exacta . . . . .	287
10.5. Estimación recursiva con el filtro de Kalman . . . . .	288
10.5.1. Modelos en el espacio de los estados . . . . .	288
10.5.2. El filtro de Kalman . . . . .	292
10.6. Propiedades de los estimadores . . . . .	295
10.7. Criterios de selección de modelos . . . . .	297
10.7.1. El criterio AIC de Akaike . . . . .	299
10.7.2. El criterio BIC . . . . .	300
10.7.3. Comparación entre criterios . . . . .	300
10.8. Lecturas complementarias . . . . .	303
Apéndice 10.1: Algoritmos de optimización no lineal . . . . .	305
Apéndice 10.2: El criterio AIC . . . . .	309
Apéndice 10.3: El criterio BIC . . . . .	310
Apéndice 10.4: Programas para la estimación . . . . .	312
 11. Diagnosis del modelo y predicción . . . . .	313
11.1. Introducción . . . . .	313
11.2. Contrastes de autocorrelación . . . . .	314
11.2.1. Propiedades de las autocorrelaciones estimadas . . . . .	314
11.2.2. El contraste de Ljung-Box sobre las autocorrelaciones . . . . .	315
11.2.3. El contraste del determinante . . . . .	316
11.2.4. Sobreajuste: contrastes con el criterio BIC . . . . .	317
11.3. Otros contrastes . . . . .	318
11.3.1. Contrastos de media cero . . . . .	318
11.3.2. Contrastos de homocedasticidad . . . . .	319
11.3.3. Contrastos de normalidad . . . . .	319
11.4. Contraste de estabilidad del modelo . . . . .	322
11.5. Interpretación del modelo. Componentes deterministas . . . . .	323
11.5.1. Tendencias . . . . .	323
11.5.2. Estacionalidad . . . . .	325
11.5.3. Caso general . . . . .	326
11.6. Predicción . . . . .	328

## Análisis de series temporales

---

11.6.1. Predicciones puntuales . . . . .	328
11.6.2. Intervalos de predicción . . . . .	331
11.6.3. Intervalos de predicción para muestras grandes . . . . .	332
11.6.4. Intervalos mediante remuestreo . . . . .	332
11.6.5. Predicción mediante promedio de modelos . . . . .	334
Apéndice 11.1: Predicción con incertidumbre de los parámetros y del modelo	336
Apéndice 11.2: Programas para la diagnosis . . . . .	338
 12. Análisis de intervención . . . . .	340
12.1. Introducción . . . . .	340
12.2. Efectos cualitativos: variables impulso y escalón . . . . .	342
12.2.1. Variables impulso: función de respuesta a impulsos . . . . .	342
12.2.2. Variables escalón: ganancia . . . . .	345
12.2.3. Relación entre impulsos y escalones . . . . .	348
12.3. Efectos deterministas generales . . . . .	351
12.4. Construcción de modelos de intervención . . . . .	356
12.5. Estimación de valores ausentes . . . . .	358
Apéndice 12.1: Predicción de valores ausentes . . . . .	362
Apéndice 12.2: Estimación de efectos de intervención . . . . .	364
 13. Valores atípicos . . . . .	366
13.1. Introducción . . . . .	366
13.2. Atípicos aditivos . . . . .	367
13.2.1. Efectos en los residuos . . . . .	368
13.2.2. Efectos en la estimación de los parámetros . . . . .	370
13.3. Atípicos innovativos (IO) . . . . .	371
13.3.1. Efecto de un IO sobre la serie . . . . .	373
13.3.2. Efectos en la estimación . . . . .	376
13.4. Cambio de nivel . . . . .	377
13.4.1. Efectos en los residuos . . . . .	379
13.4.2. Efecto en la estimación . . . . .	380
13.5. Cambios transitorios y efecto rampa . . . . .	380
13.6. Procedimientos de estimación de atípicos . . . . .	381
13.6.1. Estimación del tamaño del atípico . . . . .	382
13.6.2. El procedimiento general . . . . .	384
13.7. Comentarios al procedimiento de detección de atípicos y lecturas complementarias . . . . .	390
Apéndice 13.1: Estimación de los cambios de nivel . . . . .	394
Apéndice 13.2: Programas para análisis de atípicos . . . . .	396
 14. Modelos no lineales . . . . .	399
14.1. Introducción . . . . .	399
14.2. Modelos no lineales en la media y en la varianza . . . . .	402
14.3. Los procesos bilineales . . . . .	404
14.4. Los procesos autorregresivos por umbrales . . . . .	409
14.5. Otros modelos no lineales . . . . .	414
14.6. Contrastes de no linealidad . . . . .	417
14.6.1. Tipos de contrastes . . . . .	417

---

## Índice

14.6.2. Contrastes de regresión entre los residuos y funciones de la variable . . . . .	418
14.6.3. Contrastes sobre los cuadrados de los residuos . . . . .	419
14.6.4. Contrastes no paramétricos sobre la dependencia de los residuos. El contraste BDS . . . . .	420
14.7. Lecturas complementarias . . . . .	423
15. Modelos de heterocedasticidad condicional . . . . .	425
15.1. Introducción . . . . .	425
15.2. Modelos ARCH . . . . .	429
15.2.1. El modelo ARCH(1) . . . . .	430
15.2.1.1. La autocorrelación de los cuadrados . . . . .	431
15.2.1.2. Curtosis . . . . .	433
15.2.2. El modelo ARCH(r) . . . . .	438
15.3. Modelos GARCH . . . . .	439
15.3.1. El modelo GARCH(1,1) . . . . .	439
15.3.1.1. Dependencia de los cuadrados . . . . .	441
15.3.1.2. Curtosis . . . . .	442
15.3.2. El modelo GARCH general . . . . .	443
15.4. Construcción de modelos ARCH y GARCH . . . . .	445
15.4.1. Identificación . . . . .	446
15.4.2. Estimación . . . . .	446
15.4.3. Diagnosis . . . . .	449
15.5. Modelos de volatilidad estocástica . . . . .	453
15.5.1. El modelo SV(1) . . . . .	453
15.6. Otros enfoques y lecturas complementarias . . . . .	454
Apéndice 15.1: Propiedades de ARCH(1) . . . . .	457
Apéndice 15.2: Programas de estimación . . . . .	458
16. Casos de series temporales univariantes . . . . .	463
16.1. Introducción . . . . .	463
16.2. La serie de consumo de gasolina . . . . .	464
16.2.1. Introducción . . . . .	464
16.2.2. Identificación del modelo univariante . . . . .	465
16.2.3. Estimación y selección del modelo . . . . .	469
16.2.4. Predicciones . . . . .	471
16.2.5. Análisis de valores atípicos . . . . .	472
16.2.6. Análisis de efectos deterministas . . . . .	474
16.2.7. Análisis en distintos períodos . . . . .	477
16.3. El índice de producción industrial . . . . .	478
16.3.1. Introducción . . . . .	478
16.3.2. El modelo univariante para el IPI . . . . .	478
16.3.3. Datos atípicos . . . . .	482
16.3.4. Predicciones . . . . .	483
16.4. El modelo para la serie del índice general de la Bolsa de Madrid . . . . .	483
16.5. Lecturas complementarias . . . . .	484
17. Regresión dinámica entre variables estacionarias . . . . .	488
17.1. Introducción . . . . .	488

## Análisis de series temporales

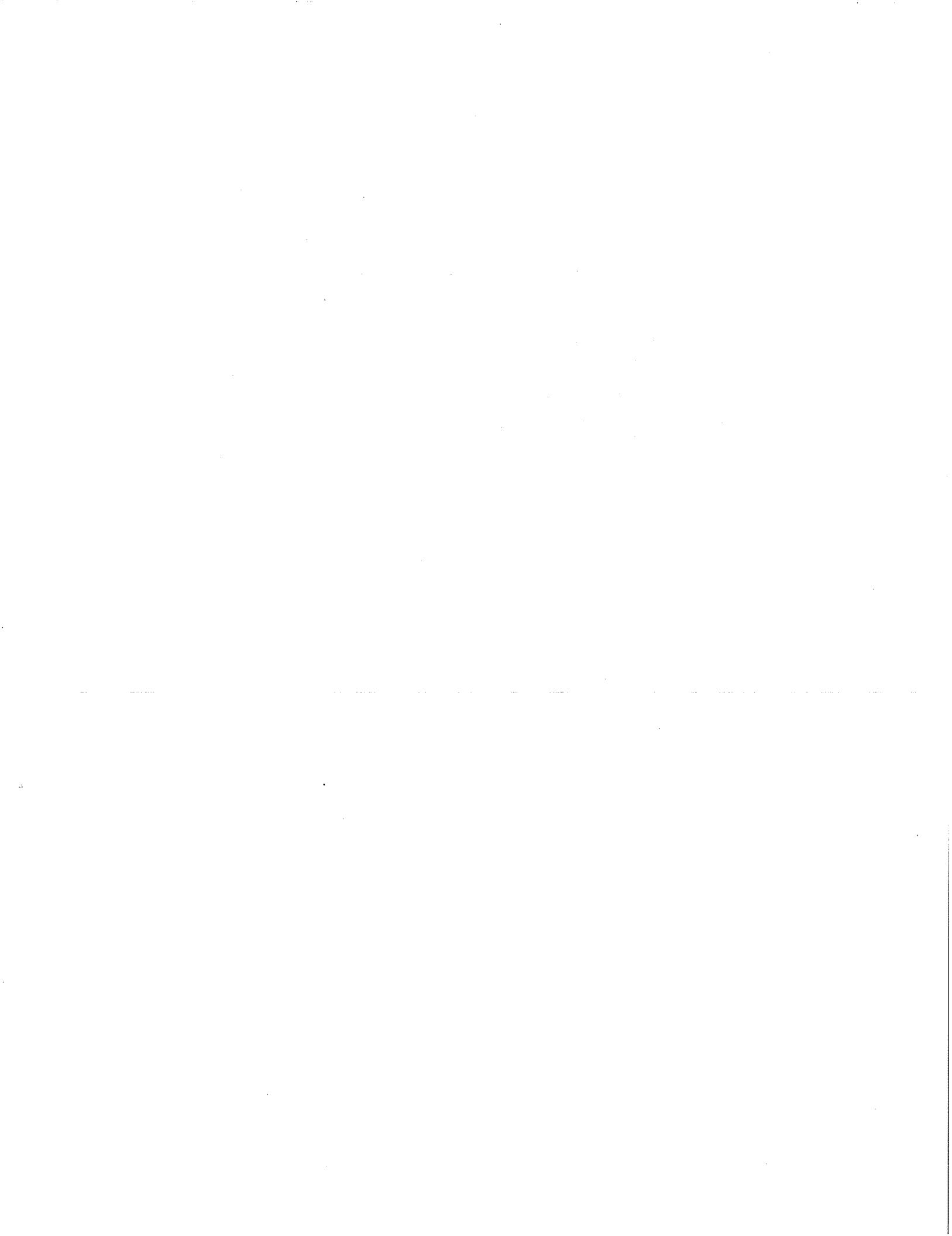
---

17.2. Relaciones entre dos series estacionarias: las funciones de covarianzas y correlaciones cruzadas . . . . .	491
17.2.1. La función de covarianzas cruzadas . . . . .	491
17.2.2. Función de autocorrelación cruzada . . . . .	494
17.2.3. Estimación de las funciones de covarianzas y correlaciones cruzadas . . . . .	495
17.3. El modelo dinámico entre dos series estacionarias . . . . .	497
17.3.1. Representación general . . . . .	497
17.3.2. Representación de la función de transferencia como cociente de polinomios . . . . .	498
17.3.3. Representación de la función de transferencia como efectos a corto y a largo plazo . . . . .	500
17.4. Construcción de modelos de regresión dinámica entre variables estacionarias . . . . .	501
17.4.1. Identificación . . . . .	501
17.4.2. Estimación y selección del modelo . . . . .	503
17.4.3. Contrastes diagnósticos . . . . .	507
17.4.3.1. Contrastes sobre los parámetros . . . . .	507
17.4.3.2. Contrastes sobre los residuos . . . . .	509
17.5. Relaciones espurias entre variables dinámicas . . . . .	512
17.5.1. Relaciones sin dinámica . . . . .	513
17.5.2. Perturbaciones autocorreladas . . . . .	514
17.6. El modelo con varias variables explicativas . . . . .	517
17.7. Predicción . . . . .	519
Apéndice 17.1: Varianza de la estimación de mínimos cuadrados con autocorrelación . . . . .	522
Apéndice 17.2: Mínimos cuadrados generalizados . . . . .	524
Apéndice 17.3: Estimación de la función de transferencia mediante preblanqueo . . . . .	526
Apéndice 17.4: Descomposición de la función de transferencia . . . . .	527
Apéndice 17.5: Programas para la estimación de funciones de transferencia	528
 18. Regresión entre variables integradas. Cointegración . . . . .	530
18.1. Introducción . . . . .	530
18.2. Cointegración . . . . .	532
18.2.1. Estimación de la relación entre variables cointegradas . . . . .	533
18.3. Construcción de modelos dinámicos. Caso general . . . . .	537
18.3.1. Variables con mismo orden de integración . . . . .	537
18.3.2. Distinto orden de cointegración . . . . .	538
18.3.3. Varios regresores . . . . .	538
18.4. Lecturas complementarias . . . . .	545
Apéndice 18.1: Propiedades de procesos integrados . . . . .	545
Apéndice 18.2: Estimación con EViews . . . . .	546
 19. Modelos multivariantes . . . . .	548
19.1. Introducción . . . . .	548
19.2. Procesos vectoriales estacionarios y sus autocorrelaciones . . . . .	548
19.2.1. Procesos vectoriales estacionarios . . . . .	548
19.2.2. Matrices de autocorrelación simple y parcial . . . . .	550

## Índice

---

19.2.3. Estimación de las matrices simples y parciales . . . . .	551
19.3. Modelos ARMA vectoriales . . . . .	552
19.3.1. Proceso VAR(1) . . . . .	554
19.3.2. Procesos VAR( $p$ ) . . . . .	556
19.3.3. Procesos VMA( $q$ ) . . . . .	557
19.3.4. Procesos VARMA( $p,q$ ) . . . . .	558
19.4. Construcción de modelos VARMA para series estacionarias . . . . .	559
19.4.1. Identificación . . . . .	559
19.4.2. Estimación . . . . .	561
19.4.3. Diagnosis . . . . .	562
19.5. Cointegración y factores comunes . . . . .	563
19.6. Especificación del modelo para variables no estacionarias . . . . .	565
19.6.1. El modelo de corrección de error . . . . .	565
19.6.2. El modelo factorial . . . . .	566
19.7. Lecturas complementarias . . . . .	569
Apéndice 19.1: Los riesgos de la sobrediferenciación en series multivariantes	569
Bibliografía . . . . .	571
Índice analítico . . . . .	585



# Prefacio

Este libro pretende enseñar al lector a construir modelos de series temporales para explicar la evolución histórica de una variable a lo largo del tiempo y predecir sus valores futuros. Puede utilizarse como texto de un curso cuatrimestral de series temporales dirigido a estudiantes de cualquier rama científica, ya que los métodos aquí presentados son de aplicación general en las ciencias económicas, las ingenierías y las ciencias biológicas y ambientales. Los conocimientos previos requeridos para seguir este libro son los fundamentos de probabilidad e inferencia que se adquieren en los cursos básicos de estadística para estas disciplinas. Es aconsejable, aunque no imprescindible, un buen conocimiento de los métodos de regresión lineal, y las ideas de regresión necesarias para seguir este libro se repasan en los capítulos iniciales.

Un primer curso cuatrimestral de predicción mediante series temporales debería cubrir la parte principal de los capítulos 1 al 13, con una breve introducción a los temas posteriores. Si va dirigido a estudiantes de Estadística o Matemáticas el curso debe cubrir la mayor parte del material presentado en estos capítulos, incluyendo los apéndices, mientras que para estudiantes de ingeniería o economía los temas 4 a 8 y 10 deben pasarse más rápidamente y concentrar el curso en los capítulos 9, 11 y 12. Un segundo curso cuatrimestral de series temporales podría comenzar revisando los primeros capítulos y concentrándose en el material de los capítulos 12 al 19.

El libro contiene numerosos ejemplos de aplicación y recomendamos

que, sea cual sea el nivel del curso, el estudiante dedique una parte del mismo a modelizar y prever series reales. Con este objetivo se ha preparado un conjunto de series temporales, que se utilizan en los ejemplos y ejercicios del libro, que pueden descargarse de la página web <http://www.alianzaeditorial.es>. Para facilitar el trabajo con los programas de ordenador muchos capítulos incluyen apéndices explicando cómo implementar los métodos estudiados con varios programas informáticos como Statgraphics, Minitab, SPSS, S-plus, Matlab, SCA, EViews o TSW.

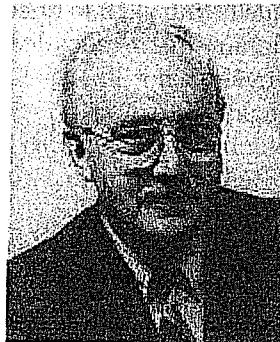
Este libro se inició como una ampliación y puesta al día de los dos capítulos de series temporales del segundo tomo de mi libro *Estadística: Modelos y Métodos*, publicado por Alianza Editorial. El proceso de revisión ha llevado a reescribirlos por completo, con el objetivo de incorporar los nuevos avances y aprovechar la rápida capacidad de cálculo disponible, que permite hoy aplicar de forma rutinaria procedimientos inimaginables hace 10 años.

En la notación decimal hemos seguido la recomendación de la Real Academia de Ciencias Exactas, Físicas y Naturales de utilizar el punto en lugar de la coma para indicar los decimales, unificando la notación española con la anglosajona que es la estándar en programas de ordenador. Esto simplifica la presentación de resultados obtenidos con programas de ordenador y evita confusiones.

Estoy muy agradecido a mis colegas que han contribuido de distintas formas a que este libro se hiciese realidad. En su largo proceso de gestación las siguientes personas han contribuido con sus comentarios y su trabajo a mi aprendizaje de series temporales y a mejorar la presentación del material de este libro: Andrés Alonso, Antonio Aznar, George Box, Ángeles Carnero, Álvaro Escribano, Rafael Flores, Clive Granger, Antonio García Ferrer, Roland Fried, Jesús Gonzalo, Juan del Hoyo, Miguel Jerez, Ana Justel, Alberto Luceño, Luis Manzanedo, José Mira, Alfonso Novales, Regina Kaiser, Pilar Poncela, Albert Prat, Javier Prieto, Dolores Redondas, Julio Rodríguez, Rosario Romera, Juan Romo, Ismael Sánchez, María Jesús Sánchez, Jaime Terceiro, George Tiao, Arthur Treadway, Ruey Tsay, Santiago Velilla y Victor Yohai. A todos les agradezco mucho su contribución desde distintos aspectos a este proyecto. En particular, estoy en deuda con Arthur Treadway por la foto de G. M. Jenkins, y con el profesor T. M. F. Smith de Southampton por la foto de M. Quenouille. Mis numerosas conversaciones con mis amigos Antoni Espasa y Agustín Maravall me han ayudado a entender mejor muchos aspectos del análisis de series temporales. Tengo una profunda deuda de gratitud con Esther Ruiz, que ha leído una primera versión de todos los capítulos y cuyas excelentes sugerencias han clarificado y mejorado la presentación de numerosas secciones en todo el libro. Pedro

Galeano ha sido una enorme ayuda en todas sus fases de realización: hemos investigado juntos algunos aspectos que el lector puede encontrar en las referencias, me ha ayudado mucho en la descripción de los programas que aparecen en los apéndices, ha rehecho y mejorado muchas de las figuras y ha preparado con gran cuidado el manuscrito en tex para su publicación, todo con extraordinaria eficacia. La perseverancia y continuo estímulo y apoyo de Belén Urrutia, de Alianza Editorial, ha sido muy importante para completar este proyecto. Por supuesto, los errores que hayan podido quedar son de mi exclusiva responsabilidad y agradezco de antemano a los lectores cualquier sugerencia o crítica, que ruego me envíen a mi dirección de correo electrónico, daniel.pena@uc3m.es.

# 1. Introducción a las series temporales



George E. P. Box (1919)

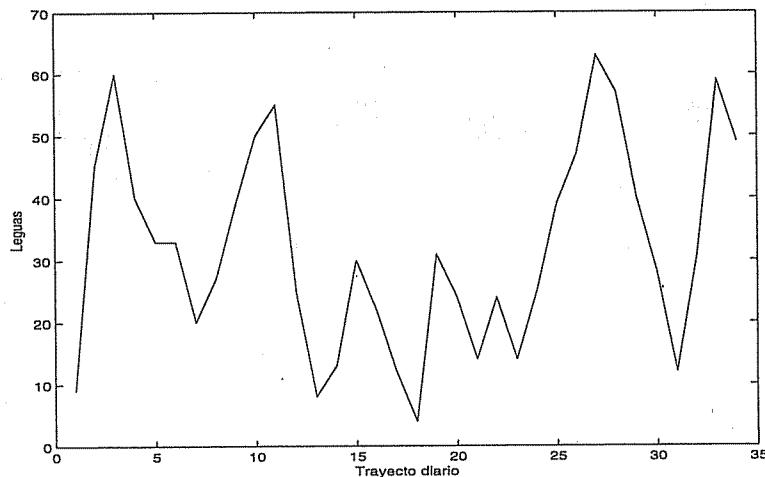
Estadístico británico. Trabajó durante ocho años en Inglaterra para una empresa química y en ese periodo inventó la operación evolutiva y el análisis de las superficies de respuesta. En 1956 emigra a EEUU, donde funda el departamento de Estadística de la Universidad de Wisconsin en Madison, que se convierte en poco tiempo en uno de los principales centros de investigación estadística en el mundo. Creador con G. Jenkins de la metodología más utilizada actualmente para el análisis de series temporales, sus contribuciones a todos los campos de la estadística le han convertido en uno de los científicos más influyentes del siglo XX.

## 1.1. Ejemplos de series temporales univariantes

Una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares (cada día, cada mes, cada año, etc.). Por ejemplo, las figuras 1.1 a 1.9 representan series temporales con distintos períodos de observación y diferentes propiedades. En las dos primeras, figuras 1.1 y 1.2, las series oscilan alrededor de un nivel constante y decimos que son estables o *estacionarias*. En las tres siguientes, representadas en las figuras 1.3, 1.4 y 1.5, las series no se mantienen en un nivel constante y decimos que son series *no estacionarias*. Las figuras 1.6, 1.8 y 1.9 representan series temporales que además de tener un nivel fijo, como la serie de la figura 1.6, o variable en el tiempo, como las de las figuras 1.8 y 1.9, tienen además un comportamiento superpuesto que se repite a lo largo del tiempo, diremos que estas series son *estacionales*. Vamos a explicar estos tres conceptos.

## 1. Introducción a las series temporales

**Figura 1.1 Leguas diarias recorridas por la flota de Cristóbal Colón en su primer viaje a América desde la isla de la Gomera a la de San Salvador**

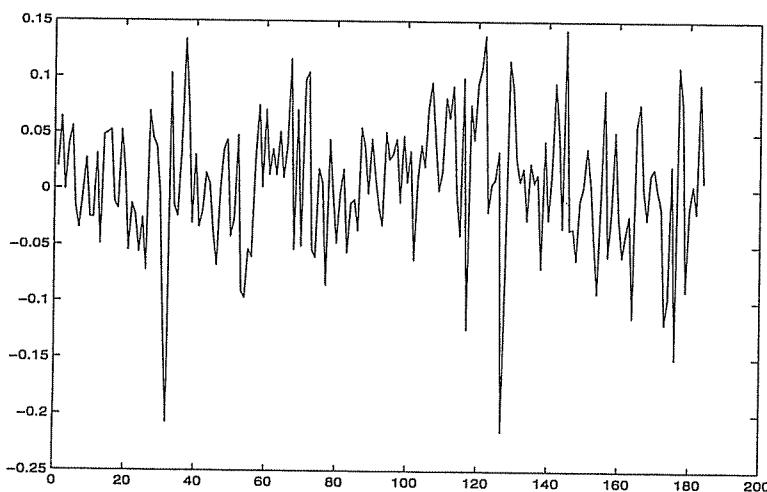


La figura 1.1 es un ejemplo de serie estacionaria. Muestra el trayecto diario navegado, medido en leguas marinas, por la flota de Cristóbal Colón en su primer viaje a América. Los datos de esta serie se encuentran en el fichero colon.dat. Se observa que esta serie es estable, con valores que oscilan alrededor de un recorrido promedio diario de unas 30 leguas. Su gráfico no muestra ninguna tendencia creciente o decreciente clara.

La figura 1.2 es otro ejemplo de una serie estacionaria. Presenta la rentabilidad promedio mensual obtenida en la Bolsa de Madrid en el periodo 1/1988 a 12/2000 medida por el índice general. La rentabilidad se ha calculado como la diferencia relativa entre el valor promedio de las acciones al final y al comienzo de cada mes. Se observa que los valores de la serie parecen moverse alrededor de una rentabilidad mensual de 0.00722, equivalente al 8.66% ( $0.00722 \times 12 = .08664$ ) anual. Como la anterior, esta serie tiene un nivel fijo, sin ninguna tendencia clara a crecer o a decrecer con el tiempo.

La figura 1.3 presenta una serie que, a diferencia de las dos anteriores, no es claramente estable en el tiempo. Llamaremos a estas series no estacionarias. La serie corresponde a la población mayor de 16 años en España al final de cada trimestre en el periodo 1/1977 al 4/2000, y se encuentra en el fichero pobmay16.dat. Se observa en el gráfico que la serie no es estable, ya que su nivel aumenta con el tiempo. Diremos que la serie tiene una clara *tendencia positiva*. La mayoría de las series económicas y sociales no son estacionarias (estables) y presentan tendencias más

**Figura 1.2 Rendimientos mensuales obtenidos en la Bolsa de Madrid de acuerdo al índice general en el periodo 1988 a 2000**



o menos constantes en el tiempo. En este caso la tendencia es aproximadamente lineal, aunque la pendiente de la recta que ajustaríamos en la primera mitad de la muestra sería algo mayor que en la segunda. Esto sugiere que quizás el crecimiento anual de la población va cambiando en el tiempo con una pauta decreciente, de manera que es mayor al principio de la muestra que al final. La tendencia de esta serie sería entonces variable en el tiempo, en lugar de constante. Esta propiedad es frecuente en las series reales, donde es raro encontrar una tendencia constante en períodos largos de observación.

La figura 1.4 representa otra serie no estacionaria, el número anual de nacimientos que se han producido en España desde 1946 hasta el año 2000. Los datos están en el fichero nacidos.dat. El gráfico muestra que esta serie anual no es estable, ya que no oscila alrededor de ningún nivel fijo. La serie no muestra una tendencia lineal creciente o decreciente en todo el periodo: los nacimientos en España crecieron en los años 50 y 60, se estabilizaron en el periodo 1965 a 1976 y a partir de 1977 comenzaron a decrecer. El decrecimiento se mantiene hasta finales de los 90, donde se observa una pequeña recuperación de la natalidad. Esta serie presenta una tendencia evolutiva o cambiante en el tiempo.

La figura 1.5 presenta otra serie no estacionaria: el precio mensual del trigo en Valladolid entre julio de 1880 y diciembre de 1890. Los datos están en el fichero trigo.dat. De nuevo la serie no presenta una evolución alrededor de un nivel fijo y tiene un nivel variable en el tiempo. El precio

## 1. Introducción a las series temporales

Figura 1.3 Población mayor de 16 años en España desde el primer cuatrimestre de 1977 al cuarto del 2000

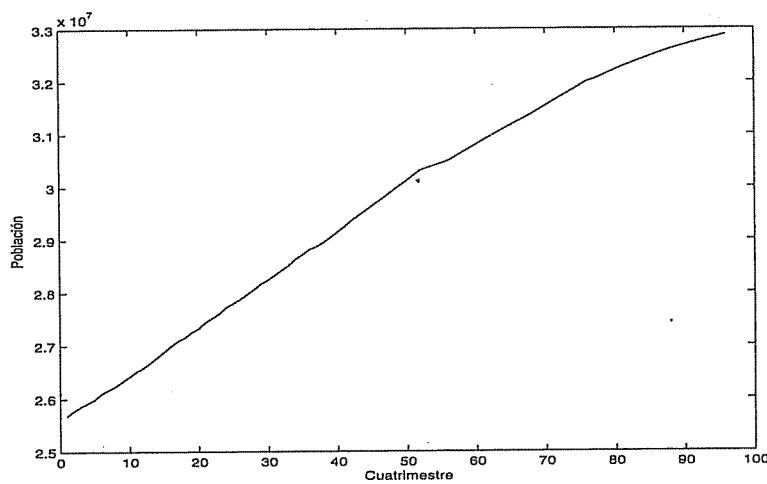
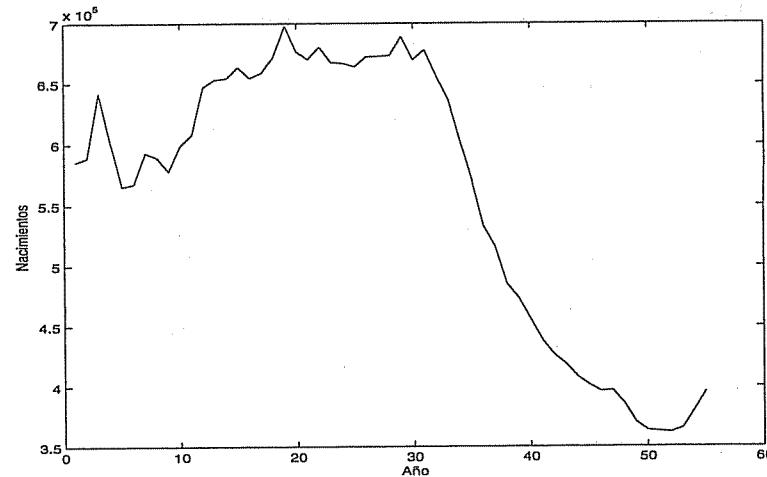


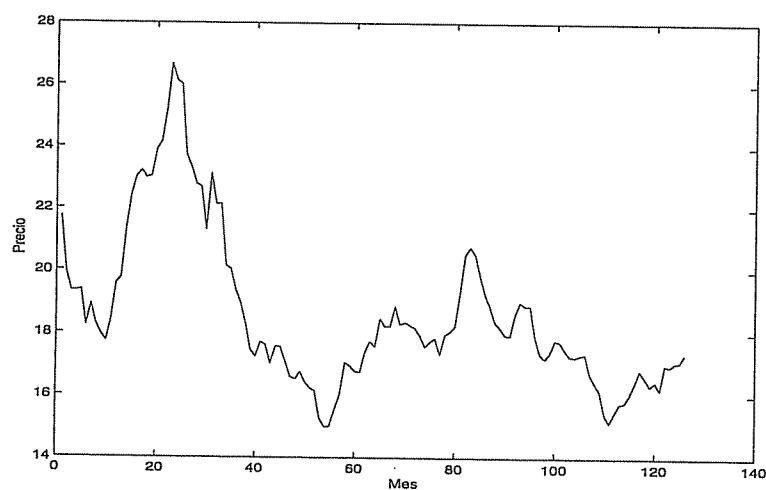
Figura 1.4 Nacimientos anuales en España 1946-2000



del trigo experimentó un gran aumento en el año 1882, pero luego decreció y en los últimos años experimenta menores fluctuaciones que en los primeros años de la muestra.

La serie de la figura 1.6, que se encuentra en el fichero lluviaSC.dat, muestra las lluvias mensuales en Santiago de Compostela durante diez años y es un ejemplo de serie estacional. Se observa que los valores de la

**Figura 1.5 Precio del trigo en Valladolid en el periodo julio de 1880 a diciembre de 1890**



serie oscilan ahora de nuevo alrededor de un valor central, pero algunos meses tienen sistemáticamente más lluvias que otros. Por ejemplo, los meses de verano tienen menos precipitaciones que los meses de invierno. Este efecto sería visible en el gráfico de la serie si marcásemos los meses, pero podemos verlo mejor haciendo gráficos para cada uno de los meses del año. Como ilustración, la figura 1.7 presenta las precipitaciones en los meses de enero y julio de distintos años. Se observa que, en ambos casos, las series oscilan alrededor de un valor central, y que hay años donde ambas son relativamente altas, años húmedos, frente a otros donde ambas son bajas, años secos. Sin embargo, lo más destacado del gráfico es que los valores medios de ambas series son muy distintos, siendo la lluvia media en enero mucho mayor que en julio. Este fenómeno, que el valor medio de la variable observada dependa del mes considerado, se denomina *estacionalidad* y es muy frecuente en las series de variables económicas, sociales o climáticas.

La figura 1.8 presenta la serie del fichero gasolauto.dat, y tiene tanto tendencia como estacionalidad. Corresponde al consumo mensual de gasolina en España. Se observa en la serie una tendencia globalmente creciente, aunque esta tendencia no es lineal y combina periodos de crecimiento con otros de estabilidad y decrecimiento. Además, la serie muestra una marcada estacionalidad, con picos y valles muy marcados en los distintos meses. En efecto, los picos del gráfico corresponden a los meses de julio y agosto, donde el consumo de gasolina se incrementa mucho por la llegada de turistas y los desplazamientos veraniegos.

## 1. Introducción a las series temporales

Figura 1.6 Lluvia en Santiago de Compostela desde enero 1988 a diciembre 1997

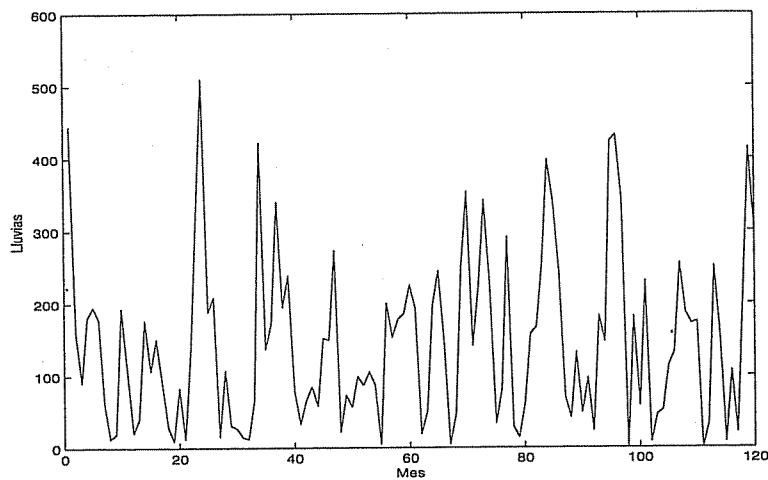
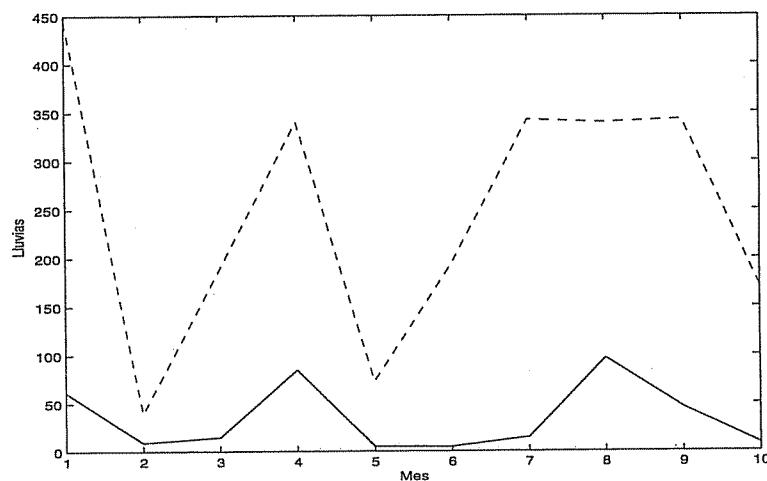
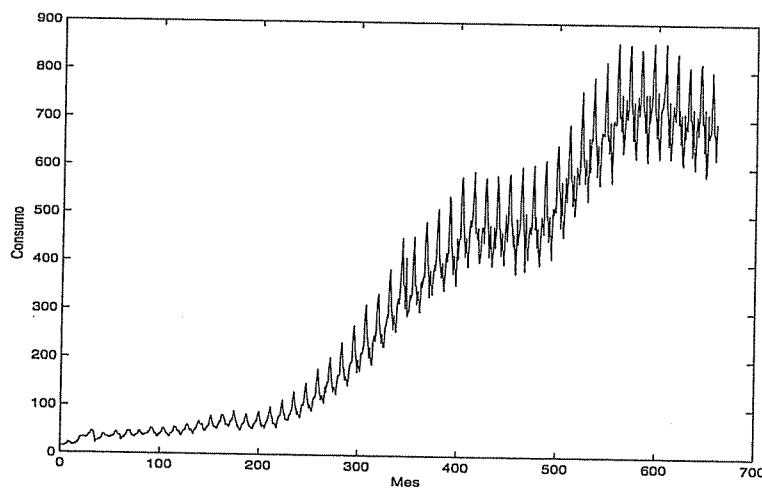


Figura 1.7 Lluvia en Santiago de Compostela en los meses de enero (línea discontinua) y julio (línea continua)

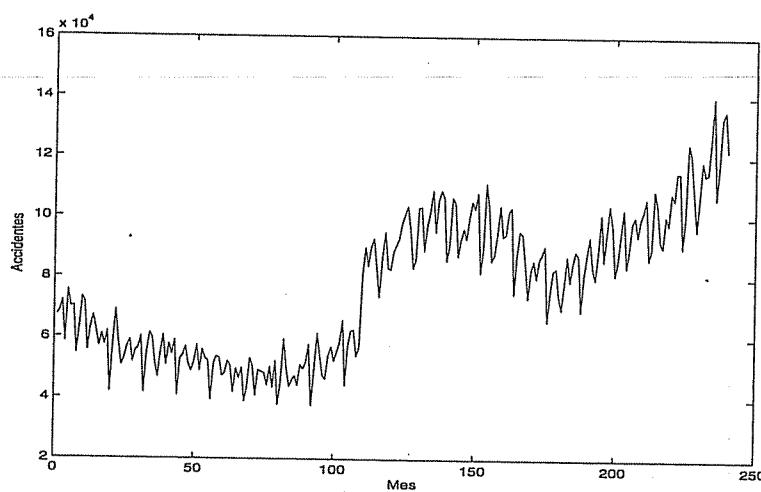


Finalmente, la figura 1.9 presenta de nuevo una serie con estacionalidad y nivel variable. Corresponde al número de accidentes laborales mensuales en España en el periodo 1979 a 1988, y se encuentra en el fichero accidentes.dat. Se observa que la tendencia de esta serie ha variado: decreciente primero, para pasar a creciente después. La serie presenta también

**Figura 1.8 Consumo de gasolina auto en España en el periodo enero de 1945 a diciembre de 1999**



**Figura 1.9 Accidentes en jornada laboral en el periodo enero de 1979 hasta diciembre de 1998**



una pauta regular anual, debida a la estacionalidad.

En resumen, las series temporales pueden tener o no un nivel estable en el tiempo, y si no lo tienen, pueden presentar tendencias más o menos constantes. Cuando el nivel de la serie no es estable decimos que la serie no es estacionaria. Un caso particular de no estacionariedad es cuando el

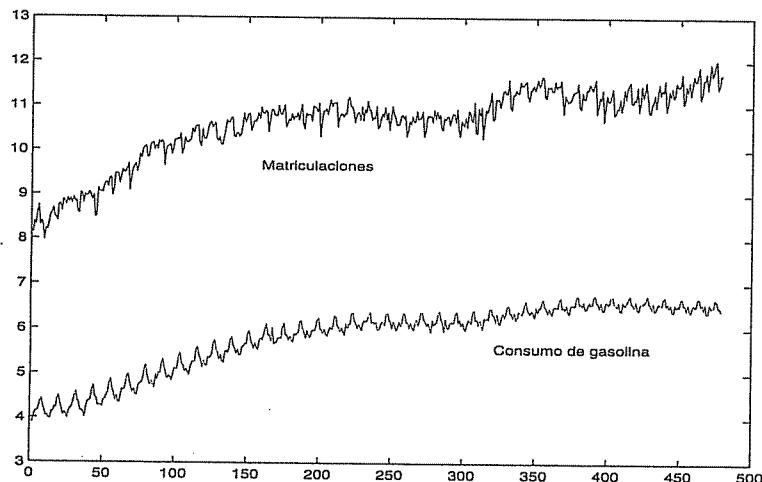
nivel varía siguiendo un ciclo, como ocurre con la temperatura mensual dentro del año. Diremos entonces que la serie es estacional y esta faceta puede combinarse con tendencias más a menos acusadas en el nivel general. El gráfico de la serie es siempre una herramienta muy valiosa para entender su comportamiento.

## 1.2. Ejemplos de series multivariantes

Además de estudiar la evolución histórica de una serie, los modelos de series temporales nos permitirán estudiar la relación dinámica entre dos o más series. Por ejemplo, la figura 1.10 presenta dos series en el mismo gráfico. La superior corresponde a las matriculaciones de vehículos automóviles, fichero matricul.dat, y la inferior al consumo de gasolina de automoción, fichero gasolaauto.dat. Los datos son mensuales y cubren el periodo 1960-1999. Ambas series son no estacionarias y estacionales, con pautas mensuales a lo largo del año bien marcadas. Se observa que el aumento de matriculaciones en los primeros años de la muestra generó un aumento claro del consumo de gasolina, y que en el último periodo los aumentos en las matriculaciones no han generado necesariamente aumentos en el consumo. Una posible explicación de este fenómeno es que en la última parte una parte importante de las matriculaciones corresponden a reposiciones de vehículos que no producen un aumento en el parque de automóviles y, por lo tanto, en el consumo de gasolina. Entre estas dos series es esperable que exista una relación donde aumentos en las matriculaciones de automóviles produzcan aumentos en el consumo de gasolina, aunque el gráfico sugiere que esta relación puede ir variando en el tiempo. Sin embargo, no es esperable una relación en sentido contrario que vaya del consumo de gasolina a las matriculaciones.

Las relaciones entre las variables dinámicas pueden ser en ambas direcciones. Por ejemplo, la figura 1.11 presenta la evolución de la cuota de mercado semanal de los dentífricos Crest y Colgate medida en las semanas entre el 1 de enero de 1958 y el mes de abril de 1963, fichero crestcolgate.dat. Ambas series son no estacionarias con un nivel que cambia en el tiempo. Por ejemplo, en el primer periodo se observa un aumento de cuota de mercado de Crest, que va acompañado por un descenso en la cuota de Colgate. Alrededor de la observación 136 se produce un aumento muy brusco de la cuota de mercado de Crest, que pasa del 15 % al 35 %, pero este aumento no produce una disminución similar en Colgate, que sigue con una tendencia decreciente. La relación entre estas dos series es compleja y no es claro si la dirección de la relación es de Crest a Colgate, al revés o en ambas direcciones. Como veremos, en este caso existe una relación bidireccional, o de realimentación, es decir, una subida de Crest

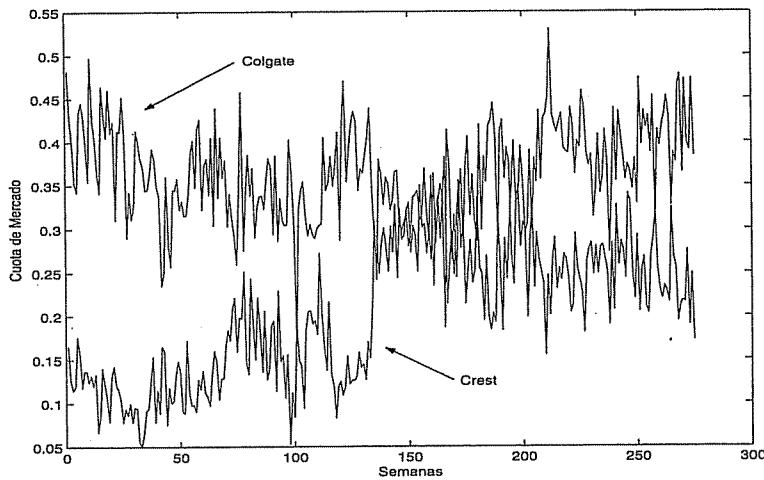
**Figura 1.10 Logaritmos del número de vehículos matriculados y del consumo de gasolina en España en el periodo 1960 a 1999**



produce una bajada de Colgate, pero también los cambios en Colgate afectan a la evolución de la cuota de Crest. Además existen intervenciones externas que afectan a las series y explican el cambio brusco de la cuota de mercado de Crest en la semana 136. En esa semana se produjo la declaración formal de la administración de Estados Unidos del resultado de las pruebas realizadas para estudiar el posible efecto positivo en la reducción de caries del fluor que Crest había comenzado a introducir en su pasta dentífrica. Se comprobó que el fluor reducía la aparición de caries, y esta declaración supuso un aumento espectacular en la cuota de mercado de Crest. Un problema importante es modelizar estos efectos externos o *efectos de intervención* dentro de la relación dinámica entre las series.

Un ejemplo adicional de dos series donde la relación dinámica se produce en ambas direcciones se presenta en las series de la figura 1.12. Las series representadas son, en la parte superior, el logaritmo de la población anual estimada de ratas almizcladas (muskrat), y en la inferior, de visones (mink) en Canadá desde 1848 a 1911. Los datos se encuentran en el fichero mink.dat. Ambos animales tienen una piel apreciada en peletería, especialmente el visón. Las dos series son no estacionarias, con un aumento gradual de la población de ratas y pequeños cambios de nivel en la población de visones. La relación entre ambas es debida a que uno es depredador del otro: como el visón se come a la rata almizclera, si aumenta la población de ratas aumenta el alimento de los visones, y su

**Figura 1.11 Cuota de mercado de los dentífricos Crest y Colgate en el periodo**



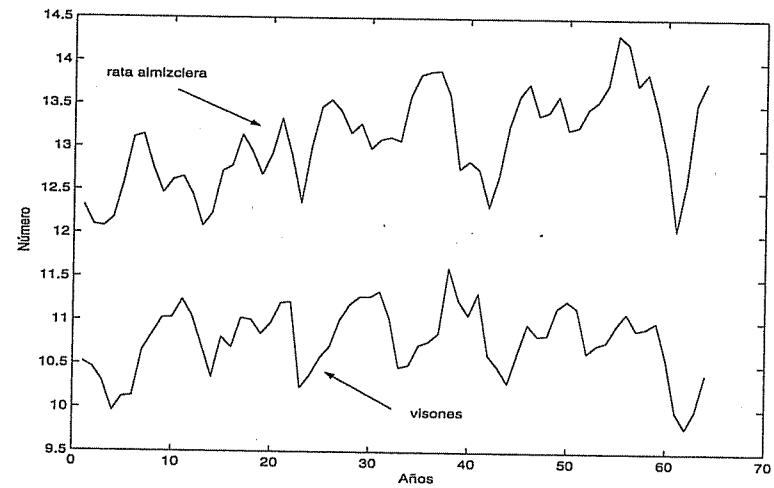
población también aumenta. Sin embargo, el aumento de la población de visones es esperable que produzca finalmente una disminución del número de ratas, con el consiguiente aumento de las dificultades de los visones para conseguir alimento, lo que se traducirá en una disminución del número de visones. Esta relación puede intuirse en el gráfico observando que ambas serie parecen tener ciclos paralelos aunque algo desplazados en el tiempo. La relación entre ambas series es bidireccional y diremos que existe realimentación entre las dos series.

Como último ejemplo, la figura 1.13 presenta los precios del trigo en cinco provincias castellanas, Ávila, Palencia, Segovia, Valladolid y Zamora. Los datos se encuentran en el fichero trigo.dat. Las series son no estacionarias y parecen evolucionar con un patrón similar. Podemos preguntarnos si existe un patrón común de evolución en todas ellas. La identificación de factores comunes en series temporales es un problema importante que abordaremos en el último capítulo de este libro.

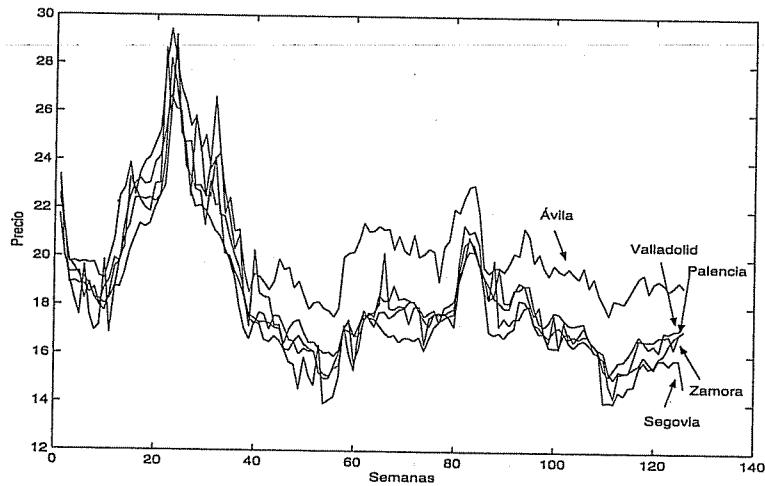
### 1.3. Contenido del libro

En este libro vamos a estudiar, en primer lugar, cómo construir un modelo para representar la evolución de una serie temporal y generar predicciones de su comportamiento futuro. Estos modelos se denominan *univariantes*, porque se basan únicamente en la historia de la propia serie, y se aplican a series como las presentadas en la sección 1.1. Las predicciones obtenidas

**Figura 1.12 Número de visones y de ratas almizcleras en Canadá (en logaritmos) desde 1848 a 1911**



**Figura 1.13 Precio medio mensual del trigo en Ávila, Segovia, Palencia, Valladolid y Zamora desde julio de 1880 a diciembre de 1890**



con estos modelos se basan en la hipótesis de que las condiciones futuras serán análogas a las pasadas y son especialmente útiles para la previsión a corto plazo.

En segundo lugar, estudiaremos métodos para encontrar la relación de

dependencia dinámica entre una serie de interés y un grupo de posibles variables explicativas. La sección 1.2 ha presentado algunos ejemplos. Las previsiones univariantes pueden mejorarse incorporando la información de la evolución de otras variables y construyendo modelos que tengan en cuenta esta dependencia. Estos modelos se conocen como modelos de *regresión dinámica* o de *función de transferencia*.

En tercer lugar presentaremos brevemente los fundamentos de los modelos para representar conjuntamente las relaciones dinámicas entre un grupo de series y obtener predicciones simultáneas de sus valores futuros. Estos modelos se conocen como modelos *multivariantes*.

En el capítulo 2 se presenta una revisión breve de métodos clásicos de análisis de series temporales. Estos métodos son casos especiales del enfoque más general que se presenta en los capítulos siguientes. El capítulo 3 introduce el marco teórico para el estudio de las series temporales, y, en general, de las variables dinámicas: los procesos estocásticos. En este capítulo se analizan los conceptos necesarios para el análisis de series temporales. Los capítulos 4 al 8 estudian la familia de procesos estocásticos más importante para el análisis univariante de series temporales: los procesos ARIMA. El capítulo 4 presenta los procesos autorregresivos y el 5 los de media móvil y la representación general de un proceso estacionario. Los capítulos 6 y 7 presentan procesos no estacionarios: los procesos integrados en el 6, y los estacionales en el 7. Además, el capítulo 6 discute un tipo de proceso que puede confundirse fácilmente con los integrados: los procesos de memoria larga. Este estudio teórico de procesos útiles para el análisis de series temporales finaliza con el capítulo 8, que estudia la generación de predicciones con modelos ARIMA.

Los tres capítulos siguientes, 9 a 11, introducen la metodología estadística para ajustar un modelo ARIMA a una serie observada. El capítulo 9 estudia la identificación de los posibles modelos, el 10 la estimación de los parámetros y selección del modelo, y el 11 la diagnosis y generación de predicciones con el modelo final. Los siguientes capítulos introducen efectos adicionales y no lineales sobre una serie temporal. El capítulo 12 estudia cómo modelar efectos deterministas, como una huelga o una catástrofe natural, de los que se conoce el instante de ocurrencia. En el capítulo 13 se relaja esta hipótesis y se pretende identificar por la historia de la serie la ocurrencia de sucesos atípicos. El capítulo 14 estudia algunos modelos no lineales en la media para series temporales y el 15 un tipo de no linealidad en la varianza que aparece mucho en series financieras y climatológicas, la heterogeneidad en la varianza condicional. Finalmente, el capítulo 16 presenta casos de estudio de series temporales univariantes y sirve de repaso y resumen al material expuesto en los capítulos anteriores.

La construcción de modelos de regresión dinámica se estudia en los capítulos 17 y 18. El primero considera las relaciones entre procesos estacionarios y extiende las ideas de regresión para estos modelos de regresión dinámica entre variables temporales. El capítulo 18 considera las relaciones entre variables no estacionarias. Finalmente el capítulo 19 introduce los modelos ARIMA vectoriales para conjuntos de series que queremos modelar conjuntamente. En este capítulo se presenta también una breve introducción al modelo factorial para series temporales.

#### 1.4. Un poco de historia

La metodología actual para analizar series temporales es, como suele ocurrir en estadística, la confluencia de varias líneas de trabajo desarrolladas en distintos campos científicos. En el caso de las series temporales pueden identificarse cinco campos de trabajo principales. El primero tiene sus raíces en los estudios de series astronómicas y climáticas, que dio lugar a la teoría de procesos estocásticos estacionarios, desarrollada por los matemáticos Kolmogorov, Wiener y Cramer en la primera mitad del siglo XX. El segundo, es el desarrollo de los métodos de alisado, inventados por investigadores operativos para prever series de producción y ventas en la década 1960-70, aprovechando las facilidades de cálculo aportadas por los primeros ordenadores. El tercero, la teoría de predicción y control de sistemas lineales, desarrollada en ingeniería de control y automática en los años 70, y estimulada por el desarrollo de la ingeniería aeronáutica y espacial. El cuarto, es la teoría de procesos no estacionarios y no lineales, desarrollada por estadísticos y económetras en los últimos veinte años del siglo XX. Finalmente, el último campo son los modelos multivariantes y los métodos de reducción de la dimensión en sistemas dinámicos, que se encuentra todavía en fase de desarrollo. De esta forma, los métodos disponibles en la actualidad para el análisis de las series temporales son deudores de las investigaciones de matemáticos, estadísticos, ingenieros, físicos y economistas durante el siglo XX para resolver problemas de predicción y control de variables y sistemas dinámicos.

Las primeras series temporales estudiadas correspondían a datos astronómicos y meteorológicos. El matemático Laplace (1749-1827), profesor de Napoleón que le hizo ministro del interior, analizó en 1823 el efecto de las fases de la luna sobre las mareas y los movimientos de aire en la tierra. Para estudiar este segundo efecto, consideró la relación entre las fases de la luna y la presión barométrica, utilizando una serie de ocho años de tres medidas diarias de la presión barométrica en París. Laplace ajustó una función sinusoidal a los datos, pero sus resultados no fueron correctos porque no tuvo en cuenta la dependencia temporal

## 1. Introducción a las series temporales

---

de las observaciones. Después de la invención del método de mínimos cuadrados por Gauss y Legandre a principios del siglo XIX, se realizan diversos ajustes de funciones sinusoidales a datos astronómicos para detectar las posibles periodicidades. Un avance importante en esta dirección fue debido a Arthur Schuster (1851-1934), físico británico que propuso en 1898 representar la amplitud de la onda ajustada a datos de espectros físicos en función de la frecuencia utilizada, y bautizó este gráfico como periodograma. La fundamentación matemática del periodograma proviene de los trabajos del matemático francés J. B. Fourier (1768-1830), que demostró a principios del siglo XIX que toda función periódica puede representarse como suma de funciones sinusoidales. El descubrimiento del concepto de regresión por F. Galton (1822-1911), primo de Darwin e investigador muy destacado en muchos campos, y del coeficiente de correlación por K. Pearson (1857-1936), permiten la utilización de estas ideas para analizar la evolución de una serie temporal. El estadístico británico G. U. Yule (1871-1951), introductor del coeficiente de correlación múltiple y de los coeficientes de autocorrelación parcial, extendió estas ideas a las series temporales proponiendo los procesos autorregresivos en 1927 para explicar la serie de manchas solares. En ese mismo año, el estadístico ruso E. E. Slutsky (1880-1948) descubrió que al tomar promedios móviles en una serie se generan artificialmente periodicidades y estudió los procesos de media móvil para representar los ciclos económicos.

Estos avances llevan a la formalización del concepto de serie temporal como proceso estocástico estacionario, debida al gran matemático ruso A. N. Kolmogorov (1903-1987), que es también el creador de la fundamentación axiomática de la probabilidad y los procesos estocásticos. Además, Komogorov extiende el análisis de su compatriota Markov (1856-1922), y estudia los procesos cuyo comportamiento futuro, conocido el presente, no depende del pasado, que se conocen en la actualidad como procesos markovianos. La primera solución general al problema de la interpolación y la predicción de series temporales para procesos estacionarios fue debida a Norbert Wiener (1894-1964) en el MIT en Boston (EE.UU.) y a A. N. Kolmogorov (1903-1987) en Moscú (Rusia) entre 1939 y 1942. Como estos descubrimientos se consideraron de gran importancia militar, se mantuvieron en secreto durante la Segunda Guerra Mundial y no se difundieron en la comunidad científica hasta los años 50. La generalización del análisis de Fourier para representar procesos estacionarios es debida al matemático sueco Cramer (1893-1985). Un estudiante suyo, Herman Wold (1908-1992), encontró en su tesis doctoral en 1938 la representación general de un proceso estacionario como una media móvil infinita. En los años posteriores a la Segunda Guerra

Mundial el estadístico británico M. S. Bartlett (1910-2002) estudió las propiedades muestrales de las autocorrelaciones de procesos estacionarios, cuyas bases estableció en 1946. Además, los trabajos de Bartlett y del estadístico estadounidense John Tukey (1915-2000), al que debemos enormes contribuciones en otros muchos campos, sentaron las bases del análisis espectral moderno. Bartlett introdujo la idea fundamental de suavizar el periodograma y Tukey, desde los laboratorios de la compañía telefónica Bell en Estados Unidos, inventa con J. W. Cooley la transformada rápida de Fourier para calcular de forma efectiva el espectro (Cooley y Tukey, 1965). Otras contribuciones importantes al campo espectral son debidas al estadounidense E. Parzen (1929), que desarrolló la estimación consistente del espectro mediante ventanas; al japonés H. Akaike (1919), que propuso la estimación del espectro ajustando modelos AR largos, y al británico P. Whittle (1929), que introduce los procedimientos de estimación de mínimos cuadrados basados en el espectro.

La aparición del ordenador impulsa el desarrollo de métodos de predicción para series reales no estacionarias basados en métodos heurísticos. Holt (1957) y Winters (1960) introducen los métodos de alisado exponencial, que tienen mucha repercusión en las aplicaciones prácticas de la predicción. La unión de los métodos de alisado y los procesos estocásticos se establece con el trabajo de Muth (1960), que demostró que los métodos de alisado eran óptimos para un tipo especial de proceso ARIMA, y, sobre todo, por las investigaciones en los años 60 de los británicos G. E. P. Box (1919) y G. Jenkins (1933-1982), que estudian el problema de la predicción y control de series temporales industriales. Fruto de sus investigaciones es su célebre libro (Box y Jenkins, 1970) que marca un hito en el análisis de las series temporales al presentar una metodología unificada para estudiar series estacionarias y no estacionarias, estacionales o no, y aplicar estos modelos en la práctica. Además, estos autores desarrollaron los fundamentos estadísticos de los sistemas de control.

Paralelamente al trabajo estadístico de Box y Jenkins en los años 60, los matemáticos e ingenieros desarrollan un enfoque general para modelar, prever y controlar sistemas dinámicos lineales. Este enfoque explica la evolución de las variables observadas en función de otras variables no observadas, que caracterizan la dinámica del sistema, y que se denominan variables de estado. R. Kalman (1930) y Busby presentan en 1962 un procedimiento mínimo cuadrático general para estimar las variables de estado y prever las observaciones futuras en sistemas lineales, hoy conocido como el filtro de Kalman. Este algoritmo generaliza resultados previos que habían aparecido para resolver casos particulares y se introduce en la estadística gracias a los trabajos de Harrison y Stevens (1976), que presentan una forma alternativa de modelar las series tem-

## 1. Introducción a las series temporales

porales mediante la formulación en el espacio de los estados y el enfoque Bayesiano. El enfoque en el espacio de los estados ha sido continuado por M. Aoki (1987) y los británicos M. West (véase West y Harrison, 1989) y A. Harvey (véase por ejemplo, Harvey, 1989) y sus colaboradores. Estos modelos aparecieron como una forma de modelar series temporales contrapuesta a los modelos ARIMA propuestos por Box y Jenkins, pero en la actualidad se reconoce el carácter complementario de ambas formulaciones. El ajuste de modelos autorregresivos a series reales planteó el problema de cómo seleccionar el orden del modelo. H. Akaike (1919), el más célebre estadístico japonés, plantea con toda generalidad este problema y proporciona una solución elegante y simple utilizando conceptos de teoría de la información, el criterio AIC (Akaike, 1976). Su trabajo, realizado inicialmente para resolver un problema de series temporales, ha tenido una influencia enorme en todas las áreas de la estadística.

Una de las contribuciones fundamentales de Box y Jenkins fue la introducción de los procesos integrados, que se convierten en estacionarios tomando diferencias. Sin embargo, la teoría de estos procesos se desarrolla posteriormente, en los años 80 y 90. Un avance significativo aparece en la tesis de J. Dickey, dirigida por W. Fuller, que desarrolla la teoría de procesos integrados de primer orden y propone un contraste de raíces unitarias (Dickey y Fuller, 1981). La teoría de los procesos integrados se ha ido completando en la última década del siglo XX, tanto desde el campo de la estadística como desde la econometría. El estudio de procesos no estacionarios vectoriales lleva al concepto de cointegración, que se define como la presencia de combinaciones lineales de series no estacionarias que forman un proceso estacionario. Esta posibilidad fue inicialmente estudiada por Box y Tiao (1977), en un trabajo pionero de gran importancia, y establecida por Granger y Engle en 1987, en una contribución con gran impacto en la economía y que ha valido a estos autores la concesión del Premio Nobel de Economía en 2003. Además, R. Engle (1943) introduce en 1982 el modelo ARCH, cuya varianza futura varía en función de los valores pasados de la serie y que ha sido generalizado por numerosos autores. Otra línea importante en los años 80 y 90 es el estudio de modelos no lineales, donde resaltaremos el trabajo de H. Tong, con los autorregresivos por umbrales, Priestley con sus modelos generales, los bilineales de Granger et al., y los modelos de memoria larga introducidos en los años 60 para datos hidrológicos por Mandelbrot y sus colaboradores. Un caso especial de no linealidad es la presencia de valores atípicos. Fox (1972) introdujo los dos tipos básicos de atípicos en series estacionarias, y G. Tiao (1983) y R. Tsay (1991) y sus colaboradores han propuesto otras clases de atípicos y diseñado métodos eficientes para modelar series con estos efectos.

El estudio de las relaciones dinámicas entre series se ha realizado con cierto retraso respecto a los análisis univariantes. La introducción de retardos en las relaciones dinámicas entre dos series es debida a R. A. Fisher, que estudió en 1923 la dependencia de la producción de trigo y la lluvia en la estación agraria de Rothamsted, en Inglaterra. Para representar los posibles retardos introdujo, por primera vez, lo que hoy conocemos como una ecuación de regresión dinámica. Paralelamente, Irving Fisher en 1925 introdujo los retardos para describir las relaciones entre variables económicas. La generalización de estas ideas para construir funciones polinómicas de retardos que pueden estimarse por mínimos cuadrados es debida a Shirley Almon, en 1965. Un trabajo pionero sobre las series temporales multivariantes fue debido a Quenouille (1924-1973), que desarrolló métodos muy avanzados para su época para el estudio de series multivariantes. Hannan (1980) estudia las series temporales multivariantes y Hannan y Deistler (1988) en el espacio de los estados. Zellner y Palm (1974) relacionan la teoría de modelos ARMA vectoriales con los modelos econométricos dinámicos desarrollados años antes. En 1980 Tiao y Box proponen un método general para construir modelos ARIMA vectoriales y en los últimos veinticinco años se han hecho avances importantes en los procedimientos para reducir la dimensión de vectores de series y modelar su dependencia temporal.

## 1.5. Programas de ordenador

Para el análisis de series temporales es imprescindible tener acceso a un paquete estadístico capaz de realizar los cálculos necesarios. Desgraciadamente, no existe un programa único de fácil acceso que permita aplicar todos los métodos presentados en este libro y en las distintas aplicaciones hemos utilizado programas distintos, que comentamos en orden creciente de complejidad.

Un programa muy simple que permite realizar el análisis univariante de una serie temporal es *Statgraphics*. Proporciona las herramientas básicas para la identificación y estimación de un modelo univariante para una serie temporal, pero no permite realizar la estimación máximo verosímil exacta, ni el tratamiento de atípicos, que veremos en los capítulos univariantes más avanzados. Tampoco permite la regresión dinámica ni el análisis multivariante. Programas algo más generales, pero también limitados al análisis univariante, son *Minitab*, que es también muy simple de usar, y *SPSS*, que tiene una complejidad similar y, en términos generales, mejores algoritmos de cálculo. Estos programas permiten construir modelos univariantes, pero no incluyen tratamiento de atípicos ni modelos no lineales. Un programa más completo es *S-plus*, que incluye muchas

posibilidades de análisis más sofisticados y que puede programarse, lo que le hace más flexible. El programa *R* es muy similar al S-plus y puede descargarse gratuitamente de la web en la dirección <http://www.r-project.org/>. El programa *TSW* está basado en los programas TRAMO y SEATS, desarrollados por Gómez y Maravall, y permite hacer análisis muy completos de series temporales univariantes. Incluye muchas capacidades, como la selección automática de modelos y el tratamiento de atípicos, y está especialmente diseñado para analizar muchas series económicas. Puede descargarse gratuitamente en la web del Banco de España en la dirección <http://www.bde.es/servicio/software/tsw.htm>. El programa *EViews* es también fácil de utilizar, incluye modelos univariantes y multivariantes y está orientado hacia la econometría. Finalmente, el programa *SCA* es el más completo de los disponibles, pero no es cómodo de usar y tiene unas capacidades gráficas limitadas, aunque incluye excelentes algoritmos para la estimación máximo verosímil, el tratamiento de atípicos y la modelización multivariante.

Los programas *Matlab* y *Gauss* son más bien lenguajes de programación muy flexibles, con muchas subrutinas estadísticas. Tienen la ventaja de ser fáciles de programar, pero son menos adecuados que los anteriores para hacer análisis rutinarios de series temporales.

En el apéndice de este capítulo se incluye una breve introducción a los programas Statgraphics, Minitab, SPSS, EViews, TSW y Matlab.

## 1.6. Lecturas complementarias

Sobre la historia de las series temporales recomendamos leer los capítulos escritos por Bartlett y Wold en el libro de Gini (1982). La biografía de Kolmogorov, de Sánchez y Valdés (2003), describe bien su trabajo en predicción. Los artículos que han publicado las revistas *Biometrika* y *JASA* durante el año 2000 para celebrar el nuevo milenio incluyen una revisión interesante del campo de las series temporales.

La literatura sobre series temporales en español es escasa. Algunas publicaciones iniciales sobre los modelos ARIMA son las monografías de Treadway et al. (1978), Espasa (1978), Peña y Sánchez-Albornoz (1983) y el libro de Uriel (1984). Dos libros recomendables con aplicaciones hacia la economía son Aznar y Trivez (1993) y Espasa y Cancelo (1993). Este último texto contiene una excelente sección sobre la modelización de relaciones entre variables económicas y análisis de coyuntura y muchos ejemplos de análisis de series temporales con datos reales. Los textos de econometría suelen contener varios capítulos de análisis de series económicas y de sus relaciones, Caridad (1998) y Novales (1993) son buenas referencias. Aznar et al. (1994) contiene numerosos ejercicios. Un

enfoque distinto de las series temporales desde el punto de vista del espacio de los estados se encuentra en la monografía de Valderrama y Ruiz (1996).

Existen excelentes textos en inglés sobre series temporales. Algunas referencias básicas, con un enfoque similar al que aquí se presenta, son Abraham y Ledolter (1983), que contiene un tratamiento muy amplio con muchos ejemplos; Anderson (1971), que tiene un enfoque teórico y se reduce a las series estacionarias; Box y Jenkins (1976), libro excelente con mucha información y una de las referencias principales sobre el tema [este libro ha sido ligeramente actualizado en Box, Jenkins y Reinsel (1994)]; Brockwell y Davis (1996), que es una excelente introducción con un nivel similar al aquí expuesto; Brockwell y Davis (1987), que contiene un tratamiento más avanzado; Gouriéroux y Monfort (1997), que está orientado a las aplicaciones económicas; Fuller (1994), que contiene muchos resultados interesantes, especialmente para procesos no estacionarios; Pankratz (1993), que incluye muchos análisis de series univariantes con datos reales, y Pankratz (1991), que contiene numerosos ejemplos de relaciones dinámicas entre dos series; Peña, Tiao y Tsay (2001), que incluye una visión amplia y actualizada de las series temporales actuales; Shumway y Stoffer (2000), un excelente libro para lectores más avanzados, y Wei (1990), que incluye un buen tratamiento introductorio de los modelos multivariantes.

Para profundizar en temas poco tratados en este libro recomendamos los de Brillinger (1975), Granger y Hatanaka (1970), Jenkins y Watts (1968) y Priestley (1981), que se dedican al análisis espectral; Harvey (1989), que presenta el enfoque en el espacio de los estados, que se discute también en Durbin y Koopman (2000) y desde el punto de vista Bayesiano en West y Harrison (1997). Un tratamiento completo de la teoría de la predicción puede encontrarse en Pourahmadi (2001). Textos orientados a las aplicaciones económicas son Hamilton (1995), Enders (1995), Franses (1998) y Hendry y Clements (1998), que se concentran en la predicción económica. Libros orientados al análisis financiero son Engle (1995), Gouriéroux (1997), Tsay (2002) y Chen (2002). Los modelos no lineales se estudian en Granger y Andersen (1980), Priestley (1988), Tong (1990) y Fang y Yao (2003). Los modelos ARMA multivariantes se estudian en Hannan (1980), Lutkepohl (1993), Reinsel (1995) y Reinsel y Velu (1998), y los modelos multivariantes en el espacio de los estados en Aoki (1988) y Hannan y Deistler (1988).

## Apéndice 1.1: Descripción de la entrada de datos con algunos programas de ordenador

Supondremos que los datos de una serie temporal se encuentran en un fichero con nombre losdatos.dat y que la serie aparece en columnas, comenzando con el dato más antiguo y finalizando con el más reciente, sin encabezamientos ni formatos especiales. Los ficheros de datos que se utilizan en el libro tienen esta estructura. Cada serie tiene asociados dos ficheros. El de los datos, que tiene como nombre el de la serie y la extensión .dat, y otro, con el mismo nombre pero extensión .txt, que describe el periodo y la fuente de los datos. Estos ficheros se denominan ascii y no tienen formatos especiales, de manera que son interpretados de igual forma por distintos programas de ordenador. Algunos de los ficheros .dat del libro contienen varias series, que entonces se encuentran siempre en columnas separadas por espacios.

Los programas que presentamos a continuación se inician haciendo doble-click sobre el ícono del programa. El programa se iniciará pero no podemos contemplar nada en la pantalla excepto una línea superior de comandos. Lo primero es introducir los datos de la serie.

### Statgraphics

Al iniciar el programa aparece un menú en la parte superior de la ventana. En el botón de *File* marcamos *Open*, seguido de *Data File*. Con esto se abre una ventana donde podemos buscar en el PC el fichero de datos. Para facilitar la búsqueda conviene poner en el recuadro de *Tipo* de fichero la opción *All Files*. Cuando se introduce el nombre del fichero aparece otra ventana donde hay que indicar cómo están los datos en el fichero. Marcar *Blank Delimited* en la opción de *File Type* y *Generate Defaults* en la opción *Variable names*. Aparece una hoja de donde la serie (o series) introducida se muestra en una columna (o en varias).

Para ver un gráfico de la serie ir a:

*Plot* → *ScatterPlots* → *Univariate Plot*.

Si hacemos click en el gráfico, este se abre y podemos editarla. En el menú *Plot* tenemos otras opciones, como hacer un histograma de los datos u otros gráficos.

### Minitab

En el menú superior marcar *File* → *Open Worksheet* y se abre una ventana donde puede seleccionarse el nombre del fichero y el tipo de fichero. En esa ventana aparece también la opción *Buscar en...* para buscar el fichero dentro

del ordenador, y la opción *Tipo* donde debe especificarse que el fichero es .dat. Una vez especificado el nombre del fichero, se activa en la ventana la opción *Options...* y en la primera ventana debe especificarse lo siguiente:

*Variable names* → *None*.

*First row of data* → *Automatic*.

*Decimal separator* → *Period*.

*Field definition* → *single character definition* → *space*.

Después de esto marcar *OK* y luego *abrir*. Aparecerán las variables en columnas con los nombres C1, C2, en la *Current Worksheet*. Para hacer un gráfico temporal de la serie ir al menú superior, marcar *Graph* → *Time series plot* y dentro de la ventana el nombre de la serie (C1,C2,...) que se quiera dibujar. Una vez que aparezca el gráfico, este puede guardarse en un fichero con la opción del menú principal *File* → *Save Graph as...*, que permite guardarle en un fichero.

## SPSS

La introducción de datos en SPSS es similar a Statgraphics y Minitab. Una vez abierto el programa, y suponiendo que existe un fichero nombre.dat que incluye los datos de una serie temporal en columnas, en el menú superior marcar: *Archivo* → *Abrir datos* y se abre una ventana donde puede seleccionarse el nombre del fichero y el tipo de fichero. En esa ventana aparece también la opción *Buscar en...* para buscar el fichero dentro del ordenador, y la opción *Tipo* donde debe especificarse que el fichero es .dat. Una vez especificado el nombre del fichero, se van abriendo ventanas donde debe especificarse el tipo de fichero como sigue:

*¿Archivo con texto predefinido?* → *No*.

*¿Cómo están organizadas las variables?* → *Delimitadas*.

*¿Está incluido el nombre de las variables?* → *No*.

A continuación hay que indicar que los datos comienzan en la primera línea y que cada línea representa un caso y que el delimitador de las variables es el espacio. Como calificador de texto poner comillas dobles. En la ventana siguiente podemos dar nombre a las variables si lo deseamos, en otro caso aparecen con los nombre genéricos VA, V2,..., También debemos asegurarnos que aparecen como numéricas y no como cadena. Para hacer un gráfico temporal de la serie ir al menú superior, marcar *Graph* → *Secuencia* y dentro de la ventana el nombre de la serie (V1,V2,...) que se quiera dibujar.

### EViews

La introducción de datos en este programa es algo distinta que en los tres anteriores. Todos los objetos con los que trabaja deben estar en un fichero de trabajo llamado workfile, por lo que el primer paso en cualquier análisis será la creación de un nuevo workfile o la apertura de un workfile creado previamente. Para definir una variable en un workfile hay que dar tres características: nombre de la serie, frecuencia y rango. La frecuencia define el intervalo temporal entre observaciones, que en EViews puede ser anual, semestral, trimestral, mensual, semanal y diario, con 5 o 7 días a la semana. El rango es un par de fechas o números de observación que describen el inicio y el fin de los datos. El programa utilizará estas dos características teniendo en cuenta un calendario interno: si un año es bisiesto y el programa está diseñado para contener dicha información. El programa permite trabajar con datos no fechados, es decir, observaciones que están simplemente numeradas consecutivamente.

La manera más sencilla de crear un workfile es seguir el siguiente camino en el menú principal: *File* → *New* → *Workfile*. Aparecerá un cuadro de diálogo. A continuación seleccionamos la frecuencia apropiada y posteriormente introducimos la fecha inicial (*Start date*) y la fecha final (*End date*). Dichas fechas pueden ser posteriormente modificadas, como veremos más adelante. Esto nos permitirá trabajar con series con diferentes rangos. Las normas para especificar los datos son las siguientes:

1. Datos anuales (Annual): se especifican los años. Por ejemplo, 1920 y 2003, para una serie cuyo primer dato sea en 1920 y el último sea en 2003.
2. Datos semestrales (Semi-annual), trimestrales (Quarterly) o mensuales (Monthly): se indica primero el año seguido de dos puntos y el número del semestre/trimestre o mes. Por ejemplo, 1920:1 y 2003:2, para datos que empiezan en el primer semestre de 1920 y acaban en el segundo semestre de 2003, y 1920:1 y 2003:12 para datos mensuales que comienzan en enero de 1920 y finalizan en diciembre de 2003.
3. Datos semanales (weekly), o diarios (daily, 5 day weeks; daily, 7 day weeks): se especifica el mes, el día y el año, cada uno separados por dos puntos. Por ejemplo, con datos semanales 9:1:2003 indica que la fecha es el 1 de septiembre de 2003, que corresponde a la primera semana de septiembre del año 2003, y con datos diarios 8:10:1997 indica que la fecha es el 10 de agosto de 1997.

Una vez especificadas las fechas marcamos en OK. Por ejemplo, especificamos Quarterly, 1955:1 y 1996:4, es decir datos trimestrales entre el primer trimestre de 1955 y el cuarto de 1996. Aparecerá el workfile creado, que como indica en su parte superior no tiene nombre (Workfile: UNTITLED). Vemos que han aparecido dos iconos, c y resid, que más adelante detallaremos.

## Análisis de series temporales

---

Aquí vemos la estructura del workfile. Una línea de comandos, información del rango y el contenido del workfile.

Empezamos salvando el workfile, lo que requiere asignarle un nombre. Para ello, en el menú principal: *File → Save as* y aparecerá una ventana de windows para asignar un nombre, por ejemplo, *practica1*. El programa creará un fichero de nombre *practica1.wfl* en el directorio que queramos salvar el workfile. Como tenemos salvado el workfile, podemos cerrarlo: *File → Close* y el workfile desaparecerá. Para abrirlo seguiremos el siguiente camino: *File → Open → Workfile → practica1.wfl* en el directorio en el que guardamos el archivo.

A continuación pasemos a variar el rango del workfile. Por ejemplo, si tenemos una serie de datos y queremos obtener predicciones, debemos ampliar el rango para dar cabida a las nuevas observaciones. Para ello, seguimos el camino en el menú del workfile: *Procs → Change Workfile Range* y especificamos las nuevas fechas.

Hay dos maneras principales de crear una serie temporal en EViews. La primera es escribiendo los datos directamente: Para crear una serie seguimos el siguiente menú en el workfile:

*Object → NewObject → Series → Name:serie*

y marcando en OK, tendremos un nuevo objeto en el workfile con el nombre serie. Haciendo doble click sobre él veremos que todos los datos aparecen con la marca NA (not available), es decir, no tenemos datos numéricos en la serie. Para introducir los datos, marcamos la celda:

*Edit +/-*

y podemos viajar por todas las celdas e incluir los valores numéricos que queramos. Una vez que hayamos acabado, volvemos a marcar en la celda de Edit y podremos cerrar la serie, que permanecerá en el workfile.

La segunda forma es importando un fichero con los datos, que puede ser de texto o de excél. El camino es:

*File → Import → Read Text-Lotus-Excel*

y abrimos el fichero anterior. Aparecerá un cuadro de diálogo. Marcamos la opción:

*Data Order - in columns*

y le asignamos el nombre de la serie. Marcamos OK y veremos cómo en el workfile aparece la serie con el nombre asignado. Haciendo doble click en ella podemos comprobar que hemos importado la serie sin problemas. Salvamos el workfile asignándole el nombre *serie.wfl*.

1. Podemos generar nuevas series a partir de una dada. Por ejemplo, si en el menú del workfile, hacemos el camino:

*Genr → logserie = log(nombre de la serie)*

## 1. Introducción a las series temporales

lo que obtenemos es una nueva serie definida como el logaritmo de la serie original. La serie estará definida en el mismo rango que la serie original. Igual que la operación logaritmo podemos utilizar otras muchas operaciones básicas: sumas, restas, productos, raíces, etc. ...

En el workfile creado en el menú podemos observar diferentes celdas. La primera es view, que está dividida en diferentes opciones.

1. La primera opción del primer bloque es: *View→SpreadSheet* y vemos el fichero con los datos.
2. La segunda opción es gráfica *View→Line Graph*, para ver el gráfico de la serie.
3. A continuación *View→Bar Graph* para ver los datos por separado, que no se permite en el gráfico anterior.
4. Con *View→Descriptive Statistics→Histogram and Stats*, vemos el histograma y algunos estadísticos descriptivos: la media, mediana, máximo, mínimo, desviación típica (normalizada por  $T - 1$ , donde  $T$  es el número de datos), y coeficientes de asimetría y curtosis.
5. En quinto lugar: *View→Tests for descriptive stats→Simple Hypothesis Tests* para contrastar hipótesis respecto a la media de la serie.
6. Finalmente: *View → Distribution Graphs→Quantile-Quantile: Normal distributions*, proporciona el gráfico de comparación de cuantiles respecto a la distribución Normal. Si los datos son normales, observamos aproximadamente una línea recta.

### TSW

El formato de entrada de este programa exige que en la primera fila del fichero de datos aparezca el nombre de la serie y en la segunda cuatro variables numéricas: número de datos en la serie, año inicial de la serie, periodo inicial de la serie y número de periodos o datos que forman un año. A continuación vienen los datos de la serie en una columna. Por ejemplo, si tenemos 100 datos de una serie que comienza en enero de 1980 y los datos son mensuales, la segunda fila del fichero de datos debe contener las variables:

100 1980 1 12

y si tenemos 28 datos cuatrimestrales comenzando en el tercer cuatrimestre de 1990, escribiremos:

28 1990 3 4

Al iniciar el programa aparece una ventana donde en la parte superior está un menú de botones y a la derecha ranuras para introducir información.

Para cargar los datos hay que seleccionar en el menú **Series** y entonces se abre una ventana que nos permite seleccionar la serie moviéndonos en las carpetas del PC. Al leer correctamente, el nombre de la serie aparece en la ventana principal central debajo de **Series list**. Esto confirma que tenemos la serie incluida en el programa. Si marcamos la serie aparece el gráfico en la pantalla y en la parte derecha su nombre y ciertos atributos que podemos cambiar. Por ejemplo, el parámetro **Iter**, si lo cambiamos a **Iter=1**, podemos ajustar a la misma serie diferentes modelos. Es conveniente marcar la opción **tramo/seats**, que es la que aparece por defecto en el programa.

### Matlab

Los datos se introducen en Matlab escribiendo la sentencia: *load losdatos.dat*, donde losdatos es el nombre del fichero que incluye los datos. Matlab lee todas las columnas del fichero y crea con ellos una matriz a la que asigna el nombre del fichero. Si la variable de interés es, por ejemplo, la tercera columna, podemos extraerla de la matriz con la sentencia: *x=losdatos(:,3)*, que interpreta como tomar todas las filas de la columna 3 de la matriz losdatos y asignarle a la variable x. Podemos hacer un gráfico temporal con la sentencia: *plot(x)*.

---

## 2. Análisis descriptivo de una serie temporal



Sir Arthur Schuster (1851–1934)

Físico inglés, nacido alemán. Fue profesor de matemáticas y física en la Universidad de Manchester. Hizo contribuciones fundamentales a la espectroscopía y al estudio de las periodicidades e inventó el periodograma. Dirigió en 1875 la expedición de la Royal Society para observar el eclipse solar en Tailandia y fué nombrado Lord por sus méritos científicos.

### 2.1. Introducción

En este capítulo presentamos los procedimientos descriptivos desarrollados entre 1940 y 1970 para analizar series temporales. Su objetivo es explicar la evolución pasada de la serie en términos de pautas simples y prever sus valores futuros. Estos procedimientos tienen poca utilidad hoy en día, ya que disponemos de métodos más generales y eficaces que se exponen en los capítulos siguientes y que los incluyen como caso particular. De todas formas, hemos considerado interesante presentarlos aquí por dos razones. En primer lugar, los métodos actuales se basan en ellos y se entienden mejor si se conocen las limitaciones de estos procedimientos más simples. En segundo lugar, están todavía disponibles como opciones de análisis en muchos programas de ordenador de uso frecuente, por lo que es conveniente que el lector conozca sus bases y sus limitaciones.

Los métodos descriptivos para series temporales son generalizaciones de los desarrollados para variables estáticas. Por ejemplo, la media de una variable estática, como las estaturas de los estudiantes en una clase, estima la media de la población de estudiantes y tiene una interpretación

clara. Sin embargo, si hacemos la media de los datos de una serie con tendencia, esta media no refleja ninguna característica constante y tiene difícil interpretación. Esto llevó a plantear los modelos de análisis de tendencias deterministas, que constituye una extensión inmediata de los métodos de regresión. Aunque estos métodos pueden ser útiles para la descripción simple de las pautas en una serie, las predicciones que proporcionan suelen tener un alto error, lo que mostró enseguida sus limitaciones. Esto es debido a que en una serie temporal la observación hoy,  $z_t$ , depende en general de sus valores previos,  $z_{t-1}, z_{t-2}, \dots$ , pero esta dependencia suele ser más fuerte con los datos más recientes y más débil con los más alejados. Como los modelos de análisis de tendencias deterministas proporcionan predicciones que no utilizan esta propiedad, se propusieron como alternativa los métodos de alisado, que realizan predicciones imponiendo una estructura donde la dependencia entre las observaciones disminuye con el tiempo. Estos métodos se introdujeron en los años 60 y se extendieron mucho en la práctica por sus buenos resultados y su facilidad de cálculo, gracias a la aparición del ordenador digital.

La extensión para series estacionales se hizo modelando separadamente la evolución de la serie y el efecto estacional. Los análisis de tendencias deterministas dieron lugar a los métodos de decomposición y los métodos de alisado a los métodos de doble alisado, uno para la evolución de la serie y otro para el efecto estacional, como el método de Holt-Winters. Paralelamente, se desarrollaron procedimientos para series cíclicas, como son las series climáticas, donde domina el efecto estacional, y la serie se representa como una suma de efectos sinusoidales. Esto dio lugar al periodograma, que es una herramienta muy útil para detectar ondas deterministas en una serie temporal.

## 2.2. El análisis de tendencias deterministas

### 2.2.1. Tipos de modelos

Comenzaremos presentando este método para series sin estacionalidad, como las de las figuras 1.1 a 1.5 del capítulo anterior. Se fundamenta en modelar la dependencia temporal mediante las ideas de regresión para relacionar variables estáticas. Para ello, una serie observada,  $z_t$ , donde  $t = 1, \dots, T$ , se representa mediante suma de dos componentes

$$z_t = \mu_t + a_t. \quad (2.1)$$

El primer componente

$$\mu_t = f(t, \beta)$$

es el *nivel de la serie* que es una función conocida determinista del tiempo que depende del instante considerado y de un vector de parámetros,  $\beta$ .

## 2. Análisis descriptivo de una serie temporal

Estos parámetros deben estimarse a partir de los datos, como veremos a continuación. El segundo componente,  $a_t$ , suele denominarse la *innovación* y es un componente aleatorio que recoge todos los demás efectos que actúan sobre la serie. Se supone que las variables aleatorias  $a_t$  tienen una estructura estable a lo largo del tiempo: media cero, varianza constante y distribución normal. Además, las observaciones correspondientes a dos períodos de tiempo distintos son independientes, es decir, conocer el valor de  $a_t$  no proporciona ninguna información sobre el posible valor de  $a_{t+1}$ : sea cual sea el valor de  $a_t$ , la variable  $a_{t+1}$  tiene una distribución normal con media cero y varianza constante.

La predicción de la serie con este modelo en un periodo futuro,  $T + k$ , se obtiene extrapolando el nivel de la serie  $\mu_t$ , ya que la predicción de la innovación es su esperanza, que es siempre cero. Llamando  $\hat{z}_T(k)$  a la predicción realizada desde el origen  $T$  para  $k$  períodos en adelante, es decir la predicción del valor  $z_{T+k}$  con la información disponible hasta el momento  $T$ , tenemos que

$$\hat{z}_T(k) = \mu_{T+k} = f(T + k, \beta). \quad (2.2)$$

La forma que establezcamos para la evolución del nivel de la serie a lo largo del tiempo determina el modelo concreto utilizado. Por ejemplo, el modelo más simple supone que el nivel de la serie es constante en el tiempo, es decir  $\mu_t = \mu_{t-1} = \mu$ , y se conoce como el modelo de nivel constante o sin *tendencia*. Entonces la ecuación (2.1) se reduce a:

$$z_t = \mu + a_t \quad (2.3)$$

y la serie oscilará alrededor de su media,  $\mu$ , que es constante. Las series de las figuras 1.1 y 1.2 no tienen tendencia y podrían explicarse por este modelo. Al ser el nivel de la serie constante, y no depender de  $t$ , y como el valor esperado de la innovación es cero, la predicción con este modelo para cualquier horizonte será la media,  $\mu$ .

Un modelo más general, que se aplica a series que tienen tendencia, creciente o decreciente, es el modelo de *tendencia lineal*. La figura 1.3 presenta una serie que podría tener esta propiedad. El modelo para  $\mu_t$  en (2.1) es:

$$\mu_t = \beta_0 + \beta_1 t, \quad (2.4)$$

donde ahora  $\beta_1$  representa la pendiente de la recta que describe la evolución de la serie. Esta pendiente corresponde con el crecimiento esperado entre dos períodos. La predicción con este modelo del valor de la serie en el instante  $T + k$  con información hasta  $T$ , será:

$$\hat{z}_T(k) = \beta_0 + \beta_1(T + k). \quad (2.5)$$

Estos dos modelos son casos particulares del modelo de *tendencia polinómica*, donde el nivel de la serie evoluciona según un polinomio de orden  $r$ :

$$\mu_t = \beta_0 + \beta_1 t + \dots + \beta_r t^r. \quad (2.6)$$

Los dos casos estudiados anteriormente corresponden a  $r = 0$ , con lo que obtenemos el modelo de nivel constante, y a  $r = 1$ , que conduce al modelo de tendencia lineal. En la práctica  $r \leq 2$ .

### 2.2.2. Estimación

El ajuste de estos modelos a una serie temporal requiere la estimación del vector de parámetros,  $\beta = (\beta_0, \dots, \beta_r)$ . Las estimaciones se obtienen con el criterio de *mínimos cuadrados*, es decir, minimizando las diferencias entre los valores observados y los previstos a horizonte uno por el modelo. Este criterio escribe como

$$\text{minimizar} \sum_{t=1}^T (z_t - \mu_t)^2. \quad (2.7)$$

Para modelos con tendencias polinómicas, los valores de los parámetros que minimizan esta expresión se obtienen fácilmente. Por ejemplo, en el modelo de nivel constante  $\mu_t = \mu$ , derivando respecto al parámetro,  $\mu$ , y llamando  $\bar{z}_t = \sum z_t / T$  a la media de los datos observados, se obtiene inmediatamente que  $\hat{\mu} = \bar{z}_t$ . En el caso de tendencia lineal la función a minimizar es:

$$\text{minimizar} \sum_{t=1}^T (z_t - \beta_0 - \beta_1 t)^2, \quad (2.8)$$

y los estimadores se obtienen derivando respecto a  $\beta_0$  y  $\beta_1$  e igualando a cero las derivadas. Así obtendremos un sistema de dos ecuaciones y dos incógnitas cuya solución proporciona los estimadores. Vamos a suponer, para simplificar la exposición, que la variable  $t$  se construye para que tenga media cero, y que tenemos  $T = 2n+1$  datos (periodos observados), de manera que podemos definir  $t$  como  $t = (-n, -n+1, \dots, -1, 0, 1, \dots, n-1, n)$ . Entonces, comenzando con la ecuación obtenida al igualar a cero la derivada de (2.8) respecto a  $\beta_0$  y llamando  $\hat{\beta}_0$  y  $\hat{\beta}_1$  a los valores que verifican esta ecuación, se obtiene:

$$\sum_{t=1}^T (z_t - \hat{\beta}_0 - \hat{\beta}_1 t) = 0. \quad (2.9)$$

Llamaremos *residuos* a las estimaciones de las innovaciones,  $\hat{a}_t$ , que según las ecuaciones (2.1) y (2.4) vienen dadas por

$$\hat{a}_t = z_t - \hat{\beta}_0 - \hat{\beta}_1 t.$$

## 2. Análisis descriptivo de una serie temporal

---

La ecuación (2.9) exige que los residuos estimados tengan media cero. Como  $\bar{t} = 0$  su solución es:

$$\hat{\beta}_0 = \bar{z}_t. \quad (2.10)$$

Esta estimación de la ordenada en el origen coincide con el nivel estimado en el modelo constante, ya que si  $\beta_1 = 0$ , entonces  $\mu_t = \beta_0 = \mu$ . Para estimar la pendiente, derivando en (2.8) respecto a  $\beta_1$  e igualando a cero el resultado, tenemos que:

$$\sum_{t=1}^T (z_t - \hat{\beta}_0 - \hat{\beta}_1 t) t = 0, \quad (2.11)$$

y esta ecuación exige que la covarianza entre los residuos y la variable tiempo sea cero, es decir, los residuos no deben tener relación lineal con el tiempo. Sustituyendo (2.10) y despejando  $\hat{\beta}_1$  se obtiene, dado que  $\bar{t} = 0$ , el estimador de la pendiente:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T (z_t - \bar{z}_t) t}{\sum_{t=1}^T t^2}. \quad (2.12)$$

A continuación, presentamos dos ejemplos de la utilización de estos modelos.

---

### Ejemplo 2.1

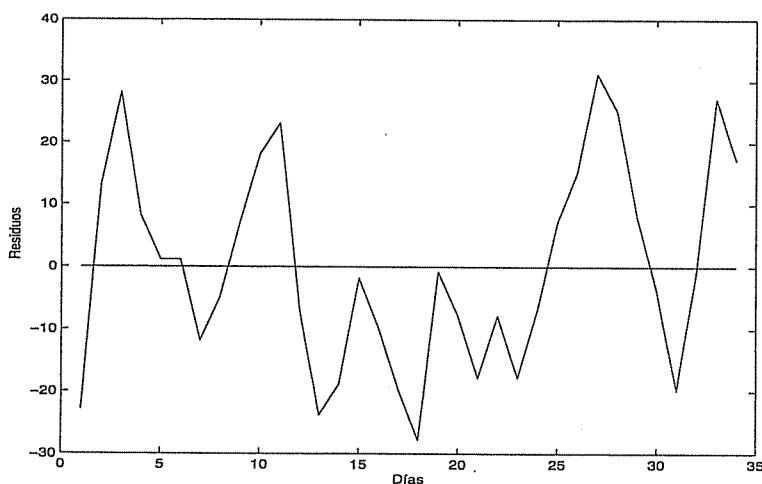
Vamos a ajustar el modelo de nivel o media constante (2.3) a la serie de las leguas diarias recorridas por la flota de Colón que se encuentra en el fichero colon.dat. La media de los datos observados de la serie es

$$\bar{z}_t = \frac{9 + \dots + 49}{34} = 31.82$$

que es, además, según el modelo (2.3) la predicción de la distancia que se recorrerá el día siguiente. Los errores de esta predicción dentro de la muestra son los residuos,  $\hat{a}_t$ , estimados como diferencia entre el valor de la serie y su media, que se presentan en la figura 2.1. La variabilidad de estos residuos mide el error esperable con el modelo utilizado. Como los errores pueden ser positivos y negativos y nos interesa su magnitud, calcularemos su desviación típica, promediando los errores al cuadrado y tomando después la raíz cuadrada. Se obtiene entonces

$$\hat{\sigma}_a = \sqrt{\frac{(-22.82)^2 + \dots + (17.18)^2}{34}} = 16.6$$

**Figura 2.1 Residuos del modelo de tendencia constante para la serie de leguas recorridas por Colón**



que indica que el error promedio en la predicción con este modelo es de 16.6 leguas diarias.

### Ejemplo 2.2

Vamos a ajustar el modelo de tendencia lineal a los datos de la población española de mayores de 16 años de la figura 1.4 que se encuentran en el fichero pobmay16.dat. Para crear una variable temporal de media cero, como hay 96 datos definimos primero una variable que tome los valores  $1, \dots, 96$  y a continuación restamos a cada uno de estos valores la media de la secuencia, 48.5, de modo que el conjunto de datos tenga media cero. Esta operación proporciona los valores de  $t = (-47.5, -46.5, \dots, -0.5, 0.5, \dots, 47.5)$ . Utilizando (2.10) y (2.12), se obtiene la ecuación de predicción

$$\hat{z}_t = 29693 + 79.7t.$$

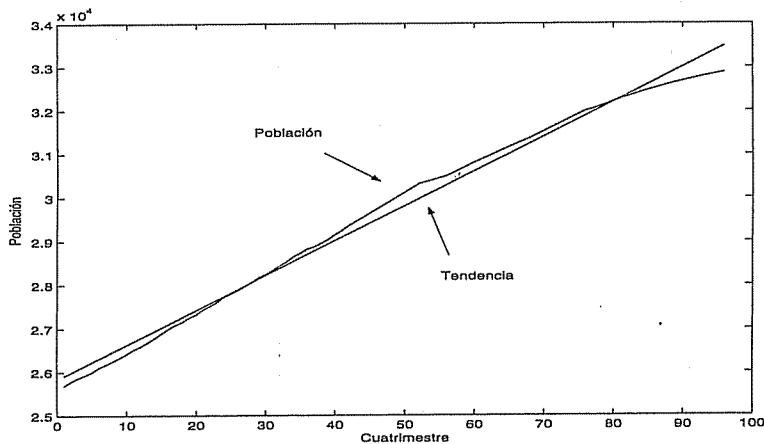
Esta recta indica que, en el periodo estudiado, el número de mayores de 16 años es en promedio 29.693 millones de personas y que cada año este número se incrementa en 79.700 personas, aproximadamente. El modelo parece tener un excelente ajuste, ya que el coeficiente de correlación es .995.

Sin embargo, la figura 2.2, que muestra los datos y el modelo ajustado, indica que un ajuste de crecimiento lineal no es bueno, porque la tendencia no es exactamente constante y ha ido cambiando suavemente en el tiempo. En particular, las predicciones generadas en el año 2000 para los dos años siguientes van considerablemente por arriba de los datos observados. Este pro-

## 2. Análisis descriptivo de una serie temporal

---

**Figura 2.2 Población mayor de 16 años en España y el modelo lineal ajustado**



blema aparece más claramente en la figura 2.3, que presenta los residuos de este modelo calculados mediante:

$$\hat{a}_t = z_t - 29693 - 79.7t.$$

Podemos observar que estos residuos tienen una pauta de variación: primero negativos, luego positivos y de nuevo negativos, lo que lleva a concluir que el modelo está mal especificado. Podría pensarse que el problema es que la tendencia sigue un polinomio de segundo grado y que la forma de los residuos, indica la necesidad de una ecuación de segundo grado para reflejar esta curvatura. Si estimamos un modelo de segundo grado, la ecuación resultante es:

$$z_t = 29910 + 79.7t - 0.2823t^2.$$

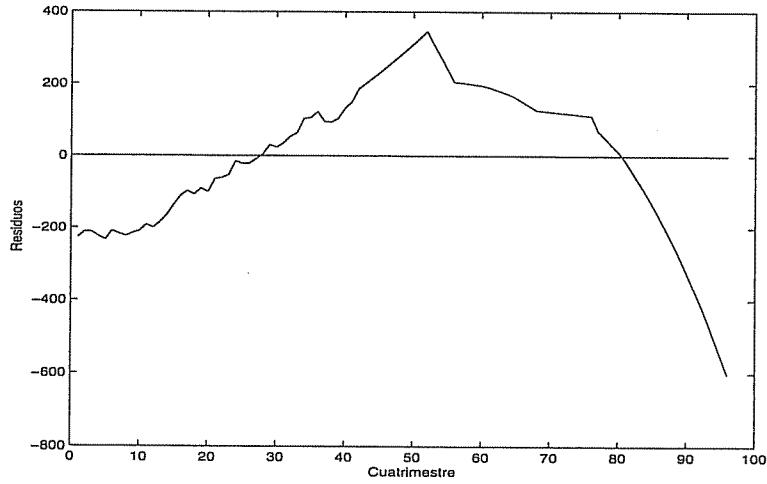
Los residuos de este modelo se presentan en la figura 2.4, donde se observa que aumentar el grado del polinomio no resuelve el problema. Los errores de predicción para los últimos datos muestrales son muy altos, y puede comprobarse que las predicciones de este modelo para los datos de 2001 son muy malas. El problema es la falta de flexibilidad de los modelos deterministas que no permiten que la predicción de un dato futuro se base principalmente en los últimos valores observados de la serie.

---

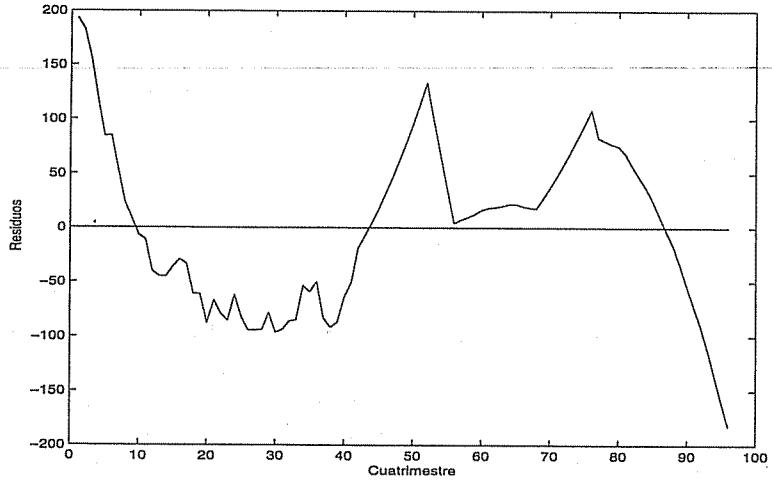
### 2.2.3. Limitaciones del ajuste de tendencias deterministas

Como se ha visto en los ejemplos anteriores, la limitación principal de estos métodos es que, aunque las series con nivel constante son frecuentes,

**Figura 2.3 Residuos del modelo lineal para la población española de mayores de 16 años**



**Figura 2.4 Residuos del ajuste cuadrático a los datos de la población mayor de 16 años**



es muy raro que una serie real tenga una tendencia determinista lineal, o en general polinómica con  $r \geq 1$ . Por ejemplo, la evolución de la serie de nacimientos en la figura 1.4 o la de los precios del trigo en la figura 1.5 no pueden representarse por tendencias polinómicas. Una posibilidad es intentar ajustar tendencias lineales por tramos, es decir dividir la serie en

## 2. Análisis descriptivo de una serie temporal

---

tramos que tengan, aproximadamente, tendencia constante y ajustar en cada tramo un modelo lineal o constante. Aunque estos modelos suponen un claro avance para explicar la evolución histórica de algunas series, son menos útiles para predecir los valores futuros, ya que no sabemos cuántas observaciones pasadas utilizar para ajustar el nivel futuro de la serie.

Un inconveniente adicional de ajustar una tendencia lineal por tramos es que el modelo de crecimiento implícito es poco razonable. Para justificar esta afirmación, veamos cuáles son las implicaciones de suponer que una serie sigue un modelo con tendencia lineal determinista en un intervalo. En este caso, la predicción del crecimiento de la serie entre  $z_t$  y  $z_{t+1}$ , medido por  $z_{t+1} - z_t$  será la pendiente estimada,  $\hat{\beta}_1$ . Hemos visto que la pendiente se estima por la ecuación (2.12), y puede demostrarse (véase Peña, 1995) que esta estimación puede siempre escribirse como una media ponderada de los crecimientos experimentados por la serie en el intervalo donde estamos ajustando la recta. Las ponderaciones que se aplican a los crecimientos pasados para prever el crecimiento futuro son tales que el mayor peso se atribuye al crecimiento en el centro del periodo observado, y el peso mínimo al crecimiento en los extremos del intervalo. Es decir, llamando  $\nabla z_t = z_t - z_{t-1}$  al crecimiento observado de la serie en el instante  $t$ , se demuestra que

$$\hat{\beta}_1 = \sum_{t=2}^T w_t \nabla z_t, \quad (2.13)$$

donde los pesos  $w_t$  son números positivos que suman uno y que son simétricos respecto al centro de la muestra. Los valores máximos de  $w_t$  corresponden a los incrementos centrales y van disminuyendo simétricamente hasta alcanzar el valor mínimo en los extremos.

Por ejemplo, supongamos una serie  $z_t$  con 5 datos en los instantes,  $t = 0, \pm 1, \pm 2$ , donde hemos construido la variable tiempo de manera que  $\bar{t} = 0$ , para simplificar la presentación. Llamemos entonces a los valores de la serie  $z_{-2}, z_{-1}, z_0, z_1, z_2$ . Aplicando las fórmulas (2.10) y (2.12) para el cálculo de la pendiente, se obtiene que:

$$\begin{aligned} \sum_{t=-2}^2 t^2 &= (-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 = 10, \\ \sum_{t=-2}^2 t z_t &= -2z_{-2} - z_{-1} + z_0 - 2z_1 = 2(z_{-1} - z_{-2}) + 3(z_0 - z_{-1}) + \\ &\quad + 3(z_1 - z_0) + 2(z_2 - z_1), \end{aligned}$$

y por tanto:

$$\hat{\beta}_1 = 0.2\nabla z_{-1} + 0.3\nabla z_0 + 0.3\nabla z_1 + 0.2\nabla z_2,$$

que es un caso particular de (2.13). Esta ecuación indica que la predicción del crecimiento futuro es una media ponderada de los crecimientos observados en cada uno de los períodos, con pesos simétricos respecto al centro y tales que el peso mínimo corresponde a los crecimientos en los períodos extremos.

La implicación de este resultado es la siguiente: si admitimos que una serie tiene una tendencia determinista lineal en un intervalo estamos diciendo que la predicción futura de su crecimiento debe hacerse ponderando los crecimientos observados, pero dando peso mínimo al último valor observado. Este peso es, además, igual al que se atribuye al crecimiento más alejado en el tiempo, es decir, al primer crecimiento observado en la muestra. Además, si aumentamos el tamaño del intervalo, el peso de los crecimientos en los últimos años disminuye, pero siempre manteniéndose igual al peso de los primeros crecimientos observados. Esto es claramente poco razonable en muchas ocasiones. Por ejemplo, si admitiésemos que la población crece linealmente y tuviésemos sólo los últimos cinco años de datos, el incremento de población con respecto al último valor observado se obtendría sumando un crecimiento previsto que se obtiene dando pesos 20 %, 30 %. 30 % y 20 % a los cuatro crecimientos anuales observados en los cinco años. Sin embargo, si aumentamos el tamaño muestral y pasamos a tener una muestra de 100 datos anuales, y ajustamos el modelo en esta muestra, el peso del último año pasa a tener un valor muy pequeño, e igual al crecimiento observado hace 100 años.

### 2.3. Métodos de alisado

Para resolver las limitaciones de los modelos con tendencias deterministas se introdujeron en los años 60 los métodos de alisado. La idea de estos métodos es permitir que los últimos datos de la serie tengan más peso en las predicciones que los valores más antiguos. Esto se consigue permitiendo que los parámetros del modelo de tendencias deterministas no sean constantes, sino que puedan ir variando en el tiempo. De esta manera vamos a ver que se obtienen modelos que generan predicciones dando mayor peso a los valores más recientes y menor a los más alejados en el tiempo.

#### 2.3.1. El modelo de alisado simple

Supongamos que hemos hecho una predicción del valor de una variable para el periodo  $T$  que llamaremos  $\hat{z}_T$  y observamos después su valor,  $z_T$ . ¿Cómo generar la próxima predicción? Holt (1956) propuso hacer una combinación lineal de la última predicción y el último valor observado,

## 2. Análisis descriptivo de una serie temporal

---

de manera que la predicción del próximo periodo,  $T+1$ , viene dada por:

$$\hat{z}_{T+1} = \theta \hat{z}_T + (1 - \theta) z_T, \quad (2.14)$$

donde  $0 < \theta < 1$  determina el peso que damos a cada uno de los dos componentes para generar las predicciones. Si tomamos  $\theta$  próximo a la unidad, las predicciones para distintos períodos son muy similares, y se modifican poco con la nueva información. En el caso límite de  $\theta$  igual a uno, el modelo de alisado produce una predicción constante, como el método de nivel constante. Por el contrario, si  $\theta$  es pequeño, próximo a cero, la predicción va adaptándose mucho en función del último valor observado. Para entender mejor este modelo, sustituymos su valor,  $\hat{z}_T$ , en la ecuación (2.14) y agrupemos términos, entonces:

$$\begin{aligned}\hat{z}_{T+1} &= \theta(\theta \hat{z}_{T-1} + (1 - \theta) z_{T-1}) + (1 - \theta) z_T = \\ &= \theta^2 \hat{z}_{T-1} + (1 - \theta)(z_T + \theta z_{T-1}),\end{aligned}$$

repitiendo este proceso de sustitución de las predicciones anteriores, obtenemos finalmente que:

$$\hat{z}_{T+1} = \theta^T \hat{z}_1 + (1 - \theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + \dots).$$

Suponiendo que  $T$  es grande y  $\theta < 1$ , el primer término será muy pequeño, y podemos escribir la ecuación de predicción como:

$$\hat{z}_{T+1} = (1 - \theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + \dots) \quad (2.15)$$

que es una media ponderada de las observaciones previas con pesos decrecientes que suman uno, ya que:

$$(1 + \theta + \theta^2 + \dots) = \frac{1}{1 - \theta}.$$

Las predicciones generadas por el modelo de alisado simple son una media ponderada de los valores previos de la serie con pesos que decrecen geométricamente.

Vamos a justificar que esta forma de generar predicciones supone que la serie se ha generado mediante el modelo:

$$z_t = \mu_t + a_t,$$

donde  $\mu_t$  es la media de  $z_t$ , que ahora no es constante, sino que puede evolucionar con el tiempo y  $a_t$  son las innovaciones que siempre suponemos de media cero. Supongamos una muestra de tamaño  $T$  y llamemos  $\hat{\mu}_T(1)$  a la estimación del nivel futuro de la serie en el instante  $T+1$ , dados los datos hasta  $T$ . Esta notación pone de manifiesto que  $T$  es la información

disponible y (1) el horizonte de la predicción, es decir, es la predicción de  $\mu_{T+1}$ . Como la media está cambiando en el tiempo, podríamos obtener la predicción minimizando la función:

$$\sum_{t=1}^T (z_t - \mu_{T+1})^2 w_t, \quad (2.16)$$

donde los  $w_t$  son pesos, es decir números positivos que suman la unidad, que permiten tener en cuenta que las observaciones más próximas son más importantes que las más lejanas para estimar la media local. Entonces, la estimación de  $\mu_{T+1}$ , derivando e igualando a cero, vendrá dada por:

$$\hat{\mu}_{T+1} = \hat{\mu}_T(1) = \sum_{t=1}^T z_t w_t.$$

Para imponer la condición de que las observaciones próximas son más importantes que las alejadas podemos elegir pesos que disminuyan de forma geométrica con el tiempo. De esta forma, la última observación,  $z_T$ , tiene peso máximo,  $c$ ; la siguiente,  $z_{T-1}$ , tiene peso  $c\theta$ , donde  $0 < \theta < 1$ ; la siguiente,  $z_{T-2}$ , peso  $c\theta^2$ , y así sucesivamente. Esto conduce a la ecuación de predicción (2.15) anterior. En efecto, la constante  $c$  debe ser tal que los pesos sumen la unidad, y utilizando la fórmula de una progresión geométrica indefinida, resulta:

$$1 = c(1 + \theta + \theta^2 + \dots) = \frac{c}{1 - \theta},$$

con lo que  $c = 1 - \theta$ . De esta forma, los pesos vienen dados, como en la ecuación (2.15), por:

$$w_t = (1 - \theta)\theta^{T-t},$$

y la ecuación de predicción es:

$$\hat{\mu}_T(1) = (1 - \theta) \sum_{t=1}^T \theta^{T-t} z_t, \quad (2.17)$$

que implica que la predicción de  $z_{T+1}$  con información hasta  $T$  es:

$$\hat{z}_T(1) = \hat{z}_{T+1} = \hat{\mu}_T(1) = (1 - \theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + \dots)$$

donde vamos a utilizar la notación  $\hat{z}_{T+1}$  para indicar la predicción de  $z_{T+1}$  con información hasta el instante anterior. Con esta notación, la predicción de  $z_T$  con información hasta  $T - 1$  es:

$$\hat{z}_{T-1}(1) = \hat{z}_T = \hat{\mu}_{T-1}(1) = (1 - \theta)(z_{T-1} + \theta z_{T-2} + \theta^2 z_{T-3} + \dots)$$

## 2. Análisis descriptivo de una serie temporal

y multiplicando la última ecuación por  $\theta$  y restándola de la anterior, se obtiene

$$\hat{z}_{T+1} - \theta \hat{z}_T = (1 - \theta) z_T,$$

que conduce a la ecuación de predicción (2.14). Esta ecuación puede también escribirse:

$$\hat{z}_{T+1} = \hat{z}_T + (1 - \theta)(z_T - \hat{z}_T), \quad (2.18)$$

que nos indica que la predicción de la próxima observación se realiza modificando la última predicción por una fracción  $(1 - \theta)$  del último error de predicción cometido.

La utilización de este modelo requiere determinar el parámetro  $\theta$ . En las primeras aplicaciones este parámetro se fijaba a priori, habitualmente entre .70 y .99, pero progresivamente se obtuvieron mejores resultados permitiendo una mayor amplitud de valores posibles y estimando su magnitud a partir de los datos con el criterio de minimizar los errores de predicción. Esto puede hacerse probando una rejilla de valores, como 0.1, 0.2, ..., 0.9 para  $\theta$ , calculando los errores de predicción dentro de la muestra  $\hat{a}_t = z_t - \hat{z}_t$  y tomando el valor de  $\theta$  que conduzca a un valor menor de  $\sum \hat{a}_t^2$ , la suma cuadrática de los residuos o errores de predicción. Si el mejor valor obtenido en la búsqueda anterior es  $\theta_0$ , podemos refinar la búsqueda y probar ahora con valores de  $\theta$  en el intervalo  $\theta_0 - .05$  y  $\theta_0 + .05$ . Por ejemplo, si  $\theta_0 = .7$  probaremos los valores del conjunto .65, .66, ..., .70, .71, ..., .75. Normalmente esta precisión es suficiente en las aplicaciones.

### Ejemplo 2.3

Vamos a determinar el mejor valor para el parámetro de alisado  $\theta$  para la serie las leguas recorridas por la flota de Colón. Utilizando la opción de alisado simple con el programa SPSS, que prueba para los valores de  $\theta$  desde 0.1 hasta 1, se obtienen los resultados siguientes para la suma de los residuos al cuadrado con cada valor de  $\theta$ :

$\theta$	1.0	.9	.8	.7	.6
SSE	8414.9	8478.6	8574.0	8686.9	8798.8
$\theta$	.5	.4	.3	.2	
SSE	8891.3	8957.1	9018.8	9141.6	

Como la suma de cuadrados de los residuos mínima se obtiene con  $\theta = 1$ , que es equivalente al modelo de tendencia constante, concluimos que la serie

oscila alrededor de una media constante y que no hay evidencia de que tomar  $\theta$  menor que uno mejore las predicciones.

---

### 2.3.2. El método de alisado doble de Holt

Las ideas anteriores pueden aplicarse a modelos con tendencia lineal. En lugar de suponer que los parámetros son fijos, podemos admitir que evolucionan en el tiempo y estimarlos dando un peso decreciente a las observaciones. El modelo es:

$$z_t = \mu_t + a_t,$$

pero ahora en lugar de suponer una tendencia determinista permitimos que el nivel evolucione linealmente en el tiempo, pero con una pendiente que puede ser distinta en los distintos períodos. Esto se consigue escribiendo el nivel de la serie en el instante  $t$  como:

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

de manera que la diferencia entre los niveles de dos instantes consecutivos,  $t - 1$  y  $t$ , es  $\beta_{t-1}$  la pendiente en el momento  $t - 1$ . Observemos que si  $\beta_{t-1} = \beta$ , constante en el tiempo, este modelo es idéntico al de tendencia lineal determinista. Al permitir que la pendiente sea variable este modelo es mucho más flexible.

La predicción de  $z_t$  con información hasta  $t - 1$ , que llamaremos  $\hat{z}_{t-1}(1)$ , se obtiene como

$$\hat{z}_{t-1}(1) = \hat{\mu}_{t|t-1} = \hat{\mu}_{t-1|t-1} + \hat{\beta}_{t-1}$$

donde la estimación del nivel de la serie en el instante  $t$  es la suma de las últimas estimaciones del nivel y de la pendiente con información hasta  $t - 1$ . La notación  $\hat{\mu}_{t|t-1}$  indica que estamos estimando el nivel en el instante  $t$ , pero con información que incluye la del instante  $t - 1$ , es decir el dato  $z_{t-1}$ .

Para construir las predicciones con este modelo vamos a generalizar lo aprendido en el modelo anterior. Las predicciones del alisado simple se obtienen con la fórmula (2.18), que indica que la predicción futura es igual a la última predicción realizada más un factor de corrección que es el producto de un coeficiente (uno menos el factor de descuento) por el último error cometido. Como ahora tenemos que estimar dos parámetros, el método de Holt generaliza esta idea introduciendo dos factores de descuento. Supongamos que la última predicción se realiza en el periodo  $T$ , después de haber observado  $z_T$ , y corresponde a la estimación de  $z_{T+1}$ ,

## 2. Análisis descriptivo de una serie temporal

---

el valor de la serie en el instante  $T + 1$ . La predicción será  $\hat{\mu}_{T+1|T} = \hat{\mu}_{T|T} + \hat{\beta}_T$ , donde  $\hat{\mu}_{T|T}$  y  $\hat{\beta}_T$  son las estimaciones del nivel y el crecimiento con información hasta el instante  $T$ . Al observar el valor  $z_{T+1}$  podemos calcular el error de predicción ( $z_{T+1} - \hat{\mu}_{T+1|T}$ ), y, como en el método de alisado simple, corregir la estimación anterior por una fracción del error cometido. En consecuencia, la estimación  $\hat{\mu}_{T+1|T+1}$  con información hasta  $T + 1$ , será:

$$\begin{aligned}\hat{\mu}_{T+1|T+1} &= \hat{\mu}_{T+1|T} + (1 - \theta)(z_{T+1} - \hat{\mu}_{T+1|T}) = \\ &= \hat{\mu}_{T|T} + \hat{\beta}_T + (1 - \theta)(z_{T+1} - \hat{\mu}_{T|T} - \hat{\beta}_T),\end{aligned}$$

donde  $\theta < 1$  es un factor de descuento. Observemos que la estructura de esta ecuación es análoga a la (2.18). La nueva estimación del crecimiento futuro con información hasta  $T + 1$ ,  $\hat{\beta}_{T+1}$  se realiza modificando la última estimación por una fracción del último error cometido:

$$\hat{\beta}_{T+1} = \hat{\beta}_T + (1 - \gamma)(\hat{\mu}_{T+1|T+1} - \hat{\mu}_{T|T} - \hat{\beta}_T),$$

donde  $\gamma < 1$  es otro factor de descuento sobre el error anterior en la estimación del crecimiento. Estas ecuaciones permiten obtener recursivamente las estimaciones a partir de unos valores iniciales. Los parámetros  $\theta$  y  $\gamma$  se determinan como en el caso anterior, probando con una rejilla de valores y escogiendo los que minimizan la suma cuadrática de los errores de predicción.

---

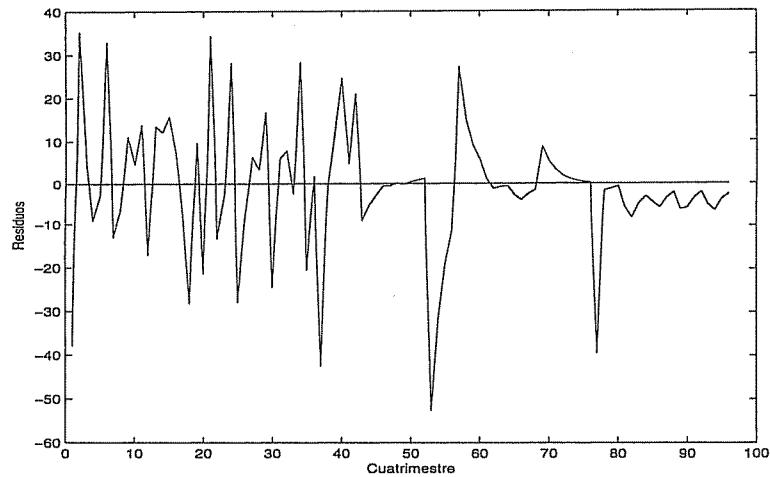
### Ejemplo 2.4

Vamos a determinar el mejor valor de los parámetros  $\theta$  y  $\gamma$  para la serie de población de mayores de 16 años en España, del fichero pobmay16.dat. Estimando los parámetros con el programa SPSS, se obtienen los siguientes mejores valores para los parámetros, en el sentido de conducir a menor error cuadrático de predicción dentro de la muestra:

$\theta$	1.00	.90	1.00	.90
$\gamma$	.40	.40	.20	.60
SSE	22499	22690	23056	23133
$\theta$	.80	.80	.80	1.00
$\gamma$	.60	.80	.40	.60
SSE	23609	23956	23963	24057

La figura 2.5 muestra los residuos del mejor modelo que corresponde a  $\theta = 1$  y  $\gamma = .4$ . Si comparamos estos resultados con el modelo lineal, vemos

**Figura 2.5 Residuos de la serie de mayores de 16 años con el método de Holt**



que los errores de predicción son mucho menores y que ahora los residuos no muestran una tendencia destacada y las predicciones son bastante buenas en muchos períodos.

---

#### 2.4. Métodos de descomposición para series estacionales

Cuando la serie además de tendencia y componente aleatorio tiene estacionalidad, los métodos de descomposición suponen que los datos se generan como suma de esos tres efectos:

$$z_t = \mu_t + S_t + a_t,$$

donde  $\mu_t$  es el nivel de la serie,  $S_t$  es el componente estacional y  $a_t$  es el componente puramente aleatorio o innovación que, como en modelos anteriores, es una secuencia de variables incorreladas de media cero y varianza constante. Los métodos clásicos de descomposición suponen que tanto el nivel como la estacionalidad son deterministas. El nivel  $\mu_t$  se modela mediante un polinomio del tiempo de orden menor o igual a dos, y la estacionalidad como una función periódica, que verifica la condición:

$$S_t = S_{t-s},$$

donde  $s$  es el periodo de la función, que depende de la estacionalidad de los datos. Una serie mensual con estacionalidad anual tiene periodo  $s = 12$  meses, ya que suponemos que los *coeficientes estacionales*,  $S_t$ , se repetirán cada 12 observaciones; una serie trimestral con estacionalidad anual tendrá periodo  $s = 4$  trimestres y una serie diaria con estacionalidad semanal tendrá  $s = 7$  días. El procedimiento de construcción del modelo para la serie se realiza en las tres etapas siguientes:

1. Se estima el nivel de la serie observada como en el modelo de tendencias deterministas. La forma de la tendencia se elige con el gráfico de la serie y se estima como se ha explicado en la sección 2.2. A continuación se resta a la serie el nivel estimado,  $\hat{\mu}_t$ , para obtener una serie residual,  $E_t = z_t - \hat{\mu}_t$ , que contendrá la estacionalidad y el componente aleatorio. Esta serie se denomina serie sin tendencia.
2. Los coeficientes estacionales,  $S_1, \dots, S_{12}$ , se definen como un conjunto de coeficientes que suman cero y que se repiten cada año. Se estiman en la serie sin tendencia como la diferencia entre la media de los periodos estacionales y la media general

$$\widehat{S}_j = \overline{E}_j - \overline{E}.$$

Por construcción, es fácil comprobar que la suma de los factores estacionales es cero. Para concretar, supongamos que tenemos  $T = 12n$  datos de una serie mensual correspondientes a  $n$  años de datos mensuales. Podemos etiquetar las observaciones para poner de manifiesto el mes y el año al que pertenecen definiendo  $t = 12i + j$ , donde la variable  $i$  toma los valores  $0, 1, \dots, n - 1$  y representa los años transcurridos y la  $j$  toma valores  $1, \dots, 12$  y representa los meses del año. Con esta notación, las observaciones de la serie sin tendencia pueden escribirse como  $E_{12i+j}$ , la media total se calcula como  $\overline{E} = \sum_{i=1}^{n-1} \sum_{j=1}^{12} E_{12i+j}/T$ , y la media de las observaciones que corresponden al mismo mes,  $\overline{E}_j$  mediante:

$$\overline{E}_j = \sum_{i=1}^n E_{12i+j}/n.$$

3. Se obtiene la serie de innovaciones estimada restando a la serie sin tendencia el factor estacional de cada observación. Por ejemplo, para datos mensuales, con la notación anterior:

$$\widehat{a}_{12i+j} = E_{12i+j} - \widehat{S}_j$$

4. La predicción de la serie se realiza sumando las estimaciones de la tendencia y del factor estacional que corresponde a cada observación este mes. Por otro lado, si restamos a la serie original el coeficiente estacional del mes se obtiene la *serie desestacionalizada*.

Como se ha visto en las secciones anteriores, existen series que claramente no tienen una tendencia constante y para las que ajustar una tendencia determinista no es adecuado. Una posibilidad alternativa es estimar el nivel de la serie localmente mediante una media móvil de doce meses como sigue: el valor del nivel en el centro del año se obtiene mediante la operación:

$$\hat{\mu}_t = \frac{z_{t-5} + \dots + z_{t+5} + z_{t+6}}{12},$$

es decir, se construye una media de doce observaciones. Esta operación se conoce como *media móvil*, ya que las observaciones que entran en el cálculo de  $\hat{\mu}_t$  van variando en el tiempo. Aplicando este método se obtiene una estimación del nivel de la serie en los instantes  $t = 6, \dots, T - 6$ . Para obtener los valores del nivel en los extremos se ajusta una recta a los últimos valores y de esta manera se completa la serie de niveles. A continuación se realiza la descomposición de la serie de la manera ya explicada mediante los pasos 2 y 3.

---

#### Ejemplo 2.5

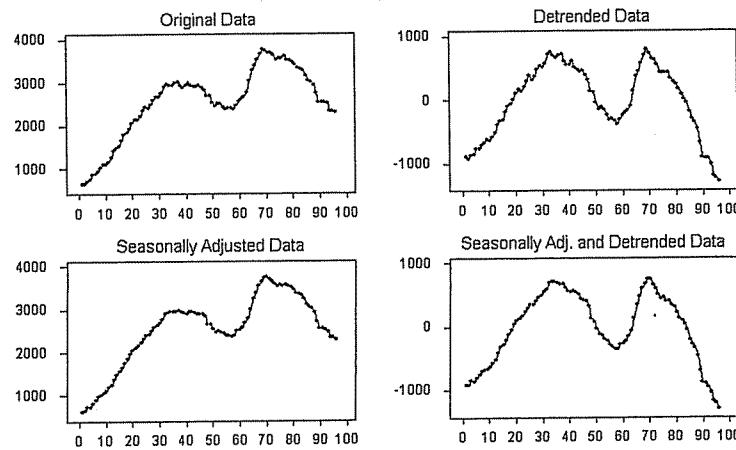
Vamos a analizar la serie de paro en España del fichero parados.dat. La figura 2.6 presenta el resultado de la descomposición con el programa SPSS. La primera serie es la serie original a la que se ha ajustado una tendencia lineal para obtener la serie sin tendencia (detrended) que se presenta a continuación. Haciendo las medias de los cuatrimestres y restando la media general se obtienen los coeficientes estacionales indicados en la figura 2.7. En esta figura se observa que el primer cuatrimestre, que corresponde al invierno, es el de más alto número de parados. Éstos disminuyen en primavera y verano y vuelven a aumentar en otoño. En los otros paneles de esta figura se observa que la variabilidad es similar en los cuatro períodos estacionales.

Para ilustrar los resultados de obtener la tendencia mediante medias móviles, las figuras 2.8, 2.9 y 2.10 presentan la tendencia obtenida por este procedimiento y los factores estacionales y residuos resultantes. Se observa que este segundo procedimiento parece dar mejores resultados en esta serie que el de tendencias deterministas, ya que la estimación del nivel de la serie mediante medias móviles de la figura 2.8 es mucho más fiel a la serie que la estimación del nivel mediante una línea recta.

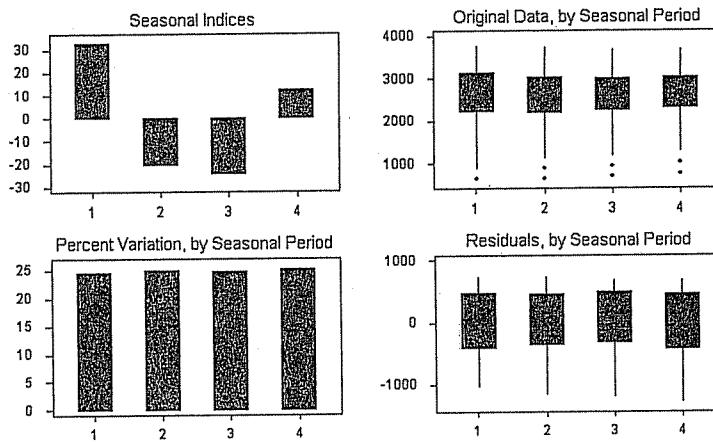
---

## 2. Análisis descriptivo de una serie temporal

**Figura 2.6 Descomposición de la serie del paro en España en el periodo primer cuatrimestre de 1977 al cuarto del 2000**



**Figura 2.7 Coeficientes estacionales para la serie del paro en España**



## 2.5. Estacionalidad y el ajuste de ciclos

### 2.5.1. Definiciones básicas

Un procedimiento alternativo para modelar la estacionalidad es representarla por una función armónica de periodo  $s$ . Suponiendo que hemos

Figura 2.8 Tendencia estimada por medias móviles para la serie de paro en España

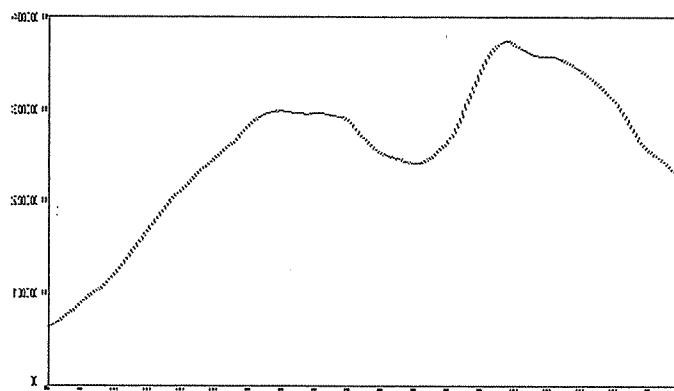
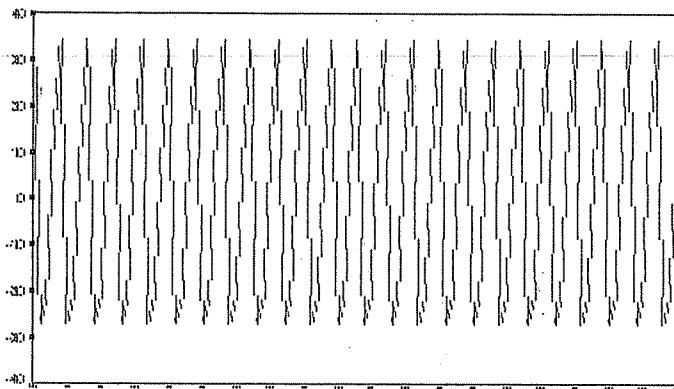


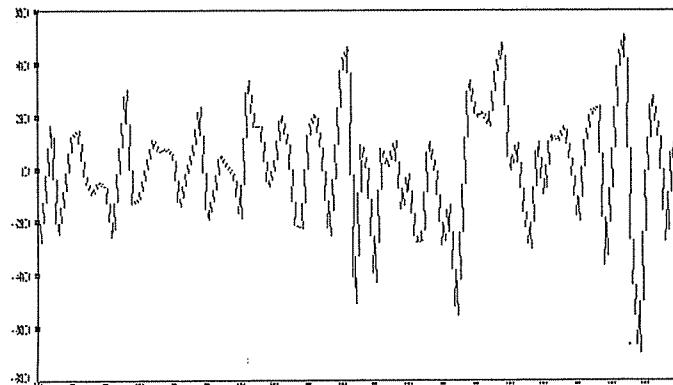
Figura 2.9 Coeficientes estacionales para la series del paro en España



eliminado la tendencia, si la tuviese, consideremos series que tienen sólo componente estacional, con estructura:

$$z_t = S_t + a_t.$$

Figura 2.10 Residuos para la serie del paro en España



La alternativa más simple para representar  $S_t$  como una función periódica, con  $S_t = S_{t-s}$ , es suponer una función armónica, como el seno o el coseno. La figura 2.11 representa las funciones  $\sin(2\pi t/s)$  y  $\cos(2\pi t/s)$  para  $t = 1, \dots, T$ , donde se ha tomado el tamaño muestral,  $T$ , exactamente igual al periodo de la función,  $s = 100$ . Es fácil comprobar que estas funciones sinusoidales se repiten exactamente cada  $s$  observaciones. Por ejemplo,  $\sin(2\pi(t+s)/s) = \sin(2\pi t/s + 2\pi) = \sin(2\pi t/s)$ .

Se denomina *frecuencia* de la función a la inversa del periodo,

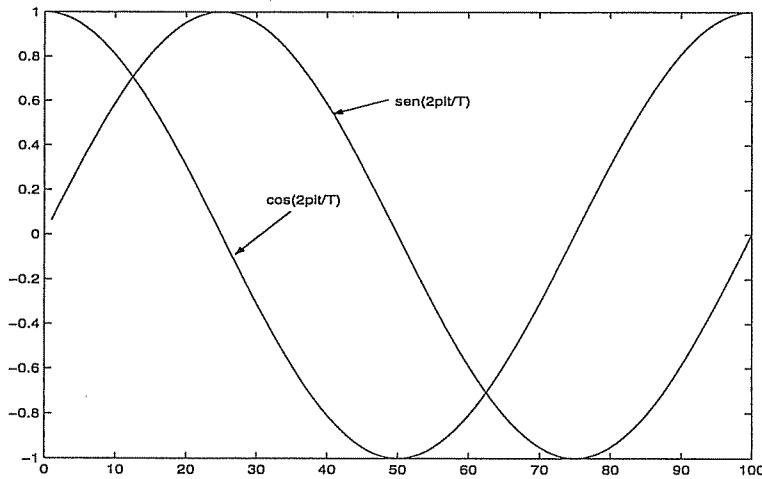
$$f = \frac{1}{s},$$

que indica la fracción de un ciclo completo que se observa entre dos unidades de tiempo. Por ejemplo, en una serie trimestral ( $s = 4$ ), la frecuencia es  $f = 1/4 = .25$ , indicando que entre dos observaciones, un trimestre, ha transcurrido .25 del periodo de la función o un 25 % de un ciclo completo. En una serie mensual ( $s = 12$ ) la frecuencia es  $1/12 = .0833$ , indicando que la unidad de tiempo de la serie, un mes, representa el 8.33 % del ciclo estacional de 12 meses. La cantidad

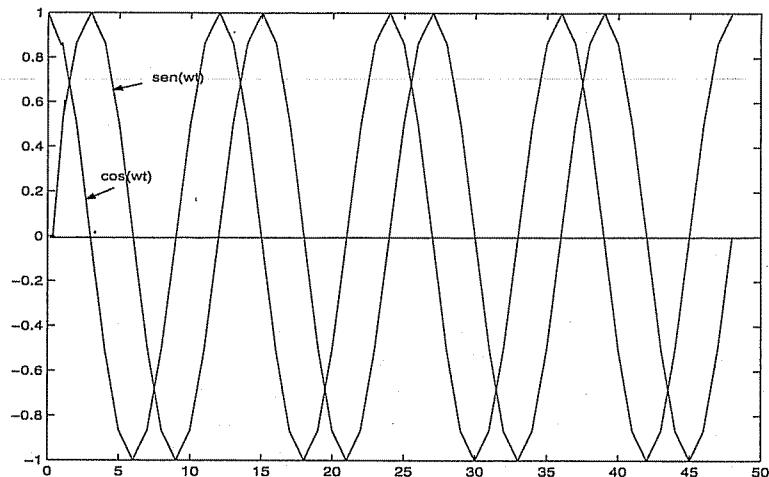
$$w = 2\pi f = 2\pi/s$$

se denomina *frecuencia angular*, e indica en radianes el ángulo recorrido en una unidad de tiempo, teniendo en cuenta que el ciclo completo son  $2\pi$  radianes. Por ejemplo, una serie trimestral tiene una frecuencia angular de  $w = 2\pi/4 = \pi/2$ , indicando que en un trimestre se recorre un ángulo de  $\pi/2$  respecto al ciclo total de  $2\pi$ .

**Figura 2.11** Representación de las funciones seno y coseno con periodo igual a los datos observados



**Figura 2.12** Tres años de datos mensuales para una serie con estacionalidad sinusoidal



En una serie real el tamaño muestral es típicamente mayor que el periodo de la función. Si el tamaño muestral fuese menor que el periodo, sólo observaremos una fracción del ciclo. Por ejemplo, con ocho datos mensuales de una serie con estacionalidad anual. Para simplificar la exposición vamos a suponer que el tamaño muestral es múltiplo exacto de

## 2. Análisis descriptivo de una serie temporal

---

$s$ , de manera que observamos  $j = T/s$  ciclos completos. Además, vamos a suponer que el tamaño muestral es un número par. Entonces,  $w = 2\pi/s = 2\pi j/T$ , con  $j$  entero. Por ejemplo, la figura 2.12 presenta una muestra de tamaño  $T = 48$  de las dos series deterministas  $x_{1t} = \sin(wt)$  y  $x_{2t} = \cos(wt)$  mensuales,  $s = 12$ , con frecuencia angular  $w = 2\pi/12 = \pi/6$ . Es intuitivo en el gráfico, y se demuestra rigurosamente en el apéndice 2.1, que al incluir el periodo muestral ciclos completos:

$$\sum_{t=1}^T \sin(wt) = \sum_{t=1}^T \cos(wt) = 0$$

y además

$$\sum_{t=1}^T \sin(wt) \cos(wt) = 0.$$

La primera condición indica que las variables  $x_{1t} = \sin(wt)$  y  $x_{2t} = \cos(wt)$  observadas en  $t = 1, \dots, T$  tienen media nula. Esto es consecuencia de que el tamaño muestral es múltiplo de ciclos completos. La segunda que, en este caso, estas variables tienen covarianzas nulas y están por tanto incorreladas en la muestra.

### 2.5.2. Representación de la estacionalidad por un ciclo

Supongamos una serie  $(z_1, \dots, z_T)$  que tiene una estacionalidad cíclica de periodo  $s$ , y en la que observamos  $j$  ciclos completos, es decir  $T = js$ , con  $j$  entero. Vamos a modelar la estacionalidad mediante una función seno con frecuencia angular  $w = 2\pi/s$ . Por ejemplo, tenemos  $j = 10$  años de datos de una serie mensual, con  $T = 120$ , y queremos ajustar un ciclo anual, de manera que  $s = 12$  y la onda sinusoidal tendrá una frecuencia angular  $w = 2\pi/12$ . Por ejemplo, la figura 2.13 presenta una serie de estas características. La primera observación de la serie no será, en general, el valor medio del ciclo, como corresponde a la función seno, sino que la onda sinusoidal que describe la estacionalidad comenzará en la primera observación con un cierto ángulo  $\theta$  de desfase, que es desconocido, con relación al comienzo del ciclo. Además, el ciclo tendrá un amplitud también desconocida, que llamaremos  $R$ . Supondremos que las oscilaciones estacionales se producen alrededor de un valor medio de la serie  $\mu$  que es constante en el tiempo. Finalmente, supondremos que existe un error aleatorio superpuesto a la estacionalidad, que representamos como en modelos anteriores por una secuencia de variables independientes,  $a_t$ , de media cero, varianza constante y distribución normal. El modelo para la serie será:

$$z_t = \mu + R \sin(wt + \theta) + a_t. \quad (2.19)$$

Para ajustar este modelo a los datos vamos a escribirlo de una forma más conveniente. Utilizaremos que el seno de la suma de dos ángulos es la suma de los productos de los senos más el producto de los cosenos, con lo que podemos escribir la expresión anterior como:

$$z_t = \mu + R \sin(wt) \sin(\theta) + R \cos(wt) \cos(\theta) + a_t,$$

y llamando  $A = R \sin \theta$  y  $B = R \cos \theta$ , tenemos que:

$$z_t = \mu + A \sin(wt) + B \cos(wt) + a_t. \quad (2.20)$$

Esta expresión es más simple que la (2.19) ya que representa la serie como suma de dos funciones sinusoidales de frecuencia angular conocida y los parámetros desconocidos,  $\theta$  y  $R$ , se han transformado en las dos amplitudes desconocidas,  $A$  y  $B$ , que estimaremos a partir de los datos. El modelo (2.20) es lineal en los tres parámetros desconocidos,  $\mu$ ,  $A$  y  $B$ , y podemos estimarlo por mínimos cuadrados. Las ecuaciones resultantes son (véase el apéndice 2.1):

$$\begin{bmatrix} \hat{\mu} \\ \hat{A} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} T & ssen & scos \\ ssen & ssen2 & ssencos \\ scos & ssencos & ssen2 \end{bmatrix}^{-1} \begin{bmatrix} \sum z_t \\ \sum z_t \sin(wt) \\ \sum z_t \cos(wt) \end{bmatrix}$$

donde  $ssen = \sum \sin(wt)$ ,  $scos = \sum \cos(wt)$ ,  $ssen2 = \sum \sin^2(wt)$  y  $ssencos = \sum \sin(wt)\cos(wt)$ . Se demuestra en el apéndice 2.1 que cuando el tamaño muestral es un múltiplo del periodo, como estamos suponiendo aquí, las variables  $\sin(wt)$  y  $\cos(wt)$  son ortogonales y tienen media cero y varianza  $1/2$ . Por tanto, la matriz de productos cruzados es diagonal, y los estimadores son:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T z_t,$$

y

$$\hat{A} = \frac{2}{T} \sum_{t=1}^T z_t \sin(wt), \quad (2.21)$$

$$\hat{B} = \frac{2}{T} \sum_{t=1}^T z_t \cos(wt). \quad (2.22)$$

Podemos obtener un estimador de la amplitud  $R$  mediante:

$$\hat{R}^2 = \hat{A}^2 + \hat{B}^2. \quad (2.23)$$

---

## 2. Análisis descriptivo de una serie temporal

Los residuos del modelo se calculan mediante:

$$\hat{a}_t = z_t - \hat{\mu} + \hat{A} \sin(wt) + \hat{B} \cos(wt),$$

y por construcción tendrán media cero. Su varianza será:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{a}_t^2.$$

El modelo estimado realiza una descomposición de la variabilidad de los datos en una parte que corresponde al componente sinusoidal y otra al componente residual. En efecto, la varianza de la variable será:

$$\frac{1}{T} \sum_{t=1}^T (z_t - \hat{\mu})^2 = \frac{1}{T} \sum_{t=1}^T (\hat{A} \sin(wt) + \hat{B} \cos(wt) + \hat{a}_t)^2$$

y sumando para todos los datos y utilizando los resultados del apéndice 2.1, donde se demuestra que las variables  $\sin(wt)$  y  $\cos(wt)$  tienen media cero, varianza  $1/2$  y están incorreladas, resulta:

$$\frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})^2 = \frac{\hat{A}^2}{2} + \frac{\hat{B}^2}{2} + \hat{\sigma}^2 = \frac{\hat{R}^2}{2} + \hat{\sigma}^2, \quad (2.24)$$

que puede interpretarse como una descomposición de la varianza en sus dos componentes ortogonales de variabilidad.

---

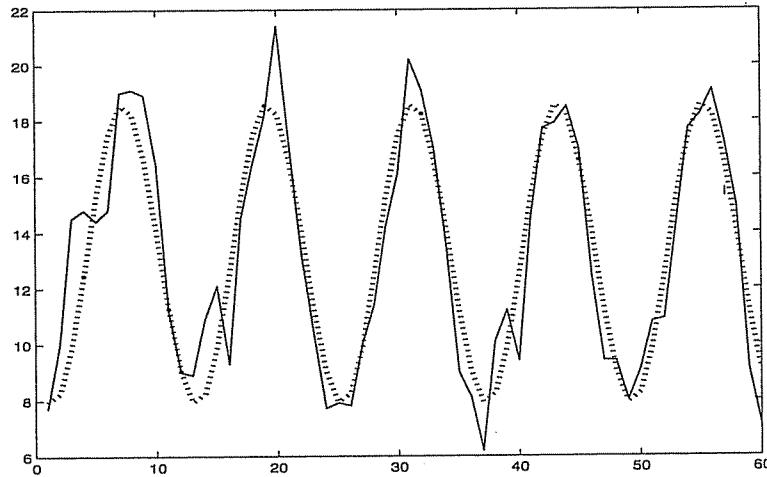
### Ejemplo 2.6

Como ejemplo vamos a ajustar una función sinusoidal de periodo 12, o frecuencia angular  $2\pi/12$ , a la temperatura media mensual en Santiago de Compostela del fichero tempsantiago.dat. La figura 2.13 ilustra la serie y su ajuste sinusoidal mediante un modelo que permite que el inicio del ciclo sea desconocido así como el nivel medio de la serie y la amplitud de la onda. Aplicando las fórmulas anteriores, el modelo estimado es:

$$\hat{z}_t = 13.2417 - 3.3351 \sin(2\pi t/12) - 4.2035 \cos(2\pi t/12).$$

Se observa que el ajuste de la serie es bastante bueno. La amplitud de la onda se calcula como  $\hat{R} = \sqrt{\hat{A}^2 + \hat{B}^2} = 5.36$ . El ángulo inicial puede también calcularse con  $-3.3351 = 5.36 \sin(\theta)$ , con lo que se obtiene que  $\theta = -0.6727$  radianes. En resumen, la temperatura en Santiago puede representarse bien por una onda seno de amplitud 5.36 grados centígrados y que comienza en enero con un desfase angular de -0.6727 radianes. Dos veces la amplitud de

Figura 2.13 Temperatura media mensual en Santiago en los años 1997 a 2001 (línea gruesa) y función sinusoidal ajustada (línea de trazos)



la onda indica la diferencia promedio máxima entre el mes más frío y el más caliente, 10.7 grados en este caso.

#### Ejemplo 2.7

Se realiza el mismo ajuste sinusoidal con una función de periodo 12 a la serie de lluvias en Santiago del fichero lluviaSC.dat. El resultado se presenta en la figura 2.14. El modelo ajustado es ahora

$$\hat{z}_t = 145.35 + 2.0858 \operatorname{sen}(2\pi t/12) + 88.2449 \cos(2\pi t/12).$$

Se observa que, aunque el modelo sinusoidal explica parte de la variabilidad de la serie, el ajuste ahora no es tan bueno como en la serie de temperaturas en el ejemplo anterior, ya que la función sinusoidal no es capaz de captar ni las asimetrías ni los picos de la serie observada.

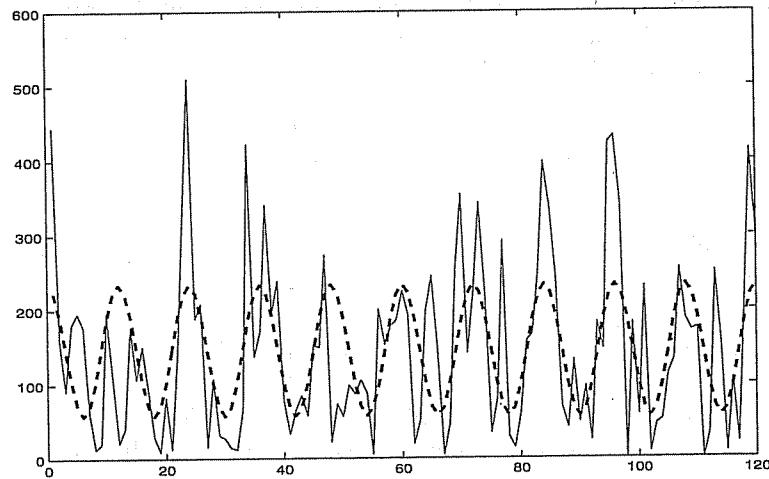
---

## 2.6. Exploración de múltiples ciclos: el periodograma

La representación de una serie estacional mediante la ecuación (2.20) es adecuada cuando la estacionalidad es exactamente sinusoidal de periodo  $s$ , pero no nos sirve para describir funciones periódicas generales.

## 2. Análisis descriptivo de una serie temporal

Figura 2.14 Lluvia mensual en Santiago de Compostela y función sinusoidal ajustada



Fourier, un matemático francés, demostró a principios del siglo XIX que toda función periódica puede representarse como suma de funciones sinusoidales de distinta amplitud y frecuencia. Este resultado sugiere generalizar el análisis realizado para un ciclo permitiendo que la función periódica sea la suma de varias funciones armónicas con distintas frecuencias.

Dada una serie de longitud  $T$ , se denominan *periodos básicos* o de Fourier a los que son fracciones exactas completas del tamaño muestral. Es decir, los periodos básicos están definidos por:

$$s_j = \frac{T}{j}, \text{ para } j = 1, 2, \dots, T/2$$

El valor máximo del periodo básico se obtiene para  $j = 1$  y es  $T$ , el tamaño muestral. Entonces observamos la onda sólo una vez, como en la figura 2.11. El valor mínimo del periodo básico se obtiene para  $j = T/2$  y es 2, ya que no podemos observar períodos que duran menos de dos observaciones. Por ejemplo, si tenemos datos diarios será imposible observar ciclos de periodo ocho horas o cualquier otro periodo. En el ajuste de ciclos suele trabajarse con las frecuencias, en lugar de con los períodos, y se definen las frecuencias básicas o de Fourier como las inversas de los períodos básicos:

$$f_j = \frac{j}{T}, \text{ para } j = 1, 2, \dots, T/2,$$

con lo que tenemos  $1/2 \geq f_j \geq 1/T$ , y el valor máximo de la frecuencia que podemos observar es  $f = .5$ . Este valor de las frecuencias básicas se conoce como la *frecuencia de Nyquist*.

Utilizando estas definiciones, podemos obtener una representación general de una función periódica como suma de ondas asociadas a todas las frecuencias básicas, mediante:

$$z_t = \mu + \sum_{j=1}^{T/2} A_j \sin(w_j t) + \sum_{j=1}^{T/2} B_j \cos(w_j t). \quad (2.25)$$

Esta ecuación contiene tantos parámetros como observaciones, por lo que siempre ajustará exactamente cualquier serie observada. En efecto, para  $j = T/2$ ,  $s_j = 2$ , y  $w_j = 2\pi/s_j = \pi$ , y la variable  $\sin \pi t$  es siempre igual a cero para  $t = 1, \dots, T$ , con lo que el coeficiente  $B_{T/2}$  no puede estimarse y tenemos  $T$  parámetros:  $\mu$  más los  $T/2$  de  $A_j$  y los  $T/2 - 1$  de  $B_j$ . En consecuencia, tenemos que encontrar un procedimiento para seleccionar las frecuencias que debemos incluir para explicar la evolución de la serie. Éste es el objetivo del periodograma.

Observemos que la ecuación (2.25) permite descomponer exactamente una serie temporal observada como una suma de componentes armónicos. Según la ecuación (2.24), la contribución de una onda a la varianza de una serie es su amplitud al cuadrado dividido por dos. Por tanto, ondas con amplitud estimada alta serán importantes para explicar la serie, mientras aquellas ondas de amplitud baja contribuyen poco a la explicación de la serie. Para seleccionar las frecuencias importantes podemos calcular los parámetros  $A_j$  y  $B_j$  para todas las frecuencias básicas y representar la contribución a la varianza de la serie, es decir la amplitud de la onda al cuadrado dividido por dos, como una función de la frecuencia. Este ajuste se realiza fácilmente, ya que las variables explicativas  $\sin(w_j t)$  y  $\cos(w_j t)$  son ortogonales (véase el apéndice 2.1), por lo que los coeficientes de regresión  $A_j$  y  $B_j$  serán iguales calculados en el modelo (2.25), que incluye todas las frecuencias básicas, y en un modelo como el (2.19), que incluye sólo la frecuencia  $w_j$ . Por tanto, podemos estimar los coeficientes  $A_j$  y  $B_j$  por (2.21) y (2.22), con  $w_j = 2\pi f_j$  y  $f = j/T$ .

Dados los coeficientes estimados  $\hat{A}_j$  y  $\hat{B}_j$  para la frecuencia  $w_j$  calculamos  $\hat{R}_j = \hat{A}_j^2 + \hat{B}_j^2$ , y utilizando la ecuación (2.25) descomponemos la varianza de la serie en los componentes asociados a cada una de las funciones armónicas. Llamando  $s_z^2$  a la varianza muestral de la serie, podemos escribir:

$$Ts_z^2 = \sum_{t=1}^T (z_t - \mu)^2 = \sum_{j=1}^{T/2} \frac{T}{2} \hat{R}_j^2. \quad (2.26)$$

## 2. Análisis descriptivo de una serie temporal

Se conoce como *periodograma* la representación de la contribución de cada frecuencia,  $T\widehat{R}_j^2/2$ , en función de la frecuencia  $w_j$  o  $f_j$ . En este segundo caso se define como:

$$I(f_j) = \frac{T\widehat{R}_j^2}{2}, \quad \text{con } 1/T \leq f_j \leq .5 \quad (2.27)$$

Observemos que, con esta representación, el valor medio de las alturas de las ordenadas del periodograma es el doble de la varianza del proceso, ya que por (2.26) y (2.27), como el número de frecuencias básicas es  $T/2$

$$\bar{I} = \frac{2}{T} \sum_{j=1}^{T/2} \frac{T}{2} \widehat{R}_j^2 = 2s_z^2.$$

El análisis anterior supone que estamos interesados únicamente en las frecuencias básicas. Esta hipótesis es poco restrictiva si el tamaño muestral es grande, ya que entonces el número de frecuencias básicas es muy grande y siempre existirá alguna frecuencia básica muy próxima a la que puede interesarnos. Por ejemplo, si tenemos 140 datos mensuales y queremos estimar la amplitud del periodo 12, el periodo básico para  $j = 11$  es  $T_{11} = 12.7$ , y para  $j = 12$ ,  $T_{12} = 11.66$ . Por tanto, podemos tener una idea aproximada de la amplitud de la onda de periodo 12 por la amplitud calculada con  $j = 11$  y 12. Naturalmente, podemos ajustar un modelo para cualquier periodo  $T_j$ ,  $2 \leq T_j \leq T/2$ , pero entonces, al perder la simetría, las fórmulas presentadas para la estimación son sólo aproximadas, aunque siempre podemos calcular los coeficientes exactamente mediante un programa de ordenador que estime regresiones múltiples.

El periodograma puede verse como una herramienta para la detección de posibles ciclos deterministas en una serie temporal. Por ejemplo, en una serie mensual estacional esperamos encontrar un valor alto del periodograma para  $f = 1/12$ , pero también podemos encontrar valores altos para  $f = j/12$ , es decir,  $1/6, 1/4, 1/3$ , que son armónicos del periodo estacional. Por otro lado, la serie puede tener otros ciclos no necesariamente ligados al periodo estacional, y el periodograma es una buena herramienta para detectar estos posibles componentes. La utilidad del periodograma aumenta si notamos que cuando estimamos la amplitud de una onda para una frecuencia determinada estamos de hecho calculando una amplitud media de todos los posibles ciclos con frecuencias próximas a la que estimamos. Como las distancias entre las frecuencias básicas son  $(j+1)/T - j/T = 1/T$ , podemos considerar la amplitud calculada para la frecuencia  $f_j$  como un promedio de las amplitudes existentes en las frecuencias situadas en el intervalo  $f_j \pm 1/2T$ . En consecuencia, en lugar de representar barras en las frecuencias básicas con alturas  $I(f_j)$ , podemos

obtener un *periodograma suavizado* construyendo rectángulos con centro en  $f_j$ , base igual a  $1/T$  y con alturas  $I(f_j) = T\hat{R}_j^2/2$ , de manera que el área del rectángulo sea igual a  $\hat{R}_j^2/2$ . De esta manera, distribuimos la variabilidad estimada para la frecuencia  $f_j$  en todo el rango de frecuencias. Este periodograma suavizado estará definido ahora para todas las frecuencias en el intervalo  $.5 \geq f_j \geq 0$  y el área total encerrada bajo esta función es, por (2.26), la varianza del proceso. Podemos dar un paso adicional y en lugar de representar una función en escalones, como la que acabamos de describir, representar una función continua. Para ello, suavizaremos la estimación anterior ponderando las estimaciones adyacentes de forma similar a como se suaviza un histograma para obtener una estimación de la función de densidad. En este periodograma suavizado la ordenada en cada punto se calcula por

$$I(f) = \sum_{f'_i-a}^{f_i+a} p_i I(f_i), \quad \text{con } .5 \geq f \geq 0$$

donde  $a$  representa la ventana utilizada y los  $p_i$  son pesos simétricos que suman la unidad.

---

### Ejemplo 2.8

Vamos a calcular el periodograma para la serie de lluvias en Santiago. Los programas de ordenador presentan siempre el periodograma suavizado como una curva continua. Para ello se estiman por mínimos cuadrados las amplitudes para las frecuencias básicas y luego se suavizan estas estimaciones aplicando medias ponderadas sobre las ordenadas adyacentes. El gráfico resultante se observa en la figura 2.15 y se observa un pico para la frecuencia estacional mensual,  $f = 1/12 = .083$ . El gráfico muestra también picos en las frecuencias  $1/6=.16$  y  $1/3=.33$ , que están asociadas a las frecuencias estacionales. Este resultado indica que la estacionalidad se modela mejor como una suma de estos tres componentes armónicos que con una onda determinista de periodo doce.

---

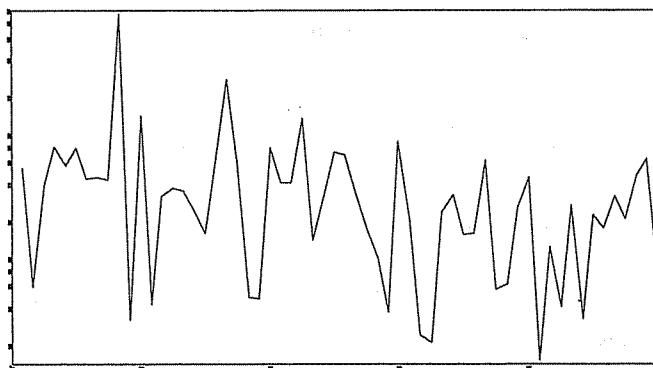
## 2.7. Conclusiones y lecturas complementarias

Hemos visto en este capítulo que los modelos de tendencias deterministas tienen muchas limitaciones para representar series reales. Los métodos adaptativos basados en dar peso decreciente al pasado funcionan mejor, pero suponen una estructura de dependencia que, aunque flexible, no

## 2. Análisis descriptivo de una serie temporal

---

Figura 2.15 Periodograma de la serie de lluvias en Santiago



es esperable que se aplique a todas las series reales. Los métodos de descomposición son útiles, pero es necesario disponer de modelos más flexibles para los componentes. El periodograma es una herramienta muy valiosa para detectar componentes sinusoidales deterministas en la serie, como efectos estacionales cíclicos.

Granger (1989) es una excelente introducción a los distintos métodos de predicción. Abraham y Ledolter (1983) es una referencia muy completa sobre los métodos de alisado, que se encuentran también claramente expuestos en el capítulo 4 de Gourieroux y Monfort (1997). Las propiedades de las funciones sinusoidales y del periodograma se presenta de forma clara en Broomfield (2000) y Wei (1990), y a un nivel más matemático en Brockwell y Davies (1987). Tres libros clásicos sobre el enfoque frecuencial que recomendamos consultar al lector interesado son Granger y Hatanaka (1964), Brillinger (1981) y Priestley (1983).

---

### Ejercicios 2

- 2.1 Estimar la tendencia de la serie de consumo de gasolina en España mediante una recta en el periodo 1945 a 1995 y generar predicciones para 24 meses. Comparar el resultado con el método de Holt.
- 2.2 Aplicar a la serie de consumo de gasolina un método de descomposición y estimar los coeficientes estacionales. Comparar los resultados al estimar

la tendencia mediante una línea recta, por el método de Holt y mediante promedios móviles.

- 2.3 Calcular el periodograma para la serie del viaje de Colón e interpretarlo.
  - 2.4 Calcular el periodograma para la serie de consumo de gasolina e interpretarlo.
  - 2.5 Demostrar que las predicciones realizadas con el método de Holt verifican la ecuación recursiva  $\hat{z}_{T+1}(1) = \hat{z}_T(1) + \hat{\beta}_T + \gamma(1 - \theta)(z_{T+1} - \hat{\mu}_T - \hat{\beta}_T)$ .
- 

## Apéndice 2.1: Regresión múltiple y la estimación del periodograma

En este apéndice resumimos la estimación de una ecuación de regresión múltiple por mínimos cuadrados y su aplicación a funciones sinusoidales.

Suponemos que se desea construir una ecuación para prever una variable  $y$ , en función de un conjunto de  $p$  variables explicativas,  $x_1, \dots, x_p$ . La ecuación a estimar es:

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \dots + \hat{\beta}_p(x_{ip} - \bar{x}_p), \quad i = 1, \dots, n \quad (2.28)$$

y es inmediato comprobar que en esta ecuación si las variables explicativas toman un valor igual a su media,  $\bar{x}_j$ , la predicción  $\hat{y}_i$  de la variable respuesta es también su media,  $\bar{y}$ . Los  $p$  coeficientes  $\hat{\beta}_j$ , para  $j = 1, \dots, p$  se determinan de manera que la ecuación proporcione, en promedio, la mejor predicción posible de los valores de  $y_i$ . Llamando  $a_i = y_i - \hat{y}_i$  a los residuos o errores de predicción, el criterio de *mínimos cuadrados* minimiza

$$M = \sum_{i=1}^n a_i^2,$$

y al derivar esta expresión respecto a los parámetros  $\hat{\beta}_j$ , se obtiene el sistema de  $p$  ecuaciones:

$$2 \sum_{i=1}^n [y_i - \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \dots + \hat{\beta}_p(x_{ip} - \bar{x}_p)] (x_{ij} - \bar{x}_j),$$

que puede escribirse:

$$\sum e_i x_{ij} = 0, \quad j = 1, \dots, p,$$

que tiene una clara interpretación intuitiva. Indica que los residuos, o errores de predicción, deben de estar incorrelados con las variables explicativas, de

## 2. Análisis descriptivo de una serie temporal

---

manera que la covarianza entre ambas variables sea cero. En efecto, si existiese relación entre ambas variables podría utilizarse para prever los errores de predicción y reducirlos, con lo que la ecuación de predicción no podría ser óptima. Geométricamente, este sistema establece que el vector de residuos debe ser ortogonal al espacio generado por las variables explicativas. Llamando

$$\tilde{\mathbf{X}} = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1p} - \bar{x}_p \\ \dots & \dots & \dots \\ x_{n1} - \bar{x}_1 & \dots & x_{np} - \bar{x}_p \end{bmatrix}$$

a la matriz de variables explicativas centradas, llamando  $\tilde{\mathbf{y}} = (y_1 - \bar{y}, \dots, y_n - \bar{y})'$  al vector de los valores de la respuesta centrados, y  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$  al vector de coeficientes, el sistema de ecuaciones para obtener los parámetros es:

$$\tilde{\mathbf{X}}' \tilde{\mathbf{y}} = \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \hat{\beta}$$

que conduce a:

$$\hat{\beta} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}}.$$

Vamos a demostrar que aplicar esta fórmula a componentes sinusoidales con frecuencias básicas equivale a utilizar variables de media cero e incorreladas, con lo que la matriz  $\tilde{\mathbf{X}}' \tilde{\mathbf{X}}$  es diagonal y la estimación de cada coeficiente es simplemente el cociente entre la covarianza de la variable y la respuesta dividida por la varianza del regresor. Es decir, si las variables explicativas son del tipo  $x_{1t} = \sin(wt)$  y  $x_{2t} = \cos(wt)$ , siendo  $w = 2\pi j/T$ , donde  $T$  es el tamaño muestral y  $j$  un entero positivo menor de  $T/2$ , tenemos  $\bar{x}_1 = \bar{x}_2 = 0$ ,  $\text{var}(x_1) = \text{var}(x_2) = T/2$  y  $\text{cov}(x_1 x_2) = 0$ .

Comencemos recordando la fórmula para la suma de una progresión geométrica. Escribiendo

$$S_T = a + \dots + a^T,$$

multiplicando por  $a$ :

$$aS_T = a^2 + \dots + a^{T+1},$$

y restando ambas ecuaciones y despejando  $S_T$ :

$$S_T = \frac{a - a^{T+1}}{1 - a}.$$

En segundo lugar, recordemos que si  $i = \sqrt{-1}$ , la exponencial de un número complejo  $iw$ , se define como

$$e^{iw} = \cos(w) + i\sin(w),$$

y aplicando la fórmula de la suma a la progresión geométrica:

$$S_T = \sum_{t=1}^T e^{iwt} = \frac{e^{iw} - e^{iw(T+1)}}{1 - e^{iw}} = \frac{e^{iw}(1 - e^{iwT})}{1 - e^{iw}}.$$

Supongamos ahora que  $w = 2\pi j/T$ , donde  $j$  es un entero positivo. Entonces  $e^{iwT} = e^{i2\pi j} = \cos(2\pi j) + i\sin(2\pi j) = \cos(2\pi) + i\sin(2\pi) = 1$ . Sustituyendo este resultado en la fórmula anterior para  $S_T$  concluimos que  $S_T = 0$ . Como:

$$S_T = \sum_{t=1}^T e^{iwt} = \sum_{t=1}^T \cos(wt) + i \sum_{t=1}^T \sin(wt) = 0,$$

tendrá que verificar que  $\sum_{t=1}^T \cos(wt) = \sum_{t=1}^T \sin(wt) = 0$ . Por otro lado, como  $\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b)$ , tenemos que, tomando  $a = b = wt$  y sumando para todos los datos muestrales:

$$\sum_{t=1}^T \cos(wt)\sin(wt) = \frac{1}{2} \sum_{t=1}^T \sin(2wt) = 0.$$

Esta ecuación indica que la covarianza entre las variables seno y coseno para las frecuencias básicas es cero. Para calcular las varianzas utilizamos que  $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$  y tomando  $a = b = wt$ :

$$\cos(2wt) = \cos^2 wt - \sin^2 wt = \cos^2 wt - (1 - \cos^2 wt),$$

y despejando  $\cos^2(wt)$  y sumando para todos los datos:

$$\sum_{t=1}^T \cos^2(wt) = \frac{1}{2} \sum_{t=1}^T (1 + \cos(2wt)) = \frac{T}{2}.$$

De la misma forma, se obtiene:

$$\sum_{t=1}^T \sin^2(wt) = \frac{1}{2} \sum_{t=1}^T (1 - \cos(2wt)) = \frac{T}{2},$$

con lo que concluimos que la varianza de las variables  $x_{1t} = \sin(wt)$  y  $x_{2t} = \cos(wt)$  es  $1/2$ .

Observemos que la prueba es válida para cualquier  $w_j = 2\pi j/T$ , donde  $j$  es un entero positivo. Vamos a comprobar que si consideramos dos frecuencias de Fourier distintas y definimos las cuatro variables  $x_{1t} = \sin(w_1 t)$ ,  $x_{2t} = \cos(w_1 t)$ ,  $x_{3t} = \sin(w_2 t)$  y  $x_{4t} = \cos(w_2 t)$  ligadas a las dos frecuencias, las cuatro variables están incorreladas. En efecto, consideremos, por ejemplo, las variables  $x_{1t}$  y  $x_{4t}$ . Entonces:

$$x_{1t}x_{4t} = \sin(w_1 t) \cos(w_2 t) = \frac{1}{2} [\sin(w_1 + w_2)t - \sin(w_1 - w_2)t]$$

y  $\sin(w_1 + w_2)t = \sin(w_3 t)$  donde  $w_3$  es también de la forma  $w_3 = 2\pi j/T$ , siendo  $j$  un entero positivo, esta variable tiene media cero. La prueba es similar en los otros casos.

## 2. Análisis descriptivo de una serie temporal

---

En resumen, hemos comprobado que si definimos variables explicativas  $x_{jt} = \sin(w_j t)$  y  $x_{ht} = \cos(w_h t)$  para las frecuencias de Fourier, estas variables tienen media cero, varianza  $1/2$  y están incorreladas. Si el tamaño muestral no es un múltiplo exacto de la frecuencia de interés, es decir, no se verifica que  $w_3 = 2\pi j/T$ , los resultados son sólo aproximados, pero si el tamaño muestral es grande, el error de utilizarlos será muy pequeño. Por ejemplo, si nos interesa  $f = 1/12$  porque tenemos datos mensuales y  $T = 121$  no podemos expresar  $2\pi/12$  como  $2\pi j/121$ . Sin embargo, como  $2\pi/12 = 2\pi 10/120$ , tomando  $j=10$ , que corresponde a la frecuencia  $f = 10/121$ , esta aproximación a la frecuencia básica será buena.

---

### 3. Series temporales y procesos estocásticos



Maurice S. Bartlett (1910-2002)

Estadístico británico. Fue profesor en Manchester, donde fundó el Departamento de Estadística, en University College de Londres, y, finalmente, en Oxford hasta su jubilación. Hizo contribuciones fundamentales a los procesos estocásticos, especialmente a la inferencia de las autocorrelaciones y el espectro. Fue elegido en 1961 miembro de la Royal Society.

#### 3.1. Introducción

Los fenómenos dinámicos que observamos mediante series temporales pueden clasificarse en dos clases. Los primeros son los que toman valores estables en el tiempo alrededor de un nivel constante, sin mostrar una tendencia a crecer o a decrecer a largo plazo. Por ejemplo, la cantidad de lluvias anuales en una región, la temperatura media del año o la proporción de nacimientos que corresponde a varones. Estos procesos se denominan estacionarios y en este capítulo comenzamos su estudio, que continuará en los siguientes. Una segunda clase de procesos, que estudiaremos en los capítulos 6 y 7, son los procesos no estacionarios, que son aquellos que pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo. Por ejemplo, la renta anual en un país, las ventas de una empresa o la demanda de energía son series que evolucionan en el tiempo con tendencias más o menos estables.

En la práctica la clasificación de una serie como estacionaria o no depende del periodo de observación, ya que la serie puede ser estable en un periodo corto y no estacionaria en un periodo mayor. Por ejemplo,

### 3. Series temporales y procesos estocásticos

la temperatura en un punto de la tierra puede parecer estable en un intervalo de 20 años, y mostrar una tendencia creciente en un periodo de miles de años. Como regla general, cuanto más amplio sea el periodo de observación, más difícil es que el proceso sea estable.

Las series temporales evolucionan normalmente con cierta inercia, que se manifiesta en dependencia entre sus valores presentes y sus valores pasados. Para medir la dependencia lineal entre el presente y el pasado se utilizan las funciones de autocovarianzas y autocorrelaciones, que definiremos en este capítulo, y que generalizan la idea de covarianza y correlación entre dos variables. El análisis que presentamos en este libro utiliza principalmente la dependencia temporal, y es el más utilizado en la práctica. Un análisis alternativo es el análisis frecuencial, que utiliza la importancia relativa de los componentes cílicos. Este segundo tipo de análisis, que tiene especialmente aplicaciones en física e ingeniería, se presenta brevemente al final del capítulo.

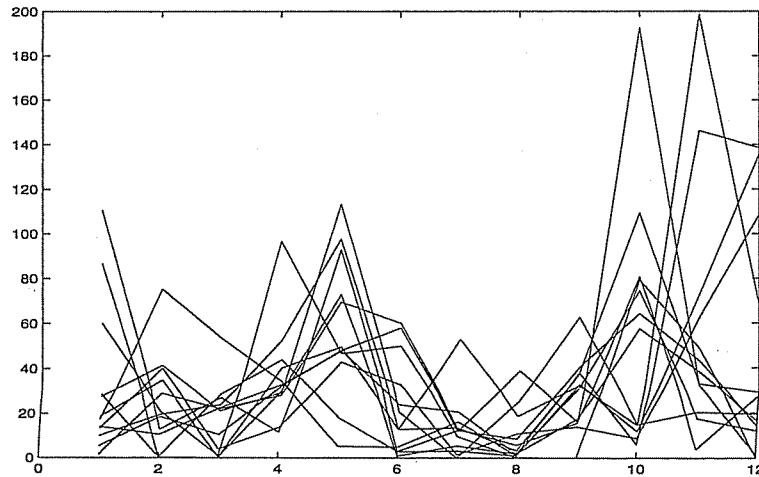
#### **3.2. El concepto de proceso estocástico**

Un proceso estocástico es un conjunto de variables aleatorias  $\{z_t\}$  donde el índice  $t$  toma valores en un cierto conjunto  $C$ . En nuestro caso, este conjunto es ordenado y corresponde a los instantes temporales (días, meses, años, etc.). Para cada valor  $t$  del conjunto  $C$  (para cada instante temporal) está definida una variable aleatoria,  $z_t$ , y los valores observados de las variables aleatorias en distintos instantes forman una serie temporal. Es decir, una serie de  $T$  datos,  $(z_1, \dots, z_t, \dots, z_T)$ , es una muestra de tamaño uno del vector de  $T$  variables aleatorias ordenadas en el tiempo correspondientes a los momentos  $t = 1, \dots, T$ , y la serie observada se considera una realización o trayectoria del proceso estocástico.

El proceso queda caracterizado si definimos la distribución de probabilidad conjunta de las variables aleatorias  $(z_1, \dots, z_t, \dots, z_T)$ , para cualquier valor de  $T$ . Estas distribuciones se denominan las *distribuciones finito-dimensionales* del proceso. Diremos que conocemos la estructura probabilística de un proceso estocástico cuando se conozcan estas distribuciones, que determinan la distribución de cualquier subconjunto de variables y, en particular, las distribuciones marginales de cada variable.

Por ejemplo, consideremos una variable climática (como la temperatura, la lluvia, la humedad, la contaminación) a las 12 horas de cada día del año en una ciudad concreta. Las 365 observaciones recogidas cada año para esta variable constituyen una realización del proceso estocástico. Al comenzar otro año se obtiene una nueva realización del proceso y así sucesivamente. Si estudiamos el valor el 9 de junio de los distintos años, tendremos la distribución de probabilidad de una variable aleatoria.

**Figura 3.1 Series mensuales de la cantidad de lluvia en Madrid en los años 1988 a 1999**



Por ejemplo, la figura 3.1 presenta 12 realizaciones (años) de la serie de lluvia en Madrid en los 12 meses del año. Tenemos por tanto 12 valores de 12 variables aleatorias, una para cada mes, y la trayectoria de los 12 valores en un año dado representa una realización del proceso estocástico. Si consideramos un mes cualquiera, los 12 valores disponibles forman una muestra de tamaño 12 de esa variable aleatoria.

Como segundo ejemplo, consideraremos el proceso estocástico definido por:

$$z_t = z_{t-1} + a_t, \quad (3.1)$$

que suponemos comienza en un instante  $t = 0$  con  $z_0 = 0$ , y donde las variables  $a_t$  son normales con media cero, varianza  $\sigma^2$  constante a lo largo del tiempo e independientes entre sí. Este proceso se conoce con el nombre de paseo aleatorio y la figura 3.2 presenta 200 realizaciones de este proceso hechas con un ordenador. El proceso comienza siempre con el valor cero, pero va alejándose de ese valor a lo largo del tiempo. La figura 3.3 proporciona histogramas de la distribución de los 200 valores de la variable del proceso disponibles para los instantes temporales  $t = 25, 50, 75$  y  $100$ . Se observa que la media de estas distribuciones es aproximadamente cero en los cuatro casos, pero la varianza de la distribución aumenta con el tiempo considerado. Este resultado coincide con la figura 3.2, donde se observa que las realizaciones del proceso tienden a alejarse del valor inicial con el tiempo.

Podemos observar varias realizaciones de un proceso estocástico en

### 3. Series temporales y procesos estocásticos

Figura 3.2 200 realizaciones de paseos aleatorios de tamaño  $T = 100$  realizadas por un ordenador

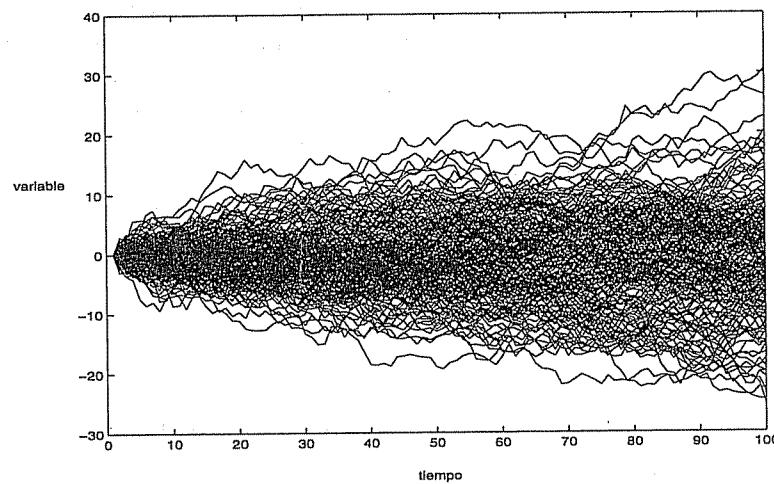
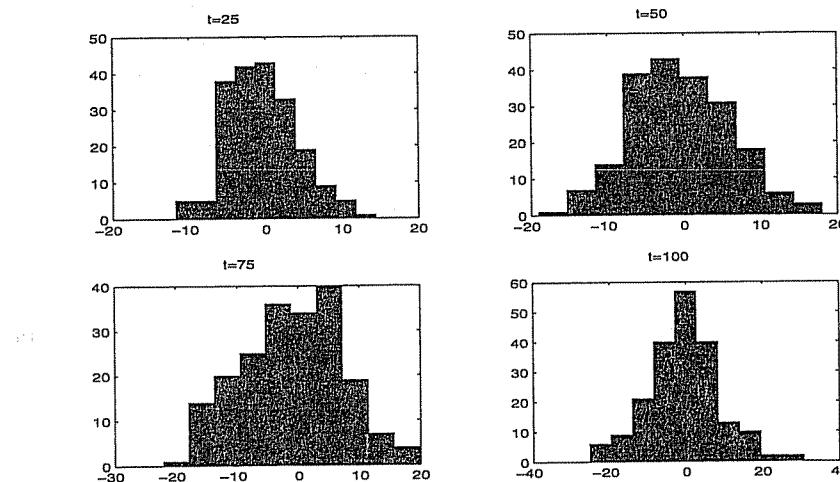


Figura 3.3 Distribución de la variable del proceso para los instantes 25, 50, 75 y 100. El histograma muestra los valores de la variable en esos instantes en las 200 realizaciones de la figura anterior



fenómenos estables repetitivos a lo largo del tiempo. Por ejemplo, con las variables climáticas, astronómicas o ambientales que siguen cada año un patrón similar como consecuencia de la rotación de la Tierra alrededor del

Sol. También, en variables que se repiten en las mismas condiciones, como el número de clientes en un puesto de servicio (supermercado, gasolinera, etc.) en la primera semana del mes en distintos meses que podemos suponer similares, o el ritmo cardiovacular de una persona a lo largo del día (en días similares), o el rendimiento obtenido por una acción en distintos días, etc. También, si definimos la ecuación de generación de las observaciones de un proceso podemos generar todas las realizaciones que queramos con un ordenador.

Determinar la distribución conjunta del proceso requiere observar un gran número de realizaciones, para estimar la probabilidad de los distintos intervalos. Esta tarea se simplifica mucho cuando podamos suponer que la distribución conjunta es normal multivariante, ya que, entonces, quedará determinada por el vector de medias y la matriz de varianzas y covarianzas entre las variables.

### 3.2.1. Propiedades de las distribuciones marginales

Llamaremos *función de medias* del proceso a una función del tiempo que proporciona las esperanzas de las distribuciones marginales  $z_t$  para cada instante:

$$E[z_t] = \mu_t. \quad (3.2)$$

Un caso particular importante, por su simplicidad, aparece cuando todas las variables tiene la misma media y entonces la función de medias es una constante. Las realizaciones del proceso no mostrarán ninguna tendencia y diremos que el proceso es estable en la media. Si, por el contrario, las medias cambian con el tiempo, las observaciones de distintos momentos mostrarán dicho cambio. Por ejemplo, los histogramas de la figura 3.3 muestran que la media de las variables del paseo aleatorio parece constante y próxima a cero. Esto puede deducirse knowing la ecuación del proceso, (3.1), ya que para la primera variable podemos escribir

$$E[z_1] = 0 + E(a_1) = 0,$$

y para las demás utilizar que si la esperanza de  $z_{t-1}$  es cero, tenemos que también lo es la de  $z_t$ , ya que:

$$E[z_t] = E[z_{t-1}] + E(a_t) = 0.$$

En muchas ocasiones sólo tenemos una realización del proceso estocástico y debemos deducir de ella si la función de medias del proceso es o no constante en el tiempo. Por ejemplo, la figura 3.4 presenta la serie temporal de rentabilidades en la Bolsa de Madrid, que parece oscilar alrededor de un valor constante, por lo que concluimos que esta serie

### 3. Series temporales y procesos estocásticos

Figura 3.4 Rentabilidad mensual de la Bolsa de Madrid en el periodo enero de 1988 hasta septiembre del 2001

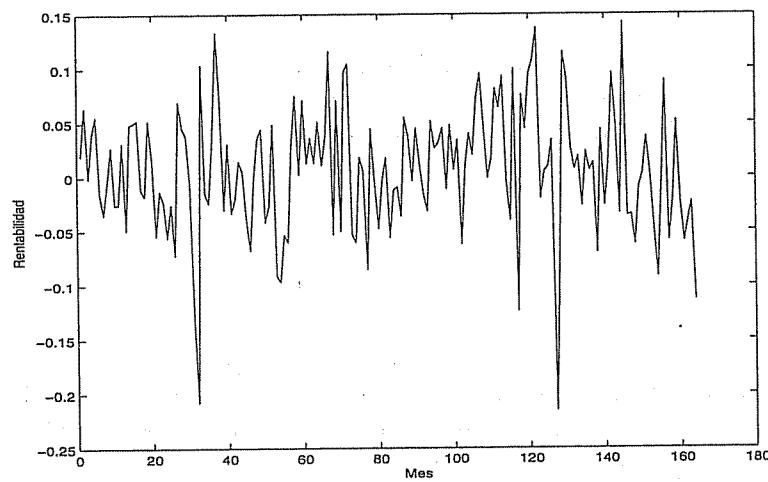
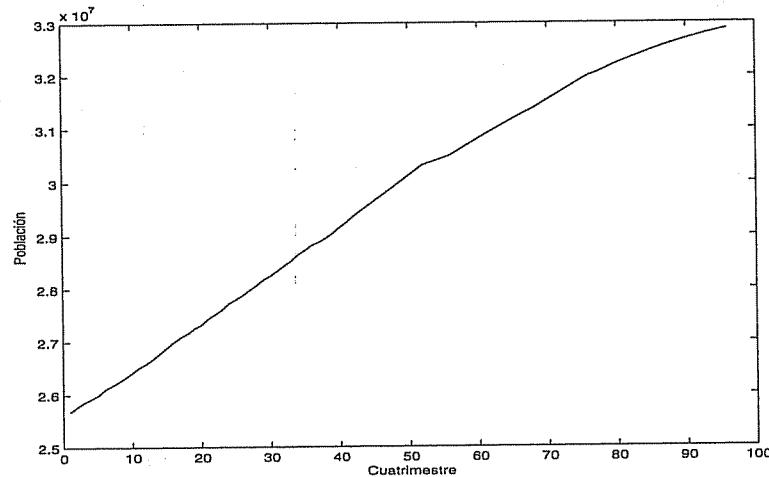


Figura 3.5 Serie de población mayor de 16 años en España en el periodo primer cuatrimestre de 1977 al cuarto del 2000



parece tener una media constante a lo largo del tiempo en el periodo estudiado. Por el contrario, la serie de la figura 3.5 muestra claramente una media no estable. Finalmente, la serie de lluvias de la figura 3.1 tampoco tiene media constante, ya que la lluvia media es distinta en los diferentes meses del año.

Se denomina *función de varianzas del proceso* a la que proporciona las varianzas en cada instante temporal:

$$Var(z_t) = \sigma_t^2 \quad (3.3)$$

y diremos que el proceso es estable en la varianza si ésta es constante en el tiempo. Un proceso puede ser estable en la media y no en la varianza y al revés. Por ejemplo, el paseo aleatorio tiene media constante, como hemos visto, pero la varianza no es constante en el tiempo. Esta propiedad se manifiesta en la figura 3.2, y puede deducirse fácilmente. En efecto, supongamos que la varianza de  $z_1$  es  $\sigma^2$ . Entonces, la variable  $z_2$  verificará:

$$Var(z_2) = E(z_2^2) = E(z_1^2 + a_2^2 + 2z_1a_2) = 2\sigma^2$$

ya que las variables  $z_1$  y  $a_2$  son independientes porque  $z_1$  depende sólo de  $a_1$ , que es independiente de  $a_2$ . Aplicando esta ecuación sucesivamente para  $t = 3, 4, \dots$ , es fácil deducir que la varianza de  $z_t$  es  $t\sigma^2$ , y la varianza aumenta linealmente con el tiempo.

Cuando tenemos una única realización, la variabilidad aparente de la serie puede ser aproximadamente constante por tramos o modificarse en el tiempo. Por ejemplo, la serie de la figura 3.4 parece tener media constante pero parece haber mayor variabilidad en ciertos tramos que en otros, con lo que la varianza de la serie podría no ser constante (veremos más adelante métodos para investigarlo).

La estructura de dependencia lineal entre las variables aleatorias del proceso se representa por las funciones de covarianza y correlación. Llamaremos *función de autocovarianzas* del proceso a la función de dos argumentos que describe las covarianzas entre dos variables del proceso en dos instantes cualesquiera:

$$\gamma(t, t+j) = Cov(z_t, z_{t+j}) = E[(z_t - \mu_t)(z_{t+j} - \mu_{t+j})]. \quad (3.4)$$

En particular, tenemos que

$$\gamma(t, t) = Var(z_t) = \sigma_t^2.$$

La función de medias y la de autocovarianzas cumplen en un proceso estocástico el mismo papel que la media y la varianza para una variable escalar.

Las autocovarianzas tienen dimensiones, las del cuadrado de la serie, por lo que no son convenientes para comparar series medidas en unidades distintas. Podemos obtener una medida adimensional de la dependencia lineal generalizando la idea del coeficiente de correlación lineal entre dos variables. Llamaremos *coeficiente de autocorrelación* de orden  $(t, t+j)$

### 3. Series temporales y procesos estocásticos

al coeficiente de correlación entre las variables  $z_t, z_{t+j}$  y *función de autocorrelación* a la función de dos argumentos que describe estos coeficientes para dos valores cualesquiera de las variables. Esta función será:

$$\rho(t, t+j) = \frac{\text{Cov}(t, t+j)}{\sigma_t \sigma_{t+j}} = \frac{\gamma(t, t+j)}{\gamma^{1/2}(t, t)\gamma^{1/2}(t+j, t+j)}. \quad (3.5)$$

#### 3.2.2. Propiedades de las distribuciones condicionadas

Además de estudiar las distribuciones marginales en los procesos estocásticos, suele ser de gran interés el estudio de las distribuciones condicionadas. Un tipo importante de procesos son los *procesos de Markov* (o markovianos) que tienen la propiedad de que:

$$f(z_{t+1}|z_t, \dots, z_1) = f(z_{t+1}|z_t), \quad t = 1, 2, \dots,$$

es decir, la distribución de la variable aleatoria en cualquier instante dados los valores previos del proceso sólo depende del último valor observado. Intuitivamente, un proceso es de Markov si conocido el valor actual del proceso la distribución del valor futuro sólo depende de ese valor, y no del camino recorrido para llegar hasta él. Un ejemplo de proceso de Markov es el paseo aleatorio, dado por (3.1) y estudiado en la sección anterior, donde:

$$f(z_{t+1}|z_t, \dots, z_1) = f(z_{t+1}|z_t) = N(z_t, \sigma^2)$$

es decir, la distribución de una variable generada por un paseo aleatorio en el instante  $t+1$  condicionada a su valor anterior,  $z_t$ , es siempre normal, con media del último valor observado,  $z_t$ , y varianza constante  $\sigma^2$ , que no depende del instante considerado.

Una propiedad más débil que la markoviana es que la esperanza condicionada sólo dependa del último valor observado. En particular, cuando se verifica que:

$$E(z_t|z_{t-1}, \dots, z_1) = E(z_t|z_{t-1}) = z_{t-1} \quad (3.6)$$

el proceso se denomina una *martingala*. Por ejemplo, el paseo aleatorio es una martingala. Observemos que una martingala no es necesariamente un proceso de Markov, porque para ello no sólo la esperanza sino toda la distribución debe depender únicamente del último valor observado. Además, en un proceso de Markov no es necesario que se verifique la condición (3.6), y  $E(z_t|z_{t-1}, \dots, z_1)$  puede ser cualquier función  $g(z_{t-1})$  del último valor observado.

Es interesante resaltar las diferencias entre las distribuciones condicionadas y las marginales estudiadas en el apartado anterior. La distribución marginal de  $z_t$  representa lo que podemos conocer de esta variable sin saber nada de su trayectoria hasta el instante  $t$ . La distribución condicionada de  $z_t$  dada  $z_{t-1}, \dots, z_{t-k}$  representa lo que podemos decir de esta variable cuando conocemos los  $k$  valores anteriores del proceso. Por ejemplo, en el paseo aleatorio (3.1) la media marginal es constante e igual a cero, mientras que la media condicionada a los valores anteriores es igual al último valor observado y la varianza marginal crece con el tiempo, mientras que la varianza condicionada es constante. Intuitivamente, si comenzamos un paseo aleatorio en  $t = 0$  y nos preguntamos por los valores futuros del proceso para un valor de  $t$  fijo, es decir por su distribución marginal para ese instante, podemos decir que la distribución será normal, con media (promedio sobre todas las posibles realizaciones) igual a cero y varianza  $t\sigma^2$ . Es decir, la probabilidad de que  $z_t$  tome valores alejados de su media crece con el tiempo, como vemos en la figura 3.2. Sin embargo, si calculamos la distribución condicionada de  $z_t$  a un valor fijo de  $z_{t-1}$ , es decir tomamos sólo aquellas realizaciones que en el instante  $t - 1$  han tomado el valor  $z_{t-1}$  fijado, la distribución del valor siguiente  $z_t$  será normal ahora con media  $z_{t-1}$  y varianza constante  $\sigma^2$ .

En las series temporales las distribuciones condicionadas van a tener mayor interés que las marginales, porque nos definen las predicciones que podemos hacer del futuro, conocido el pasado y la incertidumbre de estas predicciones.

### 3.3. Procesos estacionarios

#### 3.3.1. Definición

La obtención de las distribuciones de probabilidad del proceso es posible en ciertas situaciones, por ejemplo con variables climáticas, donde podemos suponer que cada año se observa una realización del mismo proceso, o técnicas, que pueden generarse en un laboratorio. Sin embargo, en muchas situaciones de interés, como ocurre con variables económicas o sociales, sólo podemos observar una realización del proceso. Por ejemplo, si observamos la serie de crecimientos anuales de la riqueza de un país no es posible volver atrás en el tiempo para generar otra realización. El proceso estocástico existe conceptualmente, pero no es posible obtener muestras sucesivas o realizaciones independientes del mismo. Para poder estimar las características «transversales» del proceso (medias, varianzas, etc.) a partir de su evolución «longitudinal» es necesario suponer que las propiedades «transversales» (distribución de las variables en cada

### 3. Series temporales y procesos estocásticos

instante) son estables a lo largo del tiempo. Esto conduce al concepto de estacionariedad, que definimos a continuación.

Diremos que un proceso estocástico (serie temporal) es *estacionario en sentido estricto* si:

1. las distribuciones marginales de todas las variables son idénticas;
2. las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas.

La primera condición establece que, en particular, la media y la varianza de todas las variables son las mismas. También lo son los coeficientes de asimetría y curtosis de las distribuciones marginales, ya que estas distribuciones son las mismas para todos los retardos. La segunda condición impone que la dependencia entre las variables sólo depende de sus retardos, es decir, la misma dependencia existe entre las variables  $z_t, z_{t+j}, z_{t+j+h}$  que entre las variables  $z_{t+k}, z_{t+k-j}, z_{t+k-j-h}$ . Estas dos condiciones pueden resumirse estableciendo que la distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo, es decir:

$$F(z_i, z_j, \dots, z_k) = F(z_{i+h}, z_{j+h}, \dots, z_{k+h}).$$

La estacionariedad estricta es una condición muy fuerte, ya que para contrastarla es necesario disponer de las distribuciones conjuntas para cualquier selección de variables del proceso. Una propiedad más débil, pero más fácil de contrastar en la práctica, es la *estacionariedad en sentido débil*, que implica la estabilidad de la media, la varianza y la estructura de covarianzas a lo largo del tiempo. Un proceso es estacionario en sentido débil si, para todo  $t$ :

1.  $\mu_t = \mu = cte$ ,
2.  $\sigma_t^2 = \sigma^2 = cte$ ,
3.  $\gamma(t, t - k) = E[(z_t - \mu)(z_{t-k} - \mu)] = \gamma_k \quad k = 0, \pm 1, \pm 2$ .

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera, que la covarianza entre dos variables depende sólo de su separación. En un proceso estacionario las autocovarianzas y autocorrelaciones sólo dependen del retardo entre las observaciones y, en particular, la relación entre  $z_t$  y  $z_{t-k}$ , es siempre igual a la relación entre  $z_t$  y  $z_{t+k}$ . En consecuencia, en los procesos estacionarios:

$$Cov(z_t, z_{t+k}) = Cov(z_{t+j}, z_{t+j+k}) = \gamma_k, \quad j = 0 \pm 1, \pm 2, \dots$$

y también, para las autocorrelaciones:

$$\rho_k = \frac{Cov(z_t, z_{t-k})}{\sqrt{var(z_t)var(z_{t-k})}} = \frac{\gamma_k}{\gamma_0}.$$

En resumen, en los procesos estacionarios  $\gamma_0 = \sigma^2$ , y  $\gamma_k = \gamma_{-k}$ . Para las autocorrelaciones  $\rho_k = \rho_{-k}$ .

Llamaremos *matriz de covarianzas* del proceso estacionario de orden  $k$ ,  $\Gamma_k$ , a la matriz cuadrada y simétrica de orden  $k$  que tiene en la diagonal principal las varianzas y en las diagonales siguientes las autocovarianzas. Tendremos:

$$\begin{aligned} \Gamma_k &= E \left[ \begin{bmatrix} z_t - \mu \\ z_{t-1} - \mu \\ \dots \\ z_{t-k} - \mu \end{bmatrix} \begin{bmatrix} z_t - \mu & z_{t-1} - \mu & \dots & z_{t-k} - \mu \end{bmatrix} \right] = \\ &= \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \dots & \gamma_0 \end{bmatrix} \end{aligned} \quad (3.7)$$

a las matrices cuadradas que, como esta matriz  $\Gamma_k$ , tienen los mismos elementos en cada diagonal se las denomina matrices de *Toeplitz*.

Se denomina *función de autocorrelación simple (fas)* a la representación de los coeficientes de autocorrelación del proceso en función del retardo y *matriz de autocorrelación* a la matriz cuadrada y simétrica de Toeplitz con unos en la diagonal y los coeficientes de autocorrelación fuera de la diagonal:

$$R_k = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix}. \quad (3.8)$$

La estacionaridad débil no garantiza la estabilidad completa del proceso. Por ejemplo, la distribución de las variables  $z_t$  puede estar cambiando en el tiempo. Sin embargo, si suponemos que estas variables tienen *conjuntamente* una distribución normal  $n$ -dimensional, como ésta queda determinada por las medias, las varianzas y las covarianzas, todas las distribuciones marginales serán idénticas, y también lo serán las distribuciones de subconjuntos de hasta  $n-1$  variables del tipo  $(z_{t_1}, z_{t_1+j}, z_{t_1+k})$  y  $(z_{t_2}, z_{t_2+j}, z_{t_2+k})$  para cada  $t$ . Por tanto, para procesos con distribuciones conjuntamente normal la estacionaridad débil coincide con la estricta.

En adelante utilizaremos para simplificar la expresión proceso estacionario, para referirnos a un proceso estacionario en sentido débil, indicando expresamente cuando el proceso es estacionario en sentido estricto.

### 3.3.2. Combinaciones de procesos estacionarios (\*)

Una propiedad importante de los procesos estacionarios es que son estables ante combinaciones lineales, es decir, los procesos obtenidos mediante combinaciones lineales de procesos estacionarios son también estacionarios. En particular, la combinación lineal formada por los incrementos de un proceso estacionario es una serie estacionaria. Es decir, si  $z_t$  es estacionario, el proceso  $w_t$  definido por:

$$w_t = z_t - z_{t-1},$$

es también estacionario. En efecto, vamos a comprobar las tres condiciones de la sección anterior para la estacionariedad en sentido débil. Es inmediato que la esperanza de  $w_t$  es siempre cero. Su varianza será constante, ya que, llamando  $\sigma_a^2$  a la varianza de  $z_t$  y  $\gamma_1$  a la covarianza entre observaciones contiguas:

$$Var(w_t) = Var(z_t) + Var(z_{t-1}) - 2Cov(z_t, z_{t-1}) = 2(\sigma_a^2 - \gamma_1),$$

que no depende de  $t$ . En tercer lugar, su función de autocovarianza es:

$$Cov(w_t, w_{t+k}) = E[(z_t - z_{t-1})(z_{t+k} - z_{t+k-1})] = 2\gamma_k - \gamma_{k+1} - \gamma_{k-1}$$

y depende sólo del retardo  $k$  y no del instante  $t$ .

Vamos a demostrar el resultado general de que toda combinación lineal de procesos estacionarios es un nuevo proceso estacionario (esta prueba puede omitirse sin pérdida de continuidad). Sea  $\mathbf{z}_t = (z_{1t}, \dots, z_{kt})$  un vector de  $k$  procesos estacionarios donde suponemos que las autocovarianzas sólo dependen del retardo y las covarianzas entre dos componentes en dos instantes de tiempo sólo dependen de los dos componentes considerados y del retardo entre los instantes temporales. En estas condiciones, el vector de series es estacionario. Consideraremos el proceso escalar definido por el vector de constantes  $\mathbf{c}' = (c_1, \dots, c_k)$ :

$$y_t = \mathbf{c}' \mathbf{z}_t = c_1 z_{1t} + \dots + c_k z_{kt},$$

que será una combinación lineal de los componentes del vector  $\mathbf{z}_t$ . La esperanza de ese proceso es

$$E(y_t) = c_1 E(z_{1t}) + \dots + c_k E(z_{kt}) = \mathbf{c}' \boldsymbol{\mu},$$

donde  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$  es el vector de las medias de los componentes. Como las esperanzas  $E(z_{it}) = \mu_i$  son constantes, también lo es  $E(y_t)$ . La varianza del proceso  $y_t$  será:

$$var(y_t) = E(\mathbf{c}'(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_t - \boldsymbol{\mu})\mathbf{c}) = \mathbf{c}' \boldsymbol{\Gamma}_z \mathbf{c} \quad (3.9)$$

donde  $\Gamma_z$  es la matriz de covarianzas entre los componentes del vector en el mismo instante. Como los componentes son estacionarios, la matriz de covarianzas entre ellos es también constante. Análogamente, se comprueba que

$$\text{cov}(y_t y_{t+k}) = \mathbf{c}' \Gamma_z(k) \mathbf{c}$$

donde  $\Gamma_z(k)$  contiene las covarianzas entre los componentes en distintos instantes que, por hipótesis, sólo dependen del retardo. Por tanto, el proceso  $y_t$  es estacionario.

Una consecuencia de este resultado es que las autocovarianzas tienen que cumplir ciertas condiciones para que correspondan a un proceso estacionario. Si definimos la combinación lineal de los valores retardados del proceso escalar,  $z_t$ , mediante:

$$y_t = c_1 z_t + c_2 z_{t-1} + \dots + c_k z_{t-k} = \mathbf{c}' \mathbf{z}_{t,k}$$

donde  $\mathbf{z}_{t,k} = (z_t, z_{t-1}, \dots, z_{t-k})'$ . La varianza de la variable  $y_t$  viene dada por  $\mathbf{c}' \Gamma_k \mathbf{c}$ , donde  $\Gamma_k$  tiene la expresión (3.7), será un número no negativo. Esto es equivalente a que la matriz de autocovarianzas de orden  $k$  del proceso  $z_t$  sea definida no negativa. Si consideramos un proceso de varianza unidad, o estandarizamos el proceso dividiendo por su desviación típica, las autocovarianzas coinciden con las autocorrelaciones, por lo que esta propiedad también se aplica a la matriz de autocorrelaciones  $\mathbf{R}_k$  definida por (3.8), que debe ser también definida no negativa.

### 3.4. Proceso de ruido blanco

Un proceso estacionario muy importante es el definido por las condiciones:

- 1a.  $E[z_t] = 0, \quad t = 1, 2, \dots$
- 2a.  $Var(z_t) = \sigma^2, \quad t = 1, 2, \dots$
- 3a.  $Cov(z_t, z_{t-k}) = 0, \quad k = \pm 1, \pm 2, \dots$

Este proceso se denomina *proceso de ruido blanco*. La condición 1a establece que la esperanza es siempre constante e igual a cero, la condición 2a que la varianza es constante y la condición 3a que las variables del proceso están incorreladas para todos los retardos. Por ejemplo, si tiramos una moneda en instantes  $t = 1, 2, \dots$  y definimos  $z_t = -1$  si se obtiene cara y  $z_t = +1$  si se obtiene cruz, se obtiene un proceso de ruido blanco: la esperanza es cero, la varianza es constante (igual a la unidad) y las covarianzas son nulas. Como segundo ejemplo, si apuntamos el número que se obtiene en la extracción de un bombo que contiene el mismo

### 3. Series temporales y procesos estocásticos

número de dígitos entre el cero y el nueve y consideramos la secuencia de valores  $z_t = 4.5$ , llamando  $z_t$  al valor observado en la  $t$ -ésima extracción,  $t = 1, 2, \dots$ , se obtiene también un proceso de ruido blanco. Como tercer ejemplo, si generamos números aleatorios normales de media cero y varianza constante con un ordenador y los colocamos en secuencia, obtenemos, de nuevo, un proceso de ruido blanco. En estos procesos conocer los valores pasados no proporciona ninguna información sobre el futuro, ya que el proceso «no tiene memoria».

Una condición equivalente a la de ruido blanco es suponer un proceso estacionario con varianza marginal finita  $\sigma^2$  y la condición:

$$1b. E(z_t|z_{t-1}, \dots, z_1) = 0$$

ya que puede demostrarse (véase el apéndice 3.2) que entonces se verifican automáticamente las condiciones 1a y 3a anteriores. Cuando el proceso se define mediante la condición 1b suele denominarse, por razones históricas, proceso de *diferencia de martingalas*. La razón es que si el proceso  $y_t$  es una martingala, definida por la condición (3.6), el proceso  $z_t = y_t - y_{t-1}$ , verifica:

$$E(z_t|z_{t-1}, \dots, z_1) = E(y_t|z_{t-1}, \dots, z_1) - E(y_{t-1}|z_{t-1}, \dots, z_1) = 0.$$

En efecto, condicionar al conjunto de variables  $(z_{t-1}, \dots, z_1)$  equivale a condicionar al conjunto equivalente  $(y_{t-1}, \dots, y_1, y_0)$ , ya que uno determina el otro. Al ser  $y_t$  una martingala, la condición (3.6) implica que  $E(y_t|y_{t-1}, \dots, y_0) = y_{t-1}$ , y también  $E(y_{t-1}|y_{t-1}, \dots, y_0) = y_{t-1}$ , con lo que ambas esperanzas son iguales a  $y_{t-1}$  y su diferencia es cero, que es la condición 1b.

Un proceso de ruido blanco no es necesariamente estacionario en sentido estricto ni debe estar formado por variables independientes, ya que sólo se exige la incorrelación. Si imponemos la condición adicional de que las variables del proceso sean independientes, y no sólo incorreladas, llamaremos al proceso de *ruido blanco estricto*. Si suponemos que las variables tienen distribución normal, la incorrelación garantiza la independencia, y la normalidad la igualdad de las distribuciones marginales, con lo que el proceso será de ruido blanco estricto y con variables normales. Llamaremos al proceso resultante *proceso de ruido blanco normal*. La normalidad es una condición fuerte y podemos tener procesos de ruido blanco estricto con variables que tienen distribuciones no normales. Por ejemplo, un proceso de variables uniformes independientes es un ruido blanco estricto, pero no un ruido blanco normal.

### 3.5. Estimación de los momentos de procesos estacionarios

Supongamos un proceso estacionario con media  $\mu = E(z_t)$ , varianza  $\sigma^2 = \gamma_0 = Var(z_t)$  y covarianzas  $\gamma_k = Cov(z_t, z_{t-k})$  del que se observa una realización  $(z_1, \dots, z_T)$ . Vamos a estudiar cómo estimar la media, la varianza, las covarianzas y las autocorrelaciones del proceso a partir de esta única realización disponible.

#### 3.5.1. Estimación de la media

Un estimador centrado de la media poblacional es la media muestral. En efecto, llamando  $\bar{z}$  la media muestral:

$$\bar{z} = \frac{\sum_{t=1}^T z_t}{T},$$

se verifica que

$$E(\bar{z}) = \frac{\sum_{t=1}^T E(z_t)}{T} = \mu.$$

Para datos independientes la varianza de la media muestral como estimador de la media poblacional es  $\sigma^2/T$ . En consecuencia, al aumentar el tamaño muestral el error cuadrático medio de estimación, dado por

$$E(\bar{z} - \mu)^2$$

que coincide con la varianza de  $\bar{z}$ , al ser el estimador centrado, tiende a cero. En un proceso estocástico estacionario esta propiedad no es necesariamente cierta, y es posible que al aumentar el tamaño muestral la varianza de la estimación de la media no tienda a cero. Por ejemplo, supongamos un proceso que comienza en el instante  $t = 1$ , y la primera variable,  $z_1$ , es una variable aleatoria de media cero y varianza  $\sigma^2$ . A continuación los valores posteriores del proceso son iguales al valor inicial de manera que  $z_t = z_{t-1} = \dots = z_1$ . En este proceso estacionario la esperanza de cada variable es cero, ya que en distintas realizaciones la variable inicial irá tomando distintos valores y en promedio tomará el valor cero, pero si disponemos de una realización única del proceso siempre  $\bar{z} = z_1$  y la varianza de esta estimación es  $\sigma^2$ , sea cual sea el tamaño de la realización. En adelante, excluiremos este tipo de situaciones donde las nuevas observaciones tienen correlación uno con las anteriores y, en consecuencia, no proporcionan información. Decimos que el proceso es *ergódico* para la estimación de la media cuando se verifica que:

$$\lim_{T \rightarrow \infty} E(\bar{z} - \mu)^2 \rightarrow 0.$$

### 3. Series temporales y procesos estocásticos

---

Para identificar las condiciones para que un proceso sea ergódico para la estimación de la media, vamos a calcular el error cuadrático medio de la media muestral como estimador de la media poblacional para procesos estacionarios. Como el estimador  $\bar{z}$  es centrado para  $\mu$ , el error cuadrático medio coincide con la varianza del estimador y tenemos:

$$var(\bar{z}) = E(\bar{z} - \mu)^2 = \frac{1}{T^2} E\left(\sum_{t=1}^T (z_t - \mu)\right)^2$$

que podemos escribir como:

$$var(\bar{z}) = \frac{1}{T^2} \left[ \sum_{t=1}^T E(z_t - \mu)^2 + 2 \sum_{i=1}^T \sum_{j=i+1}^T E((z_i - \mu)(z_j - \mu)) \right],$$

El primer sumatorio dentro de los corchetes es  $T\sigma^2$ . El segundo sumatorio doble contiene  $T - 1$  veces las covarianzas de orden uno,  $T - 2$  veces las covarianzas de orden 2 y, en general,  $T - i$  veces las de orden  $i$ . En consecuencia, la varianza de la media muestral puede escribirse como:

$$var(\bar{z}) = \frac{1}{T} \left[ \sigma^2 + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \gamma_i \right]. \quad (3.10)$$

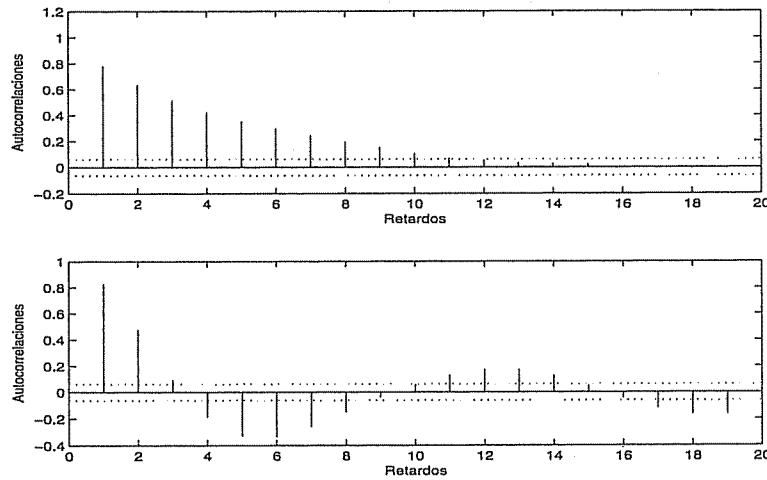
Observemos en primer lugar que si todas las covarianzas son cero, es decir, tenemos un proceso de ruido blanco, la varianza de la media muestral es  $\sigma^2/T$ . Sin embargo, cuando las covarianzas no sean cero, la varianza de la media muestral para procesos estacionarios puede ser considerablemente mayor que para observaciones independientes. En efecto, si las  $\gamma_i$  son positivas, el sumatorio puede ser muy grande. La condición para que la  $var(\bar{z})$  tienda a cero al aumentar  $T$  es que el sumatorio converja a una constante al aumentar  $T$ . Una condición necesaria (aunque no suficiente, véase el apéndice 3.1) de que la suma converja es:

$$\lim_{i \rightarrow \infty} \gamma_i \rightarrow 0,$$

lo que supone que la dependencia entre observaciones tiende a cero al aumentar el retardo. Por ejemplo, una serie constante tiene correlación uno entre las observaciones y no es ergódica. Tampoco lo será una serie con componentes deterministas periódicos, como la serie:

$$z_t = A \cos wt + a_t,$$

donde  $A$  y  $w$  son constantes para toda la realización. Las covarianzas entre las observaciones nunca van a cero al aumentar el retardo, por el

**Figura 3.6 Correlograma de dos procesos estacionarios**


caracter periódico de la serie, y aumentar el tamaño muestral no ayuda a mejorar la estimación de la media poblacional.

La ecuación (3.10) puede escribirse utilizando las autocorrelaciones como:

$$var(\bar{z}) = \frac{\gamma_0}{T} \left[ 1 + 2 \sum_{i=1}^{T-1} \left( 1 - \frac{i}{T} \right) \rho_i \right]$$

donde  $\rho_i = \gamma_i / \gamma_0$  son las autocorrelaciones del proceso. Observemos que, como ocurría con las autocovarianzas, una condición suficiente para que el sumatorio converja y el proceso sea ergódico en la estimación de la media es que  $\lim_{i \rightarrow \infty} \rho_i \rightarrow 0$ .

En resumen, la propiedad de ergodicidad, que puede aplicarse a la estimación de cualquier parámetro, se satisface si las nuevas observaciones del proceso proporcionan información adicional sobre el parámetro, de manera que al aumentar el tamaño muestral el error de estimación tiende a cero. En la estimación de la media esto no ocurrirá si existe una dependencia tan fuerte que las nuevas observaciones son previsibles desde el pasado, y no proporcionan información adicional para estimar la media.

La figura 3.6 presenta las autocorrelaciones de dos procesos estacionarios. El primero es ergódico, ya que las autocorrelaciones tienden a cero con el retardo, mientras que el segundo no lo es. En adelante supondremos que el proceso es ergódico, lo que supone en la práctica que hemos eliminado posibles términos sinusoidales deterministas de la forma  $A \cos(\omega t + \delta)$  que pudieran existir.

### 3.5.2. Estimación de las autocovarianzas y autocorrelaciones

Si la media del proceso fuese conocida, el estimador de las autocovarianzas de orden  $k$  es:

$$\tilde{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^T (z_t - \mu)(z_{t-k} - \mu), \quad (3.11)$$

y es fácil comprobar al tomar esperanzas que este estimador es centrado para estimar  $\gamma_k = E((z_t - \mu)(z_{t-k} - \mu))$ . Sin embargo, cuando  $\mu$  es desconocida y la sustituimos en la fórmula (3.11) por su estimador,  $\bar{z}$ , se demuestra que el estimador resultante no es centrado. Un estimador alternativo, que tiene mejores propiedades cuando  $\mu$  es desconocida, es:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (z_t - \bar{z})(z_{t-k} - \bar{z}), \quad (3.12)$$

que, aunque es también un estimador sesgado de la autocovarianza poblacional, tiene menor error cuadrático de estimación que el anterior. En particular, la varianza del proceso se estima mediante  $\hat{\gamma}_0$ .

Una ventaja adicional de utilizar el estimador (3.12) es que la matriz de autocovarianzas muestral:

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \dots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \dots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \dots & \hat{\gamma}_0 \end{bmatrix}$$

es siempre definida no negativa. Esto puede no ocurrir si en lugar de dividir por el tamaño muestral lo hacemos por el número de términos en la suma, como en el estimador (3.11). Ésta es una ventaja importante, porque, como vimos en la sección 3.3.2, esta propiedad es necesaria para que las covarianzas estimadas puedan corresponder a un proceso estacionario.

Las autocorrelaciones se estiman por:

$$r_k = \hat{\gamma}_k / \hat{\gamma}_0,$$

y podemos representar la función de autocorrelación dando los coeficientes de correlación muestral o estimados en función del retardo. A esta representación se la denomina *correlograma, o función de autocorrelación muestral*.

El vector de autocorrelaciones muestrales,  $r = (r_1, \dots, r_k)'$  se distribuye aproximadamente para  $T$  grande de forma normal con media  $\rho$ , el vector de autocorrelaciones teóricas y matriz de covarianzas  $\mathbf{V}_\rho / T$  donde

los términos de la matriz  $\mathbf{V}_\rho$  vienen dados por las fórmulas de Bartlett (véase Fuller, 1996, para una prueba de este resultado y el apéndice 3.3 de este capítulo para las expresiones). En particular, si sólo los primeros  $q$  coeficientes de autocorrelación son distintos de cero, las varianzas de las autocorrelaciones estimadas se aproximan por:

$$\text{var}(r_k) = \frac{T - k}{T(T + 2)}(1 + 2 \sum_{j=1}^q \rho_j^2), \quad k > q \quad (3.13)$$

Por tanto, si todas las autocorrelaciones del proceso son nulas, la varianza asintótica de los coeficientes de autocorrelación estimados o muestrales es:

$$\text{var}(r_k) = \frac{T - k}{T(T + 2)},$$

que puede aproximarse, para  $T$  grande, por  $1/T$ . En la práctica, con series reales cuando  $T$  no es muy grande podemos tener pocos términos para calcular la autocorrelación muestral de orden  $k$  y se recomienda que el orden máximo para el que se calculen autocorrelaciones muestrales sea  $k_{\max} \leq T/4$ .

En la figura 3.9 se ha representado el correlograma de una serie de ruido blanco incluyendo dos líneas paralelas al eje  $x$  de altura  $\pm 2/\sqrt{T}$ . Estas líneas proporcionan, aproximadamente, un intervalo del 95 % donde deberían estar los coeficientes de autocorrelación muestrales si la serie se ha generado como un proceso de ruido blanco. Como vemos, 19 de los 20 están dentro de los límites, confirmando que el proceso es de variables incorreladas. En efecto, como el intervalo es del 95 %, esperamos que si los intervalos son correctos el 5 % de los coeficientes salgan fuera de las bandas, que coincide con el resultado observado en la figura.

Señalaremos, por último, que las estimaciones de los coeficientes de autocorrelación y de las covarianzas están a su vez correladas entre sí, y tanto más cuanto mayor sean los coeficientes teóricos que estamos estimando (véase el apéndice 3.3). Sólo cuando todos los coeficientes de autocorrelación teóricos son cero no existe correlación entre las estimaciones.

---

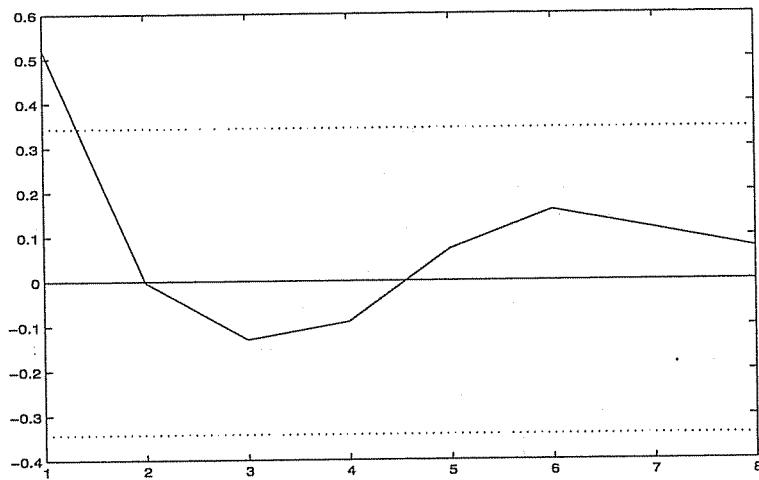
### Ejemplo 3.1

Vamos a calcular la función de autocorrelación muestral o correlograma para los datos de las leguas recorridas por Colón. Comenzaremos calculando la media de la serie:

$$\bar{x} = \frac{9 + 45 + \dots + 59 + 49}{34} = 31.83$$

### 3. Series temporales y procesos estocásticos

Figura 3.7 Función de autocorrelación para los datos de Colón



y la varianza de los datos:

$$s_z^2 = \frac{(9 - 31.83)^2 + (45 - 31.83)^2 + \dots + (49 - 31.83)^2}{34} = 266.97$$

El coeficiente de autocorrelación de primer orden se calcula como:

$$\begin{aligned} r_1 &= \frac{[(9 - 31.83)(45 - 31.83) + \dots + (59 - 31.83)(49 - 31.83)]}{34 \cdot 266.97} = \\ &= .52 \end{aligned}$$

y el de orden dos

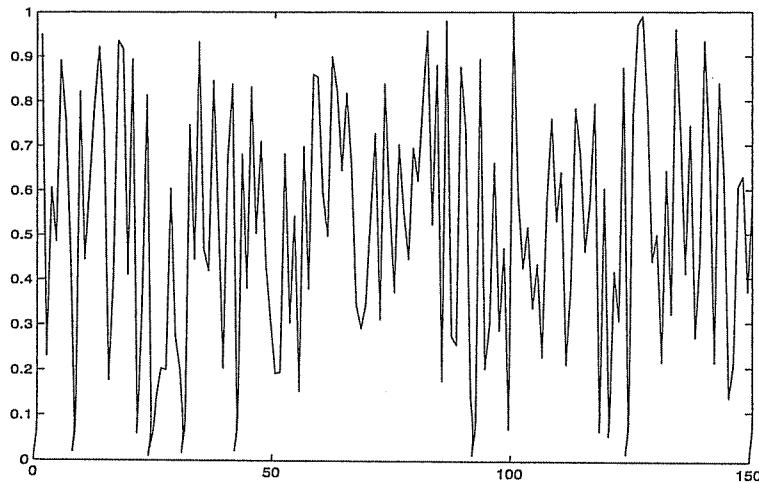
$$\begin{aligned} r_2 &= \frac{[(9 - 31.83)(60 - 31.83) + \dots + (31 - 31.83)(49 - 31.83)]}{34 \cdot 266.97} = \\ &= -.002 \end{aligned}$$

y así sucesivamente. La figura 3.7 presenta la función de autocorrelación muestral obtenida. El coeficiente de correlación de orden cero, de la variable con sí misma, es siempre uno y no se ha presentado en el gráfico que incluye las bandas de confianza a distancia  $2/\sqrt{T}$ . Únicamente el primer coeficiente de autocorrelación parece ser significativo, indicando que las leguas recorridas un día determinado dependen de las recorridas el día anterior, pero no de las de otros días previos.

#### Ejemplo 3.2

La figura 3.8 presenta una simulación de un proceso de ruido blanco y la figura 3.9 los coeficientes de autocorrelación muestrales para retardos ma-

**Figura 3.8 Simulación de 150 observaciones de un proceso de ruido blanco**



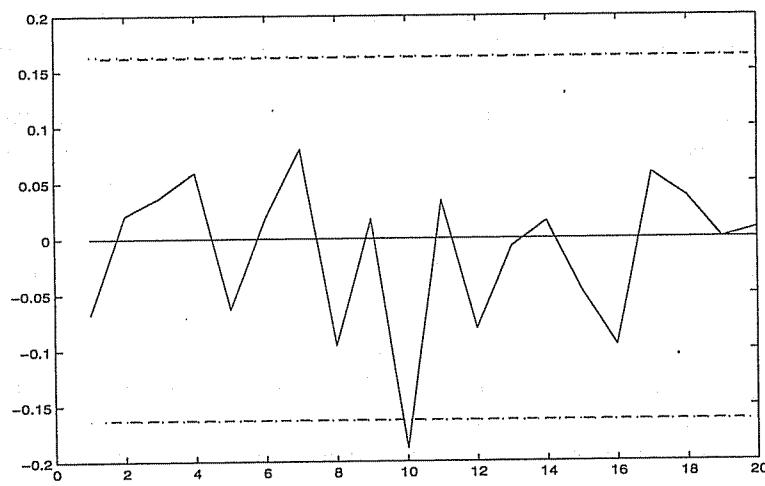
yores que cero de esta serie. Se han dibujado en el gráfico dos bandas de confianza del 95 % para las autocorrelaciones estimadas a una distancia  $2/\sqrt{T}$ , que representan el intervalo de confianza aproximado al 95 % donde deben estar los coeficientes estimados si la serie se ha generado por un proceso de ruido blanco. Observemos que un intervalo del 95 % implica que es esperable que, en promedio, cuando el proceso es verdaderamente ruido blanco una de cada 20 autocorrelaciones salga fuera de las bandas. Esto es lo que ocurre en la figura 3.9, donde sólo uno de los coeficientes estimados está fuera de las bandas.

### Ejemplo 3.3

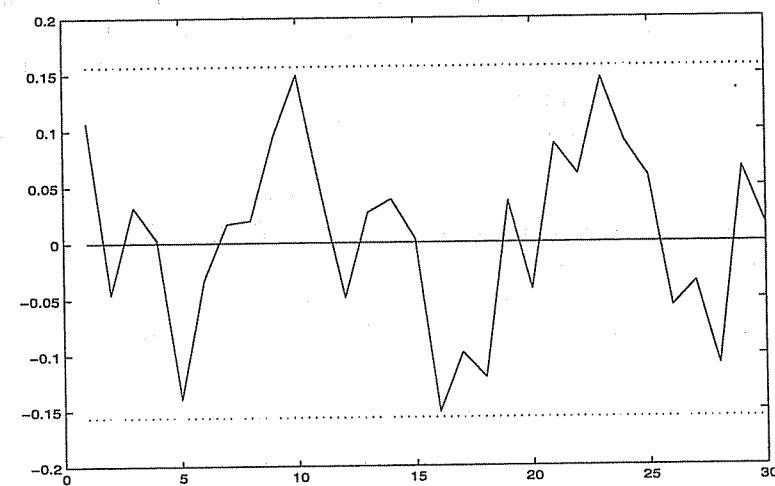
La figura 3.10 presenta la función de autocorrelación para la serie de rendimientos en la Bolsa de Madrid. Se observa que todos los coeficientes de autocorrelación son pequeños, y están dentro de las bandas de  $2/\sqrt{T}$ . Esto sugiere que el proceso es ruido blanco, y los rendimientos pasados no dan información para prever los rendimientos futuros. El lector puede pensar que la figura muestra una apariencia cíclica, pero si se generan 30 números aleatorios normales con un ordenador y realiza su gráfico, comprobará que tienen una forma parecida.

---

**Figura 3.9 Función de autocorrelación para un proceso de ruido blanco**



**Figura 3.10 Función de autocorrelación de los datos de los rendimientos con el índice general de la Bolsa de Madrid, 1988 a 2000**



### 3.6. El espectro de un proceso estacionario (\*)

Esta sección puede omitirse en primera lectura ya que no se utiliza en el resto del libro. Se introduce para facilitar al lector la comprensión de la

literatura de series que utiliza el espectro. En el capítulo 2 se introdujo el *periodograma*, como una función que nos informa sobre la importancia relativa, medida por la variabilidad explicada, de las posibles periodicidades en la serie temporal. Vimos que dada una serie temporal que se ha generado como suma de ondas armónicas deterministas, es posible descomponer su variabilidad en términos asociados a cada onda y que el periodograma para una frecuencia  $f_j$  era una medida de la varianza debida al componente de esta frecuencia. Se demuestra en el apéndice 3.4 que una expresión alternativa del periodograma de una serie es:

$$I(f_j) = 2(\hat{\gamma}_0 + 2 \sum_{k=1}^{T-1} \hat{\gamma}_k \cos 2\pi f_j k) = 2\hat{\gamma}_0(1 + 2 \sum_{k=1}^{T-1} r_j \cos 2\pi f_j k)$$

que pone de manifiesto que el periodograma es una combinación lineal de las autocovarianzas o autocorrelaciones muestrales calculadas a partir de la serie observada. El periodograma es una forma alternativa de representar la estructura de dependencia lineal observada en una serie. Conocidas las autocovarianzas muestrales, podemos obtener el periodograma según al fórmula anterior, y se demuestra que, análogamente, dado el periodograma  $I(f_j)$  podemos calcular las autocovarianzas de la serie.

Vimos también en el capítulo 2 que una forma simple de suavizar el periodograma es representarlo como un histograma, con longitudes de la base  $1/T$  y alturas de los rectángulos iguales a  $I(f_j)$ . Entonces, el área encerrada por este periodograma suavizado es la varianza de la serie. Supongamos ahora que la longitud de la serie observada tiende a infinito, de manera que las propiedades de la serie se aproximen a las propiedades del proceso estocástico que ha generado la realización observada. Las frecuencias básicas  $1/2 \geq f_j \geq 1/T$  tenderán entonces a cubrir todo el intervalo  $1/2 \geq f \geq 0$ , y el periodograma suavizado tenderá a una curva suave, que dependerá del proceso generador de la serie temporal. Si el proceso que ha generado los datos es estacionario, las autocovarianzas estimadas  $\hat{\gamma}_k$  tenderán a las teóricas del proceso estacionario,  $\gamma_k$ , y se define el espectro de este proceso mediante la expresión:

$$S(f) = 2(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi fk). \quad (3.14)$$

Esta función está definida entre  $0 \leq f \leq 1/2$  y existe para todo proceso estacionario y ergódico. Se demuestra en el apéndice 3.4 que el espectro es una transformación matemática de la función de autocovarianzas del proceso y que ambas recogen la misma información. Es decir, conocidas las autocovarianzas puede calcularse el espectro del proceso y al contrario. Los análisis basados en las covarianzas suelen denominarse *análisis en el*

### 3. Series temporales y procesos estocásticos

*dominio temporal*, ya que las covarianzas definen la relación dinámica temporal entre las observaciones del proceso. Los análisis basados en el espectro se suelen denominar *análisis en el dominio frecuencial*, ya que el espectro tiene en cuenta la contribución de los distintos armónicos a la variabilidad de la serie. La utilidad del espectro es que, como ocurre con el periodograma, podemos detectar los ciclos principales que forman el proceso. La relación entre el periodograma y el espectro es similar a la de histograma y función de densidad: la primera corresponde a una muestra y la segunda a la población. Por analogía con el periodograma, el área bajo el espectro en un intervalo de frecuencias representa la varianza que explican las funciones sinusoidales con frecuencia en ese intervalo. Observemos que un proceso de variables incorreladas tiene  $\gamma_k = 0$  para  $k \geq 1$ , con lo que su espectro será constante, y dado por

$$S(f) = 2\sigma^2$$

siendo  $\sigma^2$  la varianza del proceso. Esto justifica la denominación de ruido blanco para estos procesos, por analogía con el espectro físico de la luz blanca que es constante. (En física el espectro representa la intensidad de los componentes de una señal en función de la longitud de onda, y es de allí donde proviene la representación estadística.) Se demuestra que el área bajo el espectro proporciona siempre la varianza del proceso estacionario:

$$\gamma_0 = \int_0^{1/2} S(f) df$$

resultado que es esperable por la intuición del espectro como límite del periodograma suavizado.

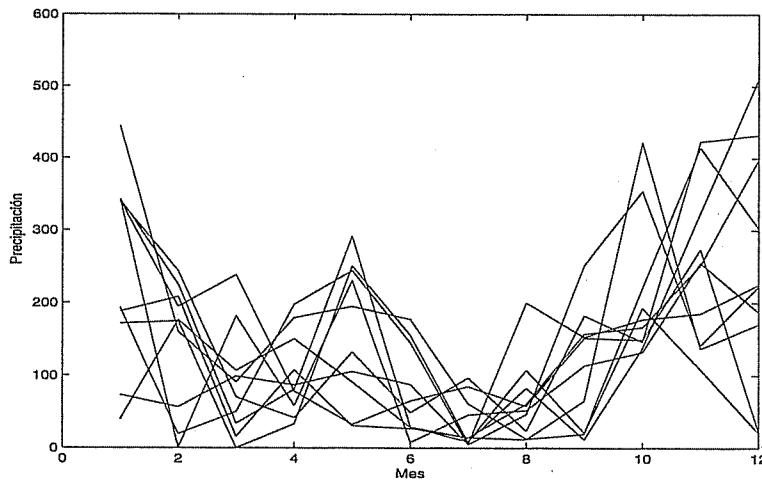
Se define la función de densidad espectral como:

$$s(f) = \frac{S(f)}{\gamma_0} = 2(1 + 2 \sum_{k=1}^{\infty} \rho_k \cos 2\pi f k).$$

El nombre de densidad proviene de que tiene las mismas propiedades que una función de densidad de probabilidad: es no negativa,  $s(f) \geq 0$ , e integra a uno en su dominio de definición (0,.5). Representa la contribución relativa de los armónicos en cada intervalo a la variabilidad total del proceso.

El análisis temporal suele ser más útil en series económicas, sociales o biológicas, donde no son esperables ciclos deterministas importantes. El análisis frecuencial se ha utilizado mucho en ingeniería eléctrica y de telecomunicaciones, en series formadas por combinaciones lineales de señales periódicas contaminadas con un ruido aditivo. El espectro del proceso permite entonces, como vimos con el periodograma, obtener la

**Figura 3.11 Precipitación mensual en Santiago de Compostela en el periodo 1988-1997**



importancia relativa de los distintos armónicos. El espectro tiene además un interés teórico, ya que puede demostrarse que podemos construir una representación de cualquier proceso estacionario como una suma infinita de funciones sinusoidales para todas las frecuencias en el intervalo  $[0, 1/2]$  con amplitudes que son variables aleatorias independientes. Esta caracterización de los procesos estacionarios permite que muchas propiedades de estos procesos pueden estudiarse mejor en el dominio frecuencial que en el temporal.

---

### Ejercicios 3

- 3.1 La figura 3.11 presenta la lluvia mensual caída en Santiago de Compostela en los años 1988-1997. Los datos están en el fichero lluviaSC.dat. Justificar las hipótesis necesarias para considerar que estas diez series son realizaciones de un proceso estocástico. ¿Sería el proceso estacionario?
- 3.2 Consideremos el proceso  $z_t = .5z_{t-1} + a_t$  donde  $a_t$  es un proceso de ruido blanco. Calcular su media marginal, su varianza y su autocovarianza de primer orden. ¿Es el proceso estacionario?
- 3.3 En el proceso del ejercicio anterior calcular la esperanza y la varianza de la distribución condicionada  $f(z_t|z_{t-1})$ . Comparar estos resultados con los obtenidos en el ejercicio anterior para la distribución marginal.

### 3. Series temporales y procesos estocásticos

- 3.4 Consideremos el proceso  $z_t = -.5a_{t-1} + a_t$ . Calcular su media y sus autocovarianzas de orden uno y dos. ¿Es el proceso estacionario?
- 3.5 Demostrar que el proceso anterior puede escribirse  $z_t = -.5(z_{t-1} + .5a_{t-2}) + a_t = -.5z_{t-1} - .25a_{t-2} + a_t$ . Utilizar esta expresión para calcular la esperanza y la varianza de la distribución condicionada  $f(z_t|z_{t-1})$ . Comparar los resultados con los de la distribución marginal.
- 3.6 Dada una serie estacionaria, demostrar que si las autocovarianzas son todas positivas, la media del proceso se estimará con mayor varianza que en el caso en que todas las autocovarianzas son nulas.
- 3.7 Demostrar que la media del proceso del ejercicio 3.4 se estima con menor variabilidad que en un proceso con datos independientes y la misma varianza marginal.

---

#### Apéndice 3.1: Condición de convergencia

Dada una secuencia de números  $a_i$ , una condición necesaria y suficiente para que la serie  $\sum a_i$  converja es que para todo  $\varepsilon > 0$  exista un número  $N$  que depende de  $\varepsilon$  y que escribiremos  $N(\varepsilon)$ , tal que para números cualesquiera,  $n, m$ , mayores que  $N(\varepsilon)$ , se verifique:

$$|a_{n+1} + a_{n+2} + \dots + a_m| < \varepsilon, \quad \text{si } m > n > N(\varepsilon)$$

En particular tomando  $m = n + 1$  se obtiene que una condición necesaria, aunque no suficiente, para la convergencia de la serie es que

$$\lim_{n \rightarrow \infty} a_n = 0$$

Por ejemplo la serie armónica  $a_n = 1/n$  verifica que  $\lim_{n \rightarrow \infty} a_n = 0$  pero  $\sum a_i$  es divergente. Una condición suficiente para la convergencia de  $\sum a_i$  es que converja  $\sum |a_i|$ , entonces se dice que la serie es absolutamente convergente.

#### Apéndice 3.2: Diferencia de martingalas

Supongamos un proceso estacionario con varianza finita que verifica  $E(z_t|z_{t-1}, \dots, z_1) = 0$ . Utilizando que, para variables con esperanza finita,

$$E(y) = E_x E(y|x),$$

tomando  $y = z_t$ ,  $x = (z_{t-1}, \dots, z_1)$  y aplicando esta relación si  $E(z_t|z_{t-1}, \dots, z_1) = 0$ , concluimos que si la esperanza condicionada es cero la esperanza

marginal de  $z_t$  es siempre cero. Además, tomando  $y = (z_t|z_{t-1}, \dots, z_{t-k})$  y  $x = (z_{t-k-1}, \dots, z_1)$  obtenemos que también  $E(z_t|z_{t-1}, \dots, z_{t-k}) = 0$ , es decir la esperanza condicionada para cualquier  $k$  es también nula.

Vamos a comprobar que la condición de esperanza condicionada nula implica que las variables están además incorreladas. Como las variables tienen media cero:

$$Cov(z_t, z_{t-k}) = E(z_t z_{t-k}) = E_{z_{t-k}} E(z_t z_{t-k} | z_{t-k}) = 0,$$

ya que  $E(z_t z_{t-k} | z_{t-k}) = z_{t-k} E(z_t | z_{t-k}) = 0$ . Por tanto, un proceso estacionario que se define como una diferencia de martingalas por la condición  $E(z_t | z_{t-1}, \dots, z_1) = 0$  es de variables incorreladas, es decir, un proceso de ruido blanco.

### Apéndice 3.3: Las fórmulas de Bartlett para las autocovarianzas estimadas

La varianza asintótica del coeficiente de autocorrelación de orden  $k$  es:

$$var(r_k) = \frac{1}{T} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i-k} \rho_{i+k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_i^2 \rho_k^2)$$

mientras que las covarianzas vienen dadas por:

$$\begin{aligned} cov(r_k, r_{k+h}) &= \frac{1}{T} \sum_{i=-\infty}^{\infty} (\rho_i \rho_{i-h} + \rho_{i-k} \rho_{i+k+h} \\ &\quad - 2\rho_{k+h} \rho_i \rho_{i-h} - 2\rho_k \rho_i \rho_{i-k-h} + 2\rho_i^2 \rho_k \rho_{k+h}). \end{aligned}$$

Si suponemos que sólo las primeras  $q$  autocorrelaciones son cero, entonces para  $k > h$  las correlaciones entre los coeficientes estimados son:

$$corr(r_k, r_{k+h}) = \frac{2 \sum_{i=0}^{\infty} \rho_i \rho_{i+k}}{1 + 2 \sum_{j=1}^q \rho_j^2}$$

y vemos que sólo cuando todos los coeficientes de autocorrelación teóricos son cero no existe correlación entre las estimaciones.

### Apéndice 3.4: El espectro y las autocovarianzas

Supongamos, para simplificar la exposición y sin pérdida de generalidad, una serie temporal de media muestral cero. En el capítulo 2 se demostró que el periodograma viene dado por:

$$I(f_j) = \frac{T \widehat{R}_j^2}{2} = \frac{T}{2} \left[ \left( \frac{2}{T} \sum_{t=1}^T z_t \sin(wt) \right)^2 + \left( \frac{2}{T} \sum_{t=1}^T z_t \cos(wt) \right)^2 \right]$$

### 3. Series temporales y procesos estocásticos

---

que puede escribirse

$$I(f_j) = \frac{2}{T} \left[ \sum_{t=1}^T \sum_{v=1}^T z_t \operatorname{sen}(wt) z_v \operatorname{sen}(wv) + \right. \\ \left. + \sum_{t=1}^T \sum_{v=1}^T z_t \cos(wt) z_v \cos(wv) \right]$$

y utilizando que  $\cos(a - b) = \cos(a)\cos(b) + \operatorname{sen}(a)\operatorname{sen}(b)$ :

$$I(f_j) = \frac{2}{T} \sum_{t=1}^T \sum_{v=1}^T z_t z_v \cos(w(t-v)) = 2 \sum_{t=1}^T \sum_{v=1}^T \left( \frac{1}{T} z_t z_v \right) \cos(w(t-v)).$$

En esta suma vemos que aparecerán las sumas de términos  $(z_t z_v / T)$ , que son las autocovarianzas muestrales. Vamos a ver los coeficientes con los que aparecerán las autocovarianzas de distintos órdenes. La covarianza  $\hat{\gamma}_0$  vendrá dada por los términos con  $t = v$ , entonces  $\cos(w(t-v)) = 1$ . Para  $\hat{\gamma}_1$  necesitaremos  $v = t-1$  o también  $v = t+1$ , por lo que tendremos dos veces la suma  $\sum z_t z_{t-1} / T$  y el coeficiente del coseno será  $\cos w$ . Análogamente, para  $\hat{\gamma}_k$  tendremos dos veces la suma  $\sum z_t z_{t-k} / T$  y el coeficiente del coseno será  $\cos wk$ , y podemos escribir:

$$I(f_j) = 2(\hat{\gamma}_0 + 2 \sum_{k=1}^T \hat{\gamma}_k \cos 2\pi f k).$$

Cuando  $T \rightarrow \infty$ , si suponemos que la serie es una realización de un proceso estocástico estacionario, las autocovarianzas muestrales convergen a las del proceso y tendremos la expresión límite del espectro:

$$S(f) = 2(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi f k). \quad (3.15)$$

Una expresión alternativa del espectro es:

$$S(f) = 2 \sum_{k=-\infty}^{\infty} \gamma_k e^{-i2\pi f k},$$

ya que  $\gamma_k = \gamma_{-k}$ ,  $\cos(x) = \cos(-x)$  y  $\operatorname{sen}(x) = -\operatorname{sen}(-x)$ , al desarrollar  $e^{-i2\pi f k}$  como  $\cos 2\pi f k + i \operatorname{sen} 2\pi f k$ , la suma de los senos se anulará y la suma de los cosenos será igual para  $k$  negativo y positivo, con lo que resulta 3.15.

El espectro se expresa a veces en función de la frecuencia angular  $w = 2\pi f$ , y en lugar de representarse en el intervalo  $(0, .5)$  se representa en el intervalo  $(0, \pi)$ . Para que se mantenga la propiedad de que la varianza del proceso sea igual al área bajo el espectro, éste se define como:

$$S(w) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-iwk}.$$

## Análisis de series temporales

---

Algunos autores prefieren definir el espectro en el intervalo  $(-\pi, \pi)$ , con  $S^*(-w) = S^*(w)$ . Su expresión entonces es:

$$S^*(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k}.$$

Estas fórmulas ponen de manifiesto que, matemáticamente, el espectro es la transformada de Fourier de la función de autocovarianzas. Por ejemplo, utilizando la representación con frecuencias angulares,  $S(w)$ , puede demostrarse que existe la transformación inversa, dada por:

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} S(w) dw.$$

En resumen, el espectro y la función de covarianzas contienen la misma información y podemos pasar de una a la otra mediante las ecuaciones anteriores.

## Apéndice 3.5: Generación de procesos con el ordenador

### Minitab

La generación de un proceso de ruido blanco normal con Minitab se hace desde el menú superior. Marcar *Cal → Random Data → Normal* y se abre una ventana donde se indica la longitud de la serie (*generate... rows*), los parámetros de la normal y la columna para almacenar la variable. Una vez generada la serie, podemos hacer un gráfico marcando en el menú: *Graph → Time series plot*. También podemos hacer un gráfico de la función de autocorrelación de los datos marcando *Stat → Time series → Autocorrelation*.

### SPSS

La generación de un ruido blanco con SPSS es similar a Minitab.

### EViews

Generemos una serie de 200 observaciones de un proceso ruido blanco de esperanza 0 y varianza 1. Para ello, procedemos creando un workfile de 200 observaciones:

File → New → Workfile

Seleccionamos la opción:

Undated or Irregular → 1 to 200

### 3. Series temporales y procesos estocásticos

Hemos creado un workfile de 200 datos sin fecha. A continuación, en el workfile marcamos:

Genr → x=nrnd

y habremos creado una serie temporal de 200 datos generada por un ruido blanco gaussiano. Haciendo doble click sobre el ícono de la serie, podemos utilizar el menú View para determinar las características principales de este proceso del cual tenemos la información procedente de la serie generada. Vemos el gráfico de la serie generada:

View → Line Graph.

En principio deberíamos ver una serie de datos que aproximadamente tenga media 0 y varianza constante. En este caso, y puesto que todos los datos han sido generados por la misma distribución, es interesante obtener un histograma y los principales estadísticos:

View → Descriptive Statistics → Histogram and Stats

Hay que comparar los valores teóricos del proceso y los obtenidos para la serie generada. Otra manera interesante de comparar la distribución de los datos con respecto la normal es el gráfico quantile-quantile:

View → Distribution Graph →  
→ Quantile-Quantile: Normal distributions

#### Matlab

Una serie de ruido blanco se genera escribiendo la sentencia:  $x=randn(n,1)$  que crea  $n$  valores de un ruido blanco normal. Podemos hacer un gráfico temporal con la sentencia:  $plot(x)$ . Para calcular las autocorrelaciones de la serie generada podemos utilizar el siguiente programa

```
function r=acf(x,k)
% Esta función proporciona las autocorrelaciones, r, para k retardos de una
% serie que se encuentra en el vector COLUMNA x.
n=length(x);
m=mean(x);
v=cov(x);
for i=1:k
    ax=x(1:n-i);
    lx=x(i+1:n);
    r(i)=(lx-m)'*(ax-m)/((n-1)*v);
end
y=r;
l=1:1:k;
plot(l,y,'-')
```

## Análisis de series temporales

---

```
hold
cero=zeros(size(y));
plot(l,cero,'-')
lim=2/sqrt(n);
ls=lim*ones(size(y));
plot(l,ls,:')
li=-ls;
plot(l,li,:')
hold off
```

---

## 4. Procesos autorregresivos



George Udny Yule (1871-1951)

Estadístico escocés. Se graduó en ingeniería en University College y estudió después física en Bonn con Hertz. A su vuelta al Reino Unido conoció a K. Pearson y decidió dedicarse a la estadística. Desarrolló el modelo de regresión con varias variables, introdujo el coeficiente de correlación múltiple y el de correlación parcial e inventó los procesos autorregresivos. Profesor en Oxford, su insaciable curiosidad le llevó a aprender a pilotar aviones al jubilarse.

### 4.1. Introducción

En este capítulo vamos a iniciar el estudio de modelos de procesos estacionarios que son útiles en la práctica para representar la dependencia de los valores de una serie temporal de su pasado. Los modelos más simples son los autorregresivos, que generalizan la idea de regresión para representar la dependencia lineal entre dos variables aleatorias. Recordemos que el modelo de regresión simple explica la evolución de una variable,  $y_t$ , como función lineal de otra variable,  $x_t$ , mediante la ecuación:

$$y_t = c + bx_t + a_t$$

donde  $c$  y  $b$  son constantes a determinar y  $a$  es una variable aleatoria normal, con media nula y varianza constante. Si aplicamos esta estructura de dependencia a las observaciones contiguas en una serie temporal,  $z_t$ , y tomamos  $y_t = z_t$  y  $x_t = z_{t-1}$ , obtenemos el modelo de dependencia que llamamos proceso autorregresivo de primer orden. Este es un proceso de Markov donde el valor presente de la serie sólo depende de forma lineal del último valor observado. Esta dependencia lineal puede generalizarse

haciendo que el valor actual de la serie,  $z_t$ , dependa no sólo de  $z_{t-1}$ , sino también de los  $p$  retardos anteriores,  $z_{t-2}, \dots, z_{t-p}$ . Entonces, se obtiene un proceso autorregresivo de orden  $p$ . Los procesos autorregresivos son los primeros procesos estacionarios que se estudiaron.

#### 4.2. El proceso autorregresivo de primer orden (AR(1))

Diremos que una serie  $z_t$  sigue un proceso autorregresivo de primer orden, o un AR(1), si ha sido generada por:

$$z_t = c + \phi z_{t-1} + a_t, \quad (4.1)$$

donde  $c$  y  $-1 < \phi < 1$  son constantes a determinar y  $a_t$  es un proceso de ruido blanco con varianza  $\sigma^2$ . Las variables  $a_t$ , que representan la nueva información que se añade al proceso en cada instante, se conocen como innovaciones.

Como ejemplo de una situación donde es esperable la aparición de este proceso, consideramos que  $z_t$  es la cantidad de agua a fin de mes en un embalse. Durante el mes llega al embalse una cantidad  $c+a_t$ , donde  $c$  es el valor medio de la cantidad que entra y  $a_t$  es la innovación, una variable aleatoria de media cero y varianza constante que hace que la entrada varíe de unos períodos a otros. Si cada mes se gasta una proporción fija de las existencias iniciales,  $(1-\phi)z_{t-1}$ , y se mantiene la proporción  $\phi z_{t-1}$ , la cantidad de agua en el embalse a final de mes seguirá el proceso (4.1).

La condición  $-1 < \phi < 1$  es necesaria para que el proceso sea estacionario. Para comprobarlo, supongamos que el proceso comienza con  $z_0 = h$ , siendo  $h$  un valor cualquiera fijo. El valor siguiente será  $z_1 = c + \phi h + a_1$ , el siguiente,  $z_2 = c + \phi z_1 + a_2 = c + \phi(c + \phi h + a_1) + a_2$  y, sustituyendo sucesivamente, podemos escribir:

$$\begin{aligned} z_1 &= c + \phi h + a_1 \\ z_2 &= c(1 + \phi) + \phi^2 h + \phi a_1 + a_2 \\ z_3 &= c(1 + \phi + \phi^2) + \phi^3 h + \phi^2 a_1 + \phi a_2 + a_3 \\ &\vdots & \vdots \\ z_t &= c \sum_{i=0}^{t-1} \phi^i + \phi^t h + \sum_{i=0}^{t-1} \phi^i a_{t-i}. \end{aligned}$$

Si calculamos la esperanza de  $z_t$ , como  $E[a_t] = 0$ ,

$$E[z_t] = c \sum_{i=0}^{t-1} \phi^i + \phi^t h.$$

Para que el proceso sea estacionario es necesario que esta función no dependa de  $t$ . La media será constante si ambos sumandos lo son, lo que

#### 4. Procesos autorregresivos

---

requiere que al aumentar  $t$  el primer término converja a una constante y el segundo se anule. Ambas condiciones se verifican si  $|\phi| < 1$ , ya que entonces  $\sum_{i=0}^{t-1} \phi^i$  es la suma de una progresión geométrica indefinida de razón  $\phi$  y converge a  $c/(1 - \phi)$ , y el término  $\phi^t$  converge a cero, con lo que la suma lo hace a la constante  $c/(1 - \phi)$ , que será la media del proceso. Con esta condición, después de un periodo transitorio inicial, cuando  $t \rightarrow \infty$ , todas las variables  $z_t$  tendrán la misma esperanza,  $\mu = c/(1 - \phi)$ , independiente de las condiciones iniciales.

Observemos también que en este proceso la innovación  $a_t$  está incorrelada con los valores previos del proceso,  $z_{t-k}$  para  $k$  positivo. En efecto, el valor  $z_{t-k}$  depende de los valores de las innovaciones hasta ese instante,  $a_1, \dots, a_{t-k}$ , pero no de sus valores futuros. Como la innovación es un proceso de ruido blanco, sus valores futuros están incorrelados con los pasados y, por tanto, con los valores previos de proceso,  $z_{t-k}$ .

El proceso AR(1) puede escribirse utilizando la notación del operador de retardo,  $B$ , definido por:

$$Bz_t = z_{t-1}, \quad (4.2)$$

cuyas propiedades se estudian con detalle en el apéndice 4.1, como:

$$(1 - \phi B)\tilde{z}_t = a_t, \quad (4.3)$$

ya que  $B\tilde{z}_t = \tilde{z}_{t-1}$ . Es decir, una serie sigue un proceso AR(1) si al aplicarle el operador  $(1 - \phi B)$  se obtiene un proceso de ruido blanco. El operador  $(1 - \phi B)$  puede interpretarse como un filtro que aplicado a la serie la convierte en una serie sin información, el proceso de ruido blanco. Si consideramos el operador como una ecuación en  $B$ , el coeficiente  $\phi$  se denomina el factor de la ecuación. La condición de estacionaridad es que este factor sea menor que la unidad en valor absoluto. Alternativamente, podemos hablar de la raíz de la ecuación del operador, que se obtiene igualando el operador a cero y resolviendo la ecuación con  $B$  como incógnita:

$$1 - \phi B = 0,$$

con el resultado  $B = 1/\phi$ . La condición de estacionaridad es entonces que la raíz del operador sea, en valor absoluto, mayor que uno.

##### 4.2.1. Esperanza y varianza

Tomando esperanzas en (4.1) supuesto  $|\phi| < 1$ , de manera que  $E[z_t] = E[z_{t-1}] = \mu$ , se obtiene que:

$$\mu = c + \phi\mu$$

$$\mu = \frac{c}{1 - \phi}, \quad (4.4)$$

sustituyendo en (4.1)  $c$  por  $\mu(1 - \phi)$ , el proceso puede escribirse en desviaciones a su media:

$$z_t - \mu = \phi(z_{t-1} - \mu) + a_t,$$

y llamando  $\tilde{z}_t = z_t - \mu$ ,

$$\tilde{z}_t = \phi\tilde{z}_{t-1} + a_t, \quad (4.5)$$

que es la expresión más utilizada del AR(1).

La varianza del proceso se obtiene elevando al cuadrado la expresión (4.5) y tomando esperanzas, con lo que tenemos:

$$E(\tilde{z}_t^2) = \phi^2 E(\tilde{z}_{t-1}^2) + 2\phi E(\tilde{z}_{t-1} a_t) + E(a_t^2).$$

Llamaremos  $\sigma_z^2$  a la varianza del proceso estacionario. El segundo término de esta expresión es cero, ya que como  $\tilde{z}_{t-1}$  y  $a_t$  son independientes la esperanza de su producto es el producto de las esperanzas, y ambas variables tienen esperanza nula. El tercero es la varianza de la innovación,  $\sigma^2$ , y concluimos que:

$$\sigma_z^2 = \phi^2 \sigma_z^2 + \sigma^2,$$

de donde obtenemos que la varianza del proceso es:

$$\sigma_z^2 = \frac{\sigma^2}{1 - \phi^2}. \quad (4.6)$$

Observemos que en esta ecuación aparece de nuevo la condición  $|\phi| < 1$  para que  $\sigma_z^2$  sea finita y positiva. Para un valor fijo de  $\sigma^2$ , la varianza es tanto mayor cuanto mayor sea  $\phi^2$ .

Es importante diferenciar la distribución marginal de cada variable de la distribución condicionada de una variable conocido el valor anterior. La distribución marginal de cada observación es la misma, ya que el proceso es estacionario: tiene media  $\mu$  y varianza  $\sigma_z^2$ . Sin embargo, la distribución condicional de  $z_t$  si conocemos el valor anterior,  $z_{t-1}$ , tiene media condicionada:

$$E(z_t | z_{t-1}) = c + \phi z_{t-1}$$

y varianza  $\sigma^2$ , que según (4.6), es siempre menor que  $\sigma_z^2$ . Si conocemos  $z_{t-1}$  reducimos la incertidumbre en la estimación de  $z_t$ , y esta reducción es tanto mayor cuanto mayor sea  $\phi^2$ . Si el parámetro AR está próximo a uno, la reducción de varianza obtenida por el conocimiento de  $z_{t-1}$  puede ser muy importante.

#### 4.2.2. Función de autocovarianzas y autocorrelación

La información sobre la dependencia lineal entre las variables de una serie temporal generada por un proceso estacionario se encuentra, como hemos visto en el capítulo anterior, en la función de autocovarianzas. Utilizando (4.5), multiplicando por  $z_{t-k}$  y tomando esperanzas se obtiene  $\gamma_k$ , la covarianza entre observaciones separadas por  $k$  periodos, o autocovarianza de orden  $k$ :

$$\gamma_k = E[(z_{t-k} - \mu)(z_t - \mu)] = E[\tilde{z}_{t-k}(\phi\tilde{z}_{t-1} + a_t)]$$

y como  $E[\tilde{z}_{t-k}a_t] = 0$ , ya que las innovaciones están incorreladas con los valores pasados de la serie:

$$\gamma_k = \phi\gamma_{k-1} \quad k = 1, 2, \dots \quad (4.7)$$

donde  $\gamma_0 = \sigma_z^2$ . Esta ecuación muestra que, como  $|\phi| < 1$ , la dependencia entre observaciones se amortigua al aumentar el retardo. En particular, utilizando (4.6):

$$\gamma_1 = \frac{\phi\sigma^2}{1 - \phi^2}. \quad (4.8)$$

##### 4.2.2.1. Función de autocorrelación simple (fas)

Las autocorrelaciones contienen la misma información que las autocovarianzas, con la ventaja de no depender de las unidades de medida. En adelante llamaremos a la función de autocorrelación del proceso la función de autocorrelación simple (fas), para diferenciarla de otras funciones ligadas a la autocorrelación que se definen al final de este capítulo. Sea  $\rho_k$  la correlación entre observaciones separadas por  $k$  periodos, o autocorrelación de orden  $k$ , definida por :

$$\rho_k = \gamma_k/\gamma_0,$$

utilizando (4.7), se obtiene que:

$$\rho_k = \phi\gamma_{k-1}/\gamma_0 = \phi\rho_{k-1}.$$

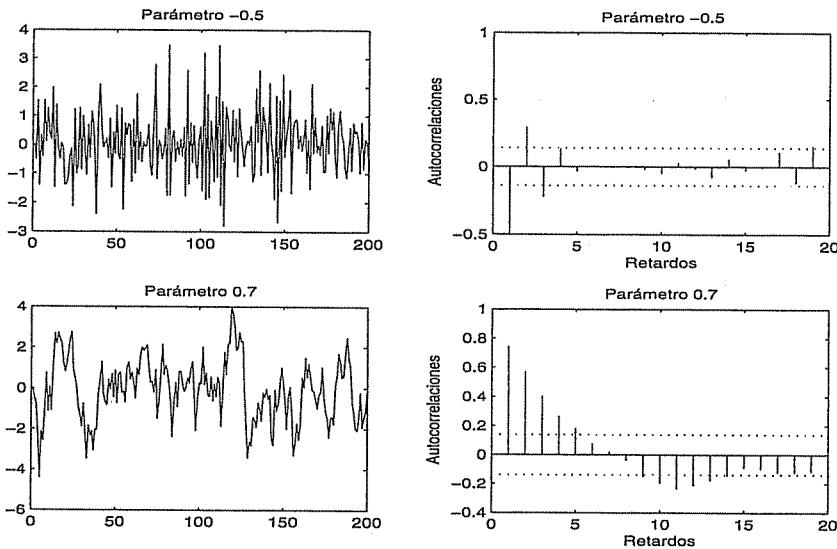
Como, según (4.6) y (4.8),  $\rho_1 = \phi$ , concluimos que:

$$\rho_k = \phi^k, \quad (4.9)$$

y cuando  $k$  es grande,  $\rho_k$  tiende a cero con rapidez que depende de  $\phi$ .

La expresión (4.9) muestra que la función de autocorrelación simple de un proceso AR(1) es igual a las potencias del parámetro AR del proceso y decrece geométricamente hacia cero. Si el parámetro es positivo, la

**Figura 4.1** Dos realizaciones del proceso AR(1) y sus correspondientes funciones de autocorrelación



dependencia lineal del presente de los valores pasados es siempre positiva, mientras que si el parámetro es negativo esta dependencia es positiva para los retardos pares y negativa para los impares. Como ilustración, la figura 4.1 presenta dos realizaciones de un proceso AR(1) con distintos valores del parámetro, el primero negativo y el segundo positivo, y sus funciones de autocorrelación teóricas. Cuando el parámetro es positivo el valor en  $t$  es parecido al valor en  $t - 1$ , por la dependencia positiva, con lo que el gráfico de la serie evoluciona suavemente. Por el contrario, cuando el parámetro es negativo el valor en  $t$  es, en general, de signo opuesto al de  $t - 1$ , con lo que el gráfico de la serie muestra muchos cambios de signo.

#### 4.2.3. Representación del proceso AR(1) como suma de innovaciones

El proceso AR(1) puede expresarse en función de los valores pasados de las innovaciones. Esta representación es útil porque revela algunas propiedades del proceso y, como veremos más adelante, sugiere cómo introducir otros procesos estacionarios importantes que estudiaremos en el capítulo siguiente. Sustituyendo en la expresión (4.5) del proceso  $\tilde{z}_{t-1}$  en función de  $\tilde{z}_{t-2}$ , tenemos:

$$\tilde{z}_t = \phi(\phi\tilde{z}_{t-2} + a_{t-1}) + a_t = a_t + \phi a_{t-1} + \phi^2 \tilde{z}_{t-2}.$$

#### 4. Procesos autorregresivos

---

Si sustituimos ahora  $\tilde{z}_{t-2}$  por su expresión como función de  $\tilde{z}_{t-3}$ , se obtiene:

$$\tilde{z}_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 \tilde{z}_{t-2},$$

y aplicando repetidamente esta sustitución, resulta que:

$$\tilde{z}_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots + \phi^{t-1} a_1 + \phi^t \tilde{z}_1.$$

Si suponemos que  $t$  es grande, como  $\phi^t$  será muy próximo a cero podemos representar la serie como función de todas las innovaciones, con pesos que decrecen geométricamente. Suponiendo que la serie comienza en el infinito pasado tenemos:

$$\tilde{z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j},$$

y esta representación se denomina la forma de media móvil de orden infinito,  $MA(\infty)$ , del proceso. Observemos que los coeficientes de las innovaciones son precisamente los coeficientes de la función de autocorrelación simple. La expresión  $MA(\infty)$  puede obtenerse también directamente multiplicando la ecuación (4.3) por el operador  $(1 - \phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \dots$ , con lo que se obtiene:

$$\tilde{z}_t = (1 - \phi B)^{-1} a_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots$$

---

#### Ejemplo 4.1

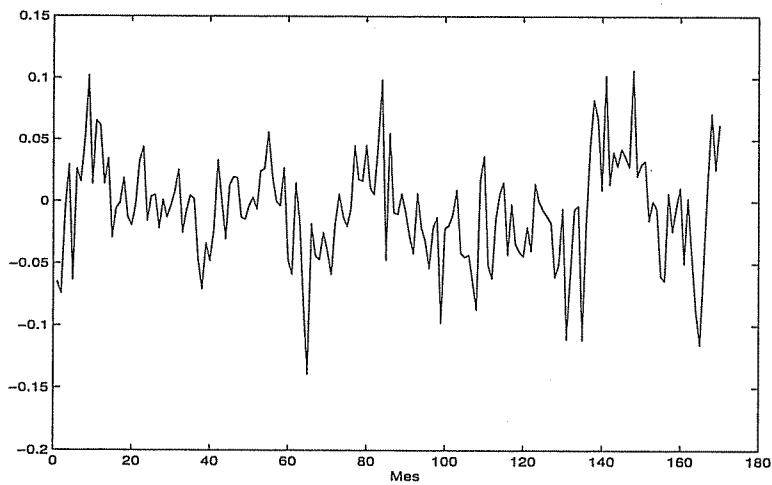
La figura 4.2 muestra la serie mensual de los cambios relativos del tipo de interés anual, definidos por  $z_t = \log(y_t/y_{t-1}) \approx (y_t - y_{t-1})/y_t$ , donde  $y_t$  es la serie de tipos de interés anual del fichero tipos88.dat. La figura 4.3 muestra la función de autocorrelación y se observa que los coeficientes de autocorrelación se amortiguan con el retardo: el primero es del orden de .4, el segundo próximo a  $.4^2 = .16$ , el tercero de un valor similar y los siguientes más pequeños y no significativos. Dado que hay variabilidad muestral, concluimos que esta estructura coincide globalmente con lo esperable para un proceso AR(1) con coeficiente positivo.

---

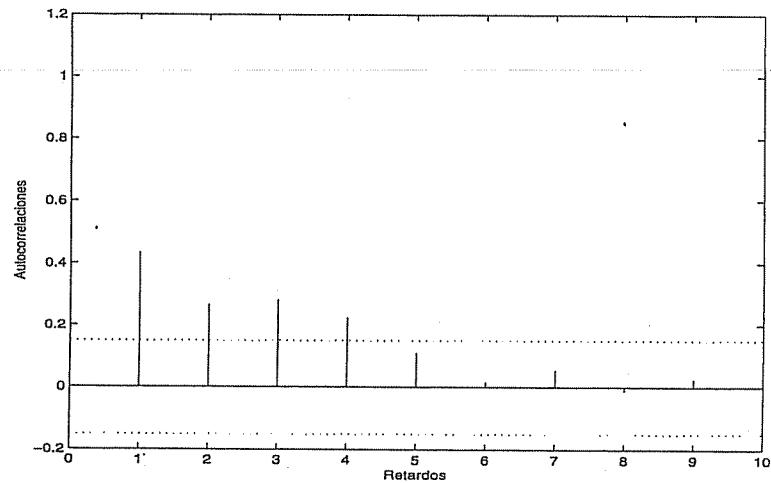
#### 4.3. El proceso AR(2)

La dependencia entre los valores presentes y los pasados que establece un proceso AR(1) puede generalizarse permitiendo que  $z_t$  dependa linealmente no sólo de  $z_{t-1}$ , sino también de  $z_{t-2}$ . Se obtiene entonces el proceso autorregresivo de segundo orden, o proceso AR(2):

**Figura 4.2 Serie mensual de cambios relativos en el tipo de interés a un año en España, 1988 a 2002**



**Figura 4.3 Función de autocorrelación de los cambios en los tipos de interés**



$$z_t = c + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t, \quad (4.10)$$

donde ahora  $c, \phi_1$  y  $\phi_2$  son constantes a determinar y  $a_t$  un proceso de ruido blanco con varianza  $\sigma^2$ .

Vamos a obtener qué condiciones tienen que verificar los parámetros

#### 4. Procesos autorregresivos

---

del proceso para que sea estacionario. Tomando esperanzas en (4.10) e imponiendo que la media sea constante, resulta:

$$\mu = c + \phi_1\mu + \phi_2\mu,$$

que implica

$$\mu = \frac{c}{1 - \phi_1 - \phi_2}, \quad (4.11)$$

y la condición para que el proceso tenga media finita es que  $1 - \phi_1 - \phi_2 \neq 0$ . Sustituyendo  $c$  por  $\mu(1 - \phi_1 - \phi_2)$  y llamando  $\tilde{z}_t = z_t - \mu$  al proceso en desviaciones a su media, el proceso AR(2) es:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t. \quad (4.12)$$

Para estudiar las propiedades del proceso es conveniente utilizar la notación de operadores del apéndice 4.1. Introduciendo el operador de retraso,  $B$ , la ecuación de este proceso es:

$$(1 - \phi_1 B - \phi_2 B^2) \tilde{z}_t = a_t.$$

El operador  $(1 - \phi_1 B - \phi_2 B^2)$  puede siempre expresarse como  $(1 - G_1 B)(1 - G_2 B)$ , donde  $G_1^{-1}$  y  $G_2^{-1}$  son las raíces de la ecuación del operador considerando  $B$  como variable y resolviendo  $1 - \phi_1 B - \phi_2 B^2 = 0$  (véase el apéndice 4.2). Esta ecuación se denomina *la ecuación característica* del operador. También se dice que  $G_1$  y  $G_2$  son los factores del polinomio característico del proceso. Estas raíces pueden ser reales o complejas conjugadas. Se demuestra en el apéndice 4.2 que la condición de estacionaridad es que  $|G_i| < 1$ ,  $i = 1, 2$ . Esta condición es análoga a la estudiada para el AR(1) y es coherente con la condición encontrada para que la media sea finita: si la ecuación  $1 - \phi_1 B - \phi_2 B^2 = 0$  tiene una raíz unidad, se verifica que  $1 - \phi_1 - \phi_2 = 0$ , y el proceso no es estacionario, ya que no tiene media finita.

##### 4.3.1. Función de autocovarianzas

Elevando al cuadrado la expresión (4.12) y tomando esperanzas, obtenemos que su varianza debe satisfacer:

$$\gamma_0 = \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + 2\phi_1 \phi_2 \gamma_1 + \sigma^2. \quad (4.13)$$

Para calcular las autocovarianzas, multiplicando la ecuación (4.12) por  $\tilde{z}_{t-k}$  y tomando esperanzas, resulta la expresión general de las autocovarianzas:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}. \quad k \geq 1 \quad (4.14)$$

Particularizando esta ecuación para  $k = 1$ , como en un proceso estacionario  $\gamma_{-1} = \gamma_1$ , se obtiene que:

$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1,$$

que proporciona  $\gamma_1 = \phi_1\gamma_0/(1 - \phi_2)$ . Sustituyendo esta expresión en (4.13), resulta la fórmula para la varianza:

$$\sigma_z^2 = \gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}. \quad (4.15)$$

Para que el proceso sea estacionario esta varianza debe ser positiva, lo que ocurrirá si el numerador y el denominador tienen el mismo signo. Puede comprobarse que los valores de los parámetros que hacen que el AR(2) sea un proceso estacionario (y por tanto con varianza siempre positiva) son los incluidos en la región:

$$-1 < \phi_2 < 1 \quad \phi_1 + \phi_2 < 1 \quad \phi_2 - \phi_1 < 1.$$

En este proceso es de nuevo importante diferenciar las propiedades marginales de las condicionadas. Supuesto que se verifican las condiciones de estacionariedad, la media marginal es  $\mu$ , dada por (4.11) y la varianza marginal es  $\sigma_z^2$ , dada por (4.15). Sin embargo, la media condicional de  $z_t$  dados los dos valores previos es:

$$E(z_t|z_{t-1}, z_{t-2}) = c + \phi_1 z_{t-1} + \phi_2 z_{t-2},$$

y la varianza condicionada es  $\sigma^2$ , la de las innovaciones, que será siempre menor que la varianza marginal del proceso,  $\sigma_z^2$ , según la ecuación (4.15).

#### 4.3.2. Función de autocorrelación simple

Dividiendo por la varianza en la ecuación (4.14), se obtiene la relación entre los coeficientes de autocorrelación:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \quad k \geq 1 \quad (4.16)$$

particularizando (4.16) para  $k = 1$ , como en un proceso estacionario  $\rho_1 = \rho_{-1}$ , obtenemos que:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad (4.17)$$

y particularizando (4.16) para  $k = 2$  y utilizando (4.17):

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \quad (4.18)$$

#### 4. Procesos autorregresivos

---

Para  $k \geq 3$  los coeficientes de autocorrelación pueden obtenerse recursivamente a partir de la ecuación en diferencias (4.16). Se demuestra en el apéndice 4.1 que la solución general de esta ecuación es:

$$\rho_k = A_1 G_1^k + A_2 G_2^k, \quad (4.19)$$

donde  $G_1$  y  $G_2$  son los factores del polinomio característico del proceso y  $A_1$  y  $A_2$  constantes a determinar a partir de las condiciones iniciales  $\rho_0 = 1$  (que implica  $A_1 + A_2 = 1$ ) y  $\rho_1 = \phi_1 / (1 - \phi_2)$ . Según (4.19) los coeficientes  $\rho_k$  serán menores o iguales a la unidad si  $|G_1| < 1$  y  $|G_2| < 1$ , que son las condiciones de estacionaridad del proceso. Si los factores  $G_1$  y  $G_2$  son complejos del tipo  $a \pm bi$ , siendo  $i = \sqrt{-1}$ , entonces esta condición es  $\sqrt{a^2 + b^2} < 1$ . Podemos encontrarnos en los casos siguientes:

1. Los dos factores  $G_1$  y  $G_2$  son reales. El decrecimiento de (4.19) será la suma de dos exponenciales y la forma de la función de autocorrelación dependerá de que  $G_1$  y  $G_2$  tengan signo igual u opuesto.
2. Los dos factores  $G_1$  y  $G_2$  son complejos conjugados. Entonces, se demuestra en el apéndice 4.1 que la función  $\rho_k$  decrecerá de forma sinusoidal.

Los cuatro tipos de funciones de autocorrelación posibles para un AR(2) se presentan gráficamente en la figura 4.4. Los dos de la izquierda corresponden a raíces reales y entonces se observa un decrecimiento geométrico similar al de un AR(1). En el superior, los coeficientes son siempre positivos, mientras que en el inferior las raíces tienen signos opuestos, siendo mayor la negativa, que es la dominante y produce un decrecimiento alterno, similar al de un AR(1). Los dos casos de la derecha corresponden a raíces complejas conjugadas, y los coeficientes decrecen de forma sinusoidal. Dependiendo del signo de la parte real, se obtiene un decrecimiento comenzando en valores positivos, como en la parte superior, o negativos, como en la parte inferior.

##### 4.3.3. Representación del AR(2) como suma de innovaciones

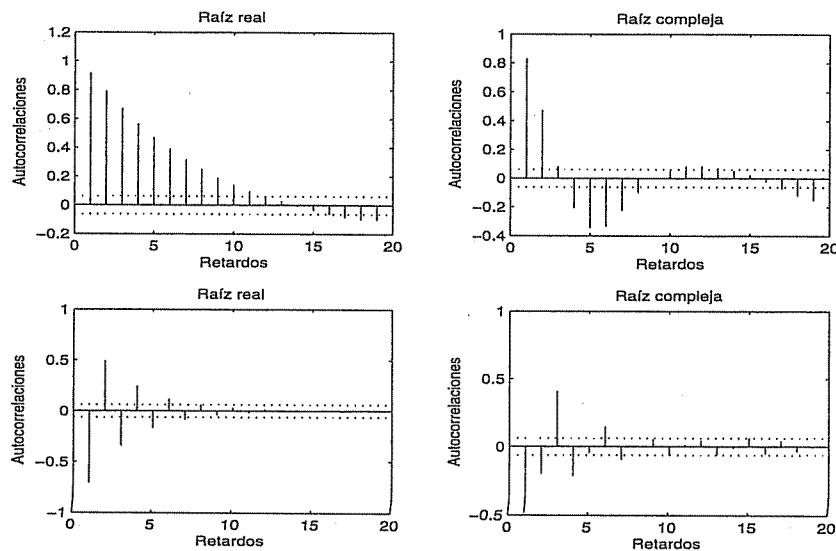
El proceso AR(2) puede representarse, de forma análoga al AR(1), como combinación lineal de las innovaciones pasadas. Escribiendo (4.3) como:

$$(1 - G_1 B)(1 - G_2 B)\tilde{z}_t = a_t$$

e invirtiendo estos operadores, tendremos que:

$$\tilde{z}_t = (1 + G_1 B + G_1^2 B^2 + \dots)(1 + G_2 B + G_2^2 B^2 + \dots)a_t \quad (4.20)$$

**Figura 4.4 Funciones de autocorrelación simple para distintos procesos AR(2)**



que conducirá a la expresión  $\text{MA}(\infty)$  del proceso:

$$\tilde{z}_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \quad (4.21)$$

Podemos obtener los coeficientes  $\psi_i$  como función de las raíces igualando potencias de  $B$  en (4.20) y (4.21). También podemos obtenerlos en función de los coeficientes  $\phi_1$  y  $\phi_2$ . Llamando  $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$  como  $\psi(B) = (1 - \phi_1 B - \phi_2 B^2)^{-1}$ , tendremos que:

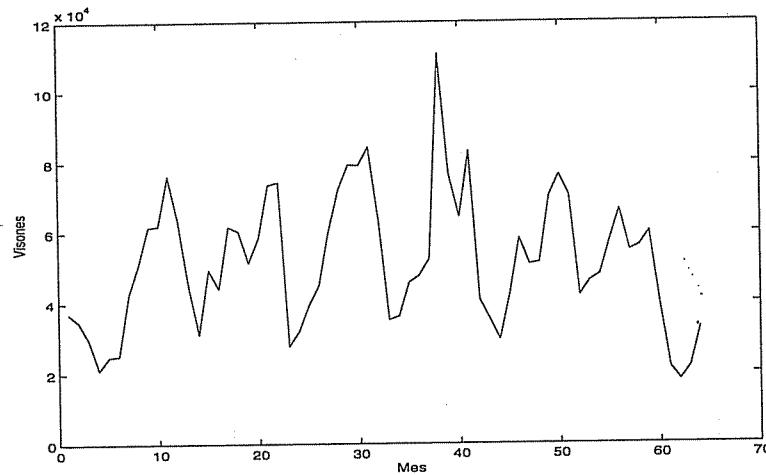
$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1. \quad (4.22)$$

Imponiendo la condición que todos los coeficientes de las potencias de  $B$  en (4.22) sean nulos, el coeficiente de  $B$  en esta ecuación es  $\psi_1 - \phi_1$ , que implica  $\psi_1 = \phi_1$ . El coeficiente de  $B^2$  es  $\psi_2 - \phi_1 \psi_1 - \phi_2$ , que implica la relación:

$$\psi_k = \phi_1 \psi_{k-1} + \phi_2 \psi_{k-2}, \quad (4.23)$$

para  $k = 2$ , ya que  $\psi_0 = 1$ . Los coeficientes de  $B^k$  para  $k \geq 2$  verifican la ecuación (4.23), que es similar a la que deben verificar los coeficientes de autocorrelación. Concluimos que la forma de los coeficientes  $\psi_i$  será similar a la de los coeficientes de autocorrelación, aunque los valores no serán idénticos porque las dos secuencias se inician de forma distinta: para los coeficientes de autocorrelación  $\rho_0 = 1$  y  $\rho_1 = \frac{\phi_1}{1-\phi_2}$ , mientras que para los coeficientes  $\psi_i$ ,  $\psi_0 = 1$  y  $\psi_1 = \phi_1$ .

**Figura 4.5 Serie del número de visones detectados en una zona de Canadá, datos anuales desde 1848 hasta 1911**



#### Ejemplo 4.2

La figura 4.5 presenta el número de visones anuales detectados en una zona de Canadá el siglo pasado. La serie, que se encuentra en el fichero mink.dat, presenta una evolución cíclica que podría explicarse por un AR(2) con raíces negativas. La figura 4.6 muestra la función de autocorrelación, que tiene una estructura sinusoidal, confirmando la posibilidad de un AR(2).

#### Ejemplo 4.3

Escribir la función de autocorrelación simple del proceso AR(2):

$$z_t = 1.2z_{t-1} - 0.32z_{t-2} + a_t.$$

La ecuación característica de dicho proceso es:

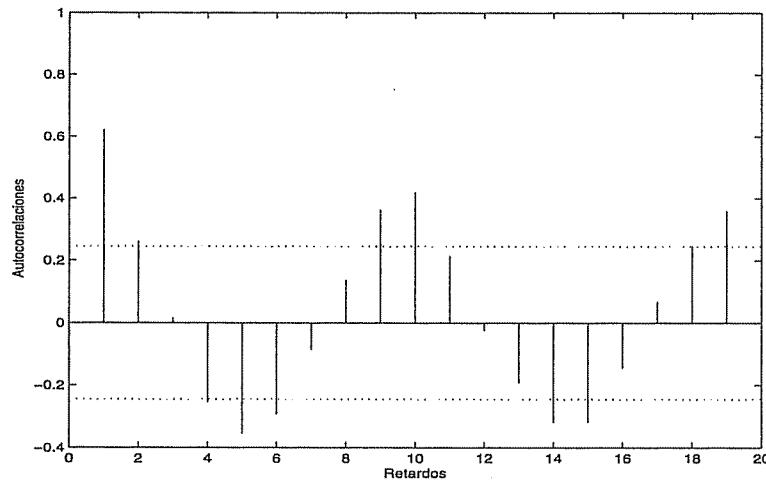
$$0.32X^2 - 1.2X + 1 = 0,$$

cuya solución es:

$$X = \frac{1.2 \pm \sqrt{1.2^2 - 4 \times 0.32}}{0.64} = \frac{1.2 \pm 0.4}{0.64}.$$

Las soluciones son  $G_1^{-1} = 2.5$  y  $G_2^{-1} = 1.25$  y los factores serán  $G_1 = 0.4$  y

**Figura 4.6 Función de autocorrelación del número de visones detectados en una zona de Canadá, datos anuales desde 1848 hasta 1911**



$G_2 = 0.8$ . En efecto, la ecuación característica puede escribirse:

$$0.32X^2 - 1.2X + 1 = (1 - 0.4X)(1 - 0.8X).$$

Por tanto, el proceso es estacionario con raíces reales y los coeficientes de autocorrelación verifican:

$$\rho_k = A_1 0.4^k + A_2 0.8^k.$$

Para determinar  $A_1$  y  $A_2$  imponemos las condiciones iniciales  $\rho_0 = 1$ ,  $\rho_1 = 1.2 / (1.322) = 0.91$ . Entonces, para  $k = 0$ :

$$1 = A_1 + A_2,$$

y para  $k = 1$ ,

$$0.91 = 0.4A_1 + 0.8A_2.$$

Resolviendo estas ecuaciones se obtiene  $A_2 = 0.51/0.4$  y  $A_1 = -0.11/0.4$ . Por tanto, la función de autocorrelación es:

$$\rho_k = -\frac{0.11}{0.4} 0.4^k + \frac{0.51}{0.4} 0.8^k$$

obteniéndose la tabla siguiente:

$k$	0	1	2	3	4	5	6	7	8
$\rho_k$	1	0.91	0.77	0.63	0.51	0.41	0.33	0.27	0.21

La representación en función de las innovaciones, escribiendo:

$$(1 - 0.4B)(1 - 0.8B)z_t = a_t,$$

e invirtiendo ambos operadores:

$$z_t = (1 + 0.4B + .16B^2 + .06B^3 + \dots)(1 + 0.8B + .64B^2 + \dots)a_t$$

resulta:

$$z_t = (1 + 1.2B + 1.12B^2 + \dots)a_t.$$


---

#### 4.4. El proceso autorregresivo general (AR(p))

Diremos que una serie temporal  $z_t$  estacionaria sigue un proceso autorregresivo de orden  $p$  si:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t, \quad (4.24)$$

donde  $\tilde{z}_t = z_t - \mu$ , siendo  $\mu$  la media del proceso estacionario  $z_t$  y  $a_t$  un proceso de ruido blanco. Utilizando la notación de operadores, la ecuación de un AR(p) es:

$$(1 - \phi_1 B - \dots - \phi_p B^p) \tilde{z}_t = a_t, \quad (4.25)$$

y llamando  $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  al polinomio de grado  $p$  en el operador de retardo, cuyo primer término es la unidad, tenemos:

$$\phi_p(B) \tilde{z}_t = a_t \quad (4.26)$$

que es la expresión general de un proceso autorregresivo.

Llamaremos *ecuación característica* del proceso a la ecuación:

$$\phi_p(B) = 0 \quad (4.27)$$

considerada como función de  $B$ . Esta ecuación tendrá  $p$  raíces  $G_1^{-1}, \dots, G_p^{-1}$ , en general distintas (véase apéndice 4.1), y podemos escribir:

$$\phi_p(B) = \prod_{i=1}^p (1 - G_i B)$$

de manera que los coeficientes  $G_i$  son los factores de la ecuación característica. Se demuestra que el proceso es estacionario si  $|G_i| < 1$ , para todo  $i$ .

#### 4.4.1. Función de autocorrelación simple

Multiplicando (4.24) por  $\tilde{z}_{t-k}$  ( $k > 0$ ), tomando esperanzas en ambos miembros y dividiendo el resultado por  $\gamma_0$ , se obtiene que los coeficientes de autocorrelación de un AR(p) verifican la siguiente ecuación en diferencias:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, \quad k > 0.$$

En las secciones anteriores hemos visto casos particulares de esta ecuación para  $p = 1$  y  $p = 2$ . Podemos concluir que los coeficientes de autocorrelación satisfacen la misma ecuación que el proceso:

$$\phi_p(B) \rho_k = 0 \quad k > 0. \quad (4.28)$$

La solución general de esta ecuación es (apéndice 4.1):

$$\rho_k = \sum_{i=1}^p A_i G_i^k, \quad (4.29)$$

donde las  $A_i$  son constantes a determinar a partir de las condiciones iniciales y los  $G_i$  son los factores de la ecuación característica. Para que el proceso sea estacionario el módulo de  $G_i$  debe ser menor que uno o, lo que es lo mismo, las raíces de la ecuación característica (4.27) deben ser mayores que uno en módulo. Para comprobarlo observemos que la condición  $|\rho_k| < 1$  exige que no exista en (4.29) ningún  $G_i$  mayor que la unidad, ya que entonces, cuando  $k$  aumenta, el término  $G_i^k$  crecería sin límite. Observemos, además, que para que el proceso sea estacionario no puede existir una raíz  $G_i$  igual a la unidad, ya que entonces su componente  $G_i^k$  no decrecería y los coeficientes  $\rho_k$  no tenderían hacia cero para ningún retardo.

La ecuación (4.29) muestra que la función de autocorrelación simple de un proceso AR(p) es una mezcla de exponentiales, debidas a los términos con raíces reales, y sinusoidales, debidas a las raíces complejas conjugadas. Su estructura puede ser, en consecuencia, muy compleja.

#### 4.4.2. Ecuaciones de Yule-Walker

Particularizando la ecuación (4.28) para  $k = 1, \dots, p$ , se obtiene un sistema de  $p$  ecuaciones que relacionan las  $p$  primeras autocorrelaciones con los parámetros del proceso. Llamaremos ecuaciones de Yule-Walker al sistema:

4. Procesos autorregresivos

---

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

Definiendo:

$$\phi' = [\phi_1, \dots, \phi_p], \quad \varrho' = [\rho_1, \dots, \rho_p], \quad R = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix},$$

el sistema anterior se escribe matricialmente:

$$\varrho = R\phi \quad (4.30)$$

y los parámetros se determinan a partir de las autocorrelaciones mediante:

$$\phi = R^{-1}\varrho.$$


---

#### Ejemplo 4.4

Obtener los parámetros de un proceso AR(3) cuyas primeras autocorrelaciones son  $\rho_1 = 0.9$ ;  $\rho_2 = 0.8$ ;  $\rho_3 = 0.5$ . ¿Es estacionario el proceso?

El sistema de ecuaciones de Yule-Walker es:

$$\begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 & 0.8 \\ 0.9 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix},$$

cuya solución es:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} 5.28 & -5 & 0.28 \\ -5 & 10 & -5 \\ 0.28 & -5 & 5.28 \end{bmatrix} \begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.89 \\ 1 \\ -1.11 \end{bmatrix}.$$

En consecuencia, el proceso AR(3) con estas correlaciones es:

$$(1 - 0.89B - B^2 + 1.11B^3) z_t = a_t.$$

Para comprobar que el proceso es estacionario tendríamos que calcular los factores de la ecuación característica. La forma más rápida es obtener las soluciones de la ecuación

$$X^3 - 0.89X^2 - X + 1.11 = 0$$

y comprobar que todas tienen módulo menor que la unidad. Las raíces de esta ecuación son  $-1.7930$ ,  $0.4515 + 0.6444i$  y  $0.4515 - 0.6444i$ . Los módulos de las raíces complejas son menores que la unidad, pero el factor real es mayor que la unidad, con lo que concluimos que no existe un proceso estacionario AR(3) que tenga estos tres coeficientes de autocorrelación.

---

#### 4.4.3. Representación del AR( $p$ ) suma de innovaciones

La forma del proceso como suma de innovaciones, o forma MA( $\infty$ ), se obtiene invirtiendo el operador AR( $p$ ). Para obtener los coeficientes de esta representación, escribiremos  $\psi(B) = \phi(B)^{-1}$ , que implica:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1,$$

y los coeficientes  $\psi_i$  se obtienen igualando potencias de  $B$  a cero. Se comprueba que deben verificar la relación:

$$\psi_k = \phi_1 \psi_{k-1} + \dots + \phi_p \psi_{k-1}$$

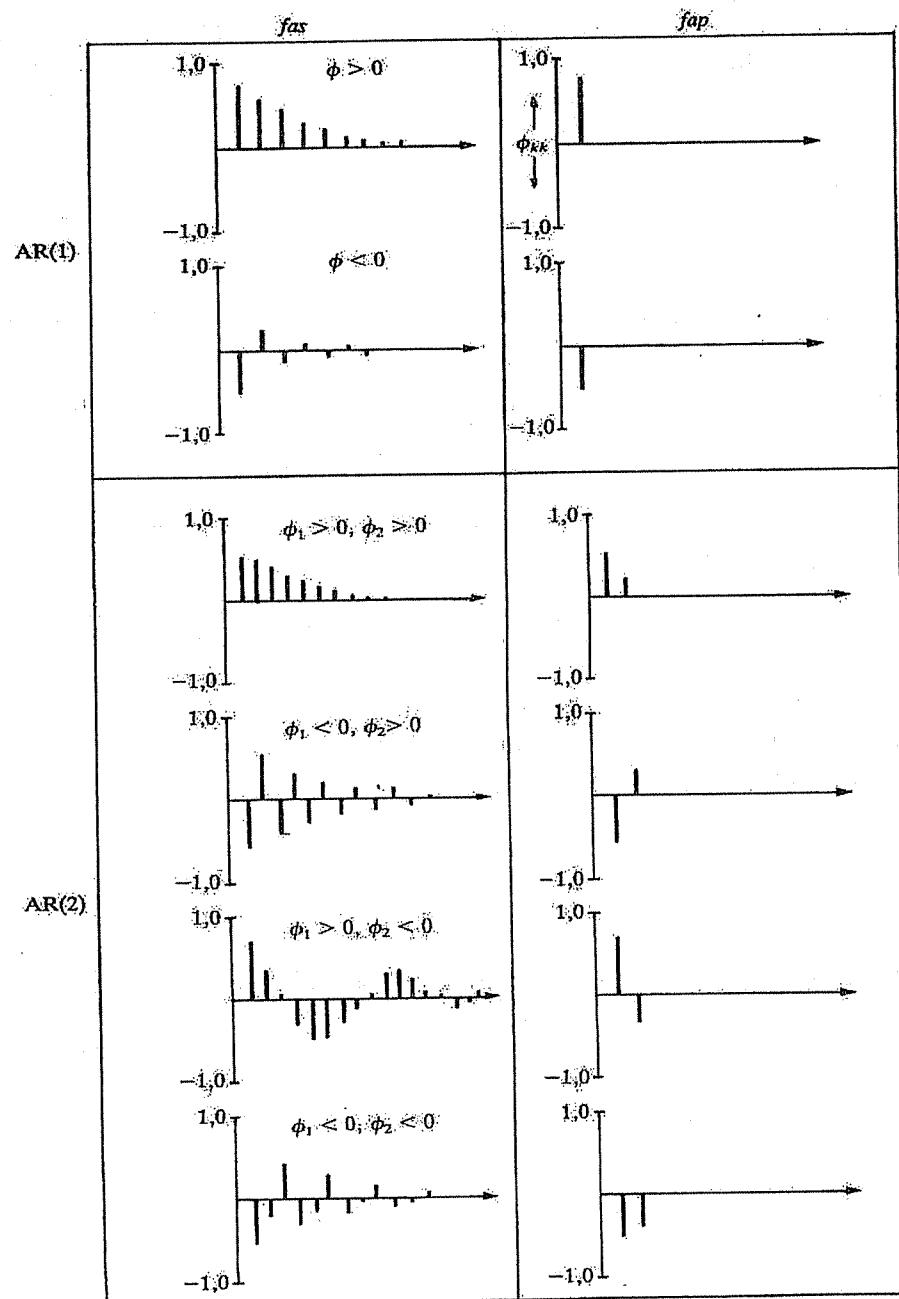
que es análoga a la que verifican los coeficientes de autocorrelación del proceso. Como hemos comentado, los coeficientes de autocorrelación,  $\rho_k$ , y los coeficientes de la estructura MA( $\infty$ ) no son idénticos: aunque ambas secuencias satisfacen la misma ecuación en diferencias y tienen la forma  $\sum A_i G_i^k$ , las constantes  $A_i$  dependen de las condiciones iniciales y serán distintas en ambas secuencias.

### 4.5. La función de autocorrelación parcial

Determinar el orden de un proceso autorregresivo a partir de su función de autocorrelación simple es difícil. Hemos visto que esta función es una mezcla de decrecimientos exponenciales y sinusoidales, que se amortiguan al avanzar el retardo, y no presenta rasgos fácilmente identificables para determinar el orden del proceso. Para resolver este problema se introduce la función de autocorrelación parcial.

Si comparamos un AR(1) con un AR(2) vemos que, aunque en ambos procesos cada observación está relacionada con las anteriores, el tipo de relación entre observaciones separadas por más de un retardo es distinto en ambos procesos. En el AR(1) el efecto de  $z_{t-2}$  sobre  $z_t$  es siempre a través de  $z_{t-1}$ , y dado  $z_{t-1}$ , el valor de  $z_{t-2}$  es irrelevante para prever  $z_t$ . Sin embargo, en un AR(2) además del efecto de  $z_{t-2}$  que se transmite a  $z_t$  a través de  $z_{t-1}$ , existe un efecto directo de  $z_{t-2}$  sobre  $z_t$ . En general, un AR( $p$ ) presenta efectos *directos* de observaciones separadas por  $1, 2, \dots, p$

Figura 4.7 Funciones de autocorrelación simple y parcial para procesos AR



retardos y los efectos *directos* de las observaciones separadas por más de  $p$  retardos son nulos. Esta idea es la clave para la utilización de la función de autocorrelación parcial.

Se define el *coeficiente de autocorrelación parcial* de orden  $k$ ,  $\rho_k^p$ , como el coeficiente de correlación entre observaciones separadas  $k$  periodos, cuando eliminamos de la relación entre las dos variables la dependencia lineal debida a los valores intermedios. Este coeficiente podría calcularse como sigue:

1. Se elimina de  $\tilde{z}_t$ , el efecto de  $\tilde{z}_{t-1}, \dots, \tilde{z}_{t-k+1}$  mediante la regresión:

$$\tilde{z}_t = \beta_1 \tilde{z}_{t-1} + \dots + \beta_{k-1} \tilde{z}_{t-k+1} + u_t,$$

donde la variable  $u_t$  recoge la parte de  $\tilde{z}_t$  no común con  $\tilde{z}_{t-1}, \dots, \tilde{z}_{t-k+1}$ .

2. Se elimina de  $\tilde{z}_{t-k}$  el efecto de  $\tilde{z}_{t-1}, \dots, \tilde{z}_{t-k+1}$ , mediante la regresión:

$$\tilde{z}_{t-k} = \gamma_1 \tilde{z}_{t-1} + \dots + \gamma_{k-1} \tilde{z}_{t-k+1} + v_t,$$

donde, de nuevo,  $v_t$  contiene la parte de  $z_{t-1}$  no común con las observaciones intermedias.

3. Se calcula el coeficiente de correlación simple entre  $u_t$  y  $v_t$  que, por definición, es el coeficiente de autocorrelación parcial de orden  $k$ .

Esta definición es análoga a la del coeficiente de correlación parcial en regresión. Puede comprobarse que las tres etapas anteriores equivalen a ajustar la regresión múltiple:

$$\tilde{z}_t = \alpha_{k1} \tilde{z}_{t-1} + \dots + \alpha_{kk} \tilde{z}_{t-k} + \eta_t$$

y entonces:

$$\rho_k^p = \alpha_{kk}.$$

Es decir, el coeficiente de autocorrelación parcial de orden  $k$  es el coeficiente  $\alpha_{kk}$  de la variable  $z_{t-k}$  al ajustar a los datos de la serie un  $AR(k)$ . Llamaremos *función de autocorrelación parcial (fap)* a la representación de los coeficientes de autocorrelación parcial en función del retardo. Por tanto, si ajustamos la familia de regresiones:

$$\begin{aligned}\tilde{z}_t &= \alpha_{11} \tilde{z}_{t-1} + \eta_{1t} \\ \tilde{z}_t &= \alpha_{21} \tilde{z}_{t-1} + \alpha_{22} \tilde{z}_{t-2} + \eta_{2t} \\ &\vdots \quad \vdots \quad \vdots \\ \tilde{z}_t &= \alpha_{k1} \tilde{z}_{t-1} + \dots + \alpha_{kk} \tilde{z}_{t-k} + \eta_{kt}\end{aligned}$$

la secuencia de coeficientes  $\alpha_{ii}$  proporciona la función de autocorrelación parcial.

De esta definición se deduce que un proceso AR( $p$ ) tendrá los  $p$  primeros coeficientes de autocorrelación parcial distintos de cero, y, por tanto, en la *fap* el número de coeficientes distintos de cero indica el orden del proceso AR. Esta propiedad va a ser clave para identificar el orden de un proceso autorregresivo. La figura 4.7 resume la *fas* y *fap* de distintos procesos AR. Además, el coeficiente de correlación parcial de orden  $p$  coincidirá siempre con el parámetro  $\phi_p$ . El apéndice 4.2 presenta métodos eficientes para estimar los coeficientes de correlación parcial.

---

#### Ejemplo 4.5

La figura 4.8 presenta la función de autocorrelación parcial para la serie de los tipos de interés del ejemplo 4.1. Concluimos que las variaciones del tipo de interés sí parecen seguir un proceso AR(1), ya que sólo hay un coeficiente significativo.

#### Ejemplo 4.6

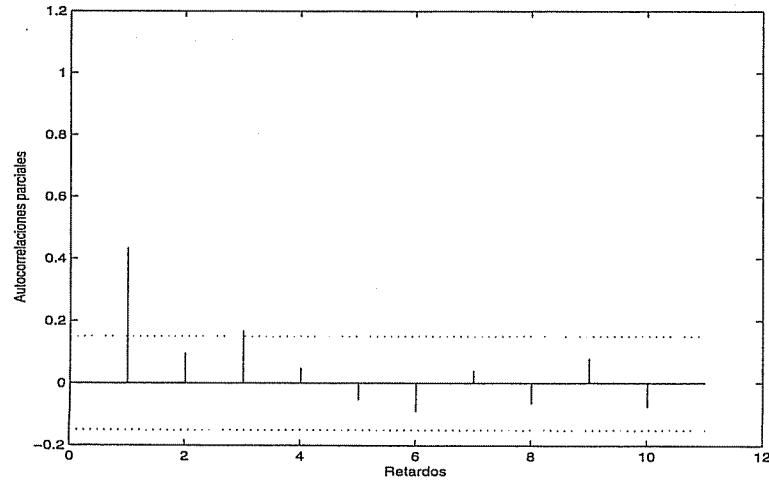
La figura 4.6 presenta la función de autocorrelación parcial para los datos de los visones del ejemplo 4.2. Esta serie presenta coeficientes de autocorrelación parcial significativos hasta el retardo cuarto, sugiriendo que el modelo es un AR(4).

---

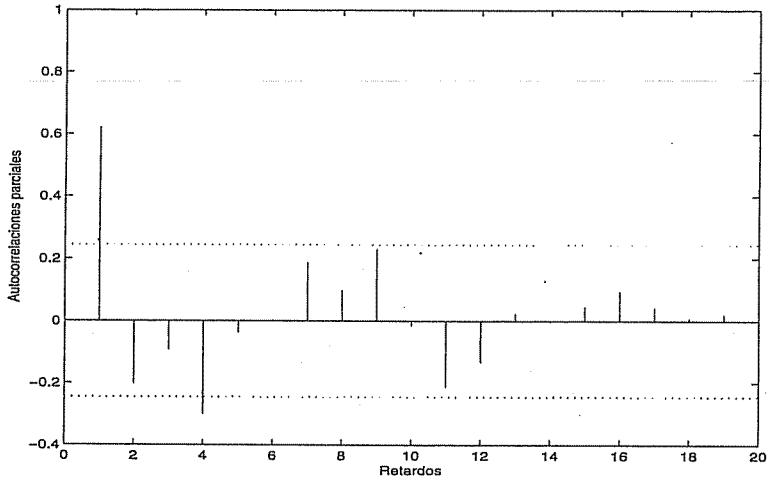
### Ejercicios 4

- 4.1 Generar muestras de un proceso AR(1) con  $\phi = 0.7$  con un programa de ordenador como sigue: (1) generar un vector  $a_t$  de 150 variables aleatorias normales (0,1); (2) tomar  $z_1 = a_1$ ; (3) para  $t = 2, \dots, 150$  calcular  $z_t = 0.7z_{t-1} + a_t$ ; (4) para evitar el efecto de las condiciones iniciales eliminar los primeros 50 datos y tomar como muestra de tamaño 100 del proceso AR(1) los valores  $z_{51}, \dots, z_{150}$ .
- 4.2 Obtener la función de autocorrelación teórica del proceso  $z_t = 0.7z_{t-1} + a_t$ , donde  $a_t$  es ruido blanco. Comparar los resultados teóricos con los observados en la muestra simulada del ejercicio anterior.
- 4.3 Obtener la función de autocorrelación simple teórica del proceso  $z_t = 0.9z_{t-1} - 0.18z_{t-2} + a_t$ , donde  $a_t$  es ruido blanco. Generar una realización del proceso mediante el ordenador y comparar la función muestral con la teórica.

**Figura 4.8 Función de autocorrelación parcial para las variaciones relativas del tipo de interés**



**Figura 4.9 Función de autocorrelación parcial para los visones en Canadá**



4.4 Expresar en notación de operadores los procesos de los ejercicios 2 y 3. Representar los procesos en la forma  $z_t = \sum \psi_i a_{t-i}$  obteniendo los operadores inversos.

4.5 Escribir la función de autocorrelación teórica del proceso  $(1 - 1.2B + .32B^2)z_t = a_t$ . Obtener la representación de este proceso como  $z_t =$

#### 4. Procesos autorregresivos

$\sum \psi_i a_{t-i}$  y comentar la relación entre los coeficientes  $\psi$  y la función de autocorrelación simple.

- 4.6 Demostrar que el proceso  $y_t = z_t - z_{t-1}$ , donde  $z_t = 0.9z_{t-1} + a_t$  es estacionario, sigue siendo estacionario.
- 4.7 Demostrar que el proceso anterior puede escribirse como  $y_t = -0.09z_{t-2} + a_t - 0.1a_{t-1}$ .
- 4.8 Calcular el operador inverso de  $(1 - 0.8B)(1 - B)$ .
- 4.9 Justificar si el proceso  $(1 - .5B)(1 - .7B)(1 - .2B)z_t = a_t$  es estacionario y escribirlo en su expresión habitual.
- 4.10 Calcular los coeficientes teóricos de autocorrelación parcial para el siguiente proceso AR(2):  $z_t = 0.7z_{t-1} - 0.5z_{t-2} + a_t$ , donde  $a_t$  es ruido blanco.
- 4.11 Demostrar que si un proceso es AR(1), entonces si hacemos la regresión  $z_t = \beta z_{t+1} + u_t$  obtenemos  $\hat{\beta} = \phi$  y  $\text{var}(u_t) = \gamma_0(1 - \phi^2)$ , donde  $\gamma_0$  es la varianza del proceso.
- 4.12 Demostrar que si un proceso es AR(1) y hacemos la regresión  $z_t = \alpha z_{t-1} + \beta z_{t+1} + u_t$ , se obtiene  $\hat{\beta} = \hat{\alpha} = \phi/(1 + \phi^2)$  y  $\text{var}(u_t) = \gamma_0(1 - \phi^2)/(1 + \phi^2)$ . Observar que la varianza de las innovaciones es ahora menor que la de un AR(1) y que la del ejercicio 4.11 e interpretar este resultado.

---

### Apéndice 4.1: Notación de operadores

#### El operador de retardo

Definimos el *operador de retardo*,  $B$ , como un operador lineal que aplicado a una función temporal proporciona esa misma función retardada un periodo. Por definición:

$$Bf(t) \equiv f(t - 1)$$

donde, para simplificar la notación, en lugar de utilizar la notación más precisa  $B[f(t)] \equiv f(t - 1)$ , escribimos el operador como un producto sobre la función. Esta notación debe interpretarse como una operación aplicada a una función,  $f(t)$ , que proporciona otra función,  $f(t - 1)$ . En particular, si aplicamos el operador de retardo a una serie temporal obtendremos la misma serie temporal retardada un periodo, y escribiremos:

$$Bz_t = z_{t-1}.$$

Una constante se interpreta como una función que es constante en el tiempo, y estableceremos que al aplicar el operador de retardo a una constante ésta no se modifica:

$$B\mu = \mu.$$

De acuerdo con la definición, si aplicamos este operador al producto de una constante,  $a$ , por una serie temporal tenemos que:

$$Baz_t = aBz_t = az_{t-1}.$$

El operador de retardo es lineal, es decir:

$$B(af(t) + bg(t)) \equiv af(t-1) + bg(t-1)$$

y puede aplicarse sucesivamente:

$$B^k z_t = \underbrace{B \cdots B}_{k} z_t = z_{t-k}.$$

Definiremos los polinomios en el operador de retardo como nuevos operadores que aplicados a una función proporcionan otra función, de acuerdo con las propiedades anteriores del operador  $B$ . Por ejemplo, el operador  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$  aplicado a una serie temporal proporciona otra serie temporal, mediante la operación:

$$(1 - \phi_1 B - \phi_2 B^2)z_t = z_t - \phi_1 Bz_t - \phi_2 B^2 z_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2}$$

y al aplicar  $\phi(B)$  a una serie,  $z_t$ , se obtiene otra serie,  $y_t$ , que se calcula como  $y_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2}$ .

### El operador diferencia

Un caso particular importante de operador polinómico es el operador  $(1 - B)$ , que se conoce como *operador diferencia*, y se denota por  $\nabla = (1 - B)$ . Entonces

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1}.$$

El resultado de aplicar el operador diferencia a una serie  $z_t$  con  $T$  observaciones es obtener una nueva serie de  $T - 1$  observaciones mediante  $y_t = z_t - z_{t-1}$ , para  $t = 2, \dots, T$ . El operador diferencia puede aplicarse más de una vez:

$$\nabla^2 z_t = (1 - B)^2 z_t = (1 - B)(z_t - z_{t-1}) = z_t - 2z_{t-1} + z_{t-2}.$$

La estacionalidad requerirá tomar diferencias entre observaciones separadas  $s$  períodos. Definimos el *operador diferencia de orden s* como:

$$\nabla_s z_t = (1 - B^s)z_t = z_t - z_{t-s}.$$

### Operadores inversos

Los operadores descritos tienen sus operadores inversos, que verifican la propiedad de que su producto por el operador inicial es la unidad. El inverso del operador de retardo es el operador de adelanto  $F = B^{-1}$ , que verifica  $FB = 1$  y está definido por:

$$Fz_t = B^{-1}z_t = z_{t+1},$$

y se verifica que  $FBz_t = z_t = BFz_t$ . Si aplicamos un operador y su inverso, en cualquier orden, la serie no se modifica.

Generalizando esta definición, podemos definir el inverso de cualquier operador  $\phi(B)$  como un operador  $\phi(B)^{-1}$  tal que  $\phi(B)\phi(B)^{-1} = 1$ . Por ejemplo, nos encontraremos frecuentemente con el operador  $\phi(B) = (1 - \phi B)$ , cuyo inverso se define para  $|\phi| < 1$  como:

$$(1 - \phi B)^{-1} = 1 + \phi B + \phi^2 B^2 + \dots$$

Observemos que esta definición es coherente, ya que

$$(1 - \phi B)(1 - \phi B)^{-1} = 1,$$

y como  $|\phi| < 1$  la suma es convergente y la definición del operador inverso es consistente con la expresión para la suma de una progresión geométrica indefinida. En general, el operador inverso puede obtenerse a partir de la condición  $\phi(B)\phi(B)^{-1} = 1$ , y será una serie infinita. Las condiciones para que esa suma sea convergente definen las situaciones en que podemos definir sin ambigüedad este operador inverso.

Si  $\phi = 1$ , el operador  $(1 - \phi B)$  es el operador diferencia y al invertirlo la suma  $S = (1 - B)^{-1} = 1 + B + B^2 + \dots$  no es convergente. Sin embargo, podemos todavía definir el inverso del operador diferencia como el operador suma:

$$Sz_t = z_t + z_{t-1} + z_{t-2} + \dots$$

ya que es fácil comprobar que  $(1 - B)(1 + B + B^2 + \dots) = 1$ .

### Apéndice 4.2: Ecuaciones en diferencias

Llamaremos ecuación en diferencias a una expresión del tipo:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_k x_{t-k} = c$$

y supondremos que los coeficientes  $\phi_i$  son constantes. Esta es una ecuación de orden  $k$ , que es el máximo retardo de la variable que aparece en la ecuación. Diremos que la ecuación en diferencias es homogénea si  $c = 0$ . Llamaremos solución de esta ecuación a una secuencia  $x_t$  que la verifica para todo  $t$ . Obtener

las soluciones de estas ecuaciones se simplifica introduciendo el operador de retardo  $Bx_t = x_{t-1}$  y escribiendo

$$\phi(B)x_t = c.$$

Llamaremos polinomio característico de una ecuación en diferencias al polinomio  $\phi(B)$  y *ecuación característica* a la ecuación  $\phi(B) = 0$  considerada como función de la variable  $B$ . Por ejemplo, la ecuación característica de una ecuación en diferencias de primer orden es:

$$1 - \phi B = 0,$$

y su solución será  $B = 1/\phi$ .

En este apéndice vamos a demostrar las siguientes propiedades de las soluciones de una ecuación en diferencias:

1. La solución de una ecuación homogénea de primer orden es  $AG^t$ , donde  $A$  es una constante que depende de las condiciones iniciales y  $G^{-1}$  es la solución de la ecuación característica.
2. La solución de una ecuación homogénea cuyo polinomio característico puede escribirse como producto de polinomios de primer grado con raíces distintas es la suma de las soluciones asociadas a cada una de las ecuaciones características de primer orden de los polinomios.
3. La solución de una ecuación homogénea de segundo grado con una raíz  $G^{-1}$  doble es  $(A_1 + A_2 t)G^t$ .
4. La solución general de una ecuación en diferencias es la suma de una solución particular de la ecuación más la solución de la ecuación homogénea.

Vamos a comprobar estas cuatro propiedades. La primera resulta al considerar la ecuación en diferencias de primer orden:

$$x_t - \phi x_{t-1} = (1 - \phi B)x_t = 0.$$

Esta ecuación produce a partir de un valor inicial,  $x_0$ , la secuencia de valores:  $x_1 = \phi x_0$ ,  $x_2 = \phi^2 x_0$ , y, en general:

$$x_t = \phi^t x_0.$$

Hemos comprobado que la solución de la ecuación es el producto de una constante, que es el valor inicial, por la inversa de la solución de la ecuación característica ( $\phi^{-1}$  en este caso) elevada a  $t$ .

Para comprobar la segunda propiedad, consideremos ahora una ecuación de segundo grado,  $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = 0$ , que escribiremos como:

$$(1 - \phi_1 B - \phi_2 B^2)x_t = 0 \quad (4.31)$$

#### 4. Procesos autorregresivos

y si  $G_1^{-1}$  y  $G_2^{-1}$  son las raíces de la ecuación característica, podemos expresar esta ecuación como:

$$(1 - G_1 B)(1 - G_2 B)x_t = 0. \quad (4.32)$$

Supongamos primero que  $G_1 \neq G_2$ . Vamos a demostrar que la solución es la suma de las soluciones de los dos términos que la componen. Es decir, la solución de esta ecuación es de la forma:

$$x_t = S_{1t} + S_{2t}, \quad (4.33)$$

donde  $S_{1t}$  verifica  $(1 - G_1 B)S_{1t} = 0$  y  $S_{2t}$  verifica  $(1 - G_2 B)S_{2t} = 0$ . En efecto, sustituyendo (4.33) en (4.32) tenemos que

$$(1 - G_1 B)(1 - G_2 B)S_{1t} + (1 - G_1 B)(1 - G_2 B)S_{2t} = 0.$$

Como la solución de  $(1 - G_i B)S_{it} = 0$  debe ser de la forma  $S_{it} = A_i G_i^t$ , donde las  $A_i$  son constantes que dependen de las condiciones iniciales, tenemos que la solución general de (4.32) debe ser de la forma:

$$x_t = A_1 G_1^t + A_2 G_2^t. \quad (4.34)$$

Por ejemplo, supongamos que tenemos la secuencia de autocorrelaciones de una AR(2) con  $\rho_0 = 1$  y  $\rho_1 = \phi_1 / (1 - \phi_2)$ . Entonces, según (4.34) la solución general es:

$$\rho_k = A_1 G_1^k + A_2 G_2^k \quad (4.35)$$

donde, por ser  $G_1$  y  $G_2$  factores de la ecuación, deberá verificarse que  $\phi_1 = G_1 + G_2$  y  $\phi_2 = -G_1 G_2$ . Para obtener las constantes  $A_1, A_2$ , particularizando para  $k = 0$  tenemos  $1 = A_1 + A_2$  y para  $k = 1$  se obtiene:

$$\rho_1 = \phi_1 / (1 - \phi_2) = (G_1 + G_2) / (1 + G_1 G_2) = A_1 G_1 + A_2 G_2.$$

De estas dos ecuaciones resulta:

$$A_1 = \frac{G_1 (1 - G_2^2)}{(1 + G_1 G_2)(G_1 - G_2)}; \quad A_2 = \frac{-G_2 (1 - G_1^2)}{(G_1 - G_2)(1 + G_1 G_2)}$$

Si las dos raíces son reales, la solución será la suma de dos decrecimientos geométricos. Sin embargo, cuando las raíces son complejas la solución (4.34) implica un comportamiento sinusoidal. En efecto, si  $G_1$  y  $G_2$  son complejos conjugados pueden escribirse como  $G_1 = r \exp(i\omega)$  y  $G_2 = r \exp(-i\omega)$ . El módulo de estos números se calcula por:

$$\begin{aligned} r^2 &= G_1 G_2 = -\phi_2 \\ r &= \sqrt{-\phi_2} \end{aligned} \quad (4.36)$$

y vemos que para que la solución sea compleja  $\phi_2$  debe ser negativo. Para obtener la frecuencia angular en función de los parámetros observemos que:

$$\phi_1 = G_1 + G_2 = r (\exp(i\omega) + \exp(-i\omega)) = 2r \cos \omega$$

llamando  $\omega = 2\pi/T$ , donde  $T$  es el periodo, y utilizando (4.36) se obtiene:

$$\cos \omega = \cos \frac{2\pi}{T_0} = \frac{\phi_1}{2\sqrt{-\phi_2}}. \quad (4.37)$$

Sustituyendo  $A_1, A_2, G_1$ , y  $G_2$  en (4.35) y operando, se obtiene:

$$\rho_k = \frac{[sg(\phi_1)]^k r^k \operatorname{sen}(2\pi k/T_0 + \alpha)}{\operatorname{sen} \alpha},$$

donde  $tg \alpha = (1+r^2) \operatorname{tag}(2\pi k/T_0)/(1-r^2)$  y  $sg(\phi_1)$  es el signo del coeficiente  $\phi_1$ . Por tanto, la función de autocorrelación será una senoide amortiguada con factor de amortiguación  $r$  y periodo  $T_0$ .

La demostración de esta propiedad para polinomios generales que pueden escribirse como:

$$(1 - G_1 B)(1 - G_2 B) \dots (1 - G_k B)x_t = 0$$

donde las raíces  $G_i$  son distintas se realiza análogamente. La solución de la ecuación será la suma de las soluciones de las ecuaciones de primer orden y vendrá dada por

$$x_t = A_1 G_1^t + \dots + A_k G_k^t, \quad (4.38)$$

donde las  $A_1, \dots, A_k$  son constantes que dependen de las condiciones iniciales y  $G_1^{-1}, \dots, G_k^{-1}$  son las raíces de la ecuación característica, o  $G_i$  son los factores de esta ecuación, de manera que:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k) = (1 - G_1 B) \dots (1 - G_k B).$$

Observemos, además, que para que  $x_t$  tienda a cero cuando  $t \rightarrow \infty$ ,  $G_i^t$  debe tender a cero, lo que requiere que el módulo de todas las soluciones sea menor que la unidad.

Vamos a demostrar la tercera propiedad. Hemos supuesto hasta ahora que las raíces de la ecuación característica son distintas. Comprobaremos que cuando las dos raíces son iguales los dos términos  $A_i G_i^t$  correspondientes pueden escribirse como  $(A_1 + A_2 t) G^t$ . En efecto, supongamos la ecuación (4.31) cuya ecuación característica es:

$$(1 - \phi_1 B - \phi_2 B^2) = 0. \quad (4.39)$$

Si existe una raíz doble en esta ecuación,  $G_0^{-1}$ , se verifica:

$$(1 - G_0 B)^2 = (1 - \phi_1 B - \phi_2 B^2),$$

que implica  $\phi_2 = -G_0^2$  y  $\phi_1 = 2G_0$ , por tanto

$$\phi_1 G_0 + 2\phi_2 = 0. \quad (4.40)$$

#### 4. Procesos autorregresivos

---

Comprobemos que si  $G_0^t$  es una solución de (4.39) también debe serlo  $tG_0^t$ . Sustituyendo:

$$(1 - \phi_1 B - \phi_2 B^2) tG_0^t = tG_0^t - \phi_1 (t-1) G_0^{t-1} - \phi_2 (t-2) G_0^{t-2}$$

y sacando factores comunes, tenemos que:

$$tG_0^{t-2} (G_0^2 - \phi_1 G_0 - \phi_2) + G_0^{t-2} (\phi_1 G_0 + 2\phi_2) = 0.$$

El primer término es cero, por ser  $G_0$  solución, y el segundo por ser doble –según (4.40)–, con lo que concluimos que la solución general es:

$$x_t = (A_1 + A_2 t) G_0^t.$$

Estos resultados se extienden para ecuaciones de cualquier orden y con raíces iguales o distintas. Supongamos la ecuación lineal homogénea en diferencias:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_k x_{t-k} = 0, \quad (4.41)$$

que escribiremos como:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_k B^k) x_t = 0.$$

Si todas las raíces son distintas, sabemos que la solución general es:

$$x_t = A_1 G_1^t + \dots + A_k G_k^t, \quad (4.42)$$

donde las  $A_1, \dots, A_k$  son constantes que dependen de las condiciones iniciales y  $G_1^{-1}, \dots, G_k^{-1}$  son las raíces de la ecuación característica. Si existen raíces repetidas  $h$  veces, es decir existe un término de la forma  $(1 - G_0 B)^h$ , es fácil comprobar que la solución asociada a ese término es:

$$x_t = (A_1 + A_2 t + \dots + A_h t^{h-1}) G_0^t. \quad (4.43)$$

Por tanto, la solución general de (4.41) tendrá: (1) por cada raíz real no repetida un término del tipo  $A_i G_i^t$ ; (2) por cada raíz real repetida  $h$  veces un término polinómico de orden  $(h-1)$  que multiplica a  $G_0^t$ ; (3) por cada par de raíces complejas conjugadas un término sinusoidal.

Comprobemos finalmente la propiedad cuarta. Supongamos la ecuación  $\phi(B)x_t = c$ . Vamos a comprobar que si  $y_t$  es una solución, también lo es  $y_t + q_t$ , donde  $q_t$  es una solución de la ecuación homogénea,  $\phi(B)q_t = 0$ . La comprobación es inmediata, ya que:

$$\phi(B)(y_t + q_t) = \phi(B)y_t = c.$$

### Validez de la solución y valores iniciales

Una variable puede verificar una ecuación en diferencias a partir de un instante definido. Entonces, la solución de la ecuación proporciona valores de la variable que verifican la ecuación no sólo desde el instante definido, sino también para los instantes previos que son necesarios para calcular la solución. Es decir, si la ecuación es válida para  $t \geq t_0$  y se necesitan  $h$  valores iniciales para resolverla, la solución se aplica a los valores desde  $t \geq t_0 - h$ , es decir, desde el origen de la ecuación menos el número de valores iniciales requeridos para calcular la solución.

Para ilustrar este resultado, comenzemos con la ecuación de primer orden:

$$x_t - \phi x_{t-1} = 0, \quad t \geq t_0.$$

Si la ecuación es válida para  $t = t_0$  tendremos que  $x_{t_0} = \phi x_{t_0-1}$ ,  $x_{t_0+1} = \phi x_{t_0} = \phi^2 x_{t_0-1}$ , y, en general,

$$x_t = \phi^{t-t_0+1} x_{t_0-1}.$$

Observemos que esta solución es válida para  $t \geq t_0 - 1$ . En efecto, para  $t = t_0 - 1$  proporciona el resultado  $x_{t_0-1} = x_{t_0-1}$ , que es trivialmente cierto. Por tanto, concluimos que la solución es válida para el intervalo de definición de la ecuación, menos el número de valores iniciales necesarios para iniciar la secuencia.

Vamos a comprobar esta regla en una ecuación de segundo grado:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = 0, \quad t \geq 1.$$

De los resultados anteriores deducimos que la solución es:

$$x_t = A_1 G_1^t + A_2 G_2^t,$$

donde  $G_1^{-1}$  y  $G_2^{-1}$  son las raíces de la ecuación característica. La solución es válida para  $t \geq -1$ . En efecto, para  $t = 0$  obtenemos

$$x_0 = A_1 + A_2$$

y para  $t = -1$ , resulta:

$$x_{-1} = A_1 G_1^{-1} + A_2 G_2^{-1}.$$

Vamos a comprobar que los tres valores  $x_1, x_0$  y  $x_{-1}$  verifican la ecuación. Sustituyendo en la ecuación de segundo grado

$$\begin{aligned} x_1 &= A_1 G_1 + A_2 G_2 = \phi_1 x_0 + \phi_2 x_{-1} = \\ &= \phi_1(A_1 + A_2) + \phi_2(A_1 G_1^{-1} + A_2 G_2^{-1}) \end{aligned}$$

#### 4. Procesos autorregresivos

y agrupando términos

$$A_1G_1 + A_2G_2 = A_1(\phi_1 + \phi_2G_1^{-1}) + A_2(\phi_1 + \phi_2G_2^{-1})$$

que es siempre cierta, como se comprueba utilizando la relación entre los parámetros y las raíces,  $\phi_1 = G_1 + G_2$  y  $\phi_2 = -G_1G_2$ , que se obtuvo al resolver la ecuación de segundo grado.

Generalizando, si tenemos una ecuación de diferencias de orden  $h$  válida para  $t \geq t_0$  serán necesarios  $h$  valores iniciales para comenzar la secuencia. Como estos valores iniciales se obtienen con la ecuación que proporciona la solución, la solución proporciona valores válidos de la secuencia para  $t \geq t_0 - h$ .

#### Apéndice 4.3: Cálculo de los coeficientes de autocorrelación parcial

El procedimiento presentado en la sección 4.6 para calcular los coeficientes de correlación parcial es equivalente a obtener el cociente de determinantes:

$$\alpha_{kk} = \frac{|\mathbf{A}(k, r_k)|}{|\mathbf{R}_k|}$$

donde  $\mathbf{R}_k$  es la matriz de coeficientes de autocorrelación de orden  $k$

$$\mathbf{R}_k = \begin{bmatrix} 1 & r_1 & \dots & r_{k-1} \\ \vdots & \vdots & & \vdots \\ r_{k-1} & r_{k-2} & \dots & 1 \end{bmatrix}$$

y  $\mathbf{A}(k, r_k)$  es el resultado de sustituir en esta matriz la última columna por el vector  $\mathbf{r}_k = (r_1, \dots, r_k)'$ . Este resultado es la regla de Cramer para resolver ecuaciones lineales. Teniendo en cuenta que podemos escribir el determinante de  $\mathbf{R}_k$  como:

$$\mathbf{R}_k = \begin{bmatrix} \mathbf{R}_{k-1} & \tilde{\mathbf{r}}'_{k-1} \\ \tilde{\mathbf{r}}_{k-1} & 1 \end{bmatrix}$$

donde  $\tilde{\mathbf{r}}_{k-1} = (r_{k-1}, \dots, r_1)'$  y entonces:

$$|\mathbf{R}_k| = |\mathbf{R}_{k-1}| (1 - \tilde{\mathbf{r}}'_{k-1} \mathbf{R}_{k-1}^{-1} \tilde{\mathbf{r}}_{k-1}).$$

Por otro lado:

$$\mathbf{A}(k, r_k) = \begin{bmatrix} \mathbf{R}_{k-1} & \mathbf{r}'_{k-1} \\ \tilde{\mathbf{r}}_{k-1} & r_k \end{bmatrix}$$

y tenemos que  $|\mathbf{A}(k, r_k)| = |\mathbf{R}_{k-1}| (r_k - \mathbf{r}'_{k-1} \mathbf{R}_{k-1}^{-1} \tilde{\mathbf{r}}_{k-1})$ , con lo que:

$$\alpha_{kk} = \frac{(r_k - \mathbf{r}'_{k-1} \mathbf{R}_{k-1}^{-1} \tilde{\mathbf{r}}_{k-1})}{(1 - \tilde{\mathbf{r}}'_{k-1} \mathbf{R}_{k-1}^{-1} \tilde{\mathbf{r}}_{k-1})},$$

que proporciona un método rápido de cálculo de los coeficientes de correlación parcial.

## Apéndice 4.4: Generación de procesos AR

La generación de un proceso autorregresivo con Minitab o SPSS es complicado, por lo que vamos a indicar cómo hacerlo con EViews y Matlab, que es mucho más simple.

### EViews

Comencemos con un proceso AR(1). Para ello creamos un workfile de, por ejemplo, 200 datos de la manera habitual. En primer lugar generamos una serie donde  $c = 0$ ,  $\phi_1 = 0.5$  y  $\sigma_a^2 = 1$ . Para ello debemos seguir los pasos:

1. *File → New → Workfile*. Ahora dentro de la ventana de Workfile marcar:
2. *Undated or Irregular → 1 to 200*. Después buscar la pestaña Genr dentro de la ventana de Workfile.
3. *Genr → a=nrnd*. (Al abrir Genr aparece una ventana donde podemos escribir la ecuación. En la parte de abajo aparece *sample* y automáticamente el programa escribe el rango 1 200 ya especificado. Esto nos define la serie de ruido blanco).
4. *Genr → x1=0*. (Al abrir Genr de nuevo aparece la ventana donde podemos escribir la ecuación. En la parte de abajo, *sample*, el programa escribe el rango 1 200 ya especificado. Con esto se define la serie como 0).
5. *Genr → Sample: 2 200 → x1=0.5\*x1(-1)+a*. (Al abrir de nuevo Genr comenzar ahora modificando en la parte de abajo, *sample*, el nuevo rango donde vamos a dar valores a la variable y escribiendo 2 200. Después escribimos la ecuación. De esta forma la serie toma sus valores definitivos).

El proceso de generación asume que  $x_1 = 0$ . Para ver la serie generada: *View → Line Graph*.

La serie debe tomar valores alrededor de la media del proceso, 0, y no tener cambios en la varianza.

Para generar un AR(2) los pasos 1 a 3 son los mismos pero ahora definimos los valores iniciales y la ecuación del proceso como:

1. *Genr → x1=1*.
2. *Genr →: Sample: 3 200 → x1=1.4\*x1(-1)-0.7\*x1(-2)+a*.

El proceso de generación asume que  $x_1 = 1$  y  $x_2 = 1$ . Podemos hacer un gráfico y mirar su función de autocorrelación simple y parcial.

#### 4. Procesos autorregresivos

##### **Matlab**

El siguiente programa genera una serie AR con vector de parámetros en el vector  $\phi$ . Primero especificamos el vector de parámetros. Por ejemplo, para simular un AR(2) con parámetros .3 y .4 escribimos en los comandos de Matlab.

```
fi=[.3 .4]
```

A continuación llamamos al programa  $y=simarma(\phi,\theta,n)$  donde  $n$  es el tamaño deseado, que nos proporciona la serie simulada en el vector  $y$ . El programa es el siguiente:

```
% Esta función simula un ARMA(p,q). La entrada son los vectores fila con los parámetros AR y MA y el tamaño muestral, n.
```

```
function y=simarma(phi,theta,n)
p=prod(size(phi));
q=prod(size(theta));
a=randn(2*n,1);
a0=randn(q,1);
xo=randn(p,1);
for i=1:2*n
    y(i)=phi*xo-theta*a0+a(i);
    xo(1)=[];
    xo=[xo y(i)]';
    a0(1)=[];
    a0=[a0 a(i)]';
end
y=y(n+1:2*n);
```

## 5. Procesos de media móvil y ARMA



Herman Wold (1908-1992)

Estadístico y económetra sueco. Ha hecho contribuciones fundamentales a la representación general de procesos estacionarios y a los problemas de identificación y causalidad en econometría. Es, además, el inventor de los mínimos cuadrados parciales (PLS), método de estimación muy utilizado en aplicaciones químicas y sociológicas. Fue profesor de estadística en la Universidad de Uppsala desde 1942 hasta 1970 y en la de Goteborg desde 1970 hasta 1975, y presidente de la Econometric Society.

### 5.1. Introducción

En el capítulo anterior hemos estudiado una clase de procesos estacionales lineales, los procesos autorregresivos, que se caracterizan por tener muchos coeficientes de autocorrelación distintos de cero y que decrecen con el retardo. Estos procesos tienen memoria relativamente larga, ya que el valor actual está correlado con todos los anteriores, aunque con coeficientes decrecientes. Como vimos en el capítulo anterior, esta propiedad se traduce en que podemos escribir un proceso AR como una función lineal de todas las innovaciones que le han generado, con pesos que tienden a cero con el retardo. Los procesos AR no pueden representar series de memoria muy corta, donde el valor actual de la serie sólo está correlado con un número pequeño de valores anteriores, de manera que la función de autocorrelación simple tenga sólo unas pocas autocorrelaciones distintas de cero. Una familia de procesos que tienen esta propiedad de «memoria muy corta» son los procesos de media móvil, o procesos MA de su nombre en inglés, *moving average*, que estudiaremos en este capítulo. Estos

## 5. Procesos de media móvil y ARMA

procesos MA son función de un número finito, y generalmente pequeño, de las innovaciones pasadas.

Después, combinaremos las propiedades de los procesos AR y MA para definir los procesos ARMA, que dan lugar a una familia muy amplia y flexible de procesos estocásticos estacionarios útiles para representar muchas series temporales. Los procesos puros AR y MA aparecen, entonces, como casos particulares de esta representación general ARMA.

### 5.2. El proceso de media móvil de orden uno (MA(1))

Se define el proceso de media móvil de orden uno, MA(1), añadiendo a un proceso de ruido blanco una dependencia del valor actual de la serie de la última innovación ocurrida. De esta manera, el proceso será una combinación lineal de las dos últimas innovaciones, de acuerdo con la ecuación:

$$\tilde{z}_t = a_t - \theta a_{t-1}, \quad (5.1)$$

donde  $\tilde{z}_t = z_t - \mu$ , siendo  $\mu$  la media del proceso y  $a_t$  es un proceso de ruido blanco con varianza  $\sigma^2$ . El proceso MA(1) puede escribirse con la notación de operadores:

$$\tilde{z}_t = (1 - \theta B) a_t. \quad (5.2)$$

Este proceso es la suma de los dos procesos estacionarios,  $a_t$  y  $-\theta a_{t-1}$  y, por tanto, siempre será estacionario, para cualquier valor del parámetro, a diferencia de los procesos AR.

En las aplicaciones de este proceso supondremos que  $|\theta| < 1$ , de manera que la innovación pasada tenga menos peso que la presente. Entonces, decimos que el proceso es *invertible* y tiene la propiedad de que el efecto de los valores pasados decrece con el tiempo. Para justificar esta propiedad, sustituimos  $a_{t-1}$  en (5.1) como función de  $z_{t-1}$ . El resultado es:

$$\tilde{z}_t = a_t - \theta (\tilde{z}_{t-1} + \theta a_{t-2}) = -\theta \tilde{z}_{t-1} - \theta^2 a_{t-2} + a_t$$

y repitiendo esta operación para  $a_{t-2}$ :

$$\tilde{z}_t = -\theta \tilde{z}_{t-1} - \theta^2 (\tilde{z}_{t-2} + \theta a_{t-3}) + a_t = -\theta \tilde{z}_{t-1} - \theta^2 \tilde{z}_{t-2} - \theta^3 a_{t-3} + a_t$$

mediante sustituciones sucesivas de  $a_{t-3}$ ,  $a_{t-4}, \dots$ , etc., se obtiene:

$$\tilde{z}_t = -\sum_{i=1}^{t-1} \theta^i \tilde{z}_{t-i} - \theta^t a_0 + a_t. \quad (5.3)$$

Se observa que si  $|\theta| < 1$ , el efecto de  $\tilde{z}_{t-k}$  tiende hacia cero con  $k$  y el proceso se denomina invertible. Si  $|\theta| \geq 1$  se daría la situación paradójica en que el efecto de las observaciones pasadas aumentaría con la distancia y, aunque el proceso seguiría siendo estacionario, parece poco adecuado para representar series reales. En adelante, supondremos que el proceso es invertible. Entonces, como  $|\theta| < 1$ , existe el operador inverso  $(1 - \theta B)^{-1}$  y podemos escribir la ecuación (5.2) como:

$$(1 + \theta B + \theta^2 B^2 + \dots) \tilde{z}_t = a_t, \quad (5.4)$$

que equivale a:

$$\tilde{z}_t = - \sum_{i=1}^{\infty} \theta^i \tilde{z}_{t-i} + a_t,$$

que es la (5.3) suponiendo que el proceso comienza en el infinito pasado. Esta ecuación representa el proceso MA(1) invertible como un AR( $\infty$ ) con coeficientes que decrecen en progresión geométrica.

La varianza del proceso se calcula a partir de (5.1). Elevando al cuadrado y tomando esperanzas, se obtiene:

$$E(\tilde{z}_t^2) = E(a_t^2) + \theta^2 E(a_{t-1}^2) - 2\theta E(a_t a_{t-1})$$

como  $E(a_t a_{t-1}) = 0$ , por ser un proceso de ruido blanco, y  $E(a_t^2) = E(a_{t-1}^2) = \sigma^2$ , tenemos que:

$$\sigma_z^2 = \sigma^2 (1 + \theta^2). \quad (5.5)$$

Esta ecuación nos indica que la varianza marginal del proceso,  $\sigma_z^2$ , es siempre mayor que la varianza de las innovaciones,  $\sigma^2$ , y tanto mayor cuanto mayor sea  $\theta^2$ .

### 5.2.1. Función de autocorrelación simple y parcial

Para obtener la función de autocorrelación simple de un proceso MA(1) comenzaremos calculando las covarianzas. La de primer orden se obtiene multiplicando la ecuación (5.1) por  $\tilde{z}_{t-1}$  y tomando esperanzas:

$$\gamma_1 = E(\tilde{z}_t \tilde{z}_{t-1}) = E(a_t \tilde{z}_{t-1}) - \theta E(a_{t-1} \tilde{z}_{t-1}).$$

En esta expresión el primer término  $E(a_t \tilde{z}_{t-1})$  es cero, ya que  $\tilde{z}_{t-1}$  según (5.1) depende de su innovación,  $a_{t-1}$ , y de la innovación previa,  $a_{t-2}$ , pero no de las innovaciones futuras, como  $a_t$ . Para calcular el segundo término, sustituyendo  $\tilde{z}_{t-1}$  por su expresión según (5.1), resulta:

$$E(a_{t-1} \tilde{z}_{t-1}) = E(a_{t-1}(a_{t-1} - \theta a_{t-2})) = \sigma^2,$$

## 5. Procesos de media móvil y ARMA

---

con lo que se obtiene:

$$\gamma_1 = -\theta\sigma^2. \quad (5.6)$$

De la misma forma, se calcula la autocovarianza de orden dos:

$$\gamma_2 = E(\tilde{z}_t \tilde{z}_{t-2}) = E(a_t \tilde{z}_{t-2}) - \theta E(a_{t-1} \tilde{z}_{t-2}) = 0,$$

ya que al estar la serie incorrelada con sus innovaciones futuras, los dos términos son nulos. El mismo resultado se obtiene para las covarianzas de orden superior a dos, ya que al multiplicar la ecuación (5.1) por  $\tilde{z}_{t-j}$ , donde  $j > 1$ , tendremos productos de la serie por sus innovaciones futuras y las esperanzas de estos términos son cero. En conclusión:

$$\gamma_j = 0, \quad j > 1. \quad (5.7)$$

Dividiendo las autocovarianzas (5.6) y (5.7) por la expresión (5.5) de la varianza del proceso, obtenemos que los coeficientes de autocorrelación de un proceso MA(1) verifican:

$$\rho_1 = \frac{-\theta}{1 + \theta^2}, \quad \rho_k = 0 \quad k > 1, \quad (5.8)$$

y la *fas* tendrá únicamente un valor distinto de cero en el primer retardo. Como  $|\theta| < 1$  el valor del coeficiente de autocorrelación en un MA(1) invertible es siempre menor que 0.5.

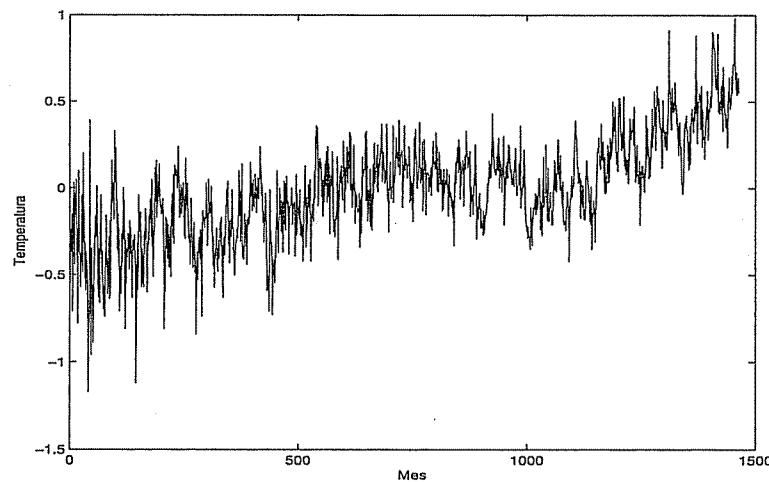
Este resultado demuestra que la función de autocorrelación simple (*fas*) de un proceso MA(1) tiene las mismas propiedades que la función de autocorrelación parcial (*fap*) de un proceso AR(1): existe un primer coeficiente distinto de cero y todos los demás son nulos. Esta dualidad entre el AR(1) y el MA(1) se presenta también en la función de autocorrelación parcial, *fap*. En efecto, según (5.4), al escribir un proceso MA(1) en forma autorregresiva hay un efecto directo de  $z_{t-k}$  sobre  $z_t$  de magnitud  $\theta^k$ , cualquiera que sea  $k$ . Por tanto, la *fap* tendrá todos los coeficientes no nulos y que decrecen geométricamente con  $k$ . Esta es la estructura de la *fas* de un AR(1) y, por tanto, concluimos que la *fap* de un MA(1) tiene la misma estructura que la *fas* de un AR(1).

---

### Ejemplo 5.1

Los datos de la serie que vamos a estudiar están en el fichero tempmundo.dat. Son datos mensuales entre los años 1881-2002 y representan la desviación entre la temperatura promedio de un mes y la media de ese mes calculada promediando las temperaturas en los 25 años entre 1951-1975. Esta serie se ha

**Figura 5.1 Temperaturas de la Tierra en desviaciones a la media del mes entre 1881 y 2002**



construido para estudiar el calentamiento de la Tierra. Su gráfico se presenta en la figura 5.1. Vemos que esta serie,  $y_t$ , tiene tendencia creciente, por lo que no es estacionaria. Sin embargo,  $z_t = y_t - y_{t-1}$ , que representa las variaciones de temperatura media en la Tierra entre un mes y el siguiente, es una serie estacionaria, como indica su gráfico que aparece en la figura 5.2.

Se observa que la serie  $z_t$  es estacionaria y parece mostrar correlación negativa. La figura 5.3 muestra su función de autocorrelación, donde se observa un solo coeficiente diferente de cero. El gráfico 5.4 muestra la fap, donde se observa un decrecimiento geométrico. Ambos gráficos sugieren un modelo MA(1) para la serie de diferencias entre meses,  $z_t$ .

---

### 5.3. El proceso MA( $q$ )

Generalizando la idea de un MA(1), podemos escribir procesos cuyo valor actual dependa no sólo de la última innovación, sino de las  $q$  últimas innovaciones. Se obtiene entonces el proceso MA( $q$ ), con representación general:

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}.$$

Introduciendo la notación de operadores:

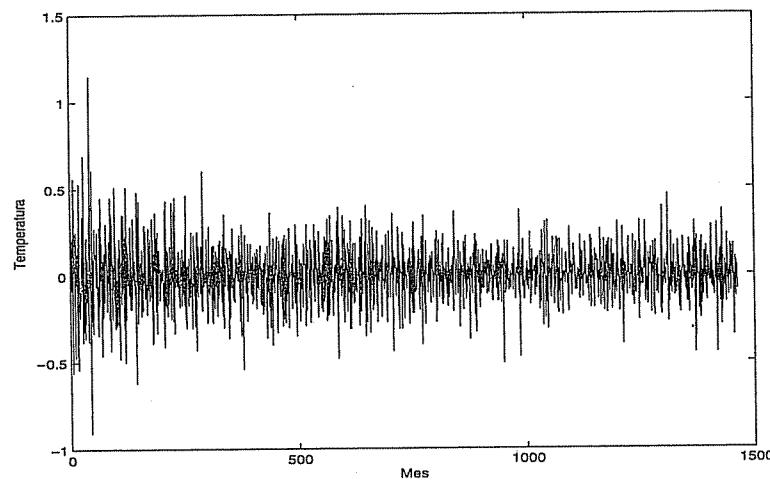
$$\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (5.9)$$

puede escribirse de forma más compacta como:

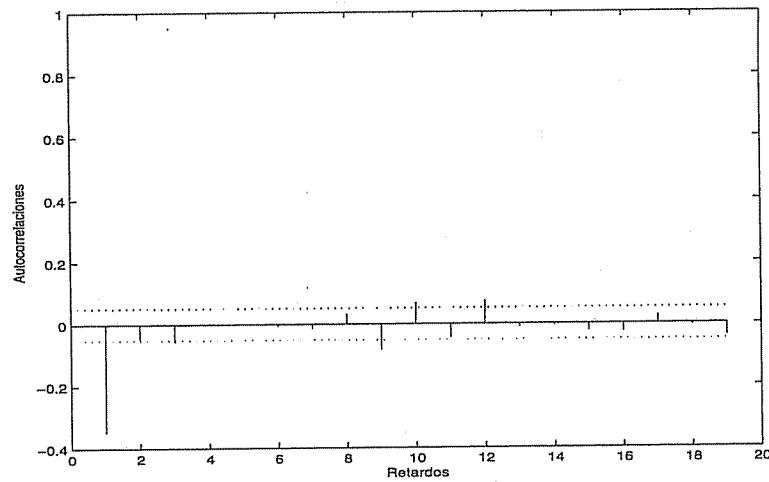
## 5. Procesos de media móvil y ARMA

---

**Figura 5.2 Variaciones mensuales sobre las medias de los meses de las temperaturas en la Tierra en el periodo 1881-2002**



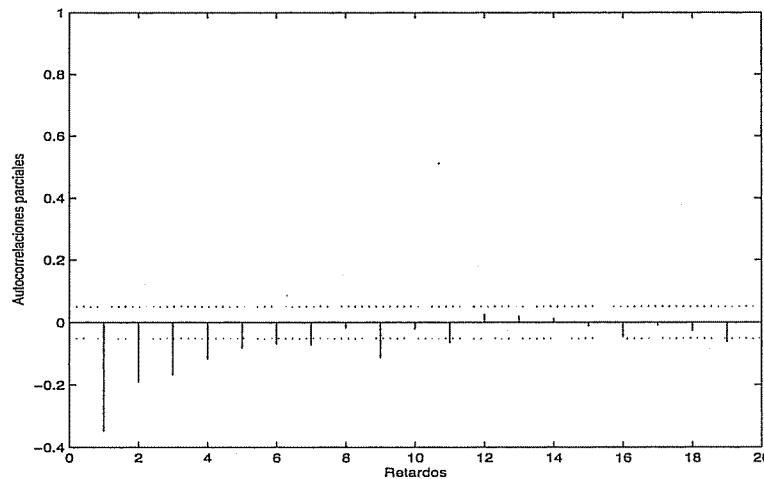
**Figura 5.3 Función de autocorrelación para la serie de desviaciones de la temperatura de la Tierra**



$$\tilde{z}_t = \theta_q(B) a_t. \quad (5.10)$$

Un proceso  $MA(q)$  es siempre estacionario, por ser la suma de procesos estacionarios. Diremos que el proceso es *invertible* si las raíces del operador  $\theta_q(B) = 0$  son, en módulo, mayores que la unidad.

**Figura 5.4 Función de autocorrelación parcial para la serie de desviaciones de la temperatura de la Tierra**



Las propiedades de este proceso se obtienen con el mismo método utilizado para el MA(1). Multiplicando (5.9) por  $\tilde{z}_{t-k}$  para  $k \geq 0$  y tomando esperanzas, se obtienen las autocovarianzas:

$$\begin{aligned}\gamma_0 &= (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2 \\ \gamma_k &= (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma^2 \quad k = 1, \dots, q, \\ \gamma_k &= 0 \quad k > q,\end{aligned}$$

resultando que un proceso MA( $q$ ) tiene exactamente los  $q$  primeros coeficientes de la función de autocovarianzas distintos de cero. Dividiendo las covarianzas por  $\gamma_0$  y utilizando una notación más compacta, la función de autocorrelación será:

$$\begin{aligned}\rho_k &= \frac{\sum_{i=0}^{i=q} \theta_i \theta_{k+i}}{\sum_{i=0}^{i=q} \theta_i^2}, \quad k = 1, \dots, q \\ \rho_k &= 0, \quad k > q,\end{aligned}\tag{5.11}$$

donde  $\theta_0 = -1$ , y  $\theta_k = 0$  para  $k \geq q + 1$ . En particular, para  $q = k = 1$  se obtiene la ecuación (5.8).

Para calcular la función de autocorrelación parcial de un MA( $q$ ) expresaremos el proceso como un AR( $\infty$ ):

$$\theta_q^{-1}(B) \tilde{z}_t = a_t,$$

## 5. Procesos de media móvil y ARMA

---

y llamando  $\theta_q^{-1}(B) = \pi(B)$ , donde:

$$\pi(B) = 1 - \pi_1 B - \dots - \pi_k B^k - \dots$$

y los coeficientes de  $\pi(B)$  se obtienen imponiendo que  $\pi(B)\theta_q(B) = 1$ . Suponer que el proceso es invertible implica que las raíces de  $\theta_q(B) = 0$  están fuera del círculo unidad, y la serie  $\pi(B)$  será convergente. Igualando las potencias de  $B$  a cero, se obtiene que los coeficientes  $\pi_i$  verifican la relación siguiente:

$$\pi_k = \theta_1 \pi_{k-1} + \dots + \theta_q \pi_{k-q}$$

donde  $\pi_0 = -1$  y  $\pi_j = 0$  para  $j < 0$ . La solución de esta ecuación en diferencias será de la forma  $\sum A_i G_i^k$ , donde ahora los  $G_i^{-1}$  son las raíces del operador de media móvil. Obtenidos los coeficientes  $\pi_i$  de la representación AR( $\infty$ ), podemos escribir el proceso MA como:

$$\tilde{z}_t = \sum_{i=1}^{\infty} \pi_i \tilde{z}_{t-i} + a_t.$$

Esta expresión permite concluir que la *fap* de un MA será no nula para todo retardo, ya que existe un efecto directo de  $\tilde{z}_{t-i}$  sobre  $\tilde{z}_t$  para todo  $i$ . La *fap* de un proceso MA tendrá pues la misma estructura que la *fas* de un proceso AR del mismo orden.

Concluimos que existe una dualidad entre procesos AR y MA, de manera que la *fap* de un MA( $q$ ) tiene la estructura de la *fas* de un AR( $q$ ), y la *fas* de un MA( $q$ ) tiene la estructura de la *fap* de un AR( $q$ ). La figura 5.5 presenta estas funciones para procesos MA(1) y MA(2).

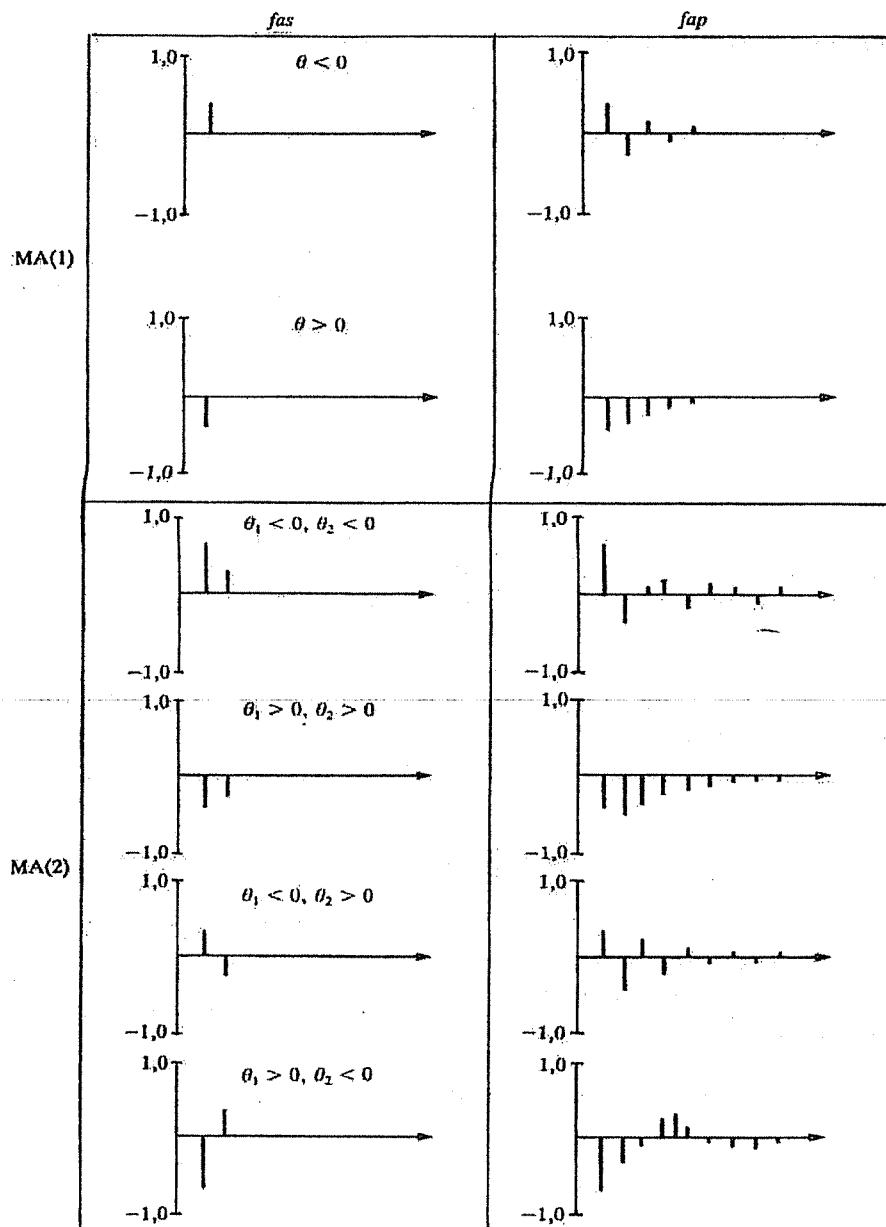
### 5.4. El proceso MA( $\infty$ ). La descomposición de Wold

Los procesos autorregresivos y de media móvil estudiados en las secciones anteriores son casos particulares de una representación general de procesos estacionarios obtenida por Wold (1938). Este autor demostró que todo proceso estocástico débilmente estacionario,  $z_t$ , de media finita,  $\mu$ , que no contenga componentes deterministas, puede escribirse como una función lineal de variables aleatorias incorreladas,  $a_t$ , como:

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad (\psi_0 = 1) \quad (5.12)$$

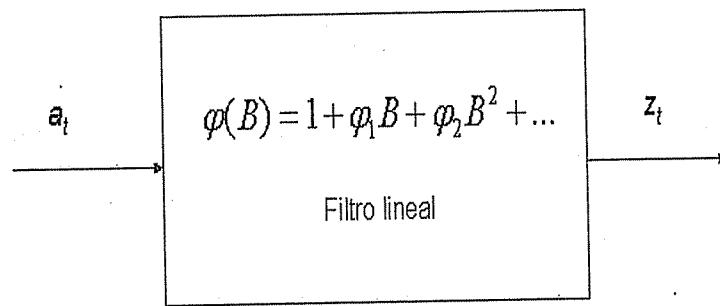
donde  $E(z_t) = \mu$ , y  $E(a_t) = 0$ ;  $Var(a_t) = \sigma^2$ ;  $E(a_t a_{t-k}) = 0$ ,  $k > 1$ . Llamando  $\tilde{z}_t = z_t - \mu$ , y utilizando el operador de retardo, podemos escribir:

Figura 5.5 Funciones de autocorrelación simple y parcial de procesos MA



$$\tilde{z}_t = \psi(B)a_t, \quad (5.13)$$

Figura 5.6 La descomposición de Wold como un filtro lineal



siendo  $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$  un polinomio indefinido en el operador de retardo  $B$ . Llamaremos a (5.13) la representación lineal general de un proceso estacionario no determinista. Esta representación es importante, porque nos garantiza que cualquier proceso estacionario admite una representación lineal. En general, las variables  $a_t$  forman un proceso de ruido blanco, es decir están incorreladas con media cero y varianza constante. En ciertos casos particulares el proceso puede escribirse en función de variables  $\{a_t\}$  normales independientes. Entonces la variable  $\tilde{z}_t$  tendrá una distribución normal, y la estacionaridad débil coincide con la estricta.

La figura 5.6 presenta una interpretación física del teorema de Wold. La serie  $\tilde{z}_t$ , puede considerarse el resultado de pasar un proceso de impulsos  $\{a_t\}$  de variables incorreladas a través del filtro lineal  $\psi(B)$  que determina la ponderación de cada «impulso» en la respuesta.

Las propiedades del proceso se obtienen como en los casos de un modelo MA. La varianza de  $z_t$  en (5.12) será:

$$Var(z_t) = \gamma_0 = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2, \quad (5.14)$$

y para que el proceso tenga varianza finita, la serie  $\{\psi_i^2\}$  debe ser convergente. Observemos que si los coeficientes  $\psi_i$  se anulan a partir de un retardo  $q$ , el modelo general se reduce a un MA( $q$ ) y la fórmula (5.14) coincide con la obtenida en el apartado anterior. Las covarianzas se obtienen con

$$\gamma_k = E(\tilde{z}_t \tilde{z}_{t-k}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k},$$

que para  $k = 0$  proporcionan, como caso particular, la fórmula (5.14) para la varianza. También, si los coeficientes  $\psi_i$  se anulan a partir de un retardo  $q$ , esta expresión proporciona las autocovarianzas de un proceso MA( $q$ ). Los coeficientes de autocorrelación vendrán dados por:

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}, \quad (5.15)$$

que generaliza la expresión (5.11) de las autocorrelaciones de un MA( $q$ ).

Una consecuencia de (5.12) es que todo proceso estacionario admite también una representación autorregresiva, que puede ser de orden infinito. Esta representación es la inversa de la de Wold, y escribiremos:

$$\tilde{z}_t = \pi_1 \tilde{z}_{t-1} + \pi_2 \tilde{z}_{t-2} + \dots + a_t,$$

que, en notación de operadores, se reduce a:

$$\pi(B) \tilde{z}_t = a_t.$$

La representación AR( $\infty$ ) es la dual de la MA( $\infty$ ) y se verifica:

$$\pi(B) \psi(B) = 1,$$

con lo que, igualando las potencias de  $B$  a cero, podemos obtener los coeficientes de una representación a partir de los de la otra.

#### 5.4.1. Los procesos AR y MA y el proceso general

Es inmediato comprobar que un proceso MA es un caso particular de la representación de Wold. También lo son los procesos AR. Por ejemplo, el proceso AR(1):

$$(1 - \phi B) \tilde{z}_t = a_t, \quad (5.16)$$

puede escribirse, como vimos en el capítulo 4, multiplicando por el operador inverso  $(1 - \phi B)^{-1}$ :

$$\tilde{z}_t = (1 + \phi B + \phi^2 B^2 + \dots) a_t,$$

que representa al proceso AR(1) como un caso particular de la forma MA( $\infty$ ) del proceso lineal general, con coeficientes  $\psi_i$  que decaen en progresión geométrica. La condición de estacionaridad y varianza finita, serie de coeficientes  $\psi_i^2$  convergente, equivale ahora a que  $|\phi| < 1$ .

Para procesos AR de orden más alto resulta más cómodo obtener los coeficientes de la representación MA( $\infty$ ) imponiendo la condición de que, al ser resultado de invertir un proceso AR, el producto de ambos

## 5. Procesos de media móvil y ARMA

---

operadores debe ser la unidad. Por ejemplo, para un AR(2) la condición será:

$$(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1.$$

Imponiendo la anulación de las potencias de  $B$ , se obtienen los coeficientes:

$$\begin{aligned}\psi_1 &= \phi_1 \\ \psi_2 &= \phi_1\psi_1 + \phi_2 \\ \psi_i &= \phi_1\psi_{i-1} + \phi_2\psi_{i-2}, \quad i \geq 2\end{aligned}$$

donde  $\psi_0 = 1$ . Análogamente, para un AR( $p$ ) los coeficientes  $\psi_i$  de la representación general se calculan mediante:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$

y para  $i \geq p$  deben verificar la condición:

$$\psi_i = \phi_1\psi_{i-1} + \dots + \phi_p\psi_{i-p}, \quad i \geq p.$$

La condición de estacionaridad implica que las raíces de la ecuación característica del proceso AR( $p$ ),  $\phi_p(B) = 0$ , deben estar fuera del círculo unidad. Escribiendo el operador  $\phi_p(B)$  como:

$$\phi_p(B) = \prod_{i=1}^p (1 - G_i B),$$

donde  $G_i^{-1}$  son las raíces de  $\phi_p(B) = 0$ , se verifica que, desarrollando en fracciones parciales:

$$\phi_p^{-1}(B) = \sum \frac{k_i}{(1 - G_i B)},$$

será convergente si  $|G_i| < 1$ .

En resumen, los procesos AR pueden considerarse casos particulares de la representación del proceso lineal general que se caracterizan porque: (1) todos los  $\psi_i$  son distintos de cero; (2) existen restricciones sobre las  $\psi_i$ , que dependen del orden del proceso. En general, verifican la secuencia  $\psi_i = \phi_1\psi_{i-1} + \dots + \phi_p\psi_{i-p}$ , con condiciones iniciales que dependen del orden del proceso.

Esta relación entre los coeficientes  $\psi_i$  y los del proceso permiten concluir que los coeficientes  $\psi_i$  tienen la misma estructura que los coeficientes de autocorrelación simple. En efecto, para un AR( $p$ ) hemos visto que satisfacen la ecuación de un proceso, como ocurría con las autocorrelaciones, y para un MA( $q$ ) el número de coeficientes de autocorrelación no nulos es el orden del proceso.

### 5.5. El proceso ARMA(1, 1)

Una conclusión de la sección anterior es que los procesos AR y MA aproximan un proceso lineal general MA( $\infty$ ) desde puntos de vista complementarios: los AR permiten estructura MA( $\infty$ ), pero imponen restricciones sobre las pautas de decrecimiento de los coeficientes  $\psi_i$ ; los MA exigen un número de términos finito, pero a cambio no imponen restricciones sobre sus valores. Desde el punto de vista de la estructura de autocorrelación, los AR permiten muchos coeficientes distintos de cero, pero con una pauta de decrecimiento fija, mientras que los MA permiten unos pocos coeficientes distintos de cero, con valores arbitrarios. Los procesos ARMA intentan combinar estas propiedades y permiten representar de forma *escueta* (utilizando pocos parámetros) procesos cuyos primeros  $q$  coeficientes son cualesquiera, mientras que los siguientes decrecen según leyes simples.

Matemáticamente, los procesos ARMA resultan de añadir estructura MA a un proceso AR o viceversa. El proceso más simple es el ARMA(1,1), que se escribe:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t - \theta_1 a_{t-1},$$

o, con notación de operadores:

$$(1 - \phi_1 B) \tilde{z}_t = (1 - \theta_1 B) a_t, \quad (5.17)$$

donde  $|\phi_1| < 1$  para que el proceso sea estacionario, y  $|\theta_1| < 1$  para que sea invertible. Además supondremos que  $\phi_1 \neq \theta_1$ . Si ambos parámetros fuesen idénticos tendríamos que, multiplicando ambos miembros por el operador  $(1 - \phi_1 B)^{-1}$ , resulta que  $\tilde{z}_t = a_t$ , y el proceso sería ruido blanco. En la formulación de los modelos ARMA supondremos siempre que no hay raíces comunes en los operadores AR y MA.

Para obtener la función de autocorrelación de un ARMA(1,1), multiplicando (5.17) por  $\tilde{z}_{t-k}$  y tomando esperanzas, se obtiene:

$$\gamma_k = \phi_1 \gamma_{k-1} + E(a_t \tilde{z}_{t-k}) - \theta_1 E(a_{t-1} \tilde{z}_{t-k}). \quad (5.18)$$

Para  $k > 1$ , el ruido  $a_t$  está incorrelado con la historia de la serie. Para comprobarlo, basta expresar el proceso como un MA( $\infty$ ) invirtiendo la parte AR:

$$\tilde{z}_t = (1 - \phi_1 B)^{-1} (1 - \theta_1 B) a_t = a_t + (\phi_1 - \theta_1) a_{t-1} + \phi_1 (\phi_1 - \theta_1) a_{t-2} + \dots$$

y al ser  $a_t$  ruido blanco estará incorrelado con los valores  $a_{t-i}$  y, por tanto, con  $\tilde{z}_{t-i}$ . En consecuencia:

$$\gamma_k = \phi_1 \gamma_{k-1}, \quad k > 1. \quad (5.19)$$

## 5. Procesos de media móvil y ARMA

---

Para  $k = 0$ ,  $E[a_t \tilde{z}_t] = \sigma^2$ , y:

$$E[a_{t-1} \tilde{z}_t] = E[a_{t-1} (\phi_1 \tilde{z}_{t-1} + a_t - \theta_1 a_{t-1})] = \sigma^2(\phi_1 - \theta_1),$$

sustituyendo estos resultados en (5.18), para  $k = 0$

$$\gamma_0 = \phi\gamma_1 + \sigma^2 - \theta_1\sigma^2(\phi_1 - \theta_1). \quad (5.20)$$

Tomando en (5.18)  $k = 1$ , resulta que  $E[a_t \tilde{z}_{t-1}] = 0$ ,  $E[a_{t-1} \tilde{z}_{t-1}] = \sigma^2$  y:

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma^2, \quad (5.21)$$

resolviendo (5.20) y (5.21) se obtiene:

$$\gamma_0 = \sigma^2 \frac{1 - 2\phi_1\theta_1 + \theta_1^2}{1 - \phi_1^2}.$$

Para obtener el primer coeficiente de autocorrelación, dividiremos (5.21) por esta expresión:

$$\rho_1 = \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - 2\phi_1\theta_1 + \theta_1^2}. \quad (5.22)$$

Observemos que si  $\phi_1 = \theta_1$ , esta autocorrelación es cero, porque, como indicamos antes, los operadores  $(1 - \phi_1 B)$  y  $(1 - \theta_1 B)$  se cancelan y resultará un proceso de ruido blanco. En el caso frecuente en que ambos coeficientes son positivos y  $\phi_1 > \theta_1$  es fácil comprobar que la correlación aumenta con  $(\phi_1 - \theta_1)$ .

El resto de los coeficientes de autocorrelación se obtienen dividiendo (5.19) por  $\gamma_0$ , que resulta en:

$$\rho_k = \phi_1 \rho_{k-1} \quad k > 1 \quad (5.23)$$

que indica que a partir del primer coeficiente la *fas* de un ARMA(1,1) tiene un decrecimiento exponencial determinado por el parámetro  $\phi_1$  de la parte AR. La diferencia con un AR(1) es que este decrecimiento comienza a partir de  $\rho_1$ , y no de  $\rho_0 = 1$ , y este primer valor de la autocorrelación de primer orden depende de la diferencia relativa entre  $\phi_1$  y  $\theta_1$ . Observemos que si  $\phi_1 \approx 1$  y  $\phi_1 - \theta_1 = \varepsilon$ , pequeño, podemos tener muchos coeficientes distintos de cero, pero todos de pequeño tamaño.

Para calcular la *fap*, escribiremos el ARMA(1, 1) en la forma AR( $\infty$ ):

$$(1 - \theta_1 B)^{-1} (1 - \phi_1 B) \tilde{z}_t = a_t,$$

y utilizando que  $(1 - \theta_1 B)^{-1} = 1 + \theta_1 B + \theta_1^2 B^2 + \dots$ , y operando, se obtiene:

$$\tilde{z}_t = (\phi_1 - \theta_1) \tilde{z}_{t-1} + \theta_1 (\phi_1 - \theta_1) \tilde{z}_{t-2} + \theta_1^2 (\phi_1 - \theta_1) \tilde{z}_{t-3} + \dots + a_t.$$

El efecto directo de  $\tilde{z}_{t-k}$  sobre  $\tilde{z}_t$  decrece geométricamente con  $\theta_1^k$  y la *fap* tendrá un decrecimiento geométrico a partir de un valor inicial.

En conclusión, en un proceso ARMA(1,1) la *fas* y la *fap* tienen una estructura similar: un valor inicial, cuya magnitud depende de  $\phi_1 - \theta_1$ , seguido de decrecimiento geométrico. La tasa de decrecimiento en la *fas* depende de  $\phi_1$ , mientras que en la *fap* depende de  $\theta_1$ . La figura 5.7 presenta las funciones de autocorrelación simple y parcial para distintos procesos ARMA (1,1).

## 5.6. Procesos ARMA( $p, q$ )

El proceso ARMA( $p, q$ ) será:

$$(1 - \phi_1 B - \dots - \phi_p B^p) \tilde{z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (5.24)$$

o, en notación compacta,

$$\phi_p(B) \tilde{z}_t = \theta_q(B) a_t.$$

El proceso será estacionario si las raíces de  $\phi_p(B) = 0$  están fuera del círculo unidad, e invertible si lo están las de  $\theta_q(B) = 0$ . Supondremos, además, que no hay raíces comunes que puedan cancelarse entre los operadores AR y MA.

Para obtener los coeficientes  $\psi_i$  de la representación general del modelo MA( $\infty$ ) escribiremos:

$$\tilde{z}_t = \phi_p(B)^{-1} \theta_q(B) a_t = \psi(B) a_t$$

e igualaremos las potencias de  $B$  en  $\psi(B) \phi_p(B)$  a las de  $\theta_q(B)$ . Análogamente, podemos representar un ARMA( $p, q$ ) como un modelo AR( $\infty$ ) haciendo:

$$\theta_q^{-1}(B) \phi_p(B) \tilde{z}_t = \pi(B) \tilde{z}_t = a_t$$

y los coeficientes  $\pi_i$  resultarán de  $\phi_p(B) = \theta_q(B) \pi(B)$ .

Para calcular las autocovarianzas, multiplicando (5.24) por  $\tilde{z}_{t-k}$  y tomando esperanzas:

$$\begin{aligned} \gamma_k - \phi_1 \gamma_{k-1} - \dots - \phi_p \gamma_{k-p} &= \\ = E[a_t \tilde{z}_{t-k}] - \theta_1 E[a_{t-1} \tilde{z}_{t-k}] - \dots - \theta_q E[a_{t-q} \tilde{z}_{t-k}] \end{aligned}$$

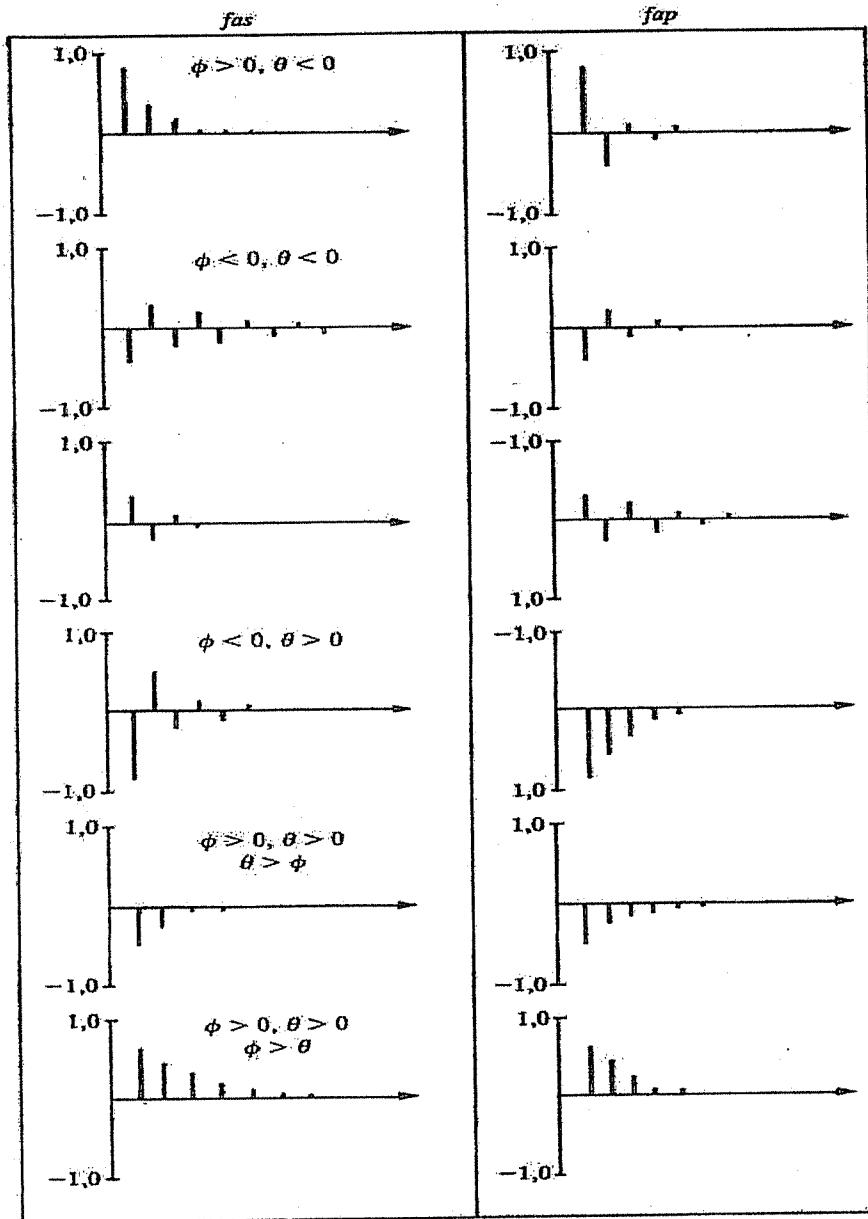
para  $k > q$  todos los términos de la derecha se anulan, y dividiendo por  $\gamma_0$ :

$$\rho_k - \phi_1 \rho_{k-1} - \dots - \phi_p \rho_{k-p} = 0,$$

es decir:

$$\phi_p(B) \rho_k = 0 \quad k > q, \quad (5.25)$$

Figura 5.7 Funciones de autocorrelación simple y parcial para procesos ARMA(1,1)



**Tabla 5.1 Resumen de la fas y fap de procesos ARMA**

	fas	fap
AR(p)	Muchos coeficientes no nulos	Primeros p no nulos, resto 0
MA(q)	Primeros q no nulos, resto 0	Muchos coeficientes no nulos
ARMA(p,q)	Muchos coeficientes no nulos	Muchos coeficientes no nulos

y concluimos que los coeficientes de autocorrelación para  $k > q$  seguirán un decrecimiento determinado únicamente por la parte autorregresiva. Los primeros  $q$  coeficientes dependen de los parámetros MA y AR y de ellos  $p$  proporcionarán los valores iniciales para el decrecimiento posterior (para  $k > q$ ) según (5.25). Por tanto, si  $p > q$  toda la *fas* mostrará un decrecimiento dictado por (5.25). En resumen, la *fas*:

1. tendrá  $q - p + 1$  valores iniciales con una estructura que depende de los parámetros AR y MA;
2. decrecerá a partir del coeficiente  $q - p$  como una mezcla de exponentiales y sinusoides, determinada exclusivamente por la parte autorregresiva.

Puede comprobarse que la *fap* tendrá una estructura similar.

### Conclusión

La *fas* y *fap* de los procesos ARMA es el resultado de la superposición de sus propiedades AR y MA: en la *fas* ciertos coeficientes iniciales que dependen del orden de la parte MA y después un decrecimiento dictado por la parte AR. En la *fap* valores iniciales dependientes del orden del AR seguidos del decrecimiento debido a la parte MA. Esta estructura compleja hace que el orden de un proceso ARMA sea difícil de identificar en la práctica.

La tabla 5.1 resume estas características.

## 5.7. Los procesos ARMA y la suma de procesos estacionarios

Una razón que explica por qué los procesos ARMA son frecuentes en la práctica es que al sumar procesos AR resulta un proceso ARMA. Para ilustrar esta idea, consideremos el caso más simple donde sumamos a un proceso AR(1) un ruido blanco. Sea

$$z_t = y_t + v_t, \quad (5.26)$$

## 5. Procesos de media móvil y ARMA

---

donde  $y_t = \phi y_{t-1} + a_t$  sigue un proceso AR(1) de media cero y  $v_t$  es un ruido blanco independiente de  $a_t$ , y, por tanto, de  $y_t$ . El proceso  $z_t$  puede interpretarse como el resultado de observar un proceso AR(1) con cierto error de medida. La varianza de este proceso suma será:

$$\gamma_z(0) = E(z_t^2) = E[(y_t^2 + v_t^2 + 2y_t v_t)] = \gamma_y(0) + \sigma_v^2, \quad (5.27)$$

como los sumandos son independientes, la varianza es la suma de las varianzas de los componentes. Para calcular las autocovarianzas tendremos en cuenta que las del proceso  $y_t$  verifican  $\gamma_y(k) = \phi^k \gamma_y(0)$  y las del proceso  $v_t$  son nulas. Entonces, para  $k \geq 1$ :

$$\gamma_z(k) = E(z_t z_{t-k}) = E[(y_t + v_t)(y_{t-k} + v_{t-k})] = \gamma_y(k) = \phi^k \gamma_y(0),$$

ya que, por la independencia de los componentes,  $E[y_t v_{t-k}] = 0$  para cualquier  $k$  y al ser  $v_t$  ruido blanco  $E[v_t v_{t-k}] = 0$ . En particular, sustituyendo en la ecuación anterior para  $k = 1$  la varianza  $\gamma_y(0)$  por su expresión (5.27), se obtiene:

$$\gamma_z(1) = \phi \gamma_z(0) - \phi \sigma_v^2, \quad (5.28)$$

mientras que para  $k \geq 2$

$$\gamma_z(k) = \phi \gamma_z(k-1). \quad (5.29)$$

Si comparamos la ecuación (5.28) con la (5.21) y la (5.29) con la (5.19), concluimos que el proceso  $z_t$  sigue un modelo ARMA(1,1) con parámetro AR igual a  $\phi$ . El parámetro  $\theta$  y la varianza de las innovaciones del proceso ARMA(1,1) dependen de la relación entre las varianzas de los sumandos. En efecto, llamando  $\lambda = \sigma_v^2 / \gamma_z(0)$  a la proporción de varianza que supone el ruido blanco, según la ecuación (5.28), la primera autocorrelación es:

$$\rho_z(1) = \phi - \phi \lambda = \phi(1 - \lambda)$$

mientras que, por (5.29), las restantes verifican, para  $k \geq 2$ :

$$\rho_z(k) = \phi \rho_z(k-1).$$

Si  $\lambda$  es muy pequeño, lo que implica que la varianza relativa del ruido adicional o error de medida es pequeño, el proceso será muy próximo a un AR(1), y el parámetro  $\theta$  será muy pequeño. Si  $\lambda$  no es muy pequeño, tendremos el ARMA(1,1) y el valor de  $\theta$  depende de  $\lambda$  y de  $\phi$ . Si  $\lambda \rightarrow 1$ , de manera que el ruido blanco sea dominante, el parámetro  $\theta$  será igual al valor de  $\phi$  y tendremos un proceso de ruido blanco.

Los resultados anteriores se generalizan para cualquier proceso AR(p).

Puede comprobarse (véase Granger y Morris, 1976) que se verifica:

$$AR(p) + AR(0) = ARMA(p, p),$$

y que, además:

$$AR(p) + AR(q) = ARMA(p + q, \max(p, q))$$

Por ejemplo, si sumamos dos procesos AR(1) independientes, se obtiene un nuevo proceso ARMA(2,1).

La suma de procesos MA es simple: al sumar procesos MA independientes obtenemos nuevos procesos MA. Supongamos:

$$z_t = x_t + y_t,$$

donde  $x_t, y_t$  tienen media cero y siguen procesos MA(1) independientes con covarianzas  $\gamma_x(k), \gamma_y(k)$ , que son cero para  $k > 1$ . La varianza del proceso suma será:

$$\gamma_z(0) = \gamma_x(0) + \gamma_y(0), \quad (5.30)$$

y la autocovarianza de orden  $k$ :

$$E(z_t z_{t-k}) = \gamma_z(k) = E[(x_t + y_t)(x_{t-k} + y_{t-k})] = \gamma_x(k) + \gamma_y(k).$$

Por tanto, todas las covarianzas  $\gamma_z(k)$  de orden mayor que uno serán cero, al serlo  $\gamma_x(k)$  y  $\gamma_y(k)$ . Dividiendo esta ecuación por  $\gamma_z(0)$  y utilizando (5.30), se obtiene que las autocorrelaciones verificarán:

$$\rho_z(k) = \rho_x(k)\lambda + \rho_y(k)(1 - \lambda)$$

donde:

$$\lambda = \frac{\gamma_x(0)}{\gamma_x(0) + \gamma_y(0)},$$

es la varianza relativa del primer sumando. En el caso particular en que uno de los dos procesos sea ruido blanco se obtendrá un modelo MA(1) con autocorrelación menor que la del proceso original. De la misma forma, es fácil demostrar que:

$$MA(q_1) + MA(q_2) = MA(\max(q_1, q_2)).$$

Para procesos ARMA se demuestra también que:

$$ARMA(p_1, q_1) + ARMA(p_2, q_2) = ARMA(a, b)$$

donde:

$$a \leq p_1 + p_2, \quad b \leq \max(p_1 + q_1, p_2 + q_2).$$

Estos resultados sugieren que siempre que observemos procesos que sean suma de otros, y alguno de ellos tenga estructura AR, es esperable observar procesos ARMA. Este resultado es a primera vista sorprendente: como la mayoría de las series reales pueden considerarse como suma de componentes, la mayoría deberían ser ARMA. Sin embargo, muchas series reales se aproximan bien mediante series AR o MA. La explicación de esta paradoja es que un proceso ARMA( $p+h, p$ ) con  $p$  raíces similares en los operadores AR y MA puede aproximarse bien por un AR( $h$ ), por la casi cancelación de raíces en ambos miembros.

---

### Ejercicios 5

- 5.1 Dado el proceso de media cero  $z_t = (1 - .7B)a_t$ : (a) calcular la función de autocorrelación; (b) escribirlo como un proceso AR( $\infty$ ).
- 5.2 Demostrar que los procesos MA(1)  $z_t = a_t - 0.5a_{t-1}$  y  $z_t = a_t - 2a_{t-1}$  tienen la misma estructura de autocorrelación pero uno es invertible y otro no.
- 5.3 Demostrar que los dos procesos  $z_t = a_t + 0.5a_{t-1}$  y  $z_t = 0.5a_t + a_{t-1}$  son indistinguibles ya que tienen la misma varianza y la misma estructura de autocorrelación.
- 5.4 Dado el proceso MA(2)  $z_t = a_t - 1.2a_{t-1} + 0.35a_{t-2}$ : (a) comprobar si es invertible; (b) calcular su estructura de autocorrelación; (c) escribirlo como un proceso AR( $\infty$ ).
- 5.5 Dado el modelo  $z_t = 5 + 0.9z_{t-1} + a_t + 0.4a_{t-1}$ : (a) calcular su estructura de autocorrelación; (b) escribirle en la forma MA( $\infty$ ); (c) escribirle en el forma AR( $\infty$ ).
- 5.6 Dado el proceso  $(1 - B + 0.21B^2)z_t = a_t - 0.3a_{t-1}$ : (a) comprobar si es estacionario e invertible; (b) obtener la función de autocorrelación simple; (c) obtener su representación AR( $\infty$ ); (d) obtener su representación MA( $\infty$ ).
- 5.7 Obtener la función de autocorrelación simple de un proceso ARMA(1,1) escribiéndolo como un MA( $\infty$ ) y utilizando la fórmula (5.15). Comprobar que se obtiene (5.22) y (5.23).
- 5.8 Demostrar que si sumamos dos procesos MA(1) se obtiene un nuevo proceso MA(1) con parámetro MA que es una combinación lineal de los parámetros MA de los dos procesos, con pesos proporcionales a los cocientes entre las varianzas de las innovaciones de los sumando con relación a la varianza de la innovación del proceso suma.

### Apéndice 5.1: Función de correlación inversa

Dado un proceso ARMA,  $\phi_p(B)\tilde{z}_t = \theta_q(B)a_t$ , se denomina proceso inverso o dual al proceso ARMA  $\theta_q(B)\tilde{z}_t = \phi_p(B)a_t$ , resultado de intercambiar los operadores AR y MA. Por ejemplo, el proceso inverso del AR(1):  $(1 - .5B)z_t = a_t$  es el proceso MA(1):  $z_t = (1 - .5B)a_t$ . Se define la función de autocorrelación inversa de un proceso ARMA como la función de autocorrelación simple del proceso inverso. Por ejemplo, la función de autocorrelación inversa del proceso AR(1):  $(1 - .5B)z_t = a_t$  tiene un solo coeficiente de autocorrelación distinto de cero en el retardo uno con valor  $\rho_1^i = -.5/(1 + .5^2)$ . La función de autocorrelación inversa de un proceso AR( $p$ ) tiene, por tanto, los  $p$  primeros coeficientes distintos de cero y el resto cero, mientras que la de un MA tendrá todos los coeficientes distintos de cero, decreciendo con una pauta geométrica o sinusoidal con el retardo.

La función de autocorrelación inversa tiene pues la misma estructura que la función de autocorrelación parcial, pero no es idéntica a ella. Se demuestra que así como el coeficiente de autocorrelación parcial de orden  $k$  representa la correlación parcial entre  $z_t$  y  $z_{t+k}$  cuando eliminamos de ambas variables el efecto de las observaciones intermedias,  $z_{t+1}, \dots, z_{t+k-1}$ , el coeficiente de autocorrelación inversa de orden  $k$  cambiado de signo representa la correlación parcial entre  $z_t$  y  $z_{t+k}$  cuando eliminamos de ambas variables el efecto de todas las otras variables del proceso distintas de  $z_t$  y  $z_{t+k}$ . Es decir, el coeficiente de autocorrelación inversa de orden  $k$  cambiado de signo representa la relación que existe entre variables separadas por  $k$  retardos cuando eliminamos de ellas el efecto de todas las restantes observaciones, tanto las previas,  $z_1, \dots, z_{t-1}$ , como las intermedias,  $z_{t+1}, \dots, z_{t+k-1}$ , como las posteriores,  $z_{t+k+1}, \dots, z_T$ .

Por ejemplo, consideremos el proceso AR(1):  $(1 - \phi B)z_t = a_t$ . El coeficiente de autocorrelación parcial de orden uno es igual al coeficiente de autocorrelación simple de orden uno,  $\rho^p(1) = \rho(1) = \phi$ . Sin embargo, el coeficiente de autocorrelación inversa es el coeficiente de autocorrelación simple del proceso MA(1):  $z_t = (1 - \phi B)a_t$ , y será, por tanto,  $\rho^i(1) = -\phi/(1 + \phi^2)$ . El coeficiente de autocorrelación inversa cambiado de signo,  $\phi/(1 + \phi^2)$ , es menor que el parcial,  $\phi$ , como sería esperable. En efecto,  $z_{t-1}$  contiene información sobre  $z_t$  y también sobre  $z_{t+1}$  y lo mismo ocurre con  $z_{t+2}$ . Si eliminamos de la relación entre  $z_t$  y  $z_{t+1}$  la parte debida a la información existente en las observaciones adyacentes, la dependencia restante es menor. Para comprobarlo, vamos a eliminar de ambas observaciones el efecto de las observaciones previas. Como  $z_t = \phi z_{t-1} + a_t$ , si eliminamos de  $z_t$  el efecto de las observaciones anteriores al instante  $t$ , que por la propiedad markoviana equivale a eliminar el efecto de  $z_{t-1}$ , obtenemos que el resultado es  $a_t$ . Si hacemos lo mismo con  $z_{t+1}$ , como  $z_{t+1} = \phi^2 z_{t-1} + \phi a_t + a_{t+1}$ , al eliminar de  $z_{t+1}$  el efecto de  $z_{t-1}$  la parte que queda es  $\phi a_t + a_{t+1}$ . La correlación simple entre  $u_t = a_t$  y  $u_{t+1} = \phi a_t + a_{t+1}$  es  $\phi/(1 + \phi^2)$ , el coeficiente de autocorrelación inversa cambiado de signo. Puede comprobarse que éste es también el resultado que

## 5. Procesos de media móvil y ARMA

se obtiene si, además del efecto de las observaciones previas, eliminamos el efecto de las posteriores, que para un AR(1) equivale a eliminar el efecto de  $z_{t+2}$ . Es decir, si eliminamos de  $z_t$  el efecto de todas las observaciones salvo  $z_{t+1}$  mediante la regresión  $z_t = \sum_{i=1}^{\infty} \alpha_i z_{t-i} + \sum_{j=2}^{\infty} \beta_j z_{t+j} + u_t$ , y hacemos lo mismo con  $z_{t+1}$  con la regresión  $z_{t+1} = \sum_{i=1}^{\infty} \alpha_i^* z_{t-i} + \sum_{j=2}^{\infty} \beta_j^* z_{t+j} + v_{t+1}$ , el coeficiente de correlación simple entre  $u_t$  y  $v_{t+1}$  es  $\phi/(1+\phi^2)$ , el coeficiente de autocorrelación inversa cambiado de signo.

Puede comprobarse que para un AR( $p$ ) los coeficientes de autocorrelación inversa son menores que los parciales, como sería de esperar. La comprobación para el último coeficiente no nulo es inmediata: el coeficiente de autocorrelación parcial de un AR( $p$ ) es  $\phi_p$ , mientras que el inverso de orden  $p$  es, utilizando la fórmula para el coeficiente de autocorrelación simple de un MA( $p$ ):

$$\rho^i(p) = \frac{-\phi_p}{\sum_{i=0}^p \phi_i^2}$$

y  $|\rho^i(p)|$  será siempre menor que  $|\phi_p|$ .

Una forma alternativa de derivar el proceso inverso, que ilustra la representación matricial de un proceso ARMA, es la siguiente. Según el teorema de Wold, una muestra de tamaño  $T$  de un proceso estacionario, sin componentes deterministas y media cero, puede escribirse según la ecuación matricial:

$$\mathbf{z} = \Psi \mathbf{a}, \quad (5.31)$$

donde  $\mathbf{z} = (z_T, z_{T-1}, \dots, z_1)'$ ,  $\mathbf{a} = (a_T, a_{T-1}, \dots, a_1)'$ , y  $\Psi$  es la matriz triangular superior:

$$\Psi = \begin{bmatrix} 1 & \psi_1 & \psi_2 & \dots & \psi_{T-1} \\ 0 & 1 & \psi_1 & \dots & \psi_{T-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \psi_1 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

En esta representación se ha resuelto el problema de las innovaciones iniciales, anteriores al instante  $t = 1$ , suponiendo que éstas son iguales a sus esperanzas, cero. Como el proceso es estacionario, lo que requiere que los coeficientes  $\psi_i$  vayan a cero para  $i$  grande, esta hipótesis sólo afectará a las primeras observaciones del proceso, que tendrán una estructura diferente del resto. Por ejemplo, si el proceso es MA(1), afectará sólo a la primera observación ya que  $z_1 = a_1$ , y  $z_t = a_t - \theta_1 a_{t-1}$  para  $t \geq 2$ .

En el capítulo 3 se definió la matriz de covarianzas del proceso de orden  $T$ ,  $\Gamma_z$ , mediante

$$\Gamma_z = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{T-2} \\ \dots & \dots & \dots & \dots & \dots \\ \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 & \gamma_1 \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 \end{bmatrix} \quad (5.32)$$

y es una matriz de Toepliz, con términos iguales en las diagonales. Podemos aproximar esta matriz mediante la matriz:

$$\mathbf{V}_z = E(\mathbf{zz}') = E(\Psi \mathbf{a} \mathbf{a}' \Psi') = \Psi E(\mathbf{a} \mathbf{a}') \Psi' = \sigma^2 \Psi \Psi', \quad (5.33)$$

que es la de covarianzas pero con las condiciones iniciales anteriores. Por tanto, las matrices  $\Gamma_z$  y  $\mathbf{V}_z$  no son idénticas. Por ejemplo, supongamos un proceso MA(1). Entonces, como  $z_1 = a_1$ , y  $z_t = a_t - \theta_1 a_{t-1}$  para  $t \geq 2$  la varianza de  $z_1$  es  $\sigma^2$  mientras que la varianza de  $z_t$  es  $\sigma^2(1 + \theta^2)$  para  $t \geq 2$ . La matriz  $\Gamma_z$  del proceso MA(1) general tiene todos los términos diagonales iguales a  $\sigma^2(1 + \theta^2)$  mientras que la matriz  $\mathbf{V}_z$  de un proceso MA(1) que suponemos comienza en  $t = 1$  tiene el último término de la diagonal igual a  $\sigma^2$ , la varianza del primer término.

Dividiendo cada término de (5.33) por  $\gamma_0 = \sigma^2(\sum_{i=0}^{\infty} \psi_i^2)$ , resulta:

$$\mathbf{C}_z = \gamma_0^{-1} E(\mathbf{zz}') = \left( \sum_{i=0}^{\infty} \psi_i^2 \right)^{-1} \Psi \Psi'. \quad (5.34)$$

que será similar a la matriz de autocorrelación del proceso,  $\mathbf{R}_T$ , salvo por las condiciones iniciales.

Si invertimos la matriz  $\Psi$ , obtenemos la representación dual del proceso como un AR( $\infty$ ):

$$\Pi \mathbf{z} = \mathbf{a} \quad (5.35)$$

donde  $\Pi = \Psi^{-1}$  será una matriz triangular inferior con coeficientes:

$$\Pi = \begin{bmatrix} 1 & -\pi_1 & -\pi_2 & \dots & -\pi_{t-1} \\ 0 & 1 & -\pi_1 & \dots & -\pi_{t-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 & -\pi_1 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

En efecto, es fácil comprobar que  $\Pi \Psi = \mathbf{I}$ . Tenemos que:

$$\Pi \Psi = \begin{bmatrix} 1 & \psi_1 - \pi_1 & \psi_2 - \psi_1 \pi_1 - \pi_2 & \dots & \psi_{t-1} - \sum \psi_j \pi_i \\ 0 & 1 & \psi_1 - \pi_1 & \dots & \psi_{t-1} - \sum \psi_j \pi_i \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & & 1 & \psi_1 - \pi_1 \\ 0 & 0 & \dots & & 1 \end{bmatrix}$$

## 5. Procesos de media móvil y ARMA

---

y como  $\psi_k = \sum_{j=1}^k \psi_{k-j} \pi_j$ , comprobamos que la matriz  $\pi$  es la inversa de la  $\psi$ . La representación (5.35) se basa en las mismas hipótesis que la (5.31) y la distribución de los primeros valores es distinta de la del resto como consecuencia de las condiciones iniciales establecidas.

Vamos a definir el proceso inverso a dual de uno dado ARMA( $p, q$ ) como el proceso ARMA( $q, p$ ) obtenido al intercambiar los operadores AR por los MA y viceversa. En la representación MA( $\infty$ ) (5.31) el proceso dual o inverso se obtendrá colocando la matriz MA( $\infty$ ),  $\Psi$ , como matriz AR( $\infty$ ). La representación general del proceso dual es:

$$\Psi z^i = a$$

o invirtiendo este operador, el proceso inverso o dual del  $z = \Psi a$  es el

$$z^i = \Pi a.$$

La matriz de covarianzas con las condiciones iniciales generales anteriores aplicadas al proceso inverso,  $V_z^i$ , vendrá dada por:

$$V_z^i = \sigma^2 \Pi \Pi' = \sigma^2 \Psi^{-1} (\Psi')^{-1}$$

a los términos de esta matriz se les denomina covarianzas inversas. Análogamente, la función de autocorrelación del proceso inverso o dual será:

$$C_z^i = \left( \sum_{i=0}^{\infty} \pi_i^2 \right)^{-1} \Pi \Pi'.$$

---

## 6. Procesos integrados y de memoria larga



Norbert Wiener (1894-1964)

Matemático genial nacido en Estados Unidos de padres judíos rusos. Se doctoró en Harvard a los 18 años y tras ampliar estudios en Cambridge con Russell y en Göttingen con Hilbert fue profesor de matemáticas en el MIT en Boston desde 1919 hasta su jubilación en 1959. Descubrió el movimiento browniano y desarrolló la teoría de predicción de procesos estacionarios y el análisis espectral. Inventor de la cibernetica, sus aportaciones han sido muy influyentes en muchos campos científicos.

### 6.1. Introducción

En este capítulo comenzamos el estudio de los procesos no estacionarios. Un proceso puede ser no estacionario en la media, en la varianza, en las autocorrelaciones, o en otras características de la distribución de las variables. Cuando el nivel de la serie no es estable en el tiempo, pudiendo en particular tener tendencia creciente o decreciente, diremos que la serie es no estacionaria en la media. Cuando la variabilidad o las autocorrelaciones se modifican con el tiempo, diremos que la serie es no estacionaria en la varianza o en las autocorrelaciones. Finalmente, si la distribución de la variable en cada instante varía con el tiempo, diremos que la serie no es estacionaria en distribución. Los procesos no estacionarios más importantes son los procesos integrados, que tienen la propiedad fundamental de que al diferenciarlos se obtienen procesos estacionarios. En este capítulo presentaremos con detalle estos procesos. Una propiedad importante que diferencia a los procesos integrados de los estacionarios es la forma en que desaparece la dependencia con el tiempo. En los procesos

## 6. Procesos integrados y de memoria larga

estacionarios ARMA las autocorrelaciones disminuyen geométricamente, y se hacen prácticamente cero a los pocos retardos. En los procesos integrados las autocorrelaciones disminuyen linealmente con el tiempo y es posible encontrar coeficientes de autocorrelación distintos de cero hasta para retardos muy altos.

Existe una clase de procesos estacionarios donde las autocorrelaciones disminuyen mucho más lentamente con el tiempo que en el caso de los procesos ARMA, y también que en los procesos integrados. Estos procesos se conocen como procesos de memoria larga, y en este capítulo revisaremos también brevemente sus propiedades. Veremos que estos procesos son casos particulares de la representación de Wold para procesos estacionarios, pero donde los coeficientes decrecen muy lentamente, lo que implica que las autocorrelaciones entre las observaciones también decrecen muy lentamente. Estos procesos, además de su interés teórico, pueden en la práctica aproximar bien el comportamiento observado en series largas climatológicas y financieras.

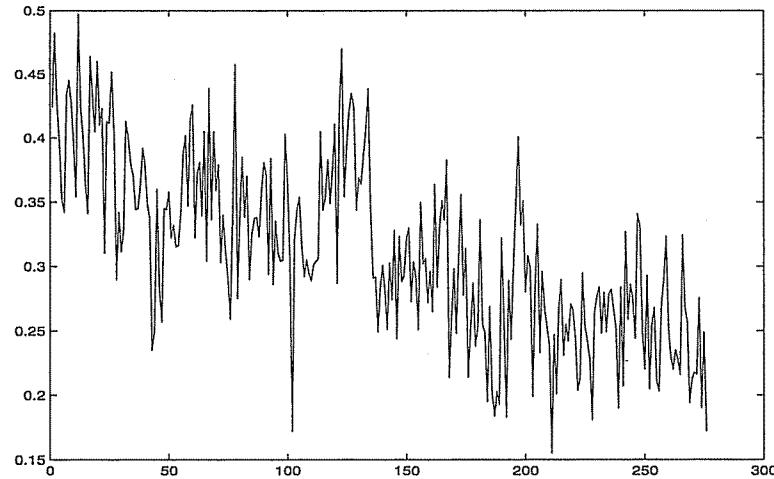
### 6.2. Procesos integrados

La mayoría de las series reales no son estacionarias, y su nivel medio varía con el tiempo. Por ejemplo, en el capítulo 1 vimos varios ejemplos de series no estacionarias. Una de estas series, la cuota semanal de mercado del dentífrico Colgate, se presenta en la figura 6.1. La serie, que llamaremos  $z_t$ , tiene una clara tendencia decreciente y por lo tanto no es estacionaria.

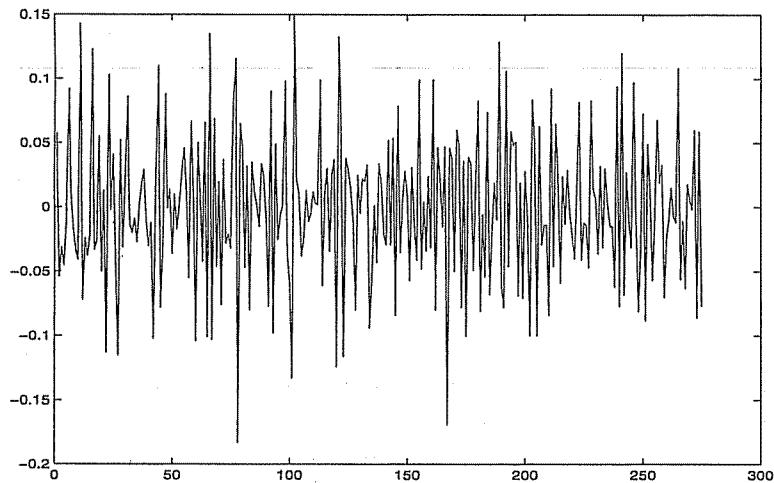
La figura 6.2 presenta la primera diferencia de esta serie, es decir, la serie de variaciones de cuota de mercado entre una semana y la siguiente. Si llamamos  $w_t = \nabla z_t$  a esta nueva serie, vemos que sus valores oscilan alrededor de una media constante y parecen corresponder a una serie estacionaria. Concluimos que la serie  $z_t$  de cuota de mercado de Colgate parece una serie integrada, que se transforma en estacionaria mediante una diferencia. Diremos entonces que la serie es integrada de orden uno, siendo el orden de integración el número de diferencias necesarias para obtener un proceso estacionario.

Frecuentemente las series económicas no son estacionarias pero sus diferencias relativas, o las diferencias cuando medimos la variable en logaritmos, son estacionarias. Por ejemplo, la figura 6.3 presenta otra serie no estacionaria que es el logaritmo de la serie del índice general en la Bolsa de Madrid, cuyos datos se encuentran en el fichero indicebolsa98-03.dat. La serie no es estacionaria en la media ya que su nivel varía a lo largo del tiempo. Si tomamos una diferencia en esta serie, se obtiene la serie de rendimientos en la Bolsa de Madrid con el índice general. En

**Figura 6.1 Cuota semanal de mercado de Colgate en el periodo enero de 1958 a abril de 1963**



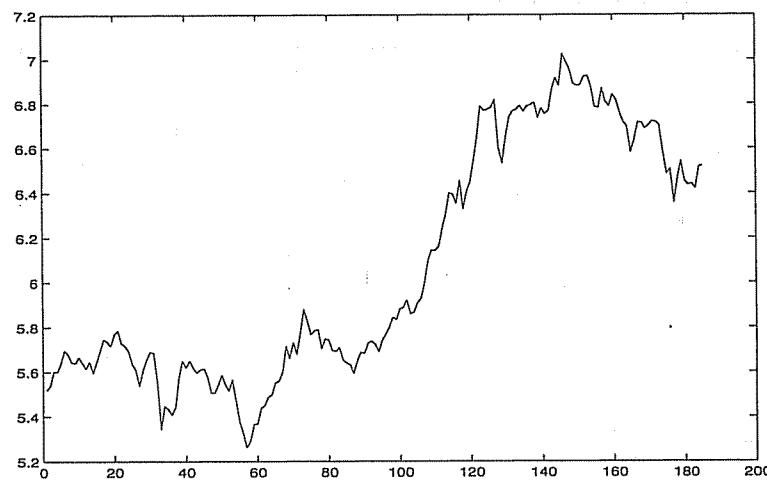
**Figura 6.2 Primera diferencia de la serie de cuota de mercado de Colgate**



efecto, si  $z_t$  es la serie del índice general y  $\log z_t$  la presentada en la figura 6.3,  $\nabla \log z_t$  es:

$$\nabla \log z_t = \log \frac{z_t}{z_{t-1}} = \log(1 + \frac{z_t - z_{t-1}}{z_{t-1}}) \approx \frac{\nabla z_t}{z_{t-1}}$$

**Figura 6.3 Logaritmo del índice general de la Bolsa de Madrid, enero de 1988 a mayo de 2003**



y comprobamos que la primera diferencia del logaritmo de una variable es aproximadamente igual a su crecimiento relativo. Cuando la variable original,  $z_t$ , es una serie de precios de un activo financiero, la serie  $\nabla \log z_t$  se define como los rendimientos de este activo. La figura 6.4 muestra que estos rendimientos mensuales son estables en el tiempo.

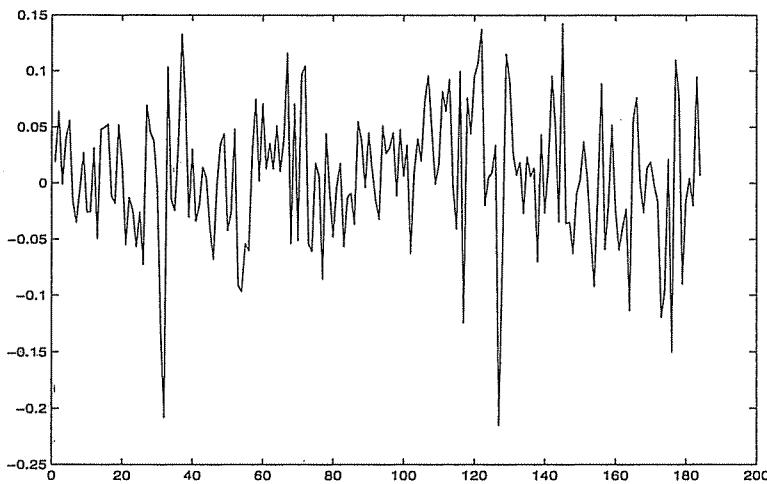
Estos dos ejemplos muestran procesos integrados de orden uno, es decir, procesos no estacionarios pero cuyos crecimientos, o primeras diferencias, son estacionarios.

A veces es necesario diferenciar más de una vez una serie para obtener un proceso estacionario. Por ejemplo, consideremos la serie de la figura 6.5. Esta serie es no estacionaria y su primera diferencia, indicada en la figura 6.6, tampoco lo es, ya que muestra una clara tendencia decreciente. Sin embargo, la diferencia de esta última serie, que corresponde a la segunda diferencia de la original y se muestra en la figura 6.7, sí tiene un nivel estable. Si llamamos  $z_t$  a la serie original, sus variaciones (o crecimientos) medidos por  $w_t = \nabla z_t$  no son estacionarias, pero las variaciones de los crecimientos, medidos por  $\nabla^2 z_t$ , sí lo son. Es decir, la serie:

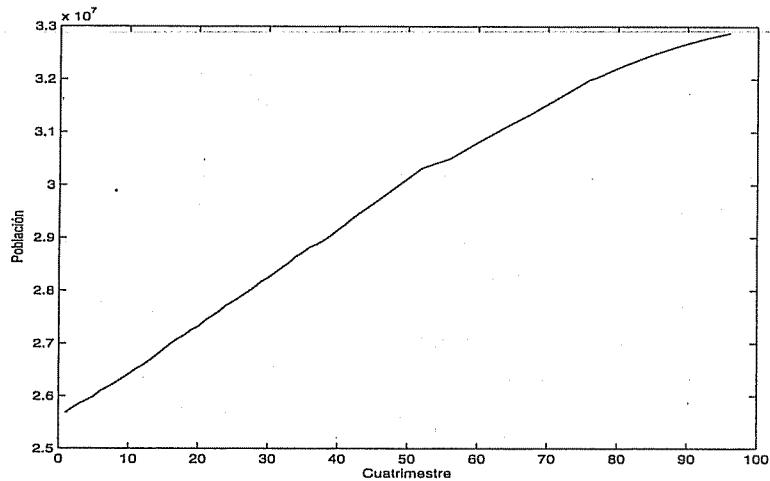
$$y_t = \nabla w_t = w_t - w_{t-1} = z_t - 2z_{t-1} + z_{t-2} = \nabla^2 z_t \quad (6.1)$$

que representa los crecimientos de la serie  $w_t$  será estacionaria. Esta serie  $y_t$  se denomina proceso «segunda diferencia» de la serie original,  $z_t$ , y diremos que  $z_t$  es integrada de orden dos, o con incrementos de segundo orden estacionarios.

**Figura 6.4 Rendimientos mensuales obtenidos en la Bolsa de Madrid de acuerdo al índice general en el periodo enero de 1988 a mayo de 2003**



**Figura 6.5 Población mayor de 16 años en España entre el primer cuatrimestre de 1977 al cuarto de 2000**



Generalizando, diremos que un proceso es *integrado de orden  $h \geq 0$* , y lo representaremos por  $I(h)$ , cuando al diferenciarlo  $h$  veces se obtiene un proceso estacionario. Un proceso estacionario es, por tanto, siempre  $I(0)$ . En la práctica la mayoría de las series no estacionarias que son

## 6. Procesos integrados y de memoria larga

Figura 6.6 Primera diferencia de la serie de población mayor de 16 años

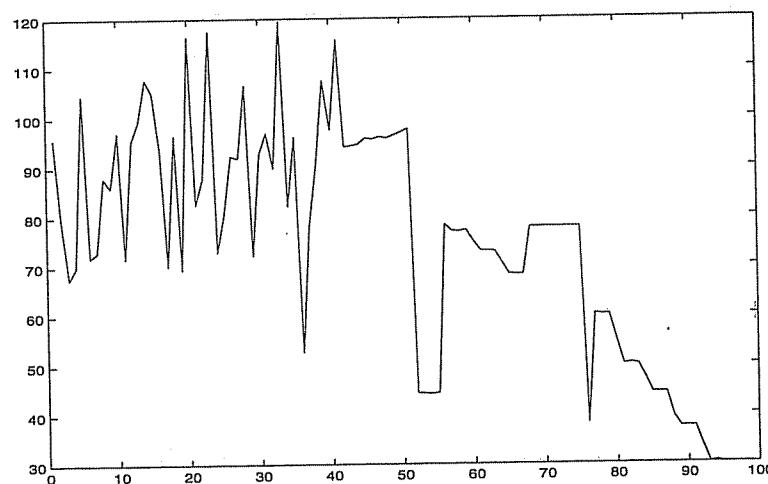
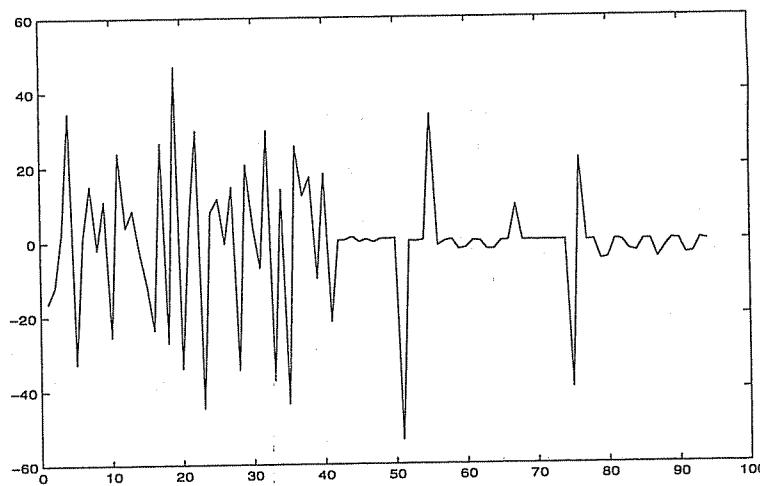
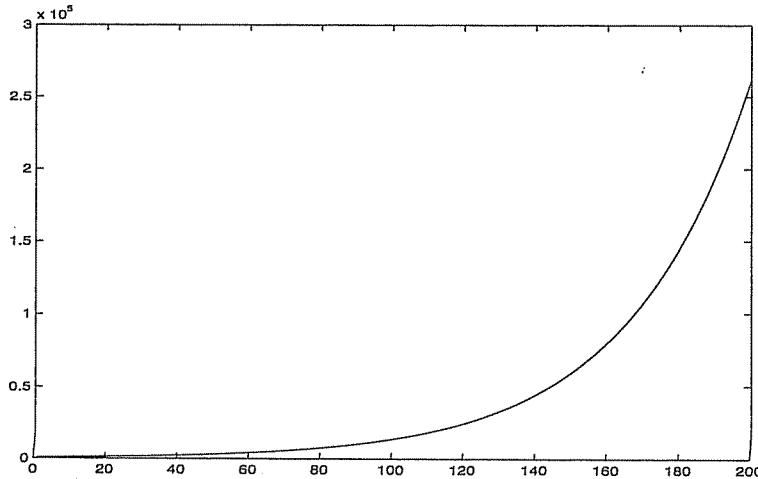


Figura 6.7 Segunda diferencia de la serie de la población mayor de 16 años en España



integradas tienen un orden  $h \leq 3$ .

Figura 6.8 Simulación de un AR con parámetro 1.03



### 6.3. El paseo aleatorio

Hemos visto que los procesos MA finitos son siempre estacionarios y que los AR sólo lo son si las raíces de  $\phi(B) = 0$  están fuera del círculo unidad. Consideremos el AR(1):

$$z_t = c + \phi z_{t-1} + a_t. \quad (6.2)$$

Si  $|\phi| < 1$  el proceso es estacionario. Si  $|\phi| > 1$  es fácil comprobar que se obtiene un proceso explosivo, donde los valores de la variable crecen sin límite hacia infinito. Por ejemplo, la figura 6.8 presenta una simulación de este proceso con  $c = 0$  y  $\phi = 1.03$ . Como los procesos explosivos no son frecuentes en la práctica, valores del parámetro AR mayores que la unidad no suelen ser útiles para representar series reales. Un caso interesante es cuando  $|\phi| = 1$ . Entonces, el proceso no es estacionario, pero tampoco es explosivo, y pertenece a la clase de procesos integrados de orden uno. En efecto, es inmediato que su primera diferencia:

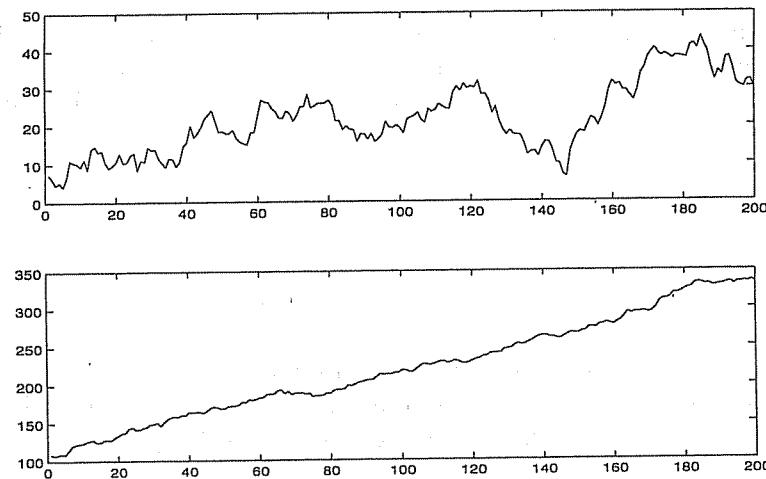
$$w_t = \nabla z_t = c + a_t,$$

sí es un proceso estacionario. Este proceso se denomina paseo aleatorio, y fue utilizado como ejemplo de proceso de Markov y de martingala en el capítulo 3.

Una característica importante que diferencia los procesos estacionarios y los no estacionarios es el papel de las constantes. En un proceso

## 6. Procesos integrados y de memoria larga

Figura 6.9 Dos realizaciones de un paseo aleatorio. La primera simulación se ha hecho sin constante y la segunda con constante  $c=1$



estacionario la constante no es importante, y en los capítulos anteriores hemos restado a las observaciones su media y trabajado con procesos de media cero. La forma del proceso y sus propiedades básicas son las mismas si tiene media cero o distinta de cero. Sin embargo, en un proceso no estacionario las constantes, si existen, son muy importantes y representan alguna propiedad permanente del proceso. Por ejemplo, la figura 6.9 presenta dos simulaciones del modelo (6.2). En la primera  $c = 0$ , se dice que el proceso no tiene deriva, y el nivel de la serie oscila en el tiempo. En la segunda  $c = 1$ , se dice que el proceso tiene una deriva igual a la unidad, y el proceso muestra una tendencia lineal de pendiente  $c$ . En efecto, el nivel de la serie es el del periodo anterior más  $c$ , lo que supone un crecimiento lineal determinista. Vemos que el gráfico de la serie es totalmente distinto en uno y otro caso.

Para calcular la función de autocorrelación de este proceso supondremos que comienza en  $t = 0$ . Entonces, sustituyendo sucesivamente  $z_t$  por  $z_{t-1}$  tendremos:

$$z_t = ct + a_t + a_{t-1} + a_{t-2} + \dots + a_1,$$

y tomando esperanzas:

$$E(z_t) = ct.$$

El proceso tiene una media que aumenta linealmente con el tiempo. Observemos que cuando no tiene deriva,  $c = 0$ , la media sí es constante e

igual a cero. Su varianza será:

$$Var(z_t) = E(a_t + a_{t-1} + a_{t-2} + \dots + a_1)^2 = \sigma^2 t \quad (6.3)$$

y aumenta con el tiempo y tiende a infinito con  $t$ . Esta propiedad indica que al aumentar el tiempo aumenta la incertidumbre sobre la situación del proceso. Para calcular las autocovarianzas utilizaremos la expresión del proceso para  $t+k$ :

$$z_{t+k} = c(t+k) + a_{t+k} + \dots + a_t + \dots + a_1,$$

con lo que se obtiene:

$$Cov(t, t+k) = E[(z_t - ct)(z_{t+k} - c(t+k))] = \sigma^2 t. \quad (6.4)$$

Observemos que las autocovarianzas también aumentan con el tiempo y no son sólo función del retardo, como en los procesos estacionarios, sino que dependen de los instantes en que se calculan. En particular:

$$Cov(t, t-k) = \sigma^2(t-k) \neq Cov(t, t+k).$$

La función de autocorrelación se obtiene dividiendo (6.4) por las desviaciones típicas de las variables que se obtienen de la expresión (6.3), con el resultado:

$$\rho(t, t+k) = \frac{t}{\sqrt{t(t+k)}} = (1 + \frac{k}{t})^{-1/2}. \quad (6.5)$$

Esta expresión indica que si  $t$  es grande, los coeficientes de la función de autocorrelación serán próximos a uno y *decrecerán aproximadamente de forma lineal con k*. En efecto, si suponemos que el proceso comenzó en un lejano pasado, de manera que  $k/t$  es pequeño, entonces la función  $(1 + k/t)^{-1/2}$  puede aproximarse con un desarrollo de Taylor de primer orden como:

$$\rho(t, t+k) \simeq 1 - \frac{k}{2t}. \quad (6.6)$$

Esta ecuación indica que suponiendo  $t$  fijo, si representamos los coeficientes de autocorrelación como función de  $k$  obtendremos una línea recta con pendiente  $(-1/2t)$ .

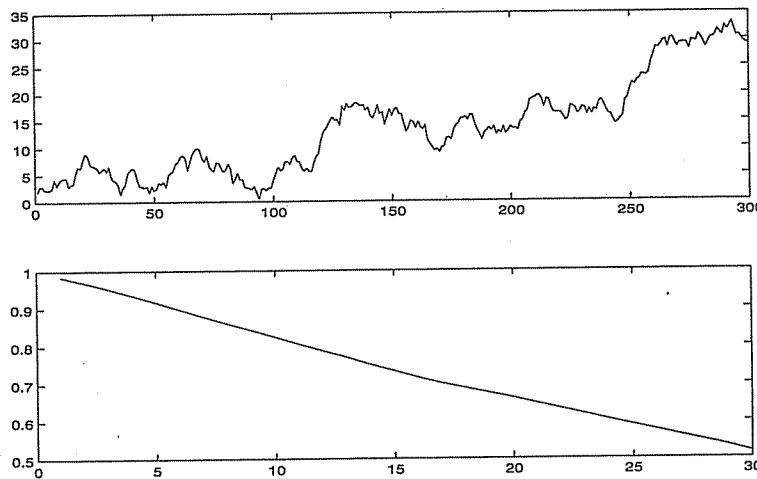
Tomando logaritmos en (6.5) tenemos que

$$\log \rho(t, t+k) = -\frac{1}{2} \log(1 + \frac{k}{t}) \approx -\frac{k}{2t}$$

y vemos que las correlaciones en logaritmos también decrecerán lineal-

## 6. Procesos integrados y de memoria larga

Figura 6.10 Generación de 300 observaciones que siguen un paseo aleatorio (gráfico superior) y su fas calculada en esta muestra (gráfico inferior)



mente con el retardo  $k$ .

Observemos que este comportamiento de la función de autocorrelación de un proceso integrado puede considerarse como un caso límite de la de un AR(1) cuando  $\phi$  se aproxima a la unidad. En efecto si  $\phi = 1 - \epsilon$ , con  $\epsilon$  pequeño, tenemos que

$$\rho_k = (1 - \epsilon)^k \approx 1 - \epsilon k$$

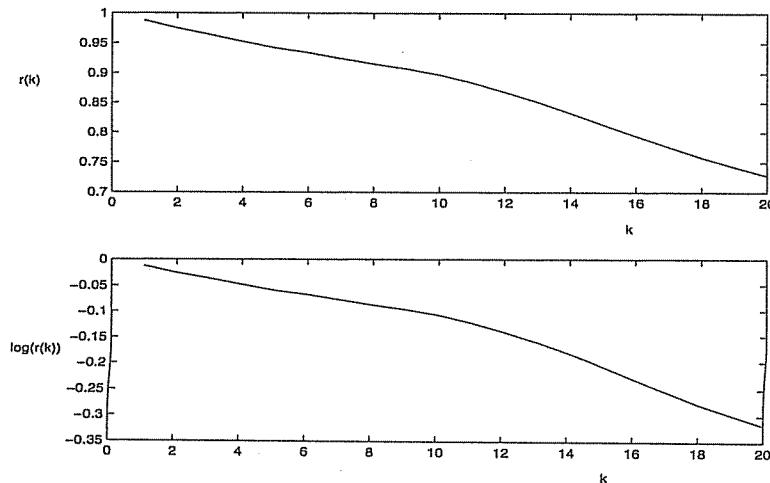
para  $\epsilon$  pequeño.

La figura 6.10 presenta un gráfico de la función de autocorrelación de una simulación de 300 observaciones que siguen un paseo aleatorio. Se observa claramente el decrecimiento lineal de la *fas*.

### Ejemplo 6.1

Vamos a calcular la función de autocorrelación muestral de la serie del fichero indicebolsa98-03.dat, que se representó en la figura 6.3. Esta serie es mensual y representa el logaritmo del índice general de la Bolsa de Madrid entre enero de 1988 y mayo de 2003. La figura 6.11 presenta su función de autocorrelación. Se observa que las correlaciones decaen linealmente, indicando un proceso no estacionario. Este comportamiento se observa tanto en las autocorrelaciones de la serie como en los logaritmos de las autocorrelaciones.

**Figura 6.11 Función de autocorrelación simple y su logaritmo como función del retardo para los datos de la Bolsa de Madrid, enero 1988 a mayo 2003**



#### 6.4. El proceso de alisado exponencial simple

Un proceso no estacionario, que se ha utilizado mucho para la previsión, y que puede considerarse una generalización del paseo aleatorio, es:

$$\nabla z_t = c + (1 - \theta B) a_t. \quad (6.7)$$

En este proceso, si hacemos  $\theta = 0$  obtenemos el paseo aleatorio. También puede verse como un caso límite del ARMA(1, 1) cuando el coeficiente autorregresivo es la unidad.

Suponiendo para simplificar  $c = 0$  y  $|\theta| < 1$  e invirtiendo la parte MA, este proceso se escribe:

$$z_t = (1 - \theta)z_{t-1} + \theta(1 - \theta)z_{t-2} + \theta^2(1 - \theta)z_{t-3} + \dots + a_t. \quad (6.8)$$

En el caso habitual donde  $0 < \theta < 1$ , todos los coeficientes de la representación AR( $\infty$ ) de la ecuación (6.8) son positivos y, además, suman la unidad. En efecto, estos coeficientes forman una progresión geométrica indefinida con primer término  $(1 - \theta)$  y razón  $\theta$ , con lo que su suma es 1. Por tanto, la ecuación (6.8) indica que el valor observado en cada instante es una media ponderada de los valores de la serie en instantes anteriores, con coeficientes que decrecen geométricamente con el retardo. Si  $\theta$  es próximo a la unidad, la media se calcula con muchos coeficientes,

## 6. Procesos integrados y de memoria larga

---

pero cada uno de ellos con poco peso. Por ejemplo, si  $\theta = .9$  los pesos de los retardos son  $.1, .09, .081, \dots$ , y los pesos decrecen muy lentamente y son similares para observaciones próximas. Sin embargo, si  $\theta$  es próximo a cero, la media se calcula ponderando sólo los últimos valores observados. Por ejemplo, si  $\theta = .1$  los pesos de los retardos son  $.9, .09, .009, \dots$ , y prácticamente sólo se tienen en cuenta los dos últimos retardos, con pesos muy distintos. Este proceso se conoce como proceso de alisado exponencial simple y fue introducido de forma intuitiva en el capítulo 2. Cuando  $c \neq 0$  el modelo está relacionado con el alisado doble de Holt, que también vimos en el capítulo 2.

Para calcular la *fas* del proceso supondremos que empezó en  $t = 0$ , con  $z_0 = a_0 = 0$  y  $c = 0$ . Entonces:

$$\begin{aligned} z_1 &= a_1 \\ z_2 &= z_1 + a_2 - \theta a_1 = a_2 + (1 - \theta) a_1 \\ z_3 &= z_2 + a_3 - \theta a_2 = a_3 + (1 - \theta) a_2 + (1 - \theta) a_1 \\ &\vdots \quad \vdots \\ z_t &= a_t + (1 - \theta) a_{t-1} + \dots + (1 - \theta) a_1. \end{aligned}$$

La media del proceso será cero, como en el paseo aleatorio cuando  $c = 0$ . Su varianza será:

$$Var(z_t) = \sigma^2(1 + (t - 1)(1 - \theta)^2),$$

y el proceso es no estacionario en la varianza, ya que esta función depende del tiempo. Las covarianzas se calculan como:

$$\begin{aligned} Cov(t, t + k) &= E(z_t z_{t+k}) = \\ &= E \left[ (a_t + (1 - \theta) \sum_{i=1}^{t-1} a_{t-i})(a_{t+k} + (1 - \theta) \sum_{i=1}^{t+k-1} a_{t+k-i}) \right] \end{aligned}$$

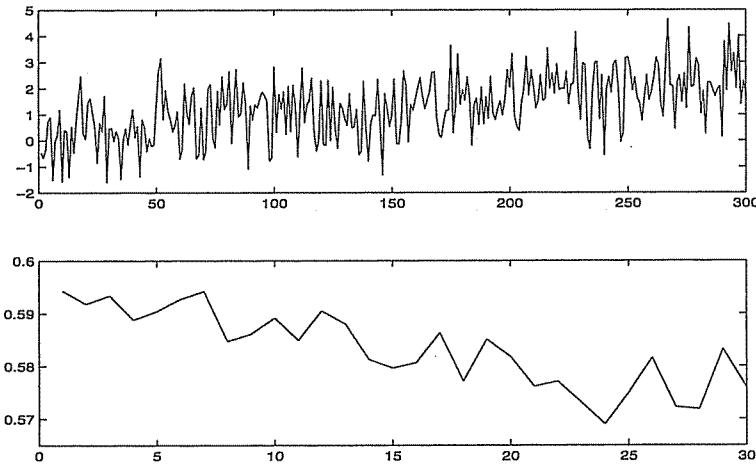
y, teniendo en cuenta que  $E(a_i a_j) = 0$ , si  $i \neq j$ , obtenemos que:

$$Cov(t, t + k) = \sigma^2(1 - \theta)(1 + (t - 1)(1 - \theta)).$$

Utilizando esta expresión y la de la varianza se obtiene la función de autocorrelación:

$$\rho(t, t + k) = \frac{(1 - \theta)(1 + (t - 1)(1 - \theta))}{\sqrt{(1 + (t - 1)(1 - \theta)^2)(1 + (t + k - 1)(1 - \theta)^2)}}.$$

**Figura 6.12** Generación de 300 observaciones que siguen un proceso IMA (1,1) con parámetro  $\theta = .95$  (gráfico superior) y sus calculadas en esta muestra (gráfico inferior)



Suponiendo que  $t$  es grande, de manera que  $t - 1 \approx t$ , y que  $\theta$  no está muy próximo a uno, de manera que  $1+t(1-\theta)^2 \approx t(1-\theta)^2$ , entonces:

$$\rho(t, t+k) = \frac{t}{\sqrt{t(t+k)}},$$

que es la misma función de autocorrelación (6.5) obtenida para el paseo aleatorio: coeficientes próximos a uno que se *amortiguan de forma aproximadamente lineal* al aumentar el retardo. Sin embargo, si  $\theta$  es próximo a uno, los coeficientes no tienen necesariamente que ser grandes, aunque siempre debemos observar el decrecimiento lineal. Por ejemplo, la figura 6.12 muestra una serie generada con  $\theta = .95$ . Se observa que aunque las autocorrelaciones ahora no son nunca próximas a uno, muestran la característica esperada de decrecer linealmente.

## 6.5. Procesos integrados de orden dos

Muchas series reales con tendencia pueden representarse con un modelo con dos diferencias, o integrado de orden dos. Un modelo simple que aparece en muchas aplicaciones es el:

$$\nabla^2 z_t = (1 - \theta B) a_t. \quad (6.9)$$

## 6. Procesos integrados y de memoria larga

Para justificar este modelo, supongamos un paseo aleatorio con una deriva que va cambiando en el tiempo, como:

$$\nabla z_t = c_t + u_t,$$

donde la media de  $\nabla z_t$ , que es el crecimiento de  $z_t$ , evoluciona en el tiempo. Sustituyendo sucesivamente en la ecuación anterior y suponiendo que el proceso comienza en  $t = 0$  y que  $z_0 = u_0 = 0$ , se obtiene:

$$z_t = (c_t + \dots + c_1) + u_t + \dots + u_1.$$

Supongamos que la evolución del coeficiente de crecimiento en cada instante,  $c_t$ , es suave, de manera que:

$$c_t = c_{t-1} + \epsilon_t,$$

donde  $\epsilon_t$  es un proceso de ruido blanco, independiente de  $u_t$ . Entonces:

$$\begin{aligned}\nabla^2 z_t &= \nabla z_t - \nabla z_{t-1} = c_t + u_t - (c_{t-1} + u_{t-1}) = \\ &= \epsilon_t + u_t - u_{t-1} = (1 - \theta B) a_t,\end{aligned}$$

ya que la suma de un ruido blanco y un MA(1) no invertible será un MA(1) invertible. Concluimos, por tanto, que el proceso (6.9) es una generalización del paseo aleatorio que permite que la deriva varíe suavemente en el tiempo. En general, los procesos integrados de orden dos puede verse como una generalización de los procesos integrados de orden uno, pero donde la pendiente de la recta de crecimiento en lugar de ser fija va cambiando con el tiempo.

### 6.6. Procesos integrados ARIMA

Los dos procesos anteriores se han obtenido admitiendo que la raíz de la parte AR de los procesos AR(1) y ARMA(1,1) es la unidad, con lo que estos procesos se convierten en no estacionarios. Esta idea puede generalizarse para cualquier proceso ARMA, permitiendo que una o varias raíces del operador AR sean la unidad. Se obtienen entonces procesos del tipo:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

que llamaremos procesos ARIMA  $(p, d, q)$ . En esta notación  $p$  es el orden de la parte autorregresiva estacionaria,  $d$  es el número de raíces unitarias (orden de integración del proceso) y  $q$  es el orden de la parte media móvil. Utilizando el operador diferencia,  $\nabla = 1 - B$ , el proceso anterior suele

escribirse:

$$\phi_p(B)\nabla^d z_t = c + \theta_q(B)a_t. \quad (6.10)$$

El nombre ARIMA proviene de las iniciales en inglés de los procesos autorregresivos integrados de media móvil (autoregressive integrated moving average), donde «integrado» indica que, llamando  $\omega_t = \nabla^d z_t$ , al proceso estacionario,  $z_t$  se obtiene como suma (integración) de  $\omega_t$ . En efecto, si  $\omega_t = (1 - B)z_t$  como:

$$(1 - B)^{-1} = 1 + B + B^2 + B^3 + \dots$$

resulta:

$$z_t = (1 - B)^{-1} \omega_t = \sum_{j=-\infty}^t \omega_j.$$

Hemos visto en las secciones anteriores dos ejemplos de procesos ARIMA: el paseo aleatorio es el modelo ARIMA(0,1,0) y el de alisado exponencial simple el ARIMA(0,1,1) o IMA(1,1). Ambos se caracterizan porque la función de autocorrelación simple tiene coeficientes que decrecen lentamente. Todos los procesos ARIMA no estacionarios tienen esta propiedad general. Para demostrarlo, recordemos que el correlograma de un ARMA(p, q) satisface para  $k > q$  la ecuación:

$$\phi_p(B)\rho_k = 0 \quad k > q$$

cuya solución es del tipo:

$$\rho_k = \sum_1^p A_i G_i^k,$$

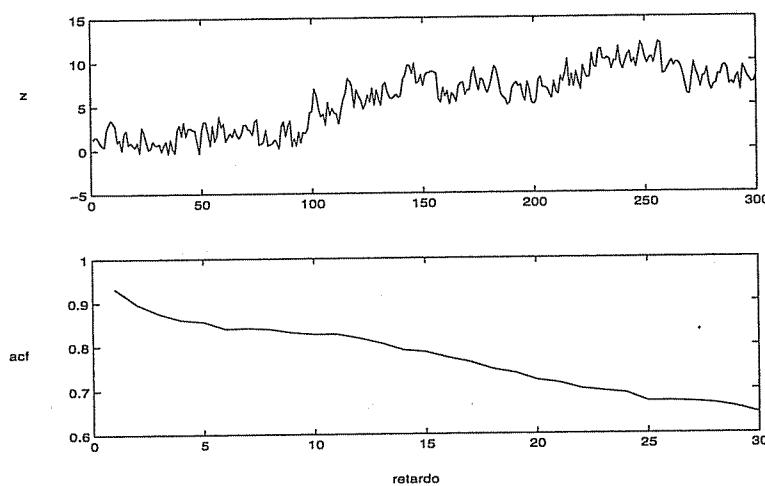
donde  $G_i^{-1}$  son las raíces de  $\phi_p(B) = 0$  y  $|G_i| < 1$ . Si una de estas raíces  $G_i$  es muy próxima a la unidad, escribiendo  $G_i = 1 - \varepsilon$ , siendo  $\varepsilon$  muy pequeño, para  $k$  grande los términos  $A_j G_j^k$  debidos a las otras raíces serán cero (ya que  $G_j^k \rightarrow 0$ ) y tendremos que, aproximadamente:

$$\rho_k = A_i(1 - \varepsilon)^k \simeq A_i(1 - k\varepsilon) \quad \text{para } k \text{ grande}$$

En consecuencia, la *fas* tendrá coeficientes positivos que se amortiguarán aproximadamente de forma lineal y pueden ser distintos de cero incluso para valores altos de  $k$ . Esta propiedad de persistencia de valores positivos en el correlograma (aunque sean pequeños) y de decrecimiento lineal caracteriza a los procesos no estacionarios. La figura 6.13 muestra una realización de la serie  $(1 - .4B)\nabla z_t = (1 - .8B)a_t$  y su función de autocorrelación observada. Se observa de nuevo el decrecimiento lineal

## 6. Procesos integrados y de memoria larga

Figura 6.13 Simulación del proceso  $(1 - .4B)\nabla z_t = (1 - .8B)a_t$  y su función de autocorrelación



de los coeficientes de autocorrelación.

### 6.7. Procesos integrados y tendencias

No todas las series no estacionarias son integradas, pero los procesos integrados cubren muchos casos interesantes que nos encontramos en la práctica. Es fácil demostrar que cualquier proceso que sea la suma de una tendencia polinómica y un proceso estacionario será integrado. Por ejemplo, consideremos el proceso de crecimiento lineal determinista:

$$z_t = b + ct + u_t \quad (6.11)$$

donde  $u_t$  es ruido blanco. Este proceso es integrado de orden uno o  $I(1)$ , ya que tomando la primera diferencia de la serie  $z_t$ :

$$\begin{aligned} \omega_t &= z_t - z_{t-1} = b + ct + u_t - (b + c(t-1) + u_{t-1}) = \\ &= c + u_t - u_{t-1} \end{aligned}$$

como  $c$  es constante y  $\nabla u_t$ , es un proceso estacionario (al ser la diferencia de dos procesos estacionarios, véase la sección 3.4), el proceso  $\omega_t$  será también estacionario. Por tanto,  $z_t$  definido por (6.11), es integrado de orden uno, y su primera diferencia seguirá el modelo (6.7) con  $a_t = u_t$  y el parámetro  $\theta$  igual a la unidad. El proceso será no invertible, y esta propiedad nos indicará que tenemos un componente lineal determinista

en el modelo. Observemos que al diferenciar el proceso la constante que queda es la pendiente del crecimiento determinista lineal.

Este resultado es válido para procesos generados como suma de una tendencia polinómica de orden  $h$  y un proceso estacionario cualquiera,  $u_t$ :

$$z_t = \mu_t + u_t, \quad (6.12)$$

donde

$$\mu_t = a + bt + ct^2 + \dots + dt^h.$$

Si diferenciamos  $h$  veces este proceso, se obtiene una serie estacionaria con una parte de media móvil no invertible. Por ejemplo, supongamos una tendencia cuadrática. Entonces, tomando dos diferencias en el proceso, resulta:

$$\nabla^2 z_t = c + \nabla^2 u_t$$

y el proceso  $\eta = \nabla^2 u_t$  tendrá un componente de media móvil no invertible. Por ejemplo, si  $u_t$  es un AR(1),  $(1 - \phi B)u_t = a_t$ , tendremos que

$$\eta = (1 - \phi B)^{-1} \nabla^2 a_t$$

y  $\eta$  será un proceso ARMA(1,2) no invertible.

Este ejemplo indica que si una serie tiene una tendencia polinómica determinista y se modela como un proceso ARIMA, obtendremos: (1) componentes de media móvil no invertibles; (2) la serie estacionaria tendrá una media distinta de cero y la constante es el coeficiente de orden mayor en la tendencia determinista. De nuevo nos encontramos con la importancia de las constantes en los procesos integrados, ya que representan efectos deterministas.

Los procesos con tendencia lineal determinista son poco frecuentes en la práctica, pero podemos generalizar un poco el modelo anterior permitiendo que la pendiente en cada instante pueda variar un poco respecto al valor anterior. Podemos escribir el proceso como en el capítulo 2, con una tendencia lineal pero con componentes variables, como:

$$z_t = \mu_t + v_t. \quad (6.13)$$

Con este modelo el valor de la serie en cada instante es la suma de su nivel,  $\mu_t$ , y un proceso de ruido blanco,  $v_t$ . Suponemos que el nivel evoluciona como:

$$\mu_t = \mu_{t-1} + c + \varepsilon_t. \quad (6.14)$$

El proceso tiene una tendencia determinista  $c$ , porque el nivel en el instante  $\mu_t$  se obtiene a partir del instante anterior sumando una constante  $c$ , que es la pendiente, más un pequeño componente aleatorio,  $\varepsilon_t$ . En este

## 6. Procesos integrados y de memoria larga

---

proceso la primera diferencia es:

$$\nabla z_t = c + \varepsilon_t + \nabla v_t,$$

y como la suma de un proceso MA(1) y un ruido blanco es un MA(1), obtenemos de nuevo el proceso (6.7), pero ahora con un parámetro  $\theta$  menor que la unidad, como en la representación (6.7).

El modelo anterior puede generalizarse un poco más haciendo que la pendiente determinista,  $c$ , cambie con el tiempo, pero con cierta inercia. Para ello, podemos sustituir la expresión (6.14) por:

$$\mu_t = \mu_{t-1} + \beta_t + \varepsilon_t,$$

y

$$\beta_t = \beta_{t-1} + v_t,$$

donde los procesos  $\varepsilon_t$  y  $v_t$  son ruidos blancos independientes. Ahora el nivel crece de forma lineal, pero con una pendiente,  $\beta_t$ , que cambia en el tiempo. En esta representación si la varianza de  $v_t$  es muy pequeña la serie tiene una tendencia casi constante, y volvemos al modelo anterior. Si tomamos una diferencia en la serie original, tenemos que:

$$\nabla z_t = \beta_t + \varepsilon_t + \nabla u_t,$$

y tomando una segunda diferencia:

$$\nabla^2 z_t = v_t + \nabla \varepsilon_t + \nabla^2 u_t = (1 - \theta_1 B - \theta_2 B^2) a_t, \quad (6.15)$$

ya que si sumamos procesos de media móvil se obtiene un nuevo proceso de media móvil igual al orden máximo de los sumandos. Dependiendo de las varianzas de los tres ruidos blancos podemos tener distintas situaciones. Si la varianza de  $u_t$  es mucho menor que las de los otros ruidos, el término  $\nabla^2 u_t$  puede despreciarse en el segundo miembro de (6.15) y la suma será un MA(1). Análogamente, si el término  $v_t$  es el dominante y despreciamos los otros dos, la serie seguirá aproximadamente el modelo  $\nabla^2 z_t = v_t$ . Por otro lado, si la varianza de  $v_t$  es cero, entonces  $\beta_t = \beta_{t-1} = \beta$  y volvemos al modelo estudiado en la sección 6.4 que tiene una tendencia determinista.

Es interesante que el lector entienda la diferencia entre los dos tipos de modelos presentados. El modelo (6.11) es determinista puro, ya que en cada instante,  $t$ , el valor esperado de la serie está determinado, es  $b + ct$ . Además, el valor esperado de la serie en  $z_t$  conocido el valor en  $z_{t-1}$  sigue siendo el mismo,  $b + ct$ . En este modelo conocer los valores previos no modifica las predicciones. El modelo (6.13) es más general, porque aunque tiene también en cada instante un crecimiento lineal igual a  $c$ , el

nivel de la serie en cada instante no está determinado desde el principio y puede variar aleatoriamente. Esto se traduce en que el valor esperado de la serie en  $z_t$  conocido el valor en  $z_{t-1}$  es  $z_{t-1} + c$ , y depende del valor observado. El modelo (6.11) es muy raro en la práctica, mientras que el (6.13) es frecuente.

En resumen, los modelos ARIMA al incluir las diferencias, incorporaran tendencias estocásticas y, como caso límite, determinísticas. En este caso la serie diferenciada tendrá una media distinta de cero y pueden aparecer términos de media móvil no invertibles.

### 6.8. Procesos de memoria larga (\*)

Existe una clase de procesos estacionarios que pueden confundirse fácilmente con los procesos no estacionarios integrados. Estos son los procesos de memoria larga, que se caracterizan por tener muchos coeficientes de autocorrelación con coeficientes pequeños y que se amortiguan muy lentamente. Esta propiedad se ha observado en ciertas series meteorológicas o ambientales estacionarias. El lento decrecimiento de las autocorrelaciones podría llevar a modelar estas series suponiendo que son integradas de orden uno, tomando una diferencia. Sin embargo el decrecimiento es distinto que para procesos integrados: mucho más rápido para los primeros retardos que en un proceso integrado, pero más lento para retardos altos. Esta estructura hace que en un proceso de memoria larga se observen muchos coeficientes de autocorrelación pequeños y que decrecen muy lentamente para retardos altos, a diferencia de lo que suele ocurrir con procesos integrados, donde el decrecimiento es lineal.

El proceso de memoria larga más simple se obtiene generalizando el paseo aleatorio con:

$$(1 - B)^d z_t = a_t, \quad (6.16)$$

donde ahora el parámetro  $d$  en lugar de ser un número entero es un número real. En el caso  $-0.5 < d < 0.5$  se obtiene un proceso estacionario que se denomina *proceso fraccional*, y que, como vamos a ver, tiene la propiedad de que sus autocorrelaciones decrecen muy lentamente. Para definir el operador  $(1 - B)^d$  cuando  $d$  no es entero, partiremos de la expresión del desarrollo del binomio de Newton. Cuando  $d$  es un número entero positivo, se verifica que:

$$(1 - B)^d = \sum_{i=0}^d \binom{d}{i} (1)^{d-i} (-B)^i = \sum_{i=0}^d \alpha_i B^i, \quad (6.17)$$

donde los coeficientes de esta suma se definen por

$$\alpha_i = \binom{d}{i} (-1)^i = \frac{d!}{(d-i)!i!} (-1)^i = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i,$$

siendo  $\Gamma(a+1) = a\Gamma(a)$  la función gamma (véase el apéndice 6A) que para valores enteros coincide con la función factorial. Para generalizar esta definición cuando  $d$  no sea entero, podemos aprovechar que la función gamma está definida para cualquier número real, y escribir para  $d$  real:

$$(1 - B)^d = \sum_{i=0}^{\infty} \alpha_i B^i, \quad (6.18)$$

donde los coeficientes de esta expansión infinita vienen dados por

$$\alpha_i = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i,$$

y puede demostrarse (véase el apéndice 6.1) que cuando  $d$  es entero esta definición es coherente con (6.17), ya que todos los términos mayores de  $d$  en este sumatorio infinito son cero, con lo que para  $d$  entero (6.18) se reduce a (6.17). Observemos que para  $i = 0$  resulta  $\alpha_0 = 1$ , y utilizando las propiedades de la función gamma se demuestra que:

$$\alpha_i = \prod_{0 < j \leq i} \frac{j-1-d}{j}, \quad \text{para } i = 1, 2, \dots$$

Utilizando esta expresión podemos escribir el proceso fraccional (6.16) mediante una representación AR( $\infty$ ) desarrollando el operador  $(1 - B)^d$ , para obtener:

$$z_t = - \sum_{i=0}^{\infty} \alpha_i z_{t-i} + a_t. \quad (6.19)$$

Puede demostrarse (véase Beran, 1994) que este proceso admite también una representación MA( $\infty$ ) de la forma

$$z_t = \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

donde los coeficientes verifican la representación de Wold de una serie MA( $\infty$ ) estacionaria con varianza finita, es decir,  $\sum \psi_i^2 < \infty$ . Estos coe-

ficientes se obtienen con

$$\psi_i = \frac{\Gamma(d+i)}{\Gamma(i+1)\Gamma(d)}.$$

Se demuestra que la función de autocorrelación del proceso fraccional de larga memoria (6.16) es

$$\rho(k) = \frac{\Gamma(d+k)\Gamma(1-d)}{\Gamma(k+1-d)\Gamma(d)}.$$

En particular para  $k = 1$ , se obtiene

$$\rho(1) = \frac{d\Gamma(d)\Gamma(1-d)}{(1-d)\Gamma(1-d)\Gamma(d)} = \frac{d}{1-d},$$

para  $k = 2$

$$\rho(2) = \frac{d(d+1)}{(2-d)(1-d)} = \rho(1) \frac{(d+1)}{(2-d)}$$

y, en general

$$\rho(k) = \rho(k-1) \frac{(d+k+1)}{(k-d)}.$$

Para  $k$  grande, podemos aproximar las correlaciones por:

$$\rho(k) \simeq \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1}$$

y tomando logaritmos

$$\log \rho(k) \simeq a + (2d-1) \log |k|.$$

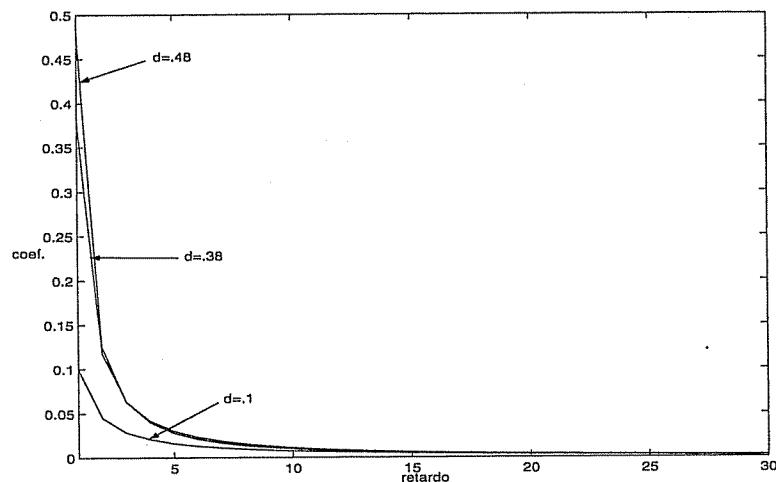
Por tanto, si representamos  $\log \rho(k)$  frente a  $\log |k|$  para valores altos de  $k$  obtenemos una recta con pendiente  $2d - 1$ . Esta propiedad diferencia un proceso de memoria larga de un proceso no estacionario integrado, donde hemos visto que  $\log \rho(k)$  decrece linealmente con  $k$  y no con  $\log k$ .

La figura 6.14 representa los coeficientes  $-\alpha_i$  de la representación AR dada por (6.19) para varios valores de  $d$  y la figura 6.15 la función de autocorrelación para estos valores de  $d$ . Se observa que los coeficientes AR son en seguida muy pequeños, y que la función de autocorrelación tiene muchos coeficientes no nulos que decrecen lentamente. Estas propiedades pueden fácilmente llevar a confundir este proceso con un AR( $p$ ) de orden alto, y si  $d$  es próximo a .5, los altos valores y el lento decrecimiento de la función de autocorrelación pueden sugerir un proceso integrado de orden uno.

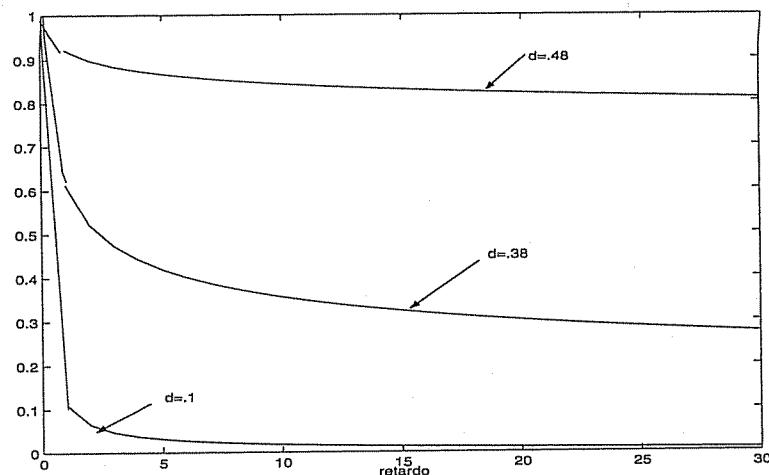
## 6. Procesos integrados y de memoria larga

---

**Figura 6.14 Coeficientes AR del proceso de memoria larga  $\nabla^d z_t = a_t$  con  $d < .5$ .**



**Figura 6.15 Función de autocorrelación del proceso  $\nabla^d z_t = a_t$  con  $d < .5$**



Generalizando lo anterior, podemos considerar procesos estacionarios ARMA pero que incluyan además memoria larga, definiendo

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

donde  $d$  no es necesariamente entero. Si  $d < 0.5$  el proceso es estacionario,

mientras que si  $d \geq 0.5$  el proceso es no estacionario. Estos modelos se denominan ARFIMA ( $p, d, q$ ), autorregresivos fraccionalmente integrados de media móvil. Si el proceso es estacionario tiene la propiedad de que para retardos altos, cuando la estructura ARMA desaparezca, mostrarán la característica de memoria larga. Si no es estacionario, al tomar una diferencia se convertirán en estacionarios, pero de memoria larga.

Los procesos de memoria larga pueden aparecer al agregar series de memoria corta bajo determinadas condiciones. Granger (1980) demostró que si sumamos N procesos AR(1) independientes pero cuyos parámetros provienen de una distribución común, si N es muy grande como límite, obtenemos un proceso de memoria larga. Recordemos que, como vimos en la sección 5.9, la agregación de procesos AR independientes conduce a procesos ARMA, y éste es el resultado esperable con pocos sumandos. El resultado de Granger indica que en el límite tendremos un proceso de memoria larga y sugiere que estos procesos pueden aproximarse por un ARMA con órdenes altos y similares de  $p$  y  $q$ , lo que se comprueba en la práctica. Con series no muy largas es difícil diferenciar un modelo de memoria larga de un ARMA. Sin embargo, si disponemos de series con muchas observaciones, como suele ocurrir con datos climatológicos y financieros, pueden proporcionar en algunos casos aproximaciones mejores que los procesos ARMA de memoria corta, aunque, como veremos en el capítulo siguiente, las diferencias en las predicciones suelen ser pequeñas.

---

### Ejemplo 6.2

La figura 6.16 presenta la serie anual de nivel mínimo de río Nilo entre los años 622 y 1284, que se encuentra en el fichero nilomin.dat. Su función de autocorrelación se presenta en la figura 6.17, y se observa que las correlaciones decrecen muy lentamente. La figura 6.18 presenta el logaritmo de las autocorrelaciones frente al logaritmo del retardo, y se observa que la relación es aproximadamente lineal. Como comparación hemos incluido en esta figura también los datos del índice de la Bolsa de Madrid. En este proceso integrado para valores pequeños de  $k$  las correlaciones decrecen más lentamente en la serie integrada que en la de memoria larga, pero para valores grandes de  $k$  los coeficientes para esta serie, que sigue un modelo ARIMA, decrecen más rápidamente de lo que correspondería a un proceso de memoria larga. Como hemos visto en la figura 6.11, si representamos el logaritmo de las autocorrelaciones como función del retardo sí se obtiene un decrecimiento lineal.

El gráfico muestra también que para diferenciar estos modelos de los ARIMA necesitamos muchos datos, ya que el comportamiento es claramente distinto para valores altos de  $k$ , que sólo pueden obtenerse con muestras grandes.

---

## 6. Procesos integrados y de memoria larga

Figura 6.16 Mínimos anuales del nivel del río Nilo desde el año 622 hasta el año 1284

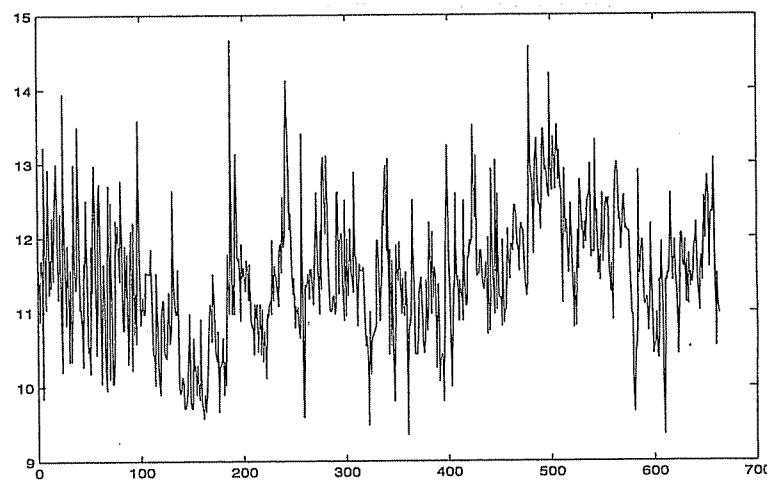
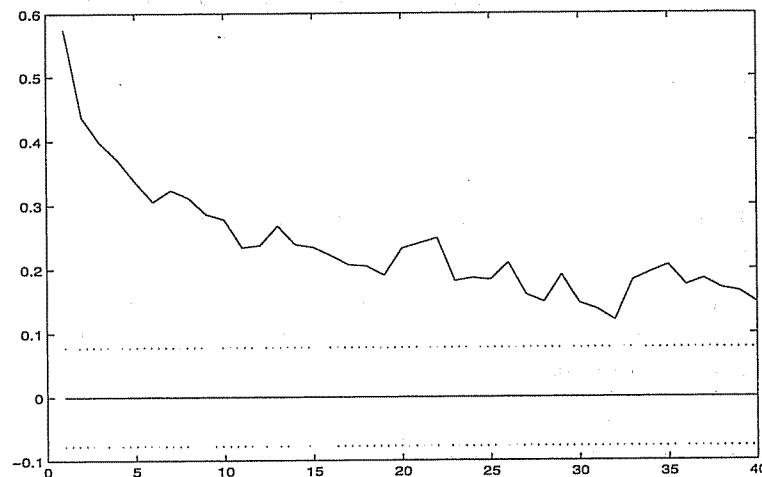


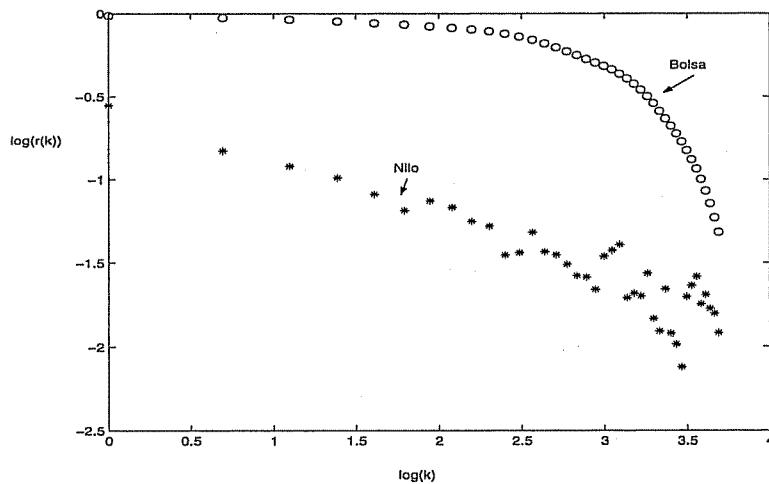
Figura 6.17 Función de autocorrelación para la serie del nivel mínimo del Nilo



### 6.9. Lecturas complementarias

Una buena explicación de los procesos integrados se encuentra en el capítulo 4 de Box, Jenkins y Reinsel (1994). Véase también el capítulo 15 de Hamilton (1994) para otra perspectiva interesante. Una buena referencia para la relación entre los procesos de alisado y los integrados es

**Figura 6.18 Relación entre el logaritmo del coeficiente de autocorrelación y el logaritmo del retardo para los datos del Nilo y los de la Bolsa**



Abraham y Ledolter (1983). Los procesos de memoria larga se presentan en Beran (1994), y su aproximación mediante modelos ARMA en Basak y otros (2001).

---

## Ejercicios 6

- 6.1 Demostrar que la suma y la diferencia de dos procesos estacionarios es estacionaria
- 6.2 Demostrar que el modelo  $z_t = a + bt + ct^2 + a_t$ , donde  $a_t$  es un proceso de ruido blanco, se convierte en un proceso estacionario y no invertible al tomar dos diferencias.
- 6.3 Demostrar que las autocorrelaciones de un paseo aleatorio pueden aproximarse por  $\rho(t, t+k) \simeq 1 - \frac{k}{2t}$ .
- 6.4 Simular un proceso ARIMA (0,1,1) con valores del parámetro  $\theta = .4, .7$  y  $.9$  y estudiar el decrecimiento de la función de autocorrelación del proceso.
- 6.5 Simular el proceso  $\nabla^2 z_t = (1 - .8B)a_t$  y estudiar el decrecimiento de la función de autocorrelación del proceso.

## 6. Procesos integrados y de memoria larga

### Apéndice 6.1: La función gamma y los procesos de memoria larga

La función gamma se define por:

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx,$$

para  $p > 0$ . Integrando por partes se obtiene

$$\Gamma(p) = [-x^{p-1} e^{-x}]_0^\infty + (p-1) \int_0^\infty x^{p-2} e^{-x} dx = (p-1)\Gamma(p-1).$$

Como  $\Gamma(1) = 1$  la función gamma coincide con la función factorial para  $p$  entero. Para cualquier valor positivo y real de  $p$  esta función generaliza la función factorial, ya que siempre podemos escribir  $p = n + r$  siendo  $n$  entero:

$$\Gamma(n+r) = (n+r-1)(n+r-2)\dots(1+r)\Gamma(1+r).$$

Esta propiedad hace que la función gamma se tabule entre 1 y 2 porque podemos calcularla para cualquier  $p$  real aplicando esta relación y sólo necesitamos el valor de  $\Gamma(1+r)$ , que corresponderá a un argumento de la función entre 1 y 2. Su valor en ese intervalo es siempre próxima a la unidad, y la función toma el valor mínimo para  $1+r = 1.4616$ , donde entonces la función vale  $\Gamma(1.4616) = .8856$ .

La función gamma puede también definirse para  $p$  negativo, mediante:

$$\Gamma(p+1) = \lim_{n \rightarrow \infty} \frac{n^p n!}{(p+1)(p+2)\dots(p+n)}$$

y se demuestra que este límite existe y que para valores positivos de  $p$  volvemos al caso anterior y el límite coincide con la integral de definición para  $p > 0$ . Además, la función así definida sigue cumpliendo la ley recurrente  $\Gamma(p) = (p-1)\Gamma(p-1)$ . Se demuestra también que esta función se hace infinita para cero y los valores enteros negativos del argumento,  $p = 0, -1, -2, -3$ .

Estos resultados llevan a definir el operador diferencia fraccional por la suma infinita:

$$(1-B)^d = \sum_{i=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i B^i$$

si  $d$  es entero cuando  $i > d$  entonces  $\Gamma(d-i+1)$  corresponde a la gamma con argumento entero negativo que, según hemos visto, es infinita. Por tanto, todos los coeficientes serán cero para  $i > d$  y obtendremos la expresión habitual del desarrollo de  $(1-B)^d$  cuando  $d$  es entero mediante el binomio de Newton. Observemos que, como:

$$\frac{\Gamma(d+1)}{\Gamma(d-i+1)} = d(d-1)(d-2)\dots(d-i+1),$$

e igualmente

$$\frac{\Gamma(i-d)}{\Gamma(-d)} = (i-d-1)(i-d-2)\dots(1-d)(-d),$$

podemos escribir los términos del desarrollo general de  $(1-B)^d$  mediante:

$$\frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)}(-1)^i = \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)},$$

que es la forma habitual de expresar los coeficientes del desarrollo cuando  $p$  no es entero.

## Apéndice 6.2: Generación de series no estacionarias EViews

Vamos a generar una serie de 400 observaciones que siga un modelo ARIMA  $(0,1,1)$ . Para ello creamos un workfile de  $n = 400$  datos de la manera habitual (*File→New→Workfile; Undated or Irregular→1 to 400*) y generamos una serie donde  $c = 0$ ,  $\theta = -0.5$  y  $\sigma_a^2 = 1$ . Para ello debemos seguir los pasos:

1. Genr→a=nrnd. (define la serie de ruido blanco).
2. Genr→x=0. (define la serie como 0).
3. Genr→ Sample: 2:400 → x=x(-1)+a+0.5\*a(-1).

La serie generada se ve con la instrucción: *View→Line Graph*. La serie no debe tener una media común y mostrar períodos de crecimiento o decrecimiento. En el correlograma (*View→Correlogram*), la función de autocorrelación estimada de la serie debe decaer lentamente y de manera aproximadamente lineal, y en la función de autocorrelación parcial debe haber un primer retardo cercano a 1.

Definimos ahora el proceso primera diferencia de este,  $y_t = x_t - x_{t-1}$ . Para ello, en el menú del wokfile:

$$y = d(x, 1) \quad \text{Sample : 2 : 400}$$

Debemos comprobar que la serie  $y$  es estacionaria: su correlograma (*View→Correlogram*), debe corresponder a un proceso MA(1).

Un proceso  $I(2)$  se genera igual. Por ejemplo, generemos una muestra del proceso  $(1-B)^2 x_t = a_t$ . Los pasos son:

## 6. Procesos integrados y de memoria larga

1. Genr→x=0. (define la serie como 0).
2. Genr→ Sample: 3:400 → x=2\*x(-1)-x(-2)+a. (damos valores a la serie).

El proceso de generación asume que  $x_1 = 0$ . Al hacer un gráfico de la serie (*View→Line Graph*) se observará que la serie cambia de nivel y es no estacionaria. El correlograma muestra que la función de autocorrelación estimada de la serie decrece muy lentamente además de manera lineal. Si definimos el proceso primera diferencia de este,  $y_t = x_t - x_{t-1}$ , mediante el menú del wokfile:

$$y = d(x, 1) \quad \text{Sample : } 2 : 400$$

comprobaremos que la serie  $y$  es no estacionaria y su correlograma (*View → Correlogram*) corresponde a un proceso no estacionario.

### Matlab

Podemos generar series ARIMA no estacionarias con el programa simarma que vimos en el capítulo 4. La entrada del programa requiere el vector de todos los parámetros AR, incluyendo las diferencias. Por ejemplo, para simular el modelo  $(1 - .5B)\nabla z_t = (1 - .7B)a_t$ , hay que calcular el operador AR ampliado:  $(1 - .5B)(1 - B) = (1 - 1.5B + .5B^2)$  e introducir como vector de parámetros AR: fi=[1.5 -.5] y como MA: teta=[.7].

## 7. Procesos ARIMA estacionales



Evgeny Evgenievich Slutsky (1880-1948)

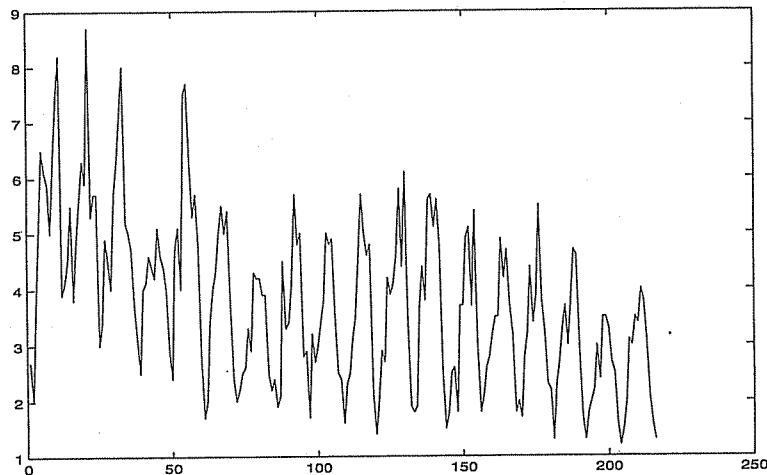
Estadístico ruso. Siendo estudiante en la Universidad de Kiev fue expulsado de todas las universidades rusas por sus actividades revolucionarias. Estudió ingeniería en Munich y después de la Revolución rusa trabajó en la administración soviética. Interesado por la estadística, demostró en 1927 que la aplicación de medias móviles a datos independientes genera secuencias periódicas, e introdujo los procesos de media móvil para explicar ciclos económicos. Sus investigaciones contribuyeron a la creación de la teoría de los procesos estacionarios y a la comprensión de los ciclos estacionales.

### 7.1. Introducción

En este capítulo continuamos el estudio de los procesos no estacionarios analizando un tipo de falta de estacionariedad en la media que es muy frecuente en la práctica: el comportamiento estacional. La estacionalidad hace que la media de las observaciones no será constante, pero evoluciona de forma previsible de acuerdo con un patrón cíclico. Por ejemplo, en una serie de temperaturas mensuales en Europa la media de las temperaturas no es constante, ya que varía con el mes, pero el mismo mes en distintos años es esperable que tenga un valor medio constante.

El caso más habitual es que podamos incorporar la estacionalidad dentro del modelo ARIMA de una forma multiplicativa, con lo que resulta el modelo ARIMA estacional multiplicativo. En este capítulo estudiaremos las propiedades de estos modelos.

**Figura 7.1 Serie de contenido de ozono en la atmósfera entre enero de 1955 y diciembre de 1972**



## 7.2. El concepto de estacionalidad y sus tipos

Diremos que una serie es estacional cuando su valor esperado no es constante, pero varía con una pauta cíclica. En concreto si:

$$E(z_t) = E(z_{t+s}),$$

diremos que la serie tiene una *estacionalidad* de periodo  $s$ . Por ejemplo, una serie mensual tiene estacionalidad si los valores esperados en distintos meses del año son distintos, pero el valor esperado en el mismo mes en distintos años es el mismo. Como ilustración, la serie de contenido de ozono en la atmósfera de la figura 7.1 presenta una pauta estacional muy fuerte, con picos y valles en los mismos meses del año. Esta pauta estacional se superpone a la tendencia global decreciente, produciendo un comportamiento cíclico que se repite los distintos años de la muestra. La pauta estacional es también visible en la serie.

El periodo estacional,  $s$ , define, como vimos en el capítulo 2 (sección 2.4), el número de observaciones que forman el ciclo estacional. Por ejemplo,  $s = 12$  para series mensuales,  $s = 4$  para series trimestrales, etc. Supondremos que el valor de  $s$  es fijo en la serie. Esto puede no ser exactamente cierto: por ejemplo, si tenemos datos diarios y el periodo estacional es la longitud del mes  $s$  será aproximadamente 30, pero variará de unos meses a otros. También puede existir más de un tipo de estacionalidad. Por ejemplo, con datos diarios podemos tener una estacionalidad

semanal, con  $s = 7$ , otra mensual, con  $s = 30$  y otra anual, con  $s = 365$ . Supondremos inicialmente que existe sólo un tipo de estacionalidad y al final del capítulo veremos cómo extender los métodos presentados cuando existan varios períodos estacionales.

El modelo más simple para la estacionalidad es el estudiado en el capítulo 2, donde la estacionalidad se modela como un efecto constante que se suma a los valores de la serie. Por ejemplo, supongamos una serie que, salvo por el efecto estacional, es estacionaria. Podemos escribir la serie como suma de un componente estacional  $S_t^{(s)}$ , y un proceso estacionario,  $n_t$ , de manera que el modelo para la serie será:

$$z_t = S_t^{(s)} + n_t. \quad (7.1)$$

Esta serie (7.1) no es estacionaria, ya que si tomamos esperanzas:

$$E(z_t) = E(S_t^{(s)}) + \mu,$$

donde  $\mu$  es la media del proceso  $n_t$ . Como, por definición, el componente estacional no toma el mismo valor en todos los períodos, la serie no será estacionaria, al no tener media constante.

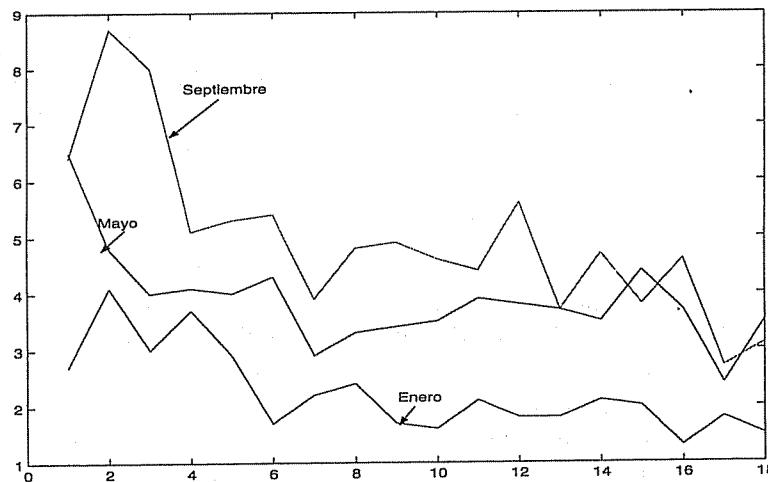
Por ejemplo, la figura 7.7 muestra la evolución del contenido en ozono de la figura 7.1 en tres meses distintos. Cada una de las series representa el ozono en ese mes en los distintos años de la muestra. Además de la tendencia decreciente, claramente visible en la figura 7.1, se observa que el contenido de ozono en enero es siempre menor que en mayo, y éste, en general, menor que en septiembre. Por tanto, cada mes tiene un comportamiento medio distinto, que es lo que caracteriza una serie estacional. El mismo fenómeno se observa en la figura 7.3: los meses de enero reciben más lluvia que los de septiembre y éstos que los de julio. La media de cada serie es distinta y existe un efecto estacional.

Podemos considerar distintas hipótesis respecto al comportamiento del proceso estacional  $S_t^{(s)}$ . La primera es que  $S_t^{(s)}$  sea un proceso determinista, es decir, una función constante para el mismo mes en distintos años:

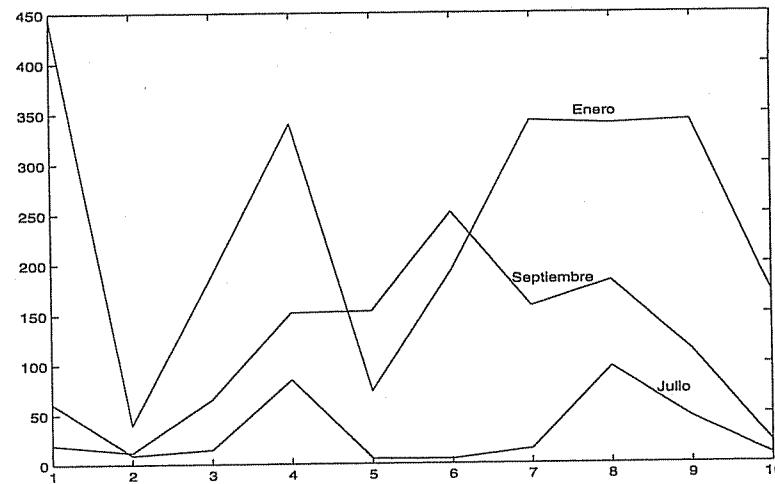
$$S_t^{(s)} = S_{t+ks}^{(s)} \quad k = \pm 1, \pm 2, \dots \quad (7.2)$$

Por ejemplo, los coeficientes estacionales pueden seguir una función sinusoidal, como ocurre con series climatológicas por la rotación de la Tierra alrededor del Sol. Esto sugiere tratar la estacionalidad con funciones sinusoidales como las estudiadas en el capítulo 2, pero estas funciones serán poco eficaces cuando la estacionalidad siga una pauta determinista pero no sinusoidal. Por ejemplo, la estacionalidad en muchas series de producción de empresas españolas muestran un valor constante todos los

**Figura 7.2** Representación de las series de contenido de ozono en los meses de enero, marzo y septiembre para distintos años en el periodo 55-72



**Figura 7.3** Evolución de la lluvia caída en Santiago de Compostela en un mes en distintos años para los meses de enero, septiembre y julio



meses con la excepción del mes de agosto, donde el valor esperado es muy bajo porque la fábrica se cierra por vacaciones. Representar esta estacionalidad mediante funciones sinusoidales sería muy ineficiente. La

estacionalidad determinista puede siempre modelizarse introduciendo 11 variables ficticias, una para cada mes del año tal como veremos en el capítulo 12.

Sin embargo, muchas series temporales no tienen una estacionalidad determinista, sino que la pauta estacional evoluciona también, como otras propiedades de la serie, a lo largo del tiempo. La segunda forma de modelar la estacionalidad es suponer que su evolución es estacionaria, es decir, los factores estacionales no son constantes, pero siguen un proceso estacionario, oscilando alrededor de un valor medio de acuerdo con la representación:

$$S_t^{(s)} = \mu^{(s)} + v_t, \quad (7.3)$$

donde  $\mu^{(s)}$  es una constante que depende del mes y representa el efecto determinista de la estacionalidad y  $v_t$  es un proceso estacionario de media cero que introduce variabilidad en la estacionalidad de cada año.

La tercera forma de modelar la estacionalidad es permitir que sea cambiante en el tiempo sin ningún valor medio fijo. En este caso, la estacionalidad sigue un proceso no estacionario. Por ejemplo, el modelo más simple es suponer que evoluciona según un paseo aleatorio:

$$S_t^{(s)} = S_{t-s}^{(s)} + v_t, \quad (7.4)$$

donde  $v_t$  es un proceso estacionario de media cero. Naturalmente, son posibles situaciones más complicadas y la estacionalidad puede seguir cualquier proceso no estacionario.

Vamos a comprobar que, en los tres casos presentados, podemos convertir una serie estacional en estacionaria aplicando una *diferencia estacional*. Definimos el operador diferencia estacional de periodo  $s$  como:

$$\nabla_s = 1 - B^s$$

(nótese que  $\nabla_s \neq \nabla^s = (1 - B)^s$ ). Si aplicamos este operador a una serie, se obtiene una serie transformada sustituyendo en cada instante,  $t$ , el valor de la serie por la diferencia entre el valor en  $t$ , y el valor en  $t-s$ . En efecto:

$$\nabla_s z_t = (1 - B^s)z_t = z_t - z_{t-s}.$$

Por tanto, si aplicamos este operador  $\nabla_s$  en (7.1), tenemos:

$$\nabla_s z_t = \nabla_s S_t^{(s)} + \nabla_s n_t$$

y vamos a comprobar que la serie  $\nabla_s z_t$  es estacionaria. Consideremos los tres casos estudiados:

- El componente  $S_t^{(s)}$  es determinista. Entonces, según (7.2):

$$\nabla_s S_t^{(s)} = S_t^{(s)} - S_{t-s}^{(s)} = 0$$

y

$$\nabla_s z_t = \nabla_s n_t.$$

Al diferenciar un proceso estacionario,  $\nabla_s n_t$ , se obtiene otro proceso estacionario, y la diferencia  $\nabla_s$  transforma la serie no estacionaria  $z_t$  dada por (7.1) en estacionaria. Observemos que en este caso, el modelo para  $\nabla_s z_t$  tiene un coeficiente de media móvil  $(1 - \Theta B^s)$  con  $\Theta = 1$ .

- El componente  $S_t^{(s)}$  sigue un proceso estacionario. Tomando una diferencia estacional en (7.3), como  $\nabla_s S_t^{(s)} = \nabla_s v_t$  será estacionario, se obtiene de nuevo un proceso estacionario, con:

$$\nabla_s z_t = \nabla_s v_t + \nabla_s n_t,$$

y utilizando las reglas para sumar dos procesos MA(1) que estudiamos en el capítulo 5, concluimos que el proceso suma,  $\nabla_s z_t$ , es un MA(1) en el operador  $B^s$  con un coeficiente  $\Theta = 1$ .

- El componente  $S_t^{(s)}$  sigue un proceso no estacionario. Si suponemos el caso más simple del modelo (7.4), tendremos que:

$$\nabla_s z_t = v_t + \nabla_s n_t,$$

y, de nuevo,  $\nabla_s z_t$  es un proceso estacionario suma de dos términos: el primero es ruido blanco y el segundo un MA(1) del tipo  $(1 - \Theta B^{12})$  con  $\Theta = 1$ . Como la autocorrelación del proceso suma es una media ponderada de las autocorrelaciones de los sumandos, según vimos en el capítulo 5, la suma dará lugar a un MA(1),  $(1 - \Theta B^{12})$ , con un coeficiente de media móvil invertible  $\Theta < 1$ .

En conclusión, en condiciones bastante generales para la estructura estacional, y tanto para procesos con estacionalidad determinista como estocástica, el operador  $\nabla_s$  convierte un proceso estacional en estacionario.

### 7.3. El modelo ARIMA estacional

Hemos visto que podemos convertir series no estacionarias en estacionarias tomando diferencias regulares, es decir, entre períodos consecutivos. En la sección anterior hemos visto también que podemos eliminar la estacionalidad mediante diferencias estacionales. Uniendo ambos resultados,

concluimos que podemos convertir una serie con estacionalidad en estacionaria mediante la transformación:

$$w_t = \nabla_s^D \nabla^d z_t,$$

donde  $D$  es el número de diferencias estacionales (si hay estacionalidad casi siempre  $D = 1$ , y si no hay estacionalidad  $D = 0$ ) y  $d$  el número de diferencias regulares ( $d \leq 3$ ).

Cuando existe dependencia estacional podemos generalizar el modelo ARMA para series estacionarias incorporando además de la dependencia *regular*, que es la asociada a los intervalos de medida de la serie, la dependencia *estacional*, que es la asociada a observaciones separadas por  $s$  periodos. Vamos a estudiar cómo modelar estos dos tipos de dependencias.

La primera solución es incorporar la dependencia estacional a la regular, añadiendo a los operadores AR o MA en el operador  $B$  términos  $B^s$ , para representar la dependencia entre observaciones separadas por  $s$  periodos. El inconveniente de esta formulación es que llevaría a polinomios muy largos en la parte AR y MA. Por ejemplo, con datos mensuales,  $s = 12$ , si un mes está relacionado con el mismo mes en tres años anteriores necesitamos un AR o MA de orden 36 para representar esta dependencia. Un enfoque más simple, y que funciona bien en la práctica, es modelar de forma separada la dependencia regular y la estacional, y construir el modelo incorporando ambas de forma multiplicativa. Se obtiene así el modelo ARIMA *estacional multiplicativo*, que tiene la forma:

$$\Phi_P(B^s) \phi_p(B) \nabla_s^D \nabla^d z_t = \theta_q(B) \Theta_Q(B^s) a_t, \quad (7.5)$$

donde  $\Phi_P(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})$  es el operador AR estacional de orden  $P$ ,  $\phi_p = (1 - \phi_1 B - \dots - \phi_p B^p)$  es el operador AR regular de orden  $p$ ,  $\nabla_s^D = (1 - B^s)^D$  representa las diferencias estacionales, y  $\nabla^d = (1 - B)^d$  las diferencias regulares,

$$\Theta_Q(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})$$

es el operador media móvil estacional de orden  $Q$ ,

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

es el operador media móvil regular de orden  $q$ , y  $a_t$  es un proceso de ruido blanco. Esta clase de modelos, introducidos por Box y Jenkins (1976), representan bien muchas series estacionales que encontramos en la práctica y se escriben de forma simplificada como el modelo ARIMA  $(P, D, Q)_s \times (p, d, q)$ .

Para justificar (7.5) consideremos una serie estacional  $z_t$  con periodo  $s$ . Como ejemplo, supondremos que los datos son mensuales, que el primer dato es de enero y que disponemos de  $h$  años completos con un total de  $T = 12h$  observaciones. Para modelar la dependencia estacional de esta serie mensual podemos dividir la serie total en 12 series anuales, cada una de ellas con  $h$  datos, el número de años disponibles. Por ejemplo, las figuras 7.2 y 7.3 presentan ejemplos de estas series donde representamos la evolución del mismo mes en distintos años. Llamemos a estas series anuales  $y_\tau^{(j)}$ , donde  $j = 1, \dots, 12$  indica el mes que define la serie y el índice temporal de la serie,  $\tau$ , es el año, que varía entre  $\tau = 1, \dots, h$ . En efecto, estas series se refieren siempre al mismo mes,  $j$ , y relacionan el valor de ese mes en distintos años. La forma de obtener estas series anuales a partir de la serie original mensual,  $z_t$ , es mediante la relación:

$$y_\tau^{(j)} = z_{j+12(\tau-1)} \quad (\tau = 1, \dots, h). \quad (7.6)$$

Por ejemplo, la serie anual del mes de enero,  $j = 1$ , comienza,  $\tau = 1$ , con el primer dato observado  $z_1$  y continúa,  $\tau = 2$ , tomando el dato 12 meses después, es decir  $z_{13}$ , que corresponderá al dato de enero en el segundo año. Los siguientes corresponden a los  $z_{25}, z_{37}, \dots$ , moviendo cada vez 12 posiciones de la serie original. En general, la serie anual del mes  $j$  se obtiene tomando los datos de la serie  $z_t$  para  $t = j + 12(\tau - 1)$  con  $\tau = 1, \dots, h$ . Aplicar el operador de retardo  $B$  a la serie  $y_\tau^{(j)}$  equivale a aplicar el operador  $B^{12}$  a la serie  $z_t$  y se verifica:

$$By_\tau^{(j)} = y_{\tau-1}^{(j)} = z_{j-12(\tau-2)} = z_{j-12+12(\tau-1)} = B^{12}z_{j+12(\tau-1)}.$$

Cada una de estas 12 series anuales seguirá un modelo ARIMA:

$$\begin{aligned} \Phi^{(j)}(B)(1-B)^D y_\tau^{(j)} &= c_j + \Theta^{(j)}(B)u_\tau^{(j)}; \\ \tau &= 1, \dots, h \end{aligned} \quad (7.7)$$

donde

$$\Phi^{(j)}(B) = (1 - \Phi_1^{(j)}B - \dots - \Phi_P^{(j)}B^P)$$

y

$$\Theta^{(j)}(B) = (1 - \Theta_1^{(j)}B - \dots - \Theta_Q^{(j)}B^Q)$$

y supongamos que tienen el mismo modelo común. Entonces, forzosamente este modelo debe ser no estacionario con  $D = 1$ . En efecto, si  $D = 0$ , las series de cada mes oscilarían alrededor de un valor  $c_j \neq 0$  y para que el modelo sea común  $c_j = c$ . Ahora bien, en un modelo estacionario la constante es proporcional a la media del proceso, y como las medias de los meses son distintas, por hipótesis, ya que estamos

suponiendo que la serie es estacional, las constantes  $c_j$  deberían ser distintas. Por tanto, no podemos tener  $D = 0$  y el mismo modelo. Sin embargo, si el modelo común tiene  $D = 1$ , al tomar esta diferencia en cada serie las medias de los meses desaparecen, y las diferencias entre el mismo mes de un año y del anterior seguirán un proceso estacionario de media cero. El modelo común de las series anuales puede escribirse en términos de los datos originales como:

$$(1 - \Phi_1 B^{12} - \dots - \Phi_P B^{12P}) (1 - B^{12}) z_t = \\ = (1 - \Theta_1 B^{12} - \dots - \Theta_Q B^{12Q}) \alpha_t \quad (7.8)$$

donde  $t = 1, \dots, T$  y ahora el modelo ARIMA se formula en  $B^{12}$  ( $B^s$  en general para datos de periodo  $s$ ) ya que estamos relacionando meses de distintos años. Como el modelo es el mismo para todos los meses, podemos aplicar este modelo a la serie original  $z_t$  y obtener la serie de residuos  $\alpha_t$ . Esta serie de residuos no será normalmente un proceso de ruido blanco, ya que no hemos tenido en cuenta la dependencia entre un dato y los inmediatamente anteriores. Suponiendo que  $\alpha_t$  sigue el proceso ARIMA regular:

$$\phi_p(B) \nabla^d \alpha_t = \theta_q(B) a_t, \quad (7.9)$$

sustituyendo este *modelo regular* (7.9) en el *modelo estacional* (7.8) se obtiene el *modelo completo* para el proceso observado, que es el dado en (7.5).

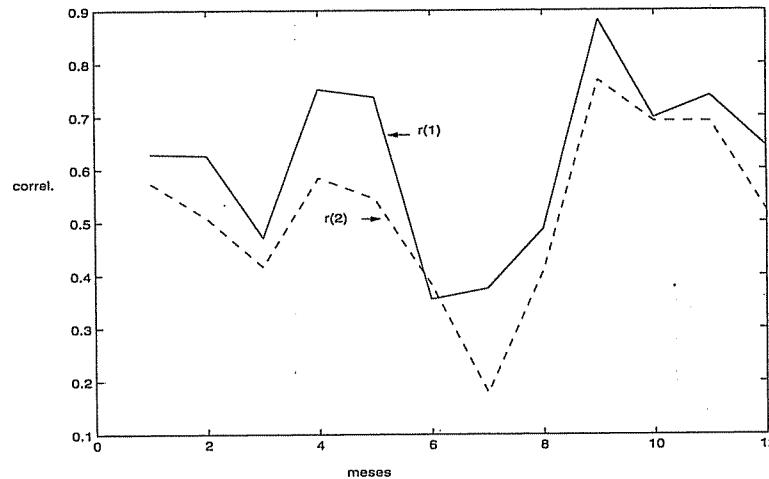
En resumen, el modelo ARIMA estacional multiplicativo se basa en la hipótesis central de que la *relación de dependencia estacional* (7.8) es la misma para todos los periodos. La experiencia indica que esta situación, aunque muy frecuente, no se cumple siempre, y conviene, siempre que se disponga de datos suficientes, contrastarla construyendo los modelos (7.8) y viendo si presentan la misma estructura.

---

### Ejemplo 7.1

Estudiemos la constancia de la estructura de autocorrelación entre las observaciones de los distintos meses para los datos de las lluvias en Santiago, fichero lluviaSC.dat, que presentamos en el capítulo 1 y que hemos utilizado en el capítulo 2. Para ello, formamos las doce series anuales de lluvia en cada uno de los meses. Por ejemplo, la figura 7.3 presenta tres de estas series. A continuación calculamos el coeficiente de autocorrelación de orden uno,  $r_1$ , y el de orden dos,  $r_2$ , para cada una de estas 12 series. La figura 7.4 muestra estos dos primeros coeficientes de autocorrelación para cada uno de los doce meses como una línea continua. Los valores menores de estos coeficientes aparecen

**Figura 7.4 Representación de los dos primeros coeficientes de autocorrelación entre meses de distintos años para los datos de las lluvias de Santiago**

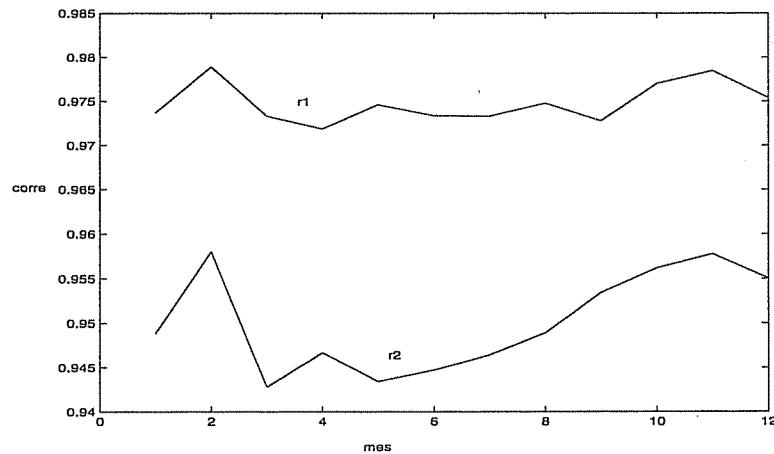


en junio, para  $r_1$ , y en julio para  $r_2$ , mientras que los más altos son para septiembre, para  $r_1$  y  $r_2$ . La figura sugiere que hay más semejanza entre los meses de otoño de distintos años que entre los meses de verano, pero como la variabilidad al estimar estos coeficientes es alta, no podemos concluir que son claramente distintos, ya que necesitariamos series más largas. Por ejemplo, la desviación típica aproximada de estimar cada coeficiente es  $1/\sqrt{T}$  y en este caso  $1/\sqrt{10} = .31$ , por lo que no podemos concluir que los coeficientes son significativamente distintos (en el apéndice 7.1 se indica un posible test para estos coeficientes).

### Ejemplo 7.2

La figura 7.5 presenta los dos primeros coeficientes de autocorrelación calculados con las 12 series anuales de cada mes del año para el consumo de gasolina en España del fichero gasolauto.dat. Como ahora la serie es muy larga, 55 años de datos, los coeficientes varían muy poco de unos meses a otros. En el gráfico del primer coeficiente de autocorrelación,  $r_1$ , para los distintos meses se observa que marzo y abril presentan menor autocorrelación que el resto de los meses. En el gráfico del segundo coeficiente,  $r_2$ , el valor mínimo de la autocorrelación corresponde al mes de marzo. Este efecto puede ser debido a la Semana Santa: esta festividad depende de la luna y puede caer en uno de esos dos meses o entre los dos. Si ocurre en un mes, aumenta el consumo de gasolina del mes por los desplazamientos de vacaciones, mientras que si ocurre

**Figura 7.5 Coeficientes de autocorrelación de orden uno y dos para las series de los meses con los datos de consumo de gasolina 1945-1999**



en ambos, afectará a los dos meses de forma proporcional. Este efecto de Semana Santa introduce cierta heterogeneidad en esos meses, lo que disminuye algo su estructura de autocorrelación. Podemos concluir que, aparte de este efecto, las autocorrelaciones del resto de los meses son similares.

---

#### 7.4. Función de autocorrelación simple de un proceso estacional

Sea  $w_t = \nabla^d \nabla_s^D z_t$  el proceso estacionario obtenido al diferenciar la serie  $d$  veces de forma regular (con  $\nabla = 1 - B$ ) y  $D$  veces de forma estacional (con  $\nabla_s = 1 - B^s$ ). Entonces,  $w_t$  sigue el proceso ARMA estacional multiplicativo:

$$\Phi_P(B^s) \phi_p(B) w_t = \theta_q(B) \Theta_Q(B^s) a_t. \quad (7.10)$$

La función de autocorrelación simple de este proceso es una mezcla de las funciones de autocorrelación correspondientes a las partes regular y estacional. Puede demostrarse (Peña, 1984) que si llamamos  $r_i$  a los coeficientes de autocorrelación del proceso ARMA  $(p, q)$  «regular»:

$$\phi_p(B) x_t = \theta_q(B) u_t, \quad (7.11)$$

## 7. Procesos ARIMA estacionales

---

$R_{si}$  a los coeficientes de autocorrelación en los retardos  $s, 2s, 3s, \dots$  del proceso ARMA  $(P, Q)$  estacional:

$$\Phi_P(B^s) y_t = \Theta_Q(B^s) v_t, \quad (7.12)$$

y  $\rho_j$  a los coeficientes de autocorrelación del proceso completo (7.10), se verifica:

$$\rho_j = \frac{r_j + \sum_{i=1}^{\infty} R_{si} (r_{si+j} + r_{si-j})}{1 + 2 \sum_{i=1}^{\infty} r_{si} R_{si}}. \quad (7.13)$$

Si suponemos  $s = 12$  y admitimos que  $r_j \simeq 0$  para retardos altos (por ejemplo, para  $j \geq 8$ ), el denominador de (7.13) es la unidad y la función de autocorrelación será:

1. En los retardos bajos ( $j = 1, \dots, 6$ ) se observará únicamente la parte regular, es decir:

$$\rho_j \simeq r_j \quad j = 1, \dots, 6.$$

2. En los retardos estacionales se observará básicamente la parte estacional, es decir:

$$\rho_{12i} \simeq R_{12i} (r_{24i} + r_0) + R_{24i} (r_{36i} + r_{12i})$$

y suponiendo que  $r_{12i} \simeq 0$  para  $i \geq 1$ , con  $r_0 = 1$  esta expresión se reduce a:

$$\rho_{12i} \simeq R_{12i} \quad (i = 1, 2, \dots).$$

3. Alrededor de los retardos estacionales observaremos la *interacción* entre la parte regular y estacional, que se manifestará en la repetición a ambos lados de cada retardo estacional de la función de autocorrelación simple de la parte regular. Concretamente, si la parte regular es media móvil de orden  $q$ , habrá a ambos lados de cada retardo estacional no nulo  $q$  coeficientes distintos de cero. Si la parte regular es autorregresiva, observaremos a ambos lados de los retardos estacionales el decrecimiento impuesto por esta estructura AR.

La figura 7.6 presenta algunos ejemplos de modelos AR(1) y MA(1) estacionales con distintas estructuras regulares. Si combinamos estas estructuras con un MA(1) regular, observaremos interacción sólo en los retardos contiguos al estacional, mientras que si lo combinamos con un AR regular, tendremos una estructura larga de interacción, donde se repite a ambos lados de los retardos estacionales la estructura regular. En la práctica, lo recomendable es fijarse en los retardos iniciales

1,2,3,4..., para identificar la estructura regular y en los retardos estacionales,  $s, 2s, 3s, \dots$ , para identificar la estructura estacional. La interacción alrededor de los coeficientes estacionales puede entonces utilizarse como confirmación de la identificación realizada.

---

### Ejemplo 7.3

La figura 7.7 presenta la fas de la serie de lluvias en Santiago. Se observa que en la parte regular hay estructura AR decreciente, mientras que en los retardos 12, 24, 36, se observa un decrecimiento lento de los coeficientes, indicando la presencia de un componente estacional de periodo 12. La figura 7.8 presenta la serie diferenciada de forma estacional. En los retardos 1, 2, 3, se observa ahora una estructura débil que puede ser un AR(1) o MA(1) con coeficiente pequeño y en la parte estacional hay coeficientes significativos en 12, 24 y 36, indicando estructura AR que debe ser al menos de orden 2, debido a que los coeficientes 12 y 24 son negativos y el 36 positivo. Los coeficientes altos que aparecen alrededor de los retardos estacionales confirman que existe estructura regular, en particular los altos coeficientes en los retardos 11 y 23 son señal clara de interacción entre la parte regular y la estacional.

### Ejemplo 7.4

La figura 7.9 presenta la fas de la primera diferencia regular de la serie de ozono. Como hemos visto en la figura 7.1, la serie original tiene una tendencia decreciente, por lo que es necesario diferenciarla para obtener una serie sin tendencia. Este gráfico muestra claramente que existe una fuerte dependencia estacional, con coeficientes altos en 12, 24, 36,... que se amortiguan lentamente con el retardo, sugiriendo la necesidad de una diferencia estacional.

La figura 7.10 presenta la fas de la serie con una diferencia regular y otra estacional. En los retardos iniciales correspondientes a la parte regular se observa o bien un MA(2) o bien un AR de al menos orden 2. En los retardos estacionales es significativo el coeficiente en 12 y está en el límite el de 24, sugiriendo una estructura estacional similar a la regular.

### Ejemplo 7.5

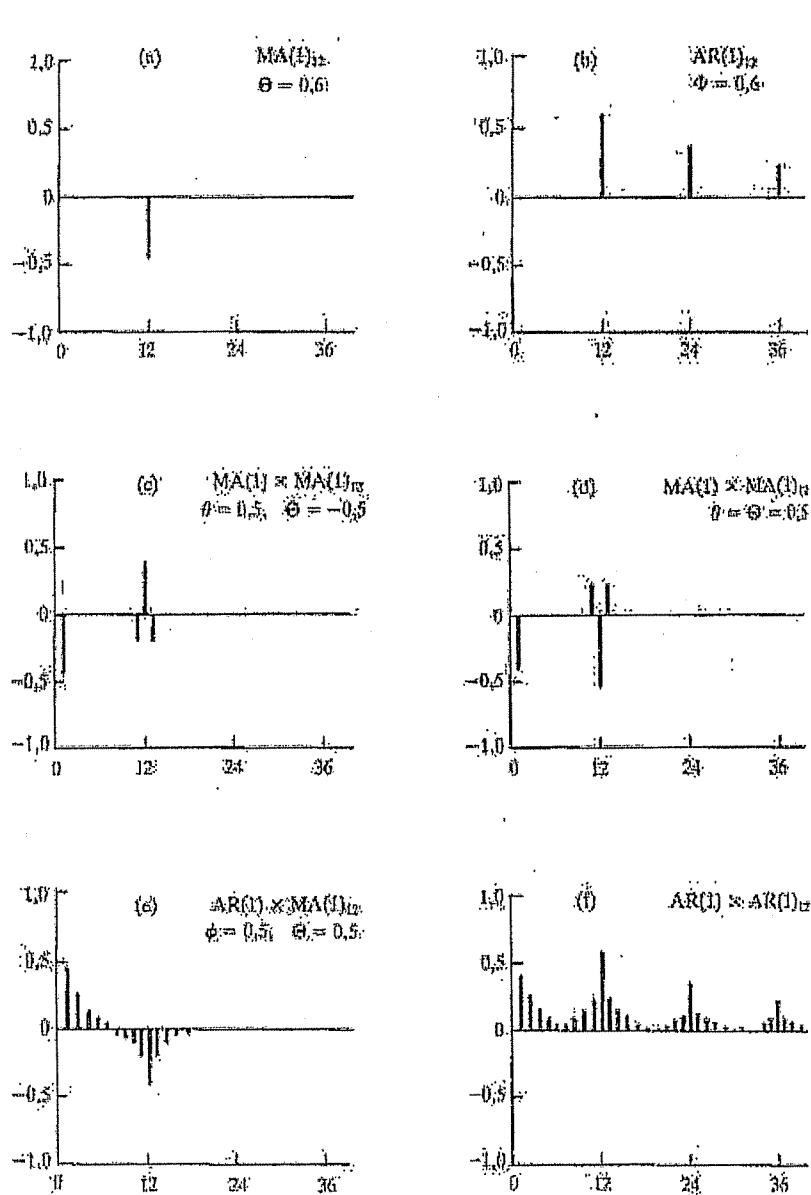
Escribir la función de autocorrelación simple teórica de la serie  $\omega_t$  obtenida al aplicar una diferencia regular y otra de periodo siete a la serie diaria,  $z_t$ , con estacionalidad semanal que sigue el modelo:

$$\omega_t = \nabla \nabla_7 z_t = (1 - 0.5B)(1 - 0.8B^7)a_t.$$

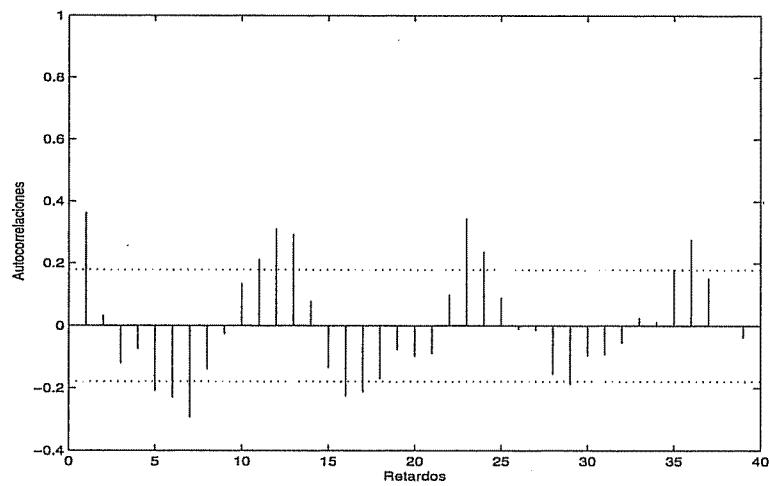
La componente regular es MA(1) y producirá un único coeficiente de autocorrelación distinto de cero. Los coeficientes generados por la parte regular

## 7. Procesos ARIMA estacionales

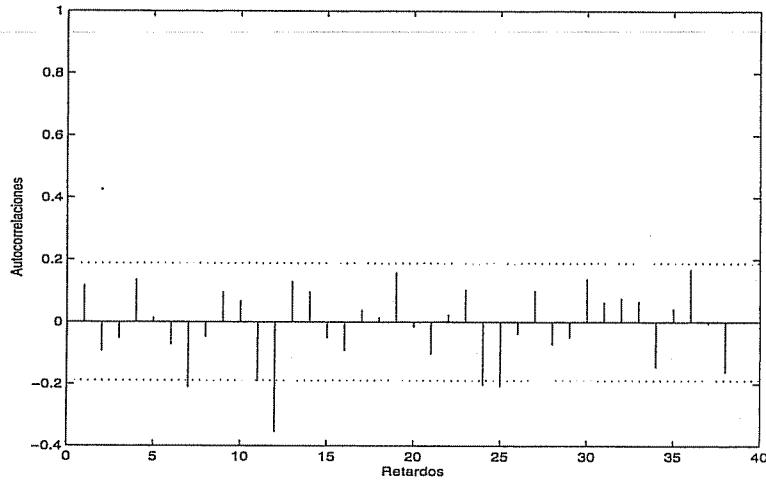
Figura 7.6 Funciones de autocorrelación para algunos procesos estacionales simples



**Figura 7.7 Función de autocorrelación de las lluvias en Santiago de Compostela. Datos mensuales desde 1/88 a 12/97**



**Figura 7.8 Función de autocorrelación de la primera diferencia estacional de la serie de lluvias en Santiago de Compostela. Datos mensuales desde 1/88 a 12/97**



son:

$$r_t = \frac{-0.5}{1 + 0.5^2} = -0.4; \quad r_j = 0 \quad j \geq 2.$$

La componente estacional es también MA(1)<sub>7</sub> con coeficientes generados:

Figura 7.9 Fas de la serie de ozono 1/1955 a 12/1974

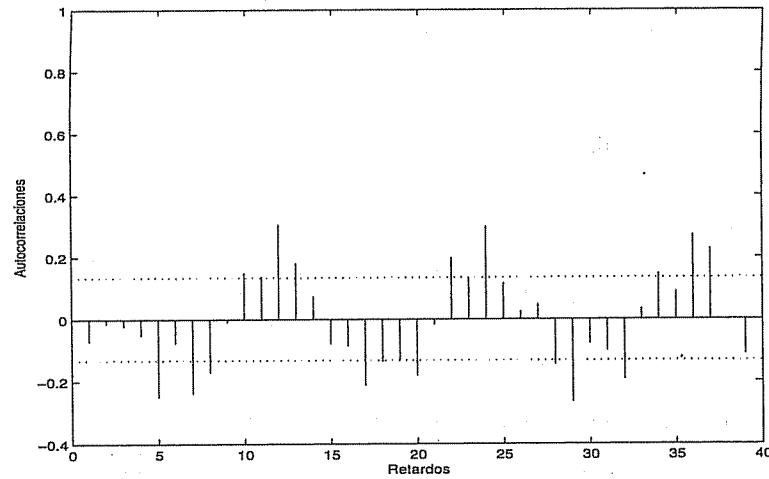
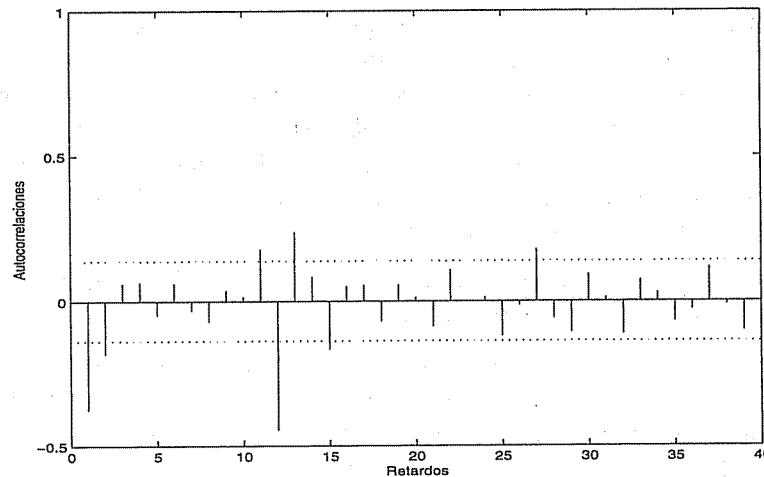


Figura 7.10 Fas estimada de la serie de ozono con una diferencia regular y otra estacional



$$R_7 = \frac{-0.8}{1.64} = -0.49; \quad R_j = 0 \quad j \neq 7$$

Para obtener los coeficientes del proceso, que serán la superposición de ambas partes, particularizando la ecuación (7.13) para  $j = 1$ :

$$\rho_1 = \frac{r_1 - R_7(r_8 + r_6) + R_{14}(r_{15} + r_{13}) + \dots}{1 + 2(r_7R_7 + r_{14}R_{14} + \dots)}$$

y como  $r_j = 0$  para  $j > 1$ , se obtiene que:

$$\rho_1 = r_1 = -0.4.$$

Aplicando la fórmula general se obtiene análogamente que:

$$\rho_j = 0; \quad j = 2, \dots, 5$$

$$\rho_6 = r_6 + R_7(r_{13} + r_1) + R_{14}(r_{20} + r_8) + \dots = R_7r_1 = 0.195$$

$$\rho_7 = r_7 + R_7(r_{14} + r_0) + \dots = R_7 = -0.49$$

$$\rho_8 = r_8 + R_7(r_{15} + r_1) + \dots = R_7r_1 = 0.195$$

$$\rho_j = 0; \quad j > 8$$

con lo que la *fas* tendrá el aspecto de la figura 7.6 (d).

---

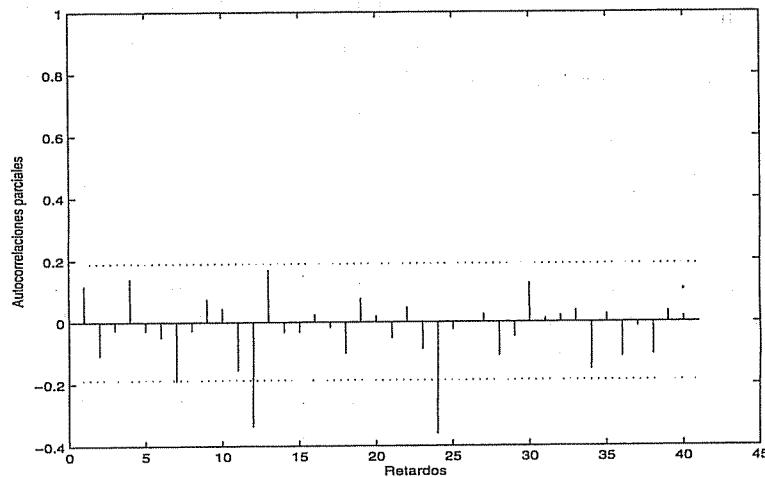
## 7.5. Función de autocorrelación parcial

La función de autocorrelación parcial de un proceso estacional multiplicativo es compleja, porque es una superposición no sólo de las funciones de autocorrelación parciales de las partes regular y estacional, (7.11) y (7.12), sino que en la *interacción* alrededor de los retardos estacionales aparece la estructura de la parte regular, como en la *fas*, pero ahora a un lado aparece la función de autocorrelación parcial y, al otro, la función de autocorrelación simple. En concreto, puede demostrarse, Hamilton y Watts (1978), que la función de autocorrelación parcial de un proceso estacional tiene la estructura siguiente:

1. En los primeros retardos aparecerá la *fap* de la estructura regular y en los estacionales la *fap* de la estacional.
2. A la derecha de cada coeficiente estacional (retardos  $js+1, js+2\dots$ ) aparecerá la *fap* de la parte regular. Si el coeficiente estacional es positivo, la *fap* regular aparece invertida, mientras que si éste es negativo la *fap* aparece con su signo.
3. A la izquierda de los coeficientes estacionales (retardos  $js-1, js-2$ ), observaremos la función de autocorrelación simple de la parte regular.

Se comprueba que en modelos estacionales los coeficientes de correlación parcial para los retardos estacionales suelen ser de magnitud mucho menor que la que correspondería a una estructura estacional pura.

**Figura 7.11 Función de autocorrelación parcial de la primera diferencia estacional de la serie de lluvias en Santiago de Compostela. Datos mensuales desde 1/88 a 12/97**



Esta complejidad recomienda fijarse sólo en los retardos iniciales 1, 2, 3, 4,... para identificar la estructura regular, y en los retardos estacionales,  $s, 2s, 3s, \dots$  para identificar la estructura estacional.

---

### Ejemplo 7.3 (continuación)

La figura 7.11 presenta la función de autocorrelación parcial de los datos de las lluvias en Santiago. Se observa en la parte regular que la estructura es AR o MA de orden bajo, probablemente uno, y en los retardos estacionales se observan dos coeficientes altos y significativos en 12 y 24 lo que sugiere un AR(2) para la parte estacional. Las interacciones explican los coeficientes relativamente altos alrededor del retardo 12.

---

## 7.6. Generalizaciones y lecturas complementarias

El modelo multiplicativo puede generalizarse fácilmente para más de un tipo de estacionalidad, pero la estimación de estos modelos requiere programas especiales ya que con los programas comerciales habituales suponen un único tipo de estacionalidad.

Los modelos multiplicativos explicados en este capítulo, aunque útiles

en la práctica, no son adecuados para todas las series y existen situaciones donde la autocorrelación es distinta en los diferentes instantes que forman el periodo estacional. Por ejemplo, con datos mensuales podemos encontrarnos que existe muy poca autocorrelación entre algunos meses y muy alta en otros. En estos casos debemos modelar cada mes independientemente y agregar después la estructura de dependencia mes a mes, como hemos comentado en la sección 7.2. Esto implica las dos etapas: (1) construir un modelo estacional para los eneros, otro para los febreros, etc.; (2) tomar los residuos de estos modelos en secuencia temporal y construir con ellos un modelo ARIMA regular. Ambas estructuras definen el modelo completo para la serie. Estos modelos se denominan *periódicos*, y el lector interesado puede acudir a Franses (1998).

Un campo no estudiado en este libro y que es importante en las aplicaciones económicas es la desestacionalización de series. Una excelente revisión del tema se encuentra en Gómez y Maravall (2001).

---

### Ejercicios 7

- 7.1 Calcular los coeficientes estacionales para una serie cuatrimestral que sigue el modelo  $z_t = 10 + \cos(\pi t/2 + \pi/8) + a_t$ .
  - 7.2 Demostrar que la serie del ejercicio anterior puede modelarse con  $\nabla_4 z_t = (1 - B^4)a_t$ .
  - 7.3 Supongamos que una serie mensual sigue el modelo  $z_t = 30 + \cos(\pi t/6 + \pi/8) + V_t + a_t$ , donde  $a_t$  es un proceso de ruido blanco con varianza  $\sigma_a^2$  y el proceso  $V_t$  verifica  $V_t = V_{t-12} + \epsilon_t$ , donde  $\epsilon_t$  es un proceso de ruido blanco con varianza  $\sigma_\epsilon^2$ . Demuestre que esta serie sigue el modelo ARIMA  $\nabla_{12} z_t = (1 - \Theta B^{12})a_t$ , donde  $\theta \leq 1$ . (Sugerencia: demuestre que el proceso  $\epsilon_t + a_t - a_{t-12}$  tiene estructura de MA(1)<sub>12</sub> y que la autocorrelación de orden doce es  $-\sigma_a^2/(\sigma_\epsilon^2 + 2\sigma_a^2)$ ).
  - 7.4 Calcular la función de autocorrelación teórica del proceso  $(1 - .4B)w_t = (1 + .5B^{12})$ .
  - 7.5 Calcular la función de autocorrelación teórica del proceso  $(1 - .4B)(1 - .8B^{12})w_t = a_t$ .
  - 7.6 Calcular la función de autocorrelación teórica del proceso  $w_t = (1 - \theta B)(1 - \Theta B^{12})a_t$  y compararlo con la del proceso no multiplicativo  $w_t = (1 - \theta B - \Theta B^{12})a_t$ .
-

### Apéndice 7.1: Un contraste simple de igualdad de la estructura estacional

Sea  $r_{ij}$  el coeficiente de autocorrelación de orden  $j$  calculado para el mes  $i$ . En la hipótesis de normalidad este coeficiente tiene una distribución asintóticamente normal, con media  $\rho_{ij}$  y varianza  $\sigma_{ij}^2$ . Podemos escribir,

$$\begin{aligned} \sum \sum (r_{ij} - \bar{r})^2 &= \sum \sum (r_{ij} - \bar{r}_{i\cdot})^2 + \\ &+ \sum \sum (r_{ij} - \bar{r}_{\cdot j})^2 + \sum \sum (r_{ij} - r_{\cdot j} - r_{i\cdot} + \bar{r})^2 \end{aligned}$$

En esta descomposición las observaciones no son independientes, ya que las estimaciones de los coeficientes de correlación son dependientes entre sí. Por otro lado, sabemos que los coeficientes de autocorrelación de distintos retardos son distintos. Para comprobar si son distintos los diferentes meses, podemos hacer un contraste aproximado mediante el estadístico

$$F = \frac{\sum \sum (r_{ij} - \bar{r}_{i\cdot})^2}{\sum \sum (r_{ij} - r_{\cdot j} - r_{i\cdot} + \bar{r})^2}$$

Bajo la hipótesis de independencia este estadístico se distribuirá aproximadamente como una  $F$  con  $I-1$  y  $(I-1)(J-1)$  grados de libertad. Cuando existe dependencia esto es sólo una aproximación que puede indicarnos si vale la pena construir modelos más complejos.

### Apéndice 7.2: Generación de series estacionales

Vamos a generar un modelo estacional mediante el programa EViews. Utilizaremos el modelo de líneas aéreas:

$$(1 - B) (1 - B^{12}) z_t = (1 - \theta B) (1 - \Theta B^{12}) a_t$$

es decir, un ARIMA(0, 1, 1)  $\times$  (0, 1, 1)<sub>12</sub> o un doble IMA para la parte regular y la estacional. Para generar una serie a partir de este modelo tomamos por ejemplo,  $\theta = -0.7$  y  $\Theta = -0.8$ . Entonces podemos escribir:

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + a_t + 0.7a_{t-1} + 0.8a_{t-12} + 0.56a_{t-13}$$

y procedemos como sigue:

1. Genr→x=0. (define la serie como 0).
2. Genr→ Sample: 14:400 →  $x = x(-1) + x(-12) - x(-13) + a + 0.7 * a(-1) + 0.8 * a(-12) + 0.56 * a(-13)$ . (genera los valores de la serie).

El proceso de generación asume que  $x_1 = 0$ . En el gráfico de la serie (*View→Line Graph*) debe apreciarse el componente estacional, y también en

## Análisis de series temporales

---

el correlograma: la función de autocorrelación estimada debe decaer lentamente, de manera lineal, y la función de autocorrelación parcial debe tener un primer retardo muy grande, cercano a 1 y el resto pequeños, salvo en los retardos 12, 13, 24, 25 y alrededores. Definimos el proceso primera diferencia de éste,  $y_t = x_t - x_{t-1}$ . Para ello, en el menú del workfile:

$$dx = d(x, 1) \quad Sample : 1 : 400$$

la serie  $dx$  que obtenemos es claramente estacional, por lo que es no estacionaria. En su correlograma (View→Correlogram), veremos cómo la fas y la fap tienen muy marcadas las autocorrelaciones de orden 12, 24, etc... Al tomar una diferencia estacional:

$$ddx = d(x, 1, 12) \quad Sample : 1 : 400$$

la serie ya es estacionaria. El correlograma tiene picos significativos en los valores iniciales y en los retardos 12, 13, etc...

---

## 8. Predicción con modelos ARIMA



Andrei Nikolaevich Kolmogorov (1903-1987)

Matemático ruso, fundador del cálculo de probabilidades moderno. Desde muy joven mostró una extraordinaria aptitud para las matemáticas. Pionero en muchas ramas de la matemática, estableció las bases rigurosas para el estudio de los procesos estocásticos y desarrolló la teoría de predicción para procesos estacionarios. En su madurez, dedicó mucho esfuerzo a la educación de niños superdotados.

### 8.1. Introducción

En este capítulo vamos a estudiar la predicción con un modelo ARIMA conocido. Definiremos los predictores óptimos como aquellos que minimizan en promedio los errores de predicción al cuadrado. Estos predictores se obtienen calculando los valores esperados de las observaciones futuras, condicionadas a los datos observados. Veremos que la función de predicción de un modelo ARIMA tiene una estructura simple: los operadores no estacionarios, es decir, las diferencias y la constante si existe, determinan la predicción a largo plazo, mientras que los operadores estacionarios, AR y MA, determinan la predicción a corto plazo.

Las predicciones tienen poca utilidad sin una medida de su precisión, y veremos cómo obtener la distribución de los errores de predicción y cómo calcular intervalos de confianza de las predicciones. Finalmente, estudiaremos cómo revisar las predicciones cuando recibimos nueva información.

## 8.2. La ecuación de predicción de un modelo ARIMA

### 8.2.1. La esperanza condicionada como predictor óptimo

Supongamos que se ha observado una realización de longitud  $T$  de una serie temporal  $\mathbf{z}_T = (z_1, \dots, z_T)$ , y se desea prever un valor futuro  $k > 0$  periodos adelante,  $z_{T+k}$ . Llamemos  $\hat{z}_T(k)$  a un predictor de  $z_{T+k}$  obtenido como función de los  $T$  valores observados, es decir, con *origen* de la predicción en  $T$ , y con *horizonte* de la predicción  $k$ . En particular, en este capítulo vamos a estudiar predictores lineales, que son aquellos que se construyen como una combinación lineal de los datos disponibles:

$$\hat{z}_T(k) = \alpha_1 z_T + \alpha_2 z_{T-1} + \dots + \alpha_T z_1.$$

Un predictor quedará definido al indicar las constantes  $\alpha_1, \dots, \alpha_T$  utilizadas para construirle. Naturalmente, queremos que la predicción esté lo más cerca posible del verdadero valor que queremos prever. Vamos a llamar  $e_T(k)$  al error de predicción de este predictor, dado por:

$$e_T(k) = z_{T+k} - \hat{z}_T(k),$$

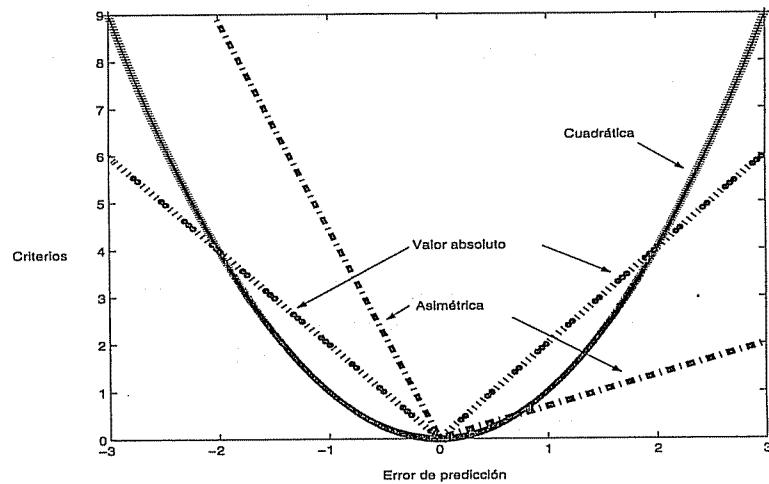
y queremos que este error sea lo más pequeño posible. Como la variable  $z_{T+k}$  es desconocida, es imposible conocer a priori el error que tendremos al predecirla. Sin embargo, si conocemos su distribución de probabilidad, es decir, sus valores posibles y sus probabilidades, definido un predictor podemos calcular la probabilidad de que su error esté entre unos valores dados. Para comparar predictores especificaremos el objetivo deseado mediante un criterio y seleccionaremos el mejor predictor de acuerdo con este criterio.

Si el objetivo es obtener pequeños errores de predicción y valoramos igual los errores de predicción positivos que los negativos, o predecir por exceso que por defecto, prescindiremos del signo del error tomando como criterio minimizar el valor esperado de una *función de pérdida* simétrica del error, como  $l(e_{T+k}) = ce_T^2(k)$ , o  $l(e_T(k)) = c |e_T(k)|$ , donde  $c$  es una constante. Sin embargo, si preferimos los errores en una dirección, errores por defecto que por exceso (o al revés), minimizaremos el valor promedio de una función de pérdida asimétrica del error, que tenga en cuenta nuestros objetivos. Por ejemplo:

$$l(e_T(k)) = \begin{cases} c_1 |e_T(k)|, & \text{si } e_T(k) \geq 0 \\ c_2 |e_T(k)|, & \text{si } e_T(k) \leq 0 \end{cases} \quad (8.1)$$

para ciertos  $c_1$  y  $c_2$  positivos. En este caso, cuanto mayor sea la relación  $c_2/c_1$ , más penalizaremos los errores por defecto.

Figura 8.1 Tres funciones del error de predicción que pueden considerarse. La función cuadrática, la de valor absoluto y una asimétrica



La figura 8.1 compara las tres funciones de pérdida que hemos comentado. Se han representado las dos funciones simétricas:

$$l_1(e_T(k)) = e_T^2(k), \quad y \quad l_2(e_T(k)) = 2|e_T(k)|,$$

y la asimétrica (8.1) con  $c_1 = 2/3$  y  $c_2 = 9/2$ . Comparando las dos funciones simétricas, vemos que la cuadrática da menos importancia a errores pequeños que la función de valores absolutos, mientras que penaliza más los errores grandes. El punto de corte de ambas funciones depende de las constantes utilizadas para definirlas y en este ejemplo los errores cuadráticos pesan menos que los absolutos si son menores que 2. La función asimétrica definida da muy poco peso a los errores positivos, pero penaliza mucho los negativos.

La función de pérdida más utilizada es la cuadrática, que conduce al criterio de minimizar el error cuadrático medio de predicción (ECMP) de  $z_{T+k}$  dada la información  $z_T$ . Tendremos que minimizar:

$$ECMP(z_{T+k}|z_T) = E [(z_{T+k} - \hat{z}_T(k))^2 | z_T] = E [e_T^2(k) | z_T]$$

donde la esperanza se toma respecto a la distribución de la variable  $z_{T+k}$  condicionada a los valores observados  $z_T$ . Vamos a demostrar que el predictor que minimiza este error cuadrático medio es la esperanza de la variable  $z_{T+k}$  condicionada a la información disponible. Por tanto, si

podemos calcular esta esperanza, obtendremos el predictor óptimo sin necesidad de conocer la distribución condicionada completa.

Para demostrarlo, llamemos:

$$\mu_{T+k|T} = E [z_{T+k}|z_T],$$

a la media de esta distribución condicionada. Restando y sumando  $\mu_{T+k|T}$  en la expresión de  $ECMP(z_{T+k})$  y desarrollando el cuadrado tenemos que:

$$ECMP(z_{T+k}|z_T) = E [(z_{T+k} - \mu_{T+k|T})^2|z_T] + \\ + E [(\mu_{T+k|T} - \hat{z}_T(k))^2|z_T] \quad (8.2)$$

ya que el doble producto se anula. En efecto, el término  $\mu_{T+k|T} - \hat{z}_T(k)$  es una constante, ya que  $\hat{z}_T(k)$  es función de los valores pasados y estamos condicionando a ellos. Por tanto, saldrá fuera de la esperanza, y:

$$E [(\mu_{T+k|T} - \hat{z}_T(k))(z_{T+k} - \mu_{T+k|T})|z_T] = \\ = (\mu_{T+k|T} - \hat{z}_T(k))E [(z_{T+k} - \mu_{T+k|T})|z_T] = 0,$$

con lo que obtenemos (8.2). Esta expresión puede escribirse:

$$ECMP(z_{T+k}|z_T) = var(z_{T+k}|z_T) + E [(\mu_{T+k|T} - \hat{z}_T(k))^2|z_T].$$

Como el primer término de esta expresión no depende del predictor, minimizaremos el  $ECMP$  del predictor haciendo cero el segundo. Esto ocurrirá si tomamos:

$$\hat{z}_T(k) = \mu_{T+k|T} = E [z_{T+k}|z_T].$$

Hemos demostrado que el predictor que minimiza el error cuadrático medio de predicción de un valor futuro se obtiene tomando su esperanza condicionada a los datos observados. Podríamos preguntarnos por el predictor que minimiza este error sea cual sea la secuencia observada, es decir, no sólo promediando para los posibles valores de  $z_{T+k}$ , sino también promediando para todas las posibles series temporales  $z_T$  que pueden observarse. En el apéndice 8.1 se demuestra que este predictor es de nuevo  $\mu_{T+k|T}$ , la esperanza condicionada a la información disponible.

---

### Ejemplo 8.1

Supongamos que disponemos de 50 datos generados por un proceso AR(1):  $z_t = 10 + 0.5z_{t-1} + a_t$ . Los dos últimos valores observados corresponden a

## 8. Predicción con modelos ARIMA

---

los instantes  $t = 50$ , y  $t = 49$ , y son  $z_{50} = 18$ , y  $z_{49} = 15$ . Queremos calcular predicciones para los dos periodos siguientes,  $t = 51$  y  $t = 52$ .

La primera observación que queremos prever es la  $z_{51}$ . Su expresión es:

$$z_{51} = 10 + 0.5z_{50} + a_{51}$$

y su esperanza, condicionada a los valores observados, es:

$$\hat{z}_{50}(1) = 10 + 0.5(18) = 19.$$

Para  $t = 52$ , la expresión de la variable es:

$$z_{52} = 10 + 0.5z_{51} + a_{52}$$

y tomando esperanzas condicionadas a los datos disponibles:

$$\hat{z}_{50}(2) = 10 + 0.5\hat{z}_{50}(1) = 10 + 0.5(19) = 19.5.$$

---

### 8.2.2. Cálculo de las predicciones

Supongamos que se dispone de una realización de tamaño  $T$ ,  $\mathbf{z}_T = (z_1, \dots, z_T)$ , de un proceso ARIMA  $(p, d, q)$ ,  $\phi(B)\nabla^d z_t = c + \theta(B)a_t$ , con parámetros conocidos. Vamos a ver cómo aplicar el resultado anterior para calcular las predicciones. Conocidos los parámetros podemos obtener todas las innovaciones  $a_t$  fijando unos valores iniciales. Por ejemplo, si el proceso es ARMA(1,1), las innovaciones,  $a_2, \dots, a_T$ , se calculan recursivamente mediante la ecuación:

$$a_t = z_t - c - \phi z_{t-1} + \theta a_{t-1}, \quad t = 2, \dots, T$$

La innovación para  $t = 1$  viene dada por:

$$a_1 = z_1 - c - \phi z_0 + \theta a_0$$

y ni  $z_0$  ni  $a_0$  son conocidas, por lo que no podemos obtener  $a_1$ . Podemos sustituirla por su esperanza  $E(a_1) = 0$ , y calcular las demás innovaciones con esta condición inicial. En consecuencia, supondremos en adelante que tanto las observaciones como las innovaciones son conocidas hasta el instante  $T$ .

De acuerdo con el resultado de la sección anterior, la predicción que minimiza el error cuadrático medio de  $z_{T+k}$ , que llamaremos para simplificar en adelante la predicción óptima de  $z_{T+k}$ , es la esperanza de la variable condicionada a los valores observados. Definamos:

$$\begin{aligned}\widehat{z}_T(j) &= E[z_{T+j} | \mathbf{z}_T] \quad j = 1, 2, \dots \\ \widehat{a}_T(j) &= E[a_{t+j} | \mathbf{z}_T] \quad j = 1, 2, \dots\end{aligned}$$

donde el subíndice  $T$  representa el *origen de la predicción*, que suponemos fijo, y  $j$  el *horizonte* de la misma, que irá cambiando para generar predicciones de distintas variables futuras desde el origen  $T$ . Llamando  $\varphi_h(B) = \phi_p(B) \nabla^d$  al operador de orden  $h = p+d$  que se obtiene multiplicando el operador AR( $p$ ) estacionario y las diferencias, la expresión de la variable  $z_{T+k}$  es:

$$z_{T+k} = c + \varphi_1 z_{t+k-1} + \dots + \varphi_h z_{T+k-h} + a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q}. \quad (8.3)$$

Tomando esperanzas condicionadas a  $\mathbf{z}_T$  en todos los términos de esta expresión, se obtiene:

$$\begin{aligned}\widehat{z}_T(k) &= c + \varphi_1 \widehat{z}_T(k-1) + \dots + \varphi_h \widehat{z}_T(k-h) \\ &\quad - \theta_1 \widehat{a}_T(k-1) - \dots - \widehat{a}_T(k-q).\end{aligned} \quad (8.4)$$

En esta ecuación algunas esperanzas se aplican a variables observadas y otras a no observadas. Cuando  $i > 0$  la expresión  $\widehat{z}_T(i)$  es la esperanza condicionada de la variable  $z_{T+i}$  que no se ha observado todavía. Sin embargo, cuando  $i \leq 0$ ,  $\widehat{z}_T(i)$  es la esperanza condicionada de la variable  $z_{T-|i|}$ , que ya se ha observado y es conocida, por lo que esta esperanza coincidirá con la observación y  $\widehat{z}_T(-|i|) = z_{T-|i|}$ . Con relación a las innovaciones, la esperanza de las innovaciones futuras condicionada a la historia de la serie es igual a su esperanza absoluta, cero, ya que la variable  $a_{T+i}$  es independiente de  $\mathbf{z}_T$ , y concluimos que para  $i > 0$ , las  $\widehat{a}_T(i)$  serán cero. Cuando  $i \leq 0$ , las innovaciones  $a_{T-|i|}$  son conocidas, por lo que sus esperanzas coinciden con los valores observados y  $\widehat{a}_T(-|i|) = a_{T-|i|}$ . De esta manera, podemos calcular recursivamente las predicciones, comenzando con  $j = 1$  y continuando con  $j = 2, 3, \dots$ , hasta el horizonte deseado.

Observemos que para  $k = 1$ , restando (8.3) de (8.4) se obtiene que:

$$a_{T+1} = z_{T+1} - \widehat{z}_T(1),$$

y las innovaciones pueden interpretarse como los errores de predicción a un periodo por delante cuando conocemos los parámetros del modelo.

La ecuación (8.4) indica que después de  $q$  valores iniciales los términos de media móvil desaparecerán, y la predicción quedará determinada exclusivamente por la parte autorregresiva del modelo. En efecto, para  $k > q$  todas las innovaciones que aparecen en la ecuación de predicción son no

observadas, sus esperanzas serán cero y desaparecerán de la ecuación de predicción. Las predicciones de (8.4) para  $k > q$  satisfacen la ecuación:

$$\hat{z}_T(k) = c + \varphi_1 \hat{z}_T(k-1) + \dots + \varphi_h \hat{z}_T(k-h). \quad (8.5)$$

Vamos a reescribir esta expresión utilizando el operador de retardo, para simplificar su manipulación. En esta ecuación el origen de la predicción siempre es el mismo, pero cambia el horizonte. Vamos a convenir que ahora el operador  $B$  actúa sobre el horizonte de la previsión, de manera que:

$$B\hat{z}_T(k) = \hat{z}_T(k-1)$$

mientras que  $B$  no afecta al origen de la predicción,  $T$ , que es fijo. Con esta notación, la ecuación (8.5) se escribe:

$$(1 - \varphi_1 B - \dots - \varphi_h B^h) \hat{z}_T(k) = \phi(B) \nabla^d \hat{z}_T(k) = c, \quad k > q. \quad (8.6)$$

Esta ecuación se denomina la *ecuación de predicción final* porque establece cómo están ligadas las predicciones para horizontes altos, cuando ha desaparecido la parte de media móvil. Observemos que la relación entre las predicciones es similar a la existente entre las autocorrelaciones de un proceso estacionario ARMA, aunque en las predicciones además del operador estacionario  $\phi(B)$  aparece también el operador no estacionario  $\nabla^d$ , y la ecuación no está igualada a cero, sino a la constante  $c$ .

### Ejemplo 8.2

Supongamos el proceso AR(1)  $z_t = c + \phi_1 z_{t-1} + a_t$ . La ecuación de predicción un periodo adelante es:

$$\hat{z}_T(1) = c + \phi_1 z_T,$$

y para dos periodos:

$$\hat{z}_T(2) = c + \phi_1 \hat{z}_T(1) = c(1 + \phi_1) + \phi_1^2 z_T.$$

Estas ecuaciones se obtuvieron para unos valores particulares de los parámetros en el ejemplo 8.1. Generalizando, para cualquier periodo  $k > 0$  tenemos:

$$\hat{z}_T(k) = c + \phi_1 \hat{z}_T(k-1) = c(1 + \phi_1 + \dots + \phi_1^k) + \phi_1^k z_T.$$

Para  $k$  grande como  $|\phi_1| < 1$  el término  $\phi_1^k z_T$  será próximo a cero y la predicción será  $c(1 + \phi_1 + \dots + \phi_1^k) = c/(1 - \phi_1)$ , que es la media del proceso. Veremos en la sección siguiente que para cualquier proceso estacionario ARMA( $p, q$ ), la predicción para un horizonte  $k$  grande es la media del proceso,

$\mu = c/(1 - \phi_1 - \dots - \phi_p)$ . Como caso particular, si  $c = 0$  la predicción a largo plazo será cero.

### Ejemplo 8.3

Supongamos el paseo aleatorio:  $\nabla z_t = c + a_t$ . La predicción a un paso con este modelo se obtiene con:

$$z_t = z_{t-1} + c + a_t.$$

Tomando esperanzas condicionadas a los datos hasta  $T$ , tenemos, para  $T + 1$ :

$$\hat{z}_T(1) = c + z_T,$$

y para  $T + 2$ :

$$\hat{z}_T(2) = c + \hat{z}_T(1) = 2c + z_T,$$

y para un horizonte  $k$  cualquiera:

$$\hat{z}_T(k) = c + \hat{z}_T(k-1) = kc + z_T.$$

Como la predicción del siguiente periodo se obtiene sumando siempre la misma constante,  $c$ , concluimos que las predicciones seguirán una línea recta con pendiente  $c$ . Si  $c = 0$  las predicciones de todos los periodos son iguales al último valor observado, y seguirán una recta horizontal. Vemos que la constante determina la pendiente de la ecuación de predicción. Sin embargo, el origen de la predicción es siempre el último dato disponible.

### Ejemplo 8.4

Supongamos el modelo  $\nabla z_t = (1 - \theta B)a_t$ . Dadas las observaciones hasta el instante  $T$ , la siguiente observación se genera por:

$$z_{T+1} = z_T + a_{T+1} - \theta a_T,$$

y tomando esperanzas condicionadas, su predicción será:

$$\hat{z}_T(1) = z_T - \theta a_T.$$

Para dos periodos adelante, como  $z_{T+2} = z_{T+1} + a_{T+2} - \theta a_{T+1}$ , tendremos que

$$\hat{z}_T(2) = \hat{z}_T(1)$$

y, en general

$$\hat{z}_T(k) = \hat{z}_T(k-1), \quad k \geq 2.$$

Para ver la relación entre estas predicciones y las generadas por el alisado exponencial simple en el capítulo 2, podemos invertir el operador  $(1 - \theta B)$  y escribir:

## 8. Predicción con modelos ARIMA

---

$$\nabla(1 - \theta B)^{-1} = 1 - (1 - \theta)B - (1 - \theta)\theta B^2 - (1 - \theta)\theta^2 B^3 - \dots$$

con lo que el proceso es:

$$z_{T+1} = (1 - \theta) \sum_{j=0}^{\infty} \theta^j z_{T-j} + a_{T+1}.$$

Esta ecuación indica que el valor futuro,  $z_{T+1}$ , es la suma de la innovación y un promedio de los valores anteriores con pesos que decrecen geométricamente. La predicción será:

$$\hat{z}_T(1) = (1 - \theta) \sum_{j=0}^{\infty} \theta^j z_{T-j}$$

y es fácil comprobar (véase ejercicio 8.7) que  $\hat{z}_T(2) = \hat{z}_T(1)$ .

### Ejemplo 8.5

Consideremos el siguiente modelo estacional para datos mensuales  $\nabla_{12}z_t = (1 - 0.8B^{12})a_t$ . La predicción para cualquier mes del año siguiente,  $j$ , donde  $j = 1, 2, \dots, 12$ , se obtiene tomando esperanzas condicionadas a los datos en:

$$z_{t+j} = z_{t+j-12} + a_{t+j} - 0.8a_{t+j-12}.$$

Para  $j \leq 12$ , el resultado es:

$$\hat{z}_T(j) = z_{T+j-12} - 0.8a_{T+j-12-1}, \quad j = 1, \dots, 12$$

y para el segundo año, como las perturbaciones  $a_t$  ya no serán observadas:

$$\hat{z}_T(12 + j) = \hat{z}_T(j).$$

Esta ecuación indica que las predicciones para todos los meses del segundo año serán idénticas a las del primero. Lo mismo ocurrirá para los años posteriores. Por tanto, la ecuación de predicción está formada por 12 coeficientes que se repiten año tras año. Para interpretarlos, sea:

$$\bar{z}_T(12) = \frac{1}{12} \sum_{j=1}^{12} \hat{z}_T(j) = \frac{1}{12} \sum_{j=1}^{12} z_{T+j-12} - 0.8 \frac{1}{12} \sum_{j=1}^{12} a_{T+j-12-1}$$

la media de las predicciones para los doce meses del primer año. Definamos:

$$S_T(j) = \hat{z}_T(j) - \bar{z}_T(12)$$

como coeficientes estacionales. Por construcción suman cero, y las predicciones

pueden escribirse:

$$\hat{z}_T(12k + j) = \bar{z}_T(12) + S_T(j), \quad j = 1, \dots, 12; k = 0, 1, \dots$$

La predicción con este modelo es la suma de un nivel constante, estimado por  $\bar{z}_T(12)$ , más el coeficiente estacional del mes. Observemos que el nivel se estima aproximadamente como el nivel medio de los últimos doce meses, con un factor de corrección que será pequeño, ya que  $\sum_{j=1}^{12} a_{T+j-12-1}$  será próximo a cero al tener los errores media cero.

---

### 8.3. Interpretación de las predicciones

#### 8.3.1. Procesos no estacionales

Hemos visto que para procesos no estacionarios la constante es muy importante y esta propiedad va a trasladarse a las predicciones. Para modelos ARIMA no estacionales hemos visto en la sección anterior que las predicciones efectuadas desde el origen  $T$  con horizonte  $k$  satisfacen, para  $k > q$ , la ecuación de predicción final (8.6).

Supongamos inicialmente el caso más simple de  $c = 0$ . Entonces, la ecuación que deben satisfacer las predicciones es:

$$\phi(B) \nabla^d \hat{z}_T(k) = 0, \quad k > q. \quad (8.7)$$

Hemos visto en el apéndice 4.2 que la solución de una ecuación en diferencias que puede expresarse como producto de operadores primos (sin raíces en común) es la suma de las soluciones correspondientes a cada uno de estos operadores. Como los polinomios  $\phi(B)$  y  $\nabla^d$  de la ecuación (8.7) son primos (no tienen ninguna raíz en común), la solución de esta ecuación será la suma de las soluciones correspondientes al operador estacionario,  $\phi(B)$ , y al no estacionario,  $\nabla^d$ . Podemos escribir:

$$\hat{z}_T(k) = P_T(k) + t_T(k) \quad (8.8)$$

donde estos componentes deben satisfacer:

$$\nabla^d P_T(k) = 0, \quad (8.9)$$

$$\phi(B) t_T(k) = 0. \quad (8.10)$$

Es inmediato comprobar que (8.8) con las condiciones (8.9) y (8.10) es la solución de (8.7). Basta sustituir la solución en la ecuación y comprobar que la verifica. La solución generada por el operador no esta-

cionario,  $P_T(k)$ , representa el *componente permanente* de la predicción y la solución del operador estacionario,  $t_T(k)$ , es el *componente transitorio*. Como la ecuación (8.7) se verifica para  $k > q$ , pero su solución depende de  $p + d$  valores iniciales, el orden máximo del operador  $B$  en la ecuación, según vimos en el apéndice 4.2 la ecuación (8.8) será válida para  $k > \max(0, q - (p + d))$ . Es decir, si  $q < (p + d)$  será válida para todas las predicciones, mientras que si  $q > (p + d)$  será válida desde  $k = q - (p + d) + 1$ .

Puede comprobarse que la solución de (8.9) es de la forma:

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)}k + \dots + \beta_{d-1}^{(T)}k^{(d-1)} \quad (8.11)$$

donde los parámetros  $\beta_j^{(T)}$  son constantes a determinar, que dependen del origen  $T$ , y se calculan utilizando las últimas observaciones disponibles. La comprobación es simple tomando diferencias en (8.11) y viendo que se obtiene (8.9). Como habitualmente  $d \leq 2$ , el componente permanente será una constante si  $d = 1$  y una tendencia lineal si  $d = 2$ . Los coeficientes de estas tendencias dependen del origen de la predicción y van variando en el tiempo.

El componente transitorio, solución de (8.10), proporciona las predicciones a corto plazo y, como obtuvimos en el apéndice 4.2, viene dado por:

$$t_T(k) = \sum_{i=1}^P A_i G_i^k$$

siendo  $G_i^{-1}$  las raíces del operador AR estacionario. Este componente tiende a cero con  $k$ , por ser todas las raíces  $|G_i| \leq 1$ , lo que justifica su nombre de componente transitorio.

En consecuencia, la ecuación de predicción final es válida para  $k > \max(0, q - (p + d))$  y está formada por la suma de dos componentes:

1. Un componente permanente, o de tendencia prevista, que es un polinomio de orden  $d - 1$ , con coeficientes que dependen del origen de la predicción.
2. Un componente transitorio, o de predicción a corto plazo, que es una mezcla de términos cíclicos y exponenciales (según las raíces de  $\phi(B) = 0$ ) y que tiende a cero con  $k$ .

Para determinar los coeficientes  $\beta_j^{(k)}$  de la ecuación de la tendencia prevista, el método más simple es generar predicciones con (8.4) para un horizonte  $k$  grande hasta que:

- a. Si  $d = 1$ , sean constantes. Entonces esa constante es  $\beta_0^{(T)}$ .
- b. Si  $d = 2$ , se verifique que la diferencia entre predicciones,  $\hat{z}_T(k) - \hat{z}_T(k-1)$  sea constante. Las predicciones están en una recta, y su pendiente es  $\beta_1^{(T)} = \hat{z}_T(k) - \hat{z}_T(k-1)$ , que va adaptándose en función de las observaciones disponibles.

Hemos supuesto que la serie estacionaria tiene media cero. Si no es así, la ecuación de predicción final es:

$$\phi(B) \nabla^d \hat{z}_T(k) = c. \quad (8.12)$$

Hemos visto que la solución de una ecuación en diferencias que no es homogénea es la solución de la homogénea (la ecuación igualada a cero), más una solución particular. Una solución particular de (8.12) debe tener la propiedad de que al diferenciarla  $d$  veces y aplicarle el término  $\phi(B)$  se obtenga la constante  $c$ . Para que esto ocurra la solución debe ser del tipo  $\beta_d k^d$ , donde el operador de retardo se aplica a  $k$  y  $\beta_d$  es una constante que se obtiene imponiendo la condición de que esta expresión verifique la ecuación, es decir:

$$\phi(B) \nabla^d (\beta_d k^d) = c.$$

Para obtener la constante  $\beta_d$  vamos a comprobar primero que al aplicar el operador  $\nabla^d$  a  $\beta_d k^d$  se obtiene  $d! \beta_d$ , donde  $d! = (d-1)(d-2)\cdots 2$ . Lo comprobaremos para los valores habituales de  $d$ . Si  $d = 1$ :

$$(1 - B) \beta_d k = \beta_d k - \beta_d (k-1) = \beta_d$$

y para  $d = 2$ :

$$\begin{aligned} (1 - B)^2 \beta_d k^2 &= (1 - 2B + B^2) \beta_d k^2 = \\ &= \beta_d k^2 - 2\beta_d (k-1)^2 + \beta_d (k-2)^2 = 2\beta_d. \end{aligned}$$

Como al aplicar  $\phi(B)$  a una constante se obtiene  $\phi(1)$  por esa constante, concluimos que:

$$\phi(B) \nabla^d (\beta_d k^d) = \phi(B) d! \beta_d = \phi(1) d! \beta_d = c$$

de donde resulta:

$$\beta_d = \frac{c}{\phi(1) d!} = \frac{\mu}{d!}, \quad (8.13)$$

ya que la constante  $c$  verifica  $c = \phi(1) \mu$ , siendo  $\mu$  la media de la serie estacionaria.

Como esta solución particular no desaparece con el retardo, la añadiremos al componente permanente (8.11), con lo que podemos escribir:

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)}k + \dots + \beta_d k^d.$$

Esta ecuación es válida en todos los casos, ya que si  $c = 0$ ,  $\beta_d = 0$ , y volvemos a la expresión (8.11). Existe una diferencia básica entre los coeficientes de la tendencia prevista hasta orden  $d - 1$ , que dependen del origen de las predicciones y van cambiando en el tiempo, y el coeficiente  $\beta_d$ , dado por (8.13), que es constante para cualquier origen de predicción, porque sólo depende de la media de la serie estacionaria. Como el comportamiento a largo plazo dependerá del término de orden más alto, las predicciones a largo plazo de un modelo con constante quedan determinadas por esta constante. Por ejemplo, si  $c \neq 0$  y  $d = 1$ , las predicciones a largo plazo desde cualquier horizonte seguirán una línea recta, consistentes con un paseo aleatorio con deriva. Sin embargo, si  $c = 0$  y  $d = 2$  las predicciones también seguirán una línea recta, pero la pendiente se va adaptando con el origen de la predicción.

### Ejemplo 8.6

Vamos a generar predicciones para la serie de población de mayores de 16 años en España que se encuentra en el fichero pobmay16.dat. Estos datos fueron utilizados en el capítulo 2 para construir un modelo determinista y otro de alisado exponencial. En el capítulo 9 comprobaremos que son necesarias dos diferencias para obtener un proceso estacionario y si calculamos la función de autocorrelación de los datos el modelo para la serie estacionaria parece ser un MA(1). Estimado este modelo con el método descrito en el capítulo 10 se obtiene:

$$\nabla^2 z_t = (1 - 0.68B)a_t$$

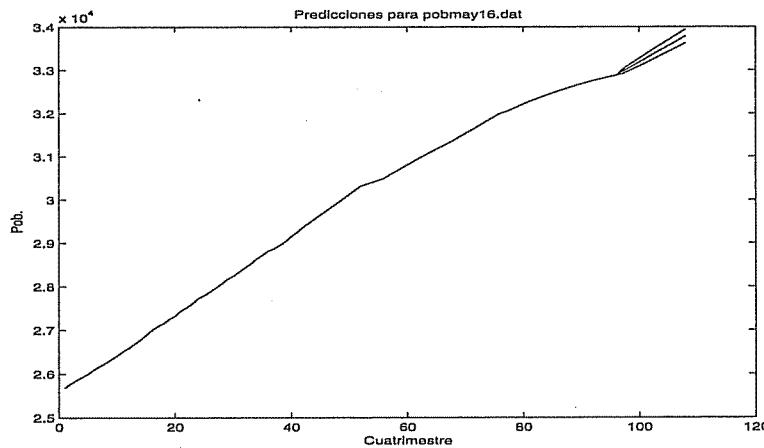
con una suma de cuadrados de los residuos de 19736, bastante menor que las obtenidas en el capítulo 2. Las predicciones generadas por este modelo para los 12 próximos cuatrimestres se dan en la figura 8.2. Las predicciones siguen una recta con pendiente 32.062 personas adicionales cada cuatrimestre.

### Ejemplo 8.7

Vamos a comparar la predicción a largo plazo de un paseo aleatorio con constante,  $\nabla z_t = c + a_t$ , con la del modelo  $\nabla^2 z_t = (1 - \theta B)a_t$ . La predicción del paseo aleatorio es, de acuerdo a lo anterior, una recta con pendiente  $c$  (véase también el ejemplo 8.3), que es la media de la serie estacionaria  $w_t = \nabla z_t$  y se estima con:

$$\hat{c} = \frac{1}{T-1} \sum_{t=2}^{t=T} \nabla z_t = \frac{z_T - z_1}{T-1}.$$

**Figura 8.2 Predicciones para tres años de la población mayor de 16 años en España**



Como:

$$\hat{z}_T(k) = \hat{c} + \hat{z}_T(k-1) = k\hat{c} + z_T,$$

la predicción del crecimiento futuro en un periodo,  $\hat{z}_T(k) - \hat{z}_T(k-1)$ , será igual al crecimiento medio observado en toda la serie  $\hat{c}$ . Observemos que esta pendiente es constante.

El modelo con dos diferencias si  $\theta \rightarrow 1$  será muy parecido al  $\nabla z_t = c + a_t$ , ya que la solución de:

$$\nabla^2 z_t = \nabla a_t$$

es

$$\nabla z_t = c + a_t.$$

Sin embargo, cuando  $\theta$  no es uno, aunque esté próxima a uno, la predicción de ambos modelos puede ser muy diferente a largo plazo. En el modelo con dos diferencias la ecuación de predicción final será también una línea recta, pero con pendiente que va cambiando con el retardo, mientras que en el modelo con una diferencia y constante la pendiente es siempre constante. La predicción para un periodo adelante del modelo con dos diferencias se obtiene de:

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta a_{t-1},$$

y al tomar esperanzas:

$$\hat{z}_T(1) = 2z_T - z_{T-1} - \theta a_T = z_T + \hat{\beta}_T,$$

donde hemos llamado

$$\hat{\beta}_T = \nabla z_T - \theta a_T,$$

## 8. Predicción con modelos ARIMA

---

a la cantidad que sumamos a la última observación, que podemos considerar como la pendiente aparente en el instante  $T$ . Para dos periodos:

$$\hat{z}_T(2) = 2\hat{z}_T(1) - z_T = z_T + 2\hat{\beta}_T,$$

y vemos que la predicción se obtiene sumando dos veces la pendiente aparente, con lo que  $z_T$ ,  $\hat{z}_T(1)$  y  $\hat{z}_T(2)$  estarán en una línea recta. De la misma forma, es fácil comprobar que:

$$\hat{z}_T(k) = z_T + k\hat{\beta}_T,$$

que muestra que todas las predicciones siguen una línea recta con pendiente  $\hat{\beta}_T$ . Observemos que la pendiente va cambiando en el tiempo, ya que depende del último crecimiento observado,  $\nabla z_T$ , y del último error de previsión cometido  $a_T$ . Sustituyendo  $a_T = (1 - \theta B)^{-1} \nabla^2 z_T$  en la definición de  $\hat{\beta}_T$ , podemos escribir:

$$\hat{\beta}_T = \nabla z_T - \theta(1 - \theta B)^{-1} \nabla^2 z_T = (1 - \theta(1 - \theta B)^{-1} \nabla) \nabla z_T,$$

y el operador resultante sobre  $\nabla z_T$  es:

$$1 - \theta(1 - \theta B)^{-1} \nabla = 1 - \theta(1 - (1 - \theta)B - \theta(1 - \theta)B^2 - \dots)$$

con lo que  $\hat{\beta}_T$  puede escribirse:

$$\hat{\beta}_T = (1 - \theta) \sum_{i=0}^{T-1} \theta^i \nabla z_{T-i}.$$

Esta expresión muestra que la pendiente de la ecuación de predicción final es una media ponderada de todos los crecimientos observados,  $\nabla z_{T-i}$ , con pesos que decrecen geométricamente. Observemos que si  $\theta$  es próximo a uno, esta media será próxima a la media aritmética, mientras que si  $\theta$  es próximo a cero la pendiente se estima únicamente con los últimos valores observados. Este ejemplo muestra la mayor flexibilidad del modelo con dos diferencias frente al modelo con una diferencia y constante.

Es interesante comparar estos modelos con uno determinista que genere también predicciones siguiendo una línea recta. Consideremos el modelo lineal determinista:

$$\hat{z}_T(k) = a + b_R(T + k),$$

supongamos para simplificar el análisis que  $T = 5$ , y escribamos  $t = (-2, -1, 0, 1, 2)$  para que  $\bar{t} = 0$ . Llamemos a los datos  $(z_{-2}, z_{-1}, z_0, z_1, z_2)$ . La pendiente de la recta se estima por mínimos cuadrados por:

$$b_R = \frac{-2z_{-2} - z_{-1} + z_1 + 2z_2}{10},$$

y esta expresión puede escribirse como:

$$b_R = .2(z_{-1} - z_{-2}) + .3(z_0 - z_{-1}) + .3(z_1 - z_0) + .2(z_2 - z_1),$$

que es una media ponderada de los crecimientos observados pero dando peso mínimo al último. Como encontramos en la sección 2.2.3 en el caso general la pendiente se escribe:

$$b_R = \sum_{t=2}^T w_t \nabla z_t,$$

donde  $\sum w_t = 1$  y los pesos son simétricos y su valor mínimo corresponde a los crecimientos al principio y al fin de la muestra. Este ejemplo muestra las limitaciones de los modelos deterministas para la predicción, que ya comentamos en el capítulo 2, y la mayor flexibilidad que puede obtenerse tomando diferencias.

---

### 8.3.2. Procesos estacionales

Si el proceso es estacional la descomposición anterior sigue siendo válida: la ecuación de predicción final tendrá un componente permanente, que dependerá de los operadores no estacionarios,  $\nabla^d$ ,  $\nabla_s$  y la media  $\mu$  del proceso estacionario, y un componente transitorio, que englobará el efecto de los operadores AR estacionarios (regular y estacional).

Para separar la tendencia de la estacionalidad en el componente permanente es necesario que los operadores asociados a estos componentes no tengan raíces en común. Como:

$$\nabla_s = (1 - B^s) = (1 + B + B^2 + \dots + B^{s-1})(1 - B)$$

el operador estacional  $(1 - B^s)$  puede interpretarse como la mezcla de dos operadores: el operador diferencia y un operador estacional puro, dado por:

$$S_s(B) = 1 + B + \dots + B^{s-1}, \quad (8.14)$$

que produce la suma de  $s$  términos consecutivos. Por ejemplo, en una serie mensual,  $s = 12$ , y esta operación suma los doce meses del año, por lo que se obtiene un valor libre del efecto estacional. Esta descomposición del operador  $\nabla_s$  ilustra que restar a una observación el valor del mismo mes en el año anterior es lo mismo que sumar a la observación los once meses anteriores, hacer la misma suma para la observación anterior y luego restar estos dos totales consecutivos.

Separando la raíz  $B = 1$  del operador  $\nabla_s$ , el modelo estacional, suponiendo una constante distinta de cero, puede escribirse:

## 8. Predicción con modelos ARIMA

---

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} z_t = c + \theta(B) \Theta(B^s) a_t$$

donde, ahora, los cuatro operadores  $\Phi(B^s)$ ,  $\phi(B)$ ,  $S_s(B)$  y  $\nabla^{d+1}$  no tienen raíces en común. La ecuación de predicción final será:

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} \hat{z}_t(k) = c, \quad (8.15)$$

y será válida para  $k > q+sQ$  y como requiere  $d+s+p+sP$  valores iniciales (orden máximo de  $B$  en el operador de la derecha), puede utilizarse para calcular predicciones para  $k > q + sQ - d - s - p - sP$ . Como ahora los operadores no estacionarios,  $S_s(B)$  y  $\nabla^{d+1}$ , no tienen raíces en común, la componente permanente puede escribirse como suma de las soluciones de cada operador. La solución de la ecuación homogénea es:

$$\hat{z}_t(k) = T_T(k) + S_T(k) + t_T(k),$$

donde  $T_T(k)$  es el componente de tendencia que se obtiene de la ecuación:

$$\nabla^{d+1} T_T(k) = 0,$$

y será un polinomio de grado  $d$  con coeficientes que se van adaptando en el tiempo. El componente estacional,  $S_T(k)$ , es la solución de:

$$S_s(B) S_T(k) = 0 \quad (8.16)$$

que tiene como solución cualquier función de periodo  $s$  con coeficientes que sumen cero. En efecto, una secuencia  $S_T(k)$  es solución de (8.16) si verifica:

$$\sum_{j=1}^s S_T(j) = \sum_{s+1}^{2s} S_T(j) = 0,$$

los coeficientes  $S_T(1), \dots, S_T(s)$  que se obtienen al resolver esta ecuación se denominan coeficientes estacionales. Finalmente, el componente transitorio incluirá las raíces de los polinomios  $\Phi(B^s)$  y  $\phi(B)$  y tendrá la expresión

$$t_T(k) = \sum_{i=1}^{p+P} A_i G_i^k,$$

donde los  $G_i^{-1}$  son las soluciones de las ecuaciones  $\Phi(B^s) = 0$  y  $\phi(B) = 0$ .

Una solución particular de la ecuación (8.15) será  $\beta_{d+1} k^{d+1}$ , donde la constante  $\beta_{d+1}$  viene determinada por la condición:

$$\Phi(1) \phi(1) s(d+1)! \beta_{d+1} = c,$$

ya que el resultado de aplicar el operador  $\nabla^{d+1}$  a  $\beta_{d+1}k^{d+1}$  es  $(d+1)!\beta_{d+1}$  y si aplicamos a una constante el operador  $S_s(B)$  obtenemos  $s$  veces esa constante. Como la media de la serie estacionaria es  $\mu = c/\Phi(1)\phi(1)$  la constante  $\beta_{d+1}$  es:

$$\beta_{d+1} = \frac{c}{\Phi(1)\phi(1)s(d+1)!} = \frac{\mu}{s(d+1)!},$$

que generaliza los resultados del modelo con constante pero sin estacionalidad. Este componente adicional se añade al término de tendencia.

En resumen, la solución general de la ecuación de predicción final de un proceso estacional tendrá tres componentes:

1. Un término de tendencia esperada, que será un polinomio de grado  $d$  con coeficientes que se van adaptando en el tiempo si no hay constante en el modelo y un polinomio de grado  $d+1$  con el coeficiente del término de mayor orden,  $\beta_{d+1}$  determinista y dado por  $\mu/s(d+1)!$ , siendo  $\mu$  la media de la serie estacionaria.
2. Un componente estacional previsto, que cambiará con las condiciones iniciales.
3. Un término transitorio de predicción a corto plazo, que vendrá determinado por las raíces de los operadores AR regulares y estacionales.

La solución general anterior puede utilizarse para obtener predicciones para horizontes  $k > q + sQ - d - s - p - sP$ .

A continuación presentamos el cálculo de estos coeficientes en un modelo muy frecuente en la práctica.

#### 8.3.2.1. El modelo de pasajeros de avión

El modelo ARIMA estacional más utilizado es el llamado modelo de pasajeros de avión:

$$\nabla\nabla_s z_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

cuya ecuación para calcular las predicciones es, para  $k > 0$

$$\begin{aligned}\widehat{z}_t(k) = & \widehat{z}_t(k-1) + \widehat{z}_t(k-12) - \widehat{z}_t(k-13) - \\ & - \theta \widehat{a}_t(k-1) - \Theta \widehat{a}_t(k-12) + \theta \Theta \widehat{a}_t(k-13).\end{aligned}$$

Por otro lado, según los resultados anteriores sabemos que la predicción de este modelo puede escribirse como:

## 8. Predicción con modelos ARIMA

---

$$\hat{z}_t(k) = \beta_0^{(t)} + \beta_1^{(t)}k + S_t^{(k)}. \quad (8.17)$$

Esta ecuación contiene 13 parámetros (recuérdese que  $\sum_{k=1}^{12} S_t^{(k)} = 0$ ), y es la suma de una tendencia lineal –que va cambiando con el origen de la predicción– y de doce coeficientes estacionales, que también se adaptan con el origen de la predicción, y que suman cero. La ecuación (8.17) será válida para  $k > q + Qs = 13$ , que es el momento futuro en que desaparecen los términos de media móvil, pero como son necesarios  $d + s = 13$  valores iniciales para determinarla, la ecuación será válida para  $k > q + Qs - d - s = 0$ , es decir, para todo horizonte.

Para determinar los coeficientes  $\beta_0^{(t)}$  y  $\beta_1^{(t)}$  correspondientes a un origen dado y los coeficientes estacionales, podemos resolver el sistema de ecuaciones obtenido al igualar las predicciones para trece períodos, a la estructura (8.17), resultando:

$$\begin{aligned}\hat{z}_t(1) &= \hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} + S_1^{(t)} \\ \vdots &\quad \vdots \quad \vdots \\ \hat{z}_t(12) &= \hat{\beta}_0^{(t)} + 12\hat{\beta}_1^{(t)} + S_{12}^{(t)} \\ \hat{z}_t(13) &= \hat{\beta}_0^{(t)} + 13\hat{\beta}_1^{(t)} + S_1^{(t)}\end{aligned}$$

y con estas ecuaciones podemos obtener  $\beta_0^{(t)}$  y  $\beta_1^{(t)}$  con la restricción  $\sum S_j^{(t)} = 0$ . Restando la primera ecuación de la última y dividiendo por 12:

$$\hat{\beta}_1^{(t)} = \frac{\hat{z}_t(13) - \hat{z}_t(1)}{12} \quad (8.18)$$

y la pendiente mensual se obtiene dividiendo por 12 el crecimiento anual, medido a su vez por la diferencia entre las predicciones de un mes cualquiera en dos años consecutivos.

Sumando las 12 primeras ecuaciones se anulan los coeficientes estacionales, y se obtiene:

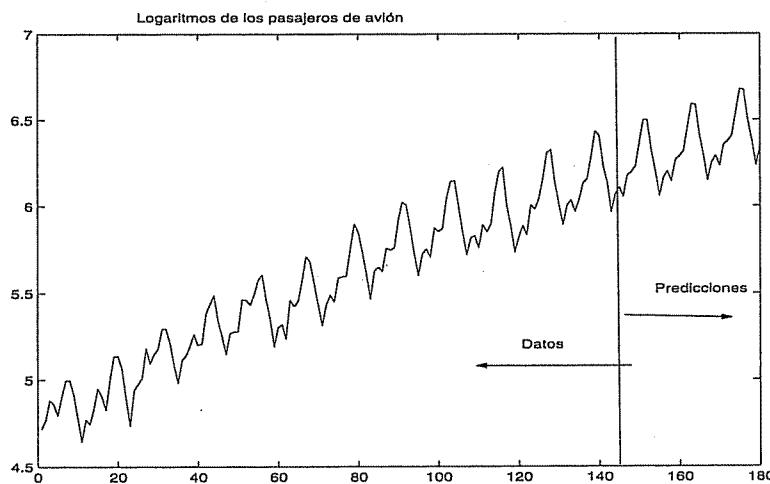
$$\bar{z}_t = \frac{1}{12} \sum_{j=1}^{12} \hat{z}_t(j) = \hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} \left( \frac{1 + \dots + 12}{12} \right)$$

que resulta en:

$$\hat{\beta}_0^{(t)} = \bar{z}_t - \frac{13}{2} \hat{\beta}_1^{(t)}.$$

Finalmente, los coeficientes estacionales se obtienen por diferencia

**Figura 8.3** Datos de los pasajeros de avión en logaritmos y predicciones para tres años



$$S_j^{(t)} = \hat{z}_t(j) - \hat{\beta}_0^{(t)} - \hat{\beta}_1^{(t)}j,$$

y es inmediato comprobar que suman cero dentro del año.

---

### Ejemplo 8.8

Vamos a calcular predicciones con los datos de pasajeros de avión del fichero airline.dat. Estos datos fueron estudiados por Box y Jenkins (1976). El modelo estimado para estos datos es:

$$\nabla\nabla_{12} \ln z_t = (1 - .39B)(1 - .61B^{12})a_t$$

y vamos a generar predicciones para tres años suponiendo que éste es el verdadero modelo para la serie. Estas predicciones se dan en la figura 8.3, conjuntamente con los datos observados.

La función de predicción es una recta más coeficientes estacionales. Para calcular estos parámetros la tabla 8.1 proporciona las predicciones para los primeros 13 datos.

Para calcular la pendiente anual de las predicciones tomamos la predicción para enero del 61,  $\hat{z}_{144}(1) = 6.1091$ , y la de enero del 62,  $\hat{z}_{144}(13) = 6.2004$ . Su diferencia es 0.0913, que corresponde a una tasa de crecimiento del 9.13 % anual. La pendiente de la recta es el crecimiento mensual, que será  $0.0913/12 = .0076083$ . Los factores estaciones se obtienen restando la tendencia a cada

## 8. Predicción con modelos ARIMA

---

**Tabla 8.1 Predicciones para los años 1955 y 1956 para la serie de pasajeros de avión**

AÑO	E	F	M	A	M	J
61	6.109	6.054	6.176	6.196	6.228	6.365
AÑO	J	A	S	O	N	D
61	6.500	6.496	6.321	6.202	6.058	6.163
AÑO	E					
62	6.200					

**Tabla 8.2 Predicciones para los años 1955 y 1956 para los datos de pasajeros de avión**

	E	F	M	A	M	J
$P_{144}(k)$	6.197	6.205	6.212	6.220	6.228	6.235
$\hat{z}_{144}(k)$	6.109	6.054	6.176	6.196	6.228	6.365
Coef. estacional	-0.088	-0.150	-0.036	-0.023	0.000	0.129
	J	A	S	O	N	D
$P_{144}(k)$	6.243	6.251	6.258	6.266	6.273	6.281
$\hat{z}_{144}(k)$	6.500	6.496	6.321	6.202	6.058	6.163
Coef. estacional	0.257	0.246	0.062	-0.063	-0.215	-0.117

predicción. La ordenada en el origen es:

$$\hat{\beta}_0 = (6.1091 + 6.0545 + \dots + 6.1637)/12 - \frac{13}{2}(.0076083) = 6.1901.$$

La tendencia de esta serie sigue la recta:

$$P_{144}(k) = 6.1901 + .0076083k$$

restando a cada predicción el valor de la tendencia se obtienen los factores estacionales, que se presentan en la tabla 8.2. Se observa que el mes más bajo es noviembre, con una caída del 21.52 % respecto a la media del año, y los más altos julio y agosto, con el 25.75 % y el 24.6 % más que la media anual.

#### 8.4. Varianza de las predicciones

La varianza de las predicciones se calcula fácilmente para procesos MA( $q$ ). En efecto, si  $z_t = \theta_q(B) a_t$ , tenemos que:

$$z_{T+k} = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q},$$

y tomando esperanzas condicionadas a las observaciones hasta el instante  $T$  y suponiendo que  $q > k$ , las innovaciones no observadas se anulan por tener esperanza cero y resulta:

$$\hat{z}_T(k) = -\theta_k a_T - \theta_{k+1} a_{T-1} - \dots - \theta_q a_{T+k-q}.$$

Restando estas dos últimas ecuaciones, se obtiene el error de predicción:

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_{k-1} a_{T+1},$$

elevando al cuadrado y tomando esperanzas se obtiene el valor esperado del error cuadrático de predicción de  $z_{T+k}$ , que será igual a la varianza de su distribución condicionada a las observaciones hasta el instante  $T$ :

$$\begin{aligned} ECMP(z_{T+k}|z_T) &= ECMP(e_T(k)) = \\ &= E[(z_{T+k} - z_T(k))^2|z_T] = \sigma^2(1 + \theta_1^2 + \dots + \theta_{k-1}^2). \end{aligned} \quad (8.19)$$

Esta idea puede extenderse a cualquier proceso ARIMA como sigue. Sea  $z_t = \psi(B) a_t$  la representación MA( $\infty$ ) del proceso. Entonces:

$$z_{T+k} = \sum_0^{\infty} \psi_i a_{T+k-i} \quad (\psi_0 = 1), \quad (8.20)$$

y la predicción óptima será, tomando esperanzas condicionadas a las primeras  $T$  observaciones,

$$\hat{z}_T(k) = \sum_0^{\infty} \psi_{k+j} a_{T-j} \quad (8.21)$$

y restando (8.21) de (8.20) se obtiene el error de predicción,

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} + \psi_1 a_{T+k-1} + \dots + \psi_{k-1} a_{T+1}$$

cuya varianza será:

$$Var(e_T(k)) = \sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2). \quad (8.22)$$

Esta ecuación muestra que la incertidumbre de la predicción es muy

## 8. Predicción con modelos ARIMA

---

diferente para modelos estacionarios y no estacionarios. En un modelo estacionario  $\psi_k \rightarrow 0$  si  $k \rightarrow \infty$ , y la varianza de la predicción a largo plazo converge a un valor constante, la varianza marginal del proceso. Esto es consecuencia de que la predicción a largo plazo es la media del proceso. Por ejemplo, para un AR(1),  $\psi_k = \phi^k$  y, cuando  $k$  es grande, tenemos que  $Var(e_T(k)) = \sigma^2/(1 - \phi^2)$ , que es la varianza marginal del proceso. Si  $\phi$  es grande, la varianza marginal será mucho mayor que la condicionada,  $\sigma^2$ , pero la incertidumbre es finita ya que, como máximo, tendremos un error cuadrático de predicción esperado igual a la varianza marginal del proceso.

Para modelos no estacionarios la serie  $\sum \psi_i^2$  no es convergente y la incertidumbre de la predicción a largo plazo crece sin límite hasta infinito. A largo plazo no podemos prever el comportamiento de un proceso no estacionario.

Si suponemos que la distribución de las innovaciones es normal, los resultados anteriores nos permiten calcular intervalos de confianza para la predicción. Entonces,  $z_{T+k}$  será una variable normal de media  $\hat{z}_T(k)$  y varianza dada por (8.22), con lo que podemos escribir

$$z_{T+k} \in (\hat{z}_T(k) \pm \lambda_{1-\alpha} \sqrt{Var(e_T(k))})$$

donde  $\lambda_{1-\alpha}$  son los percentiles de la distribución normal estándar.

---

### Ejemplo 8.9

Dado el modelo  $\nabla z_t = (1 - \theta B) a_t$ , calcular la varianza de las predicciones a distintos períodos.

Para obtener los coeficientes de la representación MA( $\infty$ ), escribiendo  $z_t = \psi(B) a_t$ , como en este modelo  $a_t = (1 - \theta B)^{-1} \nabla z_t$ , tenemos que

$$z_t = \psi(B) (1 - \theta B)^{-1} \nabla z_t$$

es decir,

$$(1 - \theta B) = \psi(B) \nabla = 1 + (\psi_1 - 1)B + (\psi_2 - \psi_1)B^2 + \dots$$

de donde obtenemos  $\psi_1 = 1 - \theta$ ,  $\psi_2 = \psi_1, \dots, \psi_k = \psi_{k-1}$ . Por tanto:

$$Var(e_T(k)) = \sigma^2 (1 + (k-1)(1 - \theta^2))$$

y la varianza de la predicción crece linealmente con el horizonte.

### Ejemplo 8.10

Dado el modelo ARIMA:

$$(1 - 0.2B) \nabla z_t = (1 - 0.8B) a_t$$

con  $\sigma^2 = 4$ , y suponiendo que  $z_{49} = 30$ ,  $z_{48} = 25$ ,  $a_{49} = z_{49} - \hat{z}_{49/48} = -2$ , obtener predicciones desde el último valor observado para  $t = 49$  y construir intervalos de confianza suponiendo normalidad.

La parte AR desarrollada es :

$$(1 - 0.2B)(1 - B) = 1 - 1.2B + 0.2B^2$$

y el modelo puede escribirse:

$$z_t = 1.2z_{t-1} - 0.2z_{t-2} + a_t - 0.8a_{t-1}.$$

Entonces:

$$\begin{aligned}\hat{z}_{49}(1) &= 1.2(30) - 0.2(25) - 0.8(-2) = 32.6, \\ \hat{z}_{49}(2) &= 1.2(32.6) - 0.2(30) = 33.12, \\ \hat{z}_{49}(3) &= 1.2(33.12) - 0.2(32.6) = 33.25, \\ \hat{z}_{49}(4) &= 1.2(33.25) - 0.2(33.12) = 33.28.\end{aligned}$$

Los intervalos de confianza de estas previsiones requieren calcular los coeficientes  $\psi(B)$ . Igualando los operadores puestos en la forma MA( $\infty$ ) tenemos  $(1 - 1.2B + 0.2B^2)^{-1}(1 - 0.8B) = \psi(B)$ , que implica:

$$(1 - 1.2B + 0.2B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - 0.8B)$$

operando en el primer miembro:

$$\begin{aligned}1 - B(1.2 - \psi_1) - B^2(1.2\psi_1 - 0.2 - \psi_2) - \\ - B^3(1.2\psi_2 - 0.2\psi_1 - \psi_3) - \dots = (1 - 0.8B).\end{aligned}$$

e igualando potencias de  $B$ , se obtiene  $\psi_1 = 0.4$ ,  $\psi_2 = 0.28$ ,  $\psi_3 = 0.33$ . Las varianzas de los errores de predicción serán:

$$\begin{aligned}Var(e_{49}(1)) &= \sigma^2 = 4, \\ Var(e_{49}(2)) &= \sigma^2(1 + \psi_1^2) = 4 \times 1.16 = 4.64, \\ Var(e_{49}(3)) &= \sigma^2(1 + \psi_1^2 + \psi_2^2) = 4.95, \\ Var(e_{49}(4)) &= \sigma^2(1 + \psi_1^2 + \psi_2^2 + \psi_3^2) = 5.38,\end{aligned}$$

y asumiendo normalidad, los intervalos aproximados del 95 % para las cuatro predicciones serán  $(32.6 \pm 1.96 \times 2)$ ,  $(33.12 \pm 1.96 \times \sqrt{4.64})$ ,  $(33.25 \pm 1.96 \times \sqrt{4.95})$ ,

y  $(33.28 \pm 1.96 \times \sqrt{5.38})$ .

---

### 8.5. Adaptación de las predicciones

Supongamos que se generan predicciones desde un instante  $T$  para períodos futuros  $T+1, \dots, T+j$ . Cuando se observa el valor  $z_{T+1}$ , queremos adaptar las previsiones  $\hat{z}_{T+2}, \dots, \hat{z}_{T+j}$  a la vista de esta nueva información. Según (8.21) la predicción de  $z_{T+k}$  con información hasta  $T$  es:

$$\hat{z}_T(k) = \psi_k a_T + \psi_{k+1} a_{T-1} + \dots$$

mientras que al observar el valor  $z_{T+1}$  y obtener el error de predicción,  $a_{T+1} = z_{T+1} - \hat{z}_T(1)$ , la nueva predicción para  $z_{T+k}$ , ahora desde el instante  $T+1$ , será:

$$\hat{z}_{T+1}(k-1) = \psi_{k-1} a_{T+1} + \psi_k a_T + \dots$$

Restando estas dos expresiones, tenemos que:

$$\hat{z}_{T+1}(k-1) - \hat{z}_T(k) = \psi_{k-1} a_{T+1}.$$

Por tanto, cuando observemos  $z_{T+1}$  y calculemos  $a_{T+1}$  podremos adaptar todas las predicciones mediante:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + \psi_{k-1} a_{T+1}, \quad (8.23)$$

que indica que las predicciones se adaptan añadiendo a las predicciones anteriores una fracción del último error de predicción obtenido. Si  $a_{T+1} = 0$  las predicciones no se modificarán.

La ecuación (8.23) tiene una interpretación interesante. Las dos variables  $z_{T+1}$  y  $z_{T+k}$  tienen, dada la información hasta  $T$ , una distribución conjuntamente normal con esperanzas,  $\hat{z}_T(1)$  y  $\hat{z}_T(k)$ , varianzas  $\sigma^2$  y  $\sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2)$  y covarianza:

$$\begin{aligned} cov(z_{T+1}, z_{T+k}) &= E[(z_{T+1} - \hat{z}_T(1))(z_{T+k} - \hat{z}_T(k))] = \\ &= E[a_{T+1}(a_{T+k} + \psi_1 a_{T+k-1} + \dots)] = \sigma^2 \psi_{k-1}. \end{aligned}$$

La mejor predicción de  $z_{T+k}$  dada  $z_{T+1}$  y la información hasta  $T$  puede calcularse por regresión, según la expresión:

$$\begin{aligned} E(z_{T+k}|z_{T+1}) &= E(z_{T+k}) + \\ &+ cov(z_{T+1}, z_{T+k}) var^{-1}(z_{T+1})(z_{T+1} - E(z_{T+1})) \end{aligned}$$

donde, para simplificar la notación, no hemos incluido en todos los términos el condicionamiento a la información hasta  $T$ , ya que aparece en todos. Sustituyendo, se obtiene:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + (\sigma^2 \psi_{k-1}) \sigma^{-2} a_{T+1}$$

que es la ecuación (8.23).

---

### Ejemplo 8.11

Adaptar las predicciones del ejercicio 8.10 suponiendo que se observa el valor  $z_{50}$  igual a 34.

Entonces:

$$a_{50} = z_{50} - \hat{z}_{49}(1) = 34 - 32.6 = 1.4,$$

y las nuevas predicciones serán,

$$\begin{aligned}\hat{z}_{50}(1) &= \hat{z}_{49}(2) + \psi_1 a_{50} = 33.12 + 0.4 \times 1.4 = 33.68, \\ \hat{z}_{50}(2) &= \hat{z}_{49}(3) + \psi_2 a_{50} = 33.25 + 0.28 \times 1.4 = 33.64, \\ \hat{z}_{50}(3) &= \hat{z}_{49}(4) + \psi_3 a_{50} = 33.28 + 0.33 \times 1.4 = 33.74,\end{aligned}$$

con nuevos intervalos de confianza  $(33.68 \pm 1.96 \times 2)$ ,  $(33.64 \pm 1.96 \times \sqrt{4.64})$  y  $(33.74 \pm 1.96 \times \sqrt{4.95})$ . Observemos que al cometerse un error por defecto en la predicción de  $z_{50}$ , las predicciones siguientes se revisan al alza.

---

## 8.6. Medidas de predecibilidad

Todo proceso estacionario ARMA,  $z_t$ , puede descomponerse como:

$$z_t = \hat{z}_{t-1}(1) + a_t,$$

que expresa el valor de la serie en cada instante como suma de dos componentes independientes: (1) la predicción desde el instante anterior conociendo los valores pasados y los parámetros del proceso y (2) las innovaciones, que son independientes del pasado de la serie. En consecuencia, podemos escribir:

$$\sigma_z^2 = \sigma_{\hat{z}}^2 + \sigma_a^2,$$

que descompone la varianza de la serie,  $\sigma_z^2$ , en las dos fuentes independientes de variabilidad: (1) la de la parte predecible,  $\sigma_{\hat{z}}^2$ , y (2) la de la parte impredecible,  $\sigma_a^2$ . Box y Tiao (1977) propusieron medir la predecibili-

## 8. Predicción con modelos ARIMA

---

bilidad de una serie estacionaria mediante el cociente entre la varianza de la parte predecible y la total:

$$P = \frac{\sigma_z^2}{\sigma^2} = 1 - \frac{\sigma^2}{\sigma_z^2}. \quad (8.24)$$

Este coeficiente es similar al de determinación en regresión, e indica la proporción de variabilidad de la serie que puede predecirse a partir de su historia. Como en un proceso ARMA  $\sigma_z^2 = \sigma^2 \sum \psi_i^2$ , el coeficiente  $P$  puede escribirse como:

$$P = 1 - (\sum \psi_i^2)^{-1}.$$

Por ejemplo, para un proceso AR(1), como  $\sigma_z^2 = \sigma^2 / (1 - \phi^2)$ , tendremos que:

$$P = 1 - (1 - \phi^2) = \phi^2,$$

y si el parámetro  $\phi$  se aproxima a cero el proceso tenderá hacia ruido blanco y la predecibilidad de la serie será próxima a cero. Si  $\phi$  se aproxima a uno, el proceso tiende a un paseo aleatorio y el coeficiente  $P$  hacia uno.

Esta medida no es útil para procesos ARIMA no estacionarios integrados porque entonces la varianza marginal tiende hacia infinito y la medida  $P$  será siempre uno. Para ver cómo podemos generalizar esta medida observamos que (8.24) compara: (1) el error cuadrático de predicción a un paso conocidos los valores previos de la serie, que es la varianza de la innovación, y (2) el error de predicción cuando carecemos de los valores previos, que es la varianza marginal de la serie. También podemos ver esta expresión como una comparación entre: (1) el error de predicción a un paso y (2) el error de predicción con horizonte infinito, ya que entonces la predicción será la media del proceso y la varianza del error de predicción la varianza marginal de la serie. Este análisis sugiere una definición más general de la predecibilidad relajando los horizontes de predicción en el numerador y en el denominador:

1. En lugar de suponer una predicción con horizonte uno podemos suponer un horizonte  $k$  general, dada la información hasta  $t$ .
2. En lugar de colocar en el denominador el error de predicción con horizonte infinito podemos colocar el error de predicción con horizonte  $k + h$ , para cierto valor de  $h \geq 1$ . Definimos la predecibilidad de una serie temporal con horizonte  $k$  obtenida por  $h$  observaciones adicionales mediante:

$$P(k, h) = 1 - \frac{var(e_t(k))}{var(e_t(k+h))}.$$

Por ejemplo, supongamos un proceso ARIMA y por conveniencia escribámoslo como  $z_t = \psi(B)a_t$ , aunque la serie  $\sum \psi_i^2$  no será en general convergente. Sin embargo, utilizando (8.21), tenemos:

$$P(k, h) = 1 - \frac{\sum_{i=0}^k \psi_i^2}{\sum_{i=0}^{k+h} \psi_i^2} = \frac{\sum_{i=k+1}^{k+h} \psi_i^2}{\sum_{i=0}^{k+h} \psi_i^2}.$$

Este estadístico  $P(k, h)$  permite ver las ventajas de disponer de  $h$  observaciones adicionales para la predicción con horizonte  $k$ . En particular, para  $k = 1$  y  $h = \infty$  se obtiene el estadístico (8.24).

---

### Ejemplo 8.12

Calcular la predecibilidad del proceso  $\nabla z_t = (1 - .8B)a_t$  para uno y dos pasos adelante en función de  $h$ .

Comencemos obteniendo los coeficientes  $\psi$  del proceso mediante:

$$(1 - B)(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - .8B)$$

que implica

$$1 - (1 - \psi_1)B - (\psi_1 - \psi_2)B^2 + \dots = (1 - .8B)$$

de donde se obtiene:

$$\psi_j = .2, \quad j \geq 1.$$

La predecibilidad un paso adelante será:

$$P(1, h) = \frac{\sum_{i=2}^{h+1} \psi_i^2}{\sum_{i=0}^{h+1} \psi_i^2} = \frac{.04h}{1 + .04(h + 1)}.$$

Esta función indica que si  $h = 1$ ,  $P(1, 1) = .04/1.08 = .037$ , y disponer de una observación adicional, o pasar de dos pasos a un paso en la predicción, reduce el error de predicción un 3.7%. Si disponemos de 10 observaciones más, la reducción del error es para  $h = 10$ , de  $P(1, 10) = .4/(1.44) = .2778$ , un 27.78 %. Si  $h = 30$ , entonces  $P(1, 30) = 1.2/(2.24) = .4375$ , y cuando  $h \rightarrow \infty$ , entonces  $P(1, h) \rightarrow 1$ .

---

### 8.7. Lecturas complementarias

Un libro recomendable sobre predicción con series temporales es Granger (1989). La función de predicción de un modelo ARMA se estudia con

## 8. Predicción con modelos ARIMA

detalle en Box, Jenkins y Reinsel (1994). La interpretación de la función de predicción que hemos presentado ha sido obtenida por Espasa y Peña (1995) y el lector puede encontrar más detalles y ejemplos en el capítulo 5 de Peña, Tiao y Tsay (2001). Un tema no tratado aquí es la combinación de predicciones generadas con distintos modelos, que se presenta muy claramente en el capítulo 8 de Granger (1989) y desde el punto de vista Bayesiano en West y Harrison (1989). Recomendamos al lector interesado en la predicción económica acudir a Espasa y Cancelo (1993) y Clements y Hendry (1998).

A veces es necesario tener en cuenta la presencia de restricciones sobre las predicciones, por ejemplo que la suma de las predicciones mensuales debe coincidir con un total anual fijado, y métodos de corregir las predicciones para que verifiquen restricciones de este tipo han sido estudiados por Cholette y Dagum (1994), Guerrero (1991) y Guerrero y Peña (2000). Un aspecto adicional es la combinación de predicciones a distintos niveles de agregación, por ejemplo mensuales y anuales, y las ventajas relativas de trabajar con datos de distinto nivel de agregación, por ejemplo anuales, frente a mensuales. El capítulo 16 del libro de Wei (1990) contiene una introducción al tratamiento de la agregación temporal que se estudia con más detalle en Wei y Stram (1990), Guerrero (1990) y Guerrero y Peña (2003). Las medidas de predecibilidad introducidas en esta sección han sido propuestas por Diebold y Kilian (2001).

---

### Ejercicios 8

- 8.1 Dado el proceso  $z_t = 2 + 0.8z_{t-1} - 0.1z_{t-2} + a_t$  y las cuatro observaciones (4, 3, 1, 2.5), generar predicciones para cuatro períodos adelante.
- 8.2 Indicar cuál será la predicción a largo plazo generada por el modelo del ejercicio anterior.
- 8.3 Suponiendo que la varianza de las innovaciones del ejercicio 8.1 es 2, calcular intervalos de confianza para las predicciones a uno y dos pasos.
- 8.4 Calcular predicciones para  $t = 100, 101$  y  $102$  y la ecuación de predicción final del proceso MA(2)  $z_t = 5 + a_t - 0.5a_{t-1}$ , sabiendo que las predicciones realizadas con información hasta  $t = 97$  han sido:  $\hat{z}_{97}(1) = 5.1$ , y  $\hat{z}_{97}(2) = 5.3$ , y que después hemos observado  $z_{98} = 4.9$  y  $z_{99} = 5.5$ .
- 8.5 Explicar la estructura de las predicciones generadas por el modelo:  $\nabla z_t = 3 + (1 - 0.7B)a_t$ .
- 8.6 Explicar la estructura de las predicciones a largo plazo con el modelo:  $\nabla\nabla_{12}z_t = (1 - 0.7B)a_t$ .

8.7 Demostrar que en el proceso IMA(1,1) que puede escribirse  $z_{T+1} = (1 - \theta) \sum_{j=0}^{\infty} \theta^j z_{T-j} + a_{T+1}$  (véase ejemplo 8.4) se verifica que para  $k \geq 2$ ,  $\hat{z}_T(k) = \hat{z}_T(k-1)$ . (Observar que  $\hat{z}_T(2) = (1 - \theta)[\hat{z}_T(1) + \theta z_T + \theta^2 z_{T-1} + \dots]$  y sustituir la expresión de  $\hat{z}_T(1)$ ).

8.8 Calcular la predicción del proceso  $z_t = 2 + 0.8z_{t-1} - 0.1z_{t-2} + a_t$ .

### Apéndice 8.1: Predicción y esperanza condicionada

Vamos a demostrar que la esperanza condicionada minimiza el error cuadrático medio de predicción en un sentido absoluto, es decir, no sólo para la realización observada, sino también para las otras realizaciones posibles que no hemos observado. Hemos demostrado en la sección 8.2 que dada una serie temporal la predicción con la esperanza condicionada minimiza el error cuadrático de las observaciones futuras, pero no es inmediato que si tenemos muchas realizaciones del mismo proceso –muchas series temporales– y prevemos cada una de ellas con este criterio minimizamos el error promedio calculado ahora para todas las series. Vamos a ver que esto es cierto.

Para demostrar este resultado, partiremos de la importante propiedad siguiente: la esperanza de una función  $g(x, y)$  respecto a la distribución conjunta de las dos variables puede obtenerse en dos etapas (a veces se conoce esta regla como de las esperanzas iteradas). Primero, se calcula la esperanza de la función respecto a la distribución de una de ellas condicionada a la otra, por ejemplo tomamos la esperanza respecto a la distribución condicionada de  $y$  dada la  $x$ , y, segundo, calculamos la esperanza del resultado anterior respecto a la distribución de la segunda, en el ejemplo de  $x$ . Escribiremos:

$$E_{xy}[g(x, y)] = E_x [E_{y|x}[g(x, y)]] \quad (8.25)$$

donde  $E_{xy}$  quiere decir la esperanza respecto a la distribución conjunta de ambas variables,  $E_{y|x}$  la esperanza respecto a la distribución condicionada de  $y$  dada  $x$  y  $E_x$  la esperanza respecto a la marginal de  $x$ .

#### Demostración

$$\begin{aligned} E_{xy}[g(x, y)] &= \int \int g(x, y) f(x, y) dx dy = \\ &= \int \left[ \int g(x, y) f(y|x) dy \right] f(x) dx = \\ &= \int E_{y|x}[g(x, y)] f(x) dx = E_x [E_{y|x}[g(x, y)]] \end{aligned}$$

## 8. Predicción con modelos ARIMA

---

y este resultado es válido para dos variables escalares o vectoriales.

Sea ahora  $z_{t+k}$  una variable aleatoria parte del proceso estocástico  $\{z_t\}$ . Cuando se carece de información sobre los valores previos del proceso, la predicción óptima de  $z_{t+k}$  es su media  $\mu_{t+k}$ , ya que la constante que minimiza  $E[(z_{t+k} - c)^2]$  es la esperanza de la variable. Sin embargo, dados  $t$  valores previos de la realización,  $\mathbf{z} = (z_1, \dots, z_t)$ , la predicción de  $z_{t+k}$  que minimiza el error cuadrático medio condicionada a los datos observados,  $\mathbf{z}$ , es la esperanza de la variable  $z_{t+k}$  condicionada a lo observado. En consecuencia, el predictor óptimo será el que minimiza:

$$E[(z_{t+k} - \hat{z}_{t+k})^2 | \mathbf{z}_t],$$

que resulta en

$$\hat{z}_{t+k} = E(z_{t+k} | \mathbf{z}_t), \quad (8.26)$$

y el predictor óptimo es la esperanza de la variable desconocida,  $z_{t+k}$ , condicionada a la información disponible,  $\mathbf{z}$ .

Vamos a demostrar que esta predicción minimiza el error cuadrático medio respecto a la distribución conjunta de  $(z_{t+k}, \mathbf{z})$ , es decir, promediando no sólo respecto a los valores futuros de  $z_{t+k}$ , sino también respecto a las posibles secuencias,  $\mathbf{z}$ . Consideremos otro predictor cualquiera  $z_{t+k}^*$ . Entonces:

$$\begin{aligned} E[(z_{t+k} - z_{t+k}^*)^2] &= E[(z_{t+k} - \hat{z}_{t+k} + \hat{z}_{t+k} - z_{t+k}^*)^2] = \\ &= E[(z_{t+k} - \hat{z}_{t+k})^2] + E[(\hat{z}_{t+k} - z_{t+k}^*)^2] + \\ &\quad + 2E[(z_{t+k} - \hat{z}_{t+k})(\hat{z}_{t+k} - z_{t+k}^*)], \end{aligned}$$

donde las esperanzas se toman respecto a la distribución conjunta de  $z_{t+k}$  y  $\mathbf{z}$ . Demostraremos que el doble producto es cero tomando esperanzas respecto a la distribución de  $z_{t+k} | \mathbf{z}$  y luego respecto a  $\mathbf{z}$  en virtud de (8.25). Como:

$$\begin{aligned} E[(\hat{z}_{t+k} - z_{t+k}^*)(z_{t+k} - \hat{z}_{t+k}) | \mathbf{z}_t] &= \\ &= (\hat{z}_{t+k} - z_{t+k}^*) E[(z_{t+k} - \hat{z}_{t+k}) | \mathbf{z}_t] = 0, \end{aligned}$$

ya que tanto  $\hat{z}_{t+k}$  como  $z_{t+k}^*$  serán funciones de  $\mathbf{z}$  que es fijo y  $\hat{z}_{t+k}$  está dado por (8.26). En consecuencia, resulta que:

$$E[(z_{t+k} - z_{t+k}^*)^2] = E[(z_{t+k} - \hat{z}_{t+k})^2] + E[(\hat{z}_{t+k} - z_{t+k}^*)^2],$$

que será mínimo cuando  $\hat{z}_{t+k} = z_{t+k}^*$ . Por tanto, la esperanza condicionada es el predictor óptimo, no sólo en un sentido condicional, sino también incondicional.

Este resultado, aunque de gran interés teórico, sólo permite obtener el predictor cuando conocemos las distribuciones condicionadas. En general, esto no ocurre, y buscaremos predictores que siendo funciones lineales de las ob-

servaciones minimicen el error de predicción. Si:

$$\hat{z}_{t+k} = b_{k,0} z_T + \dots + b_{k,T-1} z_1 = \mathbf{b}'_k \mathbf{z}_T, \quad (8.27)$$

vamos a determinar la expresión del vector  $\mathbf{b}_k = (b_{k,0}, \dots, b_{k,T-1})'$ . El error cuadrático medio de este predictor es:

$$\begin{aligned} E[(z_{t+k} - \hat{z}_{t+k})^2] &= E[(z_{t+k} - \mathbf{b}'_k \mathbf{z}_t)(z_{t+k} - \mathbf{b}'_k \mathbf{z}_t)'] \\ &= E[z_{t+k}^2 - 2z_{t+k} \mathbf{b}'_k \mathbf{z}_t + \mathbf{b}'_k \mathbf{z}_t \mathbf{z}_t' \mathbf{b}_k] \\ &= Var(z_{t+k}) - 2\mathbf{b}'_k \gamma(t, t+k) + \mathbf{b}'_k \Gamma(t) \mathbf{b}_k, \end{aligned}$$

siendo  $\gamma(t, t+k)$  el vector de covarianzas entre la variable a prever,  $z_{t+k}$ , y los predictores,  $z_T, \dots, z_1$ , y  $\Gamma(t)$  la matriz de varianzas y covarianzas de estos predictores. Para obtener el valor que minimiza esta expresión, derivando respecto a  $\mathbf{b}_k$  e igualando a cero:

$$-2\gamma(t, t+k) + 2\Gamma(t) \mathbf{b}_k = 0$$

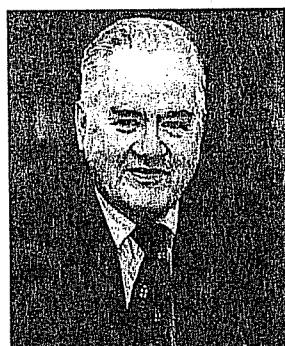
que proporciona el predictor definido por:

$$\mathbf{b}_k = \Gamma(t)^{-1} \gamma(t, t+k).$$

Este resultado es previsible ya que si consideramos (8.27) como una ecuación de regresión, la ecuación de  $\mathbf{b}_k$  es la expresión usual de estimación de los parámetros en regresión múltiple. Para procesos normales la esperanza condicionada es siempre función lineal de las observaciones y, por tanto, el mejor predictor lineal coincide con el mejor predictor global.

---

## 9. La identificación de los posibles modelos ARIMA



John Tukey (1915-2000)

Estadístico estadounidense. Uno de los creadores del análisis espectral de series temporales, inventor de la transformada rápida de Fourier. Director fundador del Departamento de estadística en Princeton, simultaneó la docencia universitaria con su trabajo en los laboratorios Bell de ATT. Pionero en los estudios de robustez y creador de numerosos términos científicos (como software o bit), sus contribuciones han influido mucho en la evolución de la estadística en la segunda mitad del siglo XX.

### 9.1. Introducción

En las secciones anteriores hemos estudiado las propiedades teóricas de los procesos ARIMA. Vamos a analizar a continuación cómo ajustar estos modelos a series reales. Box y Jenkins (1976) propusieron realizar este ajuste en tres etapas.

La primera etapa consiste en *identificar* el posible modelo ARIMA que sigue la serie, lo que requiere: (1) decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria; (2) determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes  $p$  y  $q$  de su estructura autorregresiva y de media móvil y, si el proceso es estacional, los órdenes  $P, Q$  de la estructura ARMA estacional.

Una vez seleccionado provisionalmente un modelo para la serie estacionaria, pasaremos a la segunda etapa de *estimación*, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo. La tercera etapa es la de *diagnóstico*, donde se comprueba que los residuos no tienen estruc-

tura de dependencia y siguen un proceso de ruido blanco. Si los residuos no contienen información, aceptamos el modelo como adecuado y lo utilizaremos para la previsión o la toma de decisiones. Si los residuos muestran estructura, modificaremos el modelo para incorporarla y repetiremos las tres etapas anteriores hasta obtener un modelo adecuado.

Estas tres etapas supusieron en su día un avance muy importante, ya que la estimación de los parámetros de un modelo ARMA requería mucho tiempo de cálculo, por lo que era conveniente asegurarse que el modelo a estimar podía ser el adecuado. En la actualidad, la estimación de un modelo ARIMA por máxima verosimilitud es inmediata, por lo que es más simple estimar todos los modelos que consideremos posibles para explicar la serie y después elegir entre ellos, como veremos en el capítulo siguiente. Ésta es la filosofía de los criterios automáticos de selección de modelos ARIMA, que funcionan bien en muchos casos, y son imprescindibles cuando se desea modelar y prever un gran conjunto de series. Sin embargo, cuando el número de series a modelar sea pequeño, es conveniente realizar la etapa de identificación para comprender y familiarizarse con la estructura dinámica de las series de interés. El objetivo no es ahora seleccionar un modelo, sino identificar un conjunto de modelos posibles que sean compatibles con el gráfico de la serie y sus funciones de autocorrelación simple y parcial. Estos modelos serán estimados en la etapa siguiente donde seleccionaremos el mejor. En la tercera fase, *diagnosis*, nos aseguraremos que el modelo seleccionado no tiene deficiencias detectables.

La identificación del modelo requiere identificar la estructura no estacionaria, si existe, y después la estructura ARMA estacionaria. La identificación de la estructura no estacionaria consiste en detectar qué transformaciones hay que aplicar para conseguir un proceso ARMA estacionario con varianza y media constante. En particular, consiste en determinar si es necesario: (1) transformar la serie para que tenga varianza constante; (2) diferenciar la serie para que tenga media constante.

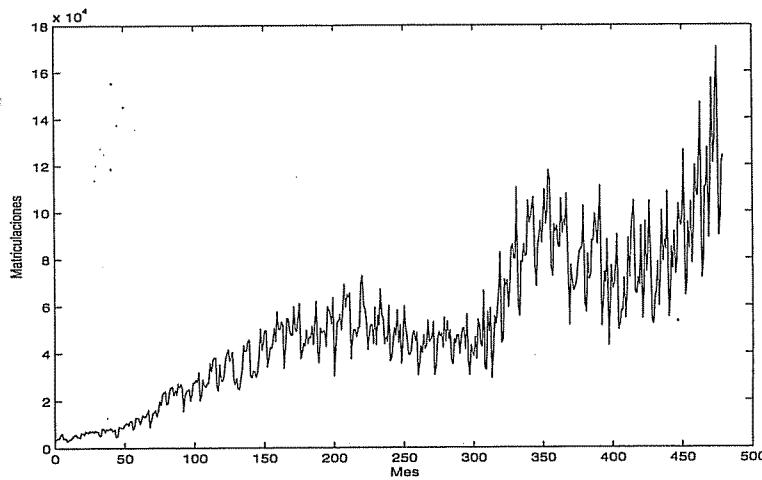
Posteriormente identificaremos el modelo ARMA para la serie estacionaria. Vamos a analizar estos aspectos.

## 9.2. Determinación de la transformación para estabilizar la varianza

Es frecuente en muchas series que la variabilidad sea mayor cuando la serie toma valores altos que cuando toma valores más bajos. Por ejemplo, en la figura 9.1 de la serie de matriculación de automóviles en España se observa que la variabilidad es mucho mayor cuando el nivel de la serie es alto, como ocurre al final de la serie, que cuando es bajo. Si obser-

## 9. La identificación de los posibles modelos ARIMA

Figura 9.1 Serie de matriculación de automóviles de turismo en España, 1960 a 1999



vamos el gráfico con atención, veremos que la variabilidad no aumenta con el tiempo, sino con el nivel de la serie: la variabilidad alrededor de la observación 200 es grande y corresponde a un máximo del nivel, y esta variabilidad decrece alrededor de la observación 300, cuando el nivel disminuye. La variabilidad máxima de la serie se observa al final, porque el nivel es también entonces mayor. Podemos confirmar esta impresión visual haciendo un gráfico entre una medida de variabilidad, como la desviación típica, y una medida del nivel, como la media local. Para hacer comparaciones homogéneas, como la serie es estacional tomaremos las 12 observaciones de cada año y calcularemos en cada año la desviación típica de las matriculaciones y la media en el año. La figura 9.2 presenta la desviación típica de cada año como función de la media del año y puede verse que existe una clara dependencia de tipo lineal entre ambas variables. Cuando la variabilidad de la serie aumente linealmente con el nivel de la serie, como en la serie de matriculaciones, tomando logaritmos se consigue una serie con variabilidad constante. La figura 9.3 de la serie en logaritmos confirma este hecho.

La dependencia de la variabilidad del nivel de la serie puede ser el resultado de que la serie se genera como producto (en lugar de la suma) de un componente sistemático o predecible, que define el nivel de la serie, y que representaremos por  $\mu_t = f(z_{t-1}, \dots)$  para indicar que es función de los valores pasados, más el término de innovación,  $u_t$ , que define la

Figura 9.2 Relación entre la desviación típica y la media de cada año para la serie de matriculación de vehículos

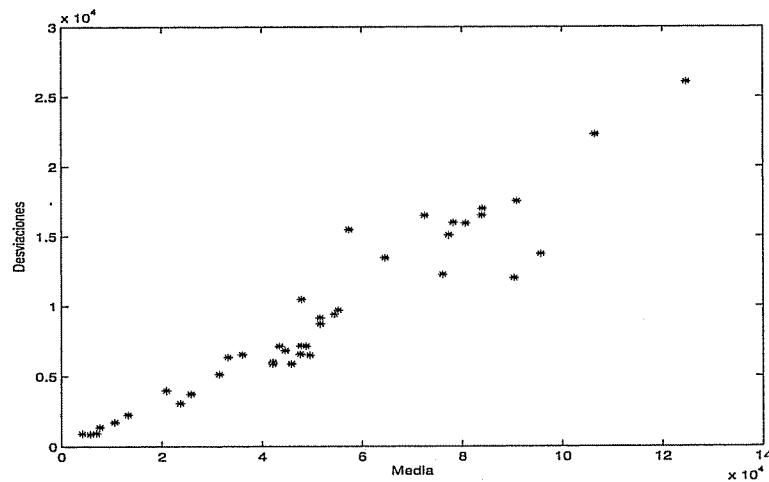
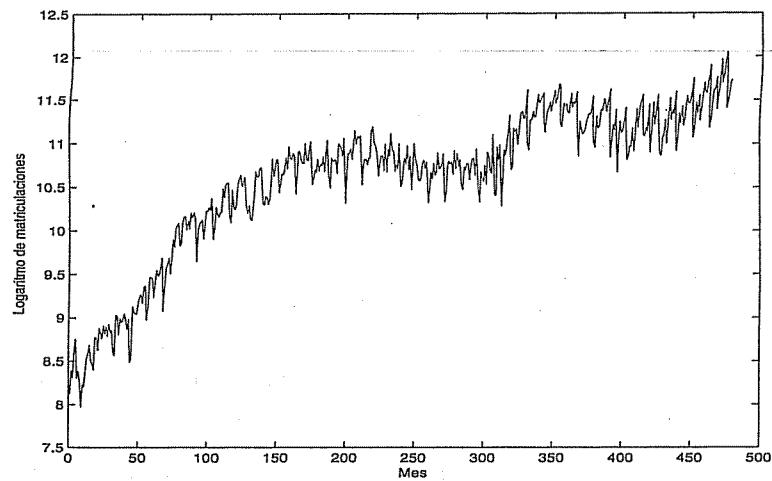


Figura 9.3 La serie de matriculación de vehículos en logaritmos



variabilidad. Entonces:

$$z_t = \mu_t u_t. \quad (9.1)$$

Supongamos que la esperanza de esta innovación es uno, de manera que

## 9. La identificación de los posibles modelos ARIMA

---

$E(z_t) = \mu_t$ . La desviación típica de la serie será:

$$\begin{aligned}\sigma_t &= [E(z_t - \mu_t)^2]^{1/2} = [E(\mu_t u_t - \mu_t)^2]^{1/2} = \\ &= |\mu_t| [E(u_t - 1)^2]^{1/2} = |\mu_t| \sigma_u\end{aligned}$$

y aunque la innovación  $u_t$  tenga variabilidad constante, la desviación típica de la serie observada  $z_t$  no es constante en el tiempo y será proporcional al nivel de la serie. Como hemos visto, este problema se resuelve tomando logaritmos, ya que entonces, llamando  $a_t = \ln u_t$ :

$$y_t = \ln z_t = \ln \mu_t + a_t$$

y se obtiene una descomposición aditiva para la variable  $y_t$ , que tendrá varianza constante.

El caso anterior puede generalizarse permitiendo que la desviación típica sea una función potencial de la media local, mediante:

$$\sigma_t = k \mu_t^\alpha, \quad (9.2)$$

y se demuestra en el apéndice 9.1 que si transformamos las variables  $z_t$  en nuevas variables  $y_t$  mediante:

$$y_t = x_t^{1-\alpha}$$

estas nuevas variables  $y_t$  tienen la misma desviación típica. Una forma un poco más general de escribir esta transformación es mediante:

$$y_t = \frac{x_t^{1-\alpha} - 1}{1 - \alpha}, \quad (9.3)$$

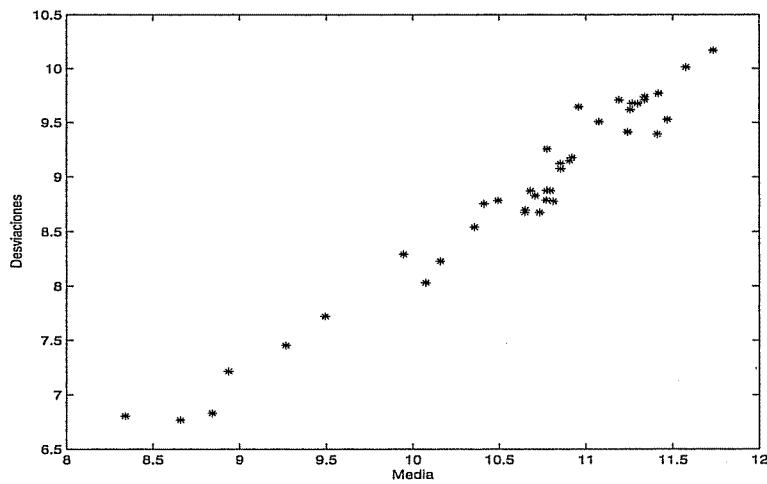
que es la familia de *transformaciones de Box-Cox*, e incluye además de las potencias el logaritmo de la variable. En efecto, cuando  $\alpha$  es muy próximo a uno puede comprobarse, tomando límites en la expresión (9.3), que la transformación equivale a tomar el logaritmo de la variable  $x_t$ .

Cuando en la serie se observa una relación entre el nivel y la variabilidad, podemos estimar el valor de  $\alpha$  necesario para conseguir variabilidad constante haciendo grupos de observaciones consecutivas, calculando en cada grupo la desviación típica,  $s_i$ , y la media  $\bar{x}_i$  y representando estas variables en un gráfico. Tomando logaritmos en (9.2), la pendiente de la regresión

$$\log s_i = c + \alpha \log \bar{x}_i \quad (9.4)$$

nos estimará el valor de  $\alpha$ , y la transformación de los datos mediante (9.3) conducirá a una serie donde la variabilidad no depende del nivel. Por ejemplo, la figura 9.4 presenta la relación entre los logaritmos de estas

Figura 9.4 Relación entre los logaritmos de las desviaciones típicas y los de las medias para la serie de matriculación de vehículos



variables para la serie de matriculaciones. Se observa que la pendiente es próxima a la unidad, y si estimamos la regresión (9.4) se obtiene  $c = -2.17$  y  $\alpha = 1.04$ . En consecuencia, una transformación con  $\alpha = 0$ , es decir, mediante logaritmos, deberá producir varianza constante con el nivel. La figura 9.3 presenta la serie de matriculación de vehículos expresada ahora en logaritmos, y se observa que con esta transformación la variabilidad es ahora aproximadamente constante.

Los gráficos entre variabilidad y media conviene hacerlos con grupos de datos lo más homogéneos posible. Con series estacionales de periodo  $s$  debemos hacer grupos de este tamaño, para incluir todos los periodos estacionales en todos los grupos. Para series no estacionales, se toma un tamaño de grupo de cinco o seis observaciones contiguas.

Existen series heterocedásticas donde la varianza cambia sin relación con el nivel. Por ejemplo, algunas observaciones tienen mucha mayor varianza que las restantes o aparecen cambios de varianza por tramos. Estas situaciones no pueden resolverse con una transformación Box-Cox y se analizarán en el capítulo 13. Además, es posible que la varianza de la serie sea constante, pero no lo sea la varianza condicional. Estas situaciones se estudiarán en el capítulo 15.

### 9.3. Determinación de la transformación para estabilizar la media

Para estabilizar la media de la serie aplicaremos diferencias regulares y estacionales. La decisión de aplicar estas diferencias puede basarse en el gráfico de la serie y en la función de autocorrelación muestral, pero también podemos hacer un contraste formal, como veremos a continuación.

#### 9.3.1. Determinar el orden de diferenciación regular

Si la serie tiene tendencia, o muestra cambios de nivel en la media, la diferenciaremos para transformarla en estacionaria. La necesidad de diferenciar es en muchos casos clara a la vista del gráfico de la serie. Por ejemplo, la serie de matriculación de vehículos de la figura 9.1 muestra claramente tendencia por tramos, y media no constante. Nótese que, como vimos en el capítulo 6 (sección 6.2):

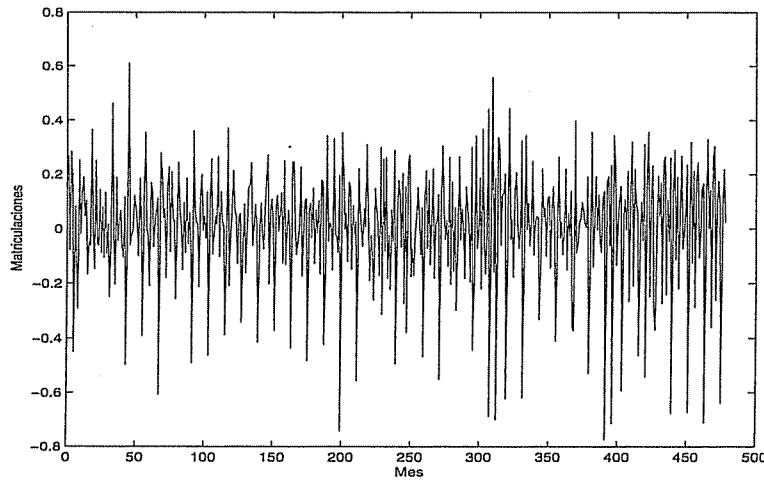
$$(1 - B) \ln z_t = \ln z_t - \ln z_{t-1} = \ln \left( 1 + \frac{z_t - z_{t-1}}{z_{t-1}} \right) \simeq \frac{z_t - z_{t-1}}{z_{t-1}}$$

donde hemos utilizado que  $\ln(1 + x) \simeq x$  si  $x$  es pequeño. Por tanto, la serie  $\nabla \ln z_t$  equivale a las tasas de incremento relativo de  $z_t$ . La figura 9.5 presenta la primera diferencia de esta serie y puede verse que contiene variaciones muy acusadas mes a mes, de hasta 0.8, es decir del 80 % de su valor. Esto ocurre por la presencia de una fuerte estacionalidad, que aparece en el gráfico con unas caídas muy fuertes, que corresponden al mes de agosto de cada año. Como consecuencia de este efecto la serie no tiene media constante. En la sección siguiente estudiaremos cómo eliminar esta pauta mediante una diferencia estacional.

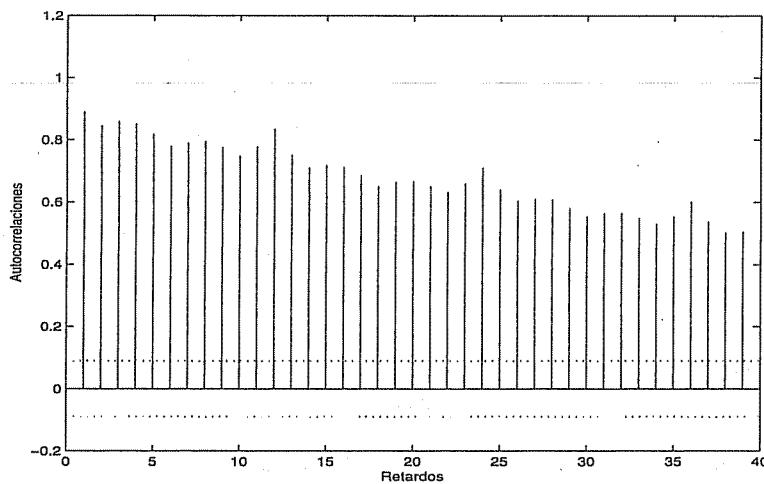
Cuando la decisión de diferenciar no sea clara al analizar el gráfico de la serie, conviene estudiar su función de autocorrelación simple, *fas*. Hemos visto que una serie no estacionaria debe mostrar autocorrelaciones positivas, con decrecimiento lento y lineal. Por ejemplo, la figura 9.6 presenta la *fas* de la serie de matriculación de vehículos: se observa un decrecimiento lento de los coeficientes de forma lineal, que indica claramente la necesidad de diferenciar.

Es importante señalar que la característica que identifica una serie no estacionaria en la *fas* estimada no es coeficientes de autocorrelación próximos a la unidad, sino decrecimiento lento de forma lineal. Puede demostrarse que el valor esperado de los coeficientes de autocorrelación estimados depende del proceso y del tamaño muestral, pudiendo ser mu-

**Figura 9.5** Primera diferencia del logaritmo de la serie de matriculaciones



**Figura 9.6** Función de autocorrelación para la serie de matriculación de vehículos



cho menor que la unidad. Por ejemplo, en el proceso IMA(1,1) si  $\theta = 0.8$  y  $T = 100$ , el valor esperado de los coeficientes de autocorrelación muestral es siempre menor que 0.5. Sin embargo, sigue siendo esperable un decrecimiento muy suave y lineal. En resumen, si la *fas* no se amortigua para retardos altos (mayores de 15 o 20), en general es necesario dife-

## 9. La identificación de los posibles modelos ARIMA

enciar para obtener un proceso estacionario. Vimos en el capítulo 6 que existen series estacionarias de memoria larga donde los coeficientes de autocorrelación decrecen muy lentamente, pero como estas series pueden aproximarse bien por modelos ARMA en la práctica, no las consideraremos en este capítulo.

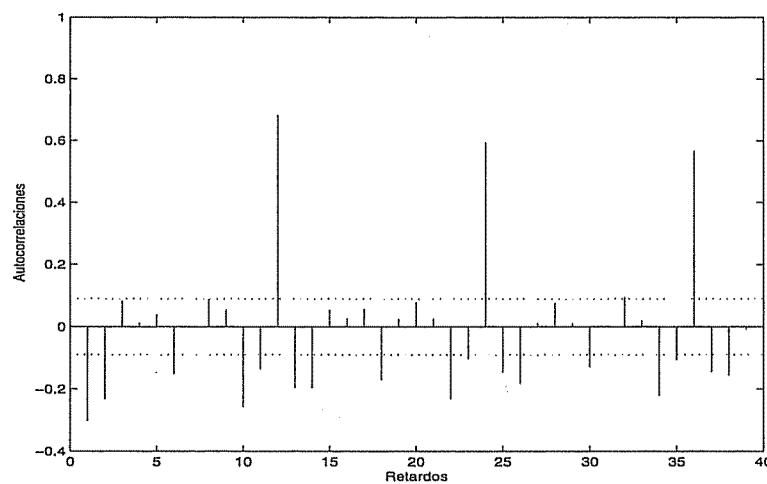
Es posible realizar un contraste para decidir si conviene o no diferenciar la serie. Si el objetivo es la predicción, estos contrastes no son muy útiles, porque, en caso de duda, conviene siempre diferenciar, ya que las consecuencias negativas de sobrediferenciar son mucho menores que las de infradiferenciar. En efecto, si suponemos que la serie es estacionaria cuando no lo es, los errores de predicción a medio plazo pueden ser enormes, ya que, según hemos visto en el capítulo 8, la predicción de una serie estacionaria a medio plazo es su media, mientras que una serie no estacionaria puede alejarse indefinidamente de dicho valor y el error de predicción no está acotado. El modelo no diferenciado será, además, poco robusto y sin capacidad de adaptación ante valores futuros. Por el contrario, si sobrediferenciamos tenemos siempre predictores adaptativos y, aunque perdamos precisión, esta pérdida está acotada. Incluir en el modelo una diferencia de más implica introducir artificialmente un término de media móvil que será no invertible. Como tendremos que estimar un parámetro adicional, perderemos algo de eficiencia en la estimación, pero, como veremos en el capítulo siguiente, los estimadores de todos los parámetros serán consistentes, y el efecto de sobrediferenciar sobre la predicción será pequeño, aunque calcularemos por exceso la incertidumbre de las predicciones. Una razón adicional de tomar diferencias en caso de duda es que puede demostrarse (véase Sánchez y Peña, 2002) que, aunque la serie sea estacionaria, si el parámetro AR está cerca de la unidad se obtienen mejores predicciones sobrediferenciando la serie que utilizando el verdadero modelo estacionario. Cuando el objetivo no sea prever, sino decidir si una variable es o no estacionaria, situación que aparece con frecuencia en aplicaciones económicas, entonces sí deben utilizarse los contrastes de raíces unitarias que estudiaremos en la sección siguiente.

### **9.3.2. Determinar el orden de diferenciación estacional**

Si la serie tiene estacionalidad, habrá que aplicar una diferencia estacional,  $\nabla_s = 1 - B^s$ , para conseguir una serie estacionaria. La estacionalidad se manifiesta:

1. En el gráfico de la serie, que presentará una pauta repetida de periodo  $s$ . Con datos mensuales veremos que algunos meses son sistemáticamente altos y otros bajos.

**Figura 9.7 Fas de la serie de matriculación en logaritmos después de una diferencia regular**



2. En la función de autocorrelación simple, que presentará coeficientes positivos que decrecen lentamente en los retardos  $s, 2s, 3s, \dots$

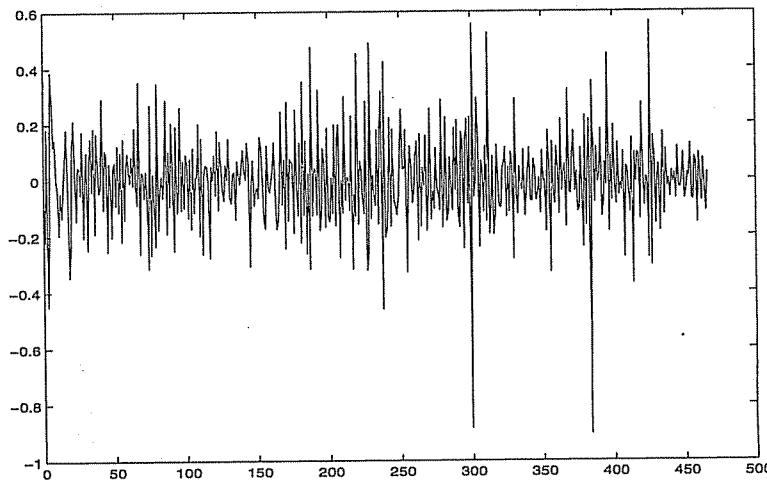
Por ejemplo, la figura 9.1 sugiere una pauta estacional, porque el valor de la serie es sistemáticamente menor en agosto. Este efecto es más claramente visible al eliminar la tendencia tomando una diferencia regular, como ilustra la figura 9.5. Además del gráfico de la serie es conveniente mirar la fas, ya que una serie con estacionalidad mostrará valores de autocorrelación positivos y altos en los retardos estacionales. Por ejemplo, la figura 9.7 muestra la fas muestral de la serie del logaritmo de la serie de matriculación de vehículos con una diferencia regular. Se observa que existen correlaciones altas y persistentes en los retardos estacionales, 12, 24 y 36. Esto sugiere la necesidad de tomar una diferencia estacional para obtener una serie estacionaria. La figura 9.8 presenta la serie con dos diferencias, una regular y otra estacional.

En principio sería teóricamente posible que una serie requiera más de una diferencia estacional, pero en la práctica esto es extremadamente raro, por lo que la decisión se reducirá a aplicar o no una diferencia estacional.

### 9.3.3. Contrastes de raíces unitarias

Si se desea decidir si un proceso es o no estacionario podemos realizar un contraste de raíces unitarias. Estos contrastes indican si debemos o

**Figura 9.8 La serie de matriculación de vehículos en logaritmos y con dos diferencias**



no tomar una diferencia adicional en una serie para convertirla en estacionaria. Un procedimiento alternativo de tomar esta decisión es utilizar los criterios de selección de modelos, que estudiaremos en el capítulo siguiente.

#### 9.3.3.1. Contraste de Dickey-Fuller

Presentaremos inicialmente el contraste en el caso más simple. Supongamos que se desea decidir entre el proceso no estacionario:

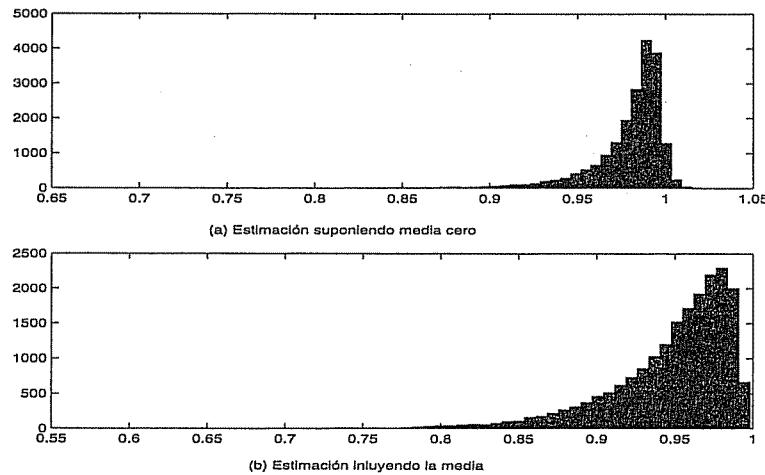
$$\nabla z_t = a_t, \quad (9.5)$$

y el estacionario:

$$(1 - \rho B) z_t = c + a_t. \quad (9.6)$$

El test de Dickey-Fuller se ha desarrollado porque el procedimiento tradicional de estimar ambos modelos y elegir el de menor varianza no es válido en este caso. En efecto, si generamos una serie que sigue un paseo aleatorio, (9.5), y ajustamos a esa serie ambos modelos, en el modelo (9.5) no estimaremos ningún parámetro mientras que en el (9.6) estimaremos los parámetros  $\rho$  y  $c$  de manera que la varianza de los residuos sea mínima. En consecuencia, el modelo estacionario proporcionará en general un mejor ajuste a los datos, ya que es posible que con un valor de  $c$  próximo a cero y con un valor de  $\rho$  próximo a uno consigamos un mejor ajuste que imponiendo  $c = 0$  y  $\rho = 1$ , y concluiremos que el modelo

**Figura 9.9 Distribución de la estimación del parámetro de un AR(1) cuando los datos se generan por un paseo aleatorio. En el caso (a) se supone que la media es cero y en el (b) se estima con la media muestral**



estacionario es el adecuado.

Para ilustrar este aspecto, la figura 9.9 presenta la distribución del estimador de mínimos cuadrados del parámetro  $\rho$  en muestras de tamaño 100 que han sido generadas por el paseo aleatorio (9.5). La distribución se ha obtenido generando 20 000 replicaciones del paseo aleatorio y estimando en cada una de ellas el parámetro suponiendo que el proceso es AR(1). Se han considerado dos estimadores. En el caso (a) se ha supuesto que la media de la serie es cero, en cuyo caso en (9.6) la constante  $c$  es cero, y el parámetro se estima mediante  $\hat{\rho} = \sum z_t z_{t-1} / \sum z_t^2$ . En el caso (b), más realista en la práctica, no se supone que la media sea cero y por tanto el parámetro se estima por  $\hat{\rho} = \sum (z_t - \bar{z})(z_{t-1} - \bar{z}) / \sum (z_t - \bar{z})^2$  y  $\hat{c} = (1 - \hat{\rho})\bar{z}$ . Se observa que la distribución es muy asimétrica y que, especialmente con el segundo estimador, el valor estimado, que es el que proporciona el mejor ajuste, puede ser bastante menor que el verdadero valor del parámetro que es la unidad. Por tanto, aunque los datos se hayan generado por un paseo aleatorio, muy frecuentemente obtendremos residuos con menor varianza utilizando un valor para el parámetro  $\hat{\rho} < 1$ . En conclusión, comparar las varianzas de ambos modelos no es un buen método para elegir entre ellos, y, especialmente, si queremos cubrirnos respecto al error de no diferenciar.

Podría pensarse que una forma mejor de decidir entre los dos modelos es estimar el modelo estacionario (9.6) por mínimos cuadrados y

## 9. La identificación de los posibles modelos ARIMA

---

contrastar si el coeficiente  $\rho$  es uno, comparando el estimador con su desviación típica estimada. El contraste sería  $H_0 : \rho = 1$  frente a la alternativa  $H_1 : \rho < 1$ , y el estadístico para el contraste:

$$t_\mu = \frac{\hat{\rho} - 1}{\hat{s}_\rho}, \quad (9.7)$$

donde  $\hat{s}_\rho$  es la desviación típica estimada del estimador de mínimos cuadrados. Podríamos realizar el contraste comparando el valor de  $t_\mu$  con las tablas habituales de la  $t$  de Student con un contraste unilateral. Sin embargo, la figura 9.9 nos indican que este método no es correcto, ya que la distribución en el muestreo del estimador  $\hat{\rho}$  no es normal con media uno, lo que es necesario para que el estadístico (9.7) sea una  $t$  de Student. Además, cuando el proceso es no estacionario tanto la estimación mínima cuadrática como la de MV exacta proporcionan un valor erróneo de la varianza del estimador  $\hat{\rho}$  y el estadístico (9.7) no sigue una distribución  $t$  de Student. La razón es que cuando la hipótesis nula es cierta y  $\rho = 1$ , el valor del parámetro se encuentra en el extremo del espacio paramétrico (0,1) y las condiciones de regularidad que son necesarias para las propiedades asintóticas del estimador MV no se verifican.

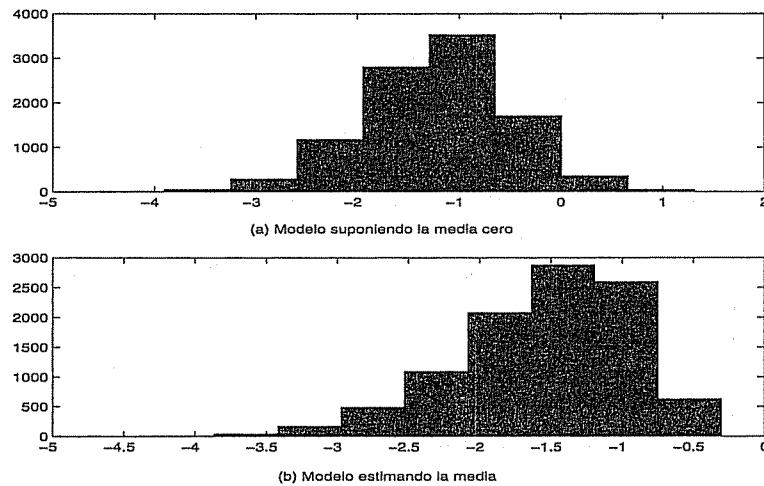
Como ilustración de este hecho la figura 9.10 presenta el resultado de calcular el estadístico (9.7) en 20 000 muestras de tamaño 100 generadas por paseos aleatorios. Se observa que la distribución del estadístico difiere mucho de la  $t$  de Student y ni siquiera tiene media cero. Por tanto, si utilizamos los valores críticos de la  $t$  de Student y rechazamos la hipótesis  $\rho = 1$  cuando el valor de (9.7) es mayor de  $-2$ , tendremos una probabilidad alta de rechazar que el proceso es un paseo aleatorio (y por tanto no estacionario) cuando realmente lo es.

El test de Dickey-Fuller se basa en los resultados anteriores. La hipótesis nula del contraste,  $H_0$ , es que la serie es no estacionaria y es necesario diferenciar,  $\rho = 1$ , y contrasta si esta hipótesis puede rechazarse a la vista de los datos. El estadístico del contraste es el (9.7), pero su valor se compara con la verdadera distribución del contraste, que no es la  $t$  de Student, sino la indicada en la figura 9.10. Una forma simple de obtener el valor de este estadístico (9.7) es escribir el modelo (9.6) que queremos estimar restando  $z_{t-1}$  a ambos miembros de la ecuación:

$$\nabla z_t = c + \alpha z_{t-1} + a_t, \quad (9.8)$$

donde  $\alpha = \rho - 1$ . Cuando escribamos un modelo incluyendo en la ecuación la variable en diferencias y en niveles, como en este caso, donde en (9.8) aparecen  $\nabla z_t$  y  $z_{t-1}$ , llamaremos al modelo un *modelo de corrección de error*. Este nombre proviene de sus primeras aplicaciones en el campo

**Figura 9.10 Distribución del estadístico t cuando se contrasta que el parámetro es uno en muestras que siguen paseos aleatorios**



de la predicción económica. En el modelo (9.8) si  $\alpha = 0$  tenemos un paseo aleatorio y si  $\alpha \neq 0$  un proceso estacionario. La hipótesis nula de que el proceso es no estacionario y  $\rho = 1$  se convierte en este modelo en la hipótesis nula  $H_0 : \alpha = 0$ , y la alternativa, que el proceso es estacionario, se convierte en  $H_1 : \alpha \neq 0$ . El contraste consiste en estimar en (9.8) el parámetro  $\alpha$  por mínimos cuadrados y rechazar que el proceso es estacionario si el valor  $t_\mu$  es significativamente pequeño. El estadístico (9.7) se escribe ahora:

$$t_\mu = \frac{\hat{\alpha}}{\hat{s}_\alpha}, \quad (9.9)$$

donde  $\hat{\alpha}$  es la estimación mínima cuadrática de  $\alpha$  en (9.8) y  $\hat{s}_\alpha$  su desviación típica estimada calculada de la forma habitual. Bajo la hipótesis de que  $\alpha = 0$  (que implica  $\rho = 1$ ) la distribución de  $t_\mu$  ha sido tabulada por Dickey y Fuller. Un extracto de sus tablas se presentan en el tabla 9.1. Observemos que la estimación de  $\hat{\alpha}$  será negativa ya que  $\rho \leq 1$ . Según indica la figura 9.10 la distribución del estadístico es distinta cuando incluimos o no una constante en el modelo y, para mayor seguridad, conviene siempre incluirla, ya que la alternativa es que el proceso es estacionario, pero no tiene por qué tener media cero. La decisión del contraste es:

Rechazar no estacionaridad si  $t_\mu \leq t_c$

donde el valor de  $t_c$  se obtiene de la tabla 9.1. Recomendamos realizar

**Tabla 9.1 Valores críticos para el contraste de Dickey-Fuller de raíz unitaria**

T	Sin constante				Con constante			
	.01	.025	.05	.1	.01	.025	.05	.1
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57
500	-2.58	-2.23	-1.95	-1.62	-3.44	-3.13	-2.87	-2.57
$\infty$	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57

el contraste con un nivel de significación pequeño, .01, de manera que la probabilidad de rechazar que el proceso es no estacionario cuando esta hipótesis sea cierta sea baja. Como hemos comentado, incluso aunque los datos se hayan generado por un AR(1) estacionario, si el parámetro está próximo a la unidad podemos obtener mejores predicciones suponiendo que el parámetro es la unidad y utilizando el paseo aleatorio que estimando el modelo estacionario, por lo que debemos cubrirnos especialmente del riesgo de rechazar que el proceso es no estacionario cuando realmente lo es.

### 9.3.3.2. El contraste de Dickey-Fuller aumentado

El contraste que hemos estudiado analiza si existe una raíz unitaria en un AR(1). Este contraste se generaliza para procesos ARMA. Comencemos estudiando el caso en el que tenemos un proceso AR( $p + 1$ ) y queremos saber si tiene una raíz unitaria. Es decir, se plantea la elección entre los dos modelos:

$$H_0 : \phi_p(B)\nabla z_t = a_t, \quad (9.10)$$

$$H_1 : \phi_{p+1}(B)z_t = c + a_t. \quad (9.11)$$

La hipótesis nula establece que la raíz mayor de un AR( $p + 1$ ) es igual a uno, modelo (9.10), y el proceso es no estacionario. La alternativa que es menor que uno, como en (9.11), y tendrímos un proceso estacionario.

Podríamos estimar directamente el modelo AR( $p + 1$ ), calcular las raíces del polinomio AR y ver si la mayor es suficientemente próxima a uno. Como en el caso anterior, la distribución del estadístico no es la  $t$  de Student y tendríamos que obtenerla por Monte Carlo. El problema ahora es más complicado, ya que para realizar un contraste necesitamos también obtener la desviación típica de la estimación de la raíz mayor

de un operador autorregresivo. En efecto, la estimación de un AR, que estudiaremos en el próximo capítulo, nos proporciona los parámetros y sus errores estándar, pero no los errores estándar de las raíces. Podríamos intentar obtener el error estándar de la raíz mayor del polinomio AR como función de los errores estándar de los coeficientes estimados, pero el proceso sería complicado. Un enfoque más simple, que es la idea de Dickey y Fuller, es intentar contrastar directamente la condición de una raíz unitaria en el operador  $\phi_{p+1}(B)$ .

Para implementar esta idea, se descompone el operador  $\phi_{p+1}(B)$  como

$$\phi_{p+1}(B) = (1 - \alpha_0 B) - (\alpha_1 B + \dots + \alpha_p B^p) \nabla. \quad (9.12)$$

Esta descomposición siempre puede hacerse, porque en ambos lados de la igualdad tenemos un polinomio en  $B$  de orden  $p+1$  con  $p+1$  coeficientes, con lo que identificando potencias de  $B$  en ambos miembros podemos obtener los  $p+1$  coeficientes  $\alpha_0, \dots, \alpha_p$ , dados los  $\phi_1, \dots, \phi_{p+1}$ . Vamos a ver que esta descomposición tiene la ventaja de trasladar la condición de una raíz unitaria en el primer miembro a una condición sobre un coeficiente que podemos estimar en el segundo miembro. En efecto, si  $\phi_{p+1}(B)$  tiene una raíz unidad, entonces  $\phi_{p+1}(1) = 0$ , y si hacemos  $B = 1$  en (9.12), el término  $(\alpha_1 B + \dots + \alpha_{p+1} B^p)(1 - B)$  se anulará y tendrá que verificarse que  $(1 - \alpha_0) = 0$ , es decir  $\alpha_0 = 1$ . La condición raíz unidad en  $\phi_{p+1}(B)$  implica  $\alpha_0 = 1$ . Vamos a ver que  $\alpha_0 = 1$  implica también una raíz unitaria: si  $\alpha_0 = 1$ , podemos escribir el polinomio de la izquierda como  $(1 - \alpha_1 B - \dots - \alpha_p B^p) \nabla$ , y el polinomio tiene una raíz unitaria. Por tanto, hemos comprobado que las dos afirmaciones siguientes son equivalentes: (1) el polinomio  $\phi_{p+1}(B)$  tiene una raíz unidad y (2) el coeficiente  $\alpha_0$  en la descomposición (9.12) es uno.

El modelo (9.11) puede escribirse, utilizando (9.12)

$$\phi_{p+1}(B) z_t = (1 - \alpha_0 B) z_t - (\alpha_1 B + \dots + \alpha_p B^p) \nabla z_t = c + a_t, \quad (9.13)$$

es decir:

$$z_t = c + \alpha_0 z_{t-1} + \sum_{i=1}^p \alpha_i \nabla z_{t-i} + a_t.$$

Para obtener directamente el estadístico del contraste podemos, como en el caso anterior, restar  $z_{t-1}$  de ambos miembros y escribir el modelo en la *forma de corrección de error*, es decir, con niveles y diferencias como regresores. Este modelo es:

$$\nabla z_t = c + \alpha z_{t-1} + \sum_{i=1}^p \alpha_i \nabla z_{t-i} + a_t \quad (9.14)$$

## 9. La identificación de los posibles modelos ARIMA

donde  $\alpha = (\alpha_0 - 1)$ . Esta ecuación (9.14) puede estimarse por mínimos cuadrados y el contraste de que la serie tiene una raíz unitaria,  $\alpha_0 = 1$  es equivalente al contraste  $\alpha = 0$ . Por tanto, el contraste es una generalización inmediata del contraste básico: estimamos un modelo donde explicamos la primera diferencia por una constante más la variable retardada, pero además añadimos como regresores los retardos de la diferencia de la variable. El contraste utiliza el mismo estadístico anterior:

$$t_\mu = \frac{\hat{\alpha}}{\hat{s}_\alpha}, \quad (9.15)$$

donde  $\hat{\alpha}$  es la estimación minimocuadrática de  $\alpha$  en (9.14) y  $\hat{s}_\alpha$  su desviación típica estimada calculada de la forma habitual. La distribución de  $t_\mu$  cuando la hipótesis de una raíz unitaria,  $\rho = 1$  es cierta, ha sido tabulada por Dickey y Fuller. De nuevo la distribución depende de que tengamos o no constante en el modelo, pero no es necesario corregir por la inclusión de los retardos adicionales, por lo que podemos utilizar la tabla 1.

Este contraste se generaliza para modelos ARMA. Supongamos que el modelo es  $\pi(B)z_t = c + a_t$ , donde  $\pi(B) = \phi(B)\theta^{-1}(B)$  y queremos contrastar si una de las raíces de  $\pi(B)$  es igual a la unidad, en cuyo caso podemos escribir  $\pi(B) = \pi^*(B)\nabla$ . Aproximando el operador  $\pi(B)$ , que puede ser infinito si existe parte MA, por un  $AR(k)$  largo, escribiremos  $\pi(B) \approx \phi_k(B)$ . La descomposición anterior, donde factorizamos un operador para separar la raíz unidad, es un caso particular de la descomposición de cualquier polinomio,  $a(B) = (1 - \alpha_1 B - \dots)$ , como:

$$a(B) = a(1)B + \nabla a^*(B) \quad (9.16)$$

donde si  $a(B)$  es infinito,  $a^*(B) = (1 - \alpha_1^* B - \dots)$  también lo es, y si  $a(B)$  es de orden  $k$ ,  $a^*(B)$  es de orden  $k - 1$ . Los coeficientes de  $a^*(B)$  se obtienen a partir de los de  $a(B)$ , igualando potencias en ambos polinomios y puede comprobarse que:

$$a_j^* = - \sum_{i=1}^p a_i; \quad j \geq 1$$

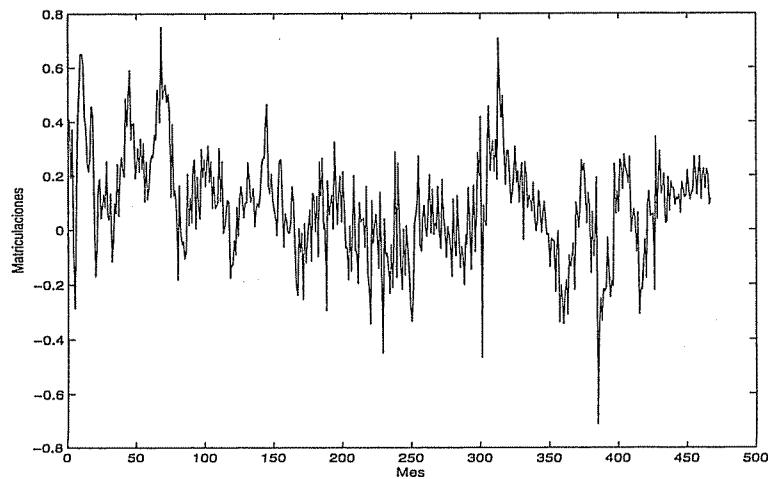
y utilizando (9.16) podemos escribir el modelo  $\pi(B)z_t = c + a_t$  como:

$$\pi(1)z_{t-1} + \pi^*(B)\nabla z_t = c + a_t,$$

o también, llamando  $\alpha = -\pi(1)$ ,

$$\nabla z_t = c + \alpha z_{t-1} + \pi_1^* \nabla z_{t-1} + \dots + \pi_p^* \nabla z_{t-p} + a_t,$$

**Figura 9.11 Gráfico de una diferencia estacional de orden 12 en el logaritmo de la serie de matriculación de vehículos**



que es análogo al modelo anterior.

En series estacionales hay que tener cuidado de introducir todos los retardos necesarios, para que al efectuar el contraste el ruido de la regresión sea ruido blanco.

---

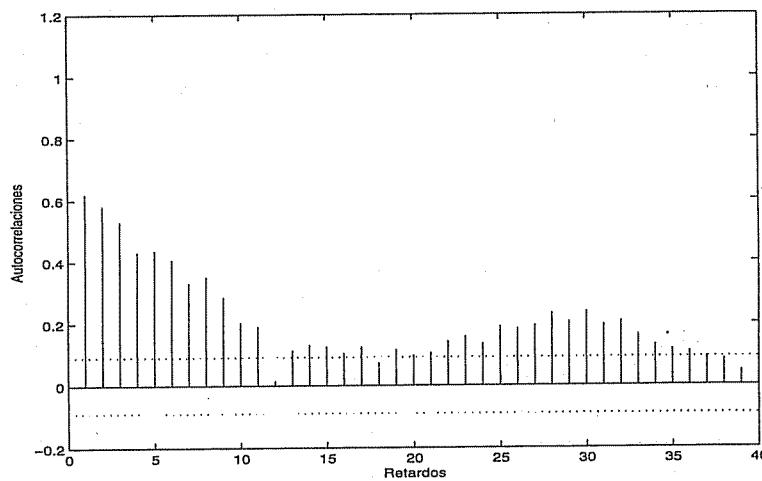
### Ejemplo 9.1

Como ilustración, vamos a realizar un contraste de raíz unitaria una vez tomada la diferencia estacional en la serie del logaritmo de matriculación de vehículos. El gráfico de la serie 9.11 muestra que la serie no tiene nivel constante, pero ya no muestra ninguna tendencia clara. La *fas* de esta serie, dada en la figura 9.12, tiene muchos coeficientes positivos que decrecen lentamente, sugiriendo la necesidad de una diferencia.

Supongamos que nos quedamos cortos en el número de retardos y, a la vista de la figura 9.12, decidimos introducir sólo los nueve retardos donde observamos coeficientes significativos al principio del gráfico. Entonces, el resultado del contraste suponiendo constante y hasta 10 retardos es  $t_\mu = -4.97$ . Como el valor crítico del estadístico, que obtenemos de la tabla 9.1 con tamaño muestral 500 y para  $\alpha = .01$ , es  $-3.44$ , concluimos que rechazamos la hipótesis de raíz unitaria, ya que el valor del estadístico es significativamente pequeño. Sin embargo, si miramos la fap obtenemos coeficientes de autocorrelación significativos para retardos tan altos como el 36, por lo que repetiremos el contraste incluyendo hasta 36 retardos. Ahora el valor del estadístico es  $t_\mu = -2.37$ , con

## 9. La identificación de los posibles modelos ARIMA

**Figura 9.12 Fas de la diferencia de orden 12 del logaritmo de la serie de matriculación de vehículos**



lo que no podemos rechazar la hipótesis nula de que la serie tiene una raíz unitaria, y es necesario diferenciar para convertirla en estacionaria.

Este ejemplo ilustra la importancia de incluir los retardos adecuados en el test.

### 9.4. La identificación de la estructura ARMA

La identificación de los órdenes  $p$  y  $q$  del modelo se realiza comparando las funciones estimadas de autocorrelación simple y parcial con las funciones teóricas de procesos ARMA.

Llamando  $\omega_t$  a la serie estacionaria,  $\omega_t = \nabla^d \nabla_s^D z_t$ , donde en la práctica  $d$  toma valores en  $(0, 1, 2)$  y  $D$  en  $(0, 1)$ , los coeficientes de autocorrelación muestrales se calculan por:

$$r_k = \frac{\sum_{t=d+sD+1}^{T-k} (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})}{\sum_{t=d+sD+1}^T (\omega_t - \bar{\omega})^2}, \quad k = 1, 2, \dots \quad (9.17)$$

y en los sumatorios hemos tenido en cuenta que al diferenciar perderemos  $d + sD$  observaciones.

Para juzgar cuándo un coeficiente  $r_k$  es distinto de cero necesitamos su error estándar, cuya determinación depende de la estructura del pro-

ceso. Una solución simple es tomar como error estándar  $1/\sqrt{T}$ , que es aproximadamente –según estudiamos en el capítulo 3– el error estándar de un coeficiente de correlación entre variables independientes. Si *todos* los coeficientes teóricos de autocorrelación fuesen nulos, las desviaciones típicas de estimación serían aproximadamente  $1/\sqrt{T}$ . Por tanto, podemos trazar bandas de confianza a  $\pm 2/\sqrt{T}$  y considerar como significativos, en primera aproximación, a los coeficientes fuera de dichas bandas.

Los coeficientes de autocorrelación parcial se obtienen con las regresiones:

$$\tilde{\omega}_t = \alpha_{k1}\tilde{\omega}_{t-1} + \dots + \alpha_{kk}\tilde{\omega}_{t-k}$$

donde  $\tilde{\omega}_t = \omega_t - \bar{\omega}$ . La secuencia  $\hat{\alpha}_{kk}$  ( $k = 1, 2, \dots$ ) de coeficientes minimocuadráticos estimados en estas regresiones es la función de autocorrelación parcial. Utilizaremos siempre en los gráficos de la *fas* los límites asintóticos  $\pm 2/\sqrt{T}$ , con la salvedad de que se trata de unos límites de referencia aproximados.

Si el proceso es estacional, estudiaremos los coeficientes de la *fas* y *fap* muestrales en los retardos  $s, 2s, 3s, \dots$ , para determinar la estructura ARMA estacional.

La identificación del modelo ARMA puede ser una tarea difícil. Con tamaños muestrales grandes y procesos puros AR o MA, la estructura de la *fas* y la *fap* muestrales suele indicar el orden requerido. Sin embargo, en general la interpretación de la *fas* y la *fap* muestral es compleja, por tres razones principales: (1) cuando existe autocorrelación las estimaciones de las autocorrelaciones están a su vez correladas, lo que introduce una pauta de variación aleatoria en la *fas* y la *fap* muestrales que se superpone a la debida al orden del modelo; (2) los límites de confianza que utilizamos para juzgar si las autocorrelaciones son distintas de cero,  $2/\sqrt{T}$ , son asintóticos y poco precisos para las primeras autocorrelaciones; (3) para procesos mixtos ARMA la estructura de la *fas* y la *fap* teórica es muy complicada, y puede ser extremadamente difícil estimar el orden del proceso, incluso conociendo los valores teóricos de las autocorrelaciones. Afortunadamente, no es necesario en la etapa de identificación decidir cuál es el orden del modelo, sino que basta con seleccionar un conjunto de modelos ARMA que parezca adecuado para representar los rasgos principales de la serie. Posteriormente estimaremos estos modelos y seleccionaremos el más adecuado.

La identificación con la función de autocorrelación simple y parcial muestral puede hacerse con las reglas siguientes:

1. Decidir cuál es el orden máximo de las partes AR y MA a partir de los rasgos obvios de la *fas* y la *fap*: coeficientes claramente significativos, pautas de decrecimiento geométricas o sinusoidales,

## 9. La identificación de los posibles modelos ARIMA

etc.

2. Evitar la identificación inicial de modelos mixtos ARMA y comenzar con modelos AR o MA, preferiblemente de orden bajo.
3. Utilizar las interacciones alrededor de los retardos estacionales, especialmente en la *fas*, para confirmar la concordancia entre la parte regular y la estacional. Como hemos visto en el capítulo 7, la *fas* de un proceso estacional presenta alrededor de los retardos estacionales la repetición de la estructura regular de autocorrelación, debido a la interacción entre ambas partes. Lo mismo ocurría con la *fap*, aunque la estructura es más complicada, como vimos en la sección 7.4.

En la práctica, la mayoría de las series reales pueden aproximarse bien con modelos ARMA con  $p$  y  $q$  menores que tres, para series no estacionales, y con  $P$  y  $Q$  menores que 2, para series estacionales. Existen procedimientos automáticos de selección de modelos, como el implantado en el programa TRAMO, que evitan la etapa de identificación y estiman todos los modelos posibles dentro de un subconjunto, que suele tomarse como  $p \leq 3, q \leq 2, P \leq 2, Q \leq 1$ . El modelo final se obtiene con los criterios de selección que estudiaremos en el próximo capítulo.

Además de seleccionar los órdenes  $(p, q)(P, Q)$  del modelo, debemos decidir en esta etapa si la serie estacionaria,  $\omega_t$ , tiene media distinta de cero. Para ello se calcula la media muestral

$$\bar{\omega} = \frac{\sum \omega_t}{T_c},$$

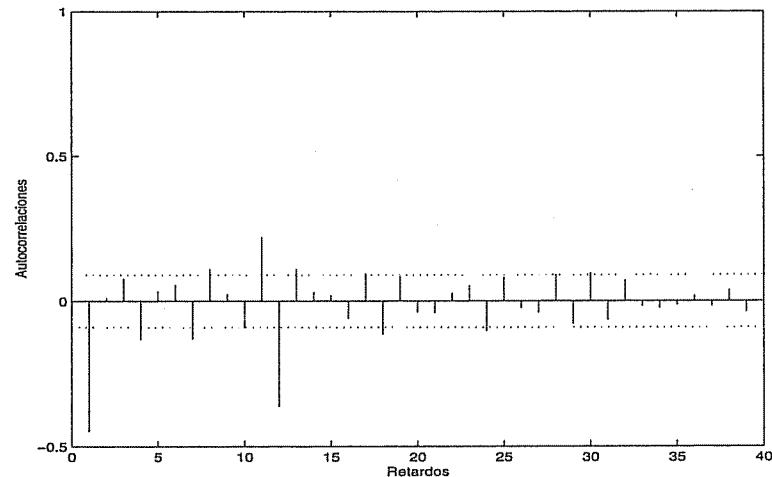
donde  $T_c$  es el número de sumandos (normalmente  $T_c = T - d - sD$ ). Su desviación típica puede aproximarse, según vimos en el capítulo 3, sección 3.5.1, por:

$$s(\bar{\omega}) \simeq \frac{s_\omega}{\sqrt{T}} (1 + 2r_1 + \dots + 2r_k)^{1/2},$$

donde  $s_\omega$  es la desviación típica de la serie estacionaria y  $r_i$  los coeficientes de autocorrelación estimados. En esta fórmula estamos suponiendo que los primeros  $k$  coeficientes autocorrelación son significativos y que  $k/T$  es despreciable. Si  $\bar{\omega} \geq 2s(\bar{\omega})$  admitiremos que la media del proceso estacionario es distinta de cero y la incluiremos como un parámetro a estimar; en caso contrario supondremos que  $E(\omega_t) = 0$ .

Como ilustración del proceso de identificación consideraremos la serie de matriculación de vehículos. Su *fas*, que se presenta en la figura 9.13, muestra como rasgos más destacados: (1) un coeficiente  $r_1$  significativo; (2) coeficientes significativos en los retardos estacionales,  $r_{12}$ ,  $r_{24}$  y  $r_{36}$ ;

**Figura 9.13 Función de autocorrelación simple de la serie de matriculaciones en logaritmos y con una diferencia regular y otra estacional**



(3) interacción alrededor de los retardos estacionales, manifestada en los valores positivos y simétricos de los coeficientes  $r_{11}$  y  $r_{13}$  y también  $r_{23}$  y  $r_{25}$  en la *fas* muestral. Además de estos rasgos destacados hay otros coeficientes significativos, de los que prescindiremos en primera instancia.

La parte regular sugiere un modelo MA(1), ya que hay un solo coeficiente distinto de cero en los primeros retardos y vemos una interacción que produce coeficientes significativos en un retardo alrededor de los retardos estacionales. La parte estacional es más compleja, ya que la estructura observada es compatible con un AR(1)<sub>12</sub> con coeficiente negativo (estructura de decrecimiento con signos alternos) y también con AR más largos o ARMA(1,1)<sub>12</sub>. El problema es que sólo tenemos tres coeficientes para obtener la estructura.

La *fap* de esta serie, que se presenta en la figura 9.14, confirma la estructura MA(1) para la parte regular: se observa un decrecimiento geométrico en los primeros retardos y también, por la interacción, que se repite después de los retardos estacionales. Los dos coeficientes significativos en los retardos estacionales nos hacen rechazar la hipótesis de un AR(1)<sub>12</sub>, pero son compatibles con un AR(2)<sub>12</sub> o también con un ARMA(1,1)<sub>12</sub> para la parte estacional. Por tanto, pasaremos a estimar modelos con parte MA(1) para la parte regular y AR(2) o ARMA(1,1) para la parte estacional.

## 9. La identificación de los posibles modelos ARIMA

Figura 9.14 Función de autocorrelación parcial para la serie de matriculación de vehículos en logaritmo y con una diferencia regular y otra estacional

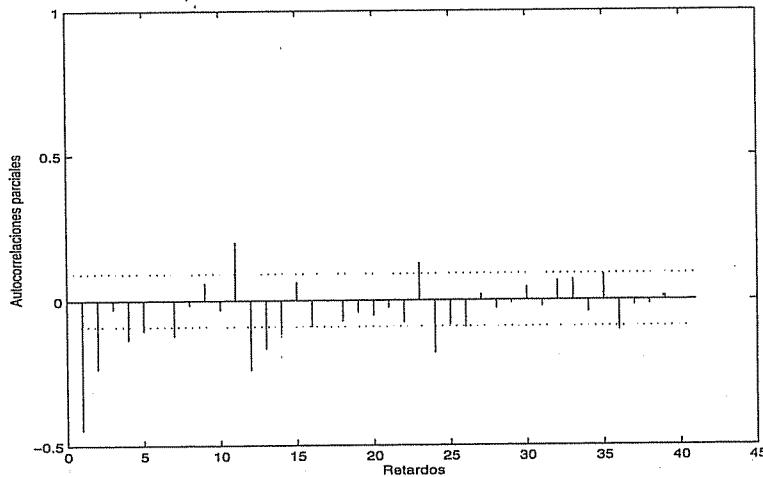
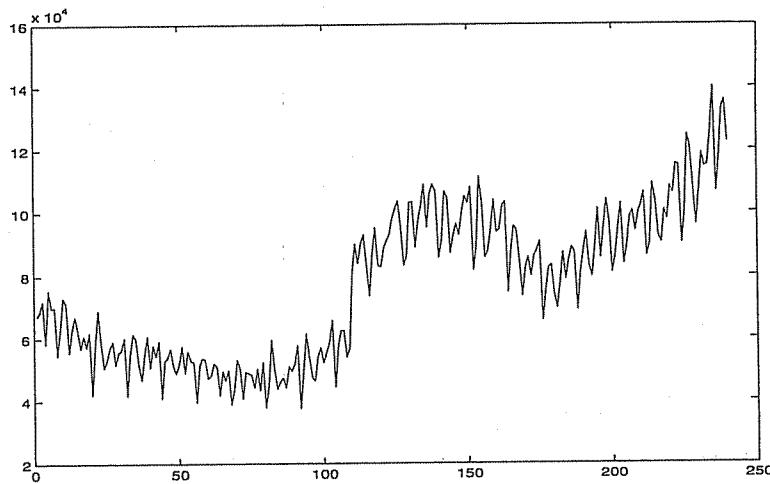


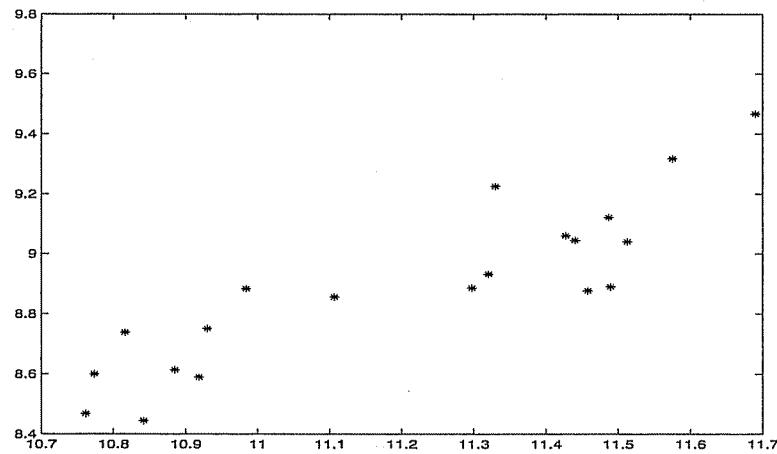
Figura 9.15 Serie de accidentes laborales en España



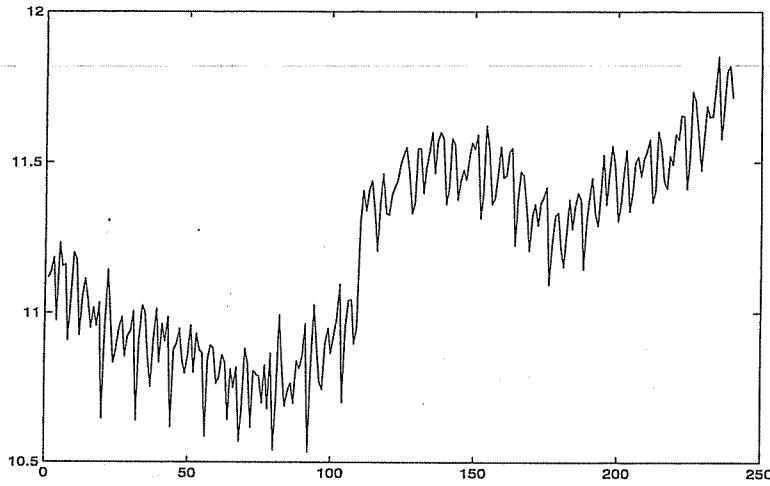
### Ejemplo 9.2

Vamos a identificar un modelo para la serie de accidentes laborales en España que se encuentra en el fichero accidentes.dat y que contiene 20 años de datos mensuales entre enero de 1979 hasta diciembre de 1998. La figura 9.15 presenta

**Figura 9.16 Relación entre el logaritmo de la variabilidad y el logaritmo de la media anual para la serie de los accidentes en España**



**Figura 9.17 Serie de accidentes laborales en logaritmos**

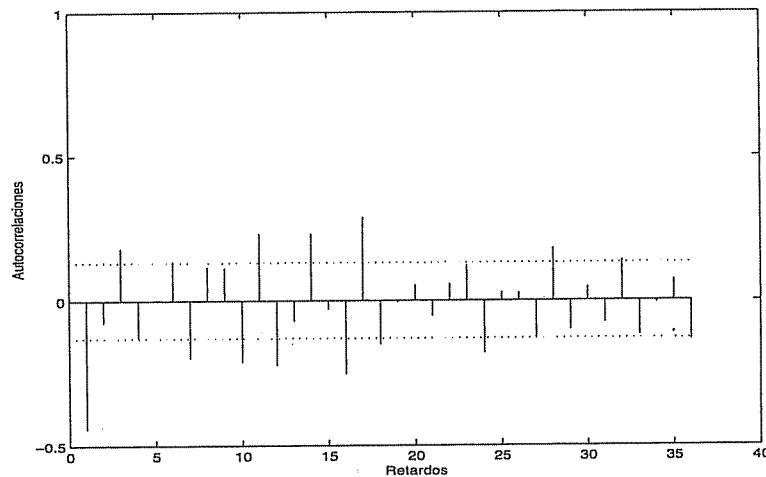


el gráfico de esta serie.

La serie parece mostrar un aumento de la variabilidad con el nivel. Para confirmar esta apreciación la figura 9.16 presenta la relación entre el logaritmo de la desviación típica cada año y el logaritmo de la media del año. Se observa una relación lineal, con pendiente un poco menor que la unidad, por lo que

## 9. La identificación de los posibles modelos ARIMA

**Figura 9.18 Fas del logaritmo de la serie de accidentes después de una diferencia regular y otra estacional**

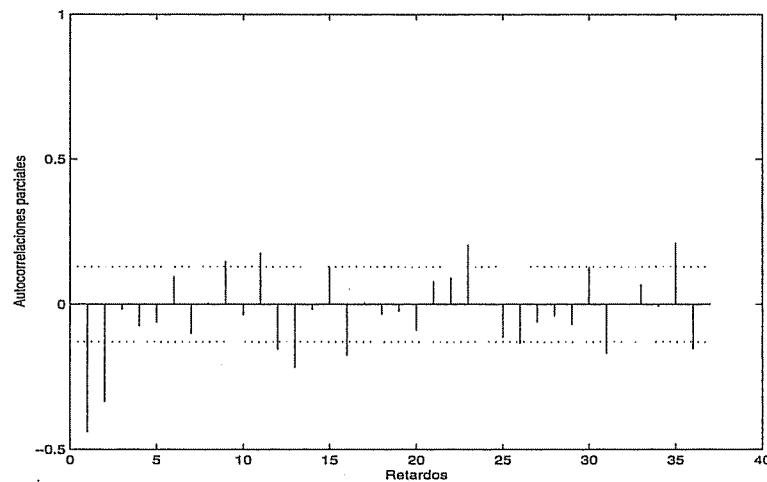


tomaremos logaritmos como primera aproximación. El gráfico de la serie en logaritmos se presenta en la figura 9.17. Es posible que la transformación logarítmica sea demasiado fuerte, porque la variabilidad de los primeros años ahora parece un poco mayor que la de los últimos. Hemos probado también con la raíz cuadrada, y el resultado es un poco mejor, pero por simplicidad en la interpretación trabajaremos con la serie en logaritmos.

El gráfico de la serie indica que es necesario al menos una diferencia regular para que la serie sea estacionaria y los picos del gráfico sugieren efecto estacional. La *fas* estimada de la transformación  $\nabla \log z_t$  muestra coeficientes altos y que se amortiguan lentamente en los retardos, 12, 24, 36,... por lo que tomamos una diferencia estacional. Las figuras 9.18 y 9.19 presentan el correlograma simple y parcial de la serie  $\nabla \nabla_{12} \log z_t$ . En la *fas* se observan coeficientes significativos en los retardos regulares 1 y 3, y en los retardos estacionales 12, 24 y 36. Además, aparecen varios coeficientes significativos alrededor de los retardos estacionales. Comenzando con la parte regular, el hecho de que aparezcan varios coeficientes significativos sugiere un proceso AR para la parte regular, lo que viene también apoyado por los numerosos coeficientes de interacción alrededor de los retardos estacionales. Con relación a la estacionalidad, existen retardos significativos en 12 y 24 y en el límite para el 36. La hipótesis más simple es una MA(2)<sub>12</sub>, pero también podría ser AR o ARMA.

Con relación a la *fap*, aparecen dos retardos significativos en la parte regular, lo que sugiere para esta parte un AR(2). En los retardos estacionales aparecen coeficientes significativos en los retardos 12 y 36, lo que sugiere que

**Figura 9.19** Fap del logaritmo de la serie de accidentes después de una diferencia regular y otra estacional



la estructura estacional puede ser o bien MA, o bien un AR mayor que dos, o, alternativamente, ARMA. Como conclusión de este análisis estimaremos en el próximo capítulo un  $AR(2) \times MA(2)_{12}$  y también modelos más complejos del tipo  $ARMA(2, 1) \times ARMA(2, 1)_{12}$ .

---

## 9.5. Lecturas complementarias

El libro de Box y Jenkins (1976) presenta un excelente análisis de la etapa de identificación de un modelo ARMA y el libro de Pankratz (1984) incluye numerosos ejemplos de identificar series reales con este enfoque. La identificación de modelos ARMA ha sido profusamente investigada en los años 80 y 90 y se han propuesto muchas herramientas valiosas con este objetivo. Una de las más utilizadas es la función de correlación extendida, propuesta por Tsay y Tiao (1984), que permite identificar modelos mixtos.

---

### Ejercicios 9

- 9.1 Un criterio antiguo para determinar el número de diferencias necesario para convertir una serie en estacionaria es el criterio de Titner, que consiste en diferenciar mientras disminuya la varianza de la serie resul-

## 9. La identificación de los posibles modelos ARIMA

---

tante y parar cuando la varianza aumente al tomar una nueva diferencia. Demostrar que si partimos de una serie estacionaria pero con autocorrelación de primer orden mayor que .5 la varianza de la serie disminuye al diferenciarla. (Sugerencia, utilice que llamando  $x_t$  a la serie original y  $n_t = \nabla x_t$ , entonces  $Var(n_t) = 2\sigma_x^2(1 - \rho_1)$ ).

- 9.2 Identificar un modelo para la serie airline.dat. Esta es la serie que da nombre al modelo de pasajeros de avión.
  - 9.3 Identificar un modelo para la serie de población de mayores de 16 años en España del fichero pobmay16.dat.
  - 9.4 Justificar que una muestra de 100 observaciones generadas por el modelo  $(1 - .2B)z_t = a_t$  puede fácilmente identificarse como generada por un MA(1) o un ARMA (1,1). (Sugerencia, exprese el modelo como MA(1) y tenga en cuenta que los límites de los coeficientes de la  $f_{as}$  y  $f_{ap}$  son  $T^{-1/2}$ ).
  - 9.5 Comprobar que el modelo de corrección de error para contrastar si el modelo verdadero es el M1:  $(1 - .7B)\nabla z_t = a_t$  o el M2:  $(1 - .7B)(1 - \phi B)z_t = a_t$ , con  $|\phi| < 1$ , es  $\nabla z_t = \alpha z_{t-1} + .7\nabla z_{t-1} + a_t$ . ¿Qué valores de  $\alpha$  indican cada uno de los dos modelos?
  - 9.6 Comprobar la equivalencia entre la condición  $\alpha_0 = 1$  y una raíz unitaria en el contraste de Dickey Fuller aumentado obteniendo los coeficientes  $\alpha_i$  como función de los  $\phi$ , igualando potencias en ambos polinomios. Comprobar que se obtiene  $\alpha_p = -\phi_{p+1}$ ,  $\alpha_{p-1} = -\phi_{p+1} - \phi_p$ , y en general  $\alpha_i = -\sum_{j=i+1}^{p+1} \phi_j$ ,  $i \geq 1$  y  $\alpha_0 = \phi_1 + \dots + \phi_{p+1}$ , que confirma que la condición  $(1 - \phi_1 - \dots - \phi_{p+1}) = 0$ , implica la condición  $\alpha_0 = 1$ .
- 

### Apéndice 9.1: La transformación Box-Cox para estabilizar la varianza

Vamos a demostrar que si existe una relación entre la media y la desviación típica de una variable aleatoria,  $x$ , del tipo:

$$\sigma_x = k\mu_x^\alpha,$$

la transformación  $y = x^{1-\alpha}$ , conduce a una nueva variable aleatoria,  $y$ , con una variabilidad constante, que no depende de la media de la variable transformada.

Para ello, supongamos que aplicamos a una variable  $x$  una transformación no lineal, pero continua y con derivadas continuas,  $y = h(x)$ . Podemos obtener

una expresión aproximada de los momentos de la variable transformada,  $y$ , en función de los de la original,  $x$ , desarrollando la función que les relaciona en serie de Taylor, alrededor de la media de la variable original:

$$y = h(x) \simeq h(\mu_x) + (x - \mu_x)h'(\mu_x) + \frac{1}{2}(x - \mu_x)^2h''(\mu_x),$$

y tomando esperanzas en ambos miembros:

$$E(y) \simeq h(\mu_x) + \frac{1}{2}\sigma_x^2h''(\mu_x),$$

y si despreciamos términos de segundo grado, podemos aproximar  $E(y)$  por  $h(\mu_x)$ . La varianza de la variable transformada será:

$$\sigma_y^2 = E(y - E(y))^2 \simeq E[(x - \mu_x)h'(\mu_x)]^2 = \sigma_x^2(h'(\mu_x))^2.$$

Supongamos que  $\sigma_x = k\mu_x^\alpha$  y tomamos como transformación  $h(x) = x^{1-\alpha}$ . Entonces, como  $h'(x) = (1-\alpha)x^{-\alpha}$ , sustituyendo en la expresión de la varianza

$$\sigma_y^2 = (k\mu_x^\alpha)^2((1-\alpha)\mu_x^{-\alpha})^2 = k^2(1-\alpha)^2$$

y la varianza de la variable transformada es constante y no depende de la media.

## Apéndice 9.2: Programas para la identificación

Vamos explicar cómo realizar la identificación del modelo para una serie temporal con algunos programas de ordenador.

### Minitab

Una vez introducidos los datos, en el menú superior marcar *Stat* → *Time series* → *Autocorrelacion* y obtendremos el correlograma de la serie. Análogamente, *Stat* → *Time series* → *Partial autocorrelacion* proporciona los parciales. En ese mismo menú, *Stat* → *Time series* se encuentra la opción diferenciar la serie y en *Cal* → *Calculator* se abre una ventana donde puede seleccionarse el tipo de transformación Box-Cox deseado para la serie. El contraste de Dickey-Fuller puede hacerse con la opción *Stat* → *regression*, pero antes hay que crear los regresores diferenciando y retardado la serie con las opciones *Stat* → *Time series* → *Differencing* y *Stat* → *Time series* → *lag*.

### SPSS

El tratamiento de este programa es similar a Minitab. En el menú principal la opción *Graph* → *Serie Temporal* → *autocorrelaciones* proporciona los simples y los parciales. Al marcar se abre un cuadro de diálogo que permite

## 9. La identificación de los posibles modelos ARIMA

indicar las diferencias que se desean aplicar y la transformación mediante el logaritmo.

### EViews

Una vez introducidos los datos, vemos el gráfico de la serie mediante *View → Linegraph*. Si se desea tomar el logaritmo, se hace con *Genr → lx = log(x)*, las diferencias regulares en el menú del workfile con : *Genr → dlx = d(lx, 1)* y las estacionales con: *ddlx = d(lx, 1, 12)*.

Para realizar el contraste de Dickey-Fuller ir a *Quick→Estimation Equation* y en la ventana introducir la primera diferencia de las variables, por ejemplo *dlx*, la constante, *c*, el retardo uno de la variable, *lx(-1)* y los retardos necesarios para tener ruido blanco, *dlx(-1 to k)*, siendo *k* el valor del máximo retardo considerado.

### Matlab

La instrucción *y=diff(x)* toma diferencias en la variable *x* y la guarda en la variable *y*, e *y = diff(x, 2)* toma dos diferencias. Matlab no incluye directamente las diferencias estacionales, pero pueden programarse fácilmente. La instrucción *z = reshape(x, s, n)* toma los datos del vector *x* por columnas y los organiza en *s* filas y *n* columnas. Por tanto esta instrucción nos permite pasar de un vector de *T = ns* datos con *n* años y *s* periodos estacionales a una matriz donde los *s* periodos estacionales de cada año aparecen en columnas y las filas nos indican el valor del mismo mes en distintos años. Si transponemos esta matriz, *tz = z'*, y la aplicamos diferencias con *dtz = diff(tz)* obtenemos las diferencias por columnas, que son las diferencias estacionales. Para volver a tener los datos como un vector columna primero transponemos esta matriz, para tener de nuevo los datos ordenados por columnas, y luego los colocamos como vector con la instrucción *dx = reshape(dtz', s \* (n - 1), 1)*.

### TSW

Para obtener la *fas* y la *fap* de una serie con el programa TSW hay que estimar un modelo que contenga sólo la media y las diferencias, ya que este programa está concebido para la estimación de muchas series. Después de leer los datos con el menú *Series* y comprobar con *Series list* que la serie se ha incluido en el programa, para identificar el modelo marcamos *+Model* en el menú. Aparece una ventana donde podemos cambiar el parámetro RSA para identificación automática. Para identificar un modelo para la serie lo dejamos en 0. Pasamos a la pestaña ARIMA model y marcamos únicamente las opciones: P=0; Q=0; D= diferencias regulares a tomar; BP=0 (orden del operador AR estacional); BQ=0 (orden del operador MA estacional); BD= diferencias estacionales a tomar; INIT=0, este parámetro sirve para inicializar la estimación de los parámetros; IMEAN=0, este parámetro sirve para corregir la media o

## Análisis de series temporales

---

no. Esto significa que no corregimos la media; LAM=-1, este parámetro dice si transformar logarítmicamente los datos. Vale 0 si se quiere transformar, 1 si no, y -1 si el programa realiza un contraste sobre la necesidad de transformar; FCT=1, relacionado con el contraste, se deja así.

Como ya hemos indicado, una propiedad muy útil del programa es que si marcamos cualquier parámetro del programa y pulsamos la tecla F1 aparece una explicación de la interpretación del parámetro.

Marcamos **Run** en el menú y en **output** obtenemos la salida del programa. Se obtienen las funciones de autocorrelación simple y parcial de la serie con la transformación fijada. En la opción **Graph** podemos obtener gráficos de estas funciones. Como no se ha ajustado ningún modelo cuando el programa indica residuos, éstos equivalen a la serie transformada.

---

## 10. Estimación y selección de modelos ARIMA



Rudolf E. Kalman (1930)

Matemático húngaro. Estudió en el MIT y fue profesor en Stanford y Florida. Desarrolló los fundamentos modernos de los sistemas dinámicos de control y obtuvo en 1961, con Richard Bucy, el procedimiento recursivo de estimación del estado de un sistema dinámico que se conoce como filtro de Kalman. Este método es de uso habitual en ingeniería de control, tanto en el campo aeroespacial como en el industrial y una de las herramientas más utilizadas para la estimación y previsión de series temporales.

### 10.1. Introducción

Este capítulo tiene dos partes. En la primera estudiaremos la estimación de los parámetros de un modelo ARMA. En la segunda, la selección de un modelo ARMA entre varios modelos estimados. Este capítulo es más técnico que los anteriores y el lector interesado en las aplicaciones puede prescindir de los detalles de la estimación sin pérdida de continuidad.

Comenzaremos suponiendo que se dispone de una serie estacionaria de  $T$  elementos  $\{\omega_t\} t = 1, \dots, T$ , y queremos estimar los parámetros de un modelo ARMA especificado. Utilizaremos la notación  $\omega_t$  porque es frecuente que la serie a estimar sea una transformación de la serie original,  $z_t$ . Por ejemplo, con datos mensuales económicos es frecuente que  $\omega_t = \nabla \nabla_{12} \ln z_t$ .

Supondremos que no se dispone de información a priori sobre los valores de los parámetros del proceso y que, como suele ser habitual, disponemos de una muestra grande. En este caso la estimación Bayesiana y la estimación clásica por máxima verosimilitud (MV) conducen a re-

sultados similares, aunque la interpretación de los resultados es distinta. Para simplificar la exposición presentamos únicamente la estimación MV, aunque el filtro de Kalman, la herramienta central de estimación de modelos ARMA, puede justificarse tanto desde el punto de vista clásico como del Bayesiano.

El estudio de la estimación comienza con el caso más simple: la estimación condicionada de procesos AR, que es similar a la estimación por mínimos cuadrados de un modelo de regresión. A continuación, veremos la estimación exacta de procesos AR que requiere la utilización de algoritmos de optimización para problemas no lineales. Después, estudiaremos la estimación de modelos MA y ARMA, que es siempre no lineal y requiere dos etapas: la primera, calcular cuanto vale la función de verosimilitud dado un valor de los parámetros; la segunda, encontrar un nuevo valor de los parámetros que haga más grande el valor de la función. La estimación consiste en iterar entre estas dos fases, hasta obtener el máximo de la función. Nos centraremos en cómo evaluar la función de verosimilitud mediante un algoritmo eficiente, el filtro de Kalman. En el apéndice 10.1 presentaremos brevemente los fundamentos de los métodos de optimización que se utilizan para obtener el máximo de la función.

La segunda parte del capítulo considera el caso en el que disponemos de varios modelos ARIMA estimados para una serie y se plantea el problema de seleccionar el más adecuado. Los criterios de ajuste no son útiles para seleccionar modelos, porque si aumentamos el número de parámetros *siempre aumentará el ajuste del modelo*. Tenemos que acudir entonces a criterios que equilibran el ajuste con el número de parámetros estimados.

## 10.2. La función de verosimilitud de un proceso ARMA

Supongamos que tenemos un proceso ARMA y se desea estimar los parámetros por máxima verosimilitud. Para ello debemos escribir la función de densidad conjunta y maximizarla respecto a los parámetros, considerando a los datos como fijos. Para escribir la densidad conjunta de las  $T$  observaciones  $\omega_T = (\omega_1, \dots, \omega_T)$ , vamos a utilizar que la distribución conjunta de dos variables cualesquiera,  $\mathbf{x}$ ,  $\mathbf{y}$ , escalares o vectoriales, puede siempre escribirse como el producto de dos distribuciones: la marginal de la primera variable y la distribución de la segunda condicionada a los valores de la primera. Es decir:

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) f(\mathbf{y}|\mathbf{x}). \quad (10.1)$$

Esta expresión sigue siendo cierta si todas las funciones de densidad van condicionadas a otra variable  $\mathbf{z}$ , de manera que:

$$f(\mathbf{x}, \mathbf{y}|\mathbf{z}) = f(\mathbf{x}|\mathbf{z})f(\mathbf{y}|\mathbf{x}, \mathbf{z}). \quad (10.2)$$

Consideremos la función de densidad conjunta de las  $T$  observaciones  $\omega_T$ . Tomando en (10.1)  $\mathbf{x} = \omega_1$  y  $\mathbf{y} = (\omega_2, \dots, \omega_T)$ , podemos escribir

$$f(\omega_T) = f(\omega_1)f(\omega_2, \dots, \omega_T|\omega_1)$$

y descomponiendo el segundo término, con (10.2) haciendo  $\mathbf{z} = \omega_1$ ,  $\mathbf{x} = \omega_2$  y  $\mathbf{y} = (\omega_3, \dots, \omega_T)$ , resulta:

$$f(\omega_T) = f(\omega_1)f(\omega_2|\omega_1)f(\omega_3, \dots, \omega_T|\omega_1, \omega_2)$$

y repitiendo este proceso, obtenemos finalmente

$$f(\omega_T) = f(\omega_1)f(\omega_2|\omega_1)f(\omega_3|\omega_2, \omega_1)\cdots f(\omega_T|\omega_{T-1}, \dots, \omega_1). \quad (10.3)$$

Esta expresión permite escribir la función de densidad conjunta de las  $T$  variables como producto de  $T$  distribuciones univariantes. La diferencia entre esta representación y la que se obtiene con datos independientes es que en lugar de tener el producto de las marginales de cada dato tenemos la marginal de la primera y el producto de las condicionadas de cada dato, dados los anteriores.

La descomposición (10.3) permite escribir la verosimilitud de un modelo ARMA, ya que si suponemos normalidad, todas las distribuciones condicionadas serán normales. Su esperanza, como hemos visto en el capítulo 8, es la predicción a un paso que minimiza el error cuadrático de predicción, y escribiremos:

$$E(\omega_t|\omega_{t-1}, \dots, \omega_1) = \hat{\omega}_{t-1}(1) = \omega_{t|t-1}.$$

La notación  $\hat{\omega}_{t-1}(1)$  es la que utilizamos en el capítulo 8 y es conveniente cuando pensamos en predicciones a distintos horizontes, porque aparece claramente el origen de la predicción y el horizonte. Como en este capítulo vamos a considerar únicamente predicciones a un paso, es más simple utilizar  $\omega_{t|t-1}$  para la esperanza de la variable  $\omega_t$  cuando disponemos de los valores previos de una realización finita,  $\omega_{t-1}, \dots, \omega_1$ , y conocemos los parámetros del proceso. Vamos a llamar  $e_t$  al error en la predicción de  $\omega_t$  hecha desde  $\omega_{t-1}$  con la información de  $\omega_{t-1}, \dots, \omega_1$  y conociendo los parámetros del proceso. Esto es, de nuevo, una simplificación de la notación  $e_{t-1}(1)$  introducida en el capítulo 8. Entonces,  $e_t = \omega_t - \omega_{t|t-1}$ .

Estos errores de predicción están muy relacionados con las innova-

ciones del proceso, aunque, como vimos en el capítulo 8, no coinciden exactamente con ellas por un problema de valores iniciales. Para ver esta diferencia, supongamos una serie,  $\omega_1, \dots, \omega_T$ , de media cero y tamaño  $T$ , generada por el proceso AR(1) con parámetro  $\phi$  conocido. Entonces, como  $\omega_{t|t-1} = \phi\omega_{t-1}$  para  $t = 2, \dots, T$  pero  $E(\omega_1) = \phi E(\omega_0) = 0$ , podemos calcular los errores de predicción a un paso mediante:

$$\begin{aligned} e_1 &= \omega_1 - \phi E(\omega_0) = \omega_1 \\ e_2 &= \omega_2 - \phi\omega_1 \\ &\vdots \quad \vdots \quad \vdots \\ e_T &= \omega_T - \phi\omega_{T-1}. \end{aligned}$$

Vemos que los errores de predicción,  $e_2, \dots, e_T$ , coinciden con las innovaciones de modelo,  $a_2, \dots, a_T$ , donde  $\omega_t = \phi\omega_{t-1} + a_t$ , para  $t = 2, \dots, T$ . La diferencia aparece en el primer error de predicción a un paso,  $e_1$ , que no es igual a la innovación en ese punto,  $a_1 = \omega_1 - \phi\omega_0$ , sino a su estimación  $\omega_1$ . Esto hace que para  $t = 2, \dots, T$  la varianza de los errores de predicción a un paso sea  $\sigma^2$ , la de la innovación, mientras que para  $t = 1$  sea distinta. En general, podemos escribir:

$$Var(\omega_t | \omega_{t-1}, \dots, \omega_1) = \sigma^2 v_{t|t-1}$$

donde para un AR(1)  $v_{t|t-1} = 1$  para  $t = 2, \dots, T$ , pero como  $Var(\omega_1) = E(\omega_1^2) = \sigma^2(1 - \phi^2)^{-1}$  concluimos que  $v_{1|0} = (1 - \phi^2)^{-1}$ . Hemos introducido la constante  $\sigma^2 = var(a_t)$  en la expresión de las varianzas condicionadas porque sale como factor común y simplifica la presentación.

Con esta notación, la función de densidad conjunta de la muestra para un proceso ARMA general puede escribirse como:

$$f(\omega_T) = \prod_{t=1}^T \sigma^{-1} v_{t|t-1}^{-1/2} (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T \frac{(\omega_t - \omega_{t|t-1})^2}{v_{t|t-1}} \right\}$$

y tomando logaritmos, llamando  $\beta = (\mu, \phi_1, \dots, \phi_q, \sigma^2)$  al vector de parámetros, la función de verosimilitud o la función soporte es:

$$L(\beta) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sum_{t=1}^T \ln v_{t|t-1} - \frac{1}{2\sigma^2} \sum_{t=1}^T \frac{e_t^2}{v_{t|t-1}} \quad (10.4)$$

donde tanto las varianzas condicionales,  $v_{t|t-1}$ , como los errores de predicción a un paso,  $e_t$ , dependen de los parámetros. Por tanto, evaluar la función de verosimilitud se reduce a calcular los errores de predicción a un paso de cada observación dadas las anteriores y sus varianzas.

## 10. Estimación y selección de modelos ARIMA

La maximización de la función de verosimilitud exacta se efectúa con un algoritmo de optimización no lineal. Se utiliza mucho el basado en el algoritmo de Gauss-Newton con una modificación debida a Marquardt, que se describe en el apéndice 10.1.

### 10.3. Procesos AR

#### 10.3.1. El proceso AR(1)

Como ilustración, consideremos el proceso AR(1) de media  $\mu$ ,  $\omega_t = c + \phi\omega_{t-1} + a_t$ , con  $c = \mu(1 - \phi)$ . En este caso, según vimos en el capítulo 4:

$$E(\omega_1) = \mu \quad (10.5)$$

y,

$$Var(\omega_1) = E(\omega_1 - \mu)^2 = \frac{\sigma^2}{1 - \phi^2}. \quad (10.6)$$

Entonces con la notación anterior,  $\omega_{1|0} = \mu$  y  $v_{1|0} = (1 - \phi^2)^{-1}$ . Para  $\omega_2$  los momentos de la distribución condicionada a  $\omega_1$  son  $\omega_{2|1} = E(\omega_2|\omega_1) = c + \phi\omega_1$  y,

$$Var(\omega_2|\omega_1) = E[(\omega_2 - c - \phi\omega_1)^2] = E(a_2^2) = \sigma^2$$

con lo que  $v_{2|1} = 1$ . De la misma forma, comprobamos que

$$\omega_{t|t-1} = E(\omega_t|\omega_{t-1}) = c + \phi\omega_{t-1}, \quad t = 2, \dots, T$$

y,

$$Var(\omega_t|\omega_{t-1}) = \sigma^2 v_{t|t-1} = \sigma^2, \quad t = 2, \dots, T.$$

En consecuencia, la función de verosimilitud será:

$$f(\omega_T) = f(\omega_1) \prod_{t=2}^T \sigma^{-1} (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^T (\omega_t - c - \phi\omega_{t-1})^2 \right\}. \quad (10.7)$$

Tomando logaritmos y utilizando que  $f(\omega_1)$  es normal, con parámetros dados por (10.6) y (10.5), resulta la función soporte:

$$\begin{aligned} L(\phi, \sigma^2 | \omega_T) &= \frac{-T}{2} \ln \sigma^2 + \frac{1}{2} \ln (1 - \phi^2) - \\ &\quad \frac{(1 - \phi^2)(\omega_1 - \mu)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{t=2}^T (\omega_t - c - \phi\omega_{t-1})^2. \end{aligned} \quad (10.8)$$

Para obtener el estimador de  $\phi$  tendremos que derivar respecto a este parámetro e igualar a cero el resultado. Se obtiene una ecuación cúbica que tiene tres raíces, y la raíz que maximice la función de verosimilitud es el estimador MV.

La expresión (10.8) muestra que, si prescindimos del primer término, la función soporte tiene la expresión habitual de un modelo lineal. Si condicionamos a la primera observación tenemos que:

$$f(\omega_2, \dots, \omega_T | \omega_1) = \prod_{t=2}^T \sigma^{-1} (2\pi)^{-1/2} \times \\ \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^T (\omega_t - c - \phi\omega_{t-1})^2 \right\}.$$

Definimos la verosimilitud condicionada como la asociada a esta función de densidad conjunta:

$$L_C(\phi, \sigma^2 | \omega_1) = \frac{-(T-1)}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T (\omega_t - c - \phi\omega_{t-1})^2$$

y el estimador del parámetro  $\phi$  que maximiza esta verosimilitud condicionada se obtiene minimizando la suma de cuadrados:

$$\sum_{t=2}^T (\omega_t - c - \phi\omega_{t-1})^2 = \sum_{t=2}^T (\tilde{\omega}_t - \phi\tilde{\omega}_{t-1})^2$$

donde  $\tilde{\omega}_t = \omega_t - \mu$ . Derivando e igualando a cero, dado el estimador  $\bar{\omega} = \hat{\mu} = \sum_{t=1}^T \omega_t / T$ , para la media que obtuvimos en el capítulo 3, el estimador de  $\phi$  es:

$$\hat{\phi} = \frac{\sum_{t=2}^T (\omega_t - \bar{\omega})(\omega_{t-1} - \bar{\omega})}{\sum_{t=2}^T (\omega_{t-1} - \bar{\omega})^2},$$

que es el de la pendiente en un modelo de regresión de  $\omega_t$  respecto a  $\omega_{t-1}$ . El estimador MV condicionado de la varianza será:

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (\omega_t - \hat{c} - \hat{\phi}\omega_{t-1})^2}{T-1},$$

donde  $\hat{c} = \bar{\omega}(1 - \hat{\phi})$ .

En conclusión, si condicionamos al primer término y escribimos la verosimilitud de las observaciones desde la 2 a la T, tenemos un modelo lineal en los parámetros. La diferencia entre el estimador obtenido con

la verosimilitud condicionada y la exacta será en general pequeña, y despreciable para muestras grandes.

### 10.3.2. Procesos AR(p)

Consideremos un proceso AR( $p$ ) general. La esperanza condicionada de  $\omega_t$ , para  $t = p+1, \dots, T$  dados los datos previos,  $\omega_{t-1}, \dots, \omega_1$  será, utilizando la ecuación del AR( $p$ ):

$$E[\omega_t | \omega_{t-1}, \dots, \omega_1] = \mu + \phi_1(\omega_{t-1} - \mu) + \dots + \phi_p(\omega_{t-p} - \mu)$$

y su varianza condicionada será:

$$\text{Var}(\omega_t | \omega_{t-1}, \dots, \omega_1) = \text{Var}(a_t) = \sigma^2.$$

En consecuencia, todas las distribuciones condicionadas para  $t = p+1, \dots, T$  son normales, con media igual a la predicción a un paso y varianza  $\sigma^2$ . La función soporte condicionada se obtendrá a partir de la densidad conjunta de las observaciones  $(\omega_{p+1}, \dots, \omega_T)$  condicionadas a las  $p$  primeras. Su expresión es:

$$\begin{aligned} L_C(\mu, \phi, \sigma^2 | \omega_1, \dots, \omega_p) &= -\frac{(T-p)}{2} \ln \sigma^2 - \\ &- \frac{1}{2\sigma^2} \sum_{t=p+1}^T (\omega_t - \mu - \sum_{i=1}^p \phi_i(\omega_{t-i} - \mu))^2. \end{aligned} \quad (10.9)$$

Maximizar esta función respecto a  $\mu$  y  $\phi$  equivale a minimizar la suma de cuadrados de los errores de predicción a un paso, que, al no depender de las condiciones iniciales serán iguales a las innovaciones y podemos escribir:

$$S = \sum_{t=p+1}^T a_t^2 = \sum_{t=p+1}^T (\omega_t - \mu - \sum_{i=1}^p \phi_i(\omega_{t-i} - \mu))^2 \quad (10.10)$$

donde  $a_t = (\omega_t - \mu - \sum_{i=1}^p \phi_i(\omega_{t-i} - \mu))$ . Por tanto, maximizar el soporte condicional equivale a mínimos cuadrados. El estimador de  $\mu$  se obtiene derivando e igualando a cero:

$$\sum_{t=p+1}^T (\omega_t - \mu - \sum_{i=1}^p \phi_i(\omega_{t-i} - \mu)) = 0$$

y suponiendo que  $\sum_{t=p+1}^T \omega_t \approx \sum_{t=p+1}^T \omega_{t-i}$ , lo que será aproximadamente cierto si  $T$  es grande, obtenemos que el estimador de la media es la media

muestral de las observaciones consideradas:

$$\hat{\mu} = \frac{\sum_{t=p+1}^T \omega_t}{T-p}.$$

Un estimador mejor de  $\mu$  es  $\bar{\omega} = \sum_{t=1}^T \omega_t / T$ , la media muestral de todas las observaciones, que estudiamos en el capítulo 3 y que hemos utilizado para el AR(1) en la sección anterior. Ambos estimadores son centrados, pero el calculado con toda la muestra tiene menor varianza (véase el ejercicio 10.1) y será el que utilicemos. Esto equivale a estimar inicialmente la media con todos los datos y después escribir la verosimilitud para las variables en desviaciones a la media.

Para obtener el estimador de  $\phi$ , sustituyendo  $\mu$  por  $\bar{\omega}$  en (10.10) y llamando  $\mathbf{x}'_t = (\omega_{t-1} - \bar{\omega}, \dots, \omega_{t-p} - \bar{\omega})$ , se obtiene el estimado habitual de mínimos cuadrados en modelos de regresión:

$$\hat{\phi} = \left( \sum_{t=p+1}^T \mathbf{x}_t \mathbf{x}'_t \right)^{-1} \left( \sum_{t=p+1}^T \mathbf{x}_t (\omega_t - \bar{\omega}) \right). \quad (10.11)$$

Esta expresión para muestras grandes será aproximadamente,

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p, \quad (10.12)$$

donde

$$\hat{\Gamma}_p = \begin{vmatrix} \hat{\gamma}_0 & \dots & \hat{\gamma}_{p-1} \\ \dots & \hat{\gamma}_0 & \dots \\ \hat{\gamma}_{p-1} & \dots & \hat{\gamma}_0 \end{vmatrix}, \quad \hat{\gamma}_p = \begin{vmatrix} \hat{\gamma}_1 \\ \dots \\ \hat{\gamma}_p \end{vmatrix},$$

que son las ecuaciones de Yule-Walker. Sin embargo, en muestras pequeñas ambos estimadores son diferentes y tanto más diferentes cuanto mayor sea el orden del proceso. Los estimadores de mínimos cuadrados consideran la muestra de  $T - p$  observaciones y todos los coeficientes de (10.11) se calculan como cocientes de sumas de  $T - p$  términos en el numerador y en el denominador. Los estimadores de Yule-Walker utilizan distintos términos en cada sumando:  $\hat{\gamma}_0$  se calcula con los  $T$  datos pero  $\hat{\gamma}_k$  con  $T - k$ . Este desequilibrio introduce sesgos importantes y puede demostrarse (Tjostheim y Paulsen, 1983) que los estimadores de mínimos cuadrados son más precisos que los de Yule-Walker.

La estimación por máxima verosimilitud exacta requiere calcular las esperanzas y varianzas condicionadas para las  $p$  primeras observaciones. Esto puede hacerse de manera similar a como se hizo para un AR(1), y veremos más adelante un procedimiento general para obtenerlas.

## 10.4. Estimación de modelos MA y ARMA

### 10.4.1. Estimación MV condicional

La estimación de modelos con componentes MA y mixtos es más complicada que la de los AR por dos razones. En primer lugar, la función de verosimilitud, tanto la condicional como la exacta, es siempre no lineal en los parámetros. En segundo, el procedimiento de condicionar a ciertos valores iniciales, que lleva a resultados simples en los AR, es más complicado para procesos MA y ARMA, haciendo el cálculo de las esperanzas y varianzas condicionadas más difícil. Para ilustrar estos problemas, consideremos el caso de un MA(1):

$$\omega_t = a_t - \theta a_{t-1}$$

con esperanza marginal cero. La esperanza de  $\omega_t$  condicionada a sus valores previos ya no es inmediata, como en los AR, y para obtenerla tenemos que expresar  $\omega_t$  en función de los valores anteriores. Comenzando con  $t = 2$ , como  $\omega_2 = a_2 - \theta a_1$ , y  $a_1 = \omega_1 + \theta a_0$ , tenemos que:

$$\omega_2 = -\theta\omega_1 + a_2 - \theta^2 a_0$$

y tomando esperanzas en esta expresión y suponiendo  $E(a_0|\omega_1) = 0$ , deducimos que la esperanza de la distribución condicionada es:

$$E(\omega_2|\omega_1) = -\theta\omega_1,$$

y la varianza:

$$var(\omega_2|\omega_1) = \sigma^2(1 + \theta^4).$$

Procediendo de esta forma para  $t = 3, 4, \dots$ , se obtiene que

$$\omega_t = -\theta\omega_{t-1} - \theta^2\omega_{t-2} - \dots - \theta^{t-1}\omega_1 + a_t - \theta^t a_0,$$

que conduce a

$$E(\omega_t|\omega_{t-1}, \dots, \omega_1) = -\theta\omega_{t-1} - \theta^2\omega_{t-2} - \dots - \theta^{t-1}\omega_1,$$

y

$$var(\omega_t|\omega_{t-1}, \dots, \omega_1) = \sigma^2(1 + \theta^{2t}).$$

Estas expresiones son no lineales en los parámetros y pesadas de calcular por este método de sustitución en procesos MA( $q$ ) generales.

Un enfoque alternativo es condicionar también en las primeras innovaciones no observadas. Observemos que para cada valor de los parámetros

$\theta$ , la expresión:

$$a_t = \omega_t + \theta a_{t-1} \quad (10.13)$$

permite calcular recursivamente las perturbaciones  $a_t$ , condicionadas a un valor inicial  $a_0$ . Tomando  $a_0 = 0$  podemos calcular todas las restantes perturbaciones a partir de los  $\omega_t$ . Entonces:

$$E(\omega_t | \omega_{t-1}, \dots, \omega_1, a_0) = -\theta a_{t-1}$$

y

$$\text{var}(\omega_t | \omega_{t-1}, \dots, \omega_1, a_0) = E[(\omega_t + \theta a_{t-1})^2] = E[a_t^2] = \sigma^2$$

que conduce a la verosimilitud condicionada en logaritmos o soporte condicional:

$$L_C(\theta | \omega_1, a_0) = \frac{-(T-1)}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T a_t^2.$$

La maximización de esta función se realiza mediante un algoritmo no lineal que se describe en el apéndice 10.1

La estimación condicionada de modelos ARMA( $p, q$ ) se realiza siguiendo los mismos principios. Llamando  $r = \max(p, q)$  y  $\beta = (\mu, \phi_1, \dots, \phi_q, \sigma^2)$  al vector de parámetros, la función soporte condicional es:

$$L_C(\beta | a_0, \omega_p) = \frac{-(T-r)}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{T=r+1}^T a_t^2 \quad (10.14)$$

donde tenemos que  $a_t^2 = a_t^2(\beta | a_0, \omega_p)$ , para resaltar que las innovaciones se calculan a partir de unos vectores  $a_0$  y  $\omega_p$  de valores iniciales. Estas innovaciones estimadas o residuos se calculan recursivamente mediante:

$$\hat{a}_t = \omega_t - c - \phi_1 \omega_{t-1} - \dots - \phi_p \omega_{t-p} + \theta_1 \hat{a}_{t-1} + \dots + \theta_q \hat{a}_{t-q} \quad t = r+1, \dots, T \quad (10.15)$$

donde  $c = \mu(1 - \phi_1 - \dots - \phi_p)$  y se supone que los primeros  $r$  residuos son cero. La maximización de (10.14) requiere un valor inicial de los parámetros que puede obtenerse con el algoritmo de Hannan y Rissanen que se presenta a continuación.

#### 10.4.1.1. El algoritmo de Hannan y Rissanen

Este algoritmo proporciona estimadores iniciales para un proceso ARMA( $p, q$ ). Si se itera, puede utilizarse para obtener estimadores de procesos ARMA utilizando sólo regresiones. El algoritmo tiene dos etapas.

*Etapa 1:* Obtenemos una estimación inicial de los residuos del modelo

## 10. Estimación y selección de modelos ARIMA

ajustando un AR largo de orden  $k > p + q$ . Sean  $\hat{\pi}_i$  los coeficientes estimados utilizando (10.11). Los residuos se calculan mediante

$$\hat{a}_t = \omega_t - \hat{c} - \sum_{i=1}^k \hat{\pi}_i \omega_{t-i}.$$

*Etapa 2:* Con los residuos obtenidos en la primera etapa, se estima la regresión

$$\omega_t = c + \phi_1 \omega_{t-1} + \dots + \phi_p \omega_{t-p} - \theta_1 \hat{a}_{t-1} - \dots - \theta_q \hat{a}_{t-q} + u_t, \quad (10.16)$$

donde las variables  $\hat{a}_{t-j}$  se construyen a partir del vector de innovaciones  $\hat{a}_t$  calculadas en la etapa anterior. La estimación de esta regresión proporciona los estimadores iniciales.

Este algoritmo puede utilizarse para obtener estimadores de modelos ARMA iterando las dos etapas anteriores, que sólo requieren regresiones. En efecto, con los parámetros estimados en la etapa 2 podemos calcular nuevos residuos y repetir la estimación de (10.16) hasta obtener convergencia. En estas condiciones obtenemos estimadores próximos a los MV. El algoritmo puede modificarse para mejorar sus propiedades asintóticas (véase Koreisha y Pukkila, 1990). Cuando se utiliza para obtener estimadores iniciales sólo se realiza el ciclo una vez, mediante las dos etapas descritas.

### 10.4.2. Estimación MV exacta

La estimación exacta requiere calcular los parámetros de las distribuciones de cada observación condicionada a las precedentes,  $\omega_{t|t-1}$  y  $\sigma^2 v_{t|t-1}$ , y viene dada por (10.4). En la sección siguiente veremos un método general para calcularla.

Es importante resaltar que la descomposición de la función de densidad conjunta como producto de condicionadas es un caso particular de una descomposición que se utiliza en muchas aplicaciones estadísticas y que se conoce como factorización de Cholesky. Observemos que llamando  $\mu = (\mu, \dots, \mu)$  al vector de medias marginales y  $M_T$  a la matriz  $T \times T$  de covarianzas entre las variables  $\omega_T$ , la expresión general de la función de densidad conjunta de  $\omega_T$  es, en logaritmos:

$$\ln f(\omega_T) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2} \ln |M_T| - \frac{1}{2\sigma^2} (\omega_T - \mu)' M_T^{-1} (\omega_T - \mu) \quad (10.17)$$

Si comparamos (10.17) y (10.4) vemos que en (10.17) se ha sustituido la forma cuadrática  $(\omega_T - \mu)' M_T^{-1} (\omega_T - \mu)$  por la expresión más simple  $e'De$ , donde  $D$  es una matriz diagonal con términos  $v_{t|t-1}^{-1}$  y

$\mathbf{e} = (e_1, \dots, e_T)'$  es el vector de errores de predicción a un paso. Además, en lugar de  $\ln|\mathbf{M}_T|$ , en (10.4) aparece el término  $\sum_{t=1}^T \ln v_{t|t-1}$ . Estos cambios pueden interpretarse como la realización de una transformación:

$$\mathbf{e} = \mathbf{L}(\boldsymbol{\omega}_T - \boldsymbol{\mu}),$$

donde la matriz  $\mathbf{L}$  es triangular inferior y tiene unos en la diagonal de manera que  $e_t$  sea sólo función de los valores actuales y previos,  $(\boldsymbol{\omega}_t, \dots, \boldsymbol{\omega}_1)$  y no de los posteriores  $(\boldsymbol{\omega}_{t+1}, \dots, \boldsymbol{\omega}_T)$ . De esta manera:

$$\mathbf{e}' \mathbf{D} \mathbf{e} = (\boldsymbol{\omega}_T - \boldsymbol{\mu})' \mathbf{L}' \mathbf{D} \mathbf{L} (\boldsymbol{\omega}_T - \boldsymbol{\mu}) = (\boldsymbol{\omega}_T - \boldsymbol{\mu})' \mathbf{M}_T^{-1} (\boldsymbol{\omega}_T - \boldsymbol{\mu})$$

y hemos descompuesto la matriz  $\mathbf{M}_T^{-1}$  como producto de dos matrices triangulares  $\mathbf{T}' \mathbf{T}$  donde  $\mathbf{T} = \mathbf{D}^{1/2} \mathbf{L}$  y donde  $\mathbf{D}^{1/2}$  es la matriz diagonal que contiene los términos  $v_{t|t-1}^{-1/2}$ . De la misma forma, el determinante de  $\mathbf{M}_T$  es el producto de los términos diagonales de  $\mathbf{D}$ , ya que el determinante de la matriz triangular con unos en la diagonal,  $\mathbf{L}$ , es la unidad.

La clave de esta descomposición es la factorización de la matriz definida positiva  $\mathbf{M}_T^{-1}$  como producto de dos matrices triangulares, y se conoce como factorización de Cholesky. En la sección siguiente veremos un algoritmo recursivo muy potente para realizar esta descomposición: el filtro de Kalman.

## 10.5. Estimación recursiva con el filtro de Kalman

El filtro de Kalman es un procedimiento recursivo computacionalmente muy rápido que tiene muchas aplicaciones en series temporales. En particular, permite evaluar rápidamente la función de verosimilitud de cualquier modelo ARMA calculando los errores de predicción a un paso y sus varianzas. El filtro fue inicialmente diseñado para resolver un problema más general: la estimación de sistemas de control en el espacio de los estados, como se explica a continuación. El filtro tiene una inmediata interpretación Bayesiana, por lo que se ha utilizado mucho en la formulación de series temporales mediante este enfoque.

### 10.5.1. Modelos en el espacio de los estados

Supongamos que observamos un sistema que puede representarse mediante una ecuación de observación:

$$\mathbf{z}_t = \mathbf{H}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t \quad (10.18)$$

donde  $\mathbf{z}_t$  es un vector de observaciones de dimensiones  $k \times 1$ ,  $\mathbf{H}_t$  es una matriz  $k \times p$  que suponemos conocida para todo  $t$ ,  $\boldsymbol{\alpha}_t$  un vector de variables de estado de dimensión  $p \times 1$  que no se observa y  $\boldsymbol{\epsilon}_t$  es un proceso de ruido blanco que suponemos tiene distribución  $N(\mathbf{0}, \mathbf{V}_t)$ . Además, la descripción del sistema incluye una ecuación que describe la evolución dinámica de las variables de estado,  $\boldsymbol{\alpha}_t$ , llamada ecuación de estado

$$\boldsymbol{\alpha}_t = \Omega_t \boldsymbol{\alpha}_{t-1} + \mathbf{u}_t \quad (10.19)$$

donde  $\Omega_t$  es una matriz conocida de dimensión  $p \times p$  y  $\mathbf{u}_t$  otro proceso de ruido blanco, independiente del anterior, que tiene distribución  $N_p(\mathbf{0}, \mathbf{R}_t)$ . Suponemos que las matrices del sistema  $\mathbf{H}_t$  y  $\Omega_t$  se conocen para todos los instantes.

Por ejemplo, supongamos que observamos en cada instante  $t$  la posición en el espacio de un satélite mediante las variables,  $\mathbf{z}_t$ , que será un vector de dimensión tres (las coordenadas de la posición del satélite en el espacio). Suponemos que las medidas de la posición del satélite dependen de un conjunto de variables de estado,  $\boldsymbol{\alpha}_t$ , que no son observables, pero que están relacionadas con nuestras observaciones a través de la matriz  $\mathbf{H}_t$ , que es conocida. Por ejemplo, si la posición del satélite depende de su velocidad y su aceleración, el vector de estado tiene dimensión dos. Suponemos que existe en cada instante un error de observación, representado por el vector  $\boldsymbol{\epsilon}_t$ . Por otro lado, la ecuación de estado indica que el vector de variables de estado en el instante  $t$  depende de la situación de las variables de estado en el instante  $t - 1$  más un error de medida.

La representación de un sistema mediante las ecuaciones (10.18) y (10.19) no es única. Siempre es posible aumentar la dimensión del vector de estado poniendo ceros en las matrices que le multiplican. Decimos que el vector de estado tiene dimensión mínima cuando no es posible representar el sistema con menos de  $p$  variables de estado. Una vez fijada la dimensión, el vector de estado tampoco es único, ya que dado un vector de estado  $\boldsymbol{\alpha}_t$  el sistema puede igualmente representarse con el vector de estado  $\boldsymbol{\alpha}_t^* = \mathbf{A}\boldsymbol{\alpha}_t$ , donde  $\mathbf{A}$  es cualquier matriz cuadrada no singular. En efecto, escribiendo la ecuación de observación como:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{A}^{-1} \mathbf{A} \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t = \mathbf{H}_t^* \boldsymbol{\alpha}_t^* + \boldsymbol{\epsilon}_t$$

y la de evolución del estado como:

$$\boldsymbol{\alpha}_t^* = \Omega_t^* \boldsymbol{\alpha}_{t-1}^* + \mathbf{u}_t$$

donde ahora  $\Omega_t^* = \mathbf{A} \Omega_t \mathbf{A}^{-1}$ . En adelante supondremos que el sistema tiene dimensión mínima.

Cualquier modelo ARMA( $p, q$ ) puede escribirse en esta formulación

como sigue. Definamos  $m = \max(p, q+1)$  y llamemos  $\alpha_t = (\alpha_{1,t}, \alpha_{2,t}, \dots, \alpha_{m,t})'$  al vector de variables de estado, que seguirá la ecuación:

$$\begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \vdots \\ \alpha_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & \dots & 0 \\ \phi_2 & 0 & \dots & 0 \\ \vdots & 0 & \dots & 1 \\ \phi_m & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \vdots \\ \alpha_{m,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_m \end{bmatrix} a_t. \quad (10.20)$$

Observemos que en esta ecuación la matriz de estado,  $\Omega_t$ , tiene la forma:

$$\Omega_t = \begin{bmatrix} \phi_{m-1} & \mathbf{I} \\ \phi_m & \mathbf{0}' \end{bmatrix}$$

donde  $\phi_{m-1}$  es un vector columna de dimensión  $m-1$ ,  $\mathbf{I}$  es la matriz identidad y  $\mathbf{0}'$  es un vector de ceros. Por otro lado, el vector de innovaciones en esta ecuación es

$$\mathbf{u}_t = \theta a_t$$

donde  $\theta' = (1, -\theta_1, \dots, -\theta_m)$ . La matriz de varianzas y covarianzas de  $\mathbf{u}$  es

$$\mathbf{R}_t = \theta \theta' \sigma^2.$$

Vamos a comprobar que sustituyendo sucesivamente en las variables de estado se obtiene la representación del proceso ARMA. La primera ecuación es

$$\alpha_{1,t} = \phi_1 \alpha_{1,t-1} + \alpha_{2,t-1} + a_t, \quad (10.21)$$

y la segunda

$$\alpha_{2,t} = \phi_2 \alpha_{1,t-1} + \alpha_{3,t-1} - \theta_1 a_t. \quad (10.22)$$

Sustituyendo  $\alpha_{2,t-1}$  en (10.21) de acuerdo con la expresión (10.22), tenemos que:

$$\alpha_{1,t} = \phi_1 \alpha_{1,t-1} + \phi_2 \alpha_{1,t-2} + \alpha_{3,t-2} + a_t - \theta_1 a_{t-1}. \quad (10.23)$$

La tercera ecuación es

$$\alpha_{3,t} = \phi_3 \alpha_{1,t-1} + \alpha_{4,t-1} - \theta_2 a_t$$

y sustituyendo ahora en (10.23)  $\alpha_{3,t-2}$  por su expresión anterior, vamos recuperando el proceso ARMA en la variable  $\alpha_{1,t}$ . La ecuación de observación sirve simplemente para hacer la variable observada,  $z_t$ , que es escalar, igual a la primera componente del vector de estado:

$$z_t = (1, 0, \dots, 0) \alpha_t. \quad (10.24)$$

Las ecuaciones (10.20) y (10.24) constituyen una forma de representar el

## 10. Estimación y selección de modelos ARIMA

---

modelo ARMA en el espacio de los estados. Observemos que son un caso particular de las (10.18) y (10.19). En la ecuación de observación (10.18) el vector de datos es ahora un escalar, el valor de la serie observada en cada instante; el vector de estado es un vector de dimensiones  $m = \max(p, q + 1)$ ; la matriz  $\mathbf{H}_t$  es siempre el vector  $(1, 0, \dots, 0)$  y no existe error de medida o ruido en la matriz de observación. En la ecuación de estado la matriz  $\Omega_t$  es invariante en el tiempo y la matriz de covarianzas de  $\mathbf{u}_t$  es singular de rango uno.

La representación de un modelo ARMA en el espacio de los estados no es única, véase Box, Jenkins y Reinsel (1994), Brockwell y Davies (1996) y Peña, Tiao y Tsay (2002) para representaciones alternativas.

Una forma de escribir modelos de series temporales en el espacio de los estados es mediante los modelos estructurales, estudiados por Harrison y Stevens (1976), West and Harrison (1989) y Harvey (1989), entre otros. En estos modelos la serie se expresa como suma de componentes asociados a la tendencia y la estacionalidad más un ruido blanco, como:

$$z_t = \mu_t + S_t + a_t.$$

Un formulación habitual de la tendencia es:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + v_t \quad (10.25)$$

$$\beta_t = \beta_{t-1} + u_t \quad (10.26)$$

donde  $\beta_t$  representa el ritmo de crecimiento de la tendencia y  $u_t$  y  $v_t$  son procesos de ruido blanco independientes con varianzas  $\sigma_u^2$  y  $\sigma_v^2$ . Si  $\sigma_u^2$ , la varianza de  $u_t$  en (10.26), es cero, la tendencia es constante o determinista. Si además  $\beta_t$  es cero, la serie no tiene tendencia. Si la varianza  $\sigma_u^2$  es grande, la tendencia puede cambiar mucho de un periodo al siguiente. La varianza  $\sigma_v^2$  determina como se mueve el nivel de la serie de un periodo al siguiente.

La estacionalidad se modela con coeficientes estacionales constantes que pueden ser arbitrarios o seguir una estructura trigonométrica, como vimos en el capítulo 2. Suponiendo una estacionalidad constante, los coeficientes  $S_t$  deben verificar:

$$\sum_{j=1}^s S_{t-j} = \epsilon_t, \quad (10.27)$$

y las variables  $v_t$ ,  $u_t$  y  $\epsilon_t$  son ruidos blancos independientes. Este modelo tiene un vector de estado de dimensión 13, formado por las variables de

estado  $\mu_t$ ,  $\beta_t$  y los 11 coeficientes estacionales. Definiendo:

$$\alpha_t = (\mu_t, \beta_t, S_t, S_{t-1}, \dots, S_{t-10})'$$

el modelo se escribe en el espacio de los estados con una ecuación de observación

$$z_t = (1, 0, 1, 0, \dots, 0) \alpha_t + a_t$$

y una ecuación de estado:

$$\begin{bmatrix} \mu_t \\ \beta_t \\ S_t \\ \vdots \\ S_{t-9} \\ S_{t-10} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ S_{t-1} \\ \vdots \\ S_{t-10} \\ S_{t-11} \end{bmatrix} + \begin{bmatrix} v_t \\ u_t \\ \epsilon_t \\ \dots \\ 0 \\ 0 \end{bmatrix}$$

Las tres primeras ecuaciones del vector de estado son las (10.25), (10.26) y (10.27). Las restantes son identidades, del tipo  $S_{t-j} = S_{t-j}$ .

### 10.5.2. El filtro de Kalman

El filtro de Kalman es un algoritmo recursivo de predicción. Vamos a presentarlo en su formulación general y después indicaremos su particularización para calcular la función de verosimilitud de un proceso ARMA.

El algoritmo funciona en tres pasos. En el primero, estimamos el estado futuro a partir de la información del estado actual. En el segundo, predecimos nuevas observaciones. En el tercero, que se realiza cuando llega una nueva observación al sistema, se revisa la estimación del estado en ese instante a la vista de la nueva información. Vamos a revisar estas tres etapas.

La primera etapa es la estimación del estado futuro a partir de una estimación del estado actual. Supongamos que disponemos de los datos  $Z_{t-1} = \{z_1, \dots, z_{t-1}\}$  y de un estimador del vector de estado,  $\hat{\alpha}_{t-1}$ , y deseamos estimar  $\hat{\alpha}_{t|t-1}$ , el próximo valor del estado utilizando los datos observados,  $Z_{t-1}$ . Esta estimación se calcula tomando esperanzas en (10.19) condicionadas a  $Z_{t-1}$  y obtendremos:

$$\hat{\alpha}_{t|t-1} = \Omega_t \hat{\alpha}_{t-1} \quad (10.28)$$

donde hemos utilizado la notación  $\hat{\alpha}_{t-1|t-1} = \hat{\alpha}_{t-1}$ . Llamaremos  $S_{t|t-1}$  a la matriz de covarianzas de esta estimación:

$$S_{t|t-1} = E [(\alpha_t - \hat{\alpha}_{t|t-1})(\alpha_t - \hat{\alpha}_{t|t-1})' | Z_{t-1}] .$$

## 10. Estimación y selección de modelos ARIMA

---

Para calcularla, utilizaremos que restando de (10.19) la ecuación (10.28):

$$\alpha_t - \hat{\alpha}_{t|t-1} = \Omega_t (\alpha_{t-1} - \hat{\alpha}_{t-1}) + u_t,$$

y sustituyendo esta expresión en la definición de  $S_{t|t-1}$  y llamando  $S_{t-1} = S_{t-1/t-1}$ , se obtiene que

$$S_{t|t-1} = \Omega_t S_{t-1} \Omega_t' + R_t. \quad (10.29)$$

Esta ecuación tiene un clara interpretación intuitiva: la incertidumbre al estimar el nuevo estado con información hasta  $t-1$ , es la suma de la incertidumbre que teníamos respecto al estado anterior, medida por  $S_{t-1}$ , y la incertidumbre del ruido en la ecuación de estado,  $R_t$ . La matriz  $\Omega_t$  aparece como en (10.19) para relacionar los componentes del estado en el instante  $t-1$  y en el instante  $t$ . Si esta matriz fuese la identidad, el estado evoluciona como un paseo aleatorio,  $\Omega_t = I$ , y la incertidumbre de estimación del estado aumentaría continuamente a través de la suma de la matriz  $R_t$ . Como en general esta matriz no es la identidad, el aumento de incertidumbre depende de su estructura. Por ejemplo, supongamos un AR(1). Entonces, el vector de estado es escalar, y  $\Omega_t = \phi < 1$ . La varianza de la estimación sigue el proceso

$$s_{t|t-1} = \phi^2 s_{t-1} + \sigma^2$$

y sólo una parte de la incertidumbre en  $t-1$  se traslada al instante  $t$ .

El segundo paso del filtro es la predicción de la nueva observación  $z_t$  dada la información hasta  $t-1$ . Esta predicción se calcula de nuevo con la esperanza condicional dada  $Z_{t-1}$  y obtenemos:

$$\hat{z}_{t|t-1} = E(z_t | Z_{t-1}) = H_t \hat{\alpha}_{t|t-1}. \quad (10.30)$$

Esta predicción tendrá una incertidumbre que mediremos por la matriz de varianzas y covarianzas de los errores de predicción:

$$e_t = z_t - \hat{z}_{t|t-1},$$

definida por:

$$P_{t|t-1} = E[e_t e_t'].$$

Para calcular esta matriz, restando la predicción (10.30) de la ecuación de observación (10.18), tenemos que:

$$e_t = z_t - \hat{z}_{t|t-1} = H_t (\alpha_t - \hat{\alpha}_{t|t-1}) + \epsilon_t \quad (10.31)$$

y sustituyendo esta expresión en la definición de  $\mathbf{P}_{t|t-1}$ , se obtiene que

$$\mathbf{P}_{t|t-1} = \mathbf{H}_t \mathbf{S}_{t|t-1} \mathbf{H}'_t + \mathbf{V}_t. \quad (10.32)$$

Esta ecuación indica que la incertidumbre de la predicción acumula la incertidumbre sobre el estado y la del error de medida de las observaciones. El error de predicción que viene de la estimación del estado se modula dependiendo de la matriz  $\mathbf{H}_t$ . Si esta matriz es la identidad, lo que supone que las observaciones  $\mathbf{z}_t$  son mediciones de las variables de estado más un error aleatorio, al error de las variables de estado se añade el error de medición de las observaciones.

El tercer y último paso del filtro es revisar la estimación del estado a la vista de la nueva información. Supongamos que se ha observado  $\mathbf{z}_t$  con lo que la información disponible pasa a ser  $\mathbf{Z}_t = (\mathbf{Z}_{t-1}, \mathbf{z}_t)$ . La nueva estimación del estado,  $\hat{\alpha}_t = \hat{\alpha}_{t|t} = E(\alpha_t | \mathbf{Z}_t)$ , se calcula por regresión (o con las propiedades de la normal multivariante) con:

$$E(\alpha_t | \mathbf{Z}_{t-1}, \mathbf{z}_t) = E(\alpha_t | \mathbf{Z}_{t-1}) + \\ + cov(\alpha_t, \mathbf{z}_t | \mathbf{Z}_{t-1}) var(\mathbf{z}_t | \mathbf{Z}_{t-1})^{-1} (\mathbf{z}_t - E(\mathbf{z}_t | \mathbf{Z}_{t-1})). \quad (10.33)$$

En esta ecuación las esperanzas  $E(\alpha_t | \mathbf{Z}_{t-1}) = \hat{\alpha}_{t|t-1}$  y  $E(\mathbf{z}_t | \mathbf{Z}_{t-1}) = \hat{\mathbf{z}}_{t|t-1}$  son conocidas, así como la matriz  $var(\mathbf{z}_t | \mathbf{Z}_{t-1}) = \mathbf{P}_{t|t-1}$ . Lo único que queda por calcular es la covarianza entre el estado y la nueva observación, que viene dada por

$$cov(\alpha_t, \mathbf{z}_t | \mathbf{Z}_{t-1}) = E [(\alpha_t - \hat{\alpha}_{t|t-1})(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})'] = \\ = E [(\alpha_t - \hat{\alpha}_{t|t-1}) \mathbf{e}'_t]$$

y sustituyendo (10.31),

$$cov(\alpha_t, \mathbf{z}_t | \mathbf{Z}_{t-1}) = E [(\alpha_t - \hat{\alpha}_{t|t-1})((\alpha_t - \hat{\alpha}_{t|t-1})' \mathbf{H}'_t + \epsilon'_t)] = \\ = \mathbf{S}_{t|t-1} \mathbf{H}'_t, \quad (10.34)$$

ya que el error de observación  $\epsilon'_t$  es ruido blanco e independiente de  $\alpha_t - \hat{\alpha}_{t|t-1}$ . Sustituyendo esta covarianza en (10.33), podemos escribir:

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + \mathbf{K}_t (\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}) \quad (10.35)$$

donde  $\mathbf{K}_t$  es la matriz de coeficientes de regresión que se llama la ganancia del filtro, y viene dada por:

$$\mathbf{K}_t = \mathbf{S}_{t|t-1} \mathbf{H}'_t \mathbf{P}_{t|t-1}^{-1}.$$

La ecuación (10.35) indica que la revisión que hacemos de la estimación

previa del estado depende del error de predicción,  $e_t = z_t - \hat{z}_{t|t-1}$ . Si este error es cero, no modificamos la estimación; en otro caso, hacemos una modificación de la estimación del estado que depende del cociente entre el error en la estimación del estado,  $S_{t|t-1}$ , y el error de predicción  $P_{t|t-1}^{-1}$ . La matriz  $H'_t$  permite comparar estas matrices. Una forma equivalente de escribir la ecuación (10.35) es:

$$\hat{\alpha}_t = (I - K_t H_t) \hat{\alpha}_{t|t-1} + K_t z_t,$$

que indica que la estimación del estado es una combinación lineal de las dos fuentes de información de que disponemos. Por un lado, la estimación previa,  $\hat{\alpha}_{t|t-1}$ , y por otra, la observación  $z_t$  que también aporta información sobre el estado. Puede demostrarse que las ponderaciones de las dos fuentes de información son iguales a su precisión relativa. La matriz de covarianzas de esta estimación será:

$$S_t = E [(\alpha_t - \hat{\alpha}_t)(\alpha_t - \hat{\alpha}_t)' | Z_t],$$

y sustituyendo  $\hat{\alpha}_t$  por su expresión en la ecuación (10.35), tenemos que:

$$S_t = E [(\alpha_t - \hat{\alpha}_{t|t-1} - K_t e_t)(\alpha_t - \hat{\alpha}_{t|t-1} - K_t e_t)' | Z_t]$$

y utilizando (10.34) y operando se obtiene finalmente que:

$$S_t = S_{t|t-1} - S_{t|t-1} H'_t P_{t|t-1}^{-1} H_t S_{t|t-1}. \quad (10.36)$$

Las ecuaciones (10.28), (10.29), (10.30), (10.32), (10.35) y (10.36) constituyen el filtro de Kalman. Bajo la hipótesis de normalidad el filtro proporciona estimaciones y predicciones óptimas.

La aplicación del filtro para obtener la función de verosimilitud del modelo ARMA requiere escribir el modelo en el espacio de los estados, como hemos visto, y calcular los errores de predicción  $e_t = z_t - \hat{z}_{t|t-1}$ , que en este caso son escalares y sus varianzas,  $P_{t|t-1}$ . En este caso la matriz  $V_t$  es cero. Para comenzar el filtro hace falta indicar un valor inicial para las variables de estado,  $\alpha_0$ , y para su matriz de covarianzas  $S_0$ . Estos valores iniciales no son cruciales, porque el filtro depende poco de las condiciones iniciales. El lector interesado en la implantación del algoritmo puede acudir a Harvey (1989), y Gómez y Maravall (1994).

## 10.6. Propiedades de los estimadores

Puede demostrarse que las propiedades asintóticas del método de máxima verosimilitud son válidas, en condiciones de regularidad generales, para

los estimadores MV de modelos ARMA. Estas condiciones exigen que el proceso sea estacionario y que el modelo ARMA que estimamos no contiene factores comunes en su parte AR y MA. Ya vimos en el capítulo anterior que si un proceso AR(1) tiene el parámetro igual a la unidad, el valor verdadero está en la frontera del intervalo paramétrico [0,1], y no se cumplen las condiciones habituales de regularidad necesarias para obtener las propiedades asintóticas del estimador MV.

Para procesos estacionarios en muestras grandes los estimadores MV tendrán distribución aproximadamente normal y serán centrados y eficientes. En particular, la matriz de segundas derivadas del soporte en su máximo proporciona directamente las varianzas y covarianzas de los estimadores:

$$\text{Var}(\hat{\beta}_{MV}) = - \left[ \frac{\partial^2 L(\hat{\beta}_{MV})}{\partial \beta \partial \beta'} \right]^{-1}$$

Este resultado se basa en que la función de verosimilitud es aproximadamente cuadrática en el máximo. Para comprobarlo, conviene siempre que sea posible estudiar la forma de la función soporte alrededor del estimador MV, calculando numéricamente y dibujando sus curvas de nivel.

La condición de que no existan factores comunes en la parte AR y MA es importante. Por ejemplo, si  $\omega_t$  es ruido blanco y estimamos el modelo

$$(1 - \phi B)\omega_t = (1 - \theta B)a_t$$

todos los valores de los parámetros con la condición  $\phi = \theta$  son compatibles con los datos y puede demostrarse que la varianza de los estimadores es infinita. En general, si el modelo está sobreparametrizado y contiene simultáneamente factores redundantes AR y MA tendremos una situación de fuerte multicolinealidad que puede dar lugar, primero, a múltiples máximos en la función de verosimilitud y, segundo, a que la aproximación cuadrática en el máximo sea inadecuada.

---

### Ejemplo 10.1

Vamos a estimar los modelos identificados para la serie de matriculación de vehículos en España que se analizó en el capítulo anterior. Utilizaremos tres programas. El programa TSW, que utiliza el filtro de Kalman y máxima verosimilitud exacta, el SCA que también utiliza MV exacta y Minitab que lo hace condicional. Los resultados se presentan en la tabla 10.1. Se observa en la tabla que los programas TSW y SCA que utilizan estimación MV exacta

## 10. Estimación y selección de modelos ARIMA

---

**Tabla 10.1 Estimación de modelos para la serie de matriculación de vehículos con distintos programas**

Progr.	Modelo	$\hat{\sigma}_a$
TSW	$\nabla \nabla_{12} \ln M_t = (1 - 0.61B)(1 - 0.78B^{12}) a_t$ (0.04) (0.03)	0.123
TSW	$(1 - 0.21B^{12}) \nabla \nabla_{12} \ln M_t = (1 - 0.61B)(1 - 0.89B^{12}) a_t$ (0.06) (0.06) (0.05)	0.121
SCA	$\nabla \nabla_{12} \ln M_t = (1 - 0.61B)(1 - 0.78B^{12}) a_t$ (0.04) (0.03)	0.122
SCA	$(1 - 0.14B^{12}) \nabla \nabla_{12} \ln M_t = (1 - 0.61B)(1 - 0.85B^{12}) a_t$ (0.05) (0.04) (0.05)	0.117
SCA	$(1 + 0.53 B^{12} + .32 B^{24}) \nabla \nabla_{12} \ln M_t = (1 - 0.62B) a_t$ (.04) (0.04) (0.04)	0.122
Minitab	$\nabla \nabla_{12} \ln M_t = (1 - 0.62B)(1 - 0.84B^{12}) a_t$ (0.03) (0.03)	0.119
Minitab	$(1 - 0.25B) \nabla \nabla_{12} \ln M_t = (1 - 0.62B)(1 - 0.95B^{12}) a_t$ (0.06) (0.06) (0.03)	0.116

conducen a valores muy similares, mientras que con Minitab la estimación de los términos estacionales es menos precisa. El programa TSW sólo permite AR estacionales de orden uno, por lo que no hemos podido estimar el AR(2) estacional con dicho programa. Se observa que el mejor modelo desde el punto de vista de la varianza residual es el que tiene un ARMA(1,1) para la parte estacional.

Una conclusión de este ejercicio es que la mayor diferencia entre las estimaciones exacta y condicional aparece cuando el modelo tiene términos de media móvil cerca del valor unitario, de manera que el proceso está cerca de la no invertibilidad.

---

### 10.7. Criterios de selección de modelos

Supongamos que hemos estimado un conjunto de modelos,  $M_1, \dots, M_m$ , y deseamos seleccionar el que mejor explica la serie observada. Seleccionar el modelo por su ajuste a una muestra dada no resulta adecuado, ya que el modelo con más parámetros siempre conducirá a una mayor verosimilitud y a una menor suma de cuadrados de los errores dentro de la muestra. Por ejemplo, si comparamos un AR( $p$ ) con un AR( $p + 1$ ), el ajuste del AR( $p + 1$ ) no puede ser peor que el del AR( $p$ ), y siempre escogeremos el modelo más complejo. Por tanto, para seleccionar entre modelos debemos acudir a otros principios.

El problema puede verse como un problema de discriminación: te-

nemos distintos modelos  $M_i$  y una serie estacionaria,  $\omega_T = (\omega_1, \dots, \omega_T)$ , y queremos seleccionar el modelo más compatible con la serie observada, y puede abordarse desde el punto de vista clásico o Bayesiano. Comenzando con el enfoque clásico, no es útil comparar la verosimilitud de distintos modelos, porque siempre el modelo con más parámetros tendrá mayor verosimilitud. Podemos, sin embargo, calcular el valor esperado de la verosimilitud para cada uno de los modelos, es decir, el valor que esperamos obtener para la verosimilitud sobre muchas realizaciones del proceso si en cada una estimamos los parámetros por máxima verosimilitud, y seleccionar aquel modelo que produzca un valor esperado más alto de esta verosimilitud esperada. Este es el enfoque que conduce al criterio de Akaike, que exponemos a continuación.

Si disponemos de probabilidades a priori para cada modelo,  $P(M_i)$ , podríamos utilizar el enfoque Bayesiano y seleccionar el modelo que tenga probabilidad máxima dados los datos. Es decir, calculamos para  $i = 1, \dots, m$

$$P(M_i|\omega_T) = \frac{P(\omega_T|M_i)P(M_i)}{\sum_{j=1}^m P(\omega_T|M_j)P(M_j)} \quad (10.37)$$

y seleccionamos el modelo más probable a la vista de los datos. Observemos que este planteamiento no requiere que la serie sea estacionaria, por lo que puede aplicarse para comparar modelos con distinto número de diferencias y utilizarse para decidir como alternativa a los contrastes de raíces unitarias. Si suponemos que las probabilidades a priori de todos los modelos son las mismas, este enfoque conduce al criterio BIC, que presentamos a continuación.

Un enfoque alternativo para seleccionar el mejor modelo es *validación cruzada*. En este enfoque estimamos el modelo en una parte de la muestra y utilizamos la otra parte para calcular el error de predicción fuera de la muestra. Un problema importante es cómo dividir la muestra en estas dos partes. Un procedimiento sugestivo es utilizar toda la muestra para estimar y prever de la forma siguiente: se elimina la nueva información que contiene el dato  $\omega_i$  sustituyéndolo por su predicción  $\hat{\omega}_i$ , se estiman los parámetros en esta serie modificada y se prevé el dato  $\omega_i$  con estos parámetros. Esto proporciona un error de predicción fuera de la muestra y puede aplicarse a todos los puntos de la muestra (véase Peña y Sánchez, 2004). Puede demostrarse que, en muestras grandes, existe una equivalencia entre los métodos de validación cruzada y los criterios de selección de modelos.

### 10.7.1. El criterio AIC de Akaike

La función de verosimilitud de un modelo ARIMA viene dada por (10.4). Multiplicando por  $-2$  y tomando esperanzas en esta expresión tenemos que

$$E(-2L(\beta)) = T \ln \sigma^2 + \sum_{t=1}^T \ln v_{t|t-1} + E \left[ \sum_{t=1}^T \frac{e_t^2}{\sigma^2 v_{t|t-1}} \right].$$

Se demuestra en el apéndice 10.2 que si: (1) suponemos que los parámetros se estiman con los datos; (2) se calcula la esperanza de esta verosimilitud respecto a futuras observaciones, se obtiene que:

$$AIC = E(-2L(\beta)) = T \ln \hat{\sigma}_{MV}^2 + 2k, \quad (10.38)$$

donde  $T$  es el tamaño muestral utilizado para estimar el modelo,  $\hat{\sigma}_{MV}^2$  el estimador MV de la varianza de las innovaciones y  $k$  el número de parámetros estimados para calcular las predicciones a un paso. Por tanto, seleccionar el modelo con verosimilitud esperada máxima equivale a escoger el que minimiza la verosimilitud con signo negativo dado por (10.38). Este criterio se conoce como criterio AIC, y es debido a Akaike.

El problema con el AIC es que tiende a sobreestimar el número de parámetros en el modelo y este efecto puede ser muy grande en pequeñas muestras. Una alternativa que corrige esta sobreestimación es el criterio AIC corregido, AICC, dado por:

$$AICC = T \ln \hat{\sigma}_{MV}^2 + T \frac{(1 + k/T)}{1 - (k + 2)/T}.$$

Si utilizamos el criterio AIC para comparar modelos ajustados a una serie es importante que  $T$ , el número efectivo de observaciones utilizado para estimar el modelo, sea el mismo para todos ellos. Este número es el de datos estacionarios, que es igual a los datos originales menos  $d - sD$ , siendo  $d$  el número de diferencias regulares,  $s$  el periodo estacional y  $D$  el número de diferencias estacionales. Si estimamos el modelo por máxima verosimilitud exacta podemos calcular residuos en todos los puntos y por tanto el número de datos efectivo es  $T - d - sD$ . Si consideramos modelos con distinto número de diferencias y llamamos  $d_{\max}$  y  $D_{\max}$  los grados de diferenciación más altos de los modelos que comparamos, entonces el número de datos efectivos es:

$$T = T_0 - d_{\max} - sD_{\max}. \quad (10.39)$$

Sin embargo, si utilizásemos un programa que realiza estimación condicionada para calcular los residuos, no tenemos estos datos sino un número

menor, ya que, por ejemplo, para un AR(1), como condicionamos al primer dato el residuo para esa observación, no puede calcularse. En general, si hacemos estimación condicional y llamamos  $r = \max(p, q)$ , y  $R = \max(P, Q)$  el número efectivo de observaciones con las que podemos calcular los residuos en un modelo dado es  $T - d - sD - r - R$ , y puede variar mucho de unos modelos a otros. Esta es una razón adicional de utilizar máxima verosimilitud exacta, de manera que siempre consideramos los residuos en las mismas  $T$  observaciones en las comparaciones entre los modelos. Si no hacemos esto, es posible que parte de las diferencias observadas entre los modelos se deban a los distintos residuos calculados.

### 10.7.2. El criterio BIC

Un criterio alternativo, propuesto por Schwarz (1978) desde el enfoque Bayesiano, es maximizar la probabilidad a posteriori del modelo dada la serie,  $P(M_i|\omega)$ , supuesto que las probabilidades a priori son las mismas para todos ellos. Como, según (10.37),  $P(M_i|\omega)$  es proporcional a  $P(\omega|M_i)P(M_i)$ , si las probabilidades a priori son las mismas, la probabilidad a posteriori es proporcional a  $P(\omega|M_i)$ . Seleccionar el modelo que maximice esta probabilidad es equivalente a seleccionar el modelo que minimiza  $-2\ln P(\omega|M_i)$ . Puede demostrarse, véase el apéndice 10.3, que sustituyendo los parámetros por sus estimaciones máximo verosímiles el modelo que minimiza asintóticamente esta cantidad es el que minimiza el criterio:

$$BIC = T \ln \hat{\sigma}_{MV}^2 + k \ln T, \quad (10.40)$$

donde, como en el caso anterior,  $T$  es el tamaño muestral,  $\hat{\sigma}_{MV}^2$  el estimador MV de la varianza y  $k$  el número de parámetros.

Si comparamos esta expresión con la (10.38) vemos que el BIC penaliza por introducir nuevos parámetros más que el AIC, con lo que tiende a elegir modelos más parsimoniosos. La diferencia entre ellos puede ser grande si  $T$  es grande.

Para comparar el BIC de varios modelos ajustados a una serie el número de observaciones en todos debe ser el mismo, para tener comparaciones homogéneas y evitar que las diferencias entre modelos se deban a diferencias entre residuos incluidos en unos modelos y no en otros. El tamaño  $T$  debe calcularse con (10.39).

### 10.7.3. Comparación entre criterios

Los criterios estudiados pueden expresarse, dividiendo por  $T$  las expresiones anteriores, como:

$$\text{minimizar}(\ln \hat{\sigma}_{MV}^2 + kc(k, T)),$$

donde  $c(k, T)$  es un término de corrección que tiende a cero al aumentar el tamaño muestral. El criterio AIC toma  $c(k, T) = 2/T$ , y el BIC  $c(k, T) = \ln(T)/T$ . Puede demostrarse que el criterio BIC es consistente, en el sentido de que cuando los datos han sido generados por un modelo ARIMA el BIC selecciona el orden adecuado del modelo con probabilidad uno. Por el contrario, el criterio AIC es eficiente, en el sentido de que si los datos han sido generados por un modelo que puede ser de orden infinito, y consideramos una secuencia de estimadores con orden que aumenta con el tamaño muestral, el predictor seleccionado es el de menor error de predicción esperado.

La diferencia entre estas dos propiedades es que el BIC supone que los datos se han generado por un modelo con un número de parámetros finito, mientras que el AIC permite que el modelo correcto pueda no ser de orden finito. Además, el BIC es una aproximación a un criterio que puede aplicarse exactamente, calcular las probabilidades a posteriori y seleccionar la más alta, mientras que AIC supone una situación más ficticia asintótica donde vamos incrementando el orden del modelo con el tamaño muestral, que puede ser menos relevante ante una muestra concreta. Existe mucha evidencia de que el criterio AIC tiende a seleccionar un orden demasiado alto, y recomendamos el criterio BIC que funciona mejor en la práctica. Por otro lado, conviene notar que no siempre el modelo correcto es el que proporciona las mejores predicciones. Por ejemplo, Sánchez y Peña (2001) han demostrado que cuando los datos se generan con un  $\text{AR}(p+1)$ , con una raíz próxima a la unidad, las predicciones obtenidas diferenciando y estimando un  $\text{AR}(p)$  tienen menor error cuadrático medio que las del  $\text{AR}(p+1)$ . La razón es que cuando el coeficiente está próximo a uno, al fijarlo en uno por la diferencia el error es menor que al estimarlo en la muestra.

### Ejemplo 10.2

Vamos a aplicar los criterios de selección de modelos para escoger el mejor de los tres estimados mediante SCA para la serie de matriculación en el ejemplo 10.1. La tabla 10.2 indica el modelo, la varianza residual, el número de parámetros y el valor del criterio de selección correspondiente. La serie tiene 479 datos menos 13 debido a las diferencias, 466, ya que hemos utilizado máxima verosimilitud exacta. En las comparaciones utilizaremos  $T = 466$ .

Se observa que aunque el segundo modelo,  $(0,1,1) \times (1,1,1)_{12}$ , tiene un menor valor de la varianza residual, el mejor modelo tanto con el criterio BIC como el AIC es el primero, que obtiene el menor valor con ambos criterios. Éste será, por tanto, el modelo elegido.

**Tabla 10.2** Valores de los criterios de selección con la serie de matriculación

Modelo	$\hat{\sigma}^2$	T	k	BIC	AIC
ARIMA(0,1,1)×(0,1,1) <sub>12</sub>	0.122 <sup>2</sup>	466	2	8.08	-0.20
ARIMA(0,1,1)×(1,1,1) <sub>12</sub>	0.117 <sup>2</sup>	466	3	14.14	1.70
ARIMA(0,1,1)×(2,1,0) <sub>12</sub>	0.122 <sup>2</sup>	466	3	14.22	1.79

**Tabla 10.3** Modelos estimados con el programa TSW para el logaritmo de la serie de accidentes

Modelo	$\hat{\sigma}$	BIC
$\nabla\nabla_{12} \ln A_t = (1 - 0.54B)(1 - 0.76B^{12}) a_t$ (0.05) (0.05)	0.06596	-5.3789
$(1 + 0.61B + 0.33B^2) \nabla\nabla_{12} \ln A_t = (1 - 0.75B^{12}) a_t$ (0.06) (0.06) (0.05)	0.06469	-5.3981
$(1 + 0.58B + 0.32B^2) (1 + 0.26B^{12}) \nabla\nabla_{12} \ln A_t =$ (0.06) (0.06) (0.09) $= (1 - 0.86B^{12}) a_t$ (0.07)	0.06345	-5.3981

### Ejemplo 10.3

Vamos a comparar algunos modelos estimados para la serie de accidentes, que se identificó en el capítulo anterior (véase el ejemplo 9.2). La tabla 10.3 indica algunos modelos estimados por MVE (máxima verosimilitud exacta) por el programa TSW. Este programa no permite AR o MA estacionales mayores que uno, por lo que hemos comenzado con los modelos más simples para la estacionalidad. Se indican los parámetros estimados y sus desviaciones típicas y la desviación típica residual y el valor del BIC para el modelo. Sólo se indican modelos donde los parámetros estimados son significativos, y debajo de cada parámetro se proporciona la desviación típica estimada.

Los tres modelos mejores son los dos últimos, de acuerdo con el BIC. El primero es una aproximación del segundo, porque  $(1 - 0.54B)^{-1} = (1 + 0.54B + 0.29B^2 + 0.16B^3 + \dots)$  y si truncamos la serie y nos quedamos con los dos primeros valores tenemos un AR(2), similar al estimado por el segundo modelo. Este segundo modelo tiene menos varianza residual que el primero, aunque también tiene un parámetro más, pero de acuerdo con el BIC la reducción de varianza es apreciable, ya que el BIC disminuye. El tercer modelo parece recoger mejor la estacionalidad, ya que el término AR(1) es significativo, y tiene el mínimo valor del BIC de los tres, aunque las diferencias entre el segundo y el tercer modelo son pequeñas.

La tabla 10.4 presenta algunas estimaciones de estos y otros modelos rea-

## 10. Estimación y selección de modelos ARIMA

---

**Tabla 10.4 Algunos modelos estimados con Minitab para el logaritmo de la serie de accidentes**

Modelo	$\hat{\sigma}$
$\nabla\nabla_{12} \ln A_t = (1 - 0.53B)(1 - 0.84B^{12}) a_t$ (0.06) (0.04)	0.0647
$(1 + 0.60B + 0.32B^2) \nabla\nabla_{12} \ln A_t = (1 - 0.80B^{12}) a_t$ (0.06) (0.06) (0.05)	0.0636
$(1 + 0.58B + 0.29B^2) (1 - 0.28B^{24} + 0.16B^{36}) \nabla\nabla_{12} \ln A_t =$ (0.06) (0.06) (0.08) (0.07) $= (1 - 0.89B^{12}) a_t$ (0.07)	0.0613
$(1 + 0.59B + 0.30B^2) \nabla\nabla_{12} \ln A_t = (1 - .61B^{12} - 0.31B^{24}) a_t$ (0.06) (0.06) (0.07) (0.06)	0.0612

lizadas con Minitab. El primer modelo es el mismo estimado antes, y aunque los parámetros estimados son muy similares, la estimación de la desviación típica de los residuos es distinta. Esto es consecuencia de que como este programa utiliza la estimación condicional, no calcula los residuos para las observaciones a las que condicionamos, que al existir parte estacional son bastantes. El segundo modelo da también resultados similares. Los otros dos modelos tienen una estacionalidad más compleja y consiguen reducciones apreciables de la varianza residual. Las diferencias entre los modelos tercero y cuarto son pequeñas. Si descomponemos el operador del modelo cuarto e invertimos una raíz se obtiene, aproximadamente, el modelo tercero. En efecto, como:  $(1 - .61B^{12} - 0.31B^{24}) = (1 - .94B^{12})(1 + .33B^{12})$  y  $(1 + .33B^{12})^{-1} \approx (1 - .33B^{12} + .11B^{24})$  obtenemos un modelo similar al tercero. Las diferencias entre el tercero y el cuarto son muy pequeñas, pero como el cuarto explica algo más que el tercero con un parámetro menos será el elegido con cualquier criterio de selección.

---

### 10.8. Lecturas complementarias

El libro de Harvey (1989) desarrolla con detalle la representación en el espacio de los estados y el filtro de Kalman. Una referencia clásica sobre este tema es Anderson y Moore (1979) y una reciente Durbin y Koopman (2000). La estimación Bayesiana de modelos en el espacio de los estados puede consultarse en West y Harrison (1997). Fuller (1996) estudia con detalle las propiedades de los estimadores. Buenas presentaciones de la estimación pueden verse en Brockwell y Davis (1991) y Shumway y Stof-

fer (2000). Para estudios más detallados de la verosimilitud de modelos ARIMA véase Luceño (1993, 1994) y Mauricio (1995, 2000).

Existe una amplia literatura sobre selección de modelos y Akaike (1973) y Schwarz (1978) son dos referencias básicas pioneras en este tema. McQuarrie y Tsai (1998) presentan una panorámica de distintos métodos, y Galeano y Peña (2005, a) una revisión desde el punto de vista de análisis discriminante. La relación entre validación cruzada y selección de modelos fue encontrada por Stone (1979). Para un procedimiento efectivo de realizar la validación cruzada con modelos ARIMA y su relación con criterios de selección de modelos, véase Peña y Sánchez (2005).

---

### Ejercicios 10

10.1 Demostrar que la varianza del estimador de  $\mu$  para un AR(p) obtenida por estimación condicionada es mayor que la varianza de la media muestral de todos los datos.

10.2 Demostrar que en un proceso AR(2) la distribución marginal de la primera observación es normal con media  $E[\omega_1] = \mu$  y  $Var[\omega_1] = \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2}$  y que para la segunda  $E[\omega_2|\omega_1] = \mu + \phi_1(\omega_1 - \mu)$  y  $Var[\omega_2|\omega_1] = \sigma^2 \frac{(1 - \phi_1^2)}{1 - \phi_1^2 - \phi_2^2}$ .

10.3 Demostrar que el estimador MV condicional de  $\hat{\sigma}_a^2$  para un proceso AR(p) es:

$$\hat{\sigma}_a^2 = \sum_{t=p+1}^T \left( \omega_t - \hat{\mu} - \sum_{i=1}^p \hat{\phi}_i (\omega_{t-i} - \hat{\mu}) \right)^2 / T.$$

10.4 Comprobar que la representación en el espacio de los estados de un modelo AR(1) tiene  $\Omega = \phi$ ,  $\mathbf{H} = 1$  y  $\mathbf{R} = \sigma^2$ .

10.5 Escribir las ecuaciones del filtro de Kalman para prever con un AR(1) y comprobar que se reducen a  $\hat{z}_{t|t-1} = \phi z_t$  con varianza  $p_{t|t-1} = \sigma^2$ .

10.6 Escribir las ecuaciones en el espacio de los estados para un MA(1) y comprobar que se verifica:

$$\Omega = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{H} = (0, 1) \quad \mathbf{R} = \sigma^2 \begin{bmatrix} 0 & -\theta \\ -\theta & \theta^2 \end{bmatrix}.$$

## 10. Estimación y selección de modelos ARIMA

10.7 Utilizando que para matrices cuadradas  $\mathbf{A}$   $\mathbf{C}$  y rectangulares con las dimensiones apropiadas  $\mathbf{B}, \mathbf{D}$  se verifica que:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1},$$

demonstrar que la ecuación de revisión de las covarianzas de estimación del estado puede escribirse como  $\mathbf{S}_t^{-1} = \mathbf{S}_{t|t-1}^{-1} + \mathbf{H}'_t \mathbf{V}_t^{-1} \mathbf{H}_t$ .

10.8 Llamando precisión a la inversa de la varianza, justificar que la expresión anterior se interpreta como que la precisión final es la suma de la precisión inicial y la precisión aportada por la última observación.

10.9 Escribir las ecuaciones del filtro de Kalman para un proceso AR(2) y relacionar el método de cálculo de las predicciones con el estudiado en el capítulo 8.

---

### Apéndice 10.1: Algoritmos de optimización no lineal

Supongamos una función  $f(\theta)$  con primeras y segundas derivadas continuas cuyo mínimo se quiere calcular. (Si se tratara de su máximo convertiríamos el problema en hallar el mínimo de  $-f(\theta)$ ). La condición necesaria y suficiente para que  $\theta^*$  sea un mínimo local es que el vector de primeras derivadas, que llamaremos vector gradiente, sea nulo en  $\theta^*$  y que la matriz de segundas derivadas (hessiana) sea definida positiva en  $\theta^*$ . Llamando:

$$\mathbf{g}(\theta) = \frac{\partial f(\theta)}{\partial \theta_i}$$

al vector gradiente y

$$\mathbf{H}(\theta) = \frac{\partial^2 f(\theta)}{\partial \theta \partial \theta'}$$

a la matriz hessiana, las condiciones de mínimo local son:

$$\mathbf{g}(\theta^*) = \mathbf{0},$$

$$\mathbf{H}(\theta^*) = \text{definida positiva.}$$

Los algoritmos de optimización no lineal son procedimientos iterativos para pasar de un valor  $\theta_i$  a otro  $\theta_{i+1}$  más próximo al mínimo, según la relación:

$$\theta_{i+1} = \theta_i + \lambda_i \mathbf{d}_i \quad (10.41)$$

siendo  $\mathbf{d}_i$  un vector de dirección y  $\lambda_i$  la "amplitud" de paso. El valor final  $\theta_{i+1}$  se convierte en el inicial de la iteración siguiente, y el proceso continúa hasta

obtener convergencia, definida por  $|\theta_{i+1} - \theta_i| < \varepsilon_1$ ,  $|f(\theta_{i+1}) - f(\theta_i)| < \varepsilon_2$  y  $|g(\theta_{i+1})| < \varepsilon_3$ , siendo  $\varepsilon_1, \varepsilon_2, \varepsilon_3$ , valores pequeños fijados previamente, que dependen de la precisión deseada.

Los algoritmos difieren principalmente por su elección de  $d_i$ , ya que  $\lambda_i$  puede determinarse siempre por una búsqueda en dicha dirección. Una posible elección de  $d_i$  es:

$$d_i = -g(\theta)$$

ya que una función disminuye siempre con un pequeño movimiento en la dirección del gradiente negativo. Este es el *método del gradiente* que consiste en calcular este vector y desplazarse en la dirección asociada a sus valores negativos, determinando  $\lambda_i$  de manera que la función disminuya lo máximo posible en cada iteración. Éste método permite avanzar rápidamente cuando nos encontramos lejos del mínimo, pero es muy lento al acercarnos a éste.

Un mejor procedimiento cuando partimos de un valor próximo al mínimo es el de *Newton-Raphson*, que se basa en que cerca del óptimo la función debe ser aproximadamente cuadrática. Aproximando la función en el punto de llegada de la iteración  $\theta_{i+1}$ , por una función cuadrática:

$$\begin{aligned} f(\theta_{i+1}) \simeq f(\theta_i) + (\theta_{i+1} - \theta_i)' g(\theta_i) + \\ + 1/2 (\theta_{i+1} - \theta_i)' H(\theta_i) (\theta_{i+1} - \theta_i) \end{aligned} \quad (10.42)$$

derivando esta expresión respecto a  $\theta_{i+1}$  y considerando constantes los términos que dependen de  $\theta_i$ , se obtiene que:

$$g(\theta_{i+1}) \simeq g(\theta_i) + H(\theta_i)(\theta_{i+1} - \theta_i) \quad (10.43)$$

es decir, aproximar la función por una cuadrática equivale a aproximar el gradiente linealmente en un entorno de  $\theta_{i+1}$ . Imponiendo la condición de que  $\theta_{i+1}$  sea un mínimo local,  $g(\theta_{i+1}) = 0$ , y despejando en (10.43) resulta el algoritmo:

$$\theta_{i+1} = \theta_i - H^{-1}(\theta_i) g(\theta_i) \quad (10.44)$$

que es el algoritmo de Newton-Raphson.

Este algoritmo puede presentar problemas cuando se comienza con un valor inicial lejos del mínimo y la aproximación cuadrática no es buena. Por ejemplo, si la función es aproximadamente lineal, la matriz hessiana será nula, y el método fallará. Además, aunque la matriz hessiana,  $H(\theta_i)$ , no sea singular, es posible que lejos del mínimo no sea definida positiva, con lo que el valor en  $\theta_{i+1}$  puede ser mayor que en  $\theta_i$ , y el método puede no converger. En contrapartida, cerca del mínimo es un buen algoritmo, aunque costoso computacionalmente, ya que requiere el cálculo de primeras y segundas derivadas.

## 10. Estimación y selección de modelos ARIMA

Se han desarrollado métodos para aproximar (10.44) utilizando una matriz  $\mathbf{H}$  que no tenga que calcularse en cada etapa. Si  $f$  es una función soporte, una solución propuesta por R. A. Fisher es tomar en todas las iteraciones el valor esperado de  $\mathbf{H}$ , que es la matriz de información. Este algoritmo, conocido por el *algoritmo de la tasa de discriminación (scoring method)* es:

$$\theta_{i+1} = \theta_i + \mathbf{IE}(\theta_i)^{-1} \mathbf{g}(\theta_i) \quad (10.45)$$

donde  $\mathbf{IE}(\theta)$  es el valor esperado de la matriz  $(-\mathbf{H}(\theta))$ .

En la estimación condicional de modelos ARMA la función a minimizar puede escribirse:

$$L(\theta) = \sum_{t=r+1}^T a_t^2(\theta)$$

donde  $r = \max(p, q)$ . Maximizar la verosimilitud condicional es equivalente a minimizar una suma de cuadrados. Para aplicar el procedimiento de Newton-Raphson a esta función es necesario calcular su gradiente, que en este caso será:

$$\mathbf{g}(\theta) = \sum_{t=r+1}^T 2a_t(\theta) \frac{\partial a_t(\theta)}{\partial \theta} = \sum_{t=r+1}^T 2a_t(\theta) \mathbf{g}(a_t) \quad (10.46)$$

llamando  $\mathbf{g}(a_t)$  al vector de primeras derivadas de  $a_t$  respecto a cada componente del vector  $\theta$ . El hessiano será, derivando de nuevo en (10.46):

$$\mathbf{H}(\theta) = 2 \sum_{t=r+1}^T \mathbf{g}(a_t) \mathbf{g}'(a_t) + 2 \sum_{t=r+1}^T a_t(\theta) \mathbf{H}(a_t)$$

siendo  $\mathbf{H}(a_t)$  la matriz hessiana de segundas derivadas de  $a_t$  respecto a los términos de  $\theta$ . Supongamos ahora que  $\mathbf{H}(a_t)$  es despreciable en cada iteración frente al primer término de primeras derivadas. Esto equivale a una aproximación lineal para los residuos del tipo:

$$a_{t,i+1} = a_t(\theta_{i+1}) \doteq a_t(\theta_i) + (\theta_{i+1} - \theta_i)' \mathbf{g}_i(a_t) \quad (10.47)$$

ya que suponemos  $\mathbf{H}(a_t)$  despreciable frente a  $\mathbf{g}(a_t)$ . Entonces, la expresión general (10.44) se reduce a:

$$\theta_{i+1} = \theta_i - \left[ \sum_{t=r+1}^T \mathbf{x}_{ti} \mathbf{x}_{ti}' \right]^{-1} \sum_{t=r+1}^T \mathbf{x}_{ti} a_{ti} \quad (10.48)$$

donde  $\mathbf{x}_{ti} = \mathbf{g}(a_{ti})$  es el vector de derivadas de los residuos respecto a los parámetros  $\theta_i$  y  $a_{ti}$  los residuos calculados con ellos. Esta ecuación indica que pasaremos de un punto al siguiente añadiendo al valor inicial el vector de parámetros resultante de realizar una regresión con variable dependiente

$a_{ti}$  ( $t = 1, \dots, n$ ) y vector de regresores  $\mathbf{x}_{ti}$ . Este mismo resultado se deduce de (10.47). Escribiendo:

$$a_{ti} = a_t(\theta_i) = \mathbf{x}'_{ti}\beta + a_{t,t+1} \quad (10.49)$$

siendo  $\beta = -(\theta_{i+1} - \theta_i)$ , es claro que  $\beta$  se obtiene por regresión y que los nuevos residuos resultantes,  $a_{t,t+1}$  tendrá siempre menor varianza que la variable dependiente  $-a_{ti}$ , con lo que avanzaremos hacia el mínimo de la suma de cuadrados. Este método se conoce como algoritmo de Gauss-Newton.

El algoritmo (10.48) puede mejorarse en dos aspectos: cuando estamos lejos del óptimo el avance a través de aproximaciones cuadráticas es muy lento mientras que, como sabemos, el método del gradiente proporciona una dirección de avance más rápida. Prescindiendo de subíndices y llamando  $\mathbf{X}'\mathbf{X} = \sum_{t=r+1}^T \mathbf{x}_t \mathbf{x}'_t$  y  $\mathbf{X}'\mathbf{Y} = \sum_{t=r+1}^T \mathbf{x}_t a_t$ , esto sugiere un esquema del tipo:

$$\theta^* = \theta - [\mathbf{X}'\mathbf{X} + \delta \mathbf{I}]^{-1} \mathbf{X}'\mathbf{Y} \quad (10.50)$$

donde  $\delta$  se determina en función del avance de la función y es muy grande al principio –para que  $\mathbf{X}'\mathbf{X}$  sea despreciable frente a  $\delta$  veces la matriz unidad  $\mathbf{I}$ , y la dirección sea la del gradiente  $\mathbf{g}(\theta)$  (10.46)– y muy pequeña cerca del mínimo, con lo que (10.50) equivale a Gauss-Newton. Las iteraciones pueden interpretarse ahora como realizar regresiones-cresta (*ridge regression*). Este algoritmo es debido a Marquard (1963).

La implantación de estos algoritmos para estimar procesos ARMA es directa. Por ejemplo, la estimación condicional se realiza como sigue:

1. Comenzar con un valor inicial  $\hat{\beta}_0$  de los parámetros calculado con el algoritmo de Hannan-Rissanen.
2. Calcular los residuos del modelo con el valor supuesto de los parámetros, mediante (10.15).
3. Calcular las derivadas primeras de los residuos respecto al vector de parámetros. Este cálculo puede realizarse numéricamente definiendo:

$$\frac{\partial a_t(\hat{\beta}_0)}{\partial \beta_i} \approx \frac{a_t(\hat{\beta}_0 + \mathbf{1}_i \varepsilon) - a_t(\hat{\beta}_0)}{\varepsilon} \quad i = 1, \dots, p+1$$

donde  $\mathbf{1}_i$  es un vector con un uno en la componente  $i$  y cero en otro caso y  $\varepsilon$  es un número pequeño (0,001 o similar). El cálculo de las derivadas requiere repetir  $(p+1)$  veces el paso (2), modificando cada vez ligeramente un componente del vector  $\hat{\beta}_0$ .

4. Estimar los coeficientes de una regresión entre los residuos calculados en (2) como la variable dependiente y un vector de  $(p+1)$  variables explicativas  $\mathbf{X}_t$ , cuyos componentes son las derivadas parciales de los residuos calculadas en (3) y cambiadas de signo. Es decir:

## 10. Estimación y selección de modelos ARIMA

$$x_{it} = -\frac{\partial a_t(\hat{\beta}_0)}{\partial \beta_i}$$

5. Sea  $\mathbf{b}$  el vector de parámetros estimado con la regresión anterior que mide el efecto lineal sobre los residuos de cambios en el vector de parámetros. Definir un nuevo estimador  $\hat{\beta}_1$  con:

$$\hat{\beta}_1 = \hat{\beta}_0 + \mathbf{b} \quad (10.51)$$

y volver al paso (2) tomando ahora  $\hat{\beta}_1$  como estimador inicial. Repetir el proceso hasta obtener convergencia, definida por  $|\beta_{i+1} - \beta_i| < \alpha$ , y  $|\partial L(\beta_i)/\partial \beta| < \gamma$ , para todos sus componentes, siendo  $\alpha$  y  $\gamma$  constantes pequeñas.

### Apéndice 10.2: El criterio AIC

Vamos a deducir el criterio AIC para comparar modelos autorregresivos. La esperanza de menos dos veces la verosimilitud para muestras grandes será, despreciando la parte que corresponde a las  $p$  primeras observaciones,

$$E(-2L(\mu, \phi, \sigma^2)) = T \ln \sigma^2 + E \left[ \sum_{t=1}^T \frac{\hat{a}_t^2}{\sigma^2} \right]. \quad (10.52)$$

Llamando  $\omega_{t-p} = (\omega_{t-1}, \dots, \omega_{t-p})'$  al vector de regresores y  $\phi = (\phi_1, \dots, \phi_p)'$  al vector de parámetros, el error de predicción cuando los parámetros se estiman con la muestra puede escribirse como:

$$\hat{a}_t = \omega_t - \hat{\phi}' \omega_{t-p} = \omega_t - \phi' \omega_{t-p} + (\phi - \hat{\phi})' \omega_{t-p} = a_t + (\phi - \hat{\phi})' \omega_{t-p}, \quad (10.53)$$

y es la suma de dos términos: el error de predicción que tendríamos aunque los parámetros fuesen conocidos,  $a_t$ , y el error de predicción debido a que los parámetros estimados pueden ser distintos de los verdaderos. Elevando al cuadrado esta expresión y sumando para todos los datos obtenemos:

$$\sum_{t=1}^T \hat{a}_t^2 = \sum_{t=1}^T a_t^2 + (\phi - \hat{\phi})' \left[ \sum_{t=1}^T \omega_{t-p} \omega_{t-p}' \right] (\phi - \hat{\phi}) + 2(\phi - \hat{\phi})' \sum_{t=1}^T \omega_{t-p} a_t.$$

Vamos a suponer que  $a_t$  y  $\hat{\phi}$  son independientes, lo que supone que los parámetros se han estimado en una muestra distinta de la utilizada en la predicción. Al tomar esperanzas, el doble producto es cero, y dividiendo todo por  $\sigma^2$ , tenemos que el error relativo esperado es:

$$E \left[ \sum_{t=1}^T \frac{\hat{a}_t^2}{\sigma^2} \right] = \frac{1}{\sigma^2} (T\sigma^2 + \sigma^2 E \left[ (\phi - \hat{\phi})' \left( \sum_{t=1}^T \omega_{t-p} \sigma^{-2} \omega'_{t-p} \right) (\phi - \hat{\phi}) \right]).$$

El segundo término del segundo miembro se distribuye como una  $\chi_p^2$ , una ji cuadrado con  $p$  grados de libertad, y su esperanza será  $p$ . Sustituyendo este resultado en (10.52) :

$$E(-2L(\mu, \phi, \sigma^2)) = T \ln \sigma^2 + T + p$$

Para calcular esta expresión necesitamos un estimador de  $\sigma^2$ . Tomando como estimador la varianza residual del modelo, que es insesgada para  $\sigma^2$ , dada por

$$\hat{\sigma}^2 = \frac{T}{T-p} \hat{\sigma}_{MV}^2,$$

donde  $\hat{\sigma}_{MV}^2$  es el estimador MV utilizando los  $T$  datos, podemos escribir:

$$\begin{aligned} E(-2L(\mu, \phi, \sigma^2)) &= -T \ln \left( \frac{T-p}{T} \right) + T \ln \hat{\sigma}_{MV}^2 + T + p \approx \\ &\approx T \ln \hat{\sigma}_{MV}^2 + T + 2p, \end{aligned}$$

donde hemos utilizado que, para  $T$  grande,  $\ln(1 - p/T) \approx -p/T$ . Esta expresión indica que para minimizar la esperanza de menos la verosimilitud (que obviamente equivale a maximizar la esperanza de la verosimilitud) debemos seleccionar el modelo que minimize  $T \ln \hat{\sigma}_{MV}^2 + 2p$ , que es el criterio AIC.

### Apéndice 10.3: El criterio BIC

Las probabilidades a posteriori de los distintos modelos se obtienen con el teorema de Bayes:

$$P(M_i | \omega_T) = \frac{P(\omega_T | M_i) P(M_i)}{P(\omega_T)},$$

suponiendo que las probabilidades a priori son las mismas, y como

$$P(\omega_T) = \sum_{i=1}^H P(\omega_T | M_i) P(M_i)$$

es una constante para todos los modelos, tenemos que

$$-2 \ln P(M_i | \omega_T) = -2 \ln P(\omega_T | M_i) + c$$

donde  $c$  es una constante que incluye las probabilidades a priori de los modelos que suponemos son iguales. Para calcular  $P(\omega | M_i)$  utilizamos que, llamando

## 10. Estimación y selección de modelos ARIMA

---

$\theta$  al vector de parámetros:

$$P(\omega_T, \theta_T | M_i) = P(\omega_T | \theta, M_i)P(\theta | M_i) = P(\theta | \omega_T, M_i)P(\omega_T | M_i)$$

y multiplicando por  $-2$  y tomando logaritmos en esta ecuación, podemos escribir:

$$\begin{aligned} -2 \ln P(\omega_T | M_i) &= -2 \ln P(\omega_T | \theta, M_i) - \\ &-2 \ln P(\theta | M_i) + 2 \ln P(\theta | \omega_T, M_i). \end{aligned} \quad (10.54)$$

El primer término del segundo miembro es menos dos veces la verosimilitud, el segundo menos dos veces la probabilidad a priori de los parámetros, que suponemos es constante sobre la región de interés, y el tercero dos veces la probabilidad a posteriori de los parámetros. Supongamos ahora, para simplificar el análisis, que tenemos procesos AR( $p$ ) de distintos órdenes y utilicemos la verosimilitud condicionada dada por (10.9). Si sustituimos los parámetros por sus estimaciones MV, tenemos que, para  $T$  grande de manera que  $T - p$  y  $T$  son similares,

$$\ln P(\omega_T | \hat{\theta}, M_i) = -\frac{T}{2} \ln \hat{\sigma}_{MV}^2 + c_1.$$

Por otro lado, la distribución a posteriori de los parámetros puede demostrarse que es normal con media  $\hat{\theta}$ , el estimador MV de los parámetros, y con matriz de varianzas y covarianzas que depende del tamaño muestral. Puede demostrarse que, sustituyendo los parámetros por sus estimaciones de MV, para tamaño muestral grande:

$$\ln P(\hat{\theta} | \omega_T, M_i) = \frac{k}{2} \ln T + c_2$$

siendo  $k$  la dimensión del vector de parámetros. Sustituyendo estas expresiones en (10.54), tenemos que, para tamaño muestral grande,

$$-2 \ln P(M_i | \omega_T) = T \ln \hat{\sigma}_{MV}^2 + k \ln T. \quad (10.55)$$

El modelo que maximiza la probabilidad a posteriori es el que minimiza  $T \ln \hat{\sigma}_{MV}^2 + k \ln T$ . Éste es el criterio BIC. Por otro lado, de la derivación anterior concluimos que si conocemos los valores del criterio BIC para varios modelos podemos deducir sus probabilidades a posteriori. De (10.55) concluimos que las probabilidades a posteriori son:

$$P(M_i | \omega_T) = ce^{-\frac{1}{2}BIC_i}$$

es decir:

$$P(M_i | \omega_T) = \frac{e^{-\frac{1}{2}BIC_i}}{\sum_{j=1}^H e^{-\frac{1}{2}BIC_j}}.$$

#### Apéndice 10.4: Programas para la estimación

La estimación de un modelo ARIMA con Minitab o SPSS es muy simple. En la parte de series temporales ambos programas permiten especificar el modelo ARIMA indicando los órdenes de los operadores y realizan la estimación por máxima verosimilitud condicionada. Se obtienen los parámetros estimados y sus desviaciones típicas.

EViews también hace una estimación condicional. Por ejemplo, para estimar un modelo ARIMA(3,1,0) a la serie x, marcamos en el menú principal de EViews: *Quick → Estimate Equation* y escribimos,

$$d(x, 1) ar(1) ar(2) ar(3)$$

Lo que aparece en pantalla es un cuadro donde aparecen las estimaciones y sus errores estándar, así como un estadístico *t* de dicha estimación para cada uno de los parámetros.

Dos programas que permiten máxima verosimilitud exacta son SCA y TSW. Vamos a exponer la estimación con este último programa.

#### TSW

Supongamos que se desea estimar el modelo ARIMA(0,1,1) $\times$ (0,1,1)<sub>12</sub>, para la serie. Marcamos **+Model** en el menú y aparece una ventana donde podemos cambiar el parámetro RSA que se utiliza para identificación automática. Para estimar un modelo definido dejamos RSA=0. Pasamos a la pestaña ARIMA model y marcamos:

$$\begin{aligned} P = 0.; Q = 1.; D = 1.; BP = 0.; BQ = 1.; BD = 1.; INIT = 0.; \\ IMEAN = 1; LAM = -1; FCT = 1; TYPE = 0; UNITS = 0. \end{aligned}$$

Como indicamos antes, marcando un parámetro y pulsando la tecla F1 se obtiene una explicación de su valor. Marcamos **Others** para nuevos parámetros. De todos estos valores, sólo nos interesarán en un futuro los correspondientes a **Outliers**. Ahora marcamos IATIP=0. Marcamos OK. Marcamos el modelo en la tabla y marcamos Run. En output, obtenemos la salida del programa.

---

# 11. Diagnosis del modelo y predicción



**Hirotugu Akaike (1919)**

Estadístico japonés. Ha investigado extensamente las relaciones entre la teoría de información y la estadística y fue pionero en el desarrollo de criterios de selección de modelos en series temporales. El criterio que propuso, AIC, se ha utilizado en todas las ramas de la ciencia y ha contribuido a modificar la metodología estadística en la última parte del siglo XX. Ha hecho también contribuciones destacadas en la formulación de los modelos de series temporales y ha tenido un papel muy importante en el desarrollo de la estadística en Japón.

## 11.1. Introducción

La diagnosis del modelo requiere comprobar que las hipótesis básicas realizadas respecto a los residuos son ciertas. Estos deben tener: (1) media marginal igual a cero; (2) varianza marginal constante; (3) falta de correlación para cualquier retardo; (4) distribución normal. Además, estas propiedades deben verificarse no sólo respecto a las distribuciones marginales sino también a las distribuciones condicionadas a cualquier conjunto de información de valores pasados de la serie. Por ejemplo, para la media la condición es:

$$E(a_t|z_{t-1}, \dots, z_1) = E(a_t|a_{t-1}, \dots, a_1) = 0.$$

que es mucho más fuerte que la condición de media marginal cero. Por ejemplo, esta propiedad implica, según demostramos en el capítulo 3, las (1) y (3). Para la varianza suponemos que:

$$Var(a_t|z_{t-1}, \dots, z_1) = Var(a_t|a_{t-1}, \dots, a_1) = \sigma^2,$$

que generaliza la condición de varianza marginal constante para cualquier distribución condicionada. En el capítulo 15 presentaremos ejemplos de procesos que cumplen la condición de varianza marginal constante, pero no de varianzas condicionadas constantes.

De las cuatro condiciones establecidas para las distribuciones marginales la (1) es poco restrictiva. Es posible que el modelo sea muy incorrecto, y sin embargo se verifique (1). Por otro lado, esta condición puede no verificarse si existen valores atípicos, como estudiaremos en el capítulo 13. La condición de varianza marginal constante es más fuerte. La condición (3), falta de incorrelación para cualquier retardo, es central para asegurar que el modelo es adecuado y se analizará con detalle en las secciones siguientes. Finalmente, la condición de normalidad es conveniente, porque nos garantiza que la incorrelación implica la independencia, y que no estamos dejando información por modelizar. Cuando los residuos de la serie sean marcadamente no normales es posible que podamos mejorar el modelo de la serie analizando los datos atípicos (véase el capítulo 13) o mediante modelos no lineales (véase el capítulo 14).

La diagnosis está relacionada, pero no es idéntica, a la selección de modelos, estudiada en el capítulo anterior. Es posible que el mejor modelo seleccionado dentro de una clase conduzca a residuos que no verifican las hipótesis anteriores, y entonces tendremos que buscar un nuevo modelo en una clase más amplia. También es posible que dispongamos de varios modelos cuyos residuos verifican las hipótesis anteriores, y entonces podemos seleccionar el mejor entre ellos con un criterio de selección. Vemos, por tanto, que la diagnosis del modelo es una etapa complementaria a la selección del mejor modelo.

Además de comprobar las hipótesis básicas, la etapa de diagnosis del modelo incluye la detección de términos deterministas, si existen, que se manifiesta por términos de media móvil no invertibles. Finalmente, en este capítulo estudiaremos el cálculo de las predicciones con parámetros estimados de un modelo ARIMA y la incorporación de la incertidumbre respecto al modelo correcto que siguen los datos.

## 11.2. Contrastes de autocorrelación

### 11.2.1. Propiedades de las autocorrelaciones estimadas

El primer contraste a realizar es si los residuos estimados están incorrelados. Para ello, se calcula su *faz* mediante:

$$\hat{r}_k = \frac{\sum_{t=1}^{T-k} (\hat{a}_t - \bar{a})(\hat{a}_{t+k} - \bar{a})}{\sum_{t=1}^T (\hat{a}_t - \bar{a})^2}, \quad (11.1)$$

donde  $\bar{a}$  es la media de los  $T$  residuos. Si los residuos son independientes, los coeficientes,  $\hat{r}_k$ , para  $k$  no muy pequeño, serán aproximadamente variables aleatorias con media cero, varianza asintótica  $1/T$  y distribución normal. La varianza asintótica es válida para  $k$  grande, pero no para los primeros retardos. Por ejemplo, puede demostrarse que si la serie sigue un proceso AR(1) la desviación típica asintótica de la autocorrelación de orden uno para los residuos,  $\hat{r}_1$ , es  $\sqrt{(1 - \phi^2)/T}$ , que puede ser mucho menor que  $1/\sqrt{T}$ . En consecuencia, el valor  $1/\sqrt{T}$  debe considerarse como un límite máximo de la desviación típica de las autocorrelaciones residuales.

El procedimiento habitual de verificar la incorrelación de los residuos es dibujar dos líneas paralelas a distancia  $2/\sqrt{T}$  del origen en sus funciones de autocorrelación simple o parcial estimadas, y comprobar si todos los coeficientes  $\hat{r}_k$  están dentro de estos límites de confianza. Como estos intervalos son, aproximadamente, del 95 %, en promedio uno de cada veinte coeficientes de autocorrelación estimados saldrá fuera, por lo que la aparición de un valor significativo en un retardo elevado es esperable. Si embargo, como según lo anterior estos límites sobreestiman la varianza en los retardos pequeños, un valor próximo a los límites de confianza  $\pm 2/\sqrt{T}$  en los retardos iniciales debe considerarse un indicio claro de que el modelo es inadecuado.

### 11.2.2. El contraste de Ljung-Box sobre las autocorrelaciones

Un contraste global de que los primeros  $h$  coeficientes son cero ( $h$  debe ser grande) es el contraste de Ljung-Box. Si los residuos son realmente ruido blanco, hemos visto en el capítulo 3 que los coeficientes de correlación estimados son asintóticamente normales, con media cero y varianza  $(T-k)/T(T+2)$ . Por tanto, el estadístico:

$$Q(h) = T(T+2) \sum_{j=1}^h \frac{\hat{r}_j^2}{T-j}, \quad (11.2)$$

se distribuye, asintóticamente, como una  $\chi^2$  con grados de libertad igual al número de coeficientes en la suma ( $h$ ) menos el número de parámetros estimados,  $n$ . Para modelos no estacionales  $n = p + q + 1$ , o  $n = p + q$ , según que el modelo tenga o no constante, y para los estacionales, que

habitualmente no tienen constante,  $n = P + p + Q + q$ . Concluiremos que el modelo es inadecuado si el valor de  $Q(h)$  obtenido con (11.2) es mayor que el percentil 0.95 de la distribución  $\chi^2$  con  $h - n$  grados de libertad, que denotaremos por  $\chi^2_{.95}(h - n)$ . En general, rechazaremos la hipótesis de incorrelación de los residuos cuando la probabilidad:

$$P(\chi^2(h - n)) > Q(h)$$

sea pequeña (menor que 0.05 o 0.01).

### 11.2.3. El contraste del determinante

El contraste de Ljung-Box tiene el inconveniente de dar, aproximadamente, el mismo peso a todos los coeficientes de autocorrelación y ser invariante ante permutaciones de estos coeficientes. Sin embargo, intuitivamente deberíamos dar más peso a los coeficientes de orden bajo que a los de orden alto. Peña y Rodríguez (2003, 2005) han propuesto un contraste más potente que el de Ljung-Box que tiene esta propiedad. El contraste se basa en la matriz de autocorrelación de los residuos:

$$\mathbf{R}_m = \begin{vmatrix} 1 & \hat{r}_1 & \dots & \hat{r}_{m-1} \\ \hat{r}_1 & 1 & \dots & \hat{r}_{m-2} \\ \dots & \dots & 1 & \dots \\ \hat{r}_{m-1} & \hat{r}_{m-2} & \dots & 1 \end{vmatrix}$$

y el estadístico del contraste es:

$$D_m = -\frac{T}{m+1} \log |\hat{\mathbf{R}}_m|, \quad (11.3)$$

que sigue, asintóticamente, una distribución gamma con parámetros:

$$\alpha = \frac{3(m+1)\{m-2(p+q)\}^2}{2\{2m(2m+1)-12(m+1)(p+q)\}}$$

$$\beta = \frac{3(m+1)\{m-2(p+q)\}}{2m(2m+1)-12(m+1)(p+q)}.$$

La distribución tiene media  $\alpha/\beta = (m+1)/2 - (p+q)$  y varianza:

$$\alpha/\beta^2 = (m+1)(2m+1)/3m - 2(p+q).$$

Los percentiles del estadístico  $D_m$  para el contraste se obtienen fácilmente calculando los parámetros de la gamma con las fórmulas anteriores y utilizando tablas de esta distribución (muchos programas de ordenador incluyen esta función). Alternativamente, puede utilizarse que la variable

$$ND_m^* = (\alpha/\beta)^{-1/4} (4/\sqrt{\alpha}) \left( (D_m)^{1/4} - (\alpha/\beta)^{1/4} \left( 1 - \frac{1}{6\alpha} \right) \right), \quad (11.4)$$

se distribuye, aproximadamente, como una variable normal estándar.

Puede demostrarse que el estadístico (11.3) puede escribirse como:

$$D_m = T \sum_{i=1}^m \frac{(m+1-i)}{(m+1)} \hat{\pi}_i^2, \quad (11.5)$$

donde  $\hat{\pi}_i^2$  es el cuadrado del coeficiente de correlación parcial de orden  $i$ . Este contraste puede verse como un contraste Ljung-Box modificado, donde en lugar de utilizar los coeficientes de correlación múltiple utilizamos los parciales, pero con ponderaciones  $(m+1-i)/m$ , que decrecen linealmente con el retardo, de manera que  $\hat{\pi}_1^2$  tiene peso uno y  $\hat{\pi}_m^2$  peso  $1/m$ .

#### 11.2.4. Sobreajuste: contrastes con el criterio BIC

Una técnica complementaria a los contrastes anteriores es la técnica del sobreajuste, que consiste en estimar un modelo de orden mayor que el analizado y comprobar si se obtienen coeficientes estimados significativos. Con esto es posible captar pequeñas estructuras remanentes que pueden mejorar las predicciones, pero que quizás no se detectan claramente en el análisis de los residuos. En general, si hemos ajustado un ARIMA( $p, d, q$ ) que parece adecuado, el sobreajuste se aplica estimando los modelos ARIMA  $(p+r, d, q)$  y ARIMA  $(p, d, q+r)$ , para un valor de  $r$  bajo, normalmente 1 o 2, y comprobando si los parámetros adicionales son significativos.

No conviene ampliar el orden de las partes AR y MA a la vez, ya que puede producirse una compensación de efectos. Si el modelo ajustado inicialmente es:

$$\phi(B) z_t = \theta(B) a_t \quad (11.6)$$

obtendremos un ajuste igualmente bueno con:

$$\phi^*(B) z_t = \theta^*(B) a_t \quad (11.7)$$

siendo  $\phi^*(B) = \phi(B)(1-\phi B)$  y  $\theta^*(B) = \theta(B)(1-\theta B)$  y  $\phi$  aproximadamente igual a  $\theta$ . Por tanto, si el modelo adecuado es (11.6) y estimamos (11.7) obtendremos todos los parámetros significativos y sólo notaremos la sobreparametrización al factorizar los operadores AR y MA. En consecuencia, siempre es conveniente obtener las raíces de los ope-

radores AR y MA en modelos mixtos y comprobar que no hay factores que se cancelan en ambos lados.

Una manera automática de realizar el sobreajuste es ajustar modelos AR hasta un orden  $p_{\max}$  prefijado a los residuos del modelo y seleccionar el mejor modelo AR mediante el criterio BIC. Si el mejor modelo seleccionado es un AR(0), es decir, un proceso de ruido blanco, aceptamos que los residuos están incorrelados. En caso contrario, rechazamos que el modelo sea adecuado y el grado del mejor modelo seleccionado nos indica cómo modificar el modelo actual. Como el criterio BIC es consistente, es decir para tamaños muestrales grandes tiende a seleccionar el modelo adecuado con probabilidad que tiende hacia uno, este procedimiento funcionará bien en la práctica si tenemos muestras grandes. Con muestras pequeñas disponemos de menos garantías, y es conveniente realizar siempre el contraste del determinante sobre los residuos estimados. Observemos que este procedimiento no puede aplicarse con el criterio AIC, ya que puede comprobarse que cuando los residuos son ruido blanco todavía existe una probabilidad apreciable, del orden del 15 %, de que el AIC seleccione un modelo AR para los residuos.

### 11.3. Otros contrastes

#### 11.3.1. Contrastos de media cero

Los residuos estimados de un modelo ARIMA no están en principio sujetos a la restricción  $\sum_{t=1}^T \hat{a}_t = 0$ . Esta condición sólo se impone cuando tenemos un modelo AR y sus parámetros se han estimado por mínimos cuadrados, pero en la estimación por máxima verosimilitud exacta esta restricción no existe. Para contrastar la hipótesis de que la perturbación tiene esperanza nula en el caso general, suponiendo  $T$  residuos y  $p+q$  parámetros, calcularemos su media:

$$\bar{a} = \frac{\sum_{t=1}^T \hat{a}_t}{T},$$

y su varianza

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T (\hat{a}_t - \bar{a})^2}{T},$$

y concluiremos que  $E[\hat{a}_t] \neq 0$ , si

$$\frac{\bar{a}}{\hat{\sigma}/\sqrt{T}}$$

es significativamente grande con relación a la distribución  $N(0, 1)$ . Este

## 11. Diagnosis del modelo y predicción

contraste debe de aplicarse después de comprobar que los residuos están incorrelados, para asegurar que  $\hat{\sigma}^2$  es un estimador razonable de la varianza.

### 11.3.2. Contrastes de homocedasticidad

La estabilidad de la varianza marginal de los residuos se comprueba estudiando el gráfico de los residuos a lo largo del tiempo. En los capítulos siguientes estudiaremos contrastes adicionales para ciertas formas de heterocedasticidad condicional. Si a la vista de los residuos estimados parece existir un cambio de varianza a partir de un punto,  $t = n_1$ , podemos dividir el intervalo muestral en dos partes y aplicar un contraste de varianzas. En la hipótesis de que ambos tramos tienen la misma varianza, el estadístico:

$$F = \frac{\sum_{t=1}^{n_1} \hat{a}_t^2 / n_1}{\sum_{t=n_1+1}^T \hat{a}_t^2 / (T - n_1)} = \frac{s_1^2}{s_2^2},$$

se distribuirá, aproximadamente, como una F con  $n_1$  y  $T - n_1$  grados de libertad. Para aplicar este contraste conviene que, al menos, dispongamos de 10 observaciones en cada tramo. De la misma forma, si sospechamos  $h$  cambios de varianza en los periodos  $n_1, \dots, n_h$  el contraste de igualdad de varianzas es:

$$\lambda = T \log \hat{\sigma}^2 - \sum_{i=1}^h n_i \log s_i^2$$

donde  $\hat{\sigma}^2$  es la varianza de los residuos en toda la muestra y  $s_i^2$  la varianza en el tramo  $i$  de  $n_i$  observaciones. Bajo la hipótesis de que la varianza es la misma en todos los tramos se demuestra que este estadístico será, asintóticamente, una ji cuadrado con  $h - 1$  grados de libertad.

### 11.3.3. Contrastes de normalidad

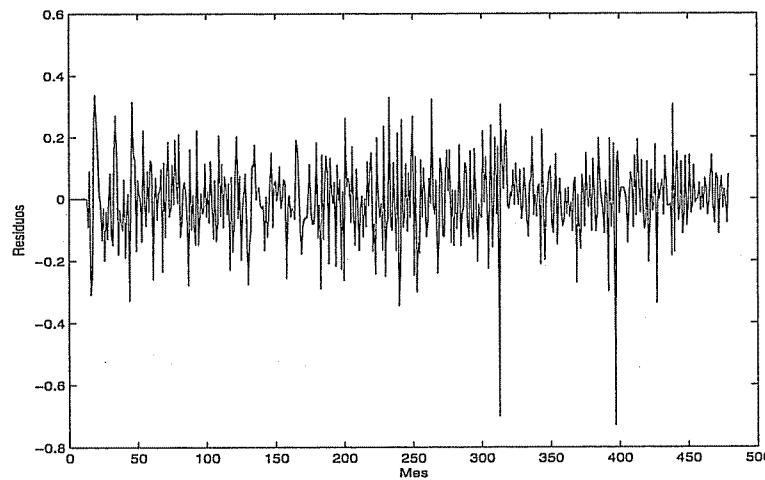
La hipótesis de que los residuos tienen una distribución normal se comprueba con cualquiera de los contrastes habituales. Un contraste simple es calcular los coeficientes de asimetría

$$\alpha_1 = \frac{\sum (\hat{a}_t - \bar{a})^3}{\hat{\sigma}^3},$$

y curtosis

$$\alpha_2 = \frac{\sum (\hat{a}_t - \bar{a})^4}{\hat{\sigma}^4},$$

**Figura 11.1 Residuos de la serie de matriculación de vehículos con el modelo  $(0,1,1)x(1,1,1)$**



de los residuos y utilizar que, bajo la hipótesis de normalidad, la variable:

$$X = \frac{T\alpha_1^2}{6} + \frac{T(\alpha_2 - 3)^2}{24}$$

es una  $\chi^2$  con dos grados de libertad.

Finalmente, conviene *siempre* estudiar el gráfico de los residuos estimados  $\hat{a}_t$  a lo largo del tiempo. Su inspección se facilita situando en él los límites  $\pm 2\hat{\sigma}$  y  $\pm 3\hat{\sigma}$ , y estudiando los puntos que salen fuera de estos límites de control. Con esto podemos detectar valores atípicos cuyo tratamiento se hará en los capítulos siguientes.

---

### Ejemplo 11.1

La figura 11.1 presenta el gráfico de los residuos de la serie de matriculación estimada con el modelo  $(0,1,1)x(1,1,1)_{12}$ . Se observan algunos valores atípicos bastante destacados. La *fas* de los residuos se presenta en la figura 11.2. Ningún coeficiente es significativo y el estadístico  $Q(24)$  es 29.5 y  $Q(39) = 41$ . Estos valores deben compararse con una  $\chi^2$  con 21 y 36 grados de libertad respectivamente y sus percentiles al 0.95 son 32.7 y 43.8, con lo que concluimos que, con este contraste, no encontramos ninguna evidencia de dependencia serial en los residuos.

La media de estos residuos es -.0017 y su desviación típica .0497, que supone

Figura 11.2 Fas muestral de la serie de residuos de matriculación de vehículos

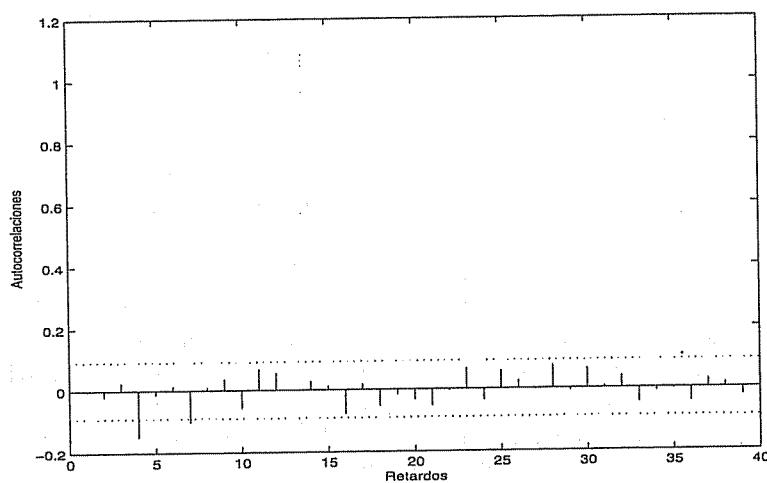
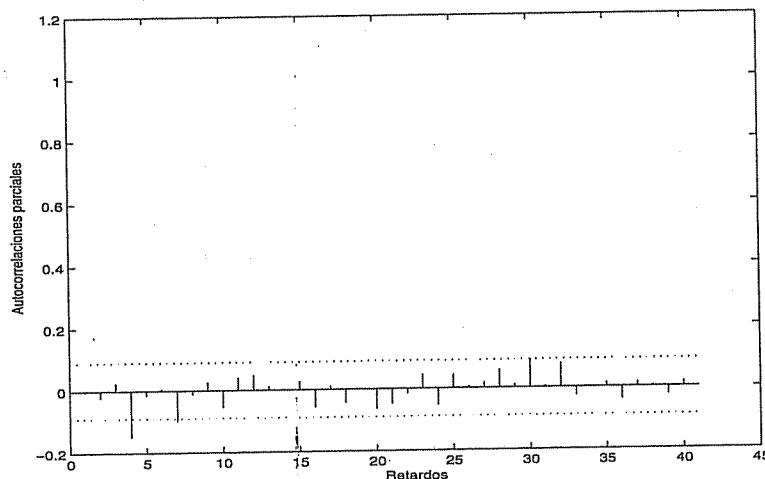


Figura 11.3 Fap muestral de los residuos de la serie de matriculación de vehículos



un error de predicción porcentual del 4.97 %. La media de los residuos no es significativamente distinta de cero, ya que:

$$t = \frac{-0.0017}{.0497/\sqrt{479}} = -.74$$

y podemos concluir que la media poblacional de los residuos es cero. La variabilidad de los residuos, salvo por unos posibles valores atípicos, parece constante en el tiempo. De hecho, si dividimos la serie en dos partes y calculamos la varianza en cada una obtenemos valores muy similares. Concluimos por tanto que, aparte de algunos valores atípicos, los residuos parecen ruido blanco.

### Ejemplo 11.2

La figura 11.4 presenta el gráfico de los residuos de la serie de accidentes en logaritmos estimada con el modelo  $(2,1,0)x(0,1,2)_{12}$  que se presentó en el capítulo anterior. Estos residuos son los del modelo estimado por MV condicional, y por tanto el primer residuo es el de la observación 14, ya que se pierden doce residuos con la diferencia estacional más dos por el AR(2). En la figura se ha dividido cada residuo por su desviación típica, para facilitar su lectura. Los residuos muestran un valor extremo de casi cinco desviaciones típicas que corresponde al cambio de nivel que se observa en el gráfico de la serie en el instante  $t = 96$ . La distribución de los residuos no tiene distribución normal, y su curtosis es 5.1723, lo que lleva a rechazar cualquier contraste de normalidad. Los correlogramas simple y parcial de los residuos muestran sin embargo que todos los residuos están dentro de las bandas, véase la figura 11.5, pero el estadístico  $Q(24)$  es 39.27, que debe compararse con una  $\chi^2$  de 20 grados de libertad, por lo que se rechaza la hipótesis de que la serie es ruido blanco. La presencia de los destacados valores atípicos sugiere investigar estos efectos antes de probar modelos más complejos para esta serie.

---

## 11.4. Contraste de estabilidad del modelo

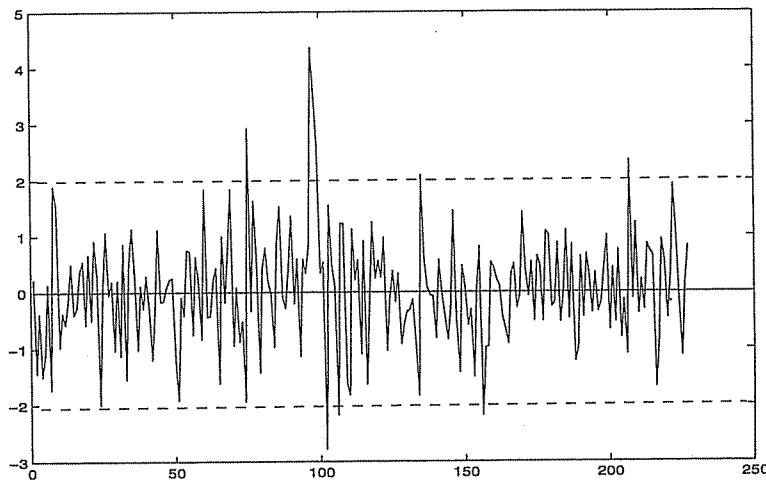
Si el modelo es adecuado, los errores de predicción para un periodo por delante son variables aleatorias normales de media cero, varianza constante e independientes. En consecuencia, si hemos estimado el modelo con datos hasta  $t = T$  y a continuación generamos predicciones que vamos readaptando al obtener los valores observados, y llamamos  $\hat{a}_{T+1}, \dots, \hat{a}_{T+h}$  los errores de predicción de un periodo adelante, la variable

$$Q = \sum_{j=1}^h \hat{a}_{T+j}^2 / \sigma^2$$

será una  $\chi^2$  con  $h$  grados de libertad. Como  $\sigma^2$  se estima mediante  $\hat{\sigma}^2$ , la variable

$$\frac{Q}{h} = \frac{\sum_{j=1}^h \hat{a}_{T+j}^2 / h}{\hat{\sigma}^2} \quad (11.8)$$

**Figura 11.4 Residuos del modelo para la serie de accidentes en logaritmos. El modelo es ARIMA(2,1,0)x  
ARIMA(0,1,2)<sub>12</sub>**



será una  $F$  con  $h$  y  $T - p - q$  grados de libertad, siendo  $T$  el número de datos iniciales y  $p + q$  el número de parámetros estimados.

En consecuencia, si  $Q/h$  es significativamente grande, esto sugiere que el modelo no es adecuado.

## 11.5. Interpretación del modelo. Componentes deterministas

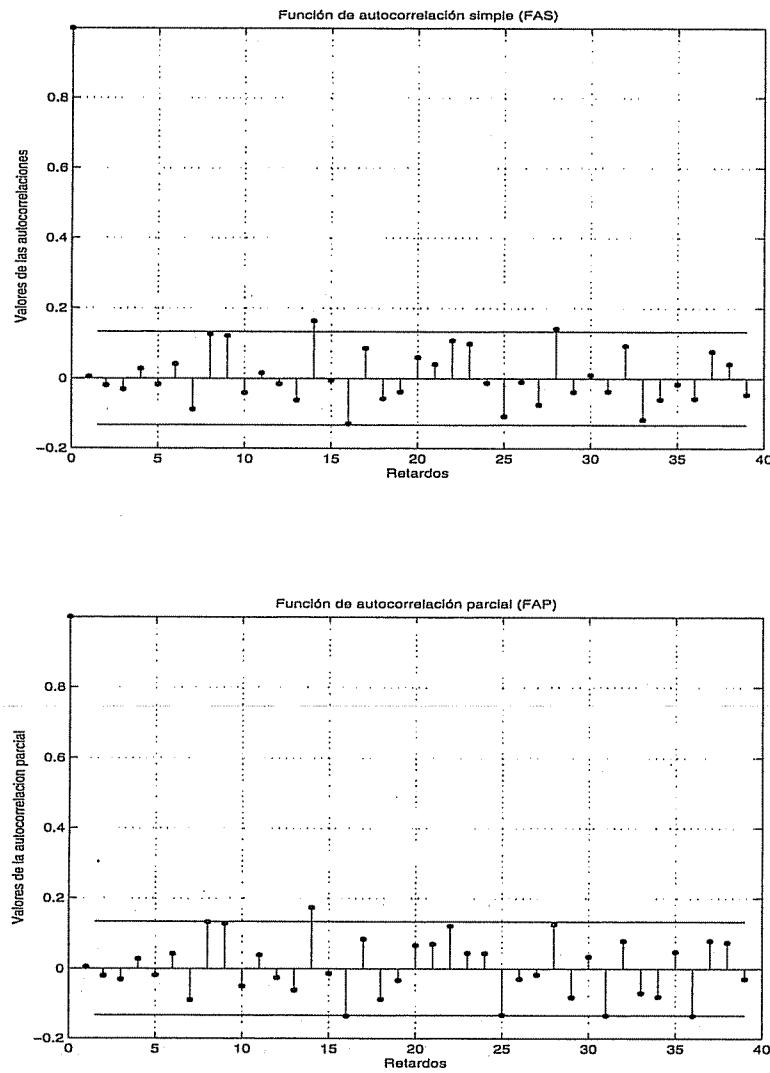
Hemos visto que los modelos ARIMA engloban como caso particular modelos con componentes deterministas, como tendencias polinómicas y estacionalidad constante. Si el modelo estimado contiene diferencias, regulares o estacionales, y estructura MA regular o estacional con alguna raíz próxima al círculo unidad, conviene investigar la posibilidad de cancelaciones de raíces que indiquen componentes deterministas. Vamos a analizar este aspecto.

### 11.5.1. Tendencias

Un caso importante de cancelación de operadores es el producido entre una diferencia regular y una media móvil regular no invertible. Por ejemplo, el modelo:

$$\nabla z_t = (1 - \theta B) a_t, \quad (11.9)$$

Figura 11.5 Fas y fap de los residuos del modelo de la serie de accidentes en logaritmos



cuando  $\theta \simeq 1$  es equivalente al modelo determinista:

$$z_t = \beta_0 + a_t, \quad (11.10)$$

donde  $\beta_0$  es una constante. La comprobación es inmediata, tomando diferencias en (11.10), ya que  $\nabla \beta_0 = 0$ . En este caso, la aparición de una raíz

## 11. Diagnosis del modelo y predicción

unitaria en la parte MA es debida a una sobrediferenciación del proceso (11.10) que no requiere diferencias de ningún tipo. Al cancelar las diferencias aparece un término determinista, que es simplemente la constante del modelo.

Análogamente, el modelo:

$$\nabla z_t = \beta_1 + (1 - \theta B) a_t \quad \text{con } \theta \simeq 1 \quad (11.11)$$

se obtiene al diferenciar el modelo con tendencia lineal determinista:

$$z_t = \beta_0 + \beta_1 t + a_t. \quad (11.12)$$

En efecto, al tomar diferencias en (11.12) resulta:

$$\nabla z_t = \beta_1 + \nabla a_t,$$

ya que  $\nabla \beta_1 t = \beta_1 t - \beta_1(t-1) = \beta_1$ . Esta propiedad es equivalente a que en (11.11) cuando  $\theta \simeq 1$  la cancelación de las diferencias conduce a (11.12) que tiene una tendencia lineal determinista.

En general, el modelo:

$$\nabla^d z_t = (1 - B)^d a_t, \quad (11.13)$$

es equivalente al modelo con tendencia determinista polinómica:

$$z_t = \beta_0 + \beta_1 t + \dots + \beta_d t^{d-1} + a_t. \quad (11.14)$$

Para demostrar este resultado basta aplicar el operador diferencia  $d$  veces en (11.14). Como caso particular, si  $d = 1$  tenemos el modelo (11.9), y si  $d = 2$  un modelo con tendencia lineal (11.13).

### 11.5.2. Estacionalidad

La cancelación de operadores estacionales produce un resultado similar a la cancelación de los operadores regulares, pero la interpretación es diferente al estar ahora las constantes ligadas a los componentes estacionales. El modelo que resulta de cancelar las diferencias estacionales en:

$$(1 - B^{12}) z_t = (1 - B^{12}) a_t \quad (11.15)$$

es

$$z_t = S_t^{(12)} + a_t, \quad (11.16)$$

donde  $S_t^{(12)}$  es cualquier función periódica cada 12 meses, que verifica

$$S_t^{(12)} = S_{t-12}^{(12)}, \quad \forall t$$

Por ejemplo, si modelamos una serie mensual de temperatura podemos obtener que sigue el modelo periódico:

$$z_t = \cos\left(\frac{2\pi}{12}t + \delta\right) + a_t \quad (11.17)$$

y tomando una diferencia estacional se reduce al (11.15). El ajuste de este modelo sinusoidal se realiza como estudiamos en el capítulo 2. También tendremos cancelación de diferencias estacionales si una serie mensual tiene una estacionalidad determinista cualquiera, que podemos representar mediante variables ficticias del tipo

$$z_t = \sum_{j=1}^{12} w_j I_t^j + a_t, \quad (11.18)$$

donde  $I_t^j = 1$  si  $t = j + 12$  y cero en otro caso. Al tomar una diferencia estacional en (11.18) se obtiene de nuevo el modelo (11.15).

La conclusión de este ejercicio es que cuando tengamos una diferencia estacional y un término de media móvil estacional con parámetro muy próximo a la unidad, conviene explorar la posibilidad de modelar la estacionalidad mediante términos deterministas, como (11.18) y (11.17), y comprobar si con este método mejoramos el ajuste del modelo.

### 11.5.3. Caso general

Cuando el modelo contiene diferencias estacionales y regulares y términos de media móvil de ambos tipos podemos tener simultáneamente tendencias y estacionalidad determinista. Por ejemplo, el modelo de pasajeros de avión con coeficientes próximos a la unidad

$$\nabla \nabla_{12} z_t = (1 - \theta B)(1 - \Theta B)a_t, \quad \theta \approx \Theta \approx 1 \quad (11.19)$$

es equivalente al modelo

$$z_t = \beta_0 + \beta_1 t + S_t^{(12)} + a_t, \quad (11.20)$$

que contiene una tendencia lineal determinista y una estacionalidad también determinista. La comprobación es inmediata tomando diferencias en (11.20), con lo que se obtiene (11.19).

Como conclusión, siempre que el modelo contenga diferencias la aparición de un componente MA(1) regular o estacional con parámetro muy

próximo a uno puede indicar un componente determinista. Un coeficiente próximo a uno de una media móvil regular sugiere una tendencia determinista, mientras que en una media móvil estacional sugiere una estacionalidad constante. Conviene ajustar ambos tipos de modelos y seleccionar entre ellos con el criterio BIC. Observemos que estas propiedades son coincidentes con la forma de la ecuación de previsión de un modelo ARIMA estudiada en el capítulo 8.

---

### Ejemplo 11.3

Vamos a analizar la temperatura máxima diaria de Melbourne, Australia. Los datos se encuentran en el fichero melbtemp.dat y se presentan en la figura 11.6. La serie es muy estacional, como sería esperable, y la variabilidad parece ser mucho mayor en el verano, cuando la temperatura es alta, que en el invierno. Para contrastar este aspecto la figura 11.7 muestra un gráfico entre la media y la desviación típica en grupos de 10 observaciones. Se observa una clara relación lineal, por lo que modelaremos la temperatura en logaritmos. Al tomar una diferencia estacional en la serie en logaritmos se obtienen las funciones *fas* y *fap* indicadas en la figura 11.8. Hay dos coeficientes distintos de cero en los parciales y un coeficiente alto en 12 para la *fas* y en la *fap* en 12, 24, 36 con valores decrecientes, lo que sugiere un MA(1) estacional. La estimación del modelo AR(2)×MA(1)<sub>12</sub> condujo a unos residuos con estructura de autocorrelación marcada, por lo que se introdujo parte AR más larga. Como muchos coeficientes se estimaron como significativos, el modelo se completó con un término MA. El modelo final es:

$$(1 - 1.46B + .615B^2 - .094B^3 - .045B^4)\nabla_{12} \log TMD_t = \\ = (1 - .887B)(1 - .999B^{12})a_t$$

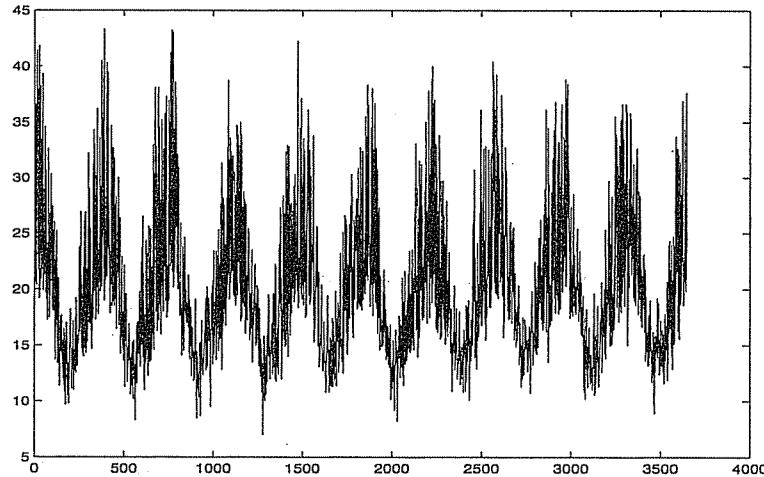
El alto coeficiente de la media móvil estacional sugiere que la estacionalidad puede explicarse bien por un componente determinista. Entonces el modelo puede escribirse como:

$$(1 - 1.46B + .615B^2 - .094B^3 - .045B^4)(\log TMD_t - S_t) = \\ = (1 - .887B)a_t$$

donde  $S_t$ , la estacionalidad, es determinista y sigue una onda cíclica. Este término se estima como vimos en el capítulo 2.

---

**Figura 11.6 Temperatura máxima diaria de Melbourne durante diez años**



## 11.6. Predicción

### 11.6.1. Predicciones puntuales

Las predicciones del modelo estimado pueden realizarse como en el capítulo 8, utilizando los parámetros estimados como si fuesen los verdaderos. Además, si el modelo es correcto los parámetros estimados, que son los que minimizan el error de predicción a un paso, son además los que minimizan el error de predicción a cualquier horizonte. Sin embargo, si el modelo no es correcto esta propiedad no es necesariamente cierta.

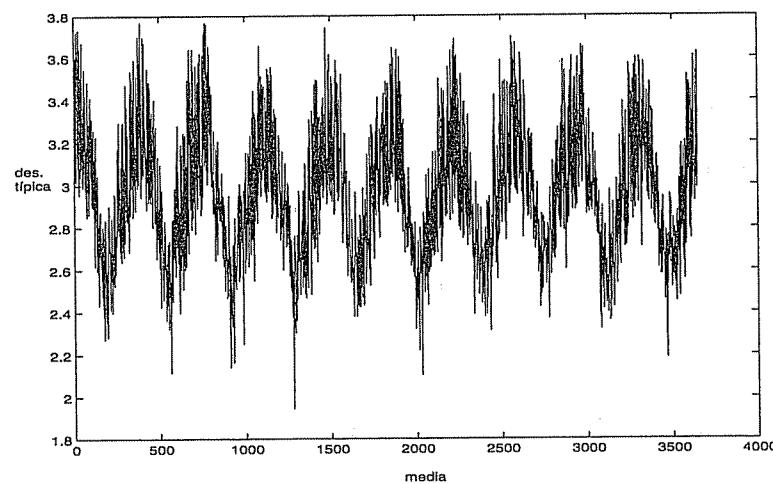
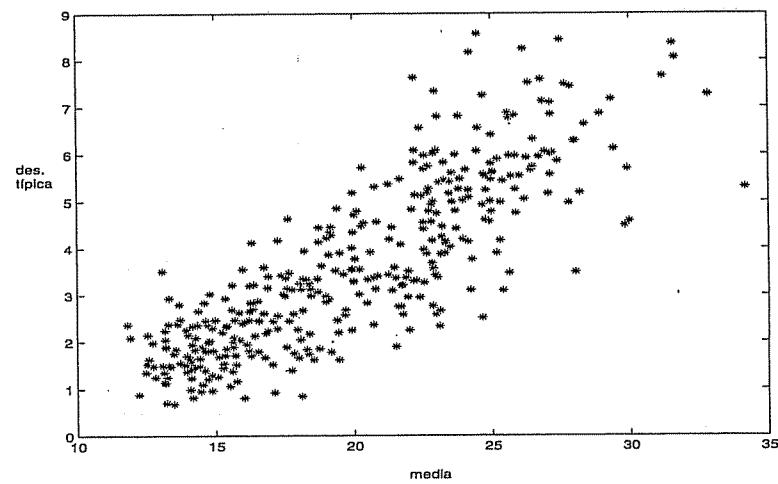
Para ilustrar este aspecto, supongamos que una serie se ha generado por un AR(2) pero se ha estimado, erróneamente, un AR(1). Si utilizamos mínimos cuadrados el parámetro se estima con:

$$\min \sum (z_t - \alpha z_{t-1})^2,$$

y el resultado es  $\hat{\alpha} = r_1$ , siendo  $r_1$  el coeficiente de autocorrelación de primer orden. Supongamos que nos interesa la predicción a dos pasos. Como suponemos que los datos se han generado por un AR(1) el predictor óptimo es, según vimos en el capítulo 8, la esperanza condicionada,  $\phi^2 z_{t-1}$ , que estimaremos con  $\hat{\alpha}^2 z_{t-1}$ . Alternativamente, podríamos obtener el coeficiente  $\beta$  de un predictor AR(1) que minimiza el error cuadrático de predicción a dos pasos, minimizando:

## 11. Diagnosis del modelo y predicción

Figura 11.7 Relación entre la desviación típica y la media para la serie de las temperaturas diarias en Melbourne y gráfico de la serie en logaritmos



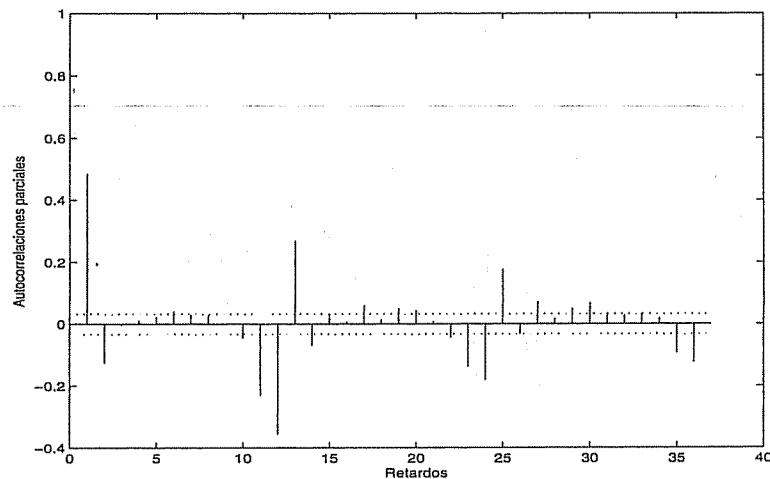
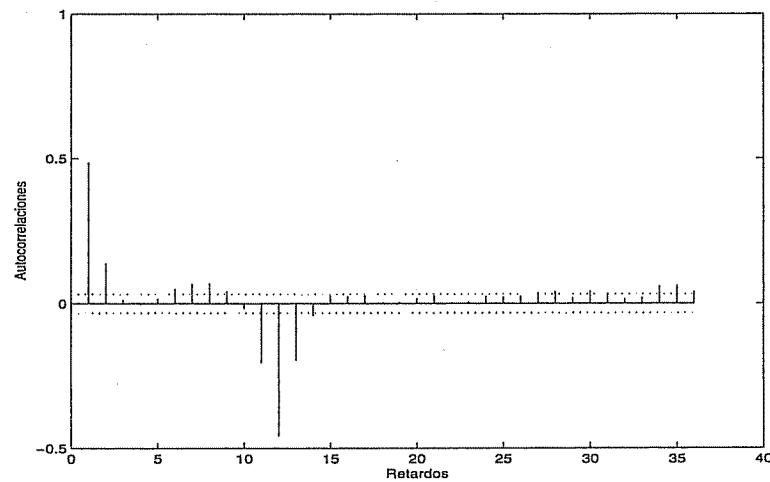
$$\min \sum (z_t - \beta z_{t-2})^2.$$

Resolviendo esta ecuación, derivando e igualando a cero, se obtiene inmediatamente que  $\hat{\beta} = r_2$ , el coeficiente de autocorrelación de orden dos.

## Análisis de series temporales

---

**Figura 11.8 Fas y fap estimadas con la serie de las temperaturas diarias en Melbourne**



Si los datos se hubiesen generado con un AR(1) como los valores teóricos verificarían  $\rho_2 = \rho_1^2 = \phi^2$ , el predictor de  $z_{t+2}$  desde  $z_{t-1}$ , suponiendo un AR(1),  $\hat{\alpha}^2 z_{t-2}$ , coindice aproximadamente para muestras grandes con  $r_2 z_{t-2}$ . Sin embargo, como el verdadero modelo es el AR(2), ambos predictores serán muy diferentes y el segundo, que se obtiene directamente, tiene un menor error cuadrático medio de predicción.

## 11. Diagnosis del modelo y predicción

En resumen, si el modelo es correcto, la previsión a cualquier horizonte se obtiene con los parámetros estimados a un horizonte. Sin embargo, si el modelo no es correcto, podemos mejorar las predicciones estimando los parámetros para cada horizonte. Esta idea ha sido investigada por algunos autores (véase Tiao y Xu, 1993). Si el modelo es correcto, los parámetros obtenidos al estimar a distintos horizontes serán aproximadamente iguales, pero si el modelo es incorrecto, obtendremos distintos parámetros para diferentes horizontes.

### 11.6.2. Intervalos de predicción

Los intervalos de predicción que estudiamos en el capítulo 8 se calcularon suponiendo los parámetros conocidos y teniendo sólo en cuenta la incertidumbre debida al desconocimiento de las innovaciones futuras. En la práctica existen cuatro tipos de incertidumbre en la predicción, asociadas al desconocimiento de:

1. Las innovaciones futuras.
2. La distribución de las innovaciones.
3. Los verdaderos valores de los parámetros.
4. El modelo que ha generado los datos.

La primera fuente de incertidumbre es inevitable y no depende del tamaño muestral. Como los valores futuros de la serie dependen de innovaciones futuras desconocidas, tendremos una incertidumbre creciente con el horizonte de predicción. Ya vimos que esta incertidumbre es acomodada para procesos estacionarios, pero crece sin límite para procesos no estacionarios.

La importancia de las tres restantes fuentes de incertidumbre depende del tamaño muestral. En general, su efecto es pequeño para muestras grandes (series largas), pero puede ser importante en muestras pequeñas (menores de 50 datos). Si tenemos una serie larga podemos hacer un contraste fiable de que la distribución de las innovaciones es normal y, si la rechazamos, estimar y utilizar la distribución real de las innovaciones a partir de los residuos. Con series cortas, la potencia de un contraste de normalidad es baja, con lo que existe siempre más incertidumbre respecto a la distribución de las innovaciones. El efecto del error en la estimación de los parámetros aumenta también con series cortas, ya que el error de estimación disminuye con el tamaño muestral. Finalmente, con series pequeñas es difícil elegir entre modelos similares, ya que los intervalos de confianza de los parámetros son amplios. Por ejemplo, si los datos

se han generado con  $(1 - .7B - .15B^2)z_t = a_t$  será difícil elegir con una muestra de  $T = 50$  entre un AR(1) y un AR(2), ya que el error de estimación de los parámetros es del orden  $1/\sqrt{T} = 1/\sqrt{50} = .14$ , de tamaño similar al parámetro, con lo que el contraste habitual indicará que este parámetro no es significativo. Con tamaños muestrales pequeños o medianos existirán habitualmente varios modelos compatibles con la serie observada, y no podemos ignorar que el modelo seleccionado puede ser erróneo.

#### 11.6.3. Intervalos de predicción para muestras grandes

Si el tamaño muestral es grande, la incertidumbre dominante es la debida a las innovaciones y las otras tres pueden despreciarse. Suponiendo que la distribución de las innovaciones es normal, se justifica en el apéndice 11.1 que la incertidumbre respecto a los parámetros será pequeña. Análogamente, como veremos en la sección siguiente, cuando el tamaño muestral es grande la incertidumbre respecto al modelo será también pequeña.

Suponiendo normalidad, podemos construir intervalos de confianza en muestras grandes tomando los estimadores como parámetros y calculando el intervalo como en el caso en que los parámetros son conocidos. Por ejemplo, para el 95 % el intervalo será:

$$z_{T+k} \in \hat{z}_T(k) \pm 1,96 \hat{\sigma} \left( 1 + \hat{\psi}_1^2 + \dots + \hat{\psi}_{k-1}^2 \right)^{1/2}.$$


---

#### Ejemplo 11.4

La figura 11.9 muestra las predicciones generadas para la serie de accidentes en la hipótesis de muestras grandes y suponiendo distribuciones normales. Se observa cómo los intervalos de confianza crecen al aumentar el horizonte de predicción.

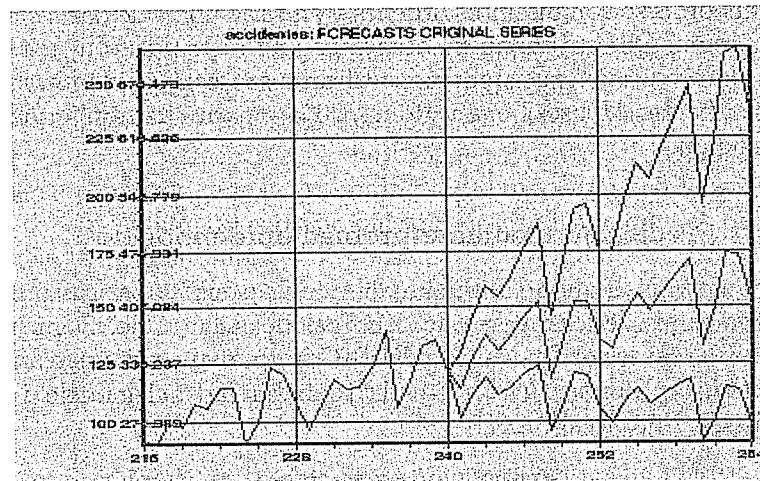
---

#### 11.6.4. Intervalos mediante remuestro

Si las innovaciones no son normales o si el tamaño muestral no es muy grande y no queremos despreciar al incertidumbre debida a la estimación de los parámetros, podemos utilizar técnicas de remuestro (*bootstrap*) para generar las predicciones y construir a partir de ellas los intervalos de confianza. Un procedimiento simple para utilizar la verdadera distribución de las innovaciones y tener en cuenta la incertidumbre por la estimación de los parámetros es el siguiente (véase Tombs y Shucany, 1993 y Pascual, Romo y Ruiz, 2004):

## 11. Diagnosis del modelo y predicción

Figura 11.9 Predicciones con la serie de accidentes y el modelo  $(2,1,0)X(1,1,1)_{12}$



1. Estimar el modelo y obtener los parámetros,  $\hat{\phi}$  y  $\hat{\theta}$  y los residuos.
2. Obtener una muestra con reemplazamiento de  $T$  residuos de la distribución de los residuos estimados. Sean  $a_1^*, \dots, a_T^*$  la muestra obtenida.
3. Reconstruir una nueva serie con los parámetros estimados y estos residuos mediante:

$$z_t^* = \hat{\phi}_1 z_{t-1}^* + \dots + \hat{\phi}_h z_{t-h}^* + a_t^* - \hat{\theta}_1^* a_{t-1}^* - \dots - \hat{\theta}_q^* a_{t-q}^*,$$

tomando como valores iniciales  $z_1 = \dots = z_h = \bar{z}$  y  $a_{-q} = \dots = a_0 = 0$ .

4. Estimar los parámetros en la serie anterior. Sean  $\hat{\phi}^*$  y  $\hat{\theta}^*$  los nuevos parámetros estimados.
5. Calcular predicciones de valores futuros de la serie,  $\hat{z}_T(k)$  utilizando los parámetros estimados en la etapa anterior,  $\hat{\phi}^*$  y  $\hat{\theta}^*$ , pero con valores de  $z$  iguales a los últimos verdaderos valores observados  $z_{T-1}, \dots, z_{T-h}$  e innovaciones iguales a las reales que hemos estimado para la serie  $a_{t-1}, \dots, a_{t-q}$ .
6. Repetir los pasos 2 a 5 un número grande de veces,  $N$ . Obtener los intervalos de confianza de la distribución de  $\hat{z}_T(k)$  a partir de los  $N$  valores obtenidos.

Este enfoque puede generalizarse para tener en cuenta también la incertidumbre en la estimación del modelo. Si podemos encontrar una distribución para los posibles órdenes del modelo, podemos generar realizaciones del proceso con distintos órdenes y en cada una calcular los errores de predicción. De esta manera obtenemos una distribución generada de errores de predicción que podemos utilizar para construir intervalos de confianza (véase Alonso, Peña y Romo, 2004a, para una implantación de este enfoque).

#### 11.6.5. Predicción mediante promedio de modelos

Supongamos que se han estimado  $k$  modelos posibles para una serie y que disponemos de los valores del BIC para cada modelo. Según vimos en el capítulo anterior, los valores del BIC son, salvo constantes aditivas,  $-2 \ln P(M_i|D)$ , donde  $P(M_i|D)$  es la probabilidad del modelo  $i$  dados los datos  $D$ . Entonces, como

$$BIC_i = -2 \ln P(M_i|D) + c$$

tenemos que

$$P(M_i|D) = c_1 e^{-\frac{1}{2} BIC_i},$$

donde  $c_1 = \exp(c)$ . Para calcular esta constante,  $c_1$ , utilizamos que la suma de las probabilidades de todos los modelos debe ser la unidad. Por tanto:

$$\sum_{j=1}^k P(M_j|D) = 1 = c_1 \sum_{j=1}^k e^{-\frac{1}{2} BIC_j},$$

y podemos transformar los valores del BIC en probabilidades a posteriori de los distintos modelos mediante:

$$P(M_i|D) = \frac{e^{-\frac{1}{2} BIC_i}}{\sum_{j=1}^k e^{-\frac{1}{2} BIC_j}}.$$

La distribución de probabilidad de una nueva observación será una distribución mezclada:

$$\sum P(M_i|D) f(z|M_i) \quad (11.21)$$

donde  $f(z|M_i)$  es la función de densidad de la nueva observación de acuerdo con el modelo  $M_i$ . Esta distribución será normal, con media igual a la predicción con dicho modelo y varianza igual a la de esta predicción. Por ejemplo, para un periodo adelante si generamos predicciones con cada modelo y llamamos  $\hat{z}_T^{(i)}(1)$  a la predicción para el periodo  $T + 1$  con el modelo  $i$ , la esperanza de la distribución (11.21), que llamaremos

---

## 11. Diagnosis del modelo y predicción

$\widehat{z}_T(1)$  será:

$$\widehat{z}_T(1) = \sum_{i=1}^k \widehat{z}_T^{(i)}(1) P(M_i|D),$$

que es el resultado de ponderar las predicciones de todos los modelos por sus probabilidades para obtener una predicción agregada única. Esta forma de calcular las predicciones se conoce como *promedio Bayesiano de modelos* (*Bayesian Model Averaging*). En general esta predicción será más precisa, en promedio, que la generada por un modelo único. Además permitirá construir intervalos de predicción más realistas que los obtenidos ignorando la incertidumbre de los parámetros y del modelo. Llamando  $\widehat{\sigma}_i^2$  a la varianza de las innovaciones del modelo  $i$ , que coincidirá con el error de predicción a un paso con este modelo, la varianza de la combinación de modelos será:

$$var(\widehat{z}_T(1)) = E [(z_{T+1} - \widehat{z}_T(1))^2],$$

donde esta esperanza se calcula respecto a la distribución (11.21). Puede demostrarse (véase ejercicio 11.3) que la varianza de la distribución mezclada es:

$$var(\widehat{z}_T(1)) = \sum_{i=1}^k \widehat{\sigma}_i^2 P(M_i|D) + \sum_{i=1}^k (\widehat{z}_T^{(i)}(1) - \widehat{z}_T(1))^2 P(M_i|D),$$

que permitirá construir intervalos más realistas para las predicciones. Estos resultados se generalizan para cualquier intervalo de predicción.

---

### Ejemplo 11.5

Vamos a ver cómo se combinarían las predicciones para la serie de matriculación de vehículos. Supongamos que los tres modelos posibles son los indicados en la tabla 11.1, que se estimaron en el capítulo anterior. A partir de los datos del BIC podemos calcular las probabilidades a posteriori de cada modelo mediante:

$$p(M_1|D) = \frac{\exp(-\frac{1}{2}8.08)}{\exp(-\frac{1}{2}8.08) + \exp(-\frac{1}{2}14.14) + \exp(-\frac{1}{2}14.22)} = .913$$

y análogamente para los otros. En este caso la incertidumbre respecto al modelo es pequeña y las predicciones obtenidas mediante la combinación serán similares a las del mejor modelo. Sin embargo, los intervalos de predicción serán más grandes, debido a la incertidumbre adicional entre los modelos.

---

**Tabla 11.1 Valores de los criterios de selección con la serie de matriculación**

Modelo	BIC	P(M/Datos)
ARIMA(0,1,1)×(0,1,1) <sub>12</sub>	8.08	.913
ARIMA(0,1,1)×(1,1,1) <sub>12</sub>	14.14	.044
ARIMA(0,1,1)×(2,1,0) <sub>12</sub>	14.22	.043

### Ejercicios 11

- 11.1 Ajustar un modelo a los datos del paro en España y analizar los residuos para comprobar que son adecuados.
  - 11.2 Demostrar que la media de la distribución  $\alpha f(\mu_1, \sigma_1^2) + (1 - \alpha)f(\mu_2, \sigma_2^2)$  es  $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$  y la varianza  $\alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 + \alpha(\mu_1 - \mu)^2 + (1 - \alpha)(\mu_2 - \mu)^2$ .
  - 11.3 Utilizar el resultado anterior para demostrar la fórmula de la media y la varianza cuando combinamos predicciones con distintos modelos.
  - 11.4 Demostrar que si el modelo más probable es el de menor varianza residual, los intervalos construidos por el promedio de modelos serán más amplios que los de un modelo único.
- 

### Apéndice 11.1: Predicción con incertidumbre de los parámetros y del modelo

Para ilustrar el error debido a la estimación de los parámetros, supongamos que hemos estimado un AR(1) de media cero y generamos predicciones con:

$$\hat{z}_T(1) = \hat{\phi}z_T,$$

donde suponemos que el parámetro  $\phi$  se estima por mínimos cuadrados. Entonces,

$$\hat{\phi} = \sum_{t=2}^T z_t z_{t-1} / \sum_{t=2}^T z_{t-1}^2$$

y la varianza de la estimación es  $var(\hat{\phi}) = \sigma^2 / \sum_{t=2}^T z_{t-1}^2$ . Llamando  $e_T(1) = a_{T+1}$ , tenemos:

$$e_T^*(1) = z_{T+1} - \hat{\phi}z_T = a_{T+1} + (\phi - \hat{\phi})z_T.$$

## 11. Diagnosis del modelo y predicción

Elevando al cuadrado esta expresión y tomando esperanzas respecto a la distribución conjunta de las observaciones, se obtiene:

$$E(e_T^*(1)^2) = E(a_{T+1}^2) + E[(\phi - \hat{\phi})^2 z_T^2],$$

ya que el doble producto se anula al ser  $a_{T+1}$  independiente de los valores muestrales observados  $y$ , por tanto, independiente de  $\hat{\phi}$ , que depende de los valores pasados, y de  $z_T$ . Para calcular la esperanza del segundo término utilizaremos la ley de las esperanzas iteradas del Apéndice 8.1. Esta esperanza respecto a la distribución conjunta de los datos puede calcularse en dos etapas. Primero, calcularemos la esperanza respecto a la distribución de  $(z_1, \dots, z_{T-1}|z_T)$ , es decir condicionada al valor  $z_T$ , y segundo, la esperanza del resultado respecto a la distribución de  $z_T$ . De esta manera, en la primera esperanza  $z_T$  se toma como constante y se considera la variabilidad debida a la estimación de  $\phi$  en distintas realizaciones del proceso que producen distintos valores del estimador. Al condicionar en  $z_T$ :

$$E[(\phi - \hat{\phi})^2 z_T^2 | z_T] = z_T^2 E(\phi - \hat{\phi})^2 = z_T^2 var(\hat{\phi})$$

donde hemos supuesto que  $E[(\phi - \hat{\phi})^2 | z_T] \approx E(\phi - \hat{\phi})^2$ , la varianza del estimador. Sustituyendo la expresión de la varianza,  $var(\hat{\phi}) = \sigma^2 / T s_z^2$ , donde  $s_z^2 = \sum_{t=2}^T z_{t-1}^2 / T$ , tenemos que

$$var(e_T^*(1) | z_T) = \sigma^2 (1 + z_T^2 / T s_z^2).$$

Ahora tenemos que tomar la esperanza de esta expresión respecto  $z_T$ . Como el cociente  $z_T^2 / s_z^2$  tiene esperanza uno, obtenemos que  $var(e_T^*(1)) = \sigma^2 (1 + 1/T)$ . Si  $T$  es grande podemos despreciar el segundo término frente al primero, lo que implica despreciar la incertidumbre asociada a la estimación de los parámetros del modelo. En general, en muestras grandes la incertidumbre adicional debido a la estimación de los parámetros es despreciable.

El error adicional debido a la incertidumbre en el modelo puede ser grande. Supongamos que hemos decidido que el modelo es AR(1), pero el verdadero modelo es AR( $p$ ). El error real de predicción será:

$$e_T^p(1) = z_{T+1} - \hat{\phi} z_T = a_{T+1} + (\phi_1 - \hat{\phi}) z_T + \dots + \phi_p z_{T-p}. \quad (11.22)$$

La varianza del error se obtiene elevando al cuadrado en (11.22) y tomando esperanzas condicionadas a los valores  $z_T, \dots, z_{T-p}$ , con lo que resulta:

$$\begin{aligned} var(e_T^p(1) | z_T, \dots, z_{T-p}) &= \sigma^2 (1 + z_T^2 / T s_z^2) + \\ &+ \sum \phi_j^2 z_{T-j}^2 + 2 \sum \phi_j \phi_i z_{T-i} z_{T-j} \end{aligned}$$

y vemos que la varianza del error de predicción puede aumentar bastante, en función de los valores de los parámetros.

La incertidumbre del modelo puede ser importante en muestras pequeñas. En muestras grandes si la probabilidad del modelo finalmente seleccionado con el criterio BIC es alta (próxima a la unidad), es claro que sus predicciones serán similares a las obtenidas combinando modelos con el procedimiento estudiado en la sección 11.6: al tener los otros modelos probabilidades muy bajas afectarán muy poco a la predicción final.

## Apéndice 11.2: Programas para la diagnosis

La diagnosis se realiza con Minitab y SPSS fácilmente, guardando los residuos del modelo como una nueva variable y analizando su gráfico y su *fas* y *fap* muestral. Ambos programas incluyen el estadístico Ljung-Box sobre las correlaciones.

En EViews la diagnosis se hace con: *Actual, fitted, residual* → *Actual, fitted, residual graph* y a continuación: *Residual tests* → *Correlogram Q - Statistic*, que proporciona la función de autocorrelación y de autocorrelación parcial de los residuos y el estadístico de Ljung-Box. Podemos también ver un histograma de los residuos en: *Residual tests* → *Histogram (Normality test)*.

Para generar predicciones hacemos: *Procs* → *Especify/Estimate*, marcamos el modelo y seleccionamos en el menú *forecast*:

<i>Forecast</i> →	Forecast of                                   x Forecast name                                xf S.E.    errores Sample range to forecast    101 110
-------------------	--

y obtenemos un gráfico con los valores previstos para los próximos 10 datos. En el workfile aparecerán dos nuevas series, xf y errores. La serie xf corresponde a la serie x, para los datos del 1 al 100, y la predicción, para los datos 101 a 110. En el gráfico se comprueba que la predicción sigue el recorrido decreciente de los últimos datos y que los errores crecen.

Como ejemplo, vamos a obtener predicciones con la serie de pasajeros de avión (airline.dat). La estimación se hará con: *Quick* → *Estimate Equation*: *d(lx,1,12) ma(1) sma(12)*. Podemos obtener predicciones para los dos últimos años con:

<i>Forecast</i> →	Forecast of                                   lx Forecast name                                lxf S.E.    errores Sample range to forecast    1961:01 1962:12
-------------------	--

## 11. Diagnosis del modelo y predicción

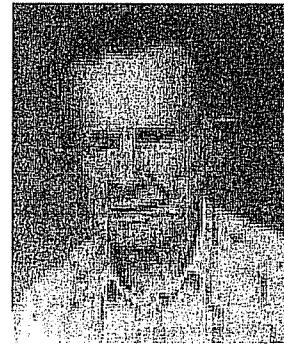
Debemos ver un gráfico con los valores predichos para los próximos dos años. En el workfile aparecerán dos nuevas series, lxf y errores. La serie lxf corresponde a la serie lx para los datos del 1941:1 al 1960:12 y la predicción para los datos 1961:01 1962:12. Podemos ver el gráfico para comprobar que la predicción sigue el recorrido anterior de los datos. Vemos en la serie errores que son cada vez más grandes.

La diagnosis con el programa TSW está incluida en el *output* de la estimación y el programa permite obtener gráficos de los residuos y de su correlograma.

Un programa en *matlab* para aplicar el contraste del determinante puede bajarse de la dirección [http://halweb.uc3m.es/daniel\\_pena](http://halweb.uc3m.es/daniel_pena).

---

## 12. Análisis de intervención



Carl Harald Cramer (1893-1985)

Estadístico sueco. Fue el primer catedrático de Estadística Matemática y Ciencia Actuarial en la Universidad de Estocolmo en 1929 y durante cerca de cuarenta años trabajó como consultor para varias compañías de seguros suecas. Hizo contribuciones fundamentales a la teoría de los procesos estocásticos estacionarios y publicó en 1945 un libro de Estadística Matemática que ha sido una de las referencias básicas durante más de treinta años. Fue presidente de la Universidad de Estocolmo desde 1950 hasta su retiro en 1961.

### 12.1. Introducción

Las series reales se ven con frecuencia afectadas por sucesos puntuales conocidos. Por ejemplo, una huelga, un año bisiesto, un cambio legal, un accidente o un cambio en una festividad. Si modelamos estos efectos en la serie podemos mejorar la precisión de la estimación de los parámetros y de las previsiones.

Un tipo de suceso conocido que ya hemos estudiado es la estacionalidad. Cuando la estacionalidad es debida a una causa sistemática, como sucede en algunas series climatológicas, produce un efecto determinista de periodo  $s$ . Por ejemplo, como hemos visto en los capítulos 2 y 11, una estacionalidad mensual,  $s = 12$ , puede modelarse con  $s - 1$  variables impulso,  $I_t^{(j)}$ , que tomen el valor uno en un mes y cero en el resto. Si el modelo tiene constante, que equivale a una variable que toma el mismo valor en todos los períodos, no podemos introducir  $s$  variables ficticias porque como su suma es siempre constante (uno) el modelo no podría estimarse al existir multicolinealidad exacta. Una estacionalidad determinista de periodo  $s$  se modela como:

$$z_t = \mu + \sum_{j=1}^{s-1} w_j I_t^{(j)} + n_t,$$

y en el caso particular que el proceso  $n_t$  sea ruido blanco resulta un modelo de regresión simple que se estima por mínimos cuadrados. Este ejemplo ilustra que la forma de modelar efectos deterministas es introducirlos como variables explicativas en el modelo. Por ejemplo, supongamos que en una serie de producción conocemos que se ha producido una huelga en un momento determinado. Podemos incluir este efecto construyendo una variable ficticia, que llamaremos variable impulso,  $I_t$ , que tome el valor cero en el instante de la huelga y cero todos los restantes instantes. El modelo será:

$$z_t = \mu + wI_t + n_t,$$

donde ahora  $n_t$  puede tener estructura temporal y seguir un modelo ARIMA. Si estudiamos la relación entre los valores de la serie y esta variable impulso, podemos estimar el efecto que la huelga ha tenido en la serie. Como segundo ejemplo, supongamos que en una serie de paro se produce un cambio legal, que modifica la definición de parado a partir de un instante. Podemos representar este efecto mediante una variable que tome el valor cero antes del cambio y uno después. Esta variable se denomina escalón,  $S_t$ , y podemos introducirla en el modelo como:

$$z_t = \mu + wS_t + n_t$$

donde, de nuevo,  $n_t$  sigue un modelo ARIMA. En este capítulo veremos cómo modelar y estimar estos efectos.

En otras situaciones queremos tener en cuenta sucesos que podemos representar mediante una serie conocida, que toma valores distintos en el tiempo. Por ejemplo, si el número de días laborables en el mes es una variable importante, como ocurre en una serie de producción mensual, podríamos tener en cuenta este hecho dividiendo las observaciones por el número de días laborables dentro de cada mes y construir una serie de producción media diaria durante el mes. Una solución mejor es introducir dentro del modelo una variable explicativa que incluya los días laborables de cada mes. Este segundo enfoque tiene la ventaja de que si los días de la semana tienen distinta actividad (por ejemplo, no son iguales las ventas en un supermercado los lunes que los viernes), podemos investigar este efecto incluyendo siete variables explicativas que tengan en cuenta el número de lunes, martes, etc., que existen en el mes. También, muchas series mensuales económicas y sociales están afectadas por festividades que ocurren de manera irregular, como la Semana Santa. Podemos medir

este efecto introduciendo en el modelo una variable impulso, que toma el valor uno en el mes del año donde aparece la mayor parte de la Semana Santa y cero en el resto de los meses. Una alternativa más sofisticada es incluir una variable cuantitativa, que tome en cada mes del año el número de días de Semana Santa que hay en ese mes (esta variable será cero en todos los meses salvo en marzo y abril).

Box y Tiao (1975) denominaron análisis de intervención a la inclusión en un modelo de series temporales de variables ficticias para representar sucesos que producen efectos deterministas. Estudiaremos primero estas variables y sus propiedades y después la estimación de sus efectos sobre la serie temporal.

## 12.2. Efectos cualitativos: variables impulso y escalón

Las variables ficticias más utilizadas para representar sucesos cualitativos que afectan a la serie son de dos tipos: *variables impulso* y *variables escalón*. Las variables impulso representan sucesos que ocurren únicamente en un instante, por ejemplo, un accidente, un error de medida o una huelga. Las variables escalón representan acontecimientos que comienzan en un instante conocido y se mantienen a partir de ese instante, por ejemplo, una subida de precios, un cambio legal, un cambio de base en un índice, etc.

### 12.2.1. Variables impulso: función de respuesta a impulsos

Supongamos una variable  $y_t$  que sigue un modelo ARIMA que representaremos de forma simplificada por:

$$y_t = \psi(B)a_t, \quad (12.1)$$

y que esta serie está afectada en un instante dado,  $t = h$ , por un suceso conocido. En consecuencia, no observamos en todos los instantes esta serie, ya que para  $t = h$  observaremos una variable  $z_t$  que está relacionada con  $y_t$  mediante:

$$z_h = w_0 + y_h,$$

donde  $w_0$  es la magnitud del efecto sobre la serie.

Para representar el instante de ocurrencia de este suceso vamos a definir una variable impulso por:

$$I_t^{(h)} = \begin{cases} 0 & t \neq h \\ 1 & t = h \end{cases}$$

## 12. Análisis de intervención

y supondremos que la serie observada,  $z_t$ , sigue el modelo:

$$z_t = w_0 I_t^{(h)} + y_t,$$

es decir, utilizando (12.1):

$$z_t = w_0 I_t^{(h)} + \psi(B) a_t. \quad (12.2)$$

Este modelo establece que la serie observada,  $z_t$ , sigue el modelo ARIMA  $z_t = \psi(B)a_t$  en todos los instantes de tiempo en que la variable impulso,  $I_t^{(h)}$ , es cero. Sin embargo para  $t = h$ , observamos la suma del valor que proviene del modelo ARIMA,  $y_h$ , que es desconocido, y del efecto determinista de tamaño  $w_0$ . La representación (12.2) es un ejemplo de *función de transferencia*, es decir, de una función que describe el efecto sobre la variable respuesta,  $z_t$ , de una variable de entrada,  $I_t^{(h)}$ . La función de transferencia en este caso es simplemente la constante  $w_0$ .

El efecto de un impulso puede ser más complejo, y distribuirse en varios períodos, aunque es siempre un efecto transitorio que desaparece pronto. Por ejemplo, en una serie diaria de producción ocurre una avería grave en el día  $h$  que afecta a la producción en varios días posteriores, hasta que se recobra el funcionamiento normal. Como segundo ejemplo, una serie de tráfico donde una lluvia torrencial en  $t = h$  afecta al tráfico en  $m$  períodos posteriores, con lo que la serie observada,  $z_t$ , verificará:

$$\begin{aligned} z_t &= y_t, & t < h \\ z_{h+j} &= w_j + y_{h+j}, & j = 0, 1, \dots, m \\ z_{h+j} &= y_{h+j}, & j > m. \end{aligned}$$

Esta situación podría representarse introduciendo  $m$  variables impulso, como:

$$z_t = \sum_{j=0}^m w_j I_t^{(h+j)} + \psi(B) a_t \quad (12.3)$$

donde  $I_t^{(h+j)}$  toma el valor uno para  $t = h + j$  y cero en caso contrario.

Para estudiar otra forma de representar efectos múltiples observemos que:

$$B I_t^{(h)} = I_{t-1}^{(h)} = I_t^{(h+1)},$$

ya que la variable  $I_{t-1}^{(h)}$  en el instante  $t = h$  toma el valor 0, mientras que en  $t = h + 1$  es igual a 1. En consecuencia, esta variable es idéntica a  $I_t^{(h+1)}$ . Concluimos que aplicar el operador de retardo a una variable impulso es equivalente a desplazar en una unidad el instante donde la variable toma el valor 1. Generalizando esta idea, es fácil comprobar que:

$$B^j I_t^{(h)} = I_{t-j}^{(h)} = I_t^{(h+j)}.$$

Con esta notación, podemos representar el modelo (12.3) como:

$$z_t = w(B) I_t^{(h)} + \psi(B) a_t, \quad (12.4)$$

donde la función  $w(B) = (w_0 + w_1 B + \dots + w_m B^m)$  describe el efecto de la variable impulso y es un polinomio de orden  $m$  con los coeficientes de la intervención en los  $m$  periodos siguientes, y  $\psi(B)a_t$  es la serie que sigue el modelo ARIMA. La representación (12.4) tiene la ventaja sobre la (12.3) de poner de manifiesto que todos los coeficientes  $w_i$  provienen de la misma causa común, la intervención  $I_t^{(h)}$ , y a la función  $w(B)$  se la denomina *función de transferencia*. La figura 12.1 presenta un ejemplo de los coeficientes de la función de transferencia,  $w_i$ , y esta representación se conoce como la *función de respuesta al impulso*.

Si el número de periodos afectados por el impulso es largo, la representación anterior puede obligar a estimar muchos parámetros. Por ejemplo, en una serie horaria de consumo de energía eléctrica una avería puede afectar durante varias horas a la distribución, pero con pesos decrecientes hasta que se establezca la normalidad. Un modelo para representar estos efectos es:

$$z_t = \frac{w_0}{1 - \delta B} I_t^{(h)} + \psi(B) a_t,$$

donde  $0 < \delta < 1$ . En este modelo la función de transferencia es  $w_0(1 - \delta B)^{-1} = w_0(1 + \delta B + \delta^2 B^2 + \dots) = (w_0 + w_1 B + \dots + w_m B^m + \dots)$  donde  $w_k = w_0 \delta^k$ . El efecto de la intervención es ahora un polinomio infinito, pero con pesos decrecientes, que tienden a cero al aumentar el retardo. Los valores de la serie posteriores a  $h$  serán:

$$\begin{aligned} z_t &= y_t, & t < h \\ z_{h+j} &= w_0 \delta^j + y_{h+j}, & j \geq 0 \end{aligned}$$

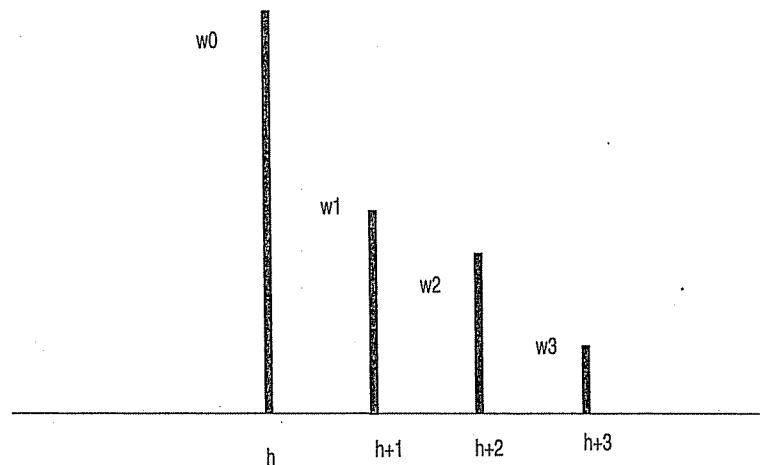
donde  $y_t = \psi(B)a_t$ , y todos los valores posteriores al momento de la intervención están afectados por ella, aunque con pesos decrecientes que tienden hacia cero. Señalemos que, como el efecto final tiende a cero, una variable impulso produce siempre un efecto transitorio sobre la serie.

---

### Ejemplo 12.1

En una serie diaria de transporte por carretera que mide el número de vehículos

**Figura 12.1 Representación de la respuesta de la serie a una intervención con función de respuesta**  
 $w_0 + w_1B + w_2B^2 + w_3B^3$



que circulan por una carretera se produce un accidente que obliga a cortes parciales durante dos días hasta que se arreglan los desperfectos. Un mes después del accidente se estima el siguiente modelo de intervención, donde  $I_t^{(h)}$  es una variable impulso que toma el valor uno en el día del accidente:

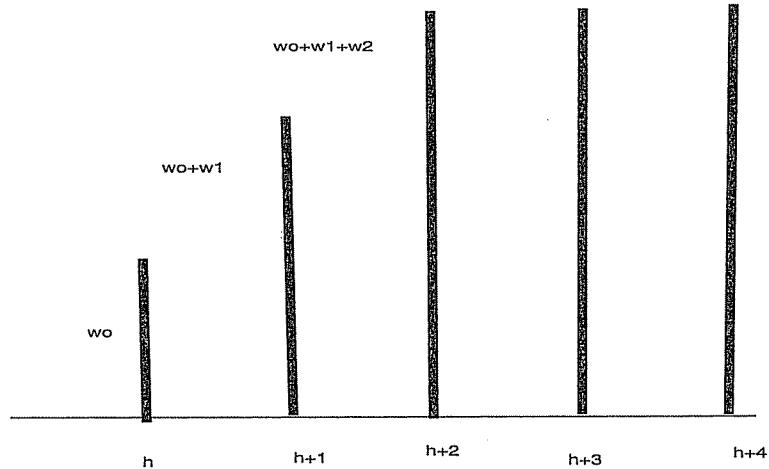
$$z_t = -(2100 + 3500B + 1000B^2 + 400B^3)I_t^{(h)} + \frac{(1 - .5B)(1 - .9B^7)}{\nabla\nabla_7}$$

Interpretemos este modelo. La función de transferencia indica que el día en que se produjo el accidente hubo una disminución de 2100 vehículos. El segundo día la disminución es de 3500 vehículos. El tercer día, que las condiciones son normales, hay todavía una disminución del tráfico de 1000 vehículos, y el cuarto día de 400. El modelo para la serie tiene estacionalidad diaria y muestra una pauta bastante regular, dado el alto valor del coeficiente de la media móvil estacional.

### 12.2.2. Variables escalón: ganancia

Estas ideas pueden extenderse para modelar intervenciones que tienen un efecto permanente sobre la serie como por ejemplo una subida de precios. Estas intervenciones se modelan con variables «escalón», que se definen mediante:

Figura 12.2 Respuesta ante un escalón de la función  
 $(w_0 + w_1B + w_2B^2)$



$$S_t^{(h)} = \begin{cases} 0 & t < h \\ 1 & t \geq h \end{cases}$$

El efecto de esta variable sobre una serie que sigue el modelo ARIMA dado por la ecuación (12.1) puede representarse mediante el modelo de intervención:

$$z_t = w_0 S_t^{(h)} + \psi(B)a_t,$$

que implica que todos los valores posteriores al instante  $h$  están efectuados por una cantidad constante,  $w_0$ . Este efecto equivale a un cambio de nivel en la serie a partir del instante  $h$ :

$$\begin{aligned} z_t &= y_t, & t < h \\ z_{h+j} &= w_0 + y_{h+j}, & j \geq 0. \end{aligned}$$

Es posible que un escalón tenga un efecto gradual. Por ejemplo, supongamos una serie de ventas de un producto que sufre una subida de precio en el instante  $h$ , y donde el primer mes hay una disminución de las ventas de  $w_0$ , en el segundo mes de  $w_1$  adicionales y en el tercero de  $w_2$  adicionales. Después, las ventas se estabilizan al nuevo nivel. Un modelo para representar estos efectos acumulativos es:

$$z_t = \sum_{j=0}^m w_j S_t^{(h+j)} + \psi(B)a_t, \quad (12.5)$$

con  $m = 2$ . Utilizando que, como ocurre con las variables impulso que  $B^j S_t^{(h)} = S_{t-j}^{(h)} = S_t^{(h+j)}$ , podemos escribir (12.5) con una única variable escalón mediante:

$$z_t = w(B)S_t^{(h)} + \psi(B)a_t,$$

donde  $w(B) = (w_0 + w_1B + \dots + w_mB^m)$  es la función de transferencia del escalón. El efecto sobre la serie será:

$$\begin{aligned} z_t &= y_t, \quad t < h \\ z_h &= w_0 + y_h, \\ z_{h+1} &= w_0 + w_1 + y_{h+1} \\ z_{h+j} &= \sum_{i=0}^j w_i + y_{h+j}, \quad j \geq 0. \end{aligned}$$

La figura 12.2 presenta un ejemplo de los efectos que sobre la serie produce una intervención en forma de escalón. Esta función se conoce como *función de respuesta a un escalón*. Se observa que un escalón siempre afecta a todos los valores posteriores a su ocurrencia. Llamaremos *ganancia* de un escalón a su efecto final a largo plazo, que es la suma de los efectos parciales, y se calcula como:

$$\text{ganancia} = w(1) = w_0 + w_1 + \dots + w_m.$$

### Ejemplo 12.2

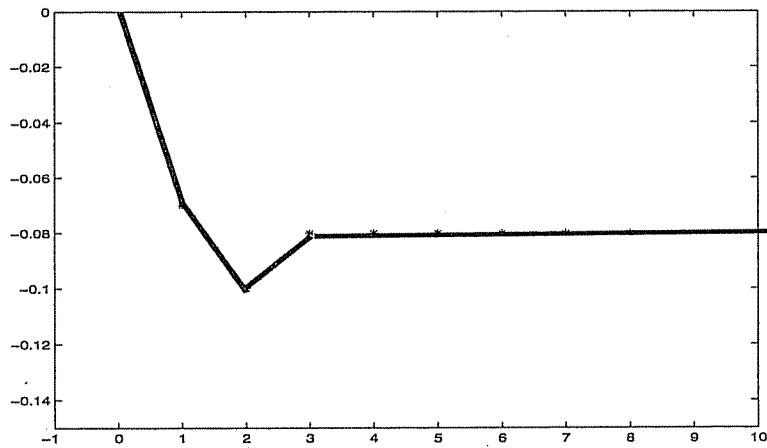
Se ha encontrado que en una serie de ventas el efecto de una subida de precios puede representarse mediante:

$$\log z_t = (-.07 - .03B + .02B^2)S_t^{(h)} + \frac{(1 - .3B)(1 - .7B^{12})}{\nabla\nabla_{12}}$$

Interpretar y dibujar estos efectos.

La función indica que en el mes de la subida se produjo una disminución de .07, que al estar la variable en logaritmos representa una disminución del 7 %. En el mes siguiente se produjo una disminución adicional del 3 % y en el tercero una subida del 2 %. El efecto total es una bajada del 8 %, que es la ganancia de la función de transferencia. La figura 12.3 representa este efecto sobre la serie.

Figura 12.3 Efecto de un escalón con ganancia .08



### 12.2.3. Relación entre impulsos y escalones

La diferencia básica entre una variable impulso y una escalón es que el efecto de la primera se reduce a un intervalo de tiempo, mientras que el de la segunda se extiende hasta el final del periodo observado. Suele decirse que el efecto de un impulso en la serie es *transitorio*, mientras que el de un escalón es *permanente*.

En determinados casos, no se conoce a priori si un efecto es permanente o transitorio. Por ejemplo, un aumento de precios en  $h$  puede, en algunos productos, disminuir las ventas a largo plazo (efecto permanente), mientras que, en otros, provoca únicamente una leve disminución de ventas en los períodos posteriores a la subida, que desaparece en pocos períodos. Para diferenciar ambas respuestas, tendremos en cuenta que un impulso es la primera diferencia de un escalón:

$$I_t^{(h)} = S_t^{(h)} - S_{t-1}^{(h)} = \nabla S_t^{(h)}. \quad (12.6)$$

Por lo tanto, podemos permitir que los datos nos indiquen si el efecto debe modelarse mediante un impulso o un escalón. Si estimamos el modelo:

$$z_t = (w_0 - w_1 B) S_t^{(h)} + \psi(B) a_t$$

y obtenemos que  $\hat{w}_0 \approx \hat{w}_1$ , resulta:

$$z_t = \hat{w}_0 \nabla S_t^{(h)} + \psi(B) a_t, \quad (12.7)$$

que puede escribirse:

$$z_t = \widehat{w}_0 I_t^{(h)} + \psi(B) a_t,$$

indicando que la serie sufre únicamente un efecto transitorio. Por el contrario, si obtenemos  $\widehat{w}_1 = 0$  o un valor de  $\widehat{w}_0$  con signo opuesto al de  $\widehat{w}_1$ , esto implica que existe un efecto permanente. Otra manera equivalente de decidir entre un efecto permanente o transitorio es partir de que el efecto es transitorio y estimar el modelo:

$$z_t = \frac{w_0}{1 - \delta B} I_t^{(h)} + \psi(B) a_t.$$

Si el parámetro  $\delta$  se estima como próximo a la unidad, tenemos que, aproximadamente, el modelo es:

$$z_t = w_0 \frac{1}{\nabla} I_t^{(h)} + \psi(B) a_t,$$

y utilizando la relación (12.6), el modelo se reduce a

$$z_t = \widehat{w}_0 S_t^{(h)} + \psi(B) a_t,$$

que representa un efecto permanente.

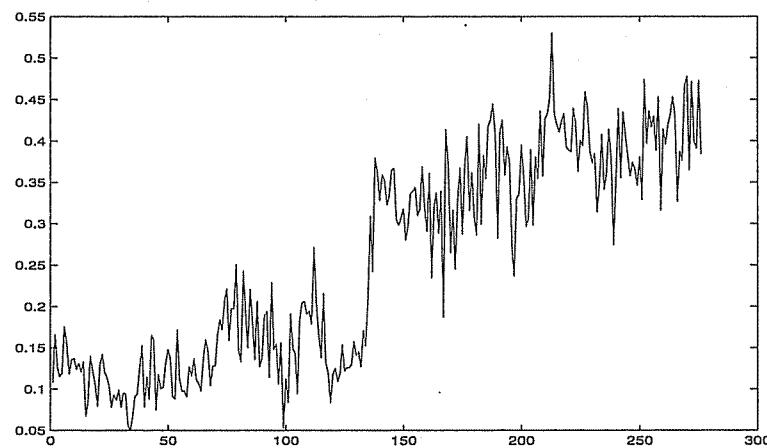
Conviene siempre que haya dudas respecto al tipo de efecto que produce la intervención y estimar una función de transferencia que permita ambas posibilidades según los valores de algún parámetro, para permitir que los datos nos indiquen el tipo de efecto. En caso de duda podemos estimar los dos modelos y seleccionar entre ambos con un criterio de selección.

### Ejemplo 12.3

Consideremos las series semanales de cuota de mercado de los dentífricos Colgate y Crest (fichero crescolgate.dat). En agosto de 1960 la Asociación Dental Americana (America Dental Association, ADA) hizo público un estudio que había demostrado un efecto positivo del fluor introducido por Crest en la reducción de caries. Esta fecha corresponde a la observación 135 en la figura 12.4, donde se observa un incremento en el nivel de la serie. Wichern y Jones (1977) encontraron que esta declaración tuvo efectos permanentes sobre el aumento de la cuota de mercado de Crest. La figura 12.4 presenta el gráfico de la serie y se observa claramente el cambio de nivel que experimenta alrededor del periodo 135.

El modelo univariante para esta serie es  $\nabla z_t = (1 - \theta B) a_t$ , que tomaremos como modelo ARIMA subyacente. Definiremos la variable escalón

**Figura 12.4 Serie semanal de cuota de mercado del dentífrico Crest en el periodo 1958-1963**



$$S_t^{(135)} = \begin{cases} 0 & t < 135 \\ 1 & t \geq 135, \end{cases}$$

y vamos a incluirla en el modelo permitiendo efectos retardados. El modelo a estimar es entonces:

$$z_t = (w_0 + w_1 B) S_t^{135} + \frac{(1 - \theta B)}{\nabla} a_t,$$

y su estimación es:

$$z_t = (0.065 + 0.112B) S_t^{135} + \frac{(1 - 0.77B)}{\nabla} a_t.$$

Este modelo indica que en la semana en que ocurrió la intervención la cuota de mercado de Crest aumentó en 6.5 puntos y en la semana siguiente en 11.2. La ganancia, o efecto total, es la suma de ambos efectos y concluimos que la declaración de la ADA supuso un aumento de la cuota de mercado total para Crest de 17.7 puntos.

Observemos que el modelo estimado equivale a introducir dos variables escalón. La primera,  $S_t^{135}$ , corresponde a la semana del anuncio y la segunda,  $S_t^{136}$ :

$$S_t^{(136)} = \begin{cases} 0 & t < 136 \\ 1 & t \geq 136 \end{cases}$$

corresponde a la semana siguiente. El modelo puede también escribirse

$$z_t = 0.065S_t^{135} + 0.112S_t^{136} + \frac{(1 - 0.77B)}{\nabla} a_t.$$


---

### 12.3. Efectos deterministas generales

Además de variables impulso y escalón podemos incluir en el modelo cualquier serie determinista,  $x_t$ , con los efectos que queremos modelar. Por ejemplo, en muchas series mensuales de actividad productiva o económica es importante el número de días laborables que existen en cada mes, que puede oscilar de un año a otro. Otro efecto similar, aunque menos importante, es el efecto de año bisiesto, que hace que febrero tenga un día más. Como tercer ejemplo, una serie de demanda de servicios puede estar afectada por los precios que, si son regulados, como ocurre con algunos servicios públicos, formarán una serie determinista.

El efecto de una variable determinista,  $x_t$ , sobre la serie de interés puede representarse por:

$$z_t = f(x_t) + y_t,$$

donde  $z_t$  es la serie observada que es la suma de dos componentes: la función de transferencia,  $f(x_t)$ , que describe el efecto de la variable explicativa o de entrada  $x_t$  sobre la serie observada, e  $y_t$ , la serie no observada que no incluye el efecto de la variable determinista  $x_t$  y que sigue el modelo ARIMA (12.1). Suponiendo que el efecto de  $x_t$  es lineal, podemos escribir la función de transferencia como :

$$f(x_t) = v_0x_t + v_1x_{t-1} + v_2x_{t-2} + \dots = (v_0 + v_1B + v_2B^2 + \dots)x_t. \quad (12.8)$$

Para interpretar esta función vamos a analizar la respuesta que proporciona cuando la serie  $x_t$  varía de una manera controlada simple. Por ejemplo, si  $x_t$  es una variable impulso, que es cero para  $t \neq h$ , el efecto sobre la variable respuesta es:

$$\begin{aligned} f(x_{h-1}) &= v_0x_{h-1} + v_1x_{h-2} + \dots = 0 \\ f(x_h) &= v_0x_h + v_1x_{h-1} + \dots = v_0 \\ f(x_{h+1}) &= v_0x_{h+1} + v_1x_h + \dots = v_1 \\ f(x_{h+2}) &= v_0x_{h+2} + v_1x_{h+1} + \dots = v_2 \\ &\vdots \quad \vdots \quad \vdots \\ f(x_{h+k}) &= v_0x_{h+k} + \dots + v_kx_h + \dots = v_k \end{aligned}$$

y cada coeficiente de la función de transferencia,  $v_i$ , representa la respuesta en  $h+i$  cuando la serie  $x_t$  sufre un impulso unitario en el instante  $h$ . La representación de  $v_k$  en función del retardo  $k$  es la *función de respuesta a impulsos (fri)*.

Otra forma de caracterizar esta respuesta es suponer que  $x_t$  es una variable escalón, cero hasta  $x_{h-1}$ , y toma el valor 1 en  $x_h$  y en todos los valores posteriores. Se obtiene entonces la respuesta:

$$\begin{aligned} f(x_{h-1}) &= v_0x_{h-1} + v_1x_{h-2} + \dots = 0 \\ f(x_h) &= v_0x_h + v_1x_{h-1} + \dots = v_0 \\ f(x_{h+1}) &= v_0x_{h+1} + v_1x_h + \dots = v_0 + v_1 \\ f(x_{h+2}) &= v_0x_{h+2} + v_1x_{h+1} + \dots = v_0 + v_1 + v_2 \\ &\vdots \quad \vdots \quad \vdots \\ f(x_{h+k}) &= v_0x_{h+k} + \dots + v_kx_h + \dots = \sum_{j=1}^k v_j, \end{aligned}$$

que es la función de respuesta a escalones. Su valor límite es la ganancia,  $g$ , dada por

$$g = \sum_{j=1}^{\infty} v_j.$$

La representación general de la función de transferencia  $v(B)$  puede contener infinitos parámetros y no es operativa. Encontramos este mismo problema en la representación del proceso lineal general, cuando estudiámos cómo aproximar  $\psi(B)$  mediante un cociente de dos operadores finitos. En particular, cuando el operador del denominador era la unidad y sólo teníamos el operador finito del numerador,  $\theta(B)$ , se obtenían los procesos MA, mientras que la combinación de ambos operadores daba lugar a los procesos ARMA. Extendiendo esta idea, podríamos pensar en aproximar  $v(B)$  mediante una representación finita del tipo:

$$w_0 + w_1B + \dots + w_mB^m = w_m(B),$$

utilizando las letras  $w$  para indicar un número finito de coeficientes. Otra posibilidad es permitir un número infinito de ellos, pero con estructura simple de amortiguamiento, por ejemplo:

$$\frac{w_0}{(1 - \delta B)} = w_0 (1 + \delta B + \delta^2 B^2 + \dots),$$

donde suponemos  $|\delta| < 1$  de manera que la serie sea convergente. Por

analogía con lo estudiado en los modelos ARMA diremos que  $w(B)$  representa una aproximación de «media móvil» a  $v(B)$ , mientras que  $w_0(1 - \delta B)^{-1}$  es del tipo «autorregresiva». Una representación general de la función de transferencia es suponer:

$$v(B) = \frac{w_m(B)}{\delta_a(B)}$$

donde

$$w_m(B) = w_0 + w_1B + \dots + w_mB^m; \quad \delta_a(B) = 1 - \delta_1B - \dots - \delta_aB^a,$$

y llamaremos  $m$  y  $a$  a los órdenes de los operadores de media móvil y autorregresivo. Observemos que el polinomio  $w_m(B)$  no comienza con 1, sino con  $w_0$ , y que sus coeficientes se toman con signo positivo, como es habitual en regresión.

Para que el sistema sea estable, es decir, para que un incremento finito en  $x_t$  no conduzca a un incremento infinito de  $y_t$ , las raíces del polinomio  $\delta_a(B) = 0$  tienen que estar fuera del círculo unidad. (Esta condición es equivalente a la del operador autorregresivo de un proceso ARMA para asegurar la estacionaridad.) No existen sin embargo restricciones sobre las raíces de  $w_m(B)$ .

Finalmente, muchos sistemas responden a un impulso en  $t$  después de cierto periodo de retardo que llamaremos  $b$ , es decir, los primeros  $b$  coeficientes  $v_i$  son cero. Tendremos en cuenta este efecto introduciendo un tercer parámetro  $b$  y escribiendo:

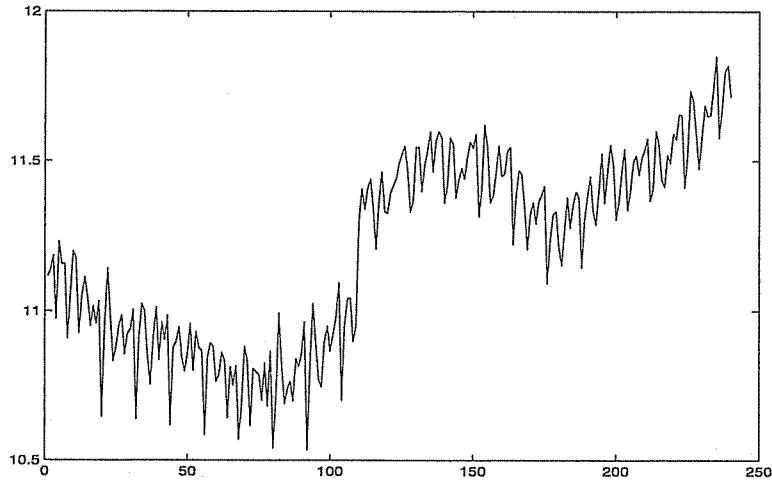
$$v(B) = \frac{w_m(B)}{\delta_a(B)} B^b \quad (12.9)$$

donde si  $b = 0$  la respuesta es instantánea y si  $b \neq 0$  comienza con  $b$  periodos de retardo. El filtro lineal  $v(B)$  queda caracterizado por los órdenes  $(m, a, b)$ .

Diremos que la relación definida por  $v(B)$  es estable si un aumento finito en  $x_t$  produce un efecto finito en  $y_t$ . En consecuencia, el sistema será estable si la ganancia es finita,  $\sum_0^\infty v_i < \infty$ , lo que exige que la serie  $v_0 + v_1B + v_2B^2 + \dots$  sea convergente para  $|B| \leq 1$ . Observemos que al sustituir  $B = 1$  en la expresión de  $v(B)$  se obtiene precisamente la ganancia. Este resultado es útil para calcular la ganancia de cualquier función de transferencia  $(m, a, b)$ , ya que:

$$g = v(1) = \frac{w_m(1)}{\delta_a(1)} (1)^b = \frac{w_0 + w_1 + \dots + w_m}{1 - \delta_1 - \dots - \delta_a} \quad (12.10)$$

**Figura 12.5 Serie de accidentes laborales en logaritmos. Datos mensuales desde enero de 1979 hasta diciembre de 1998**



y la ganancia será finita si las raíces de  $\delta(B) = 0$  están fuera del círculo unidad.

---

#### Ejemplo 12.4

Consideremos la serie  $z_t = \ln A_t$ , donde  $A_t$  es la serie de accidentes laborales, que se presenta en la figura 12.5.

Vamos a estimar cuatro efectos deterministas sobre esta serie. El primero es el número de días laborables del mes, que será una variable determinista  $D_t$ . Esta variable la introducimos porque si un mes tiene menos días laborables sólo por ese efecto debería tener menos accidentes. El segundo es el efecto de Semana Santa, que modelaremos como una variable impulso que toma el valor 1 si la Semana Santa ocurre en ese mes y cero en caso contrario. Por tanto, esta variable sólo toma valores distintos de cero en marzo y/o abril. El tercero es el efecto de año bisiesto, que toma el valor 1 en febrero de un año bisiesto y cero en el resto. La cuarta variable es un cambio de nivel que se produce en la serie en febrero de 1988 (observación 110), como consecuencia de una definición de accidente laboral. El modelo a estimar será

$$z_t = w_1 D_t + w_2 I_{1t}^{(h)} + w_3 I_{2t}^{(h)} + w_4 S_t + \psi(B)a_t$$

donde  $\psi(B)a_t$  es el modelo subyacente para la serie. Vamos a estimar el modelo con el programa TSW, que identifica automáticamente un modelo para la serie. El mejor modelo que encontramos para esta serie es el  $(1 - \phi_1 B - \phi_2 B^2)(1 -$

## 12. Análisis de intervención

$\Phi B^{12})\nabla\nabla_{12} \ln A_t = (1 - \Theta_1 B^{12} - \Theta_2 B^{24})a_t$ , pero como el programa TSW sólo permite MA estacionales de orden uno, utilizaremos este modelo simplificado para estimar los efectos. El resultado de la estimación obtenido con el programa TSW es:

$$\ln A_t = 0.0092D_t - .078I_t^{(h)} + .0287I_t^{(h)} + .297S_t^{(h)} + N_t \quad (12.14)$$

donde debajo de cada estimador se indica el valor del estadístico  $t$ . El modelo estimado para el ruido  $N_t$  es, incluyendo entre paréntesis los estadísticos  $t$ :

$$\nabla\nabla_{12}N_t = (1 - 0.53B)(1 - 0.62B^{12}). \quad (9.5) \quad (11.3)$$

Poniendo juntas estas dos ecuaciones, el modelo completo para la serie del logaritmo de los accidentes puede escribirse como:

$$\begin{aligned} \ln A_t = & 0.0092D_t - .078I_t^{(h)} + .0287I_t^{(h)} + .297S_t^{(h)} + \\ & + \frac{(1 - 0.53B)(1 - 0.62B^{12})}{\nabla\nabla_{12}} a_t. \end{aligned} \quad (12.14)$$

La serie  $N_t$  se interpreta como la variable logaritmo de los accidentes corregida de efectos deterministas mediante la ecuación:

$$\ln A_t^C = \ln A_t - (0.0092D_t - .078I_t^{(h)} + .0287I_t^{(h)} + .297S_t^{(h)}) = N_t, \quad (12.14)$$

y la serie  $\ln A_t^C$ , que contiene sólo la parte de  $\ln A_t$  que no puede explicarse por los efectos deterministas, es la serie que sigue el modelo ARIMA.

En el programa TSW las tres variables de intervención de días del mes, años bisiestos y Semana Santa se generan automáticamente. La variable  $D_t$  tiene en cuenta el número de días laborables en el mes, y el programa TSW la define como número de días laborables (contando de lunes a viernes) menos número de sábados y domingos multiplicados por  $5/2$ . De esta manera, la variable  $D_t$  será cero si hay exactamente un número fijo de semanas completas en el mes, como en un mes con 28 días que no tiene ninguna fiesta: ese mes tiene 20 días laborables y 8 festivos y  $D_t = 20 - 8 \times 5/2 = 0$ . La variable será positiva cuando haya relativamente más días laborables en el mes, como en un mes de 31 días que comienza en lunes y tiene 23 días laborables y 8 festivos,  $D_t = 23 - 8 \times 5/2 = 3$  y negativa en un mes de 30 días que comienza en sábado y  $D_t = 20 - 10 \times 5/2 = -5$ . La interpretación del coeficiente de esta variable es que un día extra laborable sobre la relación 20 y 8 produce un aumento de los accidentes en un .9%. Observemos que un aumento de 1 día sobre 20 es un aumento de .05, por lo que esos días laborables extras tienen

un efecto más que proporcional sobre los accidentes laborales.

La segunda variable,  $I_t^{(h)}$ , es una variable impulso que toma el valor uno si la Semana Santa ocurre en ese mes y cero en caso contrario. Se observa que esta semana se asocia con una disminución de los accidentes de ese mes del 7.8 %. Como la Semana Santa supone dos días adicionales de vacaciones, que son el  $2/31 = .0667$  del mes, una disminución del 6.67 % de los días de trabajo en el mes supone una reducción de accidentes del 7.8 %. La reducción debida a la Semana Santa es un poco mayor que la estrictamente proporcional.

La tercera variable toma el valor 1 cuando febrero tiene un día más por corresponder a un año bisiesto. Como  $1/29 = .0345$ , un año bisiesto supone un aumento del 3.45 % de los días laborables, y un aumento del 2.8 % en el número de accidentes, menor que el estrictamente proporcional.

El cambio de nivel en febrero de 1988 supuso un aumento de la contabilización de accidentes del 29.7 %. En efecto, mirando los datos de la serie original se observa que de un nivel de alrededor de 60000 accidentes al año se pasa a un nivel de más de 80000, que es consistente con la estimación obtenida para el cambio de nivel.

Finalmente, la desviación típica residual es 0.045, que indica que el error de predicción de la serie corregida por los efectos deterministas es el 4.5 %. Como comparación, este valor para el mismo modelo sin corrección de efectos deterministas es del 6.6 %, lo que indica que en esta serie se obtiene una reducción sustancial del error modelando efectos deterministas.

---

#### 12.4. Construcción de modelos de intervención

La estimación de un modelo de intervención se realiza por máxima verosimilitud (MV), de manera similar a la estudiada en el capítulo 10. Los parámetros a estimar son los del modelo ARIMA y los de la intervención, explicaremos el procedimiento con el modelo de intervención:

$$z_t = w(B)I_t^{(h)} + \frac{1}{\phi(B)}a_t, \quad (12.11)$$

que escribiremos:

$$\phi(B)(z_t - w(B)I_t^{(h)}) = a_t.$$

Esta formulación indica que si conociésemos los parámetros de la función de transferencia, la serie  $y_t$ , corregida de la intervención:

$$y_t = z_t - w(B)I_t^{(h)} \quad (12.12)$$

sigue un modelo ARIMA. Puede demostrarse que la estimación MV puede realizarse en tres etapas. En la primera, partiendo de una esti-

## 12. Análisis de intervención

---

mación inicial de la función de transferencia calculamos la serie  $y_t$ . En la segunda, estimamos los parámetros del modelo ARIMA con la serie  $y_t$  y calculamos los residuos,  $e_t = \phi(B)z_t$ . Con estos residuos se estima ahora el efecto de la intervención escribiendo los residuos como función de los parámetros de la función de transferencia. Para ello, se multiplica toda la ecuación por la estructura del modelo ARIMA estimado. Se escribe el modelo (12.11) como:

$$e_t = w(B)\phi(B)I_t^{(h)} + a_t.$$

Llamando  $x_t = \phi(B)I_t^{(h)}$  a la nueva variable explicativa, esta ecuación puede estimarse por regresión y obtener nuevos valores para los parámetros de la función de transferencia. Con estos nuevos valores calculamos una nueva serie corregida,  $y_t$ , con (12.12) y con ella reestimamos los parámetros del modelo ARIMA y el efecto de intervención e iteramos hasta convergencia.

Para ilustrar el procedimiento, consideremos el caso más simple:

$$z_t = w_0 I_t^{(h)} + \frac{1}{1 - \phi B} a_t, \quad (12.13)$$

y la estimación equivale a las tres etapas siguientes:

1. Partiendo de una estimación inicial del efecto de la intervención,  $\widehat{w}_0$ , se construye la serie corregida de este efecto,  $y_t$ , mediante:

$$y_t = z_t - \widehat{w}_0 I_t^{(h)}.$$

2. Con esta serie corregida se estima el parámetro  $\phi$  en el modelo AR(1):  $y_t = \phi y_{t-1} + u_t$ . Sea  $\widehat{\phi}$  el valor estimado.
3. Multiplicando por  $(1 - \widehat{\phi}B)$  el modelo (12.13), llamando  $\widehat{e}_t = z_t - \widehat{\phi}z_{t-1}$  a los residuos calculados con la serie observada,  $x_t = (1 - \widehat{\phi}B)I_t^{(h)}$  a la variable impulso modificada por la parte AR del modelo y suponiendo que aproximadamente  $\widehat{\phi} = \phi$ , tenemos que :

$$\widehat{e}_t = w_0 x_t + a_t.$$

En esta ecuación de regresión los residuos,  $\widehat{e}_t$ , son la variable dependiente; la variable independiente o regresor es la intervención modificada,  $x_t$ , y el parámetro a estimar es el tamaño del efecto,  $w_0$ . Como el regresor,  $x_t = I_t^{(h)} - \widehat{\phi} I_{t-1}^h$ , es una variable que toma siempre el valor cero menos en  $x_h = 1$  y  $x_{h+1} = -\widehat{\phi}$ , el coeficiente

de regresión se obtendrá con:

$$\widehat{w}_0 = \frac{\sum \widehat{e}_t x_t}{\sum x_t^2} = \frac{\widehat{e}_h - \widehat{\phi} \widehat{e}_{h+1}}{1 + \widehat{\phi}^2}.$$

Con este nuevo valor de  $\widehat{w}_0$  volveremos a (1), e iteraremos estas tres etapas (1) a (3), hasta obtener convergencia.

La diagnosis de un modelo de intervención se realiza de la misma forma estudiada para los modelos ARMA.

### 12.5. Estimación de valores ausentes

En algunas series los valores en determinados periodos de tiempo no se conocen. Decimos entonces que la serie tiene valores ausentes. Supondremos que estos huecos ocurren aleatoriamente, de manera independiente del valor de la serie (es decir, que no faltan precisamente los valores más altos, más bajos u otros con una pauta fija), de manera que la aparición de un valor ausente no proporciona información sobre su magnitud. En estas condiciones, puede utilizarse el análisis de intervención para estimar estos valores ausentes. Recordemos que la interpretación de una intervención es que en el instante en que ocurre no se observa el valor del proceso ARIMA, ya que el valor de la serie en ese instante incluye el efecto de la intervención. Intuitivamente, esta relación sugiere que podemos estimar un valor ausente suponiendo que ha ocurrido un suceso puntual en esa fecha, y estimando con un modelo de intervención el valor no observado de la serie en ese instante.

Para analizar cómo llevar a cabo esta idea, supongamos que la observación  $h$  está ausente. Llamaremos  $Z_{(h)} = (z_1, \dots, z_{h-1}, z_{h+1}, \dots, z_T)$  a los datos disponibles, donde el subíndice  $(h)$  indica que el valor  $h$  está ausente. Estimar el valor ausente minimizando el error cuadrático de estimación implica, como vimos al estudiar predicción, obtener su esperanza condicionada a la información disponible:

$$\widehat{z}_h = E(z_h | Z_{(h)}).$$

Si conocemos los parámetros del modelo, puede demostrarse que si la serie sigue un proceso ARMA esta esperanza viene dada por:

$$E(z_h | Z_{(h)}) = - \sum_{j=1}^{\infty} \rho_j^h (z_{h+j} + z_{h-j}) \quad (12.14)$$

donde los coeficientes  $-\rho_j^h$  miden la correlación entre observaciones de

la serie separadas por  $j$  periodos cuando eliminamos de esta relación el efecto de todas las restantes observaciones. Los  $\rho_j^i$  son los coeficientes de correlación inversa que se definieron en el apéndice 5.1 (véase también el apéndice 12.1).

Se demuestra que es posible estimar los valores ausentes mediante el método siguiente:

1. Insertar en el hueco un valor arbitrario  $z_h^*$ , por ejemplo cero, para tener una serie completa,  $z_t$ .
2. Estimar un modelo de intervención para la serie completada,

$$\mathbf{Z}^* = (z_1, \dots, z_{h-1}, z_h^*, z_{h+1}, \dots, z_T)$$

incluyendo un impulso en la posición de la observación ausente. Sea  $\hat{w}_0$  el valor estimado para el impulso.

3. Estimar el valor ausente mediante:

$$\hat{z}_h = E(z_h | \mathbf{Z}_{(h)}) = z_h^* - \hat{w}_0. \quad (12.15)$$

En efecto, al completar la serie hemos transformado el problema de valores ausentes en un problema de intervención sobre una serie completa,  $\mathbf{Z}^*$ . La serie  $z_t$  no observada en su totalidad, que sigue el modelo ARIMA, está relacionada con la construida  $z_t^*$  mediante  $z_t = z_t^*$ , si  $t \neq h$ , y  $z_h = z_h^* + w_0$ , donde  $w_0$  es desconocida. Si estimamos  $w_0$  entonces conoceremos el valor de  $z_h$ . Con este objetivo, estimamos el modelo de intervención:

$$z_t^* = w_0 I_t^{(h)} + \psi(B) a_t = w_0 I_t^{(h)} + z_t \quad (12.16)$$

Tomando esperanzas en el instante  $t = h$  condicionadas a los valores  $\mathbf{Z}_{(h)} = \mathbf{Z}_{(h)}^*$ , tenemos que

$$z_h^* = w_0 + E(z_h | \mathbf{Z}_{(h)})$$

ya que el valor de  $z_h$  es conocido, igual a  $z_h^*$ . Por tanto, la esperanza del valor del proceso ARIMA en el instante  $h$ ,  $\hat{z}_h$ , se obtiene insertando en esta ecuación la estimación  $\hat{w}_0$  lo que conduce a la ecuación (12.15).

Puede demostrarse (véase el apéndice 12.1) que al estimar el modelo (12.16) obtenemos simultáneamente la estimación de los parámetros con los datos observados, que no depende de  $z_h^*$ , y la estimación de la constante  $w_0$ . La predicción del valor ausente se obtiene con (12.15) y no depende del valor arbitrario  $z_h^*$ .

Como ilustración, supongamos que el modelo para la serie es un paseo aleatorio y que falta el valor  $z_h$ . Insertando un valor arbitrario  $z_h^*$  en esa

posición, el modelo (12.16) será:

$$z_t^* = w_0 I_t^{(h)} + \frac{1}{1-B} a_t,$$

que equivale a

$$\nabla z_t^* = w_0 x_t + a, \quad (12.17)$$

donde la variable  $x_t = \nabla I_t^{(h)}$  es cero salvo para  $x_h = 1$  y  $x_{h+1} = -1$ . El parámetro  $w_0$  en (12.17) se estimará mediante

$$\begin{aligned} \widehat{w}_0 &= \frac{\sum_t x_t \nabla z_t^*}{\sum(x_t)^2} = \frac{\nabla z_h^* - \nabla z_{h+1}^*}{2} = \\ &= \frac{z_h^* - z_{h-1} - z_{h+1} + z_h^*}{2} = z_h^* - \frac{z_{h-1} + z_{h+1}}{2}, \end{aligned}$$

y la estimación del valor ausente será, aplicando (12.15):

$$\widehat{z}_h = \frac{z_{h-1} + z_{h+1}}{2}.$$

Es decir, el valor ausente se estima como la media de los valores anterior y posterior, y no depende del valor arbitrario insertado  $z_h^*$ . En el apéndice 12.1 se justifica que este estimador es el resultado de aplicar la solución general (12.14) a un paseo aleatorio.

Este enfoque se generaliza sin dificultad si falta un bloque de  $k$  observaciones entre  $h$  y  $h+k-1$ . Sea  $\mathbf{Z}_{(h,h+k-1)} = (z_1, \dots, z_{h-1}, z_{h+k}, \dots, z_T)$  la serie disponible. Introduciendo ceros en los  $k$  valores ausentes para construir la serie completa  $\mathbf{Z}^*$ , y definiendo  $k$  variables impulso  $I_t^{(h+i-1)}$  para  $i = 1, \dots, k$ , estimaremos el modelo:

$$z_t^* = \sum_{i=1}^k w_{0,i} I_t^{(h+i-1)} + z_t, \quad (12.18)$$

y la estimación de los valores ausentes se obtiene de

$$\widehat{z}_{h+i} = E(z_{h+i} | \mathbf{Z}_{(h,h+k-1)}) = z_{h+i}^* - \widehat{w}_{0,i}.$$

Esta idea se generaliza para cualquier combinación de valores ausentes. Puede demostrarse, véase el apéndice 12.1, que si aplicamos análisis de intervención sobre la serie completada artificialmente rellenando los valores ausentes, aproximadamente los parámetros estimados son los que obtendríamos maximizando la verosimilitud de los datos observados y la estimación de los valores ausentes es su esperanza condicional.

---

## 12. Análisis de intervención

---

**Tabla 12.1 Resultado de la estimación de valores ausentes y sus valores verdaderos**

t	201	202	203	204	205
$\hat{w}$	11.396	11.543	11.530	11.407	11.367
real	11.366	11.465	11.540	11.335	11.397

### Ejemplo 12.5

Para ilustrar la estimación de valores ausentes vamos a suponer que los valores de la serie de accidentes 201 a 205 no se han observado. Podemos poner entonces cero en esas posiciones y estimar el modelo de intervención:

$$\ln A_t = \sum_{i=1}^5 w_i I_{it}^{(h_i)} + \frac{(1-\theta B)(1-\Theta B^{12})}{\nabla \nabla_{12}} a_t$$

donde las cinco variables  $I_{it}^{(h_i)}$  para  $i = 1, 2, 3, 4, 5$  toman el valor cero en todas las posiciones menos en la  $h_i$ , donde  $h_i = 201, \dots, 205$ . El resultado de la estimación de los valores ausentes se indica en la tabla 12.1. Puede verse que la estimación es razonablemente buena, incluso con el modelo simplificado que se ha utilizado.

---

### Ejercicios 12

- 12.1 Escribir la función de respuesta a impulsos y a escalones de la función de transferencia  $\frac{5-2B}{1-7B}$
  - 12.2 Escribir la función de respuesta a impulsos y a escalones de la función de transferencia  $\frac{2+3B+4B^2}{1-7B+1B^2} B^3$ .
  - 12.3 Demostrar que la función de respuesta a impulsos de la función de transferencia  $(2 - 2B)S_t^{(h)}$  sólo tiene un término distinto de cero.
  - 12.4 Calcular la ganancia de la función de transferencia del ejercicio 12.2.
  - 12.5 Dado el modelo de intervención,  $z_t = w_0 I_t^{(h)} + (1-\theta B)a_t$ , sean  $e_t = (1-\theta B)^{-1}$   $z_t = z_t + \theta z_{t-1} + \theta^2 z_{t-2} + \dots$  los residuos del modelo univariante. Justificar que la estimación del coeficiente  $w_0$  se realiza mediante  $\hat{w}_0 = (e_t + \theta e_{t+1} + \theta^2 e_{t+2} + \dots) / (1 + \theta^2 + \theta^4 + \theta^6 + \dots) = (1 - \theta^2)(e_t + \theta e_{t+1} + \theta^2 e_{t+2} + \dots)$ .
-

### Apéndice 12.1: Predicción de valores ausentes

Vamos a demostrar que el parámetro de un AR(1) con un valor ausente puede estimarse sustituyendo el valor ausente por cualquier valor arbitrario y estimando con esta serie completa un modelo AR(1) con una intervención en el valor ausente.

Supongamos que el valor  $z_h$  no se ha observado. La función de verosimilitud condicionada con este valor ausente es:

$$\begin{aligned} f(z_2, \dots, z_{h-1}, z_{h+1}, \dots, z_T | z_1) &= \\ &= f(z_2 | z_1) \dots f(z_{h+1} | z_{h-1}) f(z_{h+2} | z_{h+1}) \dots f(z_T | z_{T-1}). \end{aligned}$$

Suponiendo normalidad, hemos visto en el capítulo 10 que para  $t \geq 2$  y  $t \neq h$  la densidad  $f(z_t | z_{t-1})$  es  $N(\phi z_{t-1}, \sigma^2)$ . La distribución de  $f(z_{h+1} | z_{h-1})$ , para la observación después del hueco, es también normal, pero con otros parámetros. Para obtenerlos, escribamos:

$$z_{h+1} = \phi(\phi z_{h-1} + a_h) + a_{h+1} = \phi^2 z_{h-1} + \phi a_h + a_{h+1},$$

y tomando esperanzas condicionadas a  $z_{h-1}$  la media de esta distribución es  $\phi^2 z_{h-1}$ . Pasando este término al otro lado y elevando al cuadrado, la varianza condicionada es  $\sigma^2(1 + \phi^2)$ . Concluimos que  $f(z_{h+1} | z_{h-1})$  es  $N(\phi^2 z_{h-1}, \sigma^2(1 + \phi^2))$ . Sustituyendo estos resultados en la función de verosimilitud condicionada se obtiene:

$$\begin{aligned} l(\phi, \sigma^2 | z_1) &= -\frac{(T-2)}{2} \ln \sigma^2 - \frac{1}{2} \ln(1 + \phi^2) - \\ &\quad - \sum_{t \in A} \frac{(z_t - \phi z_{t-1})^2}{2\sigma^2} - \frac{(z_{h+1} - \phi^2 z_{h-1})^2}{2\sigma^2(1 + \phi^2)}, \end{aligned} \tag{12.19}$$

donde  $A = \{2, \dots, h-1, h+2, \dots, T\}$ .

Vamos a demostrar que si completamos la serie con un valor arbitrario y estimamos los parámetros con un modelo de intervención se obtiene una función de verosimilitud similar. Por tanto, ambos métodos llevarán a estimaciones de los parámetros similares. Sea  $z_t^*$  el valor arbitrario introducido en la posición  $h$  y llamemos  $z_t^*$  a los valores de esta serie completa, donde  $z_t^* = z_t$  para  $t \neq h$ . El modelo de intervención es:

$$z_t^* = w I_t^{(h)} + \frac{a_t}{1 - \phi B},$$

que puede escribirse:

$$z_t^* = \phi z_{t-1}^* + w I_t^{(h)} - \phi w I_{t-1}^{(h)} + a_t.$$

## 12. Análisis de intervención

Para escribir la verosimilitud condicionada de este modelo observemos que para  $t \neq h, h+1$ , las variables  $z_t^*$  condicionadas a su pasado son normales con media  $\phi z_{t-1}$  y varianza  $\sigma^2$ . La variable  $z_h^*$  tiene media  $\phi z_{h-1} + w$  y varianza  $\sigma^2$  y la  $z_{h+1}^*$  tiene media  $\phi(z_h^* - w)$  y también varianza  $\sigma^2$ . La verosimilitud para los parámetros, condicional al primer valor  $z_1$ , es, utilizando que  $z_t^* = z_t$  para  $t \neq h$ :

$$l(\omega, \phi, \sigma^2 | z_1) = -\frac{(T-1)}{2} \ln \sigma^2 - \sum_{t \in A} \frac{(z_t - \phi z_{t-1})^2}{2\sigma^2} - \frac{(z_h^* - \phi z_{h-1} - w)^2}{2\sigma^2} - \frac{(z_{h+1} - \phi(z_h^* - w))^2}{2\sigma^2}. \quad (12.20)$$

Para maximizar esta función observemos que la estimación de  $w$  sólo depende de los valores  $z_h^*$  y  $z_{h+1}$ . Es decir, sea cual sea el valor de  $\phi$  al derivar la verosimilitud respecto a  $w$ , se obtiene que el estimador de  $\omega$  debe verificar:

$$(z_h^* - \phi z_{h-1} - w) = (z_{h+1} - \phi(z_h^* - w))\phi,$$

que resulta en:

$$\hat{w} = z_h^* - \frac{\phi}{(1+\phi^2)}(z_{h+1} + z_{h-1}). \quad (12.21)$$

Este estimador puede interpretarse como la diferencia entre el valor arbitrario introducido,  $z_h^*$  y su predicción óptima con el resto de la muestra. En efecto, puede demostrarse (Grenander y Rosenblatt, 1957), que dado un proceso estocástico estacionario, la estimación mediante la esperanza condicionada de un valor ausente,  $z_h$ , viene dada por:

$$\hat{z}_h = E(z_h | Z_{(h)}) = - \sum_{j=1}^{\infty} \rho_j^i (z_{h+j} + z_{h-j}) \quad (12.22)$$

donde  $\rho_j^i$  es la función de autocorrelación inversa del proceso, definida en el apéndice 5.1. Para un AR(1) esta función es la función de autocorrelación de un MA(1), por lo que  $\rho_1^i = -\phi/(1+\phi^2)$  y  $\rho_j^i = 0$  para  $j > 1$ . Sustituyendo en (12.22):

$$\hat{z}_h = E(z_h | Z_{(h)}, AR(1)) = \frac{\phi}{(1+\phi^2)}(z_{h+1} + z_{h-1}).$$

Este estimador es una combinación de las dos fuentes de información disponibles sobre  $z_h$ , el valor anterior y el posterior. En efecto, la predicción de  $z_h$  que podemos obtener desde su pasado es  $\phi z_{h-1}$ , con varianza  $\sigma^2$ . Desde su futuro podemos obtener otra predicción mediante la relación  $z_{h+1} = \phi z_h + a_{h+1}$ , que implica:

$$z_h = \phi^{-1} z_{h+1} - \phi^{-1} a_{h+1},$$

y conduce a la predicción  $\phi^{-1} z_{h+1}$  con varianza  $\phi^{-2} \sigma^2$ . La predicción óptima

combinará estas dos fuentes de información con peso proporcional a su precisión (inversa de la varianza), para obtener:

$$\begin{aligned}\hat{z}_h &= \frac{\sigma^{-2}}{\sigma^{-2} + \phi^2\sigma^{-2}} \phi z_{h-1} + \frac{\phi^2\sigma^{-2}}{\sigma^{-2} + \phi^2\sigma^{-2}} \phi^{-1} z_{h+1} = \\ &= \frac{\phi}{(1 + \phi^2)} (z_{h+1} + z_{h-1}).\end{aligned}$$

Sustituyendo el estimador (12.21) en la verosimilitud (12.20) se obtiene una función de verosimilitud que no depende ahora del parámetro  $w$ :

$$l(\phi, \sigma^2 | z_1) = -\frac{(T-1)}{2} \ln \sigma^2 - \sum_{t \in A} \frac{(z_t - \phi z_{t-1})^2}{2\sigma^2} - \frac{(z_{h+1} - \phi^2 z_{h-1})^2}{2\sigma^2(1 + \phi^2)}, \quad (12.23)$$

y las funciones (12.20) y (12.23) serán básicamente equivalentes para tamaño muestral moderado, ya que sólo difieren en el factor  $T-1$  o  $T-2$  y en un término,  $\frac{1}{2} \ln(1 + \phi^2)$ , que no depende del tamaño muestral y cuyo efecto está acotado. Para tamaños muestrales grandes su efecto será muy pequeño y ambos procedimientos llevarán a estimadores prácticamente idénticos. Puede demostrarse que este resultado es general. Gómez, Maravall y Peña (1999) estudian su implantación y comparan este método con un procedimiento exacto para resolver el problema que utiliza el filtro de Kalman.

Para cualquier proceso ARIMA el proceso inverso es estacionario, aunque para un proceso no estacionario el proceso inverso será no invertible. Por tanto, en todo proceso ARIMA la función de autocorrelación inversa está bien definida y Peña y Maravall (1991) demostraron que (12.22) es válida también para calcular el interpolador óptimo en procesos no estacionarios.

## Apéndice 12.2: Estimación de efectos de intervención

### TSW

Veamos cómo estimar una serie con análisis de intervención con el programa TSW (Tramo-Seats de windows). Este programa identifica automáticamente estos efectos por el procedimiento de detección de atípicos que explicaremos en el capítulo siguiente. Primero se introduce la serie, seleccionando en el menú **Series** y a continuación seleccionando la serie. En **Series list**, podemos ver que la serie se ha incluido en el programa. Si marcamos la serie, aparece el nombre de la serie y ciertos atributos que podemos cambiar. Por ejemplo, si cambiamos el parámetro **Iter** a **Iter=1**, a la misma serie le podemos ajustar diferentes modelos. A continuación se marca **+Model** y se introduce el modelo a estimar. El parámetro **RSA** lo dejamos en 0. Marcamos **Others** para nuevos parámetros. De todos estos valores, sólo nos interesan los correspondientes a **Outliers**. Por ahora marcamos **IATIP=1**, lo que significa que además

## 12. Análisis de intervención

---

de estimar los parámetros, busque atípicos y estime sus efectos. Marcamos OK. Marcamos el modelo en la tabla y marcamos Run. En el fichero output aparecen los efectos estimados.

### EViews

Para estimar un cambio de nivel se genera la variable ficticia escalón como una serie. Por ejemplo para generar un escalón desde la observación 50 a la 100 se hace: *Genr → s1=0* y *Genr → s1=1, sample: 50 100*. A continuación estimamos el modelo incluyendo esta variable. Por ejemplo, se desea estimar un modelo ARIMA(0,1,1) con un cambio de nivel en la observación 50. Para ello y teniendo en cuenta que tenemos que tomar una diferencia en los dos miembros de la ecuación, el modelo a estimar es: *Quick → Estimate Equation → d(x1,1) d(s1,1) ma(1)*.

Para definir una variable impulso en el instante 55, se hace: *Genr → s2=0*, y *Genr → s2=1, sample: 55 55*, y podemos estimar el modelo con:

*Procs → Especify/Estimate → d(x1,1) d(s1,1) d(s2,1) ma(1)*.

---

## 13. Valores atípicos



George C. Tiao (1933)

Estadístico chino. Realizó el doctorado en la Universidad de Wisconsin con George Box y ha sido profesor en las Universidades de Wisconsin y Chicago, donde es actualmente profesor emérito. Ha hecho contribuciones destacadas al estudio de la estacionalidad, los atípicos en series temporales, la incorporación de efectos deterministas, la predicción y los modelos para series temporales múltiples. Ha sido miembro destacado del equipo científico interdisciplinario que descubrió los cambios en la capa de ozono de la Tierra con técnicas de series temporales.

### 13.1. Introducción

Hemos visto en el capítulo anterior cómo tener en cuenta efectos deterministas conocidos sobre una serie. Sin embargo, con mucha frecuencia ocurren en las series reales hechos puntuales que desconocemos. Por ejemplo, la serie puede haber estado sometida a cambios de base, errores de medición, etc. Las observaciones afectadas por estas intervenciones pueden presentar una estructura distinta de las demás y aparecer como datos atípicos, es decir, datos que aparentemente no han sido generados igual que las demás. En particular, si se comete un error de medida o de transcripción en un dato, este valor no mantendrá la estructura de dependencia del resto de la serie.

Es importante ser capaz de identificar estas situaciones desconocidas y separarlas de la dinámica habitual de la serie porque:

1. Si sus efectos son grandes, pueden sesgar la estimación de los parámetros, lo que producirá malas predicciones futuras.

2. Si el suceso ha ocurrido en la última parte de la serie y alguna observación afectada se utiliza para generar predicciones estas no serán buenas, incluso aunque los parámetros estén bien estimados.
3. Si estos sucesos atípicos pueden volver a aparecer en el futuro y los identificamos y estimamos sus efectos, podemos incorporar esta información en las predicciones y obtener intervalos de predicción más realistas.

En este capítulo consideraremos un valor como atípico si tiene pocas probabilidades de ocurrir dada la estructura habitual de evolución de la serie. Entonces, la serie puede considerarse como la suma de dos series distintas. La primera es la generada por las circunstancias habituales, entendiendo por ello las que están presentes en el 95 % o 99 % de los datos observados. La segunda es la formada por datos que corresponden a situaciones peculiares específicas. Como el modelo que se aplica en las circunstancias habituales puede ser muy distinto del que se aplica en las situaciones específicas, conviene separar ambas situaciones y estudiarlas separadamente. La identificación de atípicos consiste en detectar observaciones que parecen haber sido generadas de forma distinta al resto. Posteriormente, habrá que investigar las causas de estos valores atípicos y, si se desea, incorporar estas observaciones en un modelo explícito.

A continuación vamos a estudiar cómo identificar la ocurrencia de estos sucesos atípicos por las huellas que han dejado en la serie. Podemos considerar dos casos básicos de datos atípicos: los aditivos, que son análogos a los atípicos que ocurren en datos estáticos, como por ejemplo un error de medición o de transcripción de un dato, y los innovativos, que son específicos de situaciones dinámicas. Una serie puede sufrir también otros efectos, como cambios de nivel o de tendencia.

### 13.2. Atípicos aditivos

Diremos que ha ocurrido un atípico aditivo (AO) sobre una serie temporal en el instante  $h$  si el valor de la serie se genera en ese instante de manera distinta al resto. Por ejemplo, si existe un error de medida apreciable en el instante  $h$  que no aparece en las otras observaciones, el dato  $z_h$  será un atípico aditivo en la serie  $z_t$ . En general, si el dato  $z_h$  tiene propiedades distintas del resto por un cambio en las condiciones externas o del sistema de medición, diremos que este dato es atípico. El modelo que seguirá una serie observada,  $z_t$ , afectada por un AO en  $t = h$  será:

$$z_t = \begin{cases} y_t & t \neq h \\ y_t + \omega_A & t = h. \end{cases}$$

donde  $y_t$  es la serie no contaminada por atípicos que suponemos sigue un modelo ARIMA:

$$y_t = \psi(B)a_t.$$

Entonces, el modelo que sigue la serie observada,  $z_t$ , es

$$z_t = \omega_A I_t^{(h)} + \psi(B)a_t, \quad (13.1)$$

donde  $I_t^{(h)} = 0$ ,  $t \neq h$ ;  $I_h^{(h)} = 1$ . Este modelo es idéntico al de análisis de intervención con un impulso. La diferencia es que en el capítulo anterior la variable  $I_t^{(h)}$  se supone conocida, mientras que ahora el instante  $h$  es desconocido. Una forma equivalente de escribir el modelo ARIMA para la serie  $y_t$  es

$$\pi(B)y_t = a_t$$

donde se verifica que  $\psi(B)\pi(B) = 1$ , y el modelo resultante para un AO es:

$$\pi(B)(z_t - \omega_A I_t^{(h)}) = a_t. \quad (13.2)$$

Las ecuaciones (13.1) y (13.2) son intercambiables y cualquiera de las dos nos sirve para definir el AO.

### 13.2.1. Efectos en los residuos

Un atípico aditivo deja una huella en la serie debido a la alteración de su valor en un punto. Cuando desconozcamos su presencia y construyamos un modelo ARIMA, esta alteración puede detectarse con los residuos del modelo. Vamos a suponer que conocemos los verdaderos parámetros del proceso y, para simplificar, supondremos primero que el modelo es un AR(1). Entonces, si desconocemos la presencia de un atípico y calculamos los residuos suponiendo que todas las observaciones han sido generadas por el mismo modelo, mediante:

$$e_t = z_t - \phi z_{t-1},$$

donde hemos llamado  $e_t$  a los residuos que contienen el efecto del atípico para diferenciarlos de las innovaciones verdaderas,  $a_t$  que siguen un proceso de ruido blanco. La relación entre estos residuos y las innovaciones se obtiene utilizando que, según (13.2):

$$(1 - \phi B)z_t = \omega_A(1 - \phi B)I_t^{(h)} + a_t,$$

y los residuos,  $e_t = (1 - \phi B)z_t$ , están relacionados con las verdaderas innovaciones,  $a_t$ , por:

$$e_t = \omega_A x_t + a_t,$$

donde la variable  $x_t = I_t^{(h)} - \phi I_t^{(h+1)}$  toma siempre el valor cero salvo en  $x_h = 1$ ,  $x_{h+1} = -\phi$ . Podemos concluir que antes de la ocurrencia del suceso atípico en  $t = h$  los residuos  $e_t$  serán iguales a las verdaderas innovaciones  $a_t$ . Sin embargo, a partir del instante  $h$  la relación entre ambas variables es:

$$\begin{aligned} e_h &= \omega_A + a_h, \\ e_{h+1} &= -\phi\omega_A + a_{h+1}, \\ e_{h+j} &= a_{h+j}, \quad j \geq 2. \end{aligned}$$

Este análisis muestra que un AO en un AR(1) modifica dos residuos. El más afectado es en  $t = h$ , donde el residuo viene modificado por el efecto de la intervención. El residuo posterior también estará afectado, pero en sentido contrario, y por una magnitud que es el producto del tamaño del atípico y el parámetro ( $|\phi| < 1$ ).

Vamos a generalizar este resultado para un AR( $p$ ). Entonces, el modelo es:

$$\phi(B)z_t = \omega_A\phi(B)I_t^{(h)} + a_t,$$

donde  $\phi(B)$  es el operador AR de orden  $p$ . La relación entre los residuos calculados con los datos observados, que contienen el efecto de los atípicos, y las innovaciones, puede de nuevo escribirse como:

$$e_t = \omega_A x_t + a_t, \quad (13.3)$$

donde, utilizando que según vimos en el capítulo anterior  $B^j I_t^{(h)} = I_t^{(h+j)}$ , ahora la variable  $x_t$  está dada por:

$$x_t = I_t^{(h)} - \phi_1 I_t^{(h+1)} - \dots - \phi_p I_t^{(h+p)} + a_t,$$

y es cero en todo momento salvo para  $x_h = 1$ ,  $x_{h+1} = -\phi_1, \dots, x_{h+p} = -\phi_p$ . Por tanto, los  $p$  residuos posteriores a  $h$  estarán afectados, según la relación

$$e_{h+j} = a_{h+j} - \phi_j \omega_A, \quad j \geq 0$$

con  $\phi_0 = -1$ . Es importante señalar que ahora es posible que el mayor valor del efecto ocurra en un instante posterior a la ocurrencia del atípico. Por ejemplo, en el AR(2) estacionario con operador  $(1 - 1.7B + .72B^2)$  el efecto de un atípico aditivo en  $h$  sobre las innovaciones será  $\omega_A$  en  $h$  pero  $-1.7\omega_A$ , es decir un 70 % más fuerte, en el instante  $h + 1$ . Concluimos que en un AR( $p$ ) los  $p$  residuos posteriores estarán afectados de manera compleja, que depende de los parámetros AR, y que la suma de todos los efectos sobre los residuos es  $\omega_A(1 - \phi_1 - \dots - \phi_p) = \omega_A\phi_p(1)$ . Como en un proceso no estacionario  $\phi_p(1) = 0$  (por tener una raíz unitaria) en

estos procesos la suma de los efectos sobre los residuos debe ser cero, y los efectos positivos deben compensarse con los negativos.

Estos resultados se generalizan sin dificultad para procesos ARIMA. Partiendo de (13.2), los residuos serán  $e_t = \pi(B)z_t$ , y utilizando (13.2), la relación entre los residuos calculados en la serie contaminada,  $e_t$ , y las verdaderas innovaciones,  $a_t$ , es como en (13.3), pero ahora:

$$x_t = \pi(B)I_t^{(h)} = -\sum_{j=0}^{T-h} \pi_j I_t^{(h+j)}, \quad (13.4)$$

con  $\pi_0 = -1$ . Esta ecuación indica que tendremos un número de residuos afectados por el atípico igual al orden del polinomio  $\pi(B)$ . En efecto, de (13.3) y (13.4) obtenemos que:

$$e_{h+j} = -\pi_j \omega_A + a_{h+j} \quad j \geq 0$$

y si todos los  $\pi_j$  son no nulos, todos los residuos posteriores al instante de ocurrencia del atípico estarán afectados.

En todo lo anterior hemos supuesto que hay un único AO en la serie. Cuando hay varios, sus efectos se superponen y pueden modificar mucho la estructura de la serie. Por ejemplo, un proceso AR(5) que sufre un 5 % de atípicos aditivos puede llegar a tener el 25 % de los residuos afectados.

### 13.2.2. Efectos en la estimación de los parámetros

El efecto de un AO es sesgar los coeficientes de autocorrelación y la estimación de los parámetros del proceso hacia cero. Ilustraremos este efecto en el caso de un proceso AR(1) de media cero donde ocurre un AO en el instante  $h$ . La estimación del parámetro será:

$$\hat{\phi} = \frac{\sum(z_t - \bar{z})z_{t-1}}{\sum(z_t - \bar{z})^2},$$

utilizando que  $z_h = y_h + \omega_A$ , y  $z_t = y_t$ ,  $t \neq h$ , entonces  $\bar{z} = \bar{y} + \frac{1}{T}\omega_A$ . Suponiendo que  $T$  es grande, de manera que podemos suponer  $\bar{z} \approx \bar{y} \approx 0$ , podemos escribir,

$$\hat{\phi} = \frac{\sum y_t y_{t-1} + \omega_A(y_{h-1} + y_{h+1})}{\sum y_t^2 + \omega_A^2 + 2\omega_A y_h}.$$

En esta ecuación el tamaño del atípico aparece linealmente en el numerador y al cuadrado en el denominador. Esto sugiere que cuando  $\omega_A$  sea muy grande el valor de  $\hat{\phi}$  será muy pequeño. Para precisar esta intuición, dividiendo ambos miembros por  $\sum y_t^2 = Ts_y^2$  y llamando  $\hat{\phi}_0 =$

$\sum y_t y_{t-1} / \sum y_t^2$  a la estimación obtenida con la serie sin atípicos, tenemos que

$$\hat{\phi} = \frac{\hat{\phi}_0 + T^{-1} \tilde{\omega}_A (\tilde{y}_{h-1} + \tilde{y}_{h+1})}{1 + T^{-1} \tilde{\omega}_A (2\tilde{y}_h + \tilde{\omega}_A)},$$

donde hemos llamado  $\tilde{\omega}_A = \omega_A/s_y$  y  $\tilde{y}_t = y_t/s_y$  a los valores estandarizados por la desviación típica de la serie. Esta expresión muestra que si el atípico es grande con relación a la variabilidad de la serie, el coeficiente estimado irá a cero, ya que si  $\tilde{\omega}_A \rightarrow \infty$  tendremos que  $\hat{\phi} \rightarrow 0$ .

En general el efecto de un atípico aditivo depende mucho del tamaño muestral y para tamaños muestrales muy grandes un AO de tamaño moderado puede tener un efecto pequeño.

### Ejemplo 13.1

La figura 13.1 presenta una muestra de 100 observaciones de un proceso AR(1), con parámetro .6, contaminado por un AO en el instante 60 de seis desviaciones típicas. La figura 13.2 muestra la función de autocorrelación estimada con estos datos.

Se observa que todos los coeficientes de autocorrelación de los primeros retardos tienen un sesgo hacia cero. Como el parámetro AR es .6 y  $.6^6 = .04$ , para retardos altos los coeficientes reales son prácticamente cero, pero incluso en estos valores la figura muestra que las estimaciones del proceso contaminado se aproximan más a cero en valor absoluto que los del proceso original.

### 13.3. Atípicos innovativos (IO)

Diremos que ha ocurrido un atípico innovativo (IO) en una serie temporal en  $t = h$  cuando la innovación en ese punto está directamente afectada por una cantidad desconocida debida a un suceso imprevisto. Recordemos que la innovación de una serie temporal representa el efecto agregado no sistemático de todas las variables que influyen sobre la serie. Un atípico sobre la innovación puede interpretarse como un cambio imprevisible en una o varias de las variables que afectan a la serie. Como la innovación representa el error de previsión a un paso, podemos decir que un IO es un incremento del error de previsión de la serie en un punto debido a un cambio en las condiciones externas.

El modelo para la serie que sufre un atípico innovativo de magnitud  $\omega_I$  en el instante  $h$  es:

Análisis de series temporales

---

Figura 13.1 Efecto de un AO sobre un AR(1)

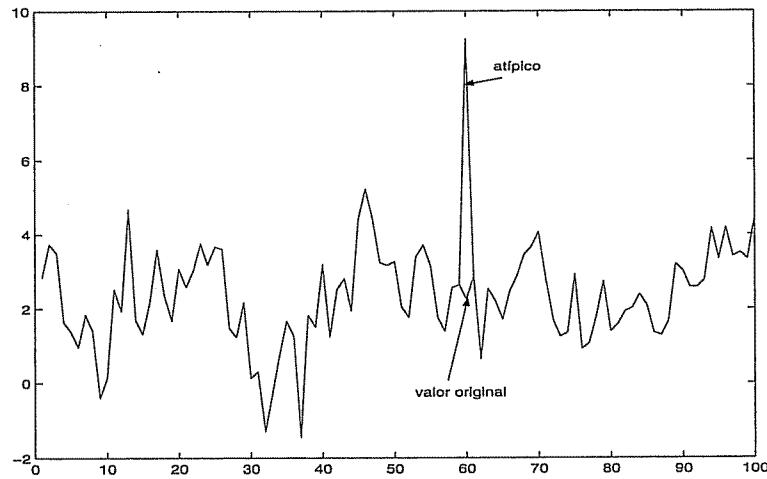
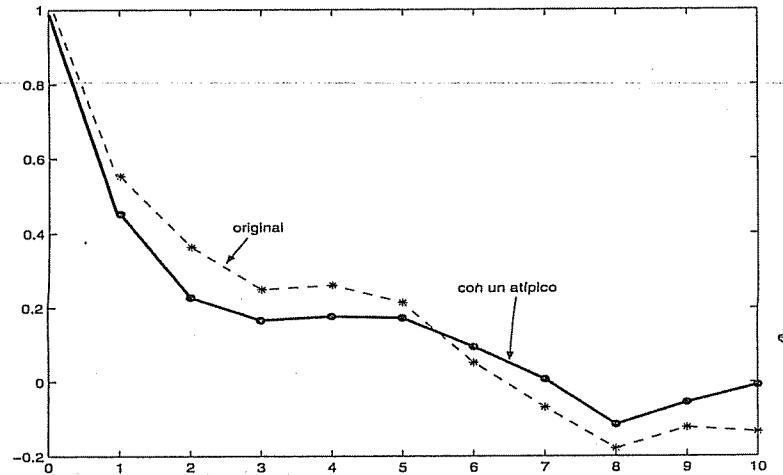


Figura 13.2 Función de autocorrelación del proceso AR(1) contaminado por un atípico



$$z_t = \psi(B)(\omega_I I_t^{(h)} + a_t) \quad (13.5)$$

o en la representación AR( $\infty$ ) :

$$\pi(B)z_t = \omega_I I_t^{(h)} + a_t,$$

y llamando  $e_t = \pi(B)z_t$  como en la sección anterior a los residuos calculados suponiendo que la serie está libre de atípicos y que conocemos los parámetros del modelo, la relación entre estos residuos y las verdaderas innovaciones es:

$$e_t = \omega_I I_t^{(h)} + a_t, \quad (13.6)$$

que podemos escribir como:

$$e_h = a_h + \omega_I, \quad (13.7)$$

y para  $j > h$

$$e_{h+j} = a_{h+j}.$$

Por tanto, cuando ocurre un IO en  $t = h$  los residuos estimados conociendo los parámetros del proceso serán igual a las innovaciones en todos los puntos, menos en el de la ocurrencia del IO. De este resultado se desprende que un AO y un IO son idénticos en una serie de ruido blanco. En otro caso, el efecto de un IO sobre la serie es muy distinto que el de un AO, ya que este último produce un efecto fijo, alteración de una observación, mientras que el efecto de un IO sobre la serie depende del modelo, como veremos a continuación.

Observemos que tanto en el caso AO como el IO los residuos observados están relacionados con las innovaciones por la ecuación:

$$e_t = \omega x_t + a_t$$

donde la serie  $x_t$  es cero antes de la ocurrencia del atípico, toma el valor uno en el instante de ocurrencia del atípico y para un AO incluye después los parámetros del proceso, mientras que para un IO es cero.

### 13.3.1. Efecto de un IO sobre la serie

Supongamos que la serie es estacionaria. Utilizando la definición (13.5) y escribiendo

$$z_t = y_t + \psi(B)\omega_I I_t^{(h)},$$

donde  $y_t = \psi(B)a_t$  es la serie sin contaminar que sigue el modelo ARMA. Según esta ecuación, la relación entre las observaciones contaminadas,  $z_t$ , y las originales,  $y_t$ , es:

$$z_t = \begin{cases} y_t & t < h \\ y_t + \omega_I \psi_j & t = h + j, \quad j \geq 0. \end{cases} \quad (13.8)$$

Esta expresión indica que

1. Para  $t = h$  el valor de la serie es siempre el valor original más la

magnitud del atípico.

2. Si la serie es estacionaria y sigue un modelo  $MA(q)$ , los  $q$  valores siguientes de la serie estarán afectados por una magnitud que depende de los coeficientes de la media móvil. En particular, si uno de estos coeficientes es mayor que la unidad, lo que es posible si  $q > 1$ , el valor más afectado de la serie ocurrirá en ese punto.
3. Si la serie es estacionaria y tiene componente AR, todos los valores posteriores al instante en que ocurre el IO estarán afectados, aunque con pesos globalmente decrecientes, ya que los coeficientes  $\psi_j$  van a cero en una serie estacionaria al aumentar el retardo. Sin embargo, como  $\psi_j$  puede ser mayor que uno, es posible, como en el caso del MA, que el valor más afectado de la serie ocurra después del instante de aparición del atípico.
4. Cuando la serie es no estacionaria el efecto de un IO puede ser muy complejo. Supongamos que la serie es  $I(1)$ . Entonces, el modelo es

$$\phi(B)\nabla z_t = \theta(B)(\omega_I I_t^{(h)} + a_t),$$

hemos visto en el capítulo anterior que un impulso es el resultado de diferenciar un escalón,  $I_t^{(h)} = \nabla S_t^{(h)}$ , y podemos escribir esta ecuación como:

$$z_t = \phi^{-1}(B)\theta(B)\omega_I S_t^{(h)} + y_t$$

donde  $y_t = \nabla^{-1}\phi^{-1}(B)\theta(B)a_t$  sigue el proceso ARIMA. Este modelo supone la superposición de un cambio de nivel y otros efectos sobre la serie, como veremos en la sección siguiente. En particular, si  $\phi(B) = \theta(B) = 1$  y el modelo es un paseo aleatorio, tendremos un cambio de nivel puro. Este resultado se obtiene también con (13.8) ya que entonces,  $\psi_j = 1$  para  $j \geq 1$  y todos los valores posteriores a  $h$  están afectados en la misma magnitud, con lo que los valores posteriores al IO sufren un escalón que les sitúa a un nivel  $\omega_I$  unidades más altas que los anteriores.

5. Si la serie es  $I(2)$ , por ejemplo del tipo  $\nabla^2 z_t = a_t$ , igualando coeficientes en  $\psi(B)\nabla^2 = 1$  obtenemos que los coeficientes  $\psi_j$  son  $\psi_1 = 2$ ,  $\psi_2 = 3$ , y en general,  $\psi_j = j+1$ . La serie  $z_t$  estará afectada por lo que se denomina un efecto rampa, que estudiaremos en las secciones siguientes. La serie sufrirá un crecimiento lineal determinista con pendiente  $\omega_I$  desde el instante en que aparece el atípico. Si, además, la serie es estacional, como por ejemplo la serie  $\nabla\nabla_{12} z_t = a_t$ , el

efecto de un atípico IO es un conjunto de cambios de nivel de la misma magnitud.

Concluimos que los atípicos innovativos pueden producir estructuras muy complejas sobre la serie y con series no estacionarias pueden confundirse fácilmente con otros efectos: cambio de nivel, para series I(1) y cambio de tendencia, para series I(2).

### Ejemplo 13.2

Dada una serie que sigue el modelo de los pasajeros de avión,  $\nabla\nabla_{12}z_t = (1 - \theta B)(1 - \Theta B^{12})a_t$ , calcular el efecto esperado de un atípico innovativo sobre los datos observados.

Para ver el efecto sobre la serie de un IO necesitamos la representación de media móvil, que se obtiene igualando potencias de  $B$  en:

$$\psi(B)\nabla\nabla_{12} = (1 - \theta B)(1 - \Theta B^{12}).$$

Escribiendo en esta ecuación las inversas de los operadores diferencia como operadores de suma indefinida:

$$\psi(B) = (1 + B + B^2 + \dots)(1 + B^{12} + B^{24} + \dots)(1 - \theta B)(1 - \Theta B^{12}).$$

lo que resulta, multiplicando la parte regular y la estacional, en:

$$\psi(B) = (1 + (1 - \theta)B + (1 - \theta)B^2 + \dots)(1 + (1 - \Theta)B^{12} + (1 - \Theta)B^{24} + \dots)$$

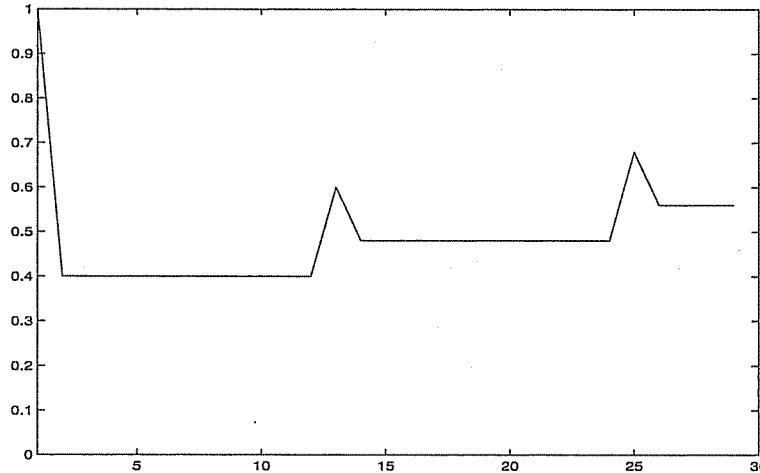
y operando:

$$\begin{aligned}\psi(B) = & 1 + (1 - \theta)B + (1 - \theta)B^2 + \dots ((1 - \Theta) + \\ & (1 - \theta))B^{12} + ((1 - \Theta)(1 - \theta) + (1 - \theta))B^{13} + \dots\end{aligned}$$

El efecto sobre la serie de un IO de tamaño 1 se representa en la figura 13.3. En el instante en que ocurre el efecto es igual al tamaño del IO, pero en los períodos siguientes el efecto cae a  $(1 - \theta)$  y se convierte en un cambio de nivel de magnitud  $(1 - \theta)$ . Al llegar al período 12 se superpone al escalón un valor de magnitud  $(1 - \Theta)$  y, a continuación, todas las observaciones de la 13 a la 23 sufren un escalón de magnitud  $(1 - \Theta)(1 - \theta) + (1 - \theta)$ . En el período 24 se superpone un valor de magnitud  $(1 - \Theta)$  y se repite la estructura como indica la figura 13.3.

Por ejemplo, si  $\Theta = .8$  y  $\theta = .6$ , y el IO ocurre cuando quedan 36 observaciones para el final de la serie, el efecto es: en el instante de ocurrencia el valor uno, después tendrá un escalón de tamaño .4. En la posición 12 tendrá un valor de .6 y entre la 13 y la 23 un escalón de tamaño  $.4 + .2 \times .4 = .48$ . Entre la

**Figura 13.3 Efectos sobre los siguientes 30 valores posteriores a su ocurrencia de un atípico innovativo en una serie que sigue el modelo de pasajeros de avión**



25 y la 35 un escalón de tamaño .56. Observemos que este efecto tan complejo puede ser aproximado por un AO en la posición 1, seguido de un cambio de nivel de alrededor de .5.

---

### 13.3.2. Efectos en la estimación

El efecto de un IO en la estimación de los parámetros depende del modelo para la serie. Para series estacionarias los efectos de un IO sobre los parámetros estimados son menores que en el caso de los AO, pero para series no estacionarias pueden ser importantes. Para ilustrar los efectos en el caso estacionario supongamos un modelo AR(1) con media cero. Entonces, el estimador del parámetro se calcula con:

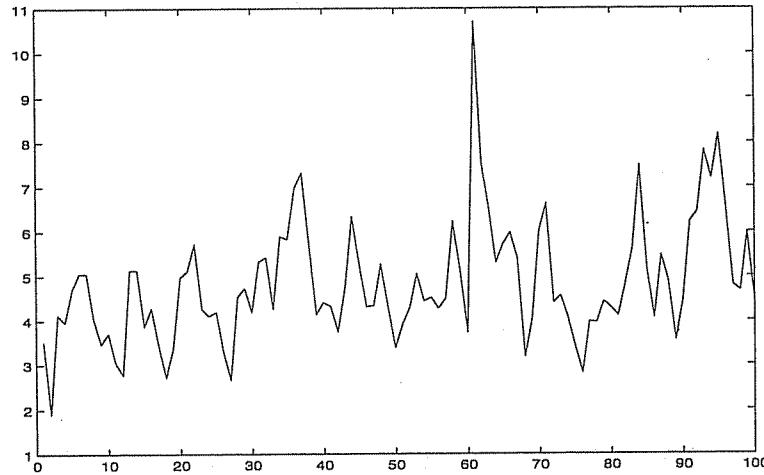
$$\hat{\phi} = \frac{\sum z_t z_{t-1}}{\sum z_t^2}$$

y reemplazando para  $t = h + j$ ,  $z_t$  por  $y_t + \omega_I \psi_j$ , este estimador puede escribirse:

$$\hat{\phi} = \frac{\hat{\phi}_0 + T^{-1} \tilde{\omega}_I S_1 + T^{-1} \tilde{\omega}_I^2 (\phi / (1 - \phi^2))}{1 + T^{-1} 2 \tilde{\omega}_I S_2 + T^{-1} \tilde{\omega}_I^2 / (1 - \phi^2)},$$

donde  $\hat{\phi}_0 = \sum y_t y_{t-1} / \sum y_t^2$  es la estimación obtenida de la serie sin atípicos y  $S_1 = \sum_{j=0}^T (\tilde{y}_{h-1+j} + \tilde{y}_{h+1+j}) \phi^j$  y  $S_2 = \sum_{j=0}^T \tilde{y}_{h+j} \phi^j$ . Esta expresión

**Figura 13.4 Un AR(1) con parámetro .6 y un IO en la posición 60 igual a seis desviaciones típicas de los residuos**



indica que si  $\tilde{\omega}_I \rightarrow \infty$  entonces  $\hat{\phi} \rightarrow \phi$ , y no existen sesgos asintóticos al estimar el parámetro.

### Ejemplo 13.3

La figura 13.4 presenta una serie AR(1) simulada con parámetro .6 contaminada por un IO en la posición 60. Se observa que los valores posteriores al instante 60 de la serie también están afectados, aunque con peso decreciente. La figura 13.5 presenta la función de autocorrelación y se ve que está muy poco afectada por el atípico.

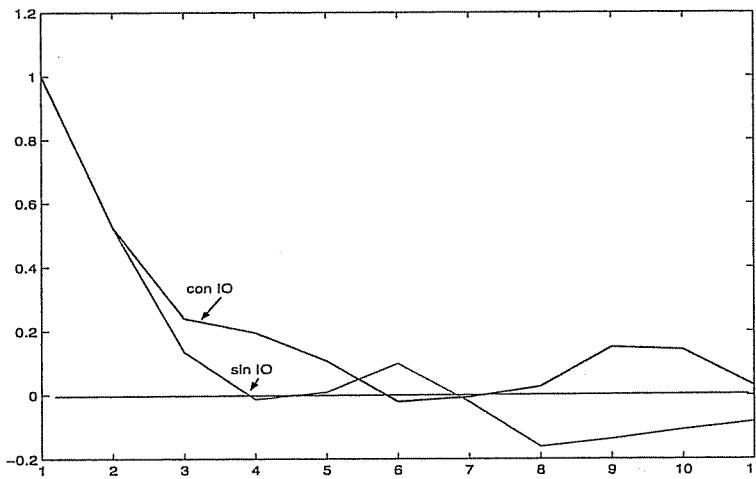
### 13.4. Cambio de nivel

Otro efecto importante que puede aparecer sobre una serie temporal es el cambio de nivel. Diremos que una serie ha sufrido un cambio de nivel en el instante  $h$  si sigue el modelo:

$$z_t = \omega_L S_t^{(h)} + \psi(B)a_t,$$

donde  $S_t^{(h)}$  es la variable escalón, definida en el capítulo 12 por  $S_t^{(h)} = 1$  si  $t \geq h$  y cero en otro caso. Los valores de la serie observada estarán

**Figura 13.5 Dos realizaciones de un proceso AR(1), una con un IO y otra sin contaminar**



relacionados con la serie sin contaminar por el cambio de nivel mediante:

$$z_t = \begin{cases} y_t & t < h \\ y_t + \omega_L & t \geq h. \end{cases}$$

La figura 13.6 muestra una serie con un cambio de nivel en la mitad del periodo muestral.

Si el proceso es estacionario, un cambio de nivel convertirá a la serie en no estacionaria, ya que la esperanza de cada observación será  $\mu$  antes del cambio de nivel y  $\mu + \omega_L$  después. Si la serie es no estacionaria el cambio de nivel puede ser equivalente a un IO, como ya hemos visto. En efecto, un paseo aleatorio con un cambio de nivel sigue el modelo:

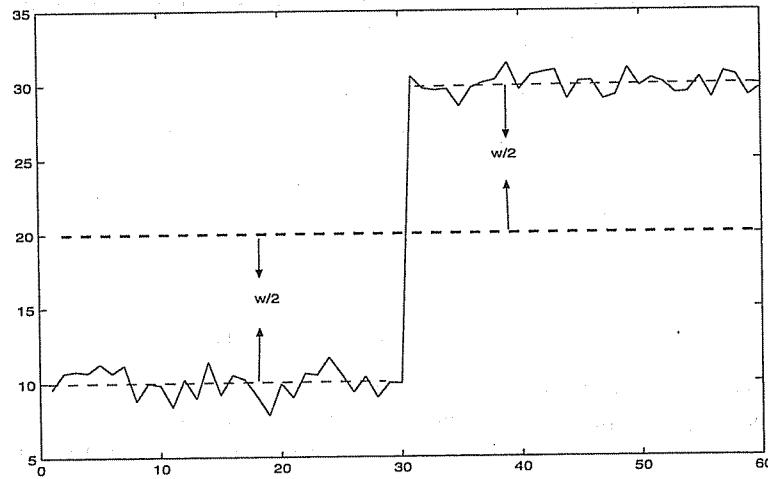
$$z_t = \omega_L S_t^{(h)} + \nabla^{-1} a_t$$

y tomando diferencias, como al diferenciar un escalón se obtiene un impulso, el modelo anterior puede también escribirse como:

$$\nabla z_t = \omega I_t^{(h)} + a_t$$

que es el modelo de un IO para esta serie.

**Figura 13.6 Una serie estacionaria con un cambio de nivel en la mitad del periodo de observación**



#### 13.4.1. Efectos en los residuos

Escribiendo el cambio de nivel como:

$$\pi(B)(z_t - \omega_L S_t^{(h)}) = a_t,$$

vemos que, como en los casos anteriores, los residuos calculados con los parámetros verdaderos están relacionados con las innovaciones mediante:

$$e_t = \omega_L x_t + a_t,$$

donde  $x_t = \pi(B)S_t^{(h)}$  es una variable que toma el valor cero antes de la intervención, el valor uno en el instante  $t = h$  y los valores  $1 - \sum_{i=1}^j \pi_i$  para  $t = h + j$ , donde  $j = 1, \dots, T - h$ . Por tanto:

$$e_t = \begin{cases} a_t & t < h \\ a_t + \omega_L(1 - \sum_{i=1}^j \pi_i) & t = h + j. \end{cases} \quad (13.9)$$

Esta expresión muestra que todos los residuos después de un cambio de nivel están afectados, pero que el efecto depende: (1) del modelo, siendo mayor el efecto para modelos estacionarios que para los no estacionarios; (2) de la distancia entre el momento de aparición,  $h$ , y el final de la serie.

### 13.4.2. Efecto en la estimación

Veamos el efecto de un cambio de nivel sobre un proceso estacionario. Supongamos que el proceso es un AR(1). Puede comprobarse que la estimación del parámetro puede escribirse como:

$$\hat{\phi} = \frac{\hat{\phi}_0 + T^{-1}S_3 + T^{-1}\tilde{\omega}_L^2(T-h)}{1 + T^{-1}2S_4 + T^{-1}\tilde{\omega}_L^2(T-h+1)},$$

donde  $\tilde{\omega}_L = \omega_L/s_y$  y  $\tilde{y}_t = y_t/s_y$ ,  $\hat{\phi}_0 = \sum y_t y_{t-1} / \sum y_t^2$  es la estimación obtenida de la serie sin atípicos y  $S_3 = \tilde{\omega}_L(\tilde{y}_{h-1} + \tilde{y}_h + 2\tilde{y}_{h+1} + \dots + 2\tilde{y}_T)$  y  $S_4 = \tilde{\omega}_L \sum_{j=0}^h \tilde{y}_{h+j}$ . Si  $T - h$  no es muy pequeño,

$$\text{si } \tilde{\omega}_L \rightarrow \infty, \Rightarrow \hat{\phi} \rightarrow 1,$$

y el efecto del cambio de nivel es hacer que el parámetro tienda a uno y el proceso parecerá no estacionario. Cuando el proceso es no estacionario el efecto de un cambio de nivel sobre los parámetros puede ser más pequeño.

### 13.5. Cambios transitorios y efecto rampa

Un efecto adicional que puede considerarse es el cambio transitorio (TC), definido por

$$z_t = \frac{\omega_{TC}}{1 - \delta B} I_t^{(h)} + \psi(B) a_t, \quad (13.10)$$

vemos que si  $\delta = 1$  el modelo es el del cambio de nivel, LS, ya que  $\nabla^{-1} I_t^{(h)} = S_t^{(h)}$ , mientras que si  $\delta = 0$  tenemos un AO. En la práctica, este tipo de atípico se utiliza fijando el valor de  $\delta$ , habitualmente .7, de manera que no pueda confundirse con el AO o LS.

Otro tipo de efecto, de interés en series no estacionarias y especialmente en series I(2), es el efecto rampa:

$$z_t = \omega_R R_t^{(h)} + \psi(B) a_t,$$

donde la variable rampa,  $R_t^{(h)}$ , se define por:

$$R_t^{(h)} = \begin{cases} 0 & t < h \\ t + 1 - h & t \geq h. \end{cases}$$

La variable rampa introduce una tendencia determinista con pendiente  $\omega_R$  en la serie a partir del instante  $h$ . Esta variable está relacionada con las variables escalón e impulso, ya que  $\nabla R_t^{(h)} = S_t^{(h)}$  y  $\nabla^2 R_t^{(h)} = I_t^{(h)}$ .

Un efecto rampa sólo es esperable en series no estacionarias y puede

confundirse fácilmente con un IO en series con dos diferencias. En efecto, supongamos que se identifica sobre la serie estacionaria  $\nabla^2 z_t$  un IO según el modelo:

$$\nabla^2 z_t = \omega I_t^{(h)} + a_t.$$

Este modelo puede escribirse en la forma equivalente:

$$\nabla z_t = \omega S_t^{(h)} + \frac{1}{\nabla} a_t,$$

que implica que la serie  $\nabla z_t$ , que es I(1), tiene un cambio de nivel. También podemos escribir:

$$z_t = \omega R_t^{(h)} + \frac{a_t}{\nabla^2},$$

donde la serie  $z_t$ , que es I(2), sufre un efecto rampa.

Muchas series mensuales requieren la diferenciación  $\nabla \nabla_{12}$  para convertirse en estacionarias y si existe un efecto rampa:

$$z_t = \omega_R R_t^{(h)} + \frac{\psi(B)}{\nabla \nabla_{12}} a_t, \quad (13.11)$$

donde el operador  $\psi(B)$  recoge todos los operadores estacionarios regulares y estacionales. Como  $\nabla \nabla_{12} = \nabla^2 S(B)$ , donde  $S(B) = 1 + B + \dots + B^{11}$  es el operador suma y  $\nabla^2 R_t^{(h)} = I_t^{(h)}$ , podemos escribir

$$\nabla \nabla_{12} z_t = \omega_R S(B) I_t^{(h)} + \psi(B) a_t,$$

y una rampa producirá sobre la serie estacionaria una racha de doce atípicos consecutivos de la misma magnitud. Vamos a comparar esta rampa pura con un IO sobre la serie. Entonces la serie estacionaria sigue el modelo

$$\nabla \nabla_{12} z_t = \psi(B) \omega_R I_t^{(h)} + \psi(B) a_t, \quad (13.12)$$

y sobre la serie estacionaria veremos una secuencia de atípicos cuya posición y tamaño dependen de los coeficientes  $\psi(B)$ . Como estos efectos pueden confundirse, cuando aparezca un IO sobre una serie I(2) conviene siempre estimar el modelo (13.11) y compararlo con el (13.12) para ver si se trata realmente de un IO o de un efecto rampa puro.

### 13.6. Procedimientos de estimación de atípicos

En la práctica, la posición y la naturaleza de los atípicos que pueden aparecer sobre una serie es desconocida y se precisa un procedimiento

para identificarlos y detectar sus efectos. En concreto se requiere, para cada dato atípico:

1. Detectar el momento de su aparición.
2. Identificar el tipo de atípico.
3. Estimar su tamaño.

Una vez conocidos estos aspectos, es posible eliminar su efecto sobre la serie.

El procedimiento que comentamos a continuación está diseñado para encontrar atípicos del tipo aditivo, AO; innovativo, IO; cambio transitorio, TC, y cambio de nivel, LS. En los cambios transitorios se supone habitualmente que el denominador es fijo, igual a .7, con lo que el modelo para este efecto es  $w/(1 - .7B)$ , y en todos los cuatro casos el atípico queda definido por un único parámetro,  $w$ . Vamos a comentar cómo realizar los tres pasos anteriores. En primer lugar, estudiaremos cómo estimar el tamaño de un atípico cuando su posición y tipo es conocida. A continuación, nos basaremos en este resultado para presentar el procedimiento general.

### 13.6.1. Estimación del tamaño del atípico

Vamos a analizar cómo estimar el tamaño del atípico cuando suponemos conocidos los parámetros del proceso, el tipo de atípico y el momento de ocurrencia. Hemos visto que la relación entre los residuos estimados cuando existe un atípico y las innovaciones verdaderas es siempre del tipo:

$$e_t = \omega x_t + a_t, \quad (13.13)$$

donde  $\omega$  es el tamaño del atípico y la serie  $x_t$  es una variable determinista que depende del tipo de atípico y del modelo que sigue la serie. Esta es una ecuación de regresión simple, y la estimación del parámetro desconocido, que es el tamaño del atípico, se realiza de la forma habitual por mínimos cuadrados:

$$\tilde{\omega}_i = \frac{\sum e_t x_t}{\sum x_t^2}, \quad (13.14)$$

y la varianza de esta estimación es:

$$var(\tilde{\omega}_i) = \frac{\sigma_a^2}{\sum x_t^2}. \quad (13.15)$$

El resultado de aplicar esta ecuación a los distintos tipos de atípicos es el siguiente:

1. Para un atípico aditivo la variable  $x_t$  es :

$$x_t = x_t^{AO} = \pi(B)I_t^{(h)} = (0, 0, \dots, 1, -\pi_1, -\pi_2, \dots, -\pi_{T-h}), \quad (13.16)$$

que depende de los parámetros de la representación AR( $\infty$ ) del proceso. Por tanto, la estimación es:

$$\hat{\omega}_A = (e_h - \pi_1 e_{h+1} - \dots - \pi_{T-h} e_T) / \rho_A^2 \quad (13.17)$$

donde  $\rho_A^2 = (1 + \pi_1^2 + \dots + \pi_{T-h}^2)$ . Este estimador es una combinación lineal de los residuos posteriores a la ocurrencia del atípico. Para interpretar la estimación observemos que cada residuo después de  $h$  contiene información sobre el efecto del atípico, ya que, como hemos visto y queda de manifiesto en (13.13) y (13.17):

$$e_{h+j} = -\omega_A \pi_j + a_{h+j}$$

con  $\pi_0 = -1$ . De esta ecuación deducimos que el estimador para el parámetro  $\omega_A$  que proporciona el residuo  $e_{h+j}$ , que llamaremos  $\hat{\omega}_A^{(j)}$ , es  $\hat{\omega}_A^{(j)} = -e_{h+j} / \pi_j$ . Estos estimadores parciales son insesgados e independientes, ya que sus errores dependen de las variables  $a_{h+j}$  que son independientes, y cada uno de ellos tiene varianza  $\sigma^2 / \pi_j^2$ . El estimador con todos los datos será una media ponderada de todos ellos con ponderaciones iguales a las precisiones (inversas de las varianzas) de los estimadores,  $\pi_j^2 / \sigma^2$ . Podemos escribir:

$$\hat{\omega}_A = \sum_{j=1}^{T-h} \frac{\pi_j^2 / \sigma^2}{\sum \pi_j^2 / \sigma^2} \hat{\omega}_A^{(j)}$$

que es equivalente a (13.17). La varianza es  $Var(\hat{\omega}_A) = \sigma_a^2 / \rho_A^2$ .

2. Para IO la variable  $x_t$  hemos visto que es:

$$x_t = x_t^{IO} = (0, 0, \dots, 1, 0, \dots, 0), \quad (13.18)$$

y el estimador es simplemente el valor del residuo en ese punto:

$$\hat{\omega}_I = e_h,$$

esta estimación tendrá varianza  $Var(\hat{\omega}_I) = \sigma_a^2$ .

3. Para el LS, todos los residuos posteriores están afectados, de acuerdo (13.9). En la ecuación (13.13) ahora  $\omega_i = \omega_L$  y

$$x_t^{LS} = \pi(B)S_t^{(h)} = (0, \dots, 1, 1 - \pi_1, 1 - \pi_1 - \pi_2, \dots, 1 - \pi_1 - \dots - \pi_{T-h}),$$

y el estimador es:

$$\hat{\omega}_L = \frac{1}{\rho_L^2} \left( \sum_{j=h}^T e_j + -\pi_1 \sum_{j=h+1}^T e_j - \dots - \pi_{T-h} e_T \right), \quad (13.19)$$

donde  $\rho_L^2 = (1 + (1 - \pi_1)^2 + \dots + (1 - \pi_1 - \dots - \pi_{T-h})^2)$ . Por ejemplo, si  $\pi(B) = 1$ , de manera que  $\pi_i = 0$  y  $\rho_L^2 = T - h + 1$ , el estimador es la media de los residuos después del cambio,  $\bar{e}_2 = \sum_{j=h}^T e_j / (T - h + 1)$  y llamando  $\bar{e}_1 = \sum_{j=1}^{h-1} e_j / (h - 1)$  a la media de los residuos antes del cambio y utilizando que  $\sum_{j=1}^T e_j / T = 0 = (h - 1)\bar{e}_1 / T + (T - h + 1)\bar{e}_2 / T$ , podemos escribir el estimador en este caso como:

$$\hat{\omega}_L = \bar{e}_2 = \bar{e}_2 - \frac{(h - 1)\bar{e}_1 + (T - h + 1)\bar{e}_2}{T} = \frac{(h - 1)}{T} (\bar{e}_2 - \bar{e}_1)$$

y el estimador es proporcional a la diferencia de medias antes y después del cambio. Este resultado es general, y se demuestra en el apéndice 13.1 que cuando el modelo es ARIMA(0,1,1) el estimador es la diferencia entre dos medias ponderadas de las observaciones antes y después del cambio.

4. Para el TC el resultado tiene una interpretación similar.

### 13.6.2. El procedimiento general

Dada una serie temporal que sigue un modelo estimado,  $\hat{\pi}(B)z_t = \hat{e}_t$ , la detección de los posibles valores atípicos se realiza encontrando el instante y tipo más probable y contrastando si es significativo. Si lo es, se limpia de la serie y se repite el proceso hasta que el efecto (instante y tipo de atípico) más probable no sea significativo, con lo que el procedimiento termina. Para implantarlo, se procede como sigue:

1. Se identifica un modelo para la serie y se estiman los parámetros. A continuación se calculan los residuos  $\hat{e}_t = \hat{\pi}(B)z_t$ , utilizando el modelo estimado.
2. Los parámetros anteriores pueden estar mal estimados por la presencia de atípicos. Para obtener una estimación inicial más robusta de los parámetros se utiliza:

$$\hat{\sigma}_R = \frac{\text{med } |e_i|}{.675}$$

donde  $\text{med } |e_i|$  es la mediana de las desviaciones de los residuos a

su mediana y el coeficiente .675 se introduce para tener un estimador consistente de la desviación típica en poblaciones normales. Se consideran sospechosos todos los residuos con  $|e_i| > 2.5\widehat{\sigma}_R$  y sea  $\Omega$  el conjunto de instantes de tiempo donde aparecen los residuos sospechosos. Se estima el modelo

$$z_t = \sum_{t \in \Omega} w_i I_{ti}^{(h_i)} + \widehat{\pi}(B)^{-1} a_t$$

y sean  $\widehat{\pi}_R$  los parámetros estimados. Se calculan de nuevo los residuos ahora con estos parámetros mediante  $\widehat{e}_t = \widehat{\pi}_R(B)z_t$ .

3. Se comienza investigando la presencia de posibles cambios de nivel en la serie. Para ello, se calcula para cada instante temporal el contraste de razón de verosimilitudes de un cambio de nivel en ese punto, dado por:

$$\lambda_t^{LS} = \frac{\widehat{\omega}_t^{LS}}{s(\widehat{\omega}_t^{LS})},$$

donde en el denominador se utiliza  $\widehat{\sigma}_R$  como estimador de la varianza de los residuos. El instante  $t = h$  donde este estadístico toma el valor máximo es el candidato a un posible cambio de nivel. Si  $\lambda_h^{LS} > 3$ , se introduce en el modelo para la serie un cambio de nivel en ese instante, y se estima el modelo

$$z_t = wS_t^{(h)} + \pi(B)^{-1} a_t.$$

Si el cambio resulta significativo, se construye una nueva serie corregida del efecto del cambio de nivel, mediante

$$z_t^c = z_t - wS_t^{(h)}$$

se obtienen los residuos y se repite el procedimiento. Si se detecta otro cambio de nivel se estiman conjuntamente los efectos de los dos cambios de nivel mediante

$$z_t = \sum w_i S_{it}^{(h_i)} + \pi(B)^{-1} a_t$$

y se vuelve a obtener la serie corregida. Es importante que la construcción de la serie corregida debe hacerse en cada iteración incluyendo todos los cambios detectados hasta ese instante. El proceso continúa hasta que no se detectan nuevos cambios de nivel. A continuación se calculan los residuos de la serie corregida.

4. Una vez limpia la serie de cambios de nivel, se identifica la presencia

de IO, AO y TC. Si comenzamos con el AO, se estima el tamaño de un posible AO,  $\widehat{\omega}_t^{AO}$ , en todos los instantes,  $t = 1, \dots, T$  utilizando los residuos  $\widehat{e}_t$  finales de la etapa anterior. Se supone que la relación entre ellos y las innovaciones es igual a la teórica presentada en la sección anterior y se utilizan los parámetros estimados  $\widehat{\pi}$  como si fuesen los verdaderos en (13.14). Se compara la estimación del tamaño del atípico con la desviación típica de la estimación, formando el estadístico:

$$\lambda_t^{AO} = \frac{\widehat{\omega}_t^{AO}}{s(\widehat{\omega}_t^{AO})}$$

A continuación se toma el máximo de este estadístico en todos los puntos muestrales. Sea  $\lambda_{AO} = \max(\lambda_t^{AO})$  este valor máximo y supongamos que corresponde al instante  $t = t_{AO}$ .

Se repite el proceso anterior para los otros tipos de atípicos. Por ejemplo, suponiendo un IO, por (13.7) la estimación del atípico será directamente el residuo y su desviación típica es  $\sigma_a$ , que se sustituye por un estimador robusto,  $\widehat{\sigma}_a$ , como comentamos en el paso anterior. El estadístico anterior se reduce a

$$\lambda_t^{IO} = \frac{\widehat{\omega}_t^{IO}}{s(\widehat{\omega}_t^{IO})} = \frac{\widehat{e}_t}{\widehat{\sigma}_a}$$

Para el cambio transitorio, se construye el estadístico utilizando lo mismos principios:

$$\lambda_t^{CT} = \frac{\widehat{\omega}_t^{CT}}{s(\widehat{\omega}_t^{CT})}$$

y se toma el valor máximo de estos estadísticos en la muestra, que llamarémos  $\lambda_{IO} = \max(\lambda_t^{IO})$ ,  $\lambda_{LS} = \max(\lambda_t^{LS})$  y  $\lambda_{TC} = \max(\lambda_t^{TC})$ . Llamaremos  $t_{IO}$ ,  $t_{LS}$  y  $t_{TC}$  a los instantes en que ocurren estos máximos. Se considera el máximo de estos tres estadísticos:

$$\lambda_{\max} = \max(\lambda_{AO}, \lambda_{IO}, \lambda_{TC}).$$

Si  $\lambda_{\max} > c_1$ , donde  $c_1$  es un valor crítico que se obtiene por simulación, concluimos que existe en la serie un atípico del tipo y en el instante que corresponde a  $\lambda_{\max}$ . En otro caso, suponemos que no hay atípicos en la serie. El valor  $c_1$  suele tomarse como 3.25 o 3.5.

5. Si se detecta un atípico en la etapa anterior se corrige en la serie, si es AO o TC, o en los residuos si es IO. La corrección de la serie por un AO detectado en  $h$  es  $z_h^c = z_h - \widehat{\omega}_{AO}$ , y para un TC en  $h$  se aplican correcciones  $z_{h+j}^c = z_{h+j} - \widehat{\omega}_{TC}(0.7)^j$ . Con esta serie

Tabla 13.1 Modelos iniciales para Crest y Colgate

	<i>d</i>	MA	st	t	$\hat{\sigma}$
Crest	1	.7661	.0392	19.54	0.047
Colgate	1	.6648	.0448	14.85	0.045

corregida se vuelve al paso 4. Para un IO se corrige simplemente el residuo en ese punto como  $\hat{e}_h^c = \hat{e}_h - \hat{\omega}_{IO}$ , y se vuelve al paso 4. El proceso se itera hasta que no se detecten más atípicos.

6. Cuando el proceso de detección ha terminado se estiman conjuntamente los efectos de los cambios de nivel, de los atípicos detectados y los parámetros del modelo. Supongamos que se han detectado cambios de nivel en los instantes  $t_1, t_2, \dots, t_k$ , y atípicos en los instantes  $t_1^*, t_2^*, \dots, t_k^*$ . El modelo a estimar es:

$$z_t = \sum_{j=1}^k \omega_j S_t^{(t_i)} + \sum_{j=1}^k \omega_j v_{t_i}(B) I_t^{(t_i)} + \psi(B) a_t,$$

donde  $v_{t_i}(B) = 1$  si AO,  $v_{h_j}(B) = \psi(B)$  si IO y  $v_{h_j}(B) = (1 - .7)^{-1}$  si es TC. Se eliminan los efectos que no sean significativos de uno en uno y se reestima el modelo hasta que todos los efectos y coeficientes estimados sean significativos.

#### Ejemplo 13.4

Como ejemplo de este procedimiento vamos a aplicárselo a las series de cuota de mercado de Crest y Colgate. Los modelos estimados inicialmente para estas series se indican en la tabla 13.1. Vamos a realizar la búsqueda con el programa SCA que no realiza las etapas 2 y 3 sino que busca todos los posibles cambios en la etapa 4. Los resultados obtenidos con este programa se indican en la tabla 13.2.

Se observa que ambas series tienen algunos atípicos comunes. En el instante 136 ambas series sufren un cambio de nivel de dirección opuesta. Una explicación de este hecho es el hecho explicado en el ejemplo 12.1 de que en esa fecha la American Dental Association declaró que la inclusión de fluor en la pasta dentífrica hecha por Crest se había demostrado eficaz en la prevención de las caries dentales. Según la estimación del cambio de nivel, esto supuso un aumento de la cuota de mercado de Crest de 16.7 puntos porcentuales y un decrecimiento de Colgate de 10.0 puntos. Se observa que en el instante 196

Análisis de series temporales

---

Tabla 13.2 Atípicos detectados para Crest y Colgate

	Crest				
t	99	136	167	196	213
tamaño	-0.102	0.167	-0.142	-0.131	0.121
est-t	-3.07	7.04	-3.69	-3.97	3.14
tipo	TC	LS	AO	TC	AO
	Colgate				
t	43	102	136	196	
tamaño	-0.130	-0.161	-0.100	0.142	
est-t	-3.77	-3.82	-4.49	4.10	
tipop	TC	AO	LS	TC	

se produjo un movimiento de sentido opuesto, aunque de carácter transitorio, correspondiente a un esfuerzo comercial de Colgate de incluir también fluor y bajar sus precios.

La tabla 13.3 presenta los modelos estimados cuando se han tenido en cuenta estos valores atípicos. Se observa que ha aumentado el coeficiente de la media móvil como corresponde a un modelo con tendencia más regular.

### Ejemplo 13.5

Vamos a aplicar el método de detección de atípicos a la serie de accidentes (accidentes.dat) que estimamos en el capítulo anterior. El modelo estimado con EViews para esta serie es:

$$(1 + .57B + .30B^2) \nabla \nabla_{12} \ln A_t = (1 - .59B^{12} - .26B^{24}) a_t.$$

con un desviación típica residual de .0625. El gráfico de los residuos muestra muy claramente el gran efecto del 2/1988 (observación 110). Si se introduce como un cambio de nivel, sale claramente significativo. A continuación se observan atípicos aditivos en las posiciones 3/1986 y 11/1988, y el modelo resultante es:

$$\begin{aligned} \ln A_t = & -.15 I_t^{(3/1986)} - .15 I_t^{(11/1988)} + .32 S_t^{(2/1988)} + \\ & (3.7) \quad (3.5) \quad (8.11) \\ & + \frac{(1 - .54B^{12} - .30B^{24})}{(1 + .73B + .40B^2) \nabla \nabla_{12}} a_t. \end{aligned}$$

Hemos dado entre paréntesis los estadísticos  $t$  para los atípicos todos los parámetros del modelo son muy significativos. La desviación típica residual ha disminuido a .052, que supone una mejora apreciable. Probablemente los efectos de los atípicos AO detectados, del 15 %, se deben a ajustes o errores

Tabla 13.3 Modelos finales para Crest y Colgate

	$d$	MA	st	t	$\hat{\sigma}$
Crest	1	.8104	.0375	21.61	0.040
Colgate	1	.8607	.0326	26.41	0.043

en los datos de la serie publicada, porque no es esperable una reducción tan importante en un mes como consecuencia de causas aleatorias.

Como comparación presentamos la estimación automática realizada con el programa TSW, que aplica un método similar al SCA, sin las etapas 2 y 3. Con este programa, incluyendo además las tres variables deterministas, días laborables,  $D_t$ , Semana Santa,  $SS_t^{(h)}$ , y años bisiestos,  $B_t^{(h)}$ , explicadas en el capítulo anterior, el resultado obtenido es:

$$\begin{aligned} \ln A_t = & 0.0092 D_t - \frac{.078 SS_t^{(h)}}{(7.0)} + \frac{.0287 B_t^{(h)}}{(1.8)} - \frac{166 I_t^{(11/88)}}{(4.9)} + \\ & \frac{(1 - 0.39B)(1 - 0.62B^{12})}{(9.5)} \\ & \frac{.298 S_t^{(2/1988)}}{(8.46)} + \frac{a_t}{(1 - 0.16B) \nabla \nabla_{12}} \end{aligned}$$

Si comparamos ambos modelos, vemos que en ambos casos se detectan bien el cambio debido al escalón y al atípico de 11/1988. TSW no detecta, sin embargo, el efecto del otro atípico importante en 3/1988. El modelo del ruido es también diferente, por las limitaciones que comentamos de este programa para los términos estacionales.

Es interesante comparar también este modelo con el estimado para esta misma serie en el ejemplo 12.4. Ahora, además de los efectos deterministas, que estimamos entonces, se ha detectado un AO en la posición 119 que supuso un reducción brusca del número de accidentes ese mes del 16.6 %.

### Ejemplo 13.6

Vamos a analizar la serie de matriculación de vehículos (matricul.dat) que hemos estimado en los capítulos siguientes. El gráfico de los residuos mostraba varios residuos muy negativos que sugerían efectos especiales. Utilizando el procedimiento automático de identificación de atípicos implantado en TSW, el modelo obtenido es:

$$\begin{aligned} \ln M_t = & 0.40 S S_t^{(7/61)} - \frac{.54}{(4.9)} I_t^{(1/86)} - \frac{.46}{(1 - .7B)} I_t^{(1/93)} + \\ & \frac{(1 - 0.61B)(1 - 0.90B^{12})}{(1 - 0.30B^{12}) \nabla \nabla_{12}} a_t, \end{aligned}$$

donde debajo de las estimaciones de los efectos de los tres atípicos detectados se indican los estadísticos t. La parte ARIMA de este modelo es similar a la estimada en el capítulo 10. Los tres atípicos detectados corresponden a un cambio de nivel con un aumento del 40 % en las matriculaciones en la observación 19, que corresponde a 7/61, y dos cambios transitorios que corresponden a fuertes caídas en la matriculación del 54 % y 46 % que se recuperan posteriormente en las observaciones 313 y 397. Estas dos fechas son meses de enero: 1/86 y 1/93. Esto indica que en enero de estos años se produjo una caída muy fuerte en las matriculaciones, pero este efecto fue desapareciendo gradualmente en los meses siguientes. Las series de residuos del modelo univariante antes de corregir de atípicos y después de esta corrección se presentan en la figura 13.7. La serie de residuos antes de corregir presentaba dos altos residuos negativos en las observaciones 313 y 397, que en el gráfico que presentamos hecho con el programa TSW aparecen en las posiciones 300 y 384, ya que los residuos comienzan en 1 y se pierden las 13 primeras observaciones al diferenciar la serie. En los residuos del modelo que incluye las intervenciones estos efectos han desaparecido. Las *fas* y *fap* de los residuos del modelo de intervención no muestran indicios de autocorrelación o estructura, por lo que concluimos que este modelo parece adecuado para la serie.

---

### 13.7. Comentarios al procedimiento de detección de atípicos y lecturas complementarias

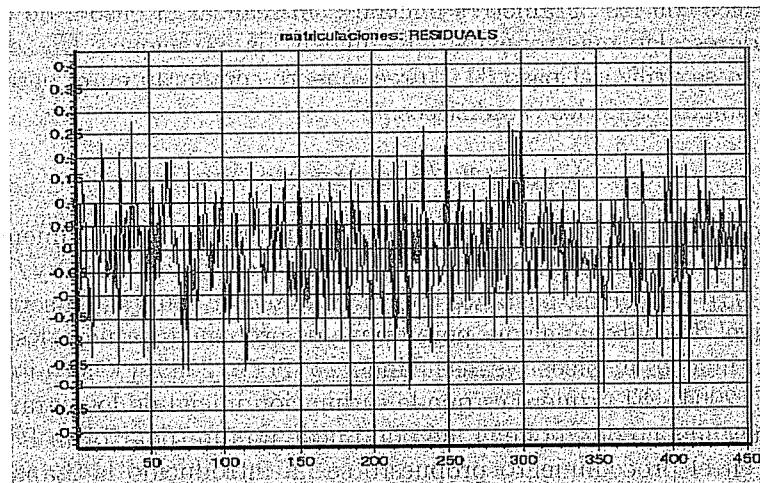
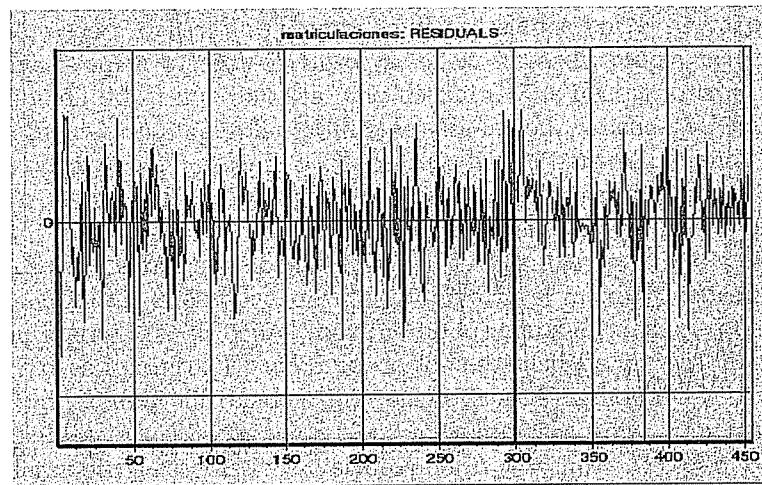
Al aplicar el procedimiento anterior es importante realizar las correcciones de los cambios de nivel de uno en uno y estimar conjuntamente sus efectos para corregir la serie. El problema que puede aparecer con varios cambios de nivel se ilustra en la figura 13.8 mediante una serie de media cero. La serie tiene dos cambios de nivel, el primero de magnitud 5 y el segundo de magnitud 10. El primer cambio de nivel que se detecta es el mayor, que ocurre en la posición 60, y su magnitud se estima por la media de los residuos después del cambio, que es alrededor de ocho. De esta manera se subestima su magnitud real debido a la presencia del primer cambio. Sin embargo, si se identifican los dos cambios de nivel y se estiman conjuntamente sus efectos, este problema no aparece.

Un problema similar de sesgos por la presencia de varios efectos puede aparecer con los restantes atípicos, aunque normalmente el sesgo en este caso no es tan grave. Para evitarlo, conviene realizar las correcciones de los atípicos de uno en uno y siempre incluir una estimación conjunta de sus efectos.

Un aspecto a tener en cuenta es que, como hemos comentado, si las series son no estacionarias los IO pueden confundirse fácilmente con otros

### 13. Valores atípicos

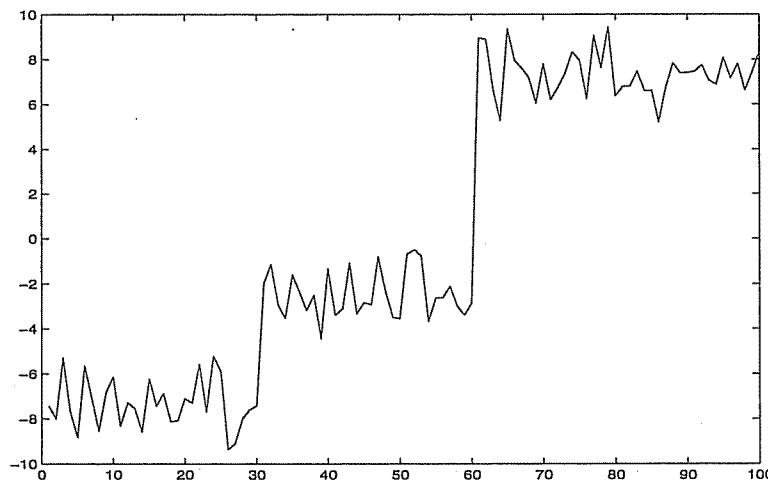
Figura 13.7 Residuos de la serie de logaritmos de las matriculaciones sin corregir y corregida de atípicos



efectos. Por ejemplo, con cambios de nivel en series I(1) y con efectos rampa, en series I(2). Conviene siempre, antes de aceptar un IO en una serie integrada, asegurarse de que no está encubriendo otro tipo de efecto de más fácil interpretación.

El procedimiento de detección de atípicos que está implantado en programas como SCA y TSW es básicamente el que hemos presentado pero

Figura 13.8 Una serie con dos cambios de nivel



con dos diferencias que pueden ser importantes en series con muchos atípicos. En primer lugar, en estos programas se comienza la búsqueda de atípicos con los parámetros iniciales, sin realizar el paso 2 de buscar estimadores algo más robustos para iniciar las iteraciones. En segundo lugar, no se utiliza el paso 3, y directamente comparan en el paso 4 los valores máximos de los estadísticos:

$$\lambda_{\max} = \max(\lambda_{AO}, \lambda_{LS}, \lambda_{IO}, \lambda_{TC})$$

para decidir el tipo de efecto. Este método fue propuesto inicialmente por Chang y Tiao (1983), Tsay (1986), Chang, Tiao y Chen (1988) y Chen y Liu (1993) y funciona bastante bien para detectar atípicos en series con bajo nivel de contaminación, pero tiene dos problemas principales en series con muchos atípicos (véanse Sánchez y Peña, 2003, y Galeano, Peña y Tsay, 2004). El primero es que la estimación inicial de los parámetros puede ser mala, con lo que el método puede no funcionar bien. Por esta razón, en el procedimiento propuesto se ha introducido la etapa 2. El segundo problema es que la comparación de los estadísticos de cambio de nivel con lo de los otros atípicos no es aconsejable, porque puede fallar en la detección de los cambios de nivel, especialmente cuando hay más de uno. La razón es que la distribución del máximo del estadístico para cambios de nivel,  $\lambda_{LS}$ , es distinto de los restantes estadísticos. Para comprender esta diferencia en la distribución de los estadísticos observemos que los estadísticos  $\lambda_t^{IO}$  o  $\lambda_t^{AO}$  están, aproximadamente, incorrelados para distintas observaciones. Por ejemplo,  $\lambda_t^{IO}$  es el residuo estandarizado, y

estará incorrelado de unos instantes a otros. El estadístico  $\lambda_t^{AO}$  depende de unos pocos residuos posteriores y, por lo tanto, existirá correlación entre valores próximos de  $\lambda_t^{AO}$ , pero este efecto es pequeño. Sin embargo, los valores de  $\lambda_t^{LS}$  están siempre muy correlados entre sí, ya que son diferencias entre medias de observaciones. Esta diferencia hace que la distribución de su máximo sea muy distinta, y es necesario tratarlos separadamente. En un amplio estudio de simulación, Sánchez y Peña (2003) encontraron que el percentil 95 % del estadístico para LS, cuando no hay atípicos, está entre 2.5 y 3 para procesos estacionarios, pero está entre 3.5 y 3.8 para procesos no estacionarios. Sin embargo, el percentil 95 % para los efectos transitorios AO, IO y TC es similar, próximo a 3.5. Este resultado indica que no es adecuado comparar el máximo del estadístico del contraste para LS con el máximo de los otros efectos, porque la distribución es distinta. El máximo de los estadísticos  $\lambda_t^{LS}$  es el máximo de variables muy correladas, que siempre será menor que el máximo de variables incorreladas. Por ejemplo, algunos autores han propuesto considerar otros estadísticos distintos del  $\lambda_{LS} = \max \lambda_t^{LS}$  para detectar cambios de nivel, como  $\sum (\lambda_t^{LS})^2$  u otros. Veáñese Andrews y Ploberger (1994), Bai y Perron (1998, 2003), Sánchez y Peña (2003), Carnero, Peña y Ruiz (2004) y Galeano, Peña y Tsay (2004).

### Ejercicios 13

- 13.1 Supongamos un proceso MA(1). Si ocurre un atípico aditivo (AO) de tamaño  $w$  para  $t = h$  y suponiendo que el parámetro  $\theta$  del MA es conocido, indicar cómo se modificarán los residuos después del cambio.
- 13.2 Supongamos un proceso AR(1). Si ocurre un atípico innovativo (IO) de tamaño  $w$  para  $t = h$  y suponiendo que el parámetro  $\phi$  es conocido, indicar cómo se modificarán los residuos después del cambio.
- 13.3 Obtener los efectos sobre los residuos de un cambio de nivel en un proceso ARMA(1,1).
- 13.4 Indicar en qué condiciones un cambio transitorio es equivalente a un IO.
- 13.5 Justificar que el modelo  $z_t = \alpha + \beta t + wR_t^{(h)} + n_t$  donde  $n_t$  es un proceso estacionario se identificará como el modelo  $\nabla z_t = \beta + wS_t^{(h)} + \eta_t$  donde  $\eta_t$  es un proceso no invertible.
- 13.6 Justificar que el modelo  $z_t = \alpha + \beta_t t + wR_t^{(h)} + n_t$ , donde además  $\beta_t = \beta_{t-1} + v_t$  y  $n_t$  y  $v_t$  son procesos estacionarios independientes, se

identificará como el modelo  $\nabla z_t = wS_t^{(h)} + \eta_t$  donde  $\eta_t$  es un proceso estacionario e invertible.

---

### Apéndice 13.1: Estimación de los cambios de nivel

El procedimiento de estimación de los cambios de nivel requiere siempre que se estimen conjuntamente los parámetros y los cambios de nivel para evitar sesgos. Para ilustrar el problema, supongamos, para simplificar, que  $z_t = \mu + a_t$ , de manera que  $\pi(B) = 1$ . Llamando  $e_t = z_t - \bar{z}$ , siendo  $\bar{z}$  la media de la serie, hemos visto que el estimador del cambio de nivel es:

$$\hat{\omega}_{LS} = \frac{(e_h + e_{h+1} + \dots + e_T)}{(T - h + 1)} = \frac{\sum_{t=h}^T (z_t - \bar{z})}{(T - h + 1)} = \bar{z}_2 - \bar{z},$$

donde hemos llamado  $\bar{z}_2$  a la media de la serie después del cambio. Sustituyendo en esta expresión la media  $\bar{z}$  como combinación lineal de las medias antes  $\bar{z}_1$ , y después,  $\bar{z}_2$ , del cambio:

$$\bar{z} = \frac{(h-1)\bar{z}_1 + (T-h+1)\bar{z}_2}{T}, \quad (13.20)$$

y la estimación de  $\hat{\omega}_{LS}$ , resulta:

$$\hat{\omega}_{LS} = \frac{h-1}{T}(\bar{z}_2 - \bar{z}_1).$$

Este estimador tiene sesgo, como consecuencia de no conocer la media verdadera del proceso, pero indicará correctamente el punto de cambio, ya que su valor máximo es esperable que ocurra en ese instante. Aunque  $(\bar{z}_2 - \bar{z}_1)$  estima en promedio la magnitud del cambio si, por ejemplo,  $h = T/2$  el valor estimado por  $\hat{\omega}_{LS}$  será la mitad del cambio real.

Sin embargo, cuando estimemos conjuntamente el tamaño del cambio y los parámetros del modelo, el problema desaparece. El modelo entonces es:

$$z_t = \mu + \omega_{LS} S_t^{(h)} + a_t, \quad (13.21)$$

y su estimación por mínimos cuadrados conduce a un valor de la constante de:

$$\hat{\mu} = \bar{z} - \hat{\omega}_{LS} \bar{S}^{(h)},$$

donde  $\bar{S}^{(h)} = \sum_{t=1}^T S_t^{(h)}/T = (T-h+1)/T$ . La magnitud del cambio de nivel se estima por:

$$\hat{\omega}_{LS} = \frac{\sum_{t=1}^T (z_t - \bar{z}) S_t^{(h)}}{\sum_{t=1}^T (S_t^{(h)} - (T-h+1)/T)^2} = \frac{\sum_{t=h}^T (z_t - \bar{z})}{(T-h+1)(h-1)/T},$$

que puede escribirse:

$$\hat{\omega}_{LS} = \frac{T}{(h-1)} (\bar{z}_2 - \bar{z})$$

y sustituyendo en esta ecuación la expresión de  $\bar{z}$  (13.20):

$$\hat{\omega}_{LS} = \bar{z}_2 - \bar{z}_1.$$

Esta ecuación estima bien el cambio de nivel, como diferencia entre las medias antes y después del cambio. La varianza de esta estimación es:

$$var(\hat{\omega}_{LS}) = \frac{\sigma_a^2}{\sum(S_t^{(h)} - \bar{S}^{(h)})^2} = \frac{\sigma_a^2}{(T-h+1)(h-1)/T},$$

que coincide con la expresión habitual de la varianza de la diferencia de medias en dos muestras independientes:

$$var(\hat{\omega}_L) = \sigma_a^2 \left( \frac{1}{(T-h+1)} + \frac{1}{h-1} \right).$$

En esta expresión  $\sigma_a^2$  se estima mediante la varianza residual de la regresión:

$$\hat{\sigma}_a^2 = \left( \sum_{t=1}^{h-1} (z_t - \bar{z}_1)^2 + \sum_{t=h}^T (z_t - \bar{z}_2)^2 \right) / T$$

que equivale a la de la regresión (13.21).

Estos resultados pueden generalizarse para otros modelos. Supongamos un modelo un poco más complejo, como el IMA(1,1), con un cambio de nivel:

$$z_t = c + \omega S_t^{(h)} + \frac{(1-\theta B)}{\nabla} a_t,$$

multiplicando por  $\nabla(1-\theta B)^{-1}$ , la ecuación a estimar es:

$$y_t = \omega x_t + a_t$$

donde:

$$y_t = (1-\theta B)^{-1} \nabla z_t, \quad (13.22)$$

y:

$$x_t = (1-\theta B)^{-1} I_t^{(h)} = (0, 0, \dots, 1, \theta, \theta^2, \dots).$$

## Análisis de series temporales

---

El estimador de la pendiente en la regresión es:

$$\hat{\omega} = \frac{\sum_{t=1}^T y_t x_t}{\sum_{t=1}^T x_t^2} = \frac{y_h + \theta y_{h+1} + \theta^2 y_{h+2} + \dots}{1 + \theta^2 + \theta^4 + \dots} = (1 - \theta^2)(1 - \theta F)^{-1} y_h$$

donde  $F = B^{-1}$  es el operador de adelanto y hemos supuesto que el tamaño muestral es grande y que  $h$  no corresponde ni al principio ni al final de la serie. Entonces, el numerador puede escribirse como  $(1 - \theta F)^{-1} y_h$ , y en el denominador tenemos una progresión geométrica indefinida con razón  $\theta^2$  y suma  $(1 - \theta^2)^{-1}$ . Sustituyendo en esta expresión (13.22):

$$\hat{\omega} = (1 - \theta^2)(1 - \theta F)^{-1}(1 - \theta B)^{-1} \nabla z_t$$

y expresando  $(1 - \theta F)^{-1} = (1 + \theta F + \theta^2 F^2 + \dots)$  y  $(1 - \theta B)^{-1} = (1 + \theta B + \theta^2 B^2 + \dots)$  y operando y agrupando términos, se obtiene que:

$$\hat{\omega} = Z(\text{antes}) - Z(\text{después})$$

donde  $Z(\text{antes})$  es una media ponderada de las observaciones antes del cambio de nivel:

$$Z(\text{antes}) = (1 - \theta)(z_{h-1} + \theta z_{h-2} + \dots)$$

y  $Z(\text{después})$  es una media ponderada de las observaciones a partir del cambio de nivel:

$$Z(\text{después}) = (1 - \theta)(z_h + \theta z_{h+1} + \dots).$$

## Apéndice 13.2: Programas para análisis de atípicos

### TSW

Veamos cómo estimar una serie con análisis de intervención con el programa TSW. Utilizaremos como ejemplo la serie crest (fichero crestcolgate.dat), que contiene 275 datos semanales que empiezan en la primera semana de enero de 1958 y acaban en la última semana de marzo de 1963. Como hemos visto en el ejemplo 13.4, esta serie se puede modelizar mediante un ARIMA(0,1,1), y en la última semana del mes de julio de 1960 se produjo un aumento de ventas debido al anuncio de que la inclusión de flúor en el dentífrico era beneficiosa para los dientes.

Vemos cómo el programa TSW busca este tipo de efectos atípicos automáticamente. Marcamos Series y a continuación seleccionamos la serie. En Series list, vemos que tenemos la serie incluida en el programa. Si marcamos la serie aparece el nombre de la serie y ciertos atributos que podemos cambiar. Por ejemplo, el parámetro Iter si lo cambiamos a Iter=1, a la misma serie le podemos ajustar diferentes modelos. Lo mismo con tramo para utilizar sólo el programa tramo. También aparece el gráfico de la serie, lo que nos per-

### 13. Valores atípicos

mite llevar a cabo una rápida inspección. Puesto que ya hemos determinado el modelo ARIMA(0,1,1), marcamos +Model. El parámetro RSA lo dejamos en 0. Pasamos a ARIMA model y marcamos, P=0; Q=1; D=1; BP=0; BQ=0; BD=0; INIT=0; IMEAN=0, este parámetro sirve para corregir la media o no. Esto significa que no corregimos la media; LAM=1, este parámetro dice si transformar logarítmicamente los datos. Vale 0 si se quiere transformar, 1 si no, y -1 si el programa realiza un contraste sobre la necesidad de transformar; FCT=1, relacionado con el contraste, se deja así; TYPE=0, para estimar MLE exacta; UNITS=0, se deja así, es para cambiar la escala si los datos son muy pequeños o grandes.

Marcamos Others para los parámetros correspondientes a Outliers. Por ahora marcamos IATIP=1, lo que significa que además de estimar los parámetros, busque atípicos y estime sus efectos. Marcamos OK. Marcamos el modelo en la tabla y marcamos Run. En output, vemos la salida. Es importante recordar que este programa supone que todos los coeficientes de los operadores AR y MA van con signo más, por lo que un coeficiente MA(1)<sub>12</sub> estimado como BMA(1)=-0.87 (en este programa se utiliza la letra B para los estacionales) debe escribirse como el operador  $(1 - .87B^{12})$ . En la parte de atípicos aparecen los siguientes valores:

135LS 166AO 195TC

El dato en 135 corresponde a un escalón (cambio de nivel), el 166 a un impulso y el dato en 195 corresponde a un cambio transitorio, o impulsos sucesivos que van decreciendo con el tiempo.

### EViews

Veamos cómo estimar una serie con análisis de intervención con el programa EViews. Abrimos el programa y creamos un workfile. La serie crest son datos semanales que empiezan en 1/1/1958 y acaban en 27/3/1963 por lo que son 275 datos. Generamos una serie tal que:

```
Genr → s1 = 0  
Genr → s1 = 1, sample: 7 : 27 : 1960, 3 : 27 : 1963
```

y pasamos a estimar el modelo ARIMA(0,1,1) con un cambio de nivel. Estimamos el siguiente modelo ARIMA(0,1,1) con un cambio de nivel en el dato 7:27:60. Para ello y teniendo en cuenta que tenemos que tomar una raíz unitaria, estimamos el modelo como sigue:

*Quick → Estimate Equation → d(x1,1) d(s1,1) ma(1)*

Obtenemos el siguiente modelo:

Análisis de series temporales

---

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(S1,1)	0.13	0.030	4.31	0
MA(1)	-0.76	0.04	-18.66	0

Además podemos ver un efecto que puede ser un impulso en el dato 1/10/1962, es decir, en la segunda semana del mes de enero. Para ello, definimos una serie como sigue:

*Genr → s2 = 0*

*Genr → s2 = 1, sample: 1 : 10 : 1962, 1 : 10 : 1962*

y podemos estimar el modelo como sigue:

*Procs → Specify/Estimate → d(x1,1) d(s1,1) d(s2,1) ma(1)*

y obtenemos el modelo siguiente:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(S1,1)	0.13	0.029	4.55	0
D(S2,1)	0.11	0.04	2.87	0.0043
MA(1)	-0.76	0.04	-18.66	0

Analizamos ahora el correlograma y el histograma de los residuos:

*Residuals test → Correlogram Statistics*

*Residuals test → Histogram*

que muestran que no existen problemas con los residuos. Observemos que un efecto escalón detectado aquí no lo ha sido con TSW.

---

## 14. Modelos no lineales



Howell Tong (1944)

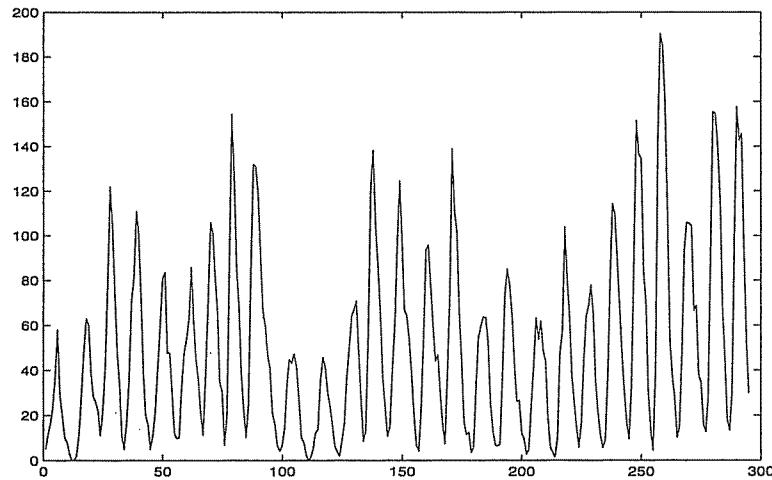
Estadístico de origen chino. Estudió en Inglaterra y es actualmente catedrático de Estadística en la London School of Economics. Inventor de los procesos autorregresivos por umbrales, es uno de los artífices principales de la teoría moderna de las series temporales no lineales.

### 14.1. Introducción

En los capítulos anteriores hemos estudiado modelos lineales. Estos modelos tienen dos propiedades importantes que los hacen muy cómodos para la predicción: (1) las predicciones, que son las esperanzas condicionadas de los valores futuros dada la historia del proceso, se construyen mediante combinaciones lineales de los valores presentes y pasados; (2) las varianzas de las predicciones dependen de los parámetros del modelo, pero no de los valores previos del proceso.

En la práctica encontramos series temporales que no cumplen estas hipótesis. Si la relación entre el valor presente de la serie y su pasado es una función no lineal, las predicciones que minimizan el error cuadrático de predicción serán funciones no lineales del pasado, y diremos que tenemos un modelo que es no lineal en la media. Aunque los modelos lineales estudiados proporcionan aproximaciones útiles a la verdadera relación, podremos mejorar la predicción si utilizamos un modelo más preciso. Por ejemplo, si la relación entre las ventas de hoy y las de ayer

Figura 14.1 Serie de manchas solares. Datos anuales de 1700 a 1994



es distinta cuando las ventas de ayer fueron altas que cuando las ventas fueron bajas, podremos mejorar las predicciones incorporando esta propiedad en el modelo. Otro ejemplo de una serie no lineal en la media es una serie de consumo de energía donde el valor presente depende linealmente del pasado, pero con parámetros que van cambiando en el tiempo.

Consideremos la serie de manchas observadas en la corteza solar, que se presenta en la figura 14.1. La serie parece estacionaria pero tiene un comportamiento pseudo cíclico con muchos valores extremos. La figura 14.2 presenta el histograma de la serie y la relación entre la serie y sus retardos  $(z_t, z_{t-1}), \dots, (z_t, z_{t-k})$ , donde hemos tomado  $k = 4$ . Se observa en este gráfico que la distribución marginal de la variable es marcadamente asimétrica, y que la relación entre la variable y sus retardos no es la esperable en un modelo lineal: hay asimetrías y aunque en el primer diagrama de dispersión entre la variable y su retardo la relación podría ser lineal, no lo es en el diagrama entre  $z_t$  y  $z_{t-4}$ .

Podría pensarse en transformar la variable para intentar obtener distribuciones marginales más simétricas y relaciones lineales entre la variable y sus retardos. La figura 14.3 presenta el histograma y los gráficos de dispersión después de transformar la serie con  $y_t = \log(z_t + 1)$ . Aunque la distribución marginal de los datos es más simétrica, las relaciones no lineales entre la variable y sus retardos se mantienen y son visibles para retardos altos.

#### 14. Modelos no lineales

Figura 14.2 Histograma y relación entre la variable y sus retardos para los datos de las manchas solares

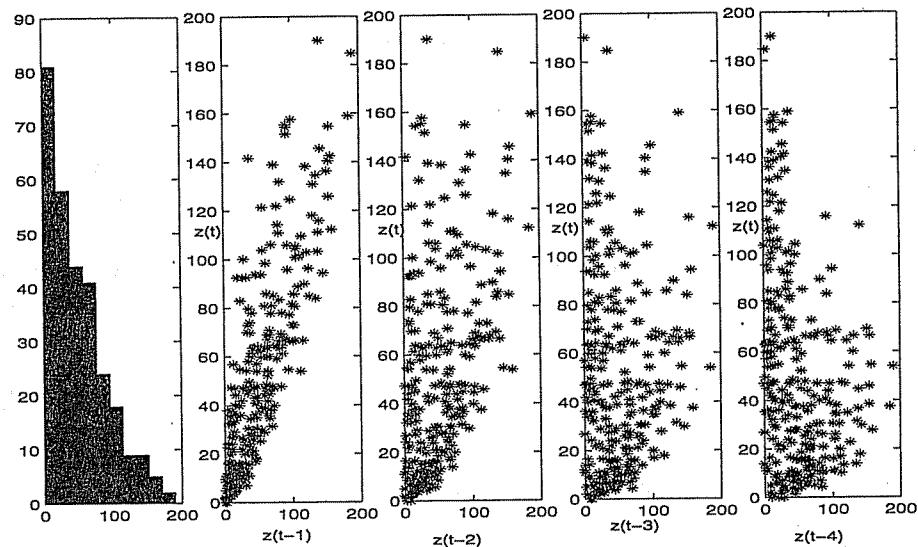
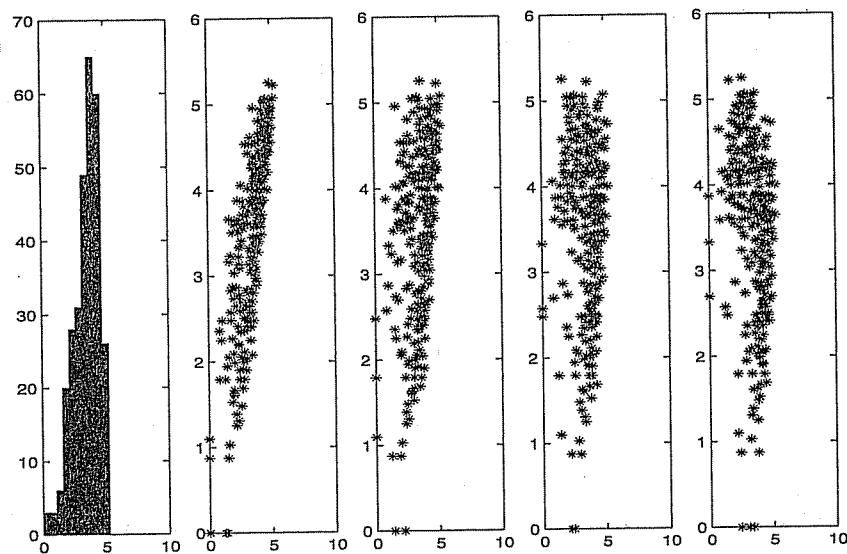


Figura 14.3 Histograma y gráficos de dispersión entre la serie y sus retardos para el logaritmo de la serie de manchas solares



## 14.2. Modelos no lineales en la media y en la varianza

Llamando  $\mathbf{z}_{t-1} = (z_{t-1}, \dots, z_1)$  al vector de valores pasados de la serie, una formulación de un modelo no lineal en la media es:

$$z_t = m(\mathbf{z}_{t-1}) + \varepsilon_t, \quad (14.1)$$

donde  $\varepsilon_t$  es un proceso de ruido blanco estricto de media cero y varianza  $\sigma^2$ . Esta condición es importante, porque según el teorema de Wold, si el proceso es estacionario podremos aproximar (14.1) mediante un modelo lineal con residuos sin autocorrelación serial, pero si el modelo generador es (14.1) los residuos no serían independientes. Para que el proceso sea estacionario en la media debe verificarse que

$$E(z_t) = E(m(\mathbf{z}_{t-1})) = \mu, \quad (14.2)$$

ya que la innovación tiene media cero. Esta condición establece fuertes restricciones en la clase de funciones  $m(\mathbf{z}_{t-1})$  que podemos considerar. Por ejemplo, la generalización más simple de un AR(1) para hacerlo no lineal es:

$$z_t = \alpha z_{t-1}^2 + \varepsilon_t,$$

pero para que este proceso sea estacionario en la media es necesario que  $E(z_t) = E(z_{t-1}^2) = E(z_t^2)$ . Esta igualdad no es en general cierta, ya que entonces  $\text{var}(z_t) = E(z_{t-1}^2)(1 - E(z_{t-1}^2))$  y como la varianza debe ser positiva sólo es posible si  $E(z_t^2) = E(z_t) < 1$ . Por tanto, para una variable general  $z_t$  este proceso no será estacionario.

Suponiendo que el proceso cumple la condición (14.2), y tiene, por tanto, media constante, la esperanza del proceso condicionada a su pasado es una función no lineal:

$$E(z_t | \mathbf{z}_{t-1}) = m(\mathbf{z}_{t-1}).$$

Sin embargo, la varianza condicionada a su pasado es constante, ya que

$$\text{Var}(z_t | \mathbf{z}_{t-1}) = \sigma^2$$

y  $\varepsilon_t$  se interpreta, igual que en un modelo lineal, como el error de predicción a un paso.

Una serie puede tener estructura no lineal no sólo en la media, sino también en la varianza. Por ejemplo, con datos financieros y climatológicos se observa con frecuencia en series estacionarias que la varianza de la serie en el instante  $t$  depende de la variabilidad de los datos anteriores, de manera que hay rachas de alta varianza, seguidas de datos con poca

varianza. Podemos generalizar (14.1) añadiendo este tipo de dependencia y obtener una formulación general para procesos no lineales, del tipo:

$$z_t = m(z_{t-1}) + u_t v(z_{t-1}), \quad (14.3)$$

donde  $u_t$  es un proceso de ruido blanco normal, formado por variables  $N(0, 1)$  independientes entre sí y de los valores  $z_{t-1}$ , y las funciones  $m$  y  $v$  son no lineales en los valores pasados de la serie. Si tomamos esperanzas condicionadas a  $z_{t-1}$  en esta expresión, obtenemos:

$$E(z_t | z_{t-1}) = m(z_{t-1}).$$

En efecto, al ser  $E(u_t | z_{t-1}) = E(u_t) = 0$ , tenemos que:

$$E(u_t v(z_{t-1}) | z_{t-1}) = E(u_t | z_{t-1}) v(z_{t-1}) = 0,$$

y si la función  $m(z_{t-1})$  es no lineal, la función de medias condicionadas del proceso,  $E(z_t | z_{t-1})$ , también lo será. Sin embargo, la varianza condicionada de este proceso es:

$$\begin{aligned} Var(z_t | z_{t-1}) &= E[(z_t - m(z_{t-1}))^2 | z_{t-1}] = \\ &= v^2(z_{t-1}) E(u_t^2 | z_{t-1}) = v^2(z_{t-1}), \end{aligned}$$

y si la función  $v^2(z_{t-1})$  es no lineal, la varianza condicional de este proceso será, en general, una función no lineal de los valores pasados. El modelo con varianza constante se obtiene suponiendo  $v^2(z_{t-1}) = \sigma^2$ .

En este capítulo nos centraremos en los procesos no lineales en la media y dejaremos para el siguiente los procesos no lineales en la varianza. Estos dos tipos de efectos no lineales pueden o no darse simultáneamente en una serie temporal.

La forma general de un proceso estacionario no lineal en la media se obtiene mediante la llamada representación de Volterra. Ésta es una generalización de la representación MA( $\infty$ ) de Wold para procesos lineales y se escribe:

$$\begin{aligned} z_t &= \mu + \sum_{i=-\infty}^{\infty} \eta_i \varepsilon_{t-i} + \sum_{i,j=-\infty}^{\infty} \eta_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \quad (14.4) \\ &\quad + \sum_{i,j,k=-\infty}^{\infty} \eta_{ijk} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_{t-k} + \dots \end{aligned}$$

donde el proceso  $\varepsilon_t$  es de ruido blanco *estricto*, es decir formado por variables no sólo incorreladas sino además independientes. Esta expresión tiene un número infinito de coeficientes y, como la representación de

Wold, no es operativa. De la misma forma que los procesos ARMA tratan de aproximar la representación de Wold, se han desarrollado muchos modelos para aproximar la representación general anterior. En este capítulo veremos algunas aproximaciones simples a esta formulación.

### 14.3. Los procesos bilineales

Los modelos habituales de procesos no lineales pueden considerarse como aproximaciones a la expansión general (14.4). Un primer enfoque es suponer que podemos expresar la dependencia en función de un número finito de retardos y escribir el modelo ARMA no lineal:

$$z_t = f(z_{t-1}, \dots, z_{t-p}; \varepsilon_{t-1}, \dots, \varepsilon_{t-q}) + \varepsilon_t$$

donde  $f$  es arbitraria y la innovación  $\varepsilon_t$  es un proceso de ruido blanco estricto, independiente de  $z_{t-k}$  para cualquier  $k > 0$ . Si  $q = 0$  tenemos un AR no lineal y si  $p = 0$  un MA no lineal. Suponiendo que  $f$  es una función suave con derivadas continuas, podemos desarrollarla en serie y, despreciando términos de orden superior, aproximarla por:

$$z_t = c + \sum_{i=1}^p \phi_i z_{t-i} + \sum_{i=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^m \sum_{j=1}^h \gamma_{ij} z_{t-i} \varepsilon_{t-j} + \varepsilon_t$$

que conduce a la clase de procesos *bilineales*. Estos modelos han sido estudiados por Granger y Anderson (1978). Puede demostrarse que los procesos bilineales aproximan la expansión de Volterra en un intervalo finito con cualquier precisión determinada. Se representan como  $BL(p, q, m, h)$ , donde  $p, q$  se refieren a la parte lineal y  $m, h$  a los órdenes de los productos bilineales.

El proceso bilineal más simple es el  $BL(1,0,1,1)$ , dado por:

$$z_t = \phi z_{t-1} + \gamma z_{t-1} \varepsilon_{t-1} + \varepsilon_t. \quad (14.5)$$

En esta ecuación a una estructura AR(1) de dependencia estacionaria,  $\phi z_{t-1}$ , se añade un término bilineal,  $\gamma z_{t-1} \varepsilon_{t-1}$ , que depende del último valor observado y del último error de predicción. Supongamos que  $\gamma \geq 0$ . Entonces, cuando  $\varepsilon_{t-1}$  es positivo, podemos escribir

$$z_t = (\phi + \gamma \varepsilon_{t-1}) z_{t-1} + \varepsilon_t,$$

y  $z_t$  se genera con un coeficiente AR igual a  $\phi + \gamma \varepsilon_{t-1}$ , que puede ser mayor que uno si  $\varepsilon_{t-1} \geq (1 - \phi)/\gamma$ . Entonces, y durante un breve periodo de tiempo, el comportamiento de la serie será como en un proceso

explosivo. Sin embargo, cuando  $\varepsilon_{t-1}$  sea negativo, el siguiente valor se genera con un coeficiente AR  $\phi - \gamma |\varepsilon_{t-1}|$ , que puede ser cero o negativo. Por tanto, este proceso genera una estructura de dependencia localmente lineal, entendiendo por ello que en un intervalo breve de tiempo la dependencia es lineal, pero que cambia continuamente. Si el valor de  $\gamma$  no es pequeño, la serie puede reflejar un comportamiento no estacionario.

Si este proceso es estacionario, su media será, tomando esperanzas en (14.5) e imponiendo la condición de media constante:

$$\mu = \phi\mu + \gamma E(z_{t-1}\varepsilon_{t-1}) = \phi\mu + \gamma\sigma_\varepsilon^2,$$

que resulta en

$$\mu = \frac{\gamma\sigma_\varepsilon^2}{1 - \phi}.$$

Hemos utilizado que  $E(z_t\varepsilon_t) = \sigma_\varepsilon^2$ , como se comprueba multiplicando (14.5) por  $\varepsilon_t$  y tomando esperanzas. Se demuestra (véase por ejemplo Granger y Andersen, 1978) que la condición para que el proceso tenga varianza positiva y constante es  $\phi^2 + \gamma^2\sigma_\varepsilon^2 < 1$ , y que la estructura de autocorrelación es similar a la de un proceso ARMA(1,1).

Este proceso puede también escribirse como:

$$\begin{aligned} z_t &= \phi_t z_{t-1} + \varepsilon_t \\ \phi_t &= \phi + \gamma\varepsilon_{t-1} \end{aligned}$$

que indica que se trata de un proceso AR(1) con parámetros cambiantes en cada instante. El proceso que gobierna los valores del parámetro es un proceso estacionario con media  $\phi$  y varianza  $\gamma^2\sigma_\varepsilon^2$ . En cada momento temporal el cambio en el parámetro es proporcional al último error de predicción cometido.

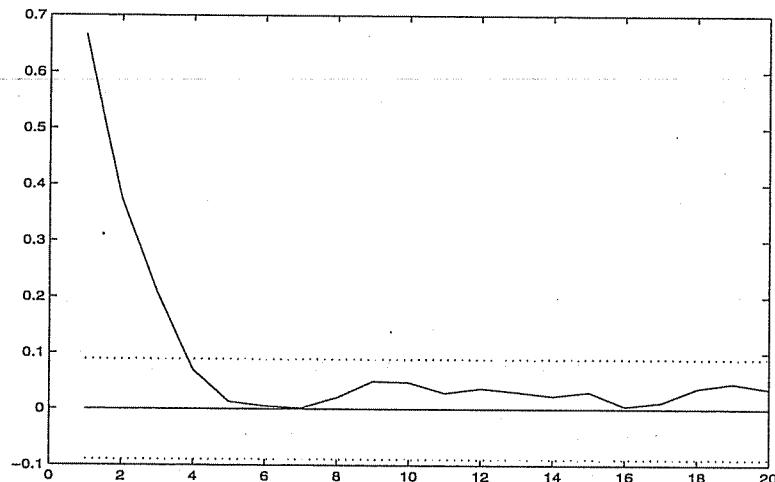
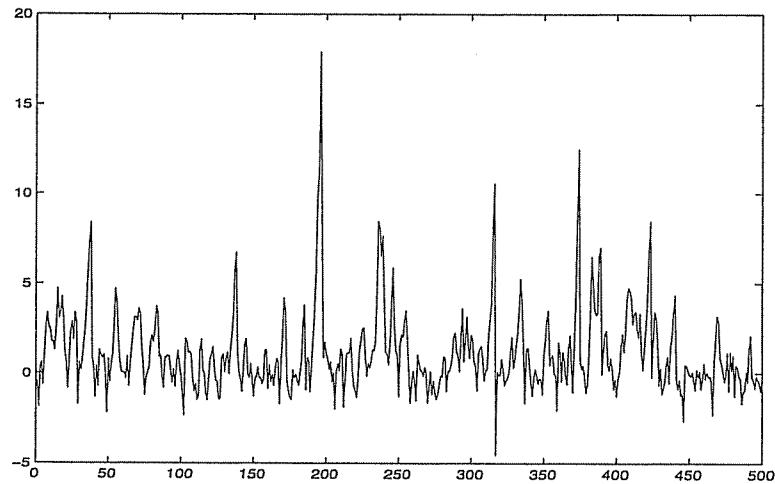
La figura 14.4 presenta una realización de este proceso y su función de autocorrelación simple. Se observa que la serie presenta muchos valores extremos, que se identificarían como atípicos en un ajuste lineal. La estructura de autocorrelación es similar a la de un ARMA(1,1), con el que podría confundirse.

La figura 14.5 muestra primero la distribución marginal del proceso, que es muy asimétrica y a continuación la relación entre la variable y sus cuatro primeros retardos. Puede comprobarse que la asimetría de la distribución marginal no desaparece tomando logaritmos. La propiedad de asimetría en la distribución marginal suele ir unida a la no linealidad: entre variables conjuntamente normales sólo pueden darse relaciones lineales y entre variables que se alejan de la normalidad es frecuente que las relaciones sean no lineales. Con relación a la dependencia entre la variable y sus retardos se observa una relación aproximadamente lineal para

## Análisis de series temporales

---

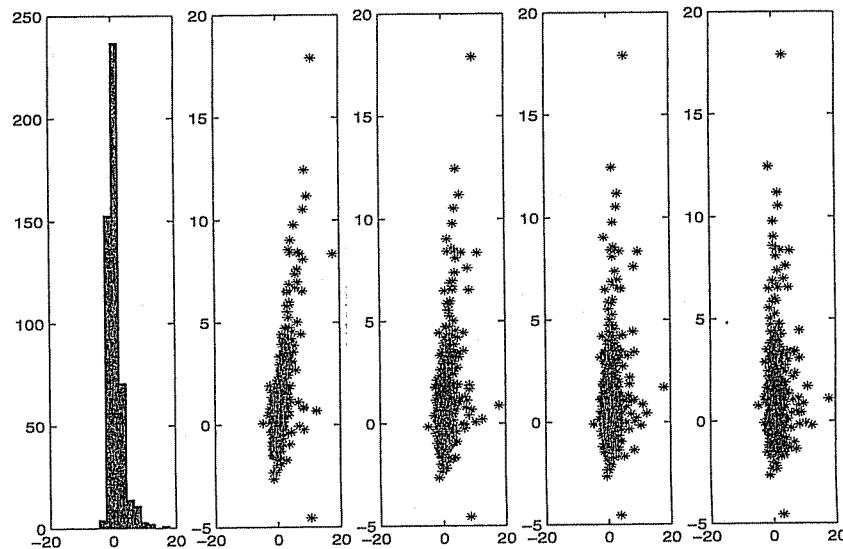
Figura 14.4 Simulación del proceso bilineal  $z(t) = .7z(t-1) + .5z(t-1)a(t-1) + a(t)$ . Gráfico del proceso y función de autocorrelación simple



valores pequeños de las variables, pero que se hace mucho más errática para valores altos de las variables, indicando un comportamiento no lineal. Esta situación es similar a la de algunas series lineales contaminadas con atípicos.

Si en lugar de modificar la estructura AR de primer orden mediante

Figura 14.5 Histograma y relación entre la variable y sus cuatro retardos en un proceso bilineal



la innovación pasada introducimos estructura de tipo AR(2) sobre una dependencia lineal de primer orden, tendremos el proceso BL(1,0,2,1):

$$z_t = \phi z_{t-1} + \beta z_{t-2} \varepsilon_{t-1} + \varepsilon_t,$$

que puede interpretarse como un proceso AR(2) donde el parámetro de primer orden es constante, pero el segundo es variable en el tiempo,  $\phi_{2t} = \beta \varepsilon_{t-1}$ . Como  $z_{t-2}$  y  $\varepsilon_{t-1}$  están incorrelados y  $E(\varepsilon_t) = 0$ , tomando esperanzas tenemos que  $\mu = \phi\mu$ , de donde concluimos que  $E(z_t) = 0$ . Las covarianzas serán, para  $k \geq 2$ ,

$$E[z_t z_{t-k}] = \phi E[z_{t-1} z_{t-k}], \quad (14.6)$$

y la varianza:

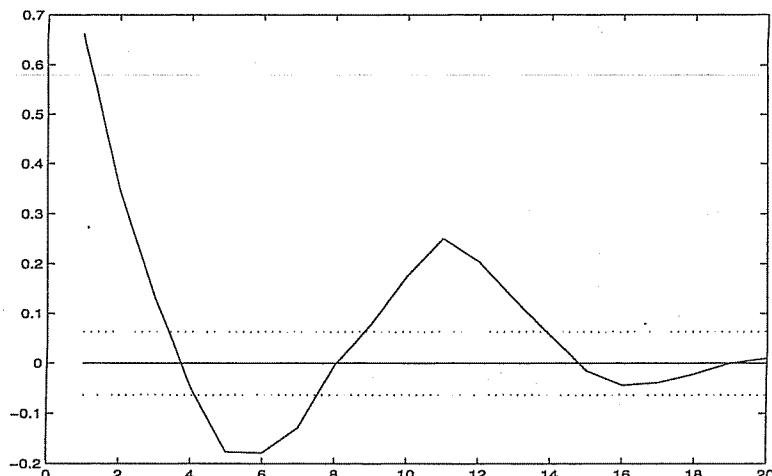
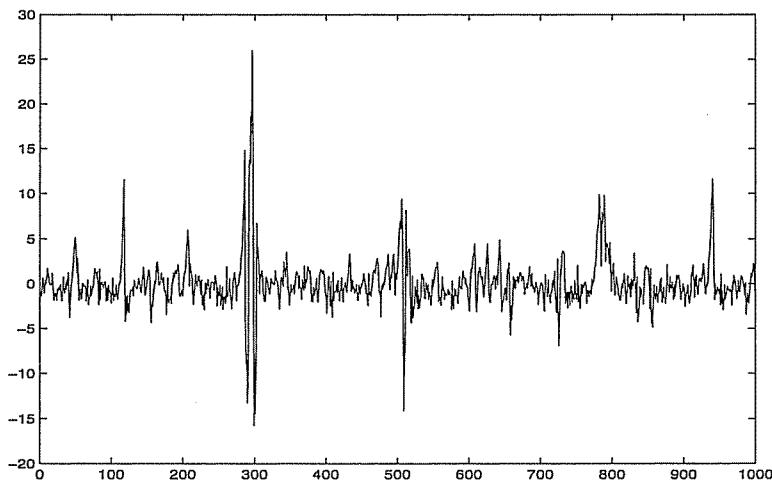
$$\sigma_z^2 = \phi^2 \sigma_z^2 + \beta^2 \sigma_z^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2,$$

de donde deducimos:

$$\sigma_z^2 = \frac{\sigma_\varepsilon^2}{(1 - \phi^2 - \beta^2 \sigma_\varepsilon^2)},$$

que será positiva si  $\phi^2 + \beta^2 \sigma_\varepsilon^2 < 1$ , que es la misma condición anterior para que el proceso sea estacionario.

**Figura 14.6 Simulación del proceso bilineal**  
 $z_t = .7z_{t-1} + .5z_{t-1}a_{t-1} + a_t$ . Gráfico del proceso  
 y función de autocorrelación simple



La figura 14.6 presenta un gráfico de este proceso y de su función de autocorrelación. Se observa que puede confundirse con un AR(2) lineal, pero con muchos atípicos. Sin embargo, si ajustamos un AR(2) lineal y calculamos sus residuos, dados por:

$$e_t = (\phi - \hat{\phi}_1)z_{t-1} + (\beta e_{t-1} - \hat{\phi}_2)z_{t-2} + \varepsilon_t,$$

es posible que estos residuos estén aproximadamente incorrelados, pero es fácil comprobar que sus cuadrados sí estarán correlados, lo que no ocurrirá si el proceso fuese lineal.

Aunque los modelos bilineales son una generalización muy natural de los lineales, se han aplicado poco en la práctica, ya que su estimación es complicada y las propiedades de los estimadores obtenidos se conocen sólo parcialmente.

#### 14.4. Los procesos autorregresivos por umbrales

Otra forma de introducir no linealidad es suponer que podemos aproximar la función que relaciona los valores presentes con los pasados por funciones lineales por tramos. Para introducir esta idea, supongamos un modelo AR(1) estacionario no lineal:

$$z_t = f(z_{t-1}) + \varepsilon_t.$$

En principio podemos aproximar la función  $f(z_{t-1})$  por tramos lineales. Por ejemplo, supongamos que la función  $f(z_{t-1}) = z_{t-1}^3$  es la representada en la figura 14.7. Esta función puede aproximarse por los tres tramos lineales que se indican en la figura y que quedan definidos por los umbrales  $d_1$  y  $d_2$ .

Utilizando esta aproximación, el modelo anterior puede escribirse como:

$$z_t = c^{(1)} + \phi^{(1)}z_{t-1} + \varepsilon_t^{(1)}, \quad z_{t-1} \leq d_1 \quad (14.7)$$

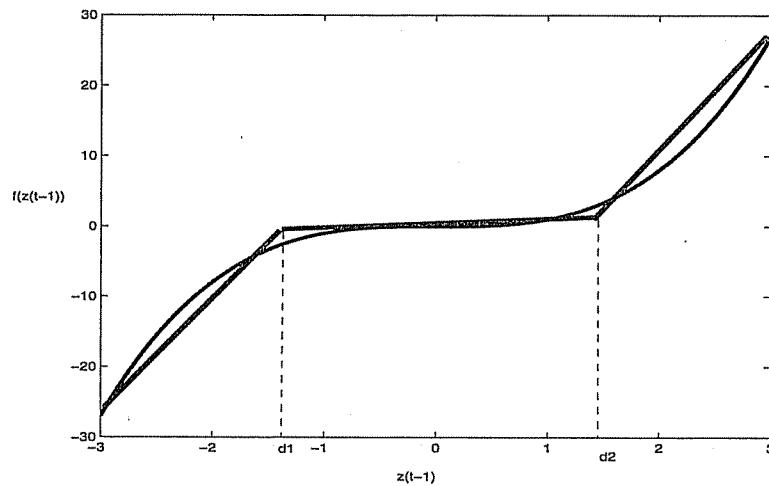
$$z_t = c^{(2)} + \phi^{(2)}z_{t-1} + \varepsilon_t^{(2)}, \quad d_1 < z_{t-1} \leq d_2$$

$$z_t = c^{(3)} + \phi^{(3)}z_{t-1} + \varepsilon_t^{(3)} \quad z_{t-1} > d_2 \quad (14.8)$$

donde  $\varepsilon_t^{(i)}$  para  $i = 1, 2, 3$  es ruido blanco con distribución  $N(0, \sigma_i^2)$ . En la figura (14.7) los parámetros  $\phi^{(1)}$  y  $\phi^{(3)}$  son iguales y el parámetro  $\phi^{(2)}$  es cero, ya que la función se approxima por una recta horizontal en ese intervalo. Por otro lado, las varianzas de las innovaciones,  $\sigma_i^2$ , pueden ser iguales en los tres intervalos, y entonces el proceso es no lineal en la media pero homocedástico, o distintas, con lo que el proceso es no lineal en la media y heterocedástico.

Esta formulación equivale a suponer que opera un modelo lineal distinto para distintos valores de la variable  $z_{t-1}$ . Estos modelos son debidos a Tong (1983) y se denominan *modelos autorregresivos por umbrales*, Threshold AutoRegressive (TAR) models, y también modelos autorre-

**Figura 14.7 Aproximación de una función no lineal por tres tramos lineales**



gresivos autoexcitados, Self-Exciting Threshold AutoRegressive (SETAR) models. La variable  $z_{t-1}$  se denomina variable de umbral y las constantes  $d_1$  y  $d_2$ , que determinan los puntos de corte, se denominan umbrales.

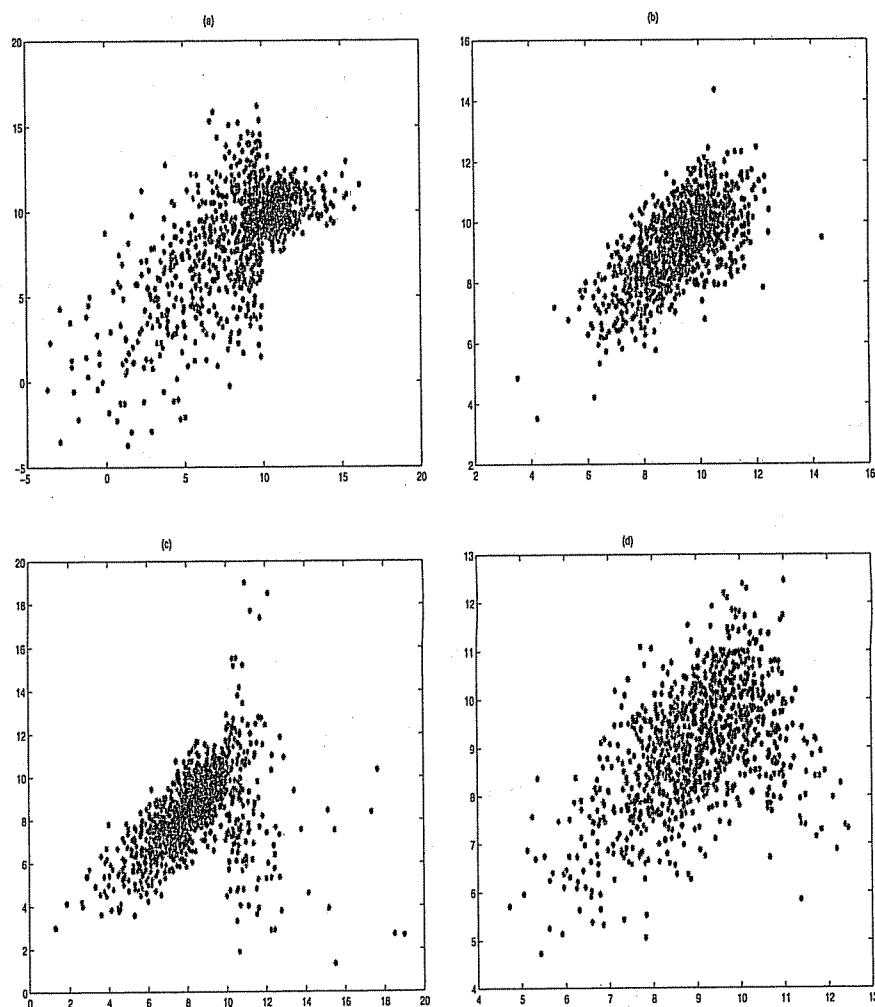
Las tres ecuaciones del modelo anterior pueden escribirse conjuntamente introduciendo tres variables auxiliares, una por cada posible régimen. Las dos primeras,  $\alpha_{1t} = P(z_{t-1} - d_1)$ ,  $\alpha_{3t} = P(z_{t-1} - d_2)$ , toman el valor 0 cuando el argumento de la función  $P(\cdot)$  es positivo y uno cuando es negativo, y la tercera  $\alpha_{2t} = 1 - \alpha_{1t} - \alpha_{3t}$ , toma el valor uno cuando las otras dos son cero. Entonces, el modelo (14.7) puede escribirse:

$$z_t = \left[ \sum_{i=1}^3 \alpha_{it} c^{(i)} \right] + \left[ \sum_{i=1}^3 \alpha_{it} \phi^{(i)} \right] z_{t-1} + \left[ \sum_{i=1}^3 \alpha_{it} \sigma_i \right] \epsilon_t$$

donde  $\epsilon_t$  es  $N(0, 1)$ .

La figura 14.8 presenta la relación entre  $z_t$  y  $z_{t-1}$  en un modelo con dos tramos con  $c^{(1)} = c^{(2)} = d = 10$ ,  $\phi^{(2)} = .8$ . Los valores de  $\phi^{(1)}$ ,  $\sigma_1$ ,  $\sigma_2$  en los cuatro casos son  $(.2, 1, 3)$ ,  $(.2, 1, 1)$ ,  $(-.8, 1, 3)$  y  $(-.8, 1, 1)$ . En el caso (a) la relación es siempre positiva pero presenta menos varianza para valores altos de  $z_{t-1}$  (mayores de 10) que para valores bajos. La pendiente también es menor para valores altos. Estas propiedades son claramente visibles en la figura (a). En el caso (b), la varianza es la misma en ambos regímenes y lo que cambia es la pendiente de la relación. En los casos (c) y (d) la relación es positiva para valores bajos, pero negativa para

**Figura 14.8** Relación entre la variable y su primer retardo para cuatro modelos TAR. En los cuatro casos  $d=10$  y el primer coeficiente AR es .8. Los modelos difieren en las varianzas y en el segundo coeficiente AR que es .2 arriba y -.8 en las de abajo



valores más altos. En el (c) la varianza es distinta en ambos regímenes, mientras que en el (d) es la misma.

Los gráficos de dispersión entre  $z_t$  y sus retardos son muy útiles para identificar la posible presencia de modelo TAR, ya que la relación entre  $z_t$  y la variable de umbral,  $z_{t-d}$ , debe ser cambiar para distintos valores

del umbral.

Los modelos TAR en el caso general se escriben como:

$$z_t = c^{(r)} + \sum_{i=1}^p \phi_i^{(r)} z_{t-i} + \varepsilon_t^{(r)}, \quad \mathbf{x}_{t-1} \in A_r, \quad r = 1, \dots, R$$

donde  $\mathbf{x}_{t-1} = (z_{t-1}, \dots, z_{t-p})'$ , es el vector de estado de la variable de umbral, y  $A_r$  la partición de los valores del vector de estado correspondiente a los  $R$  regímenes. Para cada posible estado  $r$ , tenemos distintos parámetros, tanto para los coeficientes del AR como para la varianza de las innovaciones,  $\sigma_r^2$ . La identificación de estos modelos cuando la variable que determina los umbrales es vectorial es difícil y en la práctica se supone que el vector de estado es escalar. Entonces, hay un único retardo que determina los umbrales, y los conjuntos  $A_r$  son una partición de los valores de la variable de umbral. Tong (1983) ha propuesto un procedimiento iterativo para ajustar estos modelos basado en el criterio AIC como sigue:

1. A partir de los gráficos de dispersión bivariantes entre la variable y sus retardos, se selecciona la variable  $z_{t-j}$  más prometedora como variable de umbral. Esta variable será aquella donde la relación entre  $z_t$  y  $z_{t-j}$  sea más no lineal. Una vez escogida la variable, se selecciona su valor de umbral o punto de corte,  $d$ , que es el valor que separa el gráfico en dos ecuaciones lineales. Con esta información se divide la muestra en dos submuestras, según el valor de la variable umbral, y se estiman modelos AR en cada tramo para distintos valores de  $p$ . Se selecciona el mejor modelo en cada uno de los regímenes mediante el AIC. Sea  $AIC(\hat{p}_1)$  el valor de este criterio para un modelo AR( $\hat{p}_1$ ). Se define el AIC del modelo total como la suma de los AIC correspondientes a cada régimen:

$$AIC(j, d) = AIC(\hat{p}_1) + AIC(\hat{p}_2),$$

donde  $\hat{p}_1$  y  $\hat{p}_2$  son los órdenes de los procesos AR ajustados a cada uno de los regímenes.

2. Se considera ahora un conjunto posible de valores de  $d$ ,  $-\infty \leq d_1 \leq \dots \leq d_k \leq \infty$ , dividiendo el rango de variación de  $z_{t-j}$ , la variable umbral, en tramos. Se repite el paso 1 para cada valor de  $d_j$ ,  $j = 1, \dots, k$ , y se selecciona el mejor modelo con el criterio AIC. Sea  $AIC(\hat{d})$  el valor del criterio correspondiente al mínimo de los modelos estimados.
3. Se repiten los dos pasos anteriores probando con otra posible varia-

ble de umbral,  $z_{t-h}$ , para  $h = 1, \dots, p$ . Se comparan los AIC obtenidos para distintos valores de la variable umbral y se elige como modelo final el que tenga mínimo AIC.

### Ejemplo 14.1

El modelo lineal ajustado a los datos del PIB de Estados Unidos, series  $z_t$ , que se encuentran en el fichero pibuso.dat, al periodo 4/1983-2/2003 ha sido:

$$y_t = (1 + 0.28B + .32B^2)a_t,$$

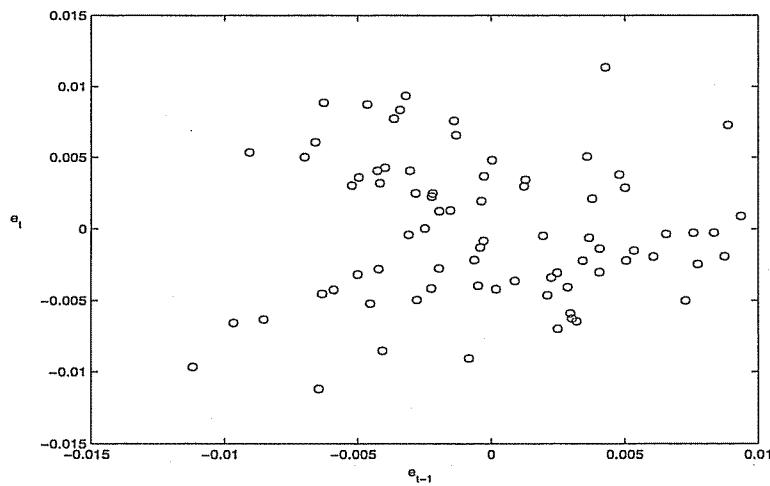
donde  $y_t = \nabla \log z_t$  es la tasa de crecimiento del PIB. La figura 14.9 muestra los residuos de este modelo con relación a su primer retardo. En esta figura se observa que la relación entre  $y_t$  y  $y_{t-1}$  es distinta cuando  $y_{t-1} \leq 0$  que cuando  $y_{t-1} > 0$ . Además, cuando  $y_{t-1} \leq 0$  se observa que los datos se agrupan alrededor de dos líneas de regresión distintas, que corresponden a dos regímenes distintos. Esto sugiere un modelo TAR. Tiao y Tsay (1994) identificaron estos dos regímenes y comprobaron que las dos rectas alrededor de las cuales se sitúan los puntos cuando  $y_{t-1} \leq 0$  corresponde a distintos valores de la variable  $y_{t-2}$ . De esta manera se obtiene un modelo TAR con cuatro umbrales, según que las variables  $y_{t-1}$  y  $y_{t-2}$  sean positivas o negativas. Véase Tiao y Tsay (1994) para un análisis de esta serie mediante un modelo por umbrales.

Peña y Rodríguez (2005b) estimaron el siguiente modelo para esta serie:

$$\begin{aligned} y_t &= .0031 + 1.0345y_{t-1} + \varepsilon_t^1 & y_{t-1} \leq .0084, \\ &\quad (.001) \quad (.1711) & \\ y_t &= .0045 + .2340y_{t-1} + \varepsilon_t^2 & y_{t-1} > .0084, \\ &\quad (.003) \quad (.2525) & \end{aligned}$$

con  $\varepsilon_t^1 \sim N(0, 1.87 \cdot 10^{-5})$  y  $\varepsilon_t^2 \sim N(0, 1.77 \cdot 10^{-5})$ . Este modelo muestra que la estructura de dependencia es distinta cuando la serie tiene una tasa de crecimiento positiva ( $y_{t-1} > .0084$ ) que cuando la tasa es casi cero o negativa ( $y_{t-1} \leq .0084$ ). En el primer caso la tasa de crecimiento sigue un modelo AR(1) con pequeña inercia, dado que el coeficiente AR es pequeño,  $\phi = .234$ . Sin embargo, en el segundo caso, cuando la tasa de crecimiento ha sido muy pequeña o negativa, la estructura de dependencia es mucho más fuerte,  $\phi = 1.0345$ , y el siguiente valor del proceso tiende a ser muy similar al de partida, con lo que la serie se mantiene en valores negativos.

**Figura 14.9** Residuos de un ajuste lineal de la serie del PIB de Estados Unidos frente a su retardo. Se observa que la dependencia por debajo del cero es distinta que por encima de cero. Por debajo de cero se aprecian además dos regímenes distintos



#### 14.5. Otros modelos no lineales

Los modelos TAR pueden generalizarse permitiendo una transición suave entre los regímenes. Se obtienen entonces los modelos STAR, o Smooth Threshold AutoRegresive. Hemos visto que un modelo TAR para un AR(1) con variable de umbral  $z_{t-1}$  puede escribirse como:

$$z_t = F(z_{t-1})(c^{(1)} + \phi^{(1)} z_{t-1} + \varepsilon_t^{(1)}) + (1 - F(z_{t-1}))(c^{(2)} + \phi^{(2)} z_{t-1} + \varepsilon_t^{(2)})$$

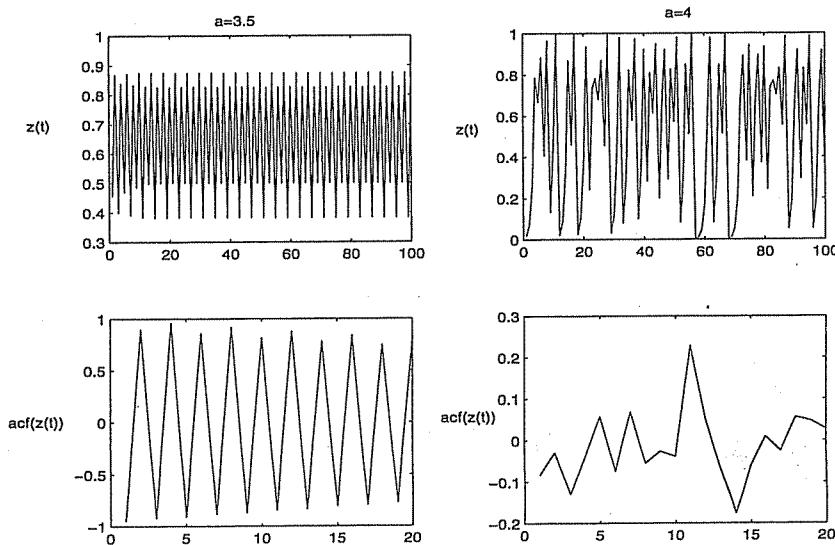
donde  $F(z_{t-1})$  es:

$$F(z_{t-1}) = \begin{cases} 0, & z_{t-1} - d \leq 0 \\ 1, & z_{t-1} - d > 0 \end{cases}$$

y vale cero para  $z_{t-1} < d$  y uno para  $z_{t-1} \geq d$ . La función  $F$  es como la función de distribución de una variable aleatoria pero tiene un salto muy brusco al llegar al punto  $z_{t-1} = d$ . Una alternativa es sustituir  $F$  por una función suave que toma valores en  $[0,1]$ . Por ejemplo, una función que se utiliza mucho es la logística

$$F(z_{t-1}) = \frac{1}{1 + e^{-\alpha(z_{t-1}-d)}},$$

**Figura 14.10** En la parte superior se encuentran dos realizaciones de la secuencia logística con distintos valores de  $a$  y en la parte inferior las acf de las series



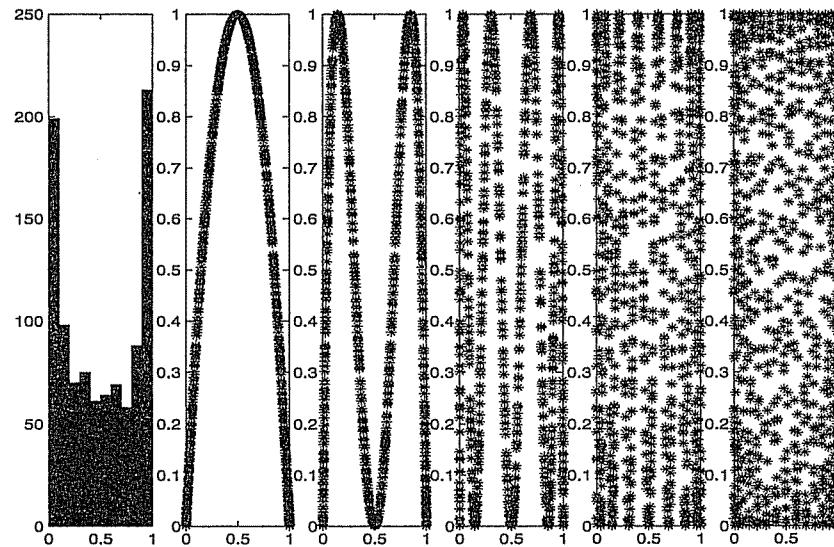
donde el parámetro  $\alpha$  es positivo. Si  $z_{t-1} < d$  tiende a menos infinito, la función tiende a cero, mientras que si  $z_{t-1} \geq d$  tiende a más infinito la función tiende a uno. La transición entre estos dos valores depende del valor de  $\alpha$ . Si este parámetro es muy grande se produce un cambio muy rápido al pasar  $z_{t-1} - d$  de negativo a positivo y tenemos un caso muy similar al TAR. Por el contrario, si  $\alpha$  es pequeño la transición será suave.

Existen series generadas por modelos deterministas pero que presentan un comportamiento aparentemente estocástico. Éstos son los modelos de caos. En estos modelos cada valor del proceso se genera de forma determinista a partir del anterior, pero su dependencia de las condiciones iniciales hace aparecer al proceso como estocástico. Por ejemplo, la figura 14.10 presenta una realización del proceso determinista:

$$z_t = az_{t-1}(1 - z_{t-1}),$$

que se denomina mapa logístico. Para  $0 \leq a \leq 4$  esta función proporciona siempre, a partir de un valor inicial entre cero y uno, valores sucesivos de la serie entre cero y uno. La figura 14.10 presenta dos realizaciones de esta secuencia. Se observa que para  $a = 3.5$  la serie presenta aspecto determinista, con gráfico de la serie muy regular y autocorrelaciones muy altas, pero para  $a = 4$  el proceso es bastante irregular y todas las auto-

**Figura 14.11 Relación entre los valores de la serie y sus primeros cinco retardos en el modelo logístico con  $a=4$**



correlaciones son cero, lo que puede confundirle con un proceso de ruido blanco.

El carácter determinista de una serie generada por un modelo caótico puede detectarse haciendo un gráfico de  $z_t$  respecto a sus retardos. Como indica la figura 14.11, al relacionar  $z_t$  con  $z_{t-1}$ , se observa claramente la relación funcional que une a ambos valores. Sin embargo, al aumentar la distancia entre los retardos, la relación determinista comienza a ser menos aparente, por el carácter caótico de la serie que la hace muy sensible a los valores iniciales. La relación entre  $z_t$  y  $z_{t-4}$  no muestra ya ningún indicio de relación entre las variables, y la relación entre  $z_t$  y  $z_{t-5}$  parece la de dos variables uniformes en  $(0,1)$  independientes.

Este ejemplo pone de nuevo de manifiesto la importancia de mirar a los gráficos entre los valores de la serie y sus retardos y no sólo a los coeficientes de autocorrelación, que son un resumen de la relación lineal presente en el gráfico, pero que no pueden dar información respecto a relaciones no lineales. Por ejemplo, en este gráfico las correlaciones entre la serie y sus retardos son nulas, aunque existe una relación fuerte de carácter no lineal entre la serie y sus retardos.

## 14.6. Contrastes de no linealidad

### 14.6.1. Tipos de contrastes

La identificación de la no linealidad se realiza habitualmente sobre los residuos de un modelo lineal. Los contrastes existentes pueden clasificarse en tres grupos. El primer grupo parte de que aunque los residuos están, por construcción, incorrelados con los valores retardados de la variable, pueden tener dependencia no lineal con los retardos. Podemos investigar esta dependencia estimando un modelo del tipo:

$$e_t = \sum \alpha_i z_{t-i}^2 + \sum \beta_{ij} z_{t-i} z_{t-j} + \epsilon_t, \quad (14.9)$$

que puede considerarse una aproximación a la relación general entre los residuos y los retardos de la variable. Observemos que no se incluyen términos lineales del tipo  $\alpha_i z_{t-i}$  por la incorrelación entre los residuos y los retardos de la variable. Si estimamos la relación (14.9) y encontramos coeficientes significativos, esto sugiere no linealidad.

El segundo tipo de contraste se basa en que los residuos de una serie no lineal a la que ajustamos un modelo lineal, aunque estén incorrelados, no serán independientes. En efecto, si por ejemplo existe la relación (14.9)  $e_t$  y  $e_{t-k}$  comparten información común y no serán independientes. Supongamos que existe una relación de dependencia del tipo  $e_t = g(e_{t-1})$ , donde  $e_{t-1} = (e_{t-1}, \dots, e_1)$ , podemos aproximar la función  $g$  mediante

$$e_t = \sum \alpha_i e_{t-i}^2 + \sum \beta_{ij} e_{t-i} e_{t-j} + \epsilon_t, \quad (14.10)$$

y ajustar una regresión entre los residuos y sus retardos al cuadrado. Si encontramos términos significativos esto sugiere no linealidad. Adicionalmente, podemos incluir términos de productos cruzados, pero inicialmente se comienza con los cuadrados de los residuos. Alternativamente, podríamos suponer que si existe una relación no lineal entre las variables, existirá una transformación de los residuos,  $h(e_t)$ , donde la relación no lineal entre los residuos y sus retardos aparezca más claramente. Entonces, buscaremos una transformación  $h$  para maximizar la dependencia:

$$h(e_t) = \sum \alpha_i h(e_{t-i}) + \epsilon_t.$$

La función  $h$  puede elegirse dentro de la familia de Box-Cox. Como caso particular, si elegimos como transformación los cuadrados estamos en el caso (14.10).

Finalmente, el tercer tipo de contraste es no paramétrico. En este caso no se supone ninguna forma funcional a la relación de los residuos con su pasado y se investiga si cuando los residuos toman valores similares, por

ejemplo en los instantes  $t_1$  y  $t_2$ , los residuos en los instantes siguientes,  $t_1+1, t_1+2, \dots$  y  $t_2+1, t_2+2, \dots$ , son también similares. En las secciones siguientes desarrollamos estos contrastes.

#### 14.6.2. Contrastos de regresión entre los residuos y funciones de la variable

Un contraste de este tipo ha sido propuesto por Tsay (1986). La idea del contraste es construir ciertas variables explicativas  $x_t$  que recojan los términos no lineales de  $z_t$  y sus retardos y relacionarlas con los residuos  $e_t$ . Si encontramos dependencia, aceptamos la presencia de estructura no lineal. El contraste se realiza como sigue:

1. Estimar un modelo autorregresivo con un retardo  $M$  largo. Sean

$$\hat{z}_t = \hat{\alpha}_0 + \sum_{i=1}^M \hat{\alpha}_i z_{t-i} \quad (14.11)$$

las predicciones generadas por el modelo estimado. Los residuos  $e_t$  de este ajuste lineal se calculan de la forma habitual con:

$$e_t = z_t - \hat{z}_t = z_t - \hat{\alpha}_0 - \sum_{i=1}^M \hat{\alpha}_i z_{t-i}. \quad (14.12)$$

2. Estimar una regresión entre los residuos  $e_t$  y las variables  $(z_{t-i} e_{t-i}, e_{t-i} e_{t-i-1})$  para  $i = 1, \dots, M$ , de la forma:

$$e_t = \sum_{i=1}^M a_i z_{t-i} + \sum_{i=1}^M b_i z_{t-i} e_{t-i} + \sum_{i=1}^M c_i e_{t-i} e_{t-i-1} + \epsilon_t.$$

Para justificar esta regresión; observemos que, según (14.12):

$$z_{t-i} e_{t-i} = z_{t-i}^2 - \hat{\alpha}_0 z_{t-i} - \sum_{j=1}^M \hat{\alpha}_j z_{t-i-j} z_{t-i}$$

y los productos entre la variable y los residuos incluyen los cuadrados y dobles productos de la variable menos los efectos lineales. Además:

$$e_{t-i} e_{t-i-1} = \hat{\alpha}_0^2 + z_{t-i} z_{t-i-1} - \hat{\alpha}_0 z_{t-i} - \hat{\alpha}_0 z_{t-i-1} + \text{dobles}$$

donde *dobles* incluye los dobles productos, incorpora también los productos cruzados como regresores. Como las variables introducidas para

explicar los residuos incluyen términos lineales con los retardos de la variable y los residuos deben estar incorrelados con ellos, para contrarrestar este efecto se introducen en la regresión entre los residuos y estas variables también los regresores  $(1, z_{t-1}, \dots, z_{t-M})$ . A continuación, se realiza el contraste F de regresión habitual y se rechaza la linealidad si este contraste es significativo. Véanse Peña, Tiao y Tsay (2000) y Peña y Rodríguez (2005a) para ejemplos y estudios de la potencia de este contraste.

#### 14.6.3. Contrastes sobre los cuadrados de los residuos

Si tenemos un modelo no lineal, los residuos del mejor ajuste lineal estarán incorrelados pero no serán independientes, ya que la información no modelizada quedará en los residuos. Si los residuos son independientes, las autocorrelaciones de sus cuadrados deben ser similares a los cuadrados de las autocorrelaciones originales. En efecto, Maravall (1983) demostró que si las  $\varepsilon_t$  son normales incorreladas y por tanto independientes:

$$\rho_k(\varepsilon_t^2) = [\rho_k(\varepsilon_t)]^2.$$

En consecuencia, como  $|\rho_k(\varepsilon_t)| \leq 1$ , si elevamos los residuos al cuadrado y calculamos sus autocorrelaciones, éstas –en la hipótesis de normalidad– no deberán aumentar con relación a las de la serie original.

McLeod y Li (1983) han propuesto un contraste de esta propiedad. Suponiendo que los residuos estimados son normales, las autocorrelaciones de sus cuadrados definidas por:

$$r_k(e_t^2) = \frac{\sum_{t=k+1}^T (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{t-k}^2 - \hat{\sigma}^2)}{\sum_{t=1}^n (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)^2}, \quad (k = 1, 2, \dots, m),$$

donde  $\hat{\sigma}^2 = \sum \hat{\varepsilon}_t^2 / T$ , serán asintóticamente variables normales incorreladas de media cero y varianza  $1/T$ . Por tanto el estadístico

$$Q_{LB}(\hat{\varepsilon}_t^2) = T(T+2) \sum_{k=1}^m (T-k)^{-1} r_k^2(\hat{\varepsilon}_t^2),$$

seguirá una distribución  $\chi^2$  con  $m$  grados de libertad. Observamos que en el contraste de los cuadrados de los residuos no hay que restar el número de parámetros estimados.

Un contraste de no linealidad más potente ha sido propuesto por Peña y Rodríguez (2005a), que utiliza que el estadístico  $D_m$  dado por

$$D_m = -\frac{T}{m+1} \log |\mathbf{R}_m|,$$

donde  $\mathbf{R}_m$  es la matriz de autocorrelación de los residuos al cuadrado, dada por

$$\mathbf{R}_m = \begin{bmatrix} 1 & r_1 & \cdots & r_m \\ r_1 & 1 & \cdots & r_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_m & r_{m-1} & \cdots & 1 \end{bmatrix}. \quad (14.13)$$

En la hipótesis de linealidad este estadístico sigue una distribución gamma  $\mathcal{G}(\alpha, \beta)$  con  $\alpha = 3m(m+1)/4(2m+1)$  y  $\beta = 3m/2(2m+1)$ .

Observemos que estos contrastes pueden resultar significativos cuando existe no linealidad en la media, pero también cuando exista no linealidad en la varianza. Por ejemplo, los modelos de heterocedasticidad condicional que estudiaremos en el próximo capítulo suelen detectarse con este contraste.

---

### Ejemplo 14.2

Vamos a analizar la serie de manchas solares (sunpot.dat). Un modelo lineal que parece representar adecuadamente esta serie es:

$$z_t = 44.77 + 0.86 z_{t-1} - 0.28 z_{t-3} + 0.18 z_{t-9} + a_t + 0.35 a_{t-1}. \quad \begin{matrix} (7.97) \\ (18.9) \\ (6.19) \\ (4.45) \\ (5.04) \end{matrix}$$

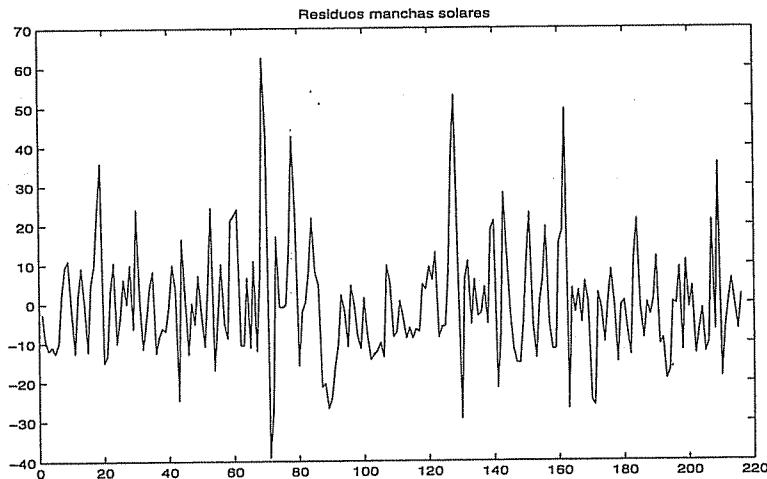
Se observa una dependencia del último mes, y después cada tres meses. Los residuos de este modelo se presentan en la figura 14.12 y parecen ruido blanco. El estadístico  $Q(31) = 37.15$ , por lo que el modelo parece adecuado. Sin embargo, si hacemos el correlograma de los residuos al cuadrado se observan coeficientes significativos en los dos primeros retardos y el estadístico  $Q_{LB}(\hat{\epsilon}_t^2)$  con 12 grados de libertad es 31.46, que indica sin lugar a dudas falta de linealidad en esta serie.

---

#### 14.6.4. Contrastos no paramétricos sobre la dependencia de los residuos. El contraste BDS

Un contraste no paramétrico que se basa en la relación no lineal de los residuos del ajuste lineal y de su pasado es debido a Brock et al. (1991, 1996), y se conoce como contraste BDS. La idea del contraste es simple: si existe relación entre los residuos y su pasado y tenemos dos bloques de datos consecutivos,  $(e_{t_1+1}, \dots, e_{t_1+k})$  y  $(e_{t_2+1}, \dots, e_{t_2+k})$ , que están próximos entre sí, es esperable que los valores futuros que siguen a ambas secuencias,  $(e_{t_1+k+1}, \dots, e_{t_1+k+g})$  y  $(e_{t_2+k+1}, \dots, e_{t_2+k+g})$ , también estén próximos. Para implementar esta idea necesitamos determinar los

Figura 14.12 Residuos de la serie de manchas solares



parámetros  $k$  y  $g$  y elegir una medida de proximidad entre bloques de observaciones. Alternativamente, podemos realizar el contraste para varios valores de  $k$  y  $g$ , con lo que el problema se reduce a definir la distancia entre bloques. En este test se toma la máxima distancia euclídea entre observaciones de ambos bloques que ocupan la misma posición. Es decir, dados dos bloques de observaciones de tamaño  $k$ ,

$$(e_{t_1+1}, \dots, e_{t_1+k}), \quad (e_{t_2+1}, \dots, e_{t_2+k})$$

se forman las distancias entre observaciones que ocupan la misma posición en el bloque,  $d_j = |e_{t_1+j} - e_{t_2+j}|$  para  $j = 1, \dots, k$ . La distancia entre bloques se define como el máximo de estas distancias, y los bloques se consideran próximos si  $\max(d_j) \leq c$ .

El test se realiza entonces como sigue: para  $t = 0, \dots, n - k$ , formamos secuencias de  $k$  elementos  $(e_{t+1}, \dots, e_{t+k})$  y para cada una de ellas contamos cuantas de las restantes secuencias están próximas a la considerada. Para resumir el resultado de la comparación entre dos secuencias de tamaño  $k$ , una comenzando en  $t$  y otra en  $s$ , se define una variable ficticia,  $C_{t,s}$ , que toma el valor uno si ambas secuencias están próximas y cero en otro caso. La proporción de secuencias de tamaño  $k$  que están próximas se calcula mediante:

$$C_{k,T} = \frac{2}{(T-k)(T-k-1)} \sum_{t=1}^{T-k} \sum_{s=t+1}^{T-k-1} C_{t,s},$$

**Tabla 14.1 Modelos para distintos periodos del PIB en Estados Unidos**

Model	Period	Size	AR	BIC	$10^{-4}\hat{\sigma}^2$	AO	TC	LS	$QLB(10)$
$M_1$	1/47-2/03	226	0.342 (0.063)	-9.28	0.900	-	-	-	11.39
$M_2$	1/47-2/03	226	0.423 (0.060)	-9.46	0.693	4/49	1/58	2/78	12.57
$M_3$	1/47-1/75	113	0.468 (0.083)	-9.09	0.989	4/49	-	-	8.26
$M_4$	2/75-2/03	113	0.477 (0.083)	-10.01	0.380	2/81	-	2/78	9.54
								2/80	

y el estadístico para el contraste BDS es el valor estandarizado de la proporción de secuencias próximas, dado por :

$$w_{k,T} = \sqrt{T-k-1} \frac{(C_{k,T} - C_{1,T-k+1}^k)}{\sigma_{k,T-k+1}}$$

Se demuestra que, bajo la hipótesis de independencia de los residuos, este estadístico sigue una distribución normal. Para utilizar este test es necesario definir  $k$ , el tamaño de la secuencia, y el nivel crítico para las distancias,  $c$ .

---

### Ejemplo 14.3

Vamos a aplicar estos contrastes a la serie del PIB en Estados Unidos. La tabla 14.1 presenta algunos modelos lineales con y sin atípicos estimados para esta serie y está tomada de Peña y Rodríguez (2005b). Se han estimado modelos para la serie completa, con y sin atípicos, y después se ha dividido la serie en dos mitades y se ha estimado un modelo para cada parte. Podemos concluir al comparar los modelos para todo el periodo,  $M_1$  con  $M_2$ , que la serie tiene un importante cambio de nivel y además otros atípicos. Comparando los modelos para el primer y el segundo periodo se observa que la varianza del segundo periodo es sólo 1/3 de la del primero, lo que sugiere un cambio estructural en esta serie.

En la tabla 14.2 se presentan los resultados obtenidos al aplicar contrastes de linealidad a los residuos de los modelos anteriores. Al aplicar el contraste a todo el periodo sin limpiar por atípicos, modelo  $M_1$ , se observa que la mayoría

**Tabla 14.2 Contrastes de no linealidad aplicados a los residuos del PIB de Estados Unidos en diferentes periodos**

Model	Period	$F_{Tsay}$	$D_m$	$Q_{LB}$	BDS
$M_1$	1/47-2/03(NO)	L	NL	NL	NL
$M_2$	1/47-2/03	L	NL	NL	L
$M_3$	1/47-1/75	L	L	L	L
$M_4$	2/75-2/03	NL	NL	NL	NL

de los contrastes detectan no linealidad. Si tenemos en cuenta los atípicos, modelo  $M_2$ , la presencia de no linealidad es algo menor, pero todavía dos de los cuatro contrastes nos la indican. Al dividir la serie en dos períodos la situación se clarifica: el primer periodo es claramente lineal, mientras que el segundo es no lineal. Los cuatro contrastes presentados se han aplicado a los residuos de los modelos estimados y en las columnas se observa el resultado de cada test. Se ha utilizado el símbolo L para indicar que no se rechaza la hipótesis de linealidad y NL para indicar que el contraste rechaza dicha hipótesis y concluimos que la serie es no lineal.

#### 14.7. Lecturas complementarias

Los libros de Priestly (1988), Tong (1990) y Fang y Yao (2002) son referencias básicas para estos modelos. El primero presenta con detalle los modelos bilineales y TAR, pero no incluye los avances recientes de contrastes de no linealidad. El libro de Tong es una referencia general muy completa para estos modelos. El libro de Fang y Yao incluye el tratamiento no paramétrico de los modelos no lineales. Granger y Anderson (1978) estudian los modelos bilineales con detalle y Tong (1983) los autorregresivos por umbrales. Franses y van Dijk (2000) investigan la aplicación de estos modelos en finanzas.

#### Ejercicios 14

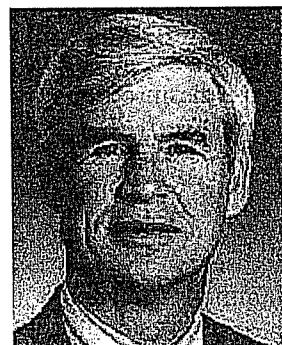
- 14.1 Justificar si es posible que el proceso  $z_t = \phi z_{t-1} z_{t-2} + a_t$  sea estacionario.
- 14.2 Demostrar que el proceso  $z_t = \phi z_{t-k} a_{t-p} + a_t$  es de ruido blanco si  $k > p$ .
- 14.3 Demostrar que en el proceso  $z_t = \phi z_{t-2} a_{t-1} + a_t$  los cuadrados del proceso tienen estructura de autocorrelación.

Análisis de series temporales

---

- 14.4 Demostrar que en el proceso TAR(1)  $z_t = -\phi z_{t-1} + a_t$  si  $z_{t-1} < 0$  y  $z_t = \phi z_{t-1} + a_t$  si  $z_{t-1} \geq 0$ , donde  $a_t$  es ruido blanco, se verifica que la función de densidad marginal es  $f(z_t) = .5N(-\phi z_{t-1}, \sigma) + .5N(\phi z_{t-1}, \sigma)$  y calcular la media y la varianza marginal.
- 14.5 Simular el proceso anterior con  $\phi = .7$  y estudiar su estructura de autocorrelación.
-

## 15. Modelos de heterocedasticidad condicional



Robert F. Engle (1942)

Económetra estadounidense. Estudió inicialmente Ciencias Físicas en Cornell, donde se doctoró en Economía en 1969. En el periodo 1975-2001 fue profesor del Departamento de Economía en la Universidad de California en San Diego y actualmente es profesor en la Universidad de Nueva York. Desarrolló con Granger el concepto de cointegración e inventó los procesos ARCH. Uno de los creadores de la econometría financiera, su trabajo ha influido en muchas áreas del análisis de series temporales y la econometría. En 2003 recibió el Premio Nobel de Economía.

### 15.1. Introducción

En los procesos lineales ARMA tanto la varianza marginal,  $\text{var}(z_t)$ , como la varianza condicional,  $\text{var}(z_t|z_{t-1}, \dots)$ , son constantes. En este capítulo vamos a estudiar procesos estacionarios que son no lineales en la varianza porque, aunque su varianza marginal es constante, la varianza condicionada a sus valores pasados no lo es. Esta no linealidad de la varianza puede darse además de otras formas de no linealidad en la media, como las estudiadas en el capítulo anterior. Como la varianza condicionada representa la incertidumbre de las predicciones, estos procesos van a reflejar un riesgo variable, entendiendo el riesgo como la incertidumbre de las predicciones en el tiempo.

Estos procesos fueron propuestos para explicar ciertas propiedades que no pueden explicarse por los modelos ARIMA y que aparecen con frecuencia en series temporales estacionarias de datos financieros y ambientales de alta frecuencia, como datos diarios, horarios, etc. Normalmente,

estas series tienen poca estructura en la media y siguen paseos aleatorios, o procesos AR de orden bajo y coeficiente pequeño. Como ejemplo, consideremos los rendimientos diarios del índice Ibex 35 medidos por la variable  $z_t = 100(\log X_t - \log X_{t-1})$ , donde  $X_t$  es el índice Ibex 35 de la Bolsa de Madrid. La figura 15.1 (a) incluye 1982 observaciones de esta serie entre enero de 1992 y diciembre de 1999. Se observa que la serie tiene poca estructura y que presenta períodos de distinta varianza. Por ejemplo, la última parte de la serie tiene mayor varianza que la parte central. La figura 15.1 (b) presenta la función de autocorrelación de estos datos. Existe una débil autodependencia, ya que  $r_1 = .11$  es significativamente distinto de cero, pero aparte de este primer coeficiente la serie no tiene estructura. Hay algún coeficiente que sale fuera de los límites del 95 %, pero esto es esperable al considerar 50 coeficientes. La figura 15.1 (c) presenta el histograma de estos datos y se observa una distribución simétrica aunque con colas más pesadas que la normal: el coeficiente de curtosis de estos datos es 6.88, más del doble de la distribución normal. Finalmente, la figura 15.1 (d) muestra la función de autocorrelación de los cuadrados de las observaciones. La función tiene muchos coeficientes distintos de cero, con una fuerte estructura de dependencia.

Podemos concluir que, aunque la serie de rendimientos parece casi ruido blanco:

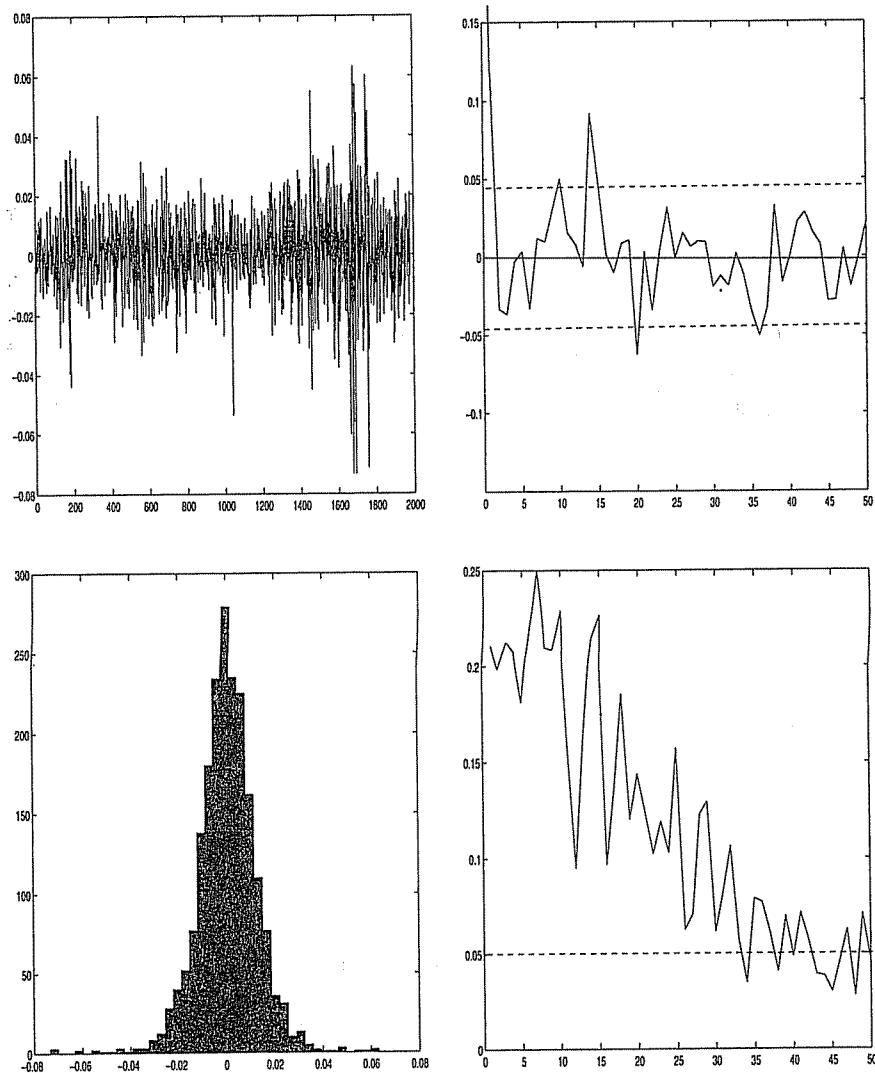
1. Su distribución no es normal, y muestra colas pesadas y alta curtosis.
2. Los datos están casi incorrelados pero al calcular las autocorrelaciones de los cuadrados se observa una fuerte estructura de dependencia.

Una tercera propiedad es que la varianza de los residuos no es constante y aparecen rachas de mayor variabilidad seguidas por otras de menor variabilidad. Esta característica, que se observa en el gráfico de los datos, se aprecia más claramente en la figura 15.2 (a), que presenta la serie de los cuadrados de los rendimientos, y en la figura 15.2 (b), que presenta las varianzas calculadas en grupos de 100 datos de la serie.

Se han propuesto distintos modelos para explicar estos efectos. Los modelos que vamos a estudiar en este capítulo suponen que, aunque la varianza marginal es constante, lo que es necesario para que el proceso sea estacionario, la varianza condicionada a los valores pasados de la serie no es constante, ya que depende de estos valores previos. A los modelos con esta propiedad se les denomina con *varianza condicional heterocedástica* (no constante) o modelos de *heterocedasticidad condicional*.

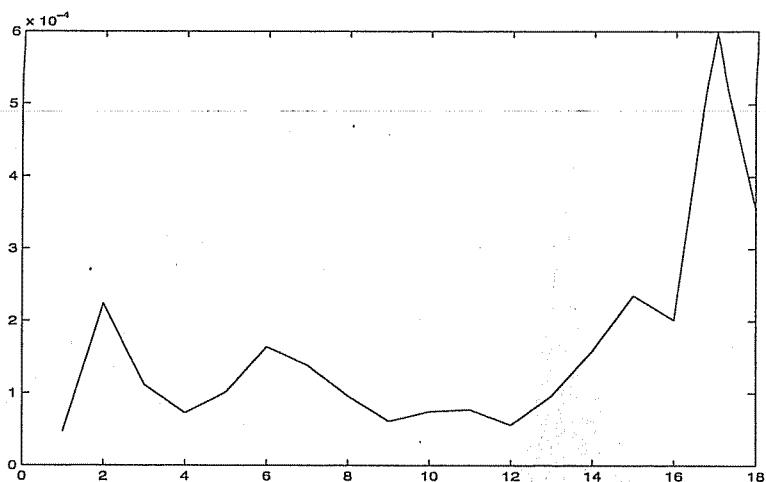
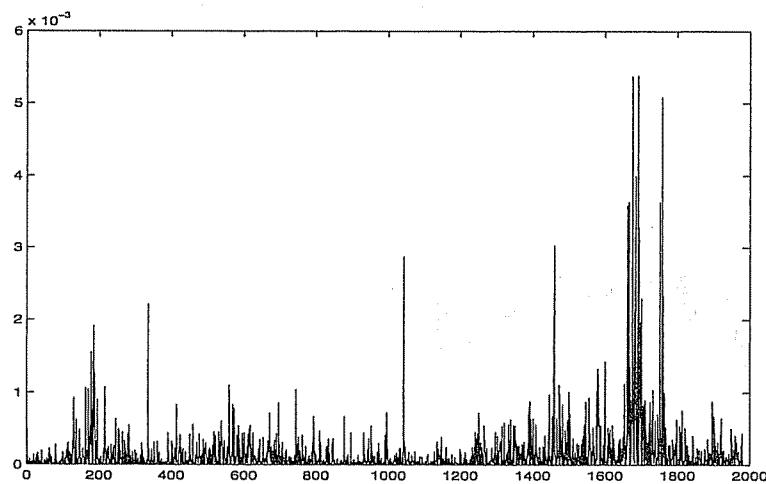
El modelo ARCH (iniciales de AutoRegressive Conditional Heteroscedastic), introducido por Engle (1982), supone que la varianza condicional

**Figura 15.1** Análisis de la serie de Ibex de rendimientos diarios:  
gráfico temporal, correlograma, histograma de la serie  
y correlograma de la serie al cuadrado



depende del pasado con estructura autorregresiva. Estos modelos fueron generalizados por Bollerslev (1986) para dar lugar a los modelos GARCH, que incorporan a esta dependencia términos de media móvil. Una clase de modelos más flexibles son los modelos de volatilidad estocástica, introducidos por Harvey, Ruiz y Shephard (1994), y Jacquier, Polson y

**Figura 15.2 Cuadrados de los rendimientos del Ibex y variación de la variabilidad**



Rossi (1994). Todos estos modelos suponen efectos simétricos, es decir, la dependencia se establece en función de las magnitudes de las observaciones pasadas y no depende de su signo. Esta hipótesis puede no ser cierta, y en la última sección mencionaremos brevemente otros modelos que no suponen esta restricción.

## 15.2. Modelos ARCH

Supongamos una serie de ruido blanco,  $e_t$ , que puede ser la serie de residuos de un modelo ARMA. Si las variables  $e_t$  fuesen normales, la incorrelación garantiza la independencia, y no puede existir información en el pasado de la serie que sea útil para explicar su futuro. Los modelos ARCH relajan la hipótesis de normalidad y permiten que tengamos procesos de ruido blanco formados por variables dependientes. Una clase de modelos con esta propiedad es:

$$e_t = \sigma_t \epsilon_t \quad (15.1)$$

donde  $\epsilon_t$  y  $\sigma_t$  son dos procesos estacionarios independientes entre sí. El proceso  $\epsilon_t$  es de ruido blanco normal estandarizado, es decir, formado por variables normales independientes de media cero y varianza unidad. El proceso  $\sigma_t$  es estacionario, pero tiene estructura dinámica, siendo su valor en  $t$  función del conjunto  $\mathbf{e}_{t-1} = (e_{t-1}, \dots, e_1)$  de los valores de la serie previos a  $t$ . La condición de independencia entre  $\epsilon_t$  y  $\sigma_t$  garantiza que la serie  $e_t$  tenga media marginal igual a cero, ya que:

$$E(e_t) = E(\sigma_t)E(\epsilon_t) = 0,$$

y también media condicional nula, ya que:

$$E(e_t | \mathbf{e}_{t-1}) = E(\sigma_t | \mathbf{e}_{t-1})E(\epsilon_t) = 0.$$

Como el proceso  $e_t$  es estacionario, tendrá una varianza marginal constante, que llamaremos  $\sigma^2$ . Esta varianza se calcula mediante:

$$E(e_t^2) = E(\sigma_t^2)E(\epsilon_t^2) = E(\sigma_t^2) = \sigma^2 \quad (15.2)$$

y coincide con la varianza del proceso  $\sigma_t$ , ya que  $E(\epsilon_t^2) = 1$ . Sin embargo, el proceso  $e_t$  tendrá una varianza condicionada que no es constante. En efecto:

$$Var(e_t^2 | \mathbf{e}_{t-1}) = E(\sigma_t^2 | \mathbf{e}_{t-1})E(\epsilon_t^2) = \sigma_t^2$$

dado que  $E(\epsilon_t^2 | \mathbf{e}_{t-1}) = E(\epsilon_t^2) = 1$ . Por tanto, el proceso  $\sigma_t^2$  tiene una interesante interpretación: representa la varianza condicionada de la serie en cada instante, que va variando en el tiempo con cierta estructura estacionaria.

La independencia entre los procesos  $\epsilon_t$  y  $\sigma_t$  garantiza que la serie  $e_t$  carece de autocorrelación y forma un proceso de ruido blanco. En efecto, las autocovarianzas de la serie son:

$$E(e_t e_{t-k}) = E(\sigma_t \epsilon_t \sigma_{t-k} \epsilon_{t-k}) = E(\epsilon_t)E(\sigma_t \sigma_{t-k}) = 0,$$

ya que el proceso  $\epsilon_t$  es independiente de su pasado y de los valores previos y actuales del proceso  $\sigma_t$ . Por tanto, el proceso  $e_t$  tendrá autocovarianzas nulas. Observemos, sin embargo, que, a diferencia de las hipótesis habituales de los modelos ARIMÁ, la serie  $e_t$  no es de variables independientes. En efecto, como  $e_t$  y  $e_{t-1}$  son función de valores comunes pasados de la serie, aunque estén incorrelados serán, en general, dependientes.

En resumen, los modelos ARCH generalizan las hipótesis habituales de un ruido blanco permitiendo dependencia, que se introduce haciendo que la varianza condicionada no sea constante. Vamos a estudiar el comportamiento de este modelo en los casos más simples.

### 15.2.1. El modelo ARCH(1)

El modelo ARCH(1) supone que la varianza condicional,  $\sigma_t^2$ , tiene una estructura similar a un AR(1), y depende sólo del último valor observado, mediante la ecuación:

$$E(e_t^2 | e_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2, \quad (15.3)$$

donde  $\alpha_0 > 0$  y  $\alpha_1 \geq 0$ , para asegurar que esta varianza es positiva. Esta ecuación establece que si el valor de  $e_{t-1}^2$  es alto, la varianza  $\sigma_t^2$  de la siguiente observación condicionada a este valor de  $e_{t-1}^2$  será también alta, haciendo más probable que el valor siguiente,  $e_t^2 = \sigma_t^2 \epsilon_t^2$ , sea alto. Este efecto va a producir, como veremos después, correlación entre los cuadrados de la serie, que se traducirá en rachas de valores de magnitud relativamente elevada, o con mayor varianza. Por otro lado, como la media marginal y condicionada es cero, aunque la varianza condicionada sea alta, siempre es posible que aparezca un valor pequeño de  $e_t^2$ , que disminuirá la varianza condicionada de la observación siguiente y facilitará que la siguiente observación sea pequeña en valor absoluto. De esta manera, la serie puede presentar rachas de valores altos, pero globalmente seguirá un proceso estacionario.

Es intuitivo que la varianza marginal de la serie, que es en definitiva según (15.2) el promedio de las varianzas condicionales, debe ser mayor que  $\alpha_0$  y será tanto mayor cuanto mayor sea el coeficiente  $\alpha_1$  que transmite el efecto de la última observación. En efecto, llamando  $\sigma^2 = E(e_t^2)$  a la varianza marginal, podemos escribir:

$$\sigma^2 = E[E(e_t^2 | e_{t-1})] = \alpha_0 + \alpha_1 E(e_{t-1}^2), \quad (15.4)$$

donde hemos utilizado la esperanza iterada que se explicó en el apéndice 8.1: podemos obtener la esperanza de una variable calculando: primero, la esperanza condicionada a otra variable (que puede ser vectorial como en este caso ya que  $e_{t-1}$  contiene todas las variables previas) y segundo,

tomando la esperanza del resultado respecto a la variable condicionante. La esperanza  $E(e_t^2|e_{t-1})$  es, por (15.3), igual a  $\alpha_0 + \alpha_1 e_{t-1}^2$ , y su esperanza respecto a la distribución de la variable  $e_{t-1} = (e_{t-1}, \dots, e_1)$ , es la misma que respecto a la distribución de  $e_{t-1}$ , ya que las restantes variables no intervienen en  $\alpha_0 + \alpha_1 e_{t-1}^2$ . La esperanza de la constante  $\alpha_0$  es ella misma y la esperanza de  $e_{t-1}^2$  respecto a su distribución es su varianza marginal. Como  $e_{t-1}$  es un proceso estacionario, tenemos que  $E(e_{t-1}^2) = E(e_t^2) = \sigma^2$ , y sustituyendo este resultado en (15.4), se obtiene que:

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1} \quad (15.5)$$

y, por lo tanto,  $0 \leq \alpha_1 < 1$ .

#### 15.2.1.1. La autocorrelación de los cuadrados

Vamos a comprobar que el modelo ARCH(1) establece una dependencia de tipo AR(1) entre los cuadrados de las observaciones. Para ello, definamos un nuevo proceso de ruido blanco,  $v_t$ , formado por variables estacionarias incorreladas de media cero y varianza marginal constante:

$$v_t = e_t^2 - \sigma_t^2 = \sigma_t^2(\epsilon_t^2 - 1), \quad (15.6)$$

las variables  $v_t$  tienen media cero, ya que:

$$E(v_t) = E(e_t^2) - E(\sigma_t^2) = \sigma^2 - \sigma^2 = 0,$$

y varianza constante:

$$var(v_t) = E(e_t^2 - \sigma_t^2)^2 = E(e_t^4) + E(\sigma_t^4) - 2E(e_t^2\sigma_t^2). \quad (15.7)$$

Para comprobar que esta expresión es una constante, observemos primero que como  $e_t = \sigma_t \epsilon_t$ , y los procesos  $\sigma_t$  y  $\epsilon_t$  son independientes:

1. La esperanza de  $E(e_t^4) = E(\epsilon_t^4 \sigma_t^4)$  será el producto de las esperanzas de  $E(\epsilon_t^4)$  y  $E(\sigma_t^4)$ . Como el proceso  $\epsilon_t$  es de variables normales e independientes se verifica que  $E(\epsilon_t^4) = 3$ , y  $E(e_t^4) = 3E(\sigma_t^4)$ .
2. La esperanza de  $E(e_t^2\sigma_t^2)$ , sustituyendo  $e_t = \sigma_t \epsilon_t$ , es igual a  $E(\epsilon_t^2\sigma_t^4) = E(\epsilon_t^2)E(\sigma_t^4) = E(\sigma_t^4)$ , ya que  $\epsilon_t$  tiene varianza unidad.

Sustituyendo estos resultados en (15.7), resulta que:

$$var(v_t) = 2E(\sigma_t^4),$$

y como  $\sigma_t$  es un proceso estacionario, sus momentos son constantes, y hemos demostrado que el proceso  $v_t$  es homocedástico. Sin embargo, el

proceso  $v_t$  es condicionalmente heterocedástico y está formado por variables incorreladas, pero no independientes. Para comprobarlo, observemos que tienen media cero y su covarianza será, utilizando (15.6):

$$E(v_t v_{t-k}) = E[\sigma_t^2(\epsilon_t^2 - 1)\sigma_{t-k}^2(\epsilon_{t-k}^2 - 1)],$$

y, por la independencia entre los procesos  $\epsilon_t^2$  y  $\sigma_t^2$ , y aplicando que al tener  $\epsilon_t$  varianza unidad  $E(\epsilon_{t-k}^2 - 1) = 0$ ,

$$E(v_t v_{t-k}) = E(\epsilon_t^2 - 1)E(\sigma_t^2 \sigma_{t-k}^2 (\epsilon_{t-k}^2 - 1)) = 0.$$

Hemos demostrado que el proceso  $v_t$  es de ruido blanco, formado por variables incorreladas de media cero y varianza marginal constante.

Escribiendo ahora, por (15.6),

$$\epsilon_t^2 = \sigma_t^2 + v_t,$$

y sustituyendo  $\sigma_t^2$  por su expresión (15.3), tenemos la siguiente relación entre los cuadrados de las observaciones de la serie:

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + v_t, \quad (15.8)$$

que establece una dependencia markoviana de primer orden, similar a un proceso AR(1), sobre los cuadrados. La ecuación (15.8) descompone la variable  $\epsilon_t^2$  en dos componentes ortogonales. Sin embargo, presenta dos diferencias principales respecto a un proceso AR(1) de variables normales. Con variables normales las innovaciones son independientes entre sí y, además, la innovación en el instante  $t$  es independiente de los valores pasados de la variable. Sin embargo, en la ecuación (15.8):

1. Las variables  $v_t$  están incorreladas, pero no son independientes entre sí.
2. Las variables  $v_t$  están incorreladas, pero no son independientes, de las variables retardadas de la ecuación,  $\epsilon_{t-1}^2$ .

En efecto, las variables  $v_t$  no van a ser independientes ya que, por su definición (15.6),  $v_t$  es función de  $\epsilon_t^2$  y  $\sigma_t^2$ , y  $v_{t-1}$  lo será de  $\epsilon_{t-1}^2$  y  $\sigma_{t-1}^2$ . Como  $\sigma_t^2$  depende de  $\epsilon_{t-1}^2$ , esto genera dependencia entre ellas. Para comprobar que  $v_t$  depende de  $\epsilon_{t-1}^2$ , observemos que en (15.8)  $\epsilon_t^2$  debe ser no negativo, por lo que la suma de  $\alpha_0 + \alpha_1 \epsilon_{t-1}^2$ , que siempre es positiva y  $v_t$ , que puede ser positiva o negativa, tiene que ser no negativa. En consecuencia, la variable  $v_t$  tendrá que satisfacer que:

$$v_t \geq -\alpha_0 - \alpha_1 \epsilon_{t-1}^2.$$

Sin embargo, es fácil comprobar que, aunque  $v_t$  no es independiente de  $e_{t-1}^2$ , está incorrelada con esta variable. Su covarianza es

$$E(v_t e_{t-1}^2) = E(\sigma_t^2 (\epsilon_t^2 - 1) \sigma_{t-1}^2 \epsilon_{t-1}^2) = E(\epsilon_t^2 - 1) E(\sigma_t^2 \sigma_{t-1}^2) = 0.$$

En el apéndice 15.1 se comprueba que el modelo AR(1) en los cuadrados establece esta estructura de autocorrelación entre los cuadrados de las variables. Llamando  $\rho_c(k)$  a la función de autocorrelación de los cuadrados de la serie, donde el subíndice  $c$  se refiere a los cuadrados, se obtiene:

$$\rho_c(k) = \alpha_1 \rho_c(k-1),$$

que indica que las autocorrelaciones de los cuadrados de las series tienen la estructura de un AR(1) con parámetro  $\alpha_1$ , de acuerdo con la ecuación (15.8).

#### 15.2.1.2. Curtosis

Se demuestra en el apéndice 15.1 que el modelo ARCH(1) conducirá a distribuciones con colas pesadas, y que el coeficiente de curtosis viene dado por:

$$K = \frac{3\alpha_0^2(1-\alpha_1)^2 + 6\alpha_0^2\alpha_1(1-\alpha_1)}{\alpha_0^2(1-3\alpha_1^2)} = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2}.$$

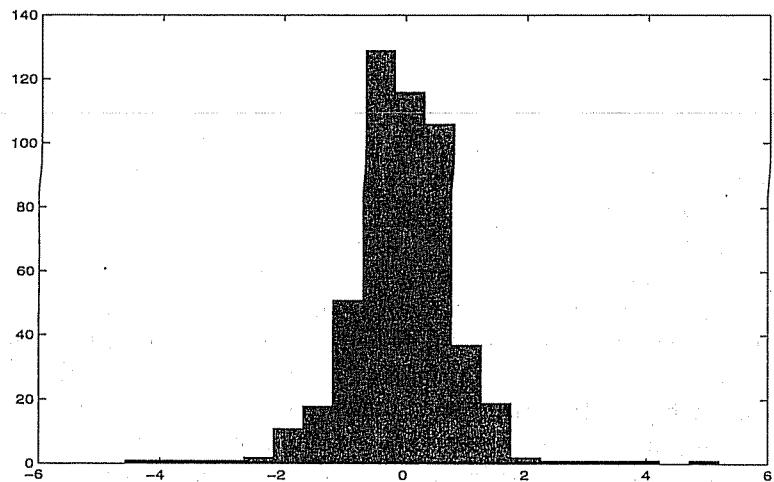
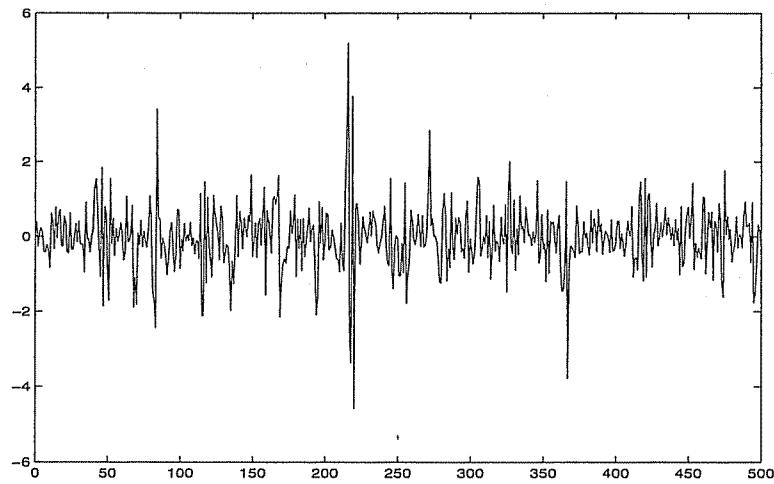
Como  $\alpha_1 > 0$ , este coeficiente de curtosis es siempre mayor que 3, y puede ser mucho mayor. Por lo tanto, la distribución marginal resultante será de colas pesadas.

La figura 15.3 presenta un ejemplo de un proceso ARCH(1) y de su distribución marginal. En el gráfico de la serie se observan períodos de mayor variabilidad y, aparentemente, bastantes valores atípicos. El histograma de los datos muestra que la distribución es simétrica, pero tiene alta curtosis o colas pesadas por la presencia de estos posibles valores típicos. La figura 15.4 presenta la función de autocorrelación muestral del proceso, que indica que éste es ruido blanco, y la *fas* muestral de los valores del proceso al cuadrado, que indica la estructura AR(1) de los cuadrados.

#### Ejemplo 15.1

Vamos a construir un modelo para la serie de datos diarios de rendimientos del indicador de Bolsa Ibex, que hemos utilizado en la introducción (dibex99.dat).

Figura 15.3 Gráfico del proceso ARCH(1) con alfa0=.2 y alfa1=.7 e histograma de los datos

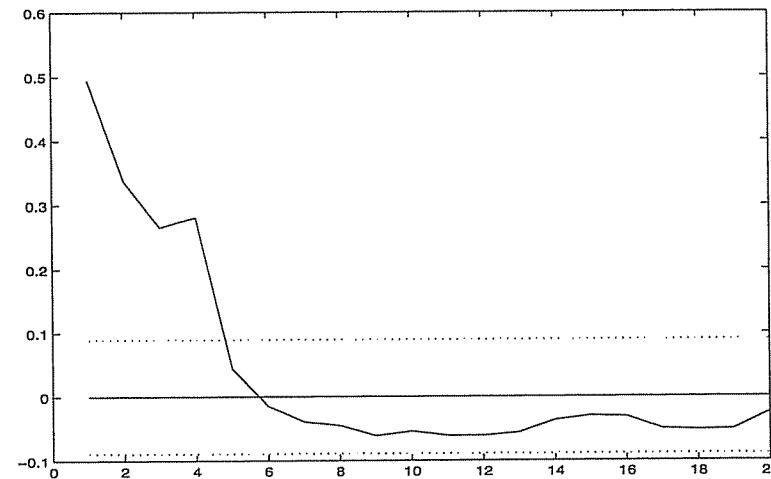
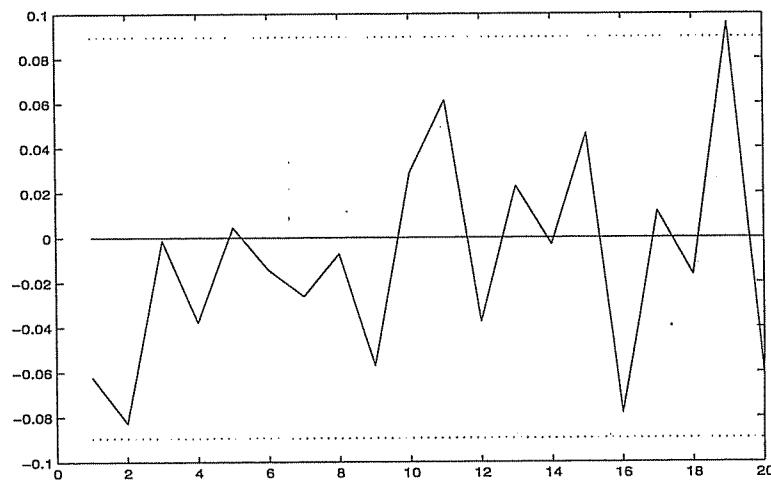


La *fas* y *fap* de los rendimientos se presenta en la figura 15.5. Se observa un coeficiente significativo en el primer retardo, y en la *fap* también el segundo. Esto parece indicar una AR(2) y el modelo estimado es:

$$z_t = .007 + .1257z_{t-1} - .0481z_{t-2} + \hat{a}_t; \quad \hat{\sigma} = .0123$$

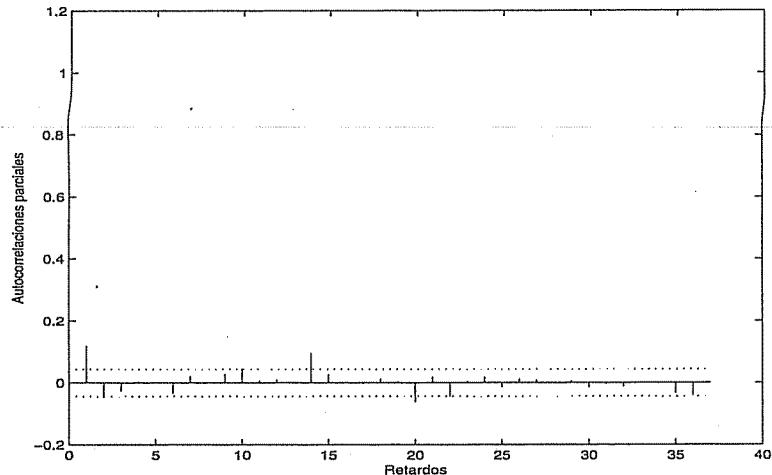
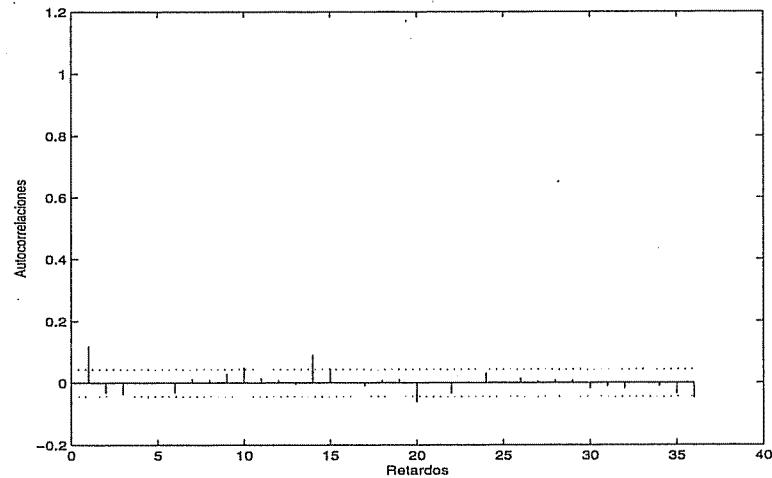
## 15. Modelos de heterocedasticidad condicional

Figura 15.4 Correlograma del proceso ARCH(1) con  $\alpha_0=.2$  y  $\alpha_1=.7$  y correlograma de los cuadrados



Los residuos de este modelo no presentan autocorrelación, por lo que la serie es ruido blanco. Sin embargo, si calculamos la autocorrelación de sus cuadrados nos encontramos que éstos no son ruido blanco. La figura 15.7, que presenta las *fas* y *fap* de los cuadrados de estos residuos, muestra en la *fap* de

Figura 15.5 Fas y fap de los rendimientos del Ibex



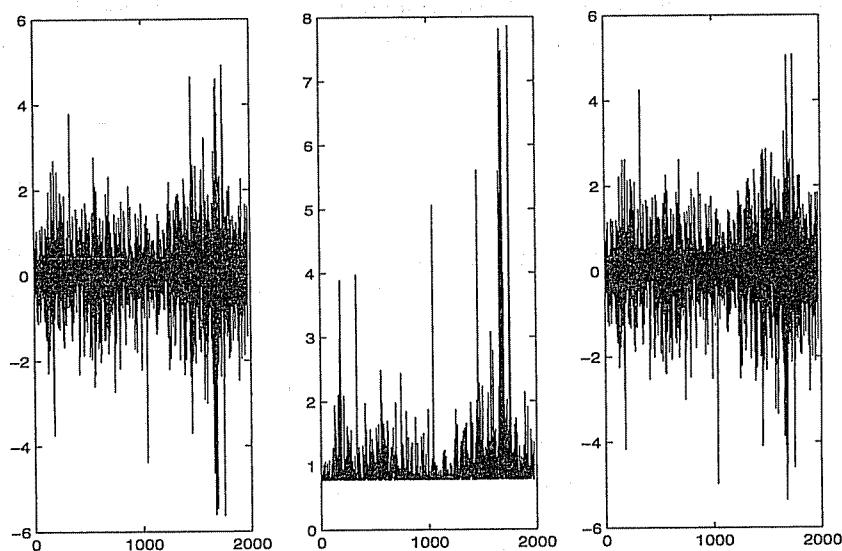
los cuadrados varios coeficientes de autocorrelación significativos. Llamando:

$$\hat{a}_t = z_t - .0007 - .1257z_{t-1} + .0481z_{t-2},$$

a los residuos del modelo ARIMA para la media de la serie, donde  $z_t$  es la

## 15. Modelos de heterocedasticidad condicional

Figura 15.6 Ruido blanco incorrelado, volatilidad y residuos estandarizados para la serie del Ibex



serie de rendimientos del Ibex, y definiendo los residuos estandarizados:

$$e_t = \frac{\hat{a}_t}{\sigma(\hat{a}_t)},$$

donde  $\sigma(\hat{a}_t) = .0123$ , vamos a estimar un modelo ARCH(1) para estos residuos,  $e_t$ , que tendrán por construcción varianza marginal igual a uno. Para ello elevamos estos residuos al cuadrado,  $e_t^2$ , y estimamos por mínimos cuadrados el modelo:

$$\hat{e}_t^2 = 0.776 + 0.22 e_{t-1}^2, \quad (13.87) \quad (10.18)$$

que muestra una relación claramente significativa entre los cuadrados de los rendimientos del Ibex.

La figura 15.6 presenta un resumen de los resultados. Hemos presentado primero la serie de ruido blanco,  $e_t$ , segundo las varianzas condicionales estimadas con esta variable que proporcionan una estimación del riesgo o volatilidad en cada instante,  $\hat{\sigma}_t^2 = 0.776 + 0.22e_{t-1}^2$ , y tercero, los valores de la serie estandarizados por sus desviaciones típicas condicionales, que son las estimaciones del ruido blanco normal  $\epsilon_t$ , calculadas como:

$$\hat{\epsilon}_t = \frac{e_t}{\sqrt{0.776 + 0.22e_{t-1}^2}}.$$

Se observa que la primera serie de este gráfico tiene rachas de mayor varianza, especialmente al final del periodo. La segunda serie muestra que la volatilidad ha sido más alta en el último periodo. En la serie estandarizada se han eliminado algunos picos de volatilidad, indicando que el modelo ajustado explica algunos componentes de los datos, pero queda bastante estructura sin explicar.

---

### 15.2.2. El modelo ARCH( $r$ )

El análisis anterior puede generalizarse permitiendo una dependencia de la varianza condicional con  $r$  retardos. Se obtiene así el proceso ARCH( $r$ ), para  $e_t = \sigma_t \epsilon_t$ , donde ahora:

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_r e_{t-r}^2. \quad (15.9)$$

En este proceso las posibilidades de rachas de alta volatilidad dependen de los  $r$  últimos valores. La varianza marginal será:

$$Var(e_t) = E(e_t^2) = E[E(e_t^2 | e_{t-1})] = \alpha_0 + \sum \alpha_i E(e_{t-i}^2),$$

que implica

$$Var(e_t) = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_r}$$

lo que exige la restricción  $\sum_{i=1}^r \alpha_i < 1$ .

Si introducimos, como en el caso del proceso ARCH(1), la variable  $v_t = e_t^2 - \sigma_t^2$ , podemos expresar la dependencia de los cuadrados de las observaciones como un proceso AR( $r$ ):

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_r e_{t-r}^2 + v_t, \quad (15.10)$$

donde la secuencia  $v_t$  tiene las mismas propiedades que en el modelo ARCH(1): variables incorreladas de medio cero, varianza constante e incorreladas con los regresores. Por otro lado, estas variables no son independientes entre sí ni de los regresores, ya que la positividad de  $e_t^2$  exige que

$$v_t > -(\alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_r e_{t-r}^2).$$

La ecuación (15.10) indica que existirá dependencia entre los cuadrados de las observaciones, y podríamos haber utilizado esta expresión en lugar de (15.9) para definir el proceso ARCH( $r$ ). En efecto, tomando esperanzas en (15.10) condicionadas a los valores observados,  $e_{t-1} = (e_{t-1}, \dots, e_1)$  se obtiene (15.9).

Puede comprobarse, de manera similar al caso del ARCH(1), que el

coeficiente de curtosis de un ARCH( $r$ ) es siempre mayor que 3.

---

### Ejemplo 15.2

Vamos a estimar un ARCH( $p$ ) para los residuos de la serie de rendimientos del Ibex (dibex99.dat). Las  $fas$  y  $fap$  de los cuadrados de los residuos se presentan en la figura 15.7 y muestran en los parciales muchos coeficientes significativos, lo que sugiere un AR de orden alto. Este mismo efecto se observa en los cuadrados de la serie original de rendimientos de la figura 15.1. Al estimar el modelo se obtienen muchos coeficientes significativos, por lo que hemos ido incrementando el orden del AR hasta encontrar un coeficiente no significativo.

La estimación conduce al modelo:

$$\hat{\sigma}_t^2 = 0.259 + .080 e_{t-1}^2 + 0.06 e_{t-10}^2 + \dots + .011 e_{t-10}^2,$$

donde todos los coeficientes son significativos. La figura 15.8 muestra la serie que estamos modelando, la volatilidad estimada y los residuos de este modelo. Ahora sí que se observa la diferencia entre la primera serie y la tercera, que tiene más aspecto de ruido blanco normal, aunque existen algunos pocos datos atípicos muy destacados.

El resultado frecuente de que es necesario ajustar ARCH de orden muy alto a las series financieras ha llevado a los modelos GARCH, que estudiamos a continuación.

---

## 15.3. Modelos GARCH

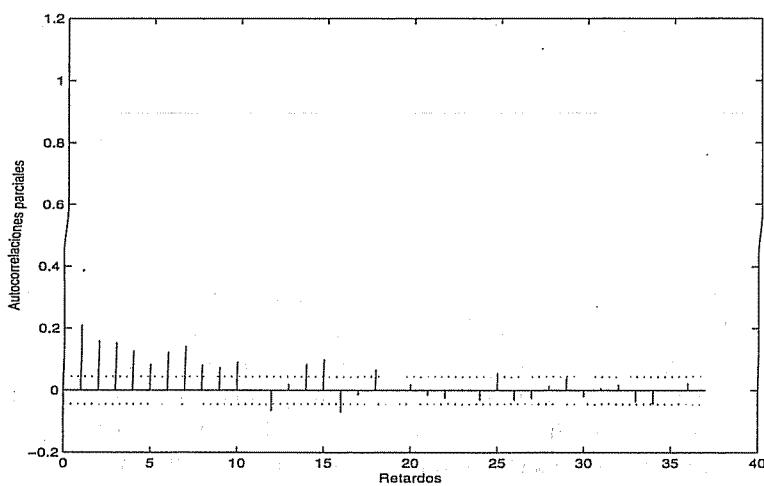
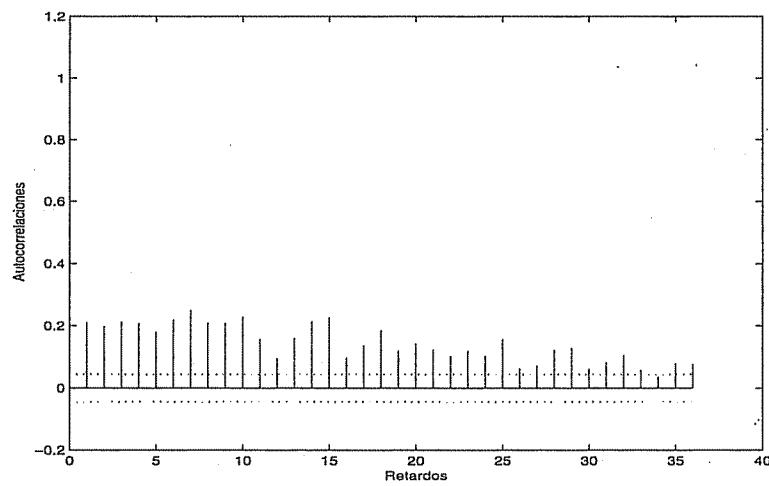
La identificación de modelos ARCH conduce generalmente a modelos de orden alto. Una idea natural es aproximar un proceso AR alto mediante una media móvil en las varianzas. Esto conduce al modelo GARCH o ARCH generalizado. Vamos a estudiar las propiedades de este modelo en el caso más simple. Supondremos, como en todo el capítulo, que la serie estudiada se ha generado mediante  $e_t = \sigma_t \epsilon_t$ , donde  $\epsilon_t$  y  $\sigma_t$  son dos procesos estacionarios independientes entre sí y con las mismas propiedades anteriores.

### 15.3.1. El modelo GARCH(1,1)

Suponemos que las varianzas condicionales siguen la ecuación:

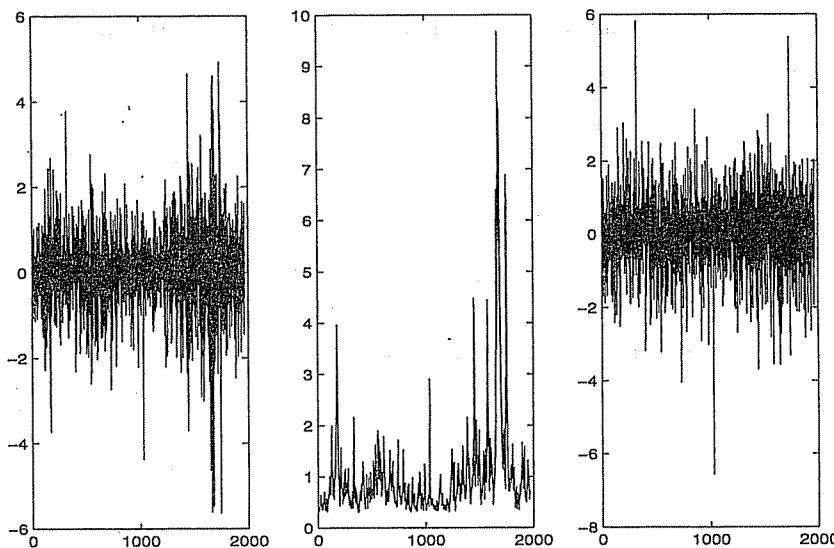
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (15.11)$$

**Figura 15.7 Fas y Fap de los cuadrados de los residuos del modelo para los rendimientos del Ibex**



con  $\alpha_1$  y  $\beta_1$  coeficientes positivos. En este modelo un valor alto para  $e_t^2$  da lugar a una mayor varianza en el periodo siguiente y el término  $\beta_1$  obliga a que la varianza cambie con cierta inercia, lo que produce rachas de mayor variabilidad.

**Figura 15.8 Serie a explicar, volatilidad estimada y residuos estandarizados de la estimación de un ARCH(10) para la serie de los residuos del Ibex**



Llamando como antes  $\sigma^2$  a la varianza marginal del proceso  $e_t$ :

$$\text{Var}(e_t) = E(e_t^2) = E(\sigma_t^2) = \sigma^2,$$

y tomando esperanzas en (15.11), tenemos que:

$$\sigma^2 = \alpha_0 + \alpha_1 E(e_{t-1}^2) + \beta_1 E(\sigma_{t-1}^2),$$

y al ser  $e_t$  estacionario,  $\text{Var}(e_t) = \text{Var}(e_{t-1})$ , que implica  $E(\sigma_t^2) = E(\sigma_{t-1}^2)$ , tenemos que:

$$\text{Var}(e_t) = \sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1},$$

y para que el proceso sea estacionario  $\alpha_1 + \beta_1 < 1$ . Es frecuente en las aplicaciones a series de rendimientos financieros que la suma de estos dos parámetros esté cerca de la unidad.

#### 15.3.1.1. Dependencia de los cuadrados

Vamos a escribir este modelo en la representación de los cuadrados de las variables introduciendo como hicimos en los procesos ARCH el proceso de ruido blanco:

$$v_t = e_t^2 - \sigma_t^2,$$

que, por construcción, es inmediato que es un proceso estacionario (por ser la diferencia de procesos estacionarios) de media cero. Se comprueba igual que en los modelos ARCH que las variables  $v_t$ , aunque incorreladas, no son independientes. Sustituyendo  $\sigma_t^2 = e_t^2 - v_t$  en la ecuación (15.11), podemos escribir

$$e_t^2 = \alpha_0 + (\alpha_1 + \beta_1)e_{t-1}^2 + v_t - \beta_1 v_{t-1}, \quad (15.12)$$

que muestra que los cuadrados de las observaciones siguen una estructura de dependencia similar a un ARMA(1,1).

Al coeficiente  $\lambda = (\alpha_1 + \beta_1)$  se le denomina *persistencia* y es frecuente, en las aplicaciones a series financieras, que se estime próximo a la unidad. Se demuestra que, llamando como en la sección anterior  $\rho_c(k)$  a los coeficientes de autocorrelación al cuadrado de la serie, se obtiene

$$\rho_c(1) = \frac{\alpha_1(1 - \alpha_1\beta_1 - \beta_1^2)}{(1 - 2\alpha_1\beta_1 - \beta_1^2)}$$

mientras que

$$\rho_c(k) = (\alpha_1 + \beta_1)^{k-1} \rho_c(1), \quad k > 1$$

que indica que el decrecimiento de estos coeficientes depende de la persistencia  $\lambda = (\alpha_1 + \beta_1)$ , y será lenta si este coeficiente es próximo a la unidad. Observemos que en el caso particular en que  $\beta = 0$  obtenemos la estructura del ARCH(1).

### 15.3.1.2. Curtosis

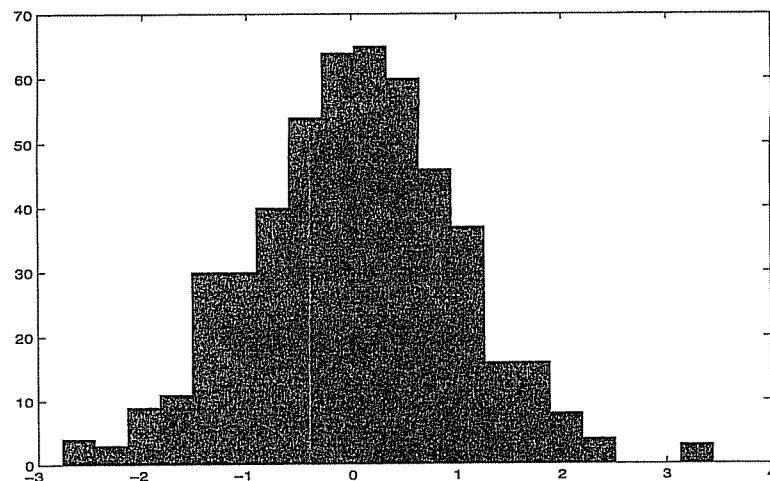
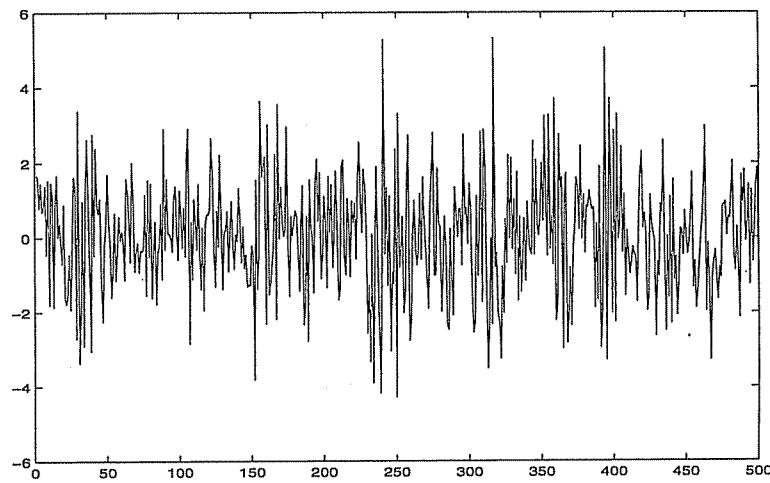
Puede demostrarse que, suponiendo que la distribución de  $\epsilon_t$  es normal, el coeficiente de curtosis viene dado por

$$K = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3,$$

que exige la restricción adicional  $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ . Por tanto, este modelo implica también colas pesadas. Observemos que esta ecuación es idéntica a la del ARCH(1) si hacemos  $\beta_1 = 0$ .

La figura 15.9 presenta el gráfico de una serie generada por este modelo y el histograma, y la 15.10 la acf de la variable y la acf de los cuadrados. El proceso es un ruido blanco, pero la acf de los cuadrados nos indica que no es de variables independientes.

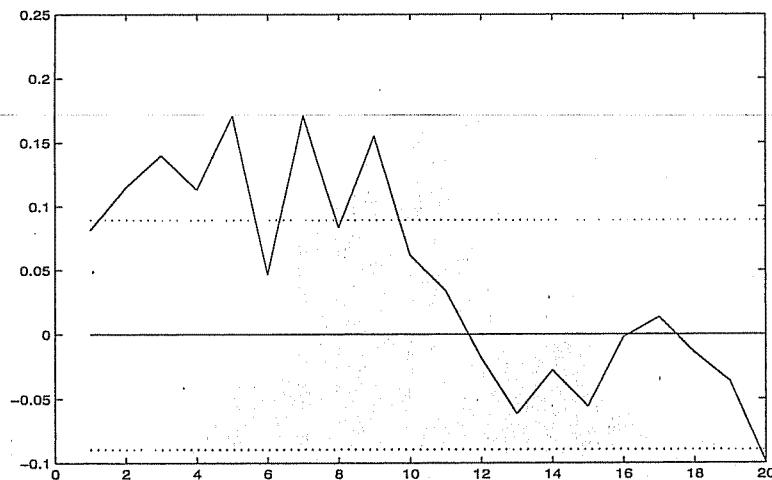
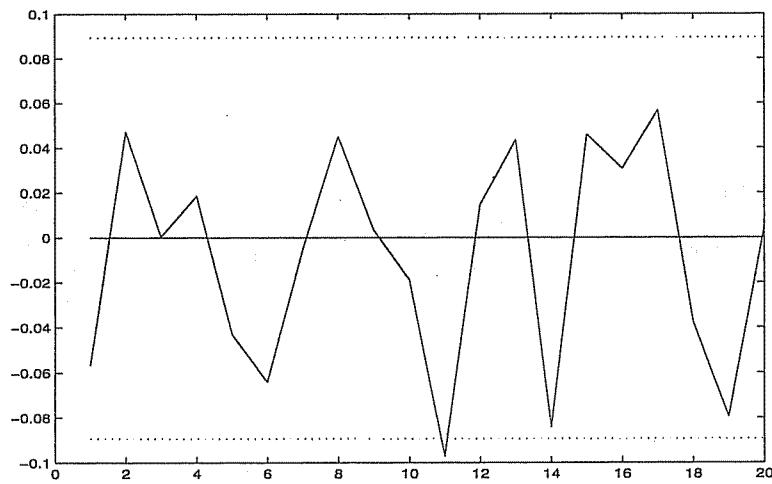
Figura 15.9 Gráfico del proceso GARCH(1,1) con alfa0=.2 y alfa1=.1 y beta=.8 e histograma de los datos



### 15.3.2. El modelo GARCH general

El modelo anterior puede generalizarse permitiendo simultáneamente una estructura AR de orden  $r$  en la dependencia de la varianza de los cuadrados de la serie y una estructura MA de orden  $s$  en la varianza condicional.

**Figura 15.10 Correlograma del proceso GARCH(1,1) con alfa0=.2 y alfa1=.1 y beta=.8 y correlograma de los cuadrados**



Se obtiene entonces el modelo GARCH( $r, s$ ), dado por

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i e_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (15.13)$$

## 15. Modelos de heterocedasticidad condicional

---

donde la secuencia  $\epsilon_t$  tiene las mismas propiedades que en los procesos ARCH. Los parámetros de esta ecuación deben verificar ciertas restricciones para que la varianza sea positiva y existan los momentos de orden superior. En particular, como la varianza debe ser positiva,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$ , y  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) < 1$ .

Definiendo como antes las variables incorreladas de media cero y varianza constante  $v_t = e_t^2 - \sigma_t^2$ , podemos escribir la dependencia de los cuadrados como:

$$e_t^2 = \alpha_0 + \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) e_{t-i}^2 + \eta_t - \sum_{j=1}^s \beta_j \eta_{t-j}, \quad (15.14)$$

ecuación que representa un proceso ARMA en los cuadrados de las observaciones. Esperamos, en consecuencia, que los cuadrados de las observaciones presenten estructura dinámica similar a la de un proceso ARMA. Por otro lado, tomando esperanzas en esta ecuación, se obtiene inmediatamente que:

$$E(e_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i)}$$

que implica restricciones sobre los parámetros del proceso para que sea estacionario.

Cuando el operador AR de la representación de los cuadrados tiene una raíz en el círculo unidad se obtienen procesos con varianza condicional que sigue un proceso integrado no estacionario, o procesos IGARCH. En particular el IGARCH(1,1) es, haciendo  $\alpha_1 + \beta_1 = 1$  en (15.12),

$$\nabla e_t^2 = \alpha_0 + v_t - \beta_1 v_{t-1},$$

con varianza condicional, poniendo  $\alpha_1 = 1 - \beta_1$  en (15.11), dada por:

$$\sigma_t^2 = \alpha_0 + (1 - \beta_1) e_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

La varianza condicional en este caso es una combinación lineal del cuadrado de la variable y de la varianza condicional anterior.

### 15.4. Construcción de modelos ARCH y GARCH

La construcción de estos modelos implica primero identificar la estructura; segundo, estimar los parámetros, y tercero, comprobar su adecuación a los datos. Vamos a revisar estos tres aspectos.

### 15.4.1. Identificación

La identificación de efectos ARCH se efectúa normalmente después de ajustar un modelo ARMA a la serie para eliminar la dependencia en la media. Si existen efectos ARCH, los residuos del modelo ARIMA estarán incorrelados pero no serán independientes y este efecto será visible en la función de autocorrelación de los residuos al cuadrado, que mostrarán correlación serial. Además, si calculamos los coeficientes de autocorrelación parcial de los residuos al cuadrado y el modelo para los residuos es ARCH puro, el número de términos distintos de cero nos indicará, aproximadamente, el orden del proceso.

Para detectar estructura en los cuadrados podemos acudir a los contrastes de McLeod y Li (1983) y Peña y Rodríguez (2005a) presentados en la sección anterior. Además de estos contrastes generales, que sirven para detectar una estructura general no lineal, pueden utilizarse contrastes específicos para detectarla. Véase por ejemplo Rodríguez y Ruiz (2005) para un buen contraste específico.

### 15.4.2. Estimación

Vamos a explicar la estimación de estos modelos en el caso de un ARCH (1). Las ideas se generalizan fácilmente para otros modelos más complejos. Los datos de la serie verifican la ecuación (15.8):

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + v_t, \quad (15.15)$$

y podríamos estimar esta ecuación por mínimos cuadrados. El ruido de esta ecuación,  $v_t$ , está incorrelado con el regresor, por lo que la estimación proporcionará estimadores consistentes (véase por ejemplo Green, 1993, pág. 419). Sin embargo, la varianza condicionada en esta ecuación no es constante y la regresión es heterocedástica, por lo que no serán eficientes. En efecto, utilizando, que según (15.6):

$$v_t^2 = \sigma_t^4 (\epsilon_t^2 - 1)^2$$

tenemos que:

$$\text{var}(v_t | e_{t-1}^2) = E(\sigma_t^4 | e_{t-1}^2) E(\epsilon_t^4 + 1 - 2\epsilon_t^2) = 2\sigma_t^4,$$

ya que

$$E(\epsilon_t^4 + 1 - 2\epsilon_t^2) = 3 + 1 - 2 = 2,$$

y  $E(\sigma_t^4 | e_{t-1}^2)$  es una constante igual a  $\sigma_t^4$  ya que  $\sigma_t^4$  sólo depende de  $e_{t-1}^2$ , que aparece condicionando. Al depender  $\sigma_t^4$  de  $e_{t-1}^2$ , que es el regresor de la ecuación, la varianza condicionada no es constante y tendríamos

que estimar la ecuación (15.15) por mínimos cuadrados generalizados. Para ello, necesitamos un estimador de la desviación típica,  $\sigma_t^2\sqrt{2}$ . Entonces, dada una estimación de  $\sigma_t^2$ , que llamaremos  $\widehat{\sigma}_t^2$ , al dividir todos los términos de esta ecuación por  $\widehat{\sigma}_t^2$  obtenemos una innovación de varianza condicional constante. Llamando  $y_t = e_t^2/\widehat{\sigma}_t^2$ ,  $z_t = 1/\widehat{\sigma}_t^2$ ,  $x_t = e_{t-1}^2/\widehat{\sigma}_t^2$  y  $u_t = v_t/\widehat{\sigma}_t^2$  podemos escribir (15.15) como:

$$y_t = \alpha_0 z_t + \alpha_1 x_t + u_t \quad (15.16)$$

y en esta ecuación el residuo  $u_t$  será aproximadamente homocedástico.

Vamos a comprobar que éste es precisamente el resultado de aplicar máxima verosimilitud a un modelo ARCH(1). La función de verosimilitud es:

$$f(e_1, \dots, e_T | \alpha_0, \alpha_1) = f(e_1)f(e_2|e_1)\cdots f(e_T|e_{T-1}).$$

dónde las funciones de densidad  $f(e_t|e_{t-1})$  son normales. En efecto, como  $e_t = \sigma_t \epsilon_t$ , si condicionamos a  $e_{t-1}$ , el valor  $\sigma_t$  es una constante y la única variable es  $\epsilon_t$  que tiene distribución normal. La media condicionada de la distribución  $f(e_t|e_{t-1})$  es cero, y la varianza  $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$ . La función de verosimilitud condicionada a  $e_1$  en logaritmos será:

$$L(e_2, \dots, e_T | \alpha_0, \alpha_1, e_1) = -\frac{1}{2} \sum_{t=2}^T \ln(\alpha_0 + \alpha_1 e_{t-1}^2) - \frac{1}{2} \sum_{t=2}^T \frac{e_t^2}{\alpha_0 + \alpha_1 e_{t-1}^2},$$

y derivando respecto a los parámetros, llamando  $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 e_{t-1}^2$  e igualando a cero se obtienen las ecuaciones:

$$\sum \frac{e_t^2}{\widehat{\sigma}_t^4} = \sum \frac{1}{\widehat{\sigma}_t^2} \quad (15.17)$$

y

$$\sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^2} = \sum \frac{e_{t-1}^2 e_t^2}{\widehat{\sigma}_t^4}. \quad (15.18)$$

Para resolverlas, multipliquemos y dividamos el primer miembro por  $\widehat{\alpha}_0 + \widehat{\alpha}_1 e_{t-1}^2$ , y escribamos (15.17) como:

$$\widehat{\alpha}_0 \sum \frac{1}{\widehat{\sigma}_t^4} + \widehat{\alpha}_1 \sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^4} = \sum \frac{e_t^2}{\widehat{\sigma}_t^4}, \quad (15.19)$$

y (15.18) como:

$$\widehat{\alpha}_0 \sum \frac{e_{t-1}^2}{\widehat{\sigma}_t^4} + \widehat{\alpha}_1 \sum \frac{e_{t-1}^4}{\widehat{\sigma}_t^4} = \sum \frac{e_{t-1}^2 e_t^2}{\widehat{\sigma}_t^4},$$

que son las ecuaciones de mínimos cuadrados para obtener los parámetros en la regresión (15.16).

Este resultado sugiere el siguiente procedimiento iterativo para maximizar la función de verosimilitud. Estimar por mínimos cuadrados la regresión (15.15) y obtener unos valores iniciales de los parámetros. Calcular con estos valores  $\hat{\sigma}_t^2$  y resolver las ecuaciones (15.18) y (15.19) para obtener nuevos estimadores de  $\hat{\alpha}_0$  y  $\hat{\alpha}_1$ . Con ellos, calcular de nuevo  $\hat{\sigma}_t^2$  y volver a las ecuaciones (15.18) y (15.19) para obtener un nuevo estimador. Iterar hasta convergencia.

En general, la función de verosimilitud de un proceso estacionario  $w_t$  cuyas variables siguen un proceso GARCH( $r, s$ ), donde suponemos  $r \geq s$  es, llamando  $W_t = (w_1, \dots, w_t)$ ,

$$\begin{aligned} & f(w_1, \dots, w_n | \alpha) = \\ &= f(w_n | W_{n-1}) f(w_{n-1} | W_{n-2}) \cdots f(w_{r+1} | W_r) f(w_1, \dots, w_r | \alpha) = \\ &= \prod_{t=r+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{(w_t - E(w_t | W_{t-1}))^2}{2\sigma_t^2}\right] \times f(w_1, \dots, w_r | \alpha) \end{aligned}$$

ya que la varianza condicional de las variables es  $\sigma_t^2$ . Condicionando a las primeras  $r$  observaciones, que tienen una distribución más complicada, la función soporte condicionada es:

$$\begin{aligned} L(w_{r+1}, \dots, w_n | \alpha, W_r) &= -\frac{1}{2} \sum_{t=r+1}^n \ln \sigma_t^2 - \\ &\quad -\frac{1}{2} \sum_{t=r+1}^n \frac{(w_t - E(w_t | W_{t-1}))^2}{\sigma_t^2} \end{aligned} \tag{15.20}$$

que puede maximizarse con un algoritmo de optimización no lineal para obtener los estimadores de los parámetros que aparecen en la media condicional y en la varianza condicional.

La estimación puede realizarse en dos etapas ya que la correlación entre los parámetros ARMA y la de los GARCH suele ser pequeña. En la primera etapa se estiman los parámetros de la media condicional, es decir, el modelo ARMA, y se construyen las innovaciones  $e_t = w_t - E(w_t | W_{t-1})$ . En la segunda se estiman los parámetros de la varianza condicional maximizando la verosimilitud de los residuos. Alternativamente, es posible estimar las ecuaciones de la media condicional y la varianza condicional conjuntamente, con lo que se obtiene una estimación más precisa.

### 15.4.3. Diagnosis

Si llamamos  $e_t$  a los residuos del modelo ARIMA y  $\hat{\sigma}_t$  a las varianzas condicionadas estimadas, los residuos estandarizados,  $e_t/\hat{\sigma}_t$ , deben seguir un proceso de ruido blanco normal y podemos aplicarles los contrastes habituales estudiados para procesos ARIMA. Además, sus cuadrados no deben mostrar dependencia, lo que podemos comprobar con los contrastes sobre las autocorrelaciones de los cuadrados descritos anteriormente.

Un punto importante a tener en cuenta en la diagnosis de estos modelos es la posible confusión entre valores atípicos y heterocedasticidad condicional. Los valores atípicos pueden interpretarse como un aumento de la varianza en ese instante y, especialmente si aparecen en rachas, pueden confundirse con efectos de heterocedasticidad condicional. Por otro lado, una serie que sigue un modelo ARCH puede mostrar muchos atípicos si se analiza como si tuviese varianza constante y siguiese un modelo ARMA. Por lo tanto, es importante diferenciar ambos fenómenos.

Un procedimiento simple que funciona bien en la práctica es limpiar la serie inicialmente de las observaciones que presentan residuos tan grandes que no pueden ser debidas a heterocedasticidad condicional y que son muy probablemente valores atípicos. Una regla simple y efectiva es tener en cuenta que es muy poco probable que la heterocedasticidad condicional pueda generar observaciones con residuos mayores de siete desviaciones típicas, y considerar estos datos como atípicos. A continuación estimaremos los efectos ARCH o GARCH con la serie corregida y construiremos la serie de innovaciones estandarizadas,  $e_t/\hat{\sigma}_t$ . Sobre esta serie de innovaciones podemos, de nuevo, buscar valores atípicos y, si se detectan, corregirlos y volver a modelar las varianzas condicionales. El proceso se itera hasta convergencia.

### Ejemplo 15.3

Vamos a construir un modelo GARCH para la serie de datos diarios del indicador de bolsa Ibex, que hemos visto podría explicarse con un ARCH(10). Vamos a estimar conjuntamente la estructura AR(2) que detectamos en el ejemplo 15.1 y la estructura GARCH(1,1) para los residuos de este modelo. Con el programa SCA la estimación conduce a las dos ecuaciones:

$$\begin{aligned} z_t &= .0007 + .1257z_{t-1} - .0481z_{t-2} + \hat{\sigma}_t \\ \hat{\sigma}_t^2 &= .025 + .087(\hat{\sigma}_{t-1}/.0123)^2 + .885\hat{\sigma}_{t-1}^2 \end{aligned}$$

y la segunda ecuación también puede escribirse, multiplicando todo por  $(.0123)^2$  y llamando de nuevo  $\hat{\sigma}_t^2$  a las nuevas varianzas condicionales (que ahora se re-

**Tabla 15.1 Modelos estimados conjuntamente para la media y la varianza condicional del Ibex**

coef	c	$\phi_1$	$\phi_2$	$\alpha_0$	$\alpha_1$	$\beta_1$
valor	.0009	0.125	-0.016	.0000004	0.09	0.88
t	3.8	5.19	-0.68	4.87	6.97	53.65

fieren a los residuos sin estandarizar)

$$\hat{\sigma}_t^2 = .0000004 + .087\hat{a}_{t-1}^2 + .885\hat{\sigma}_{t-1}^2.$$

Para comparar ambas ecuaciones, tomando esperanzas en la primera:

$$E(\hat{\sigma}_t^2) = \sigma^2 = .025 + .087 + .885\sigma^2$$

es decir

$$\sigma^2 = \frac{.112}{.115} \approx 1.$$

Como los residuos están estandarizados, la varianza marginal es uno, salvo por errores de redondeo. En la segunda, la varianza será por construcción  $(.0123)^2$ , ya que tomando esperanzas:

$$\sigma^2 = \frac{.112}{.115} (.0123)^2 \approx (.0123)^2.$$

Alternativamente, podríamos haber estimado conjuntamente todos los parámetros con el programa EViews. Los resultados de la estimación conjunta de las dos ecuaciones se presenta en la tabla 15.1.

Se observa que los resultados de ambos métodos son similares, aunque en este segundo ahora no sale significativo el segundo parámetro AR.

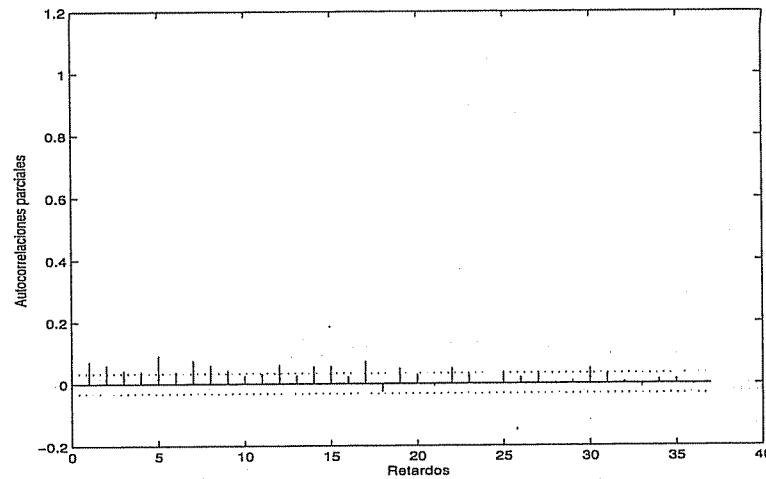
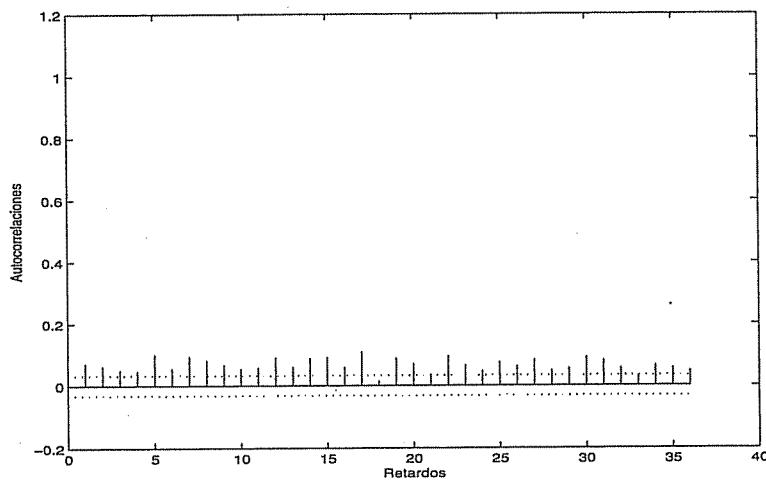
#### Ejemplo 15.4

Vamos a analizar los residuos del modelo de la temperatura máxima diaria de Melbourne, Australia, cuyo modelo estudiamos en el capítulo 11. Los datos de esta serie están en el fichero melbtemp.dat y en el ejemplo 11.3 vimos que puede modelarse como:

$$(1 - 1.46B + .615B^2 - .094B^3 - .045B^4)\nabla_{12}\log TMD_t = \\ = (1 - .887B)(1 - .999B^{12})a_t$$

con una  $\hat{\sigma} = .1738$ . Alternativamente podemos estimar un modelo con estacionalidad determinista. Todos los parámetros son significativos. Los residuos parecen ruido blanco y no muestran ninguna estructura de correlación. Por ejemplo  $Q(24) = 19.5$ . Sin embargo, los residuos al cuadrado sí presentan

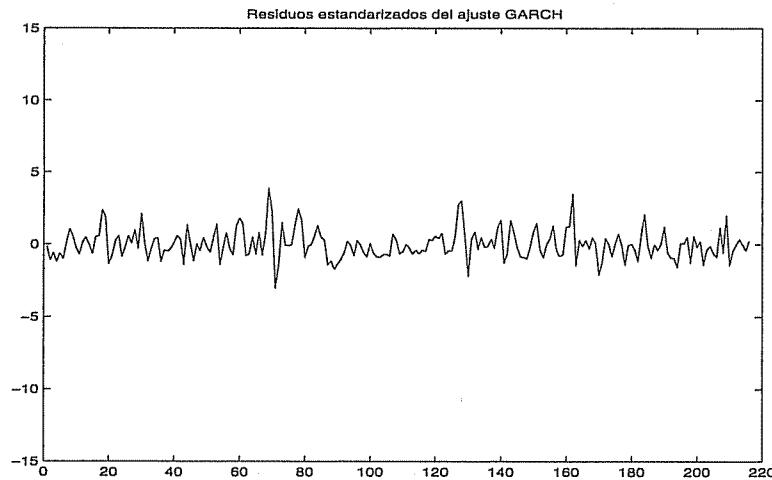
**Figura 15.11 Fas y fap de los cuadrados de los residuos del modelo para los logaritmos de las temperaturas diarias en Melbourne**



estructura, como se indica en la figura 15.11.

El gran número de coeficientes positivos y que decrecen muy lentamente podría sugerir un modelo de larga memoria para la serie de residuos o un modelo GARCH. La estimación realizada por SCA para los cuadrados de los

**Figura 15.12 Residuos estandarizados de la serie de manchas solares después de un ajuste GARCH**



residuos es:

$$(1 - .987)a_t^2 = .0004 + (1 - .948B)v_t,$$

que implica el modelo de GARCH:

$$\hat{\sigma}_t^2 = .0004 + .039\hat{a}_{t-1}^2 + .948\hat{\sigma}_{t-1}^2,$$

que no muestra deficiencias apreciables.

### Ejemplo 15.5

Vamos a estimar un modelo GARCH para la serie de manchas solares que estudiamos en el capítulo 14 y encontramos correlación en los cuadrados. El modelo estimado es:

$$\begin{aligned} z_t &= 45.50 + 0.85z_{t-1} - 0.27z_{t-3} + 0.18z_{t-9} + a_t + 0.25a_{t-1} \\ &\quad (9.28) \quad (22.9) \quad (6.73) \quad (5.18) \quad (4.33) \\ \hat{\sigma}_t^2 &= 394.47 + 0.042\hat{a}_{t-1}^2 - 0.995\hat{\sigma}_{t-1}^2 \\ &\quad (4.86) \quad (7.01) \quad (150.1) \end{aligned}$$

que parece proporcionar un ajuste razonable a estos datos. Los residuos de este modelo se indican en la figura 15.12.

### 15.5. Modelos de volatilidad estocástica

Un procedimiento alternativo para modelar la heterocedasticidad es suponer que la varianza condicional depende de factores no observados. Se obtienen así los modelos de volatilidad estocástica, o modelos SV (Stochastic volatility) que han sido estudiados por Harvey, Ruiz y Shephard (1994) y Jacquier, Polson y Rossi (1994). La ecuación estructural de estos modelos es idéntica a la de los modelos GARCH,  $e_t = \sigma_t \epsilon_t$ , donde los errores  $\epsilon_t$  son iid  $N(0, 1)$ . Sin embargo, a diferencia de los GARCH donde la varianza depende de factores observables, en este modelo se supone que los logaritmos de las varianzas condicionadas siguen un proceso  $AR(p)$ ,

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^p \alpha_i \ln(\sigma_{t-i}^2) + \xi_t, \quad (15.21)$$

donde el término de error en esta ecuación,  $\xi_t$ , es iid  $N(0, \sigma_\xi^2)$  e independiente de  $\epsilon_t$ . El coeficiente  $\alpha_0$  es una constante y los parámetros del proceso verifican las condiciones estándar de estacionariedad. Es decir, los ceros del polinomio  $1 - \sum_{i=1}^p \alpha_i B^i$  deben estar fuera del círculo unidad. Vamos a estudiar este modelo en el caso más simple.

#### 15.5.1. El modelo SV(1)

Suponiendo que el logaritmo de la volatilidad sigue un proceso AR(1), tenemos que:

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \sigma_{t-1}^2 + \xi_t,$$

donde suponemos  $|\alpha_1| < 1$ . Llamando  $h_t = \ln \sigma_t^2$  este proceso es AR(1) con ruidos normales, por lo que la distribución de la variable  $h_t$  será también normal. Los parámetros de la distribución marginal son los de un AR(1), es decir, la media es  $\mu_y = \alpha_0/(1 - \alpha_1)$  y la varianza  $\sigma_y^2 = \sigma_\xi^2/(1 - \alpha_1)$ . Entonces, la variable  $\sigma_t^2$  será lognormal y puede demostrarse que:

$$E(e_t^2) = e^{\mu_y} e^{(\sigma_y^2/2)},$$

y el coeficiente de curtosis es

$$K = 3e^{\sigma_y^2},$$

que será siempre mayor que 3. Este modelo tiene pues también la capacidad de generar distribuciones con colas pesadas. Finalmente, puede comprobarse que la estructura de correlación de los cuadrados de este modelo es

$$\rho_c(k) = \frac{\exp((\sigma_y^2 \alpha_1^k) - 1)}{[3 \exp(\sigma_y^2) - 1]}, \quad k \geq 1$$

si la varianza  $\sigma_y^2$  es pequeña, puede demostrarse que esta expresión implica un decaimiento similar a un AR(1) con parámetro  $\alpha_1$ .

### 15.6. Otros enfoques y lecturas complementarias

Es importante recordar que la presencia de autocorrelación en los cuadrados puede ser debida a muchas causas, y no sólo a la heterocedasticidad condicional. Entre ellas hay que resaltar modelos no lineales en la media y cambios por tramos en la varianza del proceso, incluyendo rachas de atípicos.

Ya vimos en el capítulo 14 que los procesos bilineales pueden tener propiedades parecidas a los GARCH. Esto también ocurre con modelos con parámetros cambiantes. Por ejemplo, consideremos el modelo simple:

$$z_t = \phi_t z_{t-1} + a_t$$

donde el parámetro AR varía en el tiempo de acuerdo con la estructura:

$$\phi_t = \phi + u_t$$

siendo  $u_t$  un proceso de ruido blanco con varianza  $\sigma_u^2$ . Con estas dos ecuaciones el modelo completo es :

$$z_t = \phi z_{t-1} + u_t z_{t-1} + a_t.$$

Los momentos de la distribución condicionada son:

$$E(z_t | z_{t-1}) = \phi z_{t-1},$$

y

$$Var(z_t | z_{t-1}) = \sigma_t^2 = \sigma_u^2 z_{t-1}^2 + \sigma_u^2$$

que es la misma estructura de un modelo ARCH(1).

La heterocedasticidad condicional aparece también si se produce un cambio de varianza mediante un escalón. Por ejemplo, es fácil demostrar (véase Peña y Rodríguez, 2005b) que si consideramos el modelo:

$$\begin{aligned} e_t &= \varepsilon_t \sigma_t \\ \sigma_t^2 &= a + b S_t^{(h)} \end{aligned}$$

donde  $S_t^{(h)}$  es una variable escalón, obtendremos correlación en los cuadra-

dos. Además, la presencia de datos atípicos puede ocultar o magnificar la heterocedasticidad, según que los atípicos aparezcan aislados, en cuyo caso la ocultan, o en rachas, en cuyo caso la magnifican. Carnero, Peña y Ruiz (2004) han estudiado este problema.

Existen además otros muchos modelos propuestos para explicar la heterocedasticidad condicional. Entre ellos están los modelos GARCH exponenciales de Nelson (1991) que conducen a efectos asimétricos de las innovaciones, los modelos ARMA de heterocedasticidad condicional introducidos por Tsay (1987), y los modelos de coeficientes aleatorios de Nicholls and Quinn (1982). Véanse Gouriéroux (1997), Peña, Tiao y Tsay (2001) y Tsay (2002) para descripciones y limitaciones de otros modelos.

El lector interesado en profundizar en este tema puede acudir a Gouriéroux (1997), Franses y van Dijk (2000), Tsay (2002) y Chan (2002).

### Ejemplo 15.6

Vamos a estudiar si existen efectos de heterocedasticidad en la serie del PIB americano, lo que podría explicar los resultados de los contrastes de linealidad del capítulo anterior. El modelo estimado para los residuos del PIB americano en el periodo 1/1947-2/2003, después de limpiar esta serie por los residuos AO y TC detectados en el capítulo anterior (pero no por los cambios de nivel) es:

$$\begin{aligned} e_t &= \varepsilon_t \sigma_t \\ \sigma_t^2 &= 4.8 \times 10^{-7} + 0.9292 \sigma_{t-1}^2 + 0.0601 e_{t-1}^2 \end{aligned}$$

y la figura 15.13, tomada de Peña y Rodríguez (2005b) muestra los residuos al cuadrado y la estimación de la varianza condicional,  $\sigma_t^2$ . Una medida del ajuste es:

$$\sum \frac{(e_i^2 - \sigma_i^2)}{T} = 1.0961 \times 10^{-8}.$$

Hemos comparado este modelo con el estimado por Peña y Rodríguez (2005b) permitiendo un cambio de nivel en la varianza:

$$\begin{aligned} e_t &= \varepsilon_t \sigma_t \\ \sigma_t^2 &= 1.0 + (.26 - 1.0) S_t^{(145)}, \end{aligned}$$

donde  $S_t^{(145)}$  es una función escalón. Este modelo implica que para  $t < 145$  la varianza residual es 1.001, mientras que para  $t \geq 146$  esta varianza es .258. La figura 15.14 muestra sus residuos y la medida de ajuste:

$$\sum \frac{(e_i^2 - \sigma_i^2)}{T} = 1.0334 \times 10^{-8}.$$

Figura 15.13 Residuos al cuadrado y varianza condicional estimada para el modelo Garch de la serie del PIB de Estados Unidos: Periodo 1/1947-2/2003

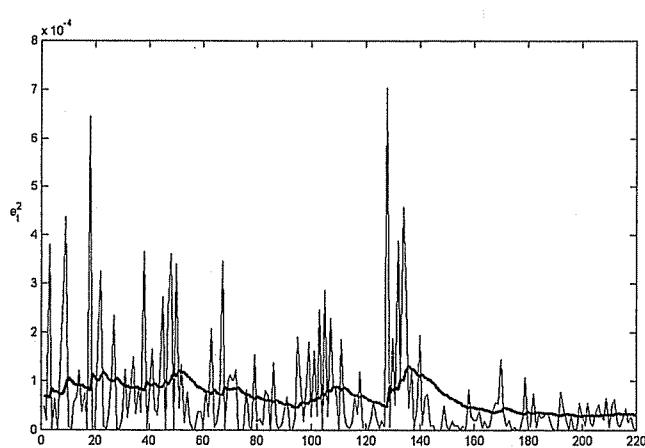
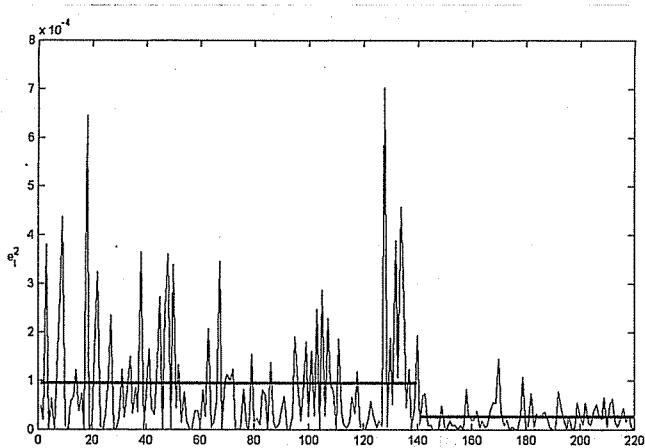


Figura 15.14 Residuos al cuadrado y estimación de la varianza para el modelo de cambio de nivel en la varianza en el PIB de Estados Unidos



Se observa que aunque ambos modelos explican bien los datos, el segundo parece dar un mejor ajuste que el primero, siendo además de más fácil explicación.

### Ejercicios 15

- 15.1 Calcular la media, varianza y coeficiente de curtosis del proceso ARCH(1) con  $\sigma_t^2 = .2 + .7e_{t-1}^2$ , que se ha representado en las figuras 15.3 y 15.4 y compararlas con las características de la realización obtenida.
- 15.2 Demostrar que el ARCH(2) implica una estructura AR(2) para el correlograma de los valores al cuadrado.
- 15.3 Demostrar que en el ARCH(1) se verifica que,

$$E(\sigma_t^4) = \alpha_0^2(1 + \alpha_1)/[(1 - 3\alpha_1^2)(1 - \alpha_1)].$$

- 15.4 Justificar las propiedades del proceso  $e_t/\sqrt{\alpha_0 + \alpha_1 e_{t-1}^2}$  y explicar su utilidad para la diagnosis del modelo ARCH.
- 15.5 Estimar modelos GARCH para los datos de rendimientos diarios del fichero bolsaEUdiario.dat (ver el apéndice 15.2 donde se ilustra cómo hacerlo con el programa EViews).
- 

### Apéndice 15.1: Propiedades de ARCH(1)

Vamos a deducir la autocorrelación entre los cuadrados de las variables para este proceso. Llamando:

$$\gamma_c(k) = E(e_t^2 e_{t-k}^2) - E(e_t^2)E(e_{t-k}^2),$$

a la función de autocovarianzas de los cuadrados de las variables del proceso, para  $k = 1$  sustituyendo  $e_t^2 = \sigma_t^2 e_t^2$ , y utilizando (15.3) y (15.5), se obtiene:

$$\gamma_c(1) = E((\alpha_0 + \alpha_1 e_{t-1}^2)\epsilon_t^2 e_{t-1}^2) - \alpha_0^2/(1 - \alpha_1)^2,$$

y como las variables  $e_t^2$  y  $e_{t-1}^2$  son independientes y  $E(\epsilon_t^2) = 1$ , resulta que:

$$\gamma_c(1) = \alpha_0^2/(1 - \alpha_1) + \alpha_1 E(e_{t-1}^4) - \alpha_0^2/(1 - \alpha_1)^2,$$

y dividiendo por  $\gamma_c(0) = E(e_{t-1}^4) - \alpha_0^2/(1 - \alpha_1)^2$  tenemos que:

$$\rho_c(1) = \frac{\gamma_c(1)}{\gamma_c(0)} = \alpha_1.$$

Esta ecuación indica que el parámetro  $\alpha_1$  es el coeficiente de correlación entre los cuadrados de las variables. Para otros retardos:

$$\gamma_c(k) = E((\alpha_0 + \alpha_1 e_{t-1}^2) e_t^2 e_{t-k}^2) - \alpha_0^2 / (1 - \alpha_1)^2,$$

y operando como antes:

$$\gamma_c(k) = \alpha_1 \gamma_c(k-1),$$

con lo que

$$\rho_c(k) = \alpha_1 \rho_c(k-1).$$

La conclusión de este ejercicio es que los cuadrados de las variables están ligados por una ecuación autorregresiva de primer orden.

Vamos a obtener el coeficiente de curtosis. Su expresión es:

$$K = \frac{E(e_t^4)}{Var(e_t)^2}, \quad (15.22)$$

como la variable  $e_t$  es normal, la variable  $e_t | e_{t-1}$  será también normal y tendrá curtosis igual a tres. Entonces,

$$E(e_t^4) = E[E(e_t^4 | e_{t-1})] = E(3\sigma_t^4) \quad (15.23)$$

y elevando al cuadrado y tomando esperanzas en (15.3):

$$E(\sigma_t^4) = \alpha_0^2 + 2\alpha_0\alpha_1 E(e_{t-1}^2) + \alpha_1^2 E(e_{t-1}^4)$$

y sustituyendo en (15.23), como  $E(e_t^4) = E(e_{t-1}^4)$  tenemos que:

$$E(e_t^4) = \frac{3\alpha_0^2 + 6\alpha_0\alpha_1 E(e_{t-1}^2)}{(1 - 3\alpha_1^2)}.$$

Observemos que esta ecuación impone una restricción adicional sobre  $\alpha_1$  ya que como el cociente debe ser positivo y el numerador siempre lo es, el denominador debe ser positivo, lo que implica  $0 \leq \alpha_1 < 1/\sqrt{3}$ . Si sustituimos esta expresión en (15.22) y utilizando la expresión de  $E(e_{t-1}^2)$  en (15.5), tenemos finalmente que

$$K = \frac{3\alpha_0^2(1 - \alpha_1)^2 + 6\alpha_0^2\alpha_1(1 - \alpha_1)}{\alpha_0^2(1 - 3\alpha_1^2)} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}.$$

## Apéndice 15.2: Programas de estimación

### EViews

Ilustraremos la estimación de estos modelos con EViews. Vamos a utilizar como ejemplo las series de los índices de mercado de cuatro países, Germany DAX (Ibis), Switzerland SMI, France CAC, y UK FTSE, que se encuentran en

## 15. Modelos de heterocedasticidad condicional

---

el fichero bolsaEUdiario.dat. Para abrir un workfile para estas cuatro series, se especifica:

*New → Workfile → Daily (5 days week) → 1:1:1991 a 2:16:1998.*

Se importan los datos y se dan a las series los nombres dax, smi, cac y ftse. Las cuatro series tienen un comportamiento muy parecido. Definimos el rendimiento de la serie como  $r_t = 100 * (\log(x_t) - \log(x_{t-1}))$ . Se generan los cuatro rendimientos de las series con las instrucciones en Genr:  $rdax = 100 * d(\log(dax), 1)$ ;  $rsmi = 100 * d(\log(smi), 1)$ ;  $rcac = 100 * d(\log(cac), 1)$ ;  $rftse = 100 * d(\log(ftse), 1)$ .

El gráfico de los rendimientos de estas cuatro series muestra cambios de varianza repentinos y el correlograma no muestra evidencia de correlación en la serie. En el histograma de los datos: *View→Descriptive Statistics→Histogram and Stats*, el contraste de Jarque-Bera no acepta la normalidad y el coeficiente de curtosis es muy elevado. Generamos el cuadrado de la serie rdax al que llamamos rdax2: ( $rdax2=rdax^2y$ ) y en su correlograma observamos correlación alta.

Vamos a estimar un modelo para la serie rdax. Como hemos visto que la serie no tiene correlación, vamos a estimar un modelo co constante: *Quick → Estimate Equation → rdax c*. La constante es significativa ( $\hat{c} = 0.065204$  con pvalor=0.0064). Si vemos el correlograma de los residuos del modelo al cuadrado, comprobamos que es el mismo que antes de estimar la constante. Vemos cómo en la *fap* y en la *fas* aparecen como significativas los dos primeros retardos.

Vamos ahora a estimar un ARCH(2):

*Procs → Especify/Estimate → Method:ARCH → ARCH=2 → GARCH=0,*

resulta en:

Parámetro	Estim.	std. error	t-stat	P-valor
C	0.0677	0.024	2.814	0.0049
Parámetro	Estim.	std. error	t-stat	P-valor
C	0.866	0.014	58.18	0.0000
ARCH(1)	0.0867	0.014	6.128	0.0000
ARCH(2)	0.090	0.022	4.022	0.0001

El gráfico de la varianza condicional es: *View→Conditional SD graph* y pasamos a ver la bondad del ajuste *View → Residual Tests → Correlogram Q-statistics*, que proporciona el correlograma de los residuos estandarizados. En el correlograma de los residuos estandarizados al cuadrado veremos como

## Análisis de series temporales

---

no queda correlación:

*View → Residual Tests → Correlogram Squared residuals.*

También vemos el histograma de los residuos estandarizados: *View→Residual Tests→Histogram, normality tests*, y su gráfico: *View→Actual, fitted, residual→Standardized residual graph*.

Se puede comprobar la presencia de un atípico al principio de la serie. Es el dato correspondiente al 2:19:1991. Creamos una variable impulso en este dato para incluirlo en la estimación: *Genr : at = 0; Genr : at = 1, sample: 2 : 19 : 1991 2 : 19 : 1991*. Lo incluimos en la estimación:

*Procs → Specify/Estimate → rdax c at*

*Method : ARCH → ARCH = 2 → GARCH = 0*

y obtenemos los siguientes resultados:

Parámetro	Estim.	std. error	t-stat	P-valor
C	0.074	0.02	3.32	0.0009
AT	-7.65	0.28	-26.56	0.0000
Parámetro	Estim.	std. error	t-stat	P-valor
C	0.766	0.025	30.19	0
ARCH(1)	0.11	0.017	6.89	0
ARCH(2)	0.12	0.025	5.02	0

El correograma de los residuos y de los residuos al cuadrado se obtiene con: *View → Residual Tests → Correlogram Q-statistics*, y *View → Residual Tests → Correlogram Squared residuals*. Comprobamos cómo los residuos al cuadrado tienen correlación. Si estimamos un GARCH(1,1) a estos datos junto con el atípico y la constante:

*Procs → Specify/Estimate → rdax c at*

*Method : ARCH → ARCH = 1 → GARCH = 1*

Aparece la siguiente información:

Parámetro	Estim.	std. error	t-stat	P-valor
C	0.076	0.020	3.69	0.0002
AT	-7.71	0.18	-41.97	0
Parámetro	Estim.	std. error	t-stat	P-valor
C	0.021	0.004	5.23	0.0000
ARCH(1)	0.073	0.009	8.120	0.0000
GARCH(1)	0.906	0.010	85.54	0.0000

Todo parece correcto. La varianza condicional se obtiene con: *View* → *Conditional SD graph*, y la bondad del ajuste: *View* → *Residual Tests* → *Correlogram Q-statistics*. En el correlograma de los residuos estandarizados al cuadrado veremos cómo no queda correlación: *View* → *Residual Tests* → *Correlogram Squared residuals*. Su histograma es: *View* → *Residual Tests* → *Histogram, normality tests* y su gráfico: *View* → *Actual, fitted, residual* → *Standardized residual graph*.

El estadístico de Jarque-Bera todavía rechaza la normalidad de los residuos. Podemos concluir que los residuos no son  $N(0,1)$  y que siguen una distribución como la  $t$  de Student, o una normal mezclada al estar contaminados por datos atípicos.

### Matlab

El siguiente programa permite estimar un modelo ARCH( $p$ ) sobre una serie  $x$  y calcular las volatilidades y residuos estandarizados:

%Este programa estima an arch(p) sobre una serie de ruido blanco  $y_0$  y devuelve los parametros en  $b$ , los estadisticos  $t$ , las volatilidades,  $y_h$ , y los residuos estandarizados a varianza uno condicional e incondicional.

```
function [b,t,yh,ep]=archpestim(y0,p)
x=y0.^2;
n=prod(size(x));
z=zeros(n-p,p+1);
for i=1:p+1
z(:,i)=x(p+2-i:n+1-i);
end
y=z(:,1);
x=z;
x(:,1)=ones(size(z(:,1)));
b=x\y;
e=y-x*b;
xi=inv(x'*x);
sr2=(e'*e)/(n-p-1);
vb=xi*sr2;
seb=sqrt(diag(vb));
t=b./seb;
yh=x*b;
y0=y0(p+1:n);
ep=y0./sqrt(yh);
subplot(1,3,1)
plot(y0)
subplot(1,3,2)
plot(yh)
```

## Análisis de series temporales

---

```
subplot(1,3,3)  
plot(ep)
```

---

## 16. Casos de series temporales univariantes



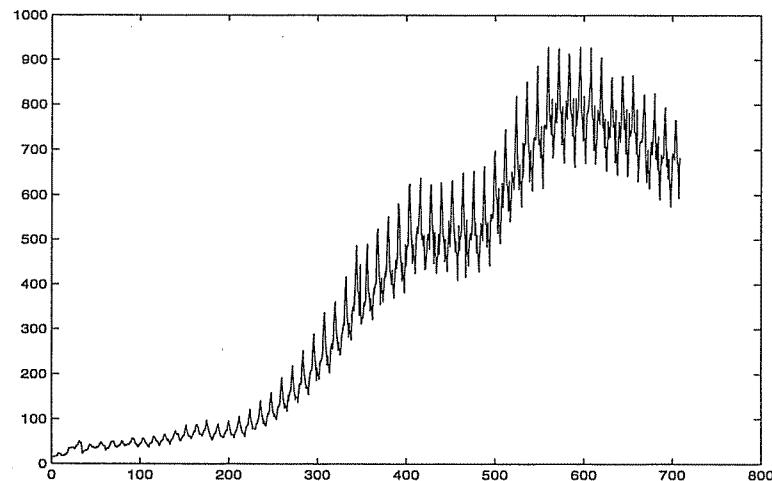
Gwilym M. Jenkins (1933 - 1982)

Estadístico británico. Fue profesor en Imperial College, Londres y en la Universidad de Lancaster. Escribió con George Box el libro de series temporales más influyente de la segunda mitad del siglo XX, basado en sus investigaciones conjuntas sobre procesos ARIMA. Abandonó la Universidad de Lancaster en 1974 para formar su propia empresa de consultoría y murió tras una larga enfermedad de cáncer a los 49 años.

### 16.1. Introducción

En este capítulo vamos a analizar con detalle tres series temporales, como repaso del material estudiado en los capítulos anteriores. Las tres elegidas son las del consumo de gasolina, el índice de producción industrial y el índice de bolsa de Madrid. La serie de consumo de gasolina en España tiene la ventaja de ser una serie muy larga y donde pueden estudiarse los efectos de las crisis del petróleo y es un buen ejemplo de una serie afectada por el entorno económico y político. La serie del índice de producción industrial se estudia sólo en el último periodo y se utiliza para ilustrar distintas correcciones de efectos deterministas. Finalmente, la serie del índice de bolsa de Madrid permite ilustrar efectos de heterocedasticidad condicional.

Figura 16.1 Serie de consumo de gasolina para automóviles en España desde 1945 hasta 2003



## 16.2. La serie de consumo de gasolina

### 16.2.1. Introducción

La serie que vamos a analizar corresponde al consumo de gasolina para automóviles en España. Disponemos de una serie muy larga, desde 1945 hasta 2003, y vamos a utilizarla para investigar los posibles cambios de modelo de una serie a lo largo del tiempo. Los datos de esta serie se encuentran en el fichero gasolauto45-03.dat y se presentan en la figura 16.1.

Como ilustración, vamos a presentar con detalle la modelización de esta serie en el periodo enero de 1966 a agosto de 1977. Hemos escogido esta época para incluir la primera crisis del petróleo, en 1974, donde los precios de los carburantes se doblaron en unos pocos días, y terminar con la segunda crisis del petróleo en 1977, que puede afectar a las predicciones realizadas con esta serie. Además el periodo 75/77 con la muerte de Franco y la llegada de la democracia fue un momento de grandes cambios en España. Los datos para este periodo, antes de actualizaciones y revisiones posteriores, se encuentran en el fichero gas66-77.dat y se han obtenido del Instituto Nacional de Estadística. Con estos datos vamos a:

1. Construir un modelo univariante y hacer predicciones.
2. Identificar atípicos y contrastar la linealidad de la serie.

## 16. Casos de series temporales univariantes

3. Comparar el modelo obtenido para este periodo con el modelo que sigue la serie en los años anteriores y posteriores.

### 16.2.2. Identificación del modelo univariante

El gráfico de la serie, que se presenta en la figura 16.2, muestra tendencia creciente, con un efecto estacional muy marcado que produce picos estacionales en los meses de julio y agosto. El gráfico muestra también que la variabilidad aumenta con el nivel, lo que sugiere transformar la serie, probablemente con el logaritmo, ya que la relación entre variabilidad y nivel parece aproximadamente lineal. Para confirmar este aspecto, la figura 16.3 presenta la desviación típica del consumo de gasolina cada año frente al consumo medio del año. La relación es aproximadamente lineal, aunque con cierta curvatura en los últimos años. Al hacer una regresión entre los logaritmos de las desviaciones típicas de los precios cada año y los precios medios del año, la pendiente estimada es .71, algo alejado del valor unidad que implicaría logaritmos. Como la serie parece tener algunos valores extremos, estudiaremos la relación entre nivel y variabilidad utilizando medidas robustas, como la media en lugar de la desviación típica para la variabilidad, y la mediana, en lugar de la media, para el nivel. El resultado obtenido ha sido en este caso muy similar: al relacionar las variables en logaritmos se obtiene una pendiente también menor que la unidad. Por razones de simplicidad, y para facilitar la interpretación de los resultados, vamos a aplicar de todas formas la transformación logaritmo.

El gráfico de la serie en logaritmos se presenta en la figura 16.4. Se observa que ha mejorado mucho la heterocedasticidad de la serie, pero existe cierta evidencia de que la hemos sobretransformado, ya que ahora la parte inicial parece tener un poco más de variabilidad que la final.

La función de correlación de la serie en logaritmos, que llamaremos  $\ln G_t$ , muestra que la *fas* decrece lentamente sugiriendo un proceso no estacionario. Como la tendencia creciente de esta serie es visible en el gráfico, tomaremos una diferencia mensual y, dada la marcada estacionalidad de la serie, observable a simple vista, también una diferencia estacional. La figura 16.5, la serie  $w_t = \nabla\nabla_{12} \ln G_t$ , representa los incrementos relativos de consumo de gasolina cada mes con relación al mismo mes del año anterior. Observemos que el primer dato del gráfico corresponde a la observación 13, ya que se pierden estos datos con las dos diferencias. Las funciones de autocorrelación simple y parcial para esta serie se presentan en la figura 16.6. En la función de autocorrelación simple observamos coeficientes significativamente distintos de cero en los retardos 1, 2, 3, 11, 12 y siguientes. Si olvidamos momentáneamente la parte estacional – definida por los retardos múltiplos de 12 y valores adyacentes– se observa

Figura 16.2 Gráfico de la serie de gasolina en el periodo 66-77

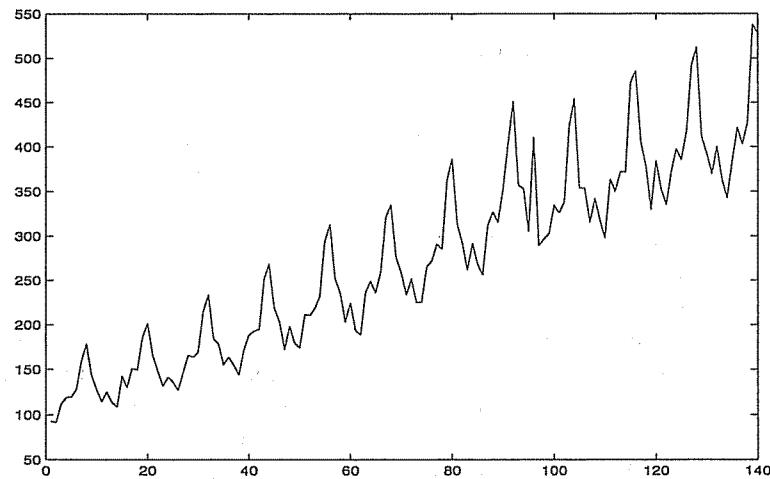
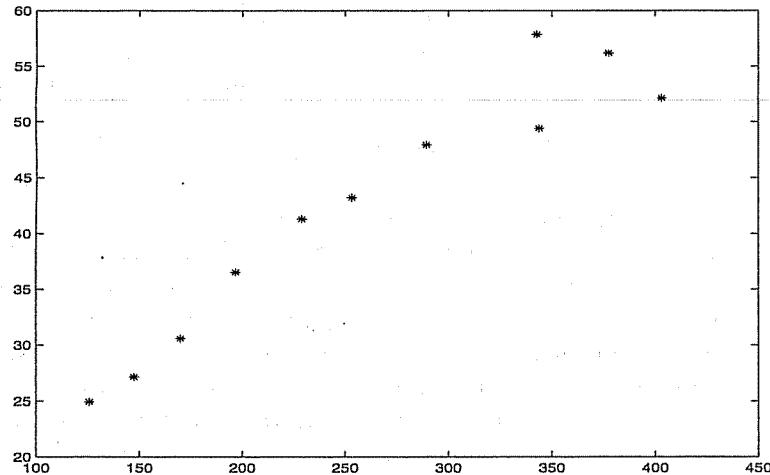


Figura 16.3 Desviación típica frente a la media en grupos de tamaño 12 para la serie de gasolina



que los dos coeficientes significativos alternan su signo de un retardo a otro. Las hipótesis más plausibles a la vista de esta estructura en la *fas* son:

1. El decrecimiento con los signos cambiados de los coeficientes podría corresponder a un AR(1) con coeficiente negativo. En esa hipótesis,

16. Casos de series temporales univariantes

Figura 16.4 Gráfico de la serie de gasolina en logaritmos

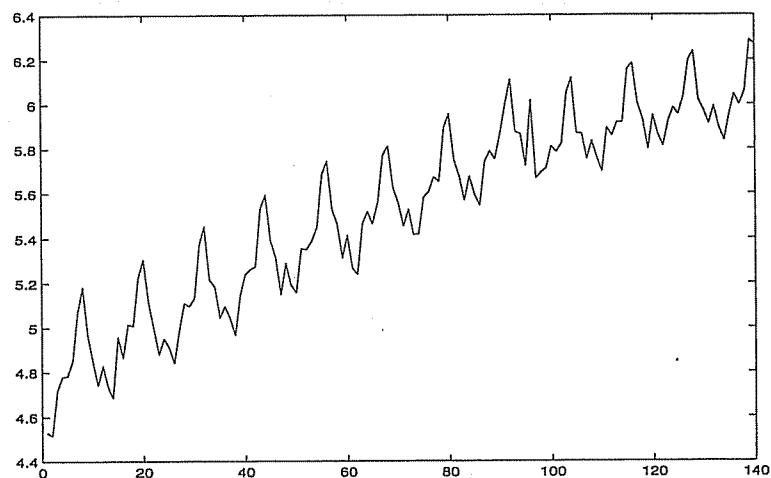
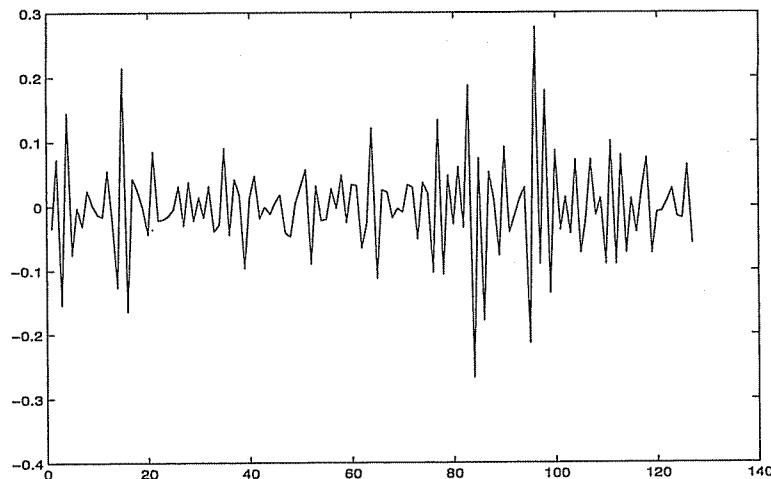


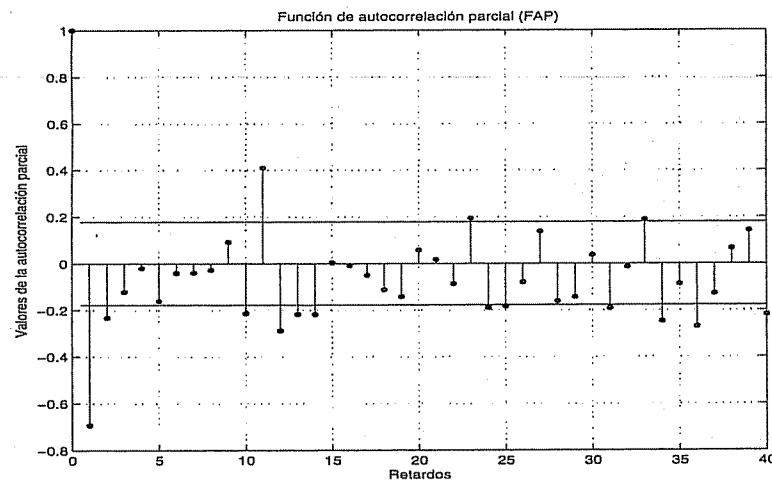
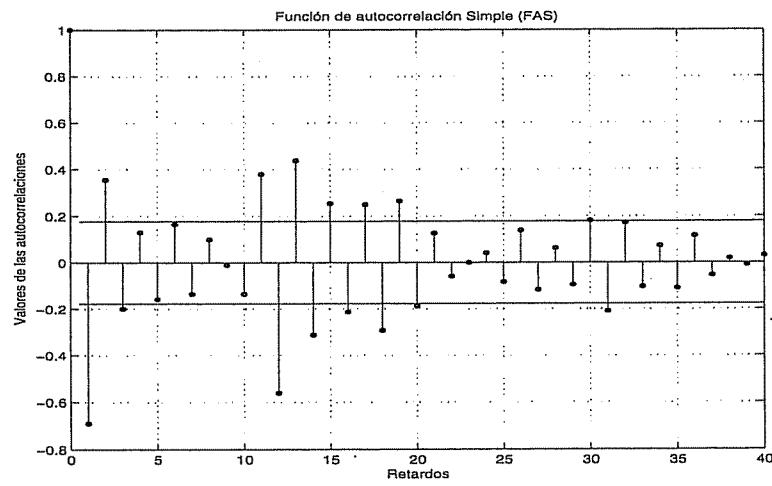
Figura 16.5 Serie del logaritmo del consumo de gasolina con una diferencia regular y otra estacional



los coeficientes deberían decrecer geométricamente en valor absoluto. El primer coeficiente es -.69, cuyo cuadrado es .48, que no está muy alejado del valor de .35 que es observado en el segundo retardo.

2. Los tres coeficientes significativos podrían corresponder a un MA(3).

**Figura 16.6 Correlograma simple y parcial para la serie en logaritmos y con dos diferencias del consumo de gasolina en España**



Respecto a la fap:

1. Los dos primeros coeficientes significativos indican un AR(2).
2. La estructura globalmente decreciente en los primeros retardos podría corresponder a un MA de orden bajo.

## 16. Casos de series temporales univariantes

Combinando ambas estructuras, rechazamos la hipótesis del AR(1), ya que el segundo coeficiente de la *fap* significativo no es compatible con esta hipótesis, y concluimos que los modelos AR(2), MA(3) y ARMA (1,1), son posibles.

Si pasamos a la parte estacional, la hipótesis más simple es una media móvil anual con coeficiente positivo. En la *fas* hay un valor significativo en el retardo 12 pero no en 24, ni en 36. En la *fap* se observan valores significativos en los coeficientes 12, 24 y 36, compatible con la estructura de MA(1) estacional. Supondremos inicialmente que la estructura estacional es MA(1) y pasaremos a estimar los tres modelos simples siguientes:  $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ ;  $ARIMA(1, 1, 1) \times (0, 1, 1)_{12}$ ;  $ARIMA(2, 1, 0) \times (0, 1, 1)_{12}$ .

Como ilustración, vamos a considerar que admitiésemos erróneamente que el proceso,  $\nabla_{12} \ln G_t$ , es estacionario y no tomásemos la diferencia mensual. Entonces, la función de autocorrelación simple muestra muchos coeficientes significativos que decrecen y la *fap* tiene tres coeficientes significativos, lo que llevaría a estimar el modelo  $ARIMA(3, 0, 0) \times (0, 1, 1)_{12}$ .

### 16.2.3. Estimación y selección del modelo

La tabla 16.1 presenta un resumen de los modelos estimados con el programa Minitab. El primer modelo tiene, en su parte autorregresiva, una raíz muy próxima a la unidad, confirmando la necesidad de tomar una diferencia regular. Formalmente, los modelos I y IV son parecidos. En ambos casos el valor del estadístico  $Q$  de Ljung-Box no muestra indicios de falta de independencia en los residuos estimados y ni el correlograma ni la función de autocorrelación parcial de los residuos muestran algún coeficiente de correlación significativamente distinto de cero. Podemos ver más claramente la similitud entre ambos modelos escribiéndolos de la misma forma. Para ello, si multiplicamos la diferencia y el operador AR(2) del modelo IV, se obtiene:

$$(1 + 0.71B + 0.12B^2)(1 - B) = (1 - 0.29B - 0.59B^2 - 0.12B^3)$$

que es un polinomio prácticamente idéntico al AR(3) del modelo 1.

Esto desecha la alternativa del modelo 1, ya que el modelo IV, con un parámetro menos tiene menor varianza residual, con lo que será preferido con cualquier criterio de selección de modelos. El estadístico  $Q$  del modelo II muestra claramente falta de independencia en los residuos y es el que tiene mayor varianza residual, con lo que queda descartado. Finalmente, el modelo III es el que presenta un menor error de previsión, aunque es virtualmente el mismo que para el modelo IV que tiene el mismo número de parámetros. Para ver la similitud más claramente, si multiplicamos

Tabla 16.1 Resumen de varios modelos estimados para la serie de gasolina

Modelo	$\hat{\sigma}$	$Q_{24}$	$Q_{36}$
I) $(1 - 0.27B - 0.58B - 0.11B^3)\nabla_{12} \ln G_t =$ $= 0.0044 + (1 - 0.81B^{12})a_t$	.0417	19.5	34.2
II) $\nabla\nabla_{12} \ln G_t = -0.0010 + (1 - 0.75B)(1 - 0.82B^{12})a_t$	.0422	47.5	59.1
III) $(1 + 0.49B)\nabla\nabla_{12} \ln G_t =$ $= -0.0017 + (1 - 0.25B)(1 - 0.84B^{12})a_t$	.0411	24.2	38.5
IV) $(1 + 0.71B + 0.12B^2)\nabla\nabla_{12} \ln G_t =$ $= -0.0021 + (1 - 0.85B^{12})a_t$	.0412	22.9	37.4

por la inversa del operador  $(1 - 0.25B)$ , tenemos que la parte regular del modelo III puede escribirse:

$$(1 + 0.49B)(1 - 0.25B)^{-1} = \\ = (1 + 0.49B)(1 + 0.25B + 0.0625B^2 + 0.0156B^3 + \dots) = \\ = (1 + 0.74B + 0.185B^2 + 0.03B^3 + \dots)$$

que es muy similar al operador AR(2) del modelo IV. La desviación típica residual del mejor modelo es .041, indicando que las predicciones a un paso tendrán un error estándar esperado del orden del 4.1%.

La tabla 16.2 compara la estimación del mejor modelo, el III, con cuatro programas. Se observa que los resultados de los dos programas que estiman por máxima verosimilitud exacta, (MVE), TSW y SCA, son muy similares, y algo distintos de los obtenidos por Minitab y EVViews, especialmente, para la media móvil estacional. Este resultado es esperable, ya que las diferencias serán mayores cuando condicionemos a muchas observaciones, como ocurre con AR largos estacionales o términos MA estacionales.

La figura (16.7) presenta la *fas* y la *fap* de los residuos del mejor modelo de la tabla 16.1, el ARIMA  $(1,1,1) \times (0,1,1)_{12}$ , obtenidos por el programa TSW. Puede observarse que, como indicaba el estadístico  $Q$ , no hay evidencia de autocorrelación residual. La figura 16.8 presenta el gráfico de los residuos. Las dos diferencias hacen que perdamos 13 observaciones, por lo que sólo disponemos de 127 datos para la estimación de los parámetros del proceso ARMA. Con esto el programa TSW cal-

## 16. Casos de series temporales univariantes

**Tabla 16.2 Comparación de la estimación del mejor modelo de la tabla 13.1 con distintos programas para la serie de gasolina**

Minitab	$(1 + 0.49B)\nabla\nabla_{12} \ln G_t = -.0017 + (1 - 0.25B)(1 - 0.84B^{12})a_t$ (.12) (.13) (.07)
EViews	$(1 + 0.61B)\nabla\nabla_{12} \ln G_t = -.0010 + (1 - 0.08B)(1 - 0.81B^{12})a_t$ (.07) (.05) (.0001)
TSW	$(1 + 0.46B)\nabla\nabla_{12} \ln G_t = -.0011 + (1 - 0.34B)(1 - 0.70B^{12})a_t$ (.11) (.13) (.07)
SCA	$(1 + 0.45B)\nabla\nabla_{12} \ln G_t = -.0015 + (1 - 0.36B)(1 - 0.68B^{12})a_t$ (.11) (.13) (.07)

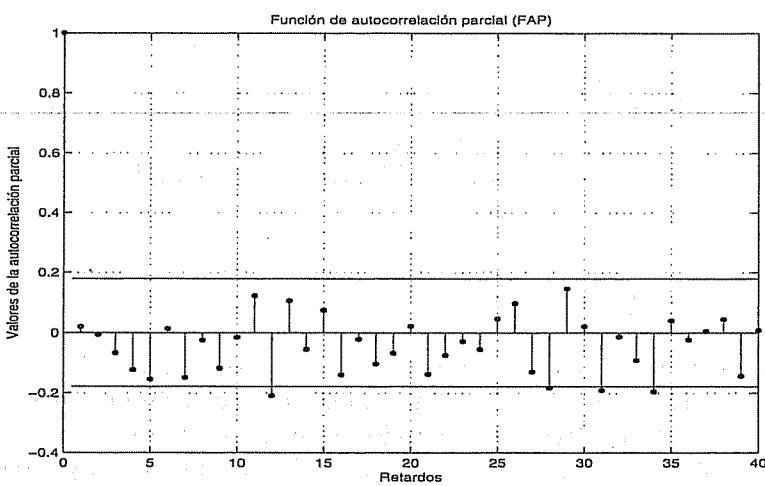
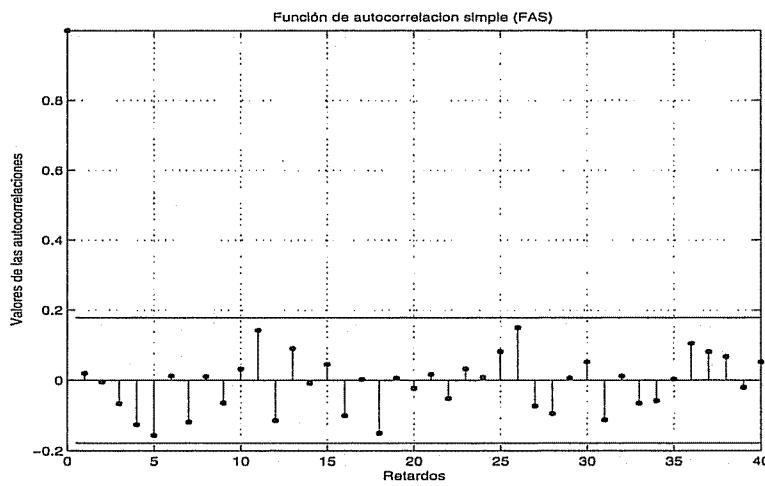
cula 126 residuos, por lo que el primer valor de la figura corresponde al 3/1967. Se observa en este gráfico un residuo muy alto en la observación 12/1973, seguido de una caída grande del 3/1974, que continúa en los meses siguientes. Los residuos altos en esta época son esperables, ya que en marzo de 1974 comenzó la primera crisis del petróleo, donde los precios del crudo se duplicaron en apenas unos días y estos aumentos se fueron trasladando paulatinamente a los consumidores. Una de las causas de esta crisis fue el fulminante éxito de Israel en la Guerra del Yom Kippur, que provocó las represalias de los productores árabes de petróleo y la OPEP aumentó a partir del 1 de enero de 1974 el precio del barril de petróleo de 5.11\$ a 11.61\$.

Prescindiendo de estos datos extremos, la distribución de los residuos es normal. El histograma es simétrico y los contrastes de Pearson y Kolmogorov Smirnov no rechazan la hipótesis de normalidad.

### 16.2.4. Predicciones

La figura 16.9 presenta las predicciones a 3 años con la serie de gasolina realizadas con el programa SCA. Como la serie tiene dos diferencias y una constante negativa, las predicciones generadas por este modelo serán a medio plazo la superposición de una tendencia cuadrática y coeficientes estacionales adaptativos. La constante del modelo es muy pequeña, por lo que la tendencia es prácticamente lineal. La tabla 16.3 indica las doce primeras predicciones, que corresponde al segundo año de crisis del petróleo, en 1977 y 1978. Se indican también los valores reales observados, los errores de predicción (error) y la desviación típica de las predicciones (std). A pesar de esta situación muy inestable, los errores de predicción son en casi todos los casos menores que dos veces la desviación típica, con la excepción de 2/1978 que es un 7% superior.

**Figura 16.7 Correlograma simple y parcial de los residuos del modelo para la serie de gasolina en España**



#### 16.2.5. Análisis de valores atípicos

El análisis de los residuos para identificar valores atípicos detecta un atípico aditivo y un cambio de nivel en los dos residuos extremos anteriores. En diciembre de 1973 (observación 96) se estima un AO de magnitud igual a .193 (sobre la serie en logaritmos). Este efecto corres-

---

## 16. Casos de series temporales univariantes

---

Figura 16.8 Residuos de la serie de gasolina

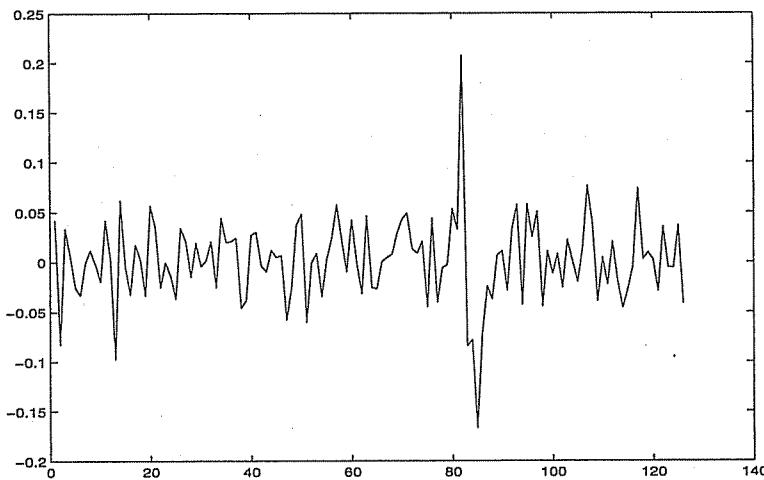
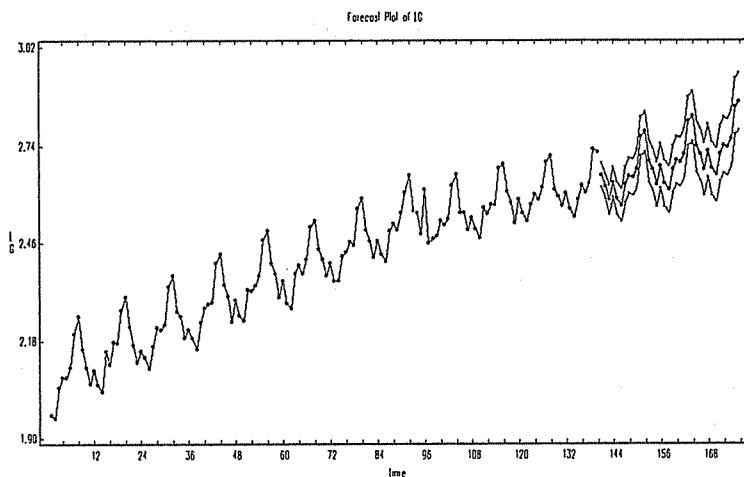


Figura 16.9 Predicción de 36 datos para la serie de los logaritmos de la gasolina



ponde a un aumento de la demanda por la prevista subida posterior de precios. El segundo valor atípico corresponde a un cambio de nivel que aparece en marzo de 1974, asociado a una importante subida del precio de la gasolina. Se observa que esta subida tuvo efectos a largo plazo y su efecto se estima en -0.091, es decir una disminución del 9 % del consumo. La subida del precio de la gasolina en ese instante fue de 17 a 20 pese-

**Tabla 16.3 Predicciones, errores estándar de predicción y valores reales para la serie de la gasolina**

	9/77	10/77	11/77	12/77	1/78	2/78
pred	6.023	5.915	6.032	5.912	5.865	5.994
real	6.038	5.933	6.031	5.937	5.869	6.116
error	0.015	0.018	-0.001	0.025	0.004	0.122
std	0.042	0.042	0.048	0.051	0.054	0.057
	3/78	4/78	5/78	6/78	7/78	8/78
pred	6.047	6.030	6.079	6.287	6.314	6.105
real	6.011	6.113	6.112	6.354	6.361	6.184
error	-0.036	0.083	0.033	0.067	0.047	0.079
std	0.060	0.063	0.065	0.068	0.070	0.073

tas, es decir del 17.6 %. Estas cifras permiten concluir que la elasticidad aparente en ese momento del precio de la gasolina era aproximadamente de .517 es decir, una subida del 17.6 % produjo una disminución del consumo a largo plazo de  $17.6\% \times .517 = 9.1\%$ .

Al estimar los parámetros del modelo incorporando estos efectos el modelo finalmente obtenido es:

$$\nabla \nabla_{12} \ln G_t = -.001 + 0.193 I_t^{(96)} - 0.091 S_t^{(99)} + \\ + \frac{(1 - 0.70B)(1 - 0.58B^{12})}{(1 + 0.32B)} a_t$$

donde la variable impulso,  $I_t^{(96)}$ , toma el valor uno en diciembre de 1973 y es cero en otro caso y la variable escalón,  $S_t^{(99)}$ , toma el valor uno a partir de marzo de 1974. La desviación típica residual de la serie sin ajustes por atípicos es de 0.041, que indica que puede preverse con un error del 4.1 %. Al incluir el efecto de estos dos atípicos, la desviación típica residual disminuye hasta 0.034, es decir un error de predicción a un paso del 3.4 %.

#### 16.2.6. Análisis de efectos deterministas

La serie de gasolina puede estar afectada por ciertas variables deterministas conocidas. Por ejemplo, si el consumo de gasolina es distinto en los distintos días de la semana, el consumo en un mes puede depender del número de fines de semana que existen en el mes. También las vacaciones y fiestas no recogidas por la estacionalidad pueden afectar al consumo de gasolina, y en especial la Semana Santa, que puede ocurrir en marzo o en

---

## 16. Casos de series temporales univariantes

---

abril y que tiene un efecto claro sobre el consumo. Estos efectos deterministas, que comentamos en el capítulo 12, pueden incluirse fácilmente con el programa TSW, que permite su incorporación automática.

La tabla 16.4 siguiente resume la estimación de estos efectos. La media de la serie estacionaria, parámetro MU, tiene un estadístico t significativamente distinto de cero, indicando un término de crecimiento lineal determinista. Los parámetros del modelo ARMA son TH1, la media móvil regular con coeficiente .59 (recordemos que en la notación que utiliza TSW los coeficientes tienen los signos cambiados respecto a la utilizada en este libro). La media móvil estacional se denota por BTH, y resulta con un coeficiente de .386. Los parámetros que vienen a continuación corresponden a la variable que tiene en cuenta el número de lunes, martes, etc., en cada mes. Este conjunto de siete variables se conoce como variables del efecto *trading day*, y pueden incorporarse automáticamente en programas como TSW y SCA. Las variables TRAD  $i$ , desde  $i = 1, \dots, 6$ , que llamaremos  $D_i$ , se definen como la diferencia entre el número de lunes, martes, ..., sábados menos el número de domingos. De esta forma se evita el problema de la multicolinealidad, ya que se incluyen seis variables, una para cada día de lunes a sábado. La variable TRAD 7, que representamos como  $D_7$ , es la suma total de días en el mes y tiene en cuenta en particular si el año es bisiesto o no. Se observa que las variables TRAD 1, 4, 6, y 7, es decir, las asociadas al lunes, jueves y sábados y la asociada al número total de días son claramente significativas. Esto implica que el lunes se utiliza más el coche, con un claro aumento del consumo de gasolina, el sábado también y el día con menos uso es el jueves, que tiene un efecto negativo. El número total de días en el mes tiene un claro efecto significativo con un coeficiente de 0.038.

La variable EAST mide el efecto de la Semana Santa y la representaremos por  $P_t$ , el efecto Pascua. Esta variable es cero salvo en los meses donde ocurre la Semana Santa y entonces pone en cada mes el número de días que corresponden a ese mes. Por ejemplo, si la Semana Santa ocurre toda en un mes, ese mes tiene un 7 y cero el resto. Sin embargo, si la Semana Santa se divide entre dos meses con 3 días en uno y 4 en otro, estos son los valores que se incluyen en los meses.

Finalmente las variables OUT corresponden a los atípicos detectados por el programa. Hay tres atípicos AO, que representaremos con las tres variables impulso  $I_t^{(h)}$ , que toman el valor 1 en el mes afectado por el atípico aditivo y cero en otro caso. En diciembre del 73 se produce un aumento del 19.1 %. En el mes siguiente se detecta un atípico de signo negativo y en julio del 77 otro ajuste positivo. Además, hay una disminución del consumo correspondiente al escalón en marzo del 74 que corresponde a una subida de precio y que produjo una disminución del consumo que ahora se estima en el 12 %.

**Tabla 16.4 Estimaciones para la serie de gasolina con el programa TSW**

parámetro	estimación	error estándar	estadístico t
MU	-11558E-02	0.00054	-2.12
TH1	-0.5937	0.07830	-7.58
BTH	-0.3863	0.09760	-3.96
TRAD 1	0.96380E-02	0.00361	2.67
TRAD 2	0.56171E-02	0.00347	1.62
TRAD 3	0.59047E-02	0.00353	1.67
TRAD 4	-71355E-02	0.00347	-2.06
TRAD 5	0.44821E-02	0.00368	1.22
TRAD 6	0.13145E-01	0.00364	3.61
TRAD 7	0.38855E-01	0.01032	3.77
EAST 1	0.56524E-01	0.00666	8.49
OUT 1 ( 96)	0.19145	0.01833	10.44 AO (12 1973)
OUT 2 ( 99)	-0.12007	0.01641	-7.32 LS ( 3 1974)
OUT 3 ( 97)	-0.72920E-01	0.01871	-3.90 AO( 1 1974)
OUT 4 (139)	0.80084E-01	0.02181	3.67 AO ( 7 1977)

El modelo estimado para esta serie tiene dos componentes. En primer lugar, cuando la serie se corrige de efectos deterministas sigue el modelo ARIMA:

$$\nabla \nabla_{12} \ln G_t^C = -0.001 + (1 - 0.59B)(1 - 0.38B^{12})a_t,$$

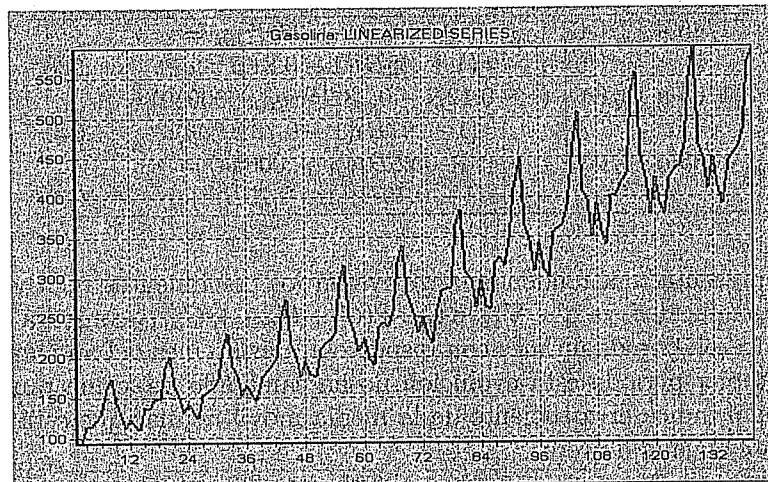
donde ahora la serie  $G_t^C$  es una serie corregida de los efectos deterministas que se describen a continuación. La desviación típica residual de esta serie es sólo 0.022, que indica que esta serie corregida puede preverse a un paso con un error de sólo el 2.2 %. El efecto de todas estas variables deterministas es pues reducir el error de predicción a un paso de esta serie en el 50 % aproximadamente, del 4.1 del modelo univariante al 2.2 de este modelo.

La serie corregida es el resultado de limpiar la serie observada de tres tipos de efectos distintos. El primero es la estructura de días laborables en el mes, el segundo la situación de la Semana Santa y el tercero los valores atípicos. Es decir:

$$\ln G_t^C = \ln G - \sum_{i=1}^7 w_i D_i - .056 P_t - \sum w_i I_t + .12 S_t^{(3/74)}.$$

Las variables  $D_i$  miden el efecto de los días en el mes. Como hemos co-

**Figura 16.10 Serie de gasolina corregida de atípicos y efectos deterministas**



mentado, las asociadas al lunes, jueves y sábados y la asociada al número total de días son claramente significativas. La variable del efecto Semana Santa,  $P_t$ , es también claramente significativa con coeficiente .056. Finalmente, tenemos la corrección por los tres AO, medidos por las variables impulso  $I_t$ , y del escalón. La figura 16.10 presenta un gráfico de esta serie  $G_t^C$ , que es la serie de gasolina corregida de todos los efectos deterministas.

#### 16.2.7. Análisis en distintos períodos

Como ilustración de los cambios en el modelo necesario para representar una serie a lo largo del tiempo, la tabla 16.5 indica modelos estimados para cada uno de los tres períodos en que hemos dividido la serie. En los tres casos  $g_t = \nabla \nabla_{12} \ln G_t$  es la serie estacionaria.

Se observa que la varianza de los residuos ha disminuido a lo largo del tiempo, haciendo la serie más predecible: del primer periodo al último se ha reducido en la mitad. En la segunda parte el alto coeficiente de la media móvil sugiere un término de crecimiento determinista, y si pasamos la media como constante al otro lado y cancelamos la diferencia regular con el término de media móvil, tendremos:

$$\nabla_{12} \ln G_t = -.00089t + \frac{(1 - 0.76B^{12})}{(1 - 0.12B - 0.38B^2)} a_t$$

Tabla 16.5 Modelos para la serie de gasolina por períodos

Período	modelo	$\hat{\sigma}_R$
1/45 a 12/64	$(1 + 0.65B + 0.26B^2)g_t =$ $= (1 + .48B)(1 - 0.81B^{12})a_t$	.066
1/65 a 12/84	$(1 - 0.12B - 0.38B^2)(g_t + .00089) =$ $= (1 - .97B)(1 - 0.76B^{12})a_t$	.040
1/85 a 12/03	$(1 + 0.28B)(g_t + .00054) =$ $= (1 - 0.79B)(1 - 0.75B^{12})a_t$	.030

que representa una tendencia negativa en los crecimientos relativos desestacionalizados, ya que  $\nabla_{12} \ln G_t \approx (G_t - G_{t-12})/G_{t-12}$  compara el crecimiento relativo de un mes con respecto al mismo mes del año anterior. En el tercer periodo esta evolución no ha cambiado de forma tan regular, como indica el coeficiente de la media móvil regular que está ahora bastante alejada de cero.

### 16.3. El índice de producción industrial

#### 16.3.1. Introducción

Vamos a construir un modelo univariante para la serie mensual del índice de producción industrial en España. En el capítulo 18 relacionaremos esta serie con un posible indicador avanzado de su evolución, que es el mercado de valores. Se supone que los movimientos en el mercado de acciones reflejan las expectativas de la economía española y que, en consecuencia, subidas en la bolsa deberían, en promedio, ir seguidas de subidas en la producción. En el capítulo 18 investigaremos estas hipótesis para la economía española

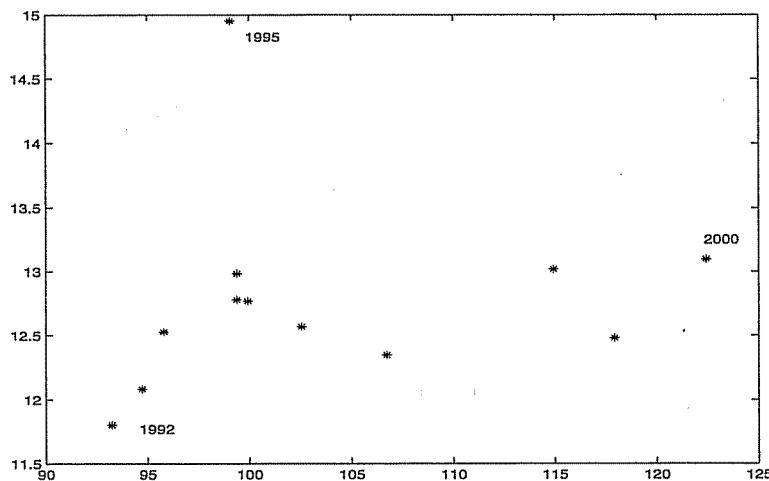
#### 16.3.2. El modelo univariante para el IPI

La figura 16.12 muestra el índice de producción industrial para el periodo 1/1988 a 12/2000. La serie se encuentra en el fichero ipies88-00.dat. La figura muestra que esta serie tiene un componente estacional muy marcado con grandes caídas en el mes de agosto y un tendencia creciente a partir de 1994, aproximadamente. No se observa una dependencia clara de la variabilidad respecto a la media, y esta impresión visual del gráfico de la serie se confirma en la figura 16.11, que representa la desviación típica respecto a la media de la serie.

La serie es claramente no estacionaria y su función de autocorrelación confirma la información del gráfico: se necesita una diferencia regular y

## 16. Casos de series temporales univariantes

Figura 16.11 Desviación típica media cada año respecto a la media para los datos del IPI en España



otra estacional para obtener una serie estacionaria. La figura 16.13 muestra las funciones de autocorrelación simple y parcial de la serie diferenciada,  $\nabla\nabla_{12}IPI_t$ . Comenzando con la *fas*, el gran número de coeficientes significativos sugiere estructura AR para la parte regular, y los dos coeficientes significativamente distintos de cero en la *fap* indican que el orden puede ser un AR(2). La estructura estacional muestra en la *fas* coeficientes significativos en 12 y 36, lo que sugiere estructura AR. En la *fap*, aunque los valores en los retardos estacionales no son significativos, la fuerte interacción alrededor de estos coeficientes sugiere que son distintos de cero, lo que implicaría estructura MA. Es especialmente llamativa la interacción en el retardo 35. Concluimos que la estacionalidad tiene probablemente estructura ARMA.

La tabla 16.6 indica los resultados obtenidos en la estimación de distintos modelos con el programa SCA, y la tabla 16.7 con EViews. Aunque las estimaciones con ambos programas difieren, en ambos casos el mejor modelo es el primero, ya que tiene menor varianza residual con menor número de parámetros, y esto se confirma con los criterios BIC y AIC.

Los residuos de este modelo y su correlograma se presentan en la figura 16.14. El gráfico de los residuos muestra un residuo muy alto, que, teniendo en cuenta los residuos perdidos por la estimación, corresponde al 4/1997. Los valores posteriores a la ocurrencia de este atípico también muestran cierta estructura de falta de aleatoriedad, por lo que conviene investigar la presencia de posibles valores atípicos. El correlograma tiene

Figura 16.12 Índice de producción industrial en España entre enero de 1988 y diciembre de 2000

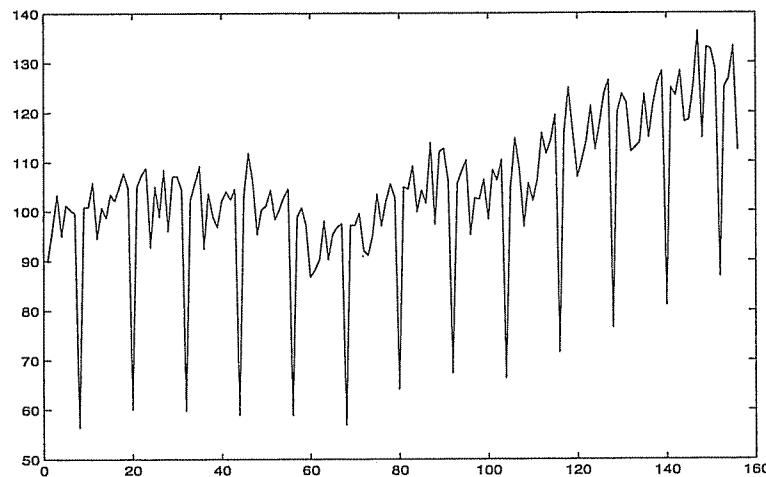


Tabla 16.6 Modelos estimados para el IPI con el programa SCA

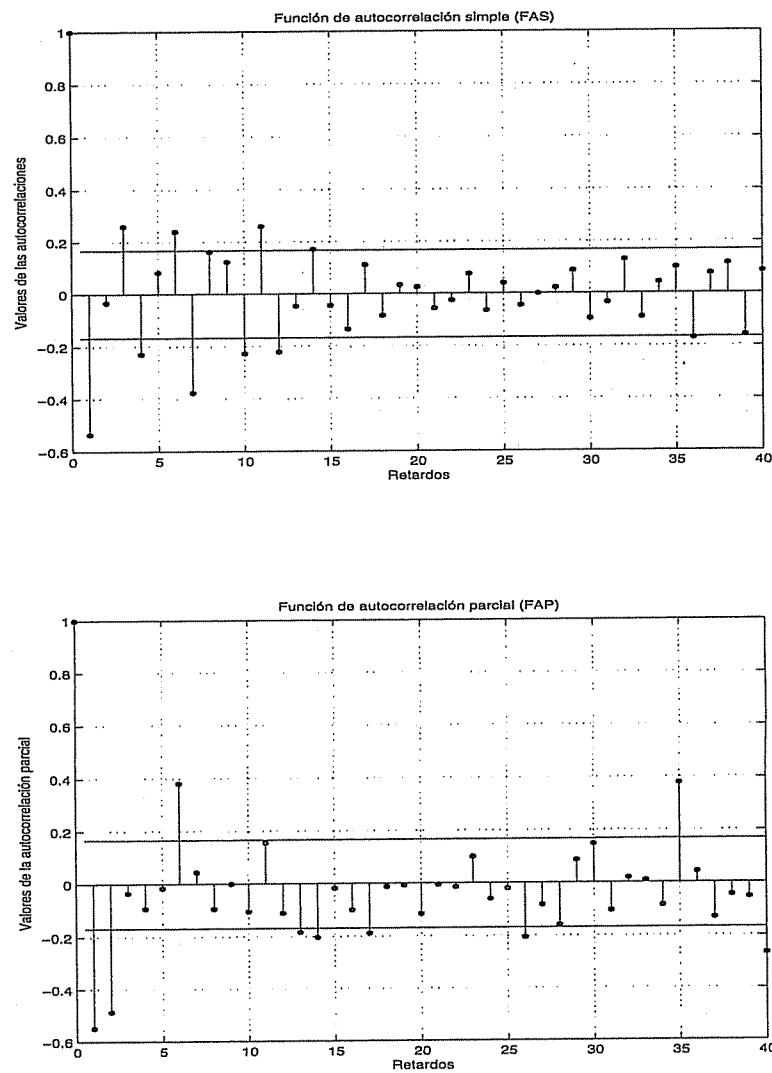
modelo	$\hat{\sigma}$	$Q(24)$	$Q(39)$
$(1 + .87B + .55B^2)\nabla\nabla_{12}IPI_t =$ $= (1 - .78B^{12})a_t$	3.37	30.3	54.4
$(1 + .43B^{12} + .26B^{24})(1 + .88B + .60B^2)\nabla\nabla_{12}IPI_t =$ $= (1 - .04B^{12})a_t$	3.71	34.3	62.3
$(1 + .50B^{12} + .40B^{24} + .28B^{36}) \times$ $\times (1 + .88B + .60B^2)\nabla\nabla_{12}IPI_t = a_t$	3.62	29.8	51.5

un coeficiente alto en el retardo 7, y coeficientes de mediano tamaño en los retardos estacionales, lo que sugiere que la estructura estacional

Tabla 16.7 Modelos estimados para el IPI con el EViews

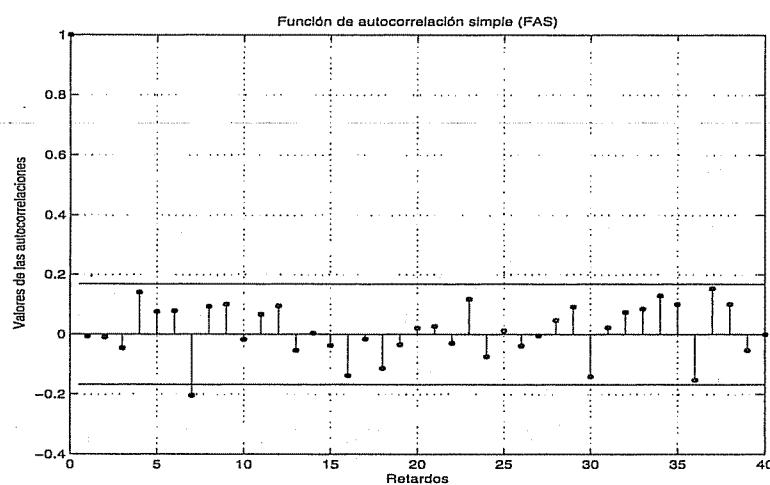
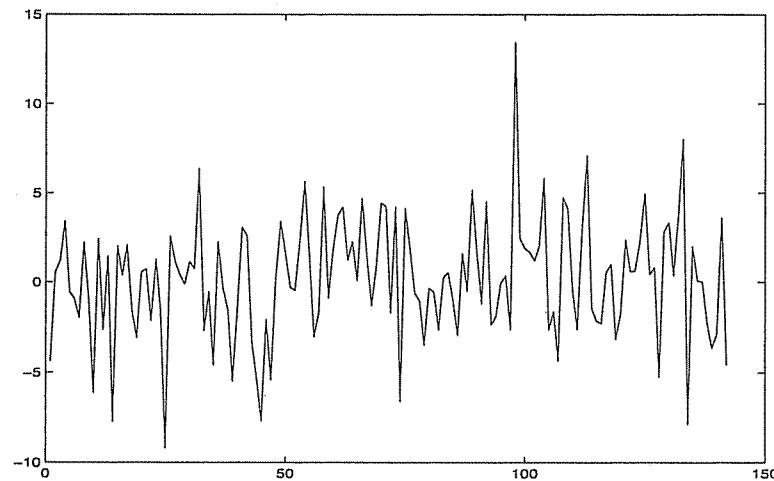
modelo	$\hat{\sigma}$	BIC	AIC
$(1 + .87B + .55B^2)\nabla\nabla_{12}IPI_t = (1 - .89B^{12})a_t$	3.27	5.29	5.23
$(1 + .11B^{12} + .11B^{24})(1 + .88B + .60B^2)\nabla\nabla_{12}IPI_t =$ $= (1 - .86B^{12})a_t$	3.36	5.42	5.30
$(1 + .21B^{12} + .19B^{24} + .23B^{36}) \times$ $\times (1 + .65B + .49B^2)\nabla\nabla_{12}IPI_t = a_t$	3.99	5.77	5.65

Figura 16.13 Fas y fap para el IPI en España



del modelo quizás puede ser mejorada. De todas formas, los valores del estadístico  $Q$  son compatibles con que la serie de residuos sea ruido blanco en el contraste de Ljung-Box al 95 %.

**Figura 16.14 Gráfico y correlograma de los residuos del modelo univariante para la serie del IPI en España**



### 16.3.3. Datos atípicos

La búsqueda de atípicos mediante el programa SCA detecta un IO en la posición 112 de valor 14.48 y un estadístico  $t$  de 4.21, lo que lo hace significativo sin ninguna duda. Como el modelo de la serie tiene una diferencia regular y otra estacional, este IO puede confundirse fácilmente con un

efecto rampa o de cambio de tendencia. Para comprobar este aspecto, supongamos que el verdadero efecto es rampa. Entonces, el modelo es:

$$IPI_t = wR_t^{(112)} + \frac{(1 - .78B^{12})}{\nabla\nabla_{12}(1 + .87B + .55B^2)} a_t$$

y como:

$$\nabla\nabla_{12}R_t^{(T)} = \nabla_{12}S_t^{(T)} = S(B)\nabla S_t^{(T)} = S(B)I_t^{(T)},$$

donde  $S(B) = 1 + B + \dots + B^{11}$ , el modelo de la rampa es equivalente al

$$\nabla\nabla_{12}IPI_t = wS(B)I_t^{(112)} + \frac{(1 - .78B^{12})}{(1 + .87B + .55B^2)} a_t$$

que implica una intervención del mismo tamaño durante 12 meses. La estimación de este modelo lleva a

$$\nabla\nabla_{12}IPI_t = (.27)S(B)I_t^{(112)} + \frac{(1 - .79B^{12})}{(1 + .87B + .55B^2)} a_t$$

y la variable de intervención no es significativa. Por lo tanto, concluimos que sobre la serie no ha ocurrido un efecto rampa puro, sino un IO en el instante 112, es decir, un cambio en las variables externas que ha afectado la evolución futura de la serie de una manera compleja, introduciendo una tendencia positiva a partir de ese punto, pero también otros efectos.

#### 16.3.4. Predicciones

La tabla 16.8 presenta las predicciones a un año del IPI obtenidas con el modelo anterior, los valores reales del IPI durante el 2001, los errores de predicción y los errores estándar de las predicciones. Se observa que los errores están siempre dentro de dos desviaciones típicas, como sería esperable si el modelo es correcto.

### 16.4. El modelo para la serie del índice general de la Bolsa de Madrid

Vamos a analizar la serie mensual del índice general de la Bolsa de Madrid en el periodo 88 a 2000. Los datos se encuentran en el fichero indicebolsa88-00.dat. La figura 16.15 presenta el índice y la serie de rendimientos  $r_t = \log z_t - \log z_{t-1}$ .

La función de autocorrelación de los rendimientos no muestra ningún coeficiente significativo y parece una serie de ruido blanco. La serie de los cuadrados de los rendimientos tiene dos coeficientes significativos, en los

**Tabla 16.8 Predicciones, valores reales, errores y errores estándar para la serie del IPI**

	1/01	2/01	3/01	4/01	5/01	6/01
pred	120.8	123.8	128.5	120.1	129.6	130.6
real	124.2	120.9	131.4	114.4	131.9	129.4
error	3.4	-2.9	2.9	-5.7	2.3	-1.2
std	3.34	3.38	3.59	4.19	4.31	4.58
	7/01	8/01	9/01	10/01	11/01	12/01
pred	131.8	86.9	128.0	131.3	132.3	119.2
real	128.0	89.7	121.5	130.6	127.0	107.4
error	-3.8	2.8	-6.5	-0.7	-5.3	-11.8
std	4.90	5.08	5.33	5.58	5.78	6.01

retardos uno y cinco, lo que sugiere un pequeño efecto de heterocedasticidad condicional.

La estimación de un AR(1) para la serie de rendimientos no conduce a resultados significativos. La de un AR(1) para la serie de los cuadrados conduce al modelo:

$$\hat{r}_t^2 = 0.0028 + 0.2169 r_{t-1}^2. \quad \begin{matrix} (4.88) & (2.74) \end{matrix}$$

La figura 16.17 presenta los rendimientos originales, las volatilidades estimadas con la ecuación anterior y los rendimientos corregidos por la heterocedasticidad condicional. Se observa que el modelo de volatilidad ha explicado uno de los dos valores atípicos. El valor extremo corresponde al dato de agosto de 1998. Al poner esta fecha en un buscador de Internet (Google) aparece la noticia de que el 12 de agosto la Bolsa pierde un 3% por la crisis asiática, y ese mes aparece destacado en la prensa, por lo que confirmamos el carácter atípico de ese dato.

## 16.5. Lecturas complementarias

La mayoría de los textos de series temporales incluidos en las referencias presentan análisis de series reales y Pankratz (1983) contiene numerosos casos de aplicación. Ejemplos de análisis de series con datos tecnológicos o industriales se encuentran en Pandit y Wu (1983), con datos biológicos en Diggle (1990), económicos en Espasa y Cancelo (1993) y financieros en Tsay (2002). El lector puede encontrar muchos otros ejemplos en artículos publicados en las revistas *Journal of Forecasting*, *International Journal of Forecasting* y *Journal of Time Series Analysis*.

16. Casos de series temporales univariantes

Figura 16.15 Gráfico de la serie del logaritmo del índice general de la Bolsa de Madrid y de su primera diferencia

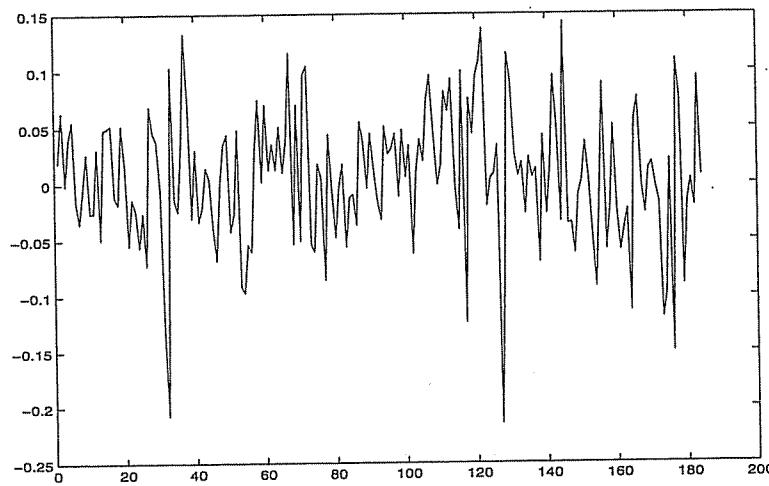
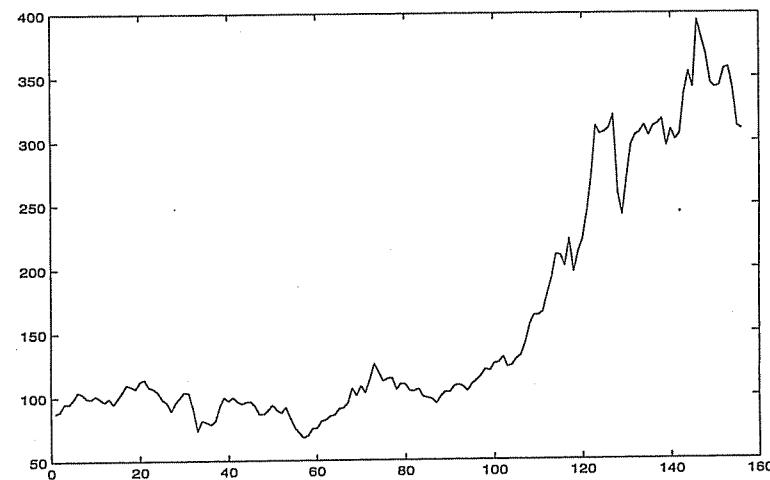


Figura 16.16 Correlograma de los rendimientos y de sus cuadrados para la Bolsa de Madrid

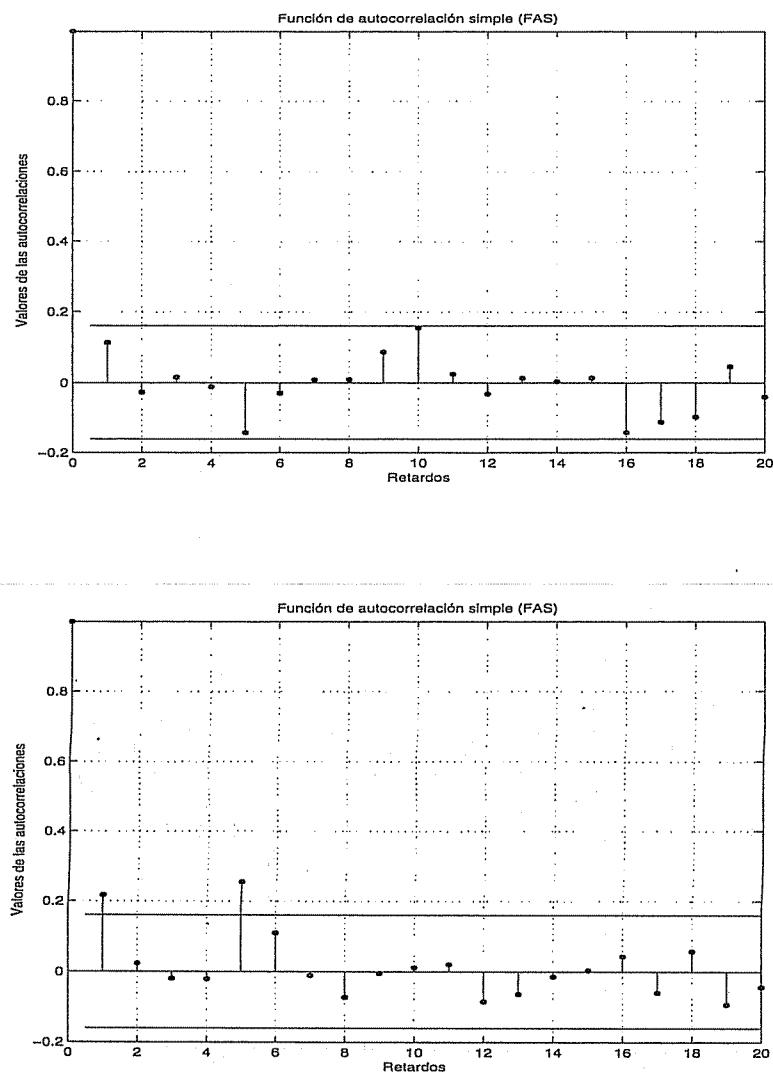
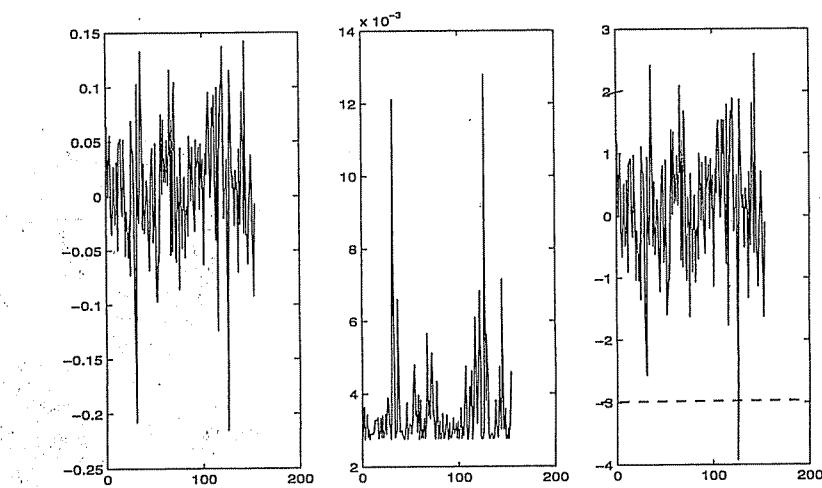


Figura 16.17 Serie de rendimientos, volatilidades estimadas y rendimientos corregidos por volatilidad



## 17. Regresión dinámica entre variables estacionarias



James Durbin (1923)

Estadístico inglés. Ha sido profesor de la London School of Economics desde 1950 hasta su retiro en 1988. Fue pionero en el estudio de la regresión entre series e introdujo el famoso estadístico de Durbin-Watson para detectar autocorrelación en los residuos. Ha hecho destacadas contribuciones al análisis de series temporales y a sus aplicaciones económicas.

### 17.1. Introducción

Un modelo de regresión dinámica es una representación de la relación existente entre dos o más series temporales. El modelo de regresión lineal simple supone dos variables,  $x_t, y_t$ , que siguen procesos de ruido blanco y expresa la media condicionada de la variable dependiente o *endógena*,  $y_t$ , en función de los valores de la explicativa o *exógena*,  $x_t$ , mediante:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (17.1)$$

donde las variables  $u_t$ , que no son observadas, son las innovaciones del modelo que siguen también un proceso de ruido blanco. Al intentar aplicar este modelo a variables que no son ruido blanco y pueden tener dependencia temporal nos encontramos con tres problemas principales.

El primero es que la ecuación (17.1) supone que la relación entre las dos variables es instantánea, cuando entre variables dinámicas la relación puede ser más compleja y transmitirse con ciertos retardos: el valor de

$y_t$  puede depender no de  $x_t$  sino de  $x_{t-k}$ , con un retardo  $k$  desconocido, o depender de todas las variables  $(x_t, x_{t-1}, \dots, x_{t-k})$ . Por ejemplo, los lectores habrán seguramente advertido que si tratamos de aumentar la temperatura del agua en la ducha abriendo la llave del agua caliente, la respuesta del sistema suele ser gradual, de manera que la temperatura de salida del agua tarda unos segundos en alcanzar el nivel deseado. Esto ocurre especialmente cuando existe cierta distancia entre la ducha y la caldera que calienta el agua. Si llamamos  $y_t$  a la temperatura del agua que sale de la ducha en el segundo  $t$  y  $x_t$  a la apertura de la llave en ese instante, un aumento de  $x_t$  en  $t = t_0$  produce efectos en la respuesta  $y_t$  en los segundos  $t_0 + 1, t_0 + 2, \dots, t_0 + k$ , aumentando la temperatura gradualmente durante  $k$  segundos hasta un cierto nivel. En consecuencia,  $y_t$  depende no sólo de la apertura de la llave en  $t$ ,  $x_t$ , sino de la apertura en los segundos previos, es decir, de las variables  $x_{t-1}, \dots, x_{t-k}$ . Como segundo ejemplo, si  $y_t$  representa los pasajeros transportados en un vuelo el día  $t$  y  $x_t$  es el precio del billete en ese día, un aumento del precio hoy tendrá poco efecto sobre los pasajeros transportados hoy, ya que la gran mayoría habrán comprado su billete antes de la subida. El incremento del precio,  $x_t$ , afectará a los pasajeros futuros,  $y_{t+1}, y_{t+2}, \dots$ , ocasionando una disminución de la demanda en los próximos días. Esto es equivalente a decir que la variable  $y_{t+k}$  dependerá no sólo de  $x_{t+k}$ , sino también de los valores previos  $x_{t+k-1}, \dots, x_t$ . El efecto dinámico será tanto más complejo cuantos más intervalos de tiempo sean necesarios para la transmisión completa del efecto. Por ejemplo, si trabajamos con datos anuales, todos los efectos que se transmitan en pocos días o meses aparecerán como efectos instantáneos, es decir todo el efecto de una subida de precios en el año  $t$  se observará en la respuesta ese año, ya que todos los efectos quedan incluidos en el periodo de observación de la serie.

El segundo problema de (17.1) es que este modelo supone implícitamente que la relación va de la variable  $x_t$  a la  $y_t$ , pero que la variable  $y_t$  no influye sobre la  $x_t$ . Decimos que la variable  $y_t$  es *endógena*, influenciada por  $x_t$ , pero que  $x_t$  es *exógena* o explicativa, lo que implica que no hay relación entre  $y_{t-k}$  y  $x_t$  para cualquier  $k > 0$ . Por ejemplo, la cantidad de lluvia influye sobre el crecimiento de una planta, pero no es esperable que el crecimiento de la planta afecte a la lluvia futura. Aunque en muchas aplicaciones físicas o técnicas el carácter exógeno de la variable  $x_t$  es evidente por el mecanismo generador de los datos, en otros conviene comprobarlo a la vista de los datos. Por ejemplo, los gastos en publicidad,  $x_t$ , es esperable que influyan sobre las ventas futuras,  $y_{t+k}$ , pero es posible que las decisiones sobre los gastos de publicidad futuros dependan de las ventas pasadas. En este caso no sólo hay una relación desde  $x_t$  hacia  $y_{t+k}$  sino que existe realimentación o causalidad bidireccional

entre ambas variables de manera que  $y_t$  también influye sobre  $x_{t+m}$ . En este capítulo supondremos que conocemos la dirección de causalidad, y, por tanto, que el modelo de regresión dinámica es adecuado para las variables. Cuando sospechemos que la variable  $y_t$  puede también influir sobre la  $x_{t+m}$  tendremos que utilizar los métodos que se presentan en el capítulo 19, que permiten causalidad bidireccional.

El tercer problema de la ecuación (17.1) es suponer que la parte de la variable respuesta no explicada por la variable independiente  $x_t$ , es un proceso  $u_t$  de variables independientes. Esta hipótesis es natural para datos estáticos, pero no suele verificarse con datos dinámicos. Si pensamos de nuevo en el ejemplo de la temperatura del agua en la ducha, la ecuación (17.1) indica que cuando  $x_t$  (apertura de la llave) permanece constante, la temperatura del agua debe variar aleatoriamente de forma imprevisible. Sin embargo, nuestra experiencia diaria nos indica que, afortunadamente, la temperatura evoluciona siempre con inercia, de manera que la temperatura en un instante,  $y_t$  será próxima a la temperatura en el instante anterior,  $y_{t-1}$ , y ésta a  $y_{t-2}$ , etc., lo que supone correlación entre las innovaciones. De la misma forma, si el número de pasajeros transportados en un vuelo depende del día de la semana, las innovaciones estarán autocorreladas, por el efecto de la estacionalidad semanal.

En este capítulo vamos a estudiar modelos para relacionar series estacionarias y en el siguiente generalizaremos estos resultados para variables no estacionarias. Los modelos que vamos a estudiar se conocen en las ciencias físicas, naturales y de ingeniería como *modelos de función de transferencia* y en las ciencias económicas como *modelos econométricos dinámicos*. Describen cómo se transmiten los efectos desde una variable  $x_t$  a otra  $y_t$  cuando no existe realimentación o causalidad bidireccional y son un instrumento muy utilizado en todos los campos científicos para evaluar respuestas dinámicas. Si las variables explicativas son controlables, estos modelos permiten simular y evaluar políticas alternativas. Si no lo son, ofrecen la posibilidad de estudiar cómo ciertos «escenarios» (definidos por posibles evoluciones de la variable explicativa) afectan a la variable respuesta.

La utilidad de estos modelos para la predicción depende del intervalo entre observaciones. Si es grande, por ejemplo datos anuales, la relación entre la variable explicativa y la respuesta no suele contener retardos y para prever  $y_t$  es necesario conocer  $x_t$ . En el caso más general en que la variable  $x_t$  es estocástica, para preverla es necesario construir su modelo univariante y es posible que la mejora en la predicción de  $y_{t+k}$  que aporte el modelo dinámico basado en la predicción de  $x_t$  respecto al modelo univariante para  $y_t$  sea pequeña, por el error de predicción de la variable exógena. Sin embargo, el modelo puede ser muy adecuado para la

predicción condicional de la respuesta en función de posibles valores de la variable explicativa. La situación es distinta cuando el periodo de observación es corto, por ejemplo diario o cada minuto, donde es frecuente que existan retardos en la relación entre las variables, de manera que  $x_{t-k}$  afecta a  $y_t$  ( $k > 0$ ). En este caso diremos que la variable  $x_t$  es un indicador avanzado de la respuesta y como el regresor que necesitamos para prever la respuesta en el instante  $t$  es conocido, el modelo de regresión dinámica suele proporcionar mejores predicciones de  $y_t$  que el univariante.

## 17.2. Relaciones entre dos series estacionarias: las funciones de covarianzas y correlaciones cruzadas

### 17.2.1. La función de covarianzas cruzadas

Supongamos que tenemos dos procesos estacionarios  $x_t, y_t$ , con medias  $\mu_x, \mu_y$ . Vamos a introducir medidas de su dependencia lineal para distintos retardos. Definimos la *función de covarianzas cruzadas*, mediante:

$$\gamma_{xy}(t, t+k) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)]. \quad (17.2)$$

Esta función depende del orden en que se toman las variables y de los instantes en que se consideran cada una de ellas. La definición establece que la variable que aparece en primer lugar en el subíndice,  $x$ , se supone en  $t$  y la segunda,  $y$ , en  $t+k$ . Por tanto  $\gamma_{xy}(t, t+k)$  mide la relación lineal entre las dos variables  $x_t$  y  $y_{t+k}$ . El valor de  $k$  puede ser positivo o negativo. Se verifica que  $\gamma_{xy}(t, t+k)$  es idéntica a  $\gamma_{yx}(t+k, t)$ .

Diremos que dos procesos estacionarios  $x_t, y_t$ , son *conjuntamente estacionarios* si cada uno de ellos es estacionario y las covarianzas cruzadas sólo dependen del retardo entre las variables,  $k$ , y no del instante inicial considerado. Entonces, escribimos la función de covarianzas cruzadas como:

$$\gamma_{xy}(t, t+k) = \gamma_{xy}(k), \quad k = 0, \pm 1, \pm 2, \dots$$

En adelante, para simplificar la exposición, diremos normalmente que dos variables son estacionarias para indicar que son conjuntamente estacionarias. La función de covarianzas cruzadas para procesos estacionarios no varía si intercambiamos las variables y el signo del retardo simultáneamente:

$$\gamma_{xy}(k) = \gamma_{yx}(-k).$$

En efecto:

$$\begin{aligned}\gamma_{xy}(k) &= E[(x_t - \mu_x)(y_{t+k} - \mu_y)] = \\ &= E[(x_{t-k} - \mu_x)(y_t - \mu_y)] = \gamma_{yx}(-k)\end{aligned}$$

ya que, por la estacionaridad, la covarianza sólo depende del retardo. Concluimos que la función  $\gamma_{xy}(k)$ , ( $k = 0, \pm 1, \dots$ ) resume toda la dependencia lineal entre ambas variables y no es necesario calcular la  $\gamma_{yx}$ . Sin embargo, la función  $\gamma_{xy}(k)$  no es simétrica respecto al retardo  $k$ , y se verifican las propiedades siguientes:

1. Para  $k > 0$  los coeficientes  $\gamma_{xy}(k)$  representan cómo los valores de  $x_t$  influyen en los valores futuros  $y_{t+k}$ , lo que interpretamos diciendo que la parte positiva de esta función representa la relación «causal» de  $x_t$  hacia  $y_{t+k}$ .
2. Para  $k < 0$  los coeficientes  $\gamma_{xy}(k)$  representan cómo los valores de  $y_t$  influyen en los valores futuros  $x_{t+k}$ , lo que interpretamos diciendo que la parte negativa de esta función representa la relación «causal» de  $y_t$  hacia  $x_{t+k}$ .

Diremos que existe relación contemporánea o instantánea entre dos variables si  $\gamma_{xy}(0) \neq 0$ .

De la misma forma que las autocovarianzas permiten identificar el orden del modelo ARMA, podría pensarse en utilizar la función de covarianzas cruzadas para identificar si la relación es en una dirección o bidireccional y también obtener el número de retardos distintos de cero en la relación entre dos variables. Sin embargo no es así, ya que esta función tiene muchas limitaciones, como ilustramos a continuación. Supongamos que la relación entre dos variables estacionarias de media cero es:

$$y_t = v_0 x_t + v_1 x_{t-1} + a_{t+k} \quad (17.3)$$

Multiplicando por  $x_{t-k}$  y tomando esperanzas, resulta:

$$\begin{aligned}\gamma_{xy}(k) &= E(x_{t-k} y_t) = v_0 E(x_{t-k} x_t) + v_1 E(x_{t-k} x_{t-1}) + E(x_{t-k} a_t) = \\ &= v_0 \gamma_x(k) + v_1 \gamma_x(k-1).\end{aligned} \quad (17.4)$$

Vemos que la función de covarianzas cruzadas es una mezcla de los coeficientes  $v_i$  de la relación entre las dos variables y de la estructura de autocovarianzas de  $x_t$ , con lo que será complicada de interpretar. Supongamos que la variable  $x_t$  sigue un proceso AR(1), de manera que  $\gamma_x(k) \neq 0$ . Entonces, si existe correlación contemporánea,  $v_0 \neq 0$ , las correlaciones cruzadas  $\gamma_{xy}(k)$  serán no nulas para  $k > 0$ , indicando relación desde  $x_{t-k}$

---

## 17. Regresión dinámica entre variables estacionarias

---

hacia  $y_t$ , aunque según (17.3) esta relación sólo existe para  $k = 0, 1$ . Además, también serán no nulas para  $k < 0$ , indicando una relación desde  $y_{t-k}$  hacia  $x_t$ , que no existe. Por esta razón no podemos utilizar las covarianzas cruzadas para determinar ni la dirección de la relación, ni el número de retardos en la ecuación. Sin embargo, si la serie  $x_t$  fuese ruido blanco, entonces este problema desaparece. Entonces, como  $\gamma_x(k) = 0$  para  $k \neq 0$ , las únicas covarianzas cruzadas no nulas son  $\gamma_{xy}(0)$  y  $\gamma_{xy}(1)$ , con valores proporcionales a los coeficientes  $v_i$  de la relación.

En resumen, si  $k < 0$ , todas las covarianzas cruzadas  $\gamma_{xy}(k)$  son cero, podemos concluir que no hay relación desde  $y_{t-k}$  hacia  $x_t$ , pero lo contrario no es cierto: es posible que la causalidad sea unidireccional, como en el ejemplo (17.3), y que encontremos coeficientes distintos de cero en esta función. Sólo cuando  $x_t$  es ruido blanco podemos interpretar simplemente esta función. Estas propiedades hacen que las covarianzas cruzadas no se utilicen para identificar el modelo, aunque tienen otra utilidad para su diagnosis, como veremos posteriormente.

---

### Ejemplo 17.1

Supongamos dos procesos estacionarios de media cero, relacionados por:

$$y_t = \phi y_{t-1} + \beta x_t + u_t,$$

donde  $x_t$  y  $u_t$  son ruidos blancos, mutuamente incorrelados. Este modelo puede también escribirse como:

$$y_t = \frac{\beta}{1 - \phi B} x_t + \frac{1}{1 - \phi B} u_t$$

que indica que los coeficientes de la relación entre las variables pueden escribirse:

$$\frac{\beta}{1 - \phi B} x_t = \beta(1 - \phi B)^{-1} x_t = \beta(1 + \phi B + \phi^2 B^2 + \dots) x_t$$

por lo que concluimos que  $v_k = \beta \phi^k$ .

Vamos a calcular la función de covarianzas cruzadas entre los dos procesos. Aplicando la definición de esta función, como  $E(x_t x_{t-k}) = 0$  para  $k \neq 0$ , tendremos que para  $k > 0$ :

$$\gamma_{xy}(-k) = E(x_t y_{t-k}) = E[x_t (\phi y_{t-k-1} + \beta x_{t-k} + u_{t-k})] = 0,$$

ya que  $x_t$  ni está relacionada con  $y_{t-k}$ , ni con sus  $x_{t-k}$  por ser ruido blanco, ni con los valores de  $u_t$ , al ser independientes por hipótesis. La función de covarianzas cruzadas identifica correctamente que la relación va de  $x_t$  hacia  $y_t$ .

y no al contrario, lo que es esperable al ser  $x_t$  ruido blanco. Observemos que si  $x_t$  no lo fuese, esto no es cierto, porque aunque  $x_t$  no esté relacionado ni con  $y_{t-k}$  ni con  $u_{t-k}$ , la autocovarianza  $E(x_t x_{t-k})$  no será cero.

Vamos a comprobar que, además, ahora la función  $\gamma_{xy}(k)$  nos indica la forma de la relación. Tendremos que:

$$\begin{aligned}\gamma_{xy}(0) &= E[x_t(\phi y_{t-1} + \beta x_t + u_t)] = \beta E(x_t^2) = \beta \sigma_x^2 \\ \gamma_{xy}(1) &= E[x_t(\phi y_t + \beta x_{t+1} + u_{t+1})] = \phi E(x_t y_t) = \phi \beta \sigma_x^2 \\ \gamma_{xy}(2) &= E[x_t(\phi y_{t+1} + \beta x_{t+2} + u_{t+2})] = \phi E(x_t y_{t+1}) = \phi \beta^2 \sigma_x^2 \\ &\vdots \\ \gamma_{xy}(k) &= \phi E(x_t y_{t+k-1}) = \phi \gamma_{xy}(k-1) \quad k \geq 2\end{aligned}$$

Este análisis muestra que la función de covarianzas cruzadas entre ambas variables,  $\gamma_{xy}(k)$ , tiene todos los términos positivos para  $k > 0$ , y las covarianzas seguirán la estructura:

$$\gamma_{xy}(k) = \beta \phi^k \sigma_x^2 = \sigma_x^2 v_k$$

que es proporcional a los coeficientes de la relación para el retardo  $k$ .

Concluimos que la función de covarianzas cruzadas muestra un decrecimiento AR para  $k > 0$  y es nula para  $k < 0$ , indicando que  $x_t$  influye sobre  $y_t$ , pero no al contrario, y que la relación entre  $y_t$  y  $x_t$  se amortigua en el tiempo con coeficiente  $\phi$ .

---

### 17.2.2. Función de autocorrelación cruzada

Definimos la función de correlación cruzada,  $\rho_{xy}(k)$ , de dos procesos estocásticos conjuntamente estacionarios,  $x_t$ ,  $y_t$ , por:

$$\rho_{xy} = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (k = 0, \pm 1, \pm 2 \dots)$$

Esta función es la estandarización de  $\gamma_{xy}(k)$  y, en consecuencia, tiene propiedades análogas. Se verifica que

1.  $|\rho_{xy}(k)| \leq 1$ ;
2.  $\rho_{xy}(k) = \rho_{yx}(-k)$ , pero en general  $\rho_{xy}(k) \neq \rho_{xy}(-k)$ , y la función no es simétrica alrededor del origen. Para  $k > 0$  mide la relación desde  $x$  hacia  $y$ , mientras que para  $k < 0$  la relación desde  $y$  hacia  $x$ .

La función de correlación cruzada, al ser la estandarización de la de covarianzas cruzadas, tiene los mismos problemas que ésta para identificar los retardos que existen en la relación entre dos variables: mezcla los coeficientes de la relación con los de autocorrelación.

### Ejemplo 17.2

Vamos a calcular la función de correlación cruzada para las series del ejemplo 17.1. La varianza de  $x_t$  es  $\sigma_x^2$ , y la varianza del proceso  $y_t$  se obtiene elevando al cuadrado la expresión de esta variable:

$$y_t^2 = \phi^2 y_{t-1}^2 + \beta^2 x_t^2 + u_t^2 + 2\phi y_{t-1}\beta x_t + 2\phi y_{t-1} u_t + 2u_t \beta x_t$$

y tomando esperanzas, como los dobles productos son nulos al ser  $x_t$  y  $u_t$  ruidos blancos independientes, se obtiene:

$$\gamma_y(0) = \frac{\beta^2 \sigma_x^2 + \sigma_u^2}{1 - \phi^2}.$$

Como  $\gamma_{xy}(k)$  es nula, para  $k < 0$  también lo será  $\rho_{xy}(k)$ . Para  $k = 0$ , tendremos que:

$$\rho_{xy}(0) = \frac{\beta \sigma_x^2}{\sigma_x \sqrt{\beta^2 \sigma_x^2 + \sigma_u^2}} \sqrt{1 - \phi^2} = \frac{r \sqrt{1 - \phi^2}}{\sqrt{r^2 + 1}},$$

donde  $r = \beta \sigma_x / \sigma_u$  es la relación entre la variabilidad de la respuesta producida por la variable  $x_t$  y la variabilidad debida al ruido. Este coeficiente  $r$  se conoce con el nombre de *relación señal-ruido*. Si  $r$  es grande, mucho mayor que uno, podemos utilizar que  $\sqrt{r^2 + 1} \approx r$ , con lo que la correlación  $\rho_{xy}(0)$  es  $\sqrt{1 - \phi^2}$ . Si  $\phi$  es pequeño, esta correlación será próxima a uno, pero si es grande la correlación puede ser pequeña. Si  $r$  es pequeño, la correlación será siempre pequeña.

Para retardo uno:

$$\rho_{xy}(1) = \frac{\sigma_x^2 \phi \beta}{\sigma_x \sqrt{\beta^2 \sigma_x^2 + \sigma_u^2}} \sqrt{1 - \phi^2} = \phi \rho_{xy}(0),$$

y, en general:

$$\rho_{xy}(k) = \phi \rho_{xy}(k-1) \quad k \geq 1.$$

Como la variable  $x_t$  es ruido blanco, la función de correlación cruzada describe bien el tipo de relación.

---

#### 17.2.3. Estimación de las funciones de covarianzas y correlaciones cruzadas

En el capítulo 3 vimos que las autocovarianzas teóricas de un proceso  $z_t$  se estiman mediante las muestrales, definidas por:

$$\widehat{\gamma}_z(k) = \frac{1}{T} \sum_{t=1}^{T-k} (z_t - \bar{z})(z_{t+k} - \bar{z}).$$

De la misma forma, dadas dos series temporales,  $x_t, y_t$ , estimaremos sus covarianzas cruzadas mediante las muestrales:

$$\widehat{\gamma}_{xy}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(y_{t+k} - \bar{y}),$$

y las correlaciones cruzadas muestrales por:

$$r_{xy}(k) = \frac{\widehat{\gamma}_{xy}(k)}{s_x s_y},$$

donde  $s_x = \widehat{\gamma}_x^{1/2}(0)$ ,  $s_y = \widehat{\gamma}_y^{1/2}(0)$ .

Las propiedades muestrales de las funciones de correlación cruzada son complicadas. Sin embargo, podemos utilizar esta función para contrastar la causalidad unidireccional de manera simple en el caso particular en que uno de los dos procesos es *ruido blanco*. En este caso particular, puede demostrarse (véase Bartlett (1973)) que si ambos procesos están incorrelados, las varianzas de los coeficientes de autocorrelación cruzada muestrales entre un proceso  $y_t$  y un ruido blanco  $a_t$  son:

$$Var[r_{ay}(k)] \simeq \frac{1}{T-k}. \quad (17.5)$$

Es decir, podemos aproximar las desviaciones típicas de las estimaciones de los coeficientes de correlación cruzada de manera idéntica a las de las autocorrelaciones mediante  $1/\sqrt{T}$ . Por tanto, si  $x_t$  es ruido blanco podemos construir bandas  $2/\sqrt{T}$  sobre las autocorrelaciones y si  $r_{ay}(k) = 0$  para  $k < 0$  concluimos que no hay relación desde  $y_t$  hacia  $x_{t+k}$ . La parte de esta función para  $k > 0$  nos permitirá investigar si existe o no relación desde  $x_{t-k}$  hacia  $y_t$ . Si la variable  $x_t$  no es ruido blanco, este contraste puede aplicarse utilizando los residuos del modelo univariante de  $x_t$ , como se explica en el apéndice 17.3.

Para interpretar la función de correlación cruzada hay que tener en cuenta que las estimaciones para distintos retardos están correladas, y para valores moderados de  $k$  las covarianzas entre los coeficientes de correlación muestrales  $r_{ya}(k)$  y  $r_{ya}(k+h)$  pueden calcularse aproximadamente por:

$$Cov[r_{ya}(k)r_{ya}(k+h)] \simeq \frac{\rho_y(h)}{(T-k)},$$

y utilizando (17.5) se obtiene que:

$$\rho[r_{xa}(k)r_{xa}(k+h)] \simeq \rho_x(h), \quad (17.6)$$

es decir, aunque  $y_t$  y  $a_t$  no estén correlados entre sí (por tanto,  $\rho_{ya}(k) \equiv 0$ ), los coeficientes estimados  $r_{ya}(k)$  sí lo estarán. La función de correlación cruzada estimada variará alrededor de cero con varianza dada por (17.5) pero, según (17.6), al estar sus coeficientes relacionados de la misma forma que las autocorrelaciones de  $y_t$  tenderá a reproducir el correlograma de  $y_t$ .

Esta propiedad hace que la función de correlación cruzada no sea fácil de interpretar, incluso en el caso más simple donde una de las series es ruido blanco y existe independencia.

### 17.3. El modelo dinámico entre dos series estacionarias

#### 17.3.1. Representación general

Los modelos que se estudian en este capítulo suponen que existe relación unidireccional de la serie  $x$  hacia la serie  $y$  pero no al contrario. En el capítulo 19 consideraremos el caso más general donde es posible la realimentación. Una relación lineal unidireccional de  $x_t$  hacia  $y_t$  entre dos procesos estocásticos estacionarios de media cero puede representarse, como vimos en el capítulo 12, mediante:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + n_t \quad (17.7)$$

donde los coeficientes  $v_i$  de la *función de transferencia* describen la relación dinámica entre las dos series y el proceso  $n_t$ , que es también estacionario, recoge el efecto de todas las otras variables que pueden tener efecto sobre  $y_t$ , y se denomina el *proceso de perturbación* o de *inercia*, o simplemente *perturbación*, de la relación. Esta ecuación puede escribirse de forma más compacta utilizando el operador de retardo:

$$y_t = v(B)x_t + n_t \quad (17.8)$$

donde

$$v(B) = v_0 + v_1 B + v_2 B^2 + \dots$$

se denomina la *función de transferencia* y a los coeficientes  $v_i$  se les denomina función de respuesta a impulsos, como vimos en el capítulo 12.

La ganancia de la función de transferencia es

$$g = v(1) = \sum_{i=0}^{\infty} v_i$$

y representa el efecto a largo plazo que experimenta  $y_t$  cuando  $x_t$  aumenta en una unidad y permanece constante a continuación.

El proceso de perturbación,  $n_t$ , seguirá un proceso ARMA estacionario:

$$\phi(B)n_t = \theta(B)a_t, \quad (17.9)$$

donde  $a_t$  es un proceso de ruido blanco. El modelo formado por las ecuaciones (17.7) y (17.9) se conoce como modelo de regresión dinámica o modelo de función de transferencia entre las dos variables.

### 17.3.2. Representación de la función de transferencia como cociente de polinomios

La expresión (17.7) de la función de transferencia no es útil para ser estimada en la práctica. Si el proceso  $x_t$  tiene alta autocorrelación, las variables  $x_t, x_{t-1}, x_{t-2}, \dots$  estarán muy correladas entre sí y será muy difícil separar sus efectos por el problema de multicolinealidad. Además, si la relación es larga tendremos que estimar muchos parámetros, con lo que el problema se complica.

Una forma alternativa de escribir la función de transferencia, que ya estudiamos en el capítulo 12, es utilizar la idea de aproximar una función infinita como cociente de polinomios y, como hicimos en los modelos ARMA, representar la función mediante:

$$v(B) = \frac{w_m(B)}{\delta_a(B)} B^b,$$

donde el término  $B^b$  tiene en cuenta que la relación puede establecerse con un retardo inicial  $b$  y:

$$w_m(B) = w_0 + w_1 B + \dots + w_m B^m,$$

es el numerador de la función de transferencia que juega un papel similar a la parte MA de un modelo ARMA, y:

$$\delta_a(B) = 1 - \delta_1 B - \dots - \delta_a B^a,$$

es el denominador, que juega el papel de la parte AR. Los factores de esta ecuación deben ser menores que la unidad para que el sistema pro-

## 17. Regresión dinámica entre variables estacionarias

porcione una respuesta finita a una modificación finita de la  $x_t$ . Es decir, si escribimos  $\delta_a(B) = (1 - \alpha_1 B) \dots (1 - \alpha_a B)$ , entonces  $|\alpha_i| < 1$ .

En la práctica, los órdenes  $m$  y  $a$  son pequeños, no mayores que 3. La ganancia de la función de transferencia se calcula como antes:

$$g = v(1) = \frac{w_m(1)}{\delta_a(1)} = \frac{\sum_{j=0}^m w_j}{1 - \sum_{i=1}^a \delta_i}$$

Una ventaja de esta representación es separar en la función de transferencia el efecto inicial de la pauta de decrecimiento, y al ser estos dos factores independientes podemos estimarlos mejor. Consideremos el caso más simple:

$$v(B) = \frac{w_0}{(1 - \delta B)} = w_0(1 + \delta B + \delta^2 B^2 + \dots) \quad (17.10)$$

donde  $|\delta| < 1$ . Esta función representa una relación donde si  $x_t$  aumenta en una unidad y luego vuelve a su valor inicial, la respuesta tiene un aumento instantáneo de  $w_0$ , seguido de un aumento en el siguiente periodo de  $w_0\delta$ , y en los periodos siguientes  $w_0\delta^k$ , con decrecimiento con tasa  $|\delta| < 1$ , hasta acumular el efecto total que es la ganancia  $g = w_0/(1 - \delta)$ . Por ejemplo, si  $\delta = .7$  y  $w_0 = 2$ , un impulso en  $x_t$ , es decir un incremento en  $t$  de una unidad que desaparece en el instante  $t+1$ , produce aumentos en  $y_t$  de 2, en  $y_{t+1}$  de  $2(.7) = 1.4$ , en  $y_{t+2}$  de .98 y en  $y_{t+k}$  de  $2(.7)^k$ . Si sumamos todos estos efectos, obtenemos la ganancia,  $2/.3 = 6.667$ .

Si la verdadera relación es del tipo (17.10) con  $\delta$  no muy pequeño, con la representación (17.7) tendríamos que estimar muchos coeficientes  $v_i$  y además muy dependientes entre sí, con lo que la estimación sería muy imprecisa. Sin embargo, con la representación (17.10) sólo necesitamos estimar dos parámetros y además independientes, con lo que podremos hacerlo con precisión.

Otro modelo muy útil en la práctica es:

$$v(B) = \frac{w_0 + w_1 B}{(1 - \delta B)} = w_0 + (w_1 + w_0\delta)B + \delta(w_1 + w_0\delta)B^2 + \dots \quad (17.11)$$

donde permitimos que el efecto inicial sea distinto del que aparece en el decrecimiento gobernado por  $\delta$ . Para otras respuestas recomendamos al lector releer la formulación de la función de respuesta a impulsos que analizamos con detalle en el capítulo 12.

### 17.3.3. Representación de la función de transferencia como efectos a corto y a largo plazo

Un método alternativo de escribir la función de transferencia es separando la respuesta a largo plazo, o ganancia, de los efectos a corto plazo. Se demuestra en el apéndice 17.4 que toda función de transferencia de orden máximo  $k$  puede escribirse como:

$$v_k(B) = g + \sum_{i=0}^{k-1} \alpha_i \nabla B^i. \quad (17.12)$$

Utilizando esta representación, en lugar de la ecuación (17.8), podemos escribir:

$$y_t = gx_t + \sum_{i=0}^{k-1} \alpha_i \nabla x_{t-i} + n_t, \quad (17.13)$$

que separa el efecto a largo plazo, medido por  $g = v(1)$ , de los efectos a corto, medidos por los coeficientes  $\alpha_i$ . Observemos que, en ambos casos, tenemos  $k+1$  coeficientes. Los  $v_i$  ( $i = 0, \dots, k$ ) en la izquierda y en la derecha los  $k$  coeficientes  $\alpha_i$  ( $i = 0, \dots, k-1$ ) más el  $g$ . La ventaja de la formulación (17.13) frente a (17.8) es que las variables  $\nabla x_{t-i}$  estarán mucho menos correladas entre sí que las  $x_t$ , con lo que sus efectos se estimarán mejor. Por ejemplo, supongamos que  $x_t$  sigue un proceso AR(1) con alto coeficiente. Entonces  $\nabla x_t$  es casi ruido blanco y los términos  $\nabla x_{t-i}$  estarán muy poco correlados entre sí y con  $x_t$ . La estimación de (17.13) será mucho más eficiente que la de (17.8).

Otra forma alternativa de separar los efectos a corto y a largo es escribiendo la relación como:

$$y_t = gx_t + \sum_{i=1}^k \beta_i \nabla^i x_t + n_t \quad (17.14)$$

donde, de nuevo, hemos separado el efecto a largo plazo,  $g$ , de las fluctuaciones a más corto plazo dadas por las diferencias  $\nabla^i x_t$ . Se demuestra en el apéndice 17.4 que, como en el caso anterior, cualquier polinomio en el operador de retardo de orden  $k$  puede escribirse como

$$v_k(B) = v(1) + \sum_{i=1}^k \beta_i \nabla^i. \quad (17.15)$$

Esta expresión indica que en la ecuación (17.14) el coeficiente de  $x_t$  es la ganancia, y los coeficientes de  $\nabla^i x_t$  representan efectos a corto plazo. Los coeficientes  $\beta_i$  representan los efectos sobre la respuesta de los incremen-

tos de  $x_t$ , o velocidad de cambio de  $x_t$ , de la aceleración en el cambio de  $x_t$  (medida por  $\nabla^2 x_{t-1}$ ), etc. Es fácil comprobar que si  $x_t$  tiene alta autocorrelación las variables  $x_t, \nabla x_{t-1}, \nabla^2 x_{t-1}$ , tienen menos autocorrelación que las variables  $x_t, x_{t-1}, x_{t-2}$ . Por tanto, la ecuación (17.14) resulta mucho más conveniente que la (17.8) cuando la serie  $x_t$  tenga alta autocorrelación. Por ejemplo, se demuestra en el ejercicio 17.2 que si la autocorrelación de primer orden de  $x_t$  es mayor que .5, las variables  $x_t$ , y  $\nabla x_{t-1}$  tienen menos correlación que las variables  $x_t$  y  $x_{t-1}$ .

Las dos representaciones (17.13) y (17.14) tienen el mismo objetivo: separar los efectos a corto de los efectos a largo y escribir la relación utilizando variables explicativas tan poco correladas como sea posible para estimar mejor los efectos. Cuál de estas dos representaciones sea más eficaz depende de la estructura de la variable  $x_t$ .

#### 17.4. Construcción de modelos de regresión dinámica entre variables estacionarias

Las etapas para el ajuste de un modelo de regresión dinámica entre dos variables estacionarias son las mismas que para un modelo ARIMA: identificación, estimación y diagnosis. Vamos a analizarlas a continuación suponiendo que la variable  $y_t$  sigue un proceso ARMA ( $p_y, q_y$ ) y que la variable  $x_t$  sigue un proceso ARMA ( $p_x, q_x$ ).

##### 17.4.1. Identificación

La identificación supone decidir la estructura de la función de transferencia y del modelo para la perturbación, que será ARMA ( $p_n, q_n$ ).

Podemos utilizar como modelo de la perturbación el de la variable  $y_t$ , suponiendo  $p_n = p_y$  y  $q_n = q_y$ . La justificación de esta elección es que, si no hay relación entre las variables y los coeficientes de la función de transferencia son cero, entonces  $y_t$  seguirá el proceso de la perturbación. En general, el modelo final para  $n_t$  será más simple que el de  $y_t$ , ya que es muy posible que  $x_t$  explique parte de la dinámica de la respuesta, y una posibilidad alternativa es tomar un modelo  $AR(1) \times AR(1)_s$  para captar la autocorrelación regular y la estacional.

El método para identificar la función de transferencia depende del grado de autocorrelación en la variable  $x_t$ . Si esta variable sigue un proceso de ruido blanco o tiene una estructura de correlación muy pequeña, podemos estimar el modelo:

$$y_t = \beta_0 + \sum_{i=0}^k v_i x_{t-i} + n_t, \quad (17.16)$$

y estudiar la estructura de los coeficientes  $v_1, \dots, v_k$  para decidir el orden del numerador y del denominador de la función de transferencia. El orden  $k$  se elige teniendo en cuenta el tamaño muestral y procurando incluir todos los retardos posibles donde pueden existir efectos. En general  $k < T/4$ . Por ejemplo, con datos anuales  $k$  será pequeño,  $k \leq 2$ , pero con datos horarios puede ser alto. Pondremos siempre en la ecuación un modelo con autocorrelación para  $n_t$  para evitar los problemas de regresión espuria que comentamos después.

Esta formulación de la función de transferencia, (17.16), no es conveniente cuando la serie  $x_t$  tenga fuerte autocorrelación, ya que entonces los coeficientes  $v_i$  serán difíciles de estimar por la alta multicolinealidad entre las variables explicativas. Es mejor estimar directamente modelos del tipo:

$$y_t = \beta_0 + \frac{w_0 + w_1 B}{(1 - \delta B)} B^b x_t + n_t, \quad (17.17)$$

y comprobar si encontramos coeficientes significativos que nos indican la forma de la respuesta. Si la correlación de la  $x_t$  es fuerte, una solución simple es estimar un modelo que nos separe los efectos a largo y a corto, del tipo (17.13) y (17.14). Con los coeficientes significativos determinados qué forma de la función de transferencia parece más adecuada y, si es posible, la escribimos como cociente de polinomios, como indica la ecuación (17.17).

En esta fase de exploración no conviene seguir reglas estrictas al eliminar coeficientes, y mantener aquellos que sean grandes aunque no sean formalmente significativos. Si se estiman varios coeficientes ( $k > 2$ ) siempre existirá cierta multicolinealidad en la estimación y además los errores estándar que utilizamos sólo serán aproximados, ya que el modelo es también una aproximación al correcto. Por tanto, recomendamos mantener todos los coeficientes con valores de  $t$  mayores que uno. Escribiendo el modelo estimado en la forma  $v(B)$  podemos decidir la estructura de la función de transferencia de forma similar a como se hizo con el correlograma o *fas* de un proceso ARMA. En concreto:

- a) Si los primeros  $b$  coeficientes  $\hat{v}_i$  parecen iguales a cero, entonces la función empieza en el retardo  $b$ .
- b) Si hay  $m$  coeficientes aparentemente significativos sin una pauta característica de variación, tomar ese grado para el numerador de la función.
- c) Si se observa una estructura de decrecimiento geométrico en los coeficientes, incluir un término de orden uno en el denominador,  $a = 1$ . Si el decrecimiento es sinusoidal suponer  $a = 2$ .

De esta manera se establece una forma provisional para la función:

$$\frac{w_m(B)}{\delta_a(B)} B^b$$

que pasaremos a estimar eficientemente en la etapa siguiente.

Una alternativa propuesta originalmente por Box y Jenkins es el preblanqueo que se expone en el apéndice 17.1. Este método no se utiliza mucho actualmente, porque es más rápido estimar directamente el modelo y, además, el método del preblanqueo no puede extenderse fácilmente para muchas variables.

#### 17.4.2. Estimación y selección del modelo

Una vez establecida la forma de la función de transferencia y del modelo para la perturbación y en la hipótesis de normalidad, la función de verosimilitud puede escribirse siguiendo el mismo procedimiento que en el caso univariante. Por ejemplo, si el modelo para el ruido es  $AR(p)$  y en la función de transferencia  $a = 0$ , llamando  $h = \max(p, m + b)$ ,  $\mathbf{y} = (y_1, \dots, y_T)'$ ,  $\mathbf{y}_h = (y_1, \dots, y_h)$ , el logaritmo de la verosimilitud o función soporte puede escribirse como:

$$L(\mathbf{y}) = \sum_{t=h+1}^T L(y_t | y_{t-1}, \dots, y_1) + L(\mathbf{y}_h).$$

Prescindiendo de constantes y despreciando el segundo término, la función soporte condicional para el vector de parámetros  $\beta$ , que incluye todos los del modelo, será:

$$L(\beta | \mathbf{y}_h) = -\frac{T-h}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=h+1}^T a_t^2$$

donde  $a_t$  son los residuos. La estimación condicional equivale, como en el caso univariante, a la minimización de la suma de cuadrados de los residuos.

La minimización de  $\sum a_t^2$  puede plantearse como un problema de mínimos cuadrados generalizados. Para ilustrar el procedimiento vamos a suponer el caso más simple:

$$y_t = \beta_0 + v_0 x_t + n_t, \quad (17.18)$$

donde la perturbación sigue el proceso  $AR(1)$

$$n_t = \rho n_{t-1} + a_t,$$

y  $a_t$  es ruido blanco. Si conociésemos el valor de  $\rho$ , multiplicando (17.18) por el operador  $(1 - \rho B)$ , tendríamos que:

$$(1 - \rho B)y_t = (1 - \rho)\beta_0 + v_0(1 - \rho B)x_t + a_t,$$

y definiendo las nuevas variables,  $\tilde{y}_t = y_t - \rho y_{t-1}$ ,  $\tilde{x}_t = x_t - \rho x_{t-1}$ , podríamos aplicar mínimos cuadrados a la regresión:

$$\tilde{y}_t = c + v_0 \tilde{x}_t + a_t$$

entre  $\tilde{y}_t$  y  $\tilde{x}_t$  y obtener  $v_0$ . Esto equivale a aplicar a las variables la transformación:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ \vdots & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}$$

que, como se justifica en el apéndice 17.2, es aproximadamente el resultado de aplicar el método de mínimos cuadrados generalizados.

En general, el modelo del proceso de inercia es desconocido, pero el procedimiento anterior puede aplicarse iterativamente: (1) estimamos por mínimos cuadrados una regresión entre ambas variables y obtenemos los residuos  $\hat{n}_t = y_t - \hat{\beta}_0 + \hat{v}_0 x_t$ ; (2) con estos residuos estimamos el AR(1),  $\hat{n}_t = \hat{\rho} \hat{n}_{t-1}$ ; (3) con el coeficiente estimado transformamos las variables en  $\tilde{y}_t = y_t - \hat{\rho} y_{t-1}$  y  $\tilde{x}_t = x_t - \hat{\rho} x_{t-1}$ , y con las nuevas variables repetimos (1), (2) y (3) hasta que los residuos de la regresión sean ruido blanco.

Este proceso puede generalizarse como sigue. Supongamos el modelo:

$$y_t = v(B)x_t + \frac{1}{\pi(B)}a_t, \quad (17.19)$$

entonces la estimación puede llevarse a cabo en las siguientes etapas:

1. Estimar una ecuación entre  $y_t$  y  $x_t$ , obtener los residuos,  $n_t$ , y estimar el modelo  $\hat{n}_t^{(0)} = \hat{\pi}^{(0)}(B)a_t^{(0)}$  con estos residuos. Alternativamente, podemos suponer como modelo para  $n_t$  el modelo univariante de  $y_t$ . Tomar  $i = 0$ , y definir  $y_t^{(0)} = y_t$ ,  $x_t^{(0)} = x_t$ , y  $\hat{n}_t^{(0)} = \hat{\pi}^{(0)}(B)a_t^{(0)}$ , que es el modelo inicial para la perturbación.
2. Hacer  $i = i + 1$ . Aplicar el modelo  $\hat{\pi}^{(i-1)}(B)$  a la ecuación (17.19) para que tenga residuos incorrelados, mediante:

$$\hat{\pi}^{(i-1)}(B)y_t = v(B)\hat{\pi}^{(i-1)}(B)x_t + [\hat{\pi}^{(i-1)}(B)/\pi(B)]a_t$$

Llamar:

$$y_t^{(i)} = \widehat{\pi}^{(i-1)}(B)y_t^{(i-1)} = y_t^{(i-1)} - \widehat{\pi}_1^{(i-1)}y_{t-1}^{(i-1)} - \dots - \widehat{\pi}_{k_i}^{(i-1)}y_{t-k_i}^{(i-1)},$$

y,  $x_t^{(i)} = \widehat{\pi}^{(i-1)}(B)x_t^{(i-1)}$ . Suponer que  $[\widehat{\pi}^{(i-1)}(B)/\pi(B)] \approx 1$  y estimar los parámetros de la función de transferencia  $v(B)$  por regresión entre  $y_t^{(i)}$  y  $x_t^{(i)}$  por mínimos cuadrados. Sea  $\widehat{y}_t^{(i)} = \widehat{v}^{(i)}(B)x_t^{(i)}$  la regresión estimada. Con estos valores obtener el proceso de inercia  $\widehat{n}_t^{(i)} = y_t^{(i)} - \widehat{v}^{(i)}(B)x_t^{(i)}$ .

3. Estimar los parámetros de la nueva serie del proceso de inercia. Sea  $\widehat{n}_t^{(i)} = \widehat{\pi}^{(i)}(B)a_t^{(i)}$  el modelo estimado. Si el modelo  $\widehat{\pi}^{(i)}(B)$  es prácticamente el mismo obtenido para  $\widehat{\pi}^{(i-1)}(B)$ , finalizar la estimación. En otro caso, hacer  $i = i + 1$  e ir a (2).

El algoritmo itera las dos etapas (2) y (3) hasta convergencia. Si en la etapa (2) los parámetros de la función de transferencia no pueden obtenerse mediante regresión, porque tenemos términos en el denominador, esta etapa se sustituye por:

- 2' Calcular recursivamente los residuos del modelo,  $\widehat{a}_t$ , obteniendo primero el proceso de ruido y, a continuación, sus residuos. Por ejemplo, si  $b = 0$ ,  $m = 0$ ,  $a = 1$ , y  $p = 1$ ,  $q = 0$ , y tenemos el modelo

$$y_t = \frac{w_0}{(1 - \delta B)}x_t + \frac{1}{(1 - \phi B)}a_t$$

entonces

$$a_t = y_t - \phi y_{t-1} + w_0 x_t + (\delta + \phi)x_{t-1} + (\delta^2 + \phi^2 + \delta\phi)x_{t-2} + \dots$$

que permite obtener los residuos a partir de unos valores iniciales. Sea  $\beta$  el vector de parámetros del modelo y  $\widehat{\beta}_0$  el valor inicial, utilizado en los cálculos anteriores. Sea

$$S(\beta) = \sum \widehat{a}_t^2(\beta)$$

y  $S(\widehat{\beta}_0)$  el valor inicial de esta función. Calcularemos un nuevo valor de los parámetros que disminuya la suma de cuadrados mediante un algoritmo tipo Gauss-Newton e iteraremos hasta obtener convergencia.

Como en el caso univariante, las varianzas y covarianzas de los estimadores se obtendrán a partir de la matriz de segundas derivadas en

el máximo. La maximización de la función de verosimilitud exacta se efectúa con estos mismos principios. Normalmente se estiman varios modelos modificando los parámetros y se elige el más adecuado con el criterio BIC.

La estimación eficiente de un modelo de función de transferencia puede hacerse con el filtro de Kalman. Véase Gómez y Maravall (1994) para una buena implantación, que es la que se utiliza en el programa TSW.

La elección entre modelos alternativos puede hacerse con los criterios de selección presentados en el capítulo 11. La expresión es la misma y sólo hay que tener en cuenta la varianza residual del modelo y el número de parámetros estimados.

### Ejemplo 17.3

Como ilustración vamos a identificar y estimar la relación entre los rendimientos de la Bolsa de Madrid y los de París. Veremos si podemos utilizar la información de los rendimientos en la Bolsa de París para predecir lo que ocurrirá en la de Madrid. Los índices de bolsa que utilizaremos se encuentran en el fichero bolsamundo.dat, y los rendimientos se obtienen del índice mediante la transformación  $r_t = \nabla \log I_t$ . Como las series de rendimientos tienen baja autocorrelación, son prácticamente ruido blanco, estimamos directamente la relación con la forma (17.16). El modelo estimado con EViews para la relación entre los rendimientos, suponiendo un AR(1) para la perturbación y fijando  $k = 6$ , aparece en la tabla 17.1. Vemos que sólo la relación instantánea y el primer retardo parecen significativos. Los coeficientes de los retardos -1 y -2 tienen valores de los estadísticos t mayores que uno y el signo correcto. El bajo coeficiente estimado para el AR(1) indica que la perturbación es ruido blanco en este caso. Reestimando el modelo y eliminando las variables no significativas se obtiene finalmente:

$$rm_t = 0.71 rp_t + 0.17 rp_{t-1} + a_t, \quad (11.9)$$

donde  $rm_t$  son los rendimientos en Madrid y  $rp_t$  los de París. Esta ecuación indica que una subida del rendimiento de la Bolsa de París en este mes de un 1% produce una subida en Madrid del .88 %. En el mismo mes la subida es del .71 % y en el mes siguiente de .17 %.

La figura 17.1 presenta la función de correlación cruzada entre la Bolsa de Madrid y la de París. Se observa en el gráfico que la relación es sobre todo instantánea, pero que aparece un pequeño coeficiente en el retardo uno en ambas direcciones. Esto sugiere que podríamos estimar un modelo equivalente entre los rendimientos en París y en Madrid, ya que la relación es sobre todo

**Tabla 17.1** Modelo estimado con EViews para la relación entre los rendimientos suponiendo un AR(1) para la perturbación y fijando  $k = 6$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000967	0.003665	0.263695	0.7924
RP	0.766655	0.062977	12.17358	0.0000
RP(-1)	0.115646	0.063064	1.833795	0.0688
RP(-2)	0.067236	0.063988	1.050762	0.2952
RP(-3)	-0.081678	0.063877	-1.278684	0.2031
RP(-4)	-0.028350	0.064013	-0.442878	0.6585
RP(-5)	-0.094768	0.062993	-1.504419	0.1347
RP(-6)	-0.054714	0.062002	-0.882459	0.3791
AR(1)	-0.075803	0.084752	-0.894401	0.3727

contemporánea. Al estimarlo, se obtiene el modelo:

$$rp_t = 0.66 rm_t + 0.01 rm_{t-1} + a_t. \quad (11.6) \quad (0.22)$$

Esta estimación confirma que la relación entre estas dos series es desde París hacia Madrid, y no al contrario, ya que además de la relación instantánea se encuentran efectos significativos del valor actual de la Bolsa de París sobre el valor futuro de la Bolsa de Madrid, pero no al contrario. Si existiese realimentación, es decir, encontrásemos que los valores actuales de Madrid influyesen sobre los futuros de París, tendríamos que acudir a los modelos que estudiaremos en el capítulo 19.

#### 17.4.3. Contrastes diagnósticos

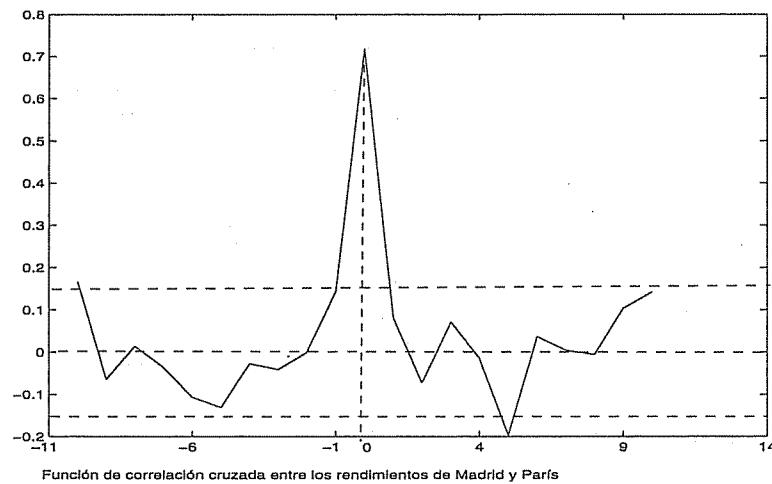
Una vez seleccionado un modelo, antes de aceptarlo conviene efectuar contrastes diagnósticos para comprobar que no presenta deficiencias detectables que pueden sugerir reformulaciones. Clasificaremos los contrastes que pueden aplicarse en: (a) sobre los parámetros estimados, y (b) sobre los residuos.

##### 17.4.3.1. Contrastos sobre los parámetros

Debe comprobarse que la ganancia concuerda con el comportamiento esperado del sistema. Conviene también examinar:

- Si el modelo puede simplificarse eliminando operadores con valores próximos en el numerador y en el denominador.

**Figura 17.1 Función de correlación cruzada entre los rendimientos de Madrid y París**



2. Si las raíces de los polinomios autorregresivos cumplen las condiciones de estabilidad o estacionaridad.
3. Si los efectos de los distintos operadores del ruido y de la función de transferencia ofrecen una interpretación coherente con el comportamiento cualitativo del sistema que estamos modelando.

Comenzaremos comprobando que no tenemos raíces comunes que pueden simplificarse, o en un cociente de polinomios o en ambos miembros de la ecuación. Por ejemplo, el modelo:

$$y_t = \frac{(5 - 2.8B)}{1 - .6B} x_t + \frac{1 - .3B}{1 - 1.1B + .2B^2} a_t,$$

puede simplificarse, cancelando raíces muy próximas, como  $(1 - .56B)$  en el numerador y  $(1 - .6B)$  en el denominador de la función de transferencia y  $(1 - .3B)$  y  $(1 - .31B)$  en el ruido, a:

$$y_t = 5x_t + \frac{1}{1 - .8B} a_t.$$

La presencia de raíces comunes o próximas puede a veces mostrarse por correlaciones altas entre los parámetros estimados, lo que sugeriría buscar una forma de parametrizar el modelo para separar mejor los distintos efectos.

## 17. Regresión dinámica entre variables estacionarias

Comprobaremos también que todos los parámetros incluidos son significativamente distintos de cero, comparando el valor obtenido con la desviación estándar estimada. Hay que recordar que estos contrastes son sólo válidos asintóticamente, por lo que conviene fijar el nivel de significación en función de la interpretación lógica del modelo, especialmente para los parámetros de la función de transferencia: si su signo y magnitud coinciden con lo esperado conviene mantenerlos, aunque el valor de su estadístico  $t$  asociado no sea exactamente significativo al 95 %.

### 17.4.3.2. Contrastos sobre los residuos

Consisten, como siempre, en comprobar si los residuos son una secuencia de variables aleatorias normales con varianza constante e independientes. Esto supone realizar contrastes para confirmar:

1. Normalidad.
2. Que su media es cero.
3. Que están incorrelados entre sí.

Estos contrastes son análogos al caso univariante.

Para diseñar contrastes de errores de especificación, veamos las consecuencias de una especificación incorrecta. Supongamos que el modelo verdadero es:

$$y_t = v(B)x_t + \psi(B)a_t,$$

pero erróneamente estimamos:

$$y_t = \hat{v}_1(B)x_t + \hat{\psi}_1(B)\hat{a}_t.$$

Entonces, igualando las dos expresiones:

$$v(B)x_t + \psi(B)a_t = \hat{v}_1(B)x_t + \hat{\psi}_1(B)\hat{a}_t$$

concluimos que los residuos estimados  $\hat{a}_t$  verificarán el modelo:

$$\hat{a}_t = \frac{v(B) - \hat{v}_1(B)}{\hat{\psi}_1(B)}x_t + \frac{\psi(B)}{\hat{\psi}_1(B)}a_t.$$

En esta expresión podemos encontrarnos en los casos siguientes:

1. Hemos especificado mal tanto la función de transferencia como el modelo para la perturbación, es decir,  $\psi(B) \neq \hat{\psi}_1(B)$  y  $v(B) \neq \hat{v}_1(B)$ . Entonces, los residuos  $\hat{a}_t$  estarán tanto autocorrelados entre sí, ya que tienen estructura dinámica, como correlados con los

valores de  $x_t$ . La autocorrelación de los residuos se detectará con su correlograma y la dependencia con la variable  $x_t$  estudiando la función de correlación cruzada entre los residuos del modelo y la variable explicativa.

2. Si la función de transferencia es incorrecta, aunque el modelo del ruido sea adecuado, también observaremos correlación cruzada entre los residuos y  $x_t$  y autocorrelación entre los residuos. En efecto, si la función de transferencia no es correcta pero está bien especificado el modelo del ruido, los residuos pueden escribirse como

$$\hat{a}_t = \frac{v(B) - \hat{v}_1(B)}{\hat{\psi}_1(B)} x_t + a_t^*,$$

y esta ecuación implica autocorrelación entre los residuos estimados  $\hat{a}_t$ .

3. Si la función de transferencia es correcta, es decir aproximadamente  $v(B) = \hat{v}_1(B)$ , pero el modelo del proceso de perturbación o de inercia es equivocado,  $\psi(B) \neq \hat{\psi}_1(B)$ , habrá autocorrelación en los residuos  $\hat{a}_t$ , pero no existirá correlación cruzada entre  $\hat{a}_t$  y  $x_t$ .
4. Si tanto la función de transferencia como el modelo del proceso de inercia son correctos, no habrá ni autocorrelación, ni correlación cruzada.

En resumen, la función de correlación cruzada entre los residuos estimados  $\hat{a}_t$  y la variable explicativa  $x_t$  es muy útil para detectar posibles errores en el modelo. Si encontramos correlaciones cruzadas entre los residuos estimados y la  $x_t$  esto sugiere que la función de transferencia no está bien especificada. Además, en este caso es posible que necesitemos también reformular el modelo del ruido. Si no hay correlaciones cruzadas entre los residuos estimados y la variable  $x_t$  concluimos que la función de transferencia es correcta y miraremos el correlograma de los residuos. Si hay autocorrelación, el modelo de la perturbación debe modificarse. En otro caso, el modelo es correcto.

Dado que la función de correlación entre  $\hat{a}_t$  y  $x_t$  puede no ser muy clara debido a la propia estructura de dependencia de  $x_t$ , como ya comentamos en el caso general, es más operativo estudiar la correlación entre  $\hat{a}_t$  y  $a_t$ , siendo  $a_t$  los residuos del modelo univariante de  $x_t$ , ya que, al ser el proceso  $a_t$  ruido blanco, la imagen será mucho más clara.

---

**Ejemplo 17.4**

La figura 17.2 presenta los residuos de la relación entre las Bolsas de París y Madrid que estudiamos en el ejemplo 17.3. Aunque la media de la serie es cero, parece que los primeros años, 1989-1993, la media es más baja que en los siguientes, lo que sugeriría un cambio de nivel que podría investigarse. Ni el correlograma de los residuos ni el de los residuos al cuadrado muestran coeficientes significativos ni evidencia de errores de especificación. La segunda parte de esta figura muestra la función de correlación cruzada entre estos residuos y los rendimientos de París y se observa que ninguna correlación es significativa. La figura 17.3 muestra que el histograma de los residuos es simétrico y con curtosis próxima a 3, con lo que podemos considerar que la distribución es aproximadamente normal.

**Ejemplo 17.5**

Vamos a estimar la relación entre las nuevas matriculaciones de vehículos y el consumo de gasolina. Suponemos que nuestro interés se centra en relacionar las variables estacionarias,  $g_t = \nabla \nabla_{12} \log G_t$  y  $mv_t = \nabla \nabla_{12} \log M_t$ , donde  $G_t$  es la serie de consumo de gasolina y  $M_t$  la de matriculación de vehículos. En el capítulo siguiente veremos cómo relacionar las variables  $G_t$  y  $M_t$  directamente. Las series  $g_t$  y  $mv_t$  son las tasas de crecimiento teniendo en cuenta las estacionalizadas. Estimamos el modelo:

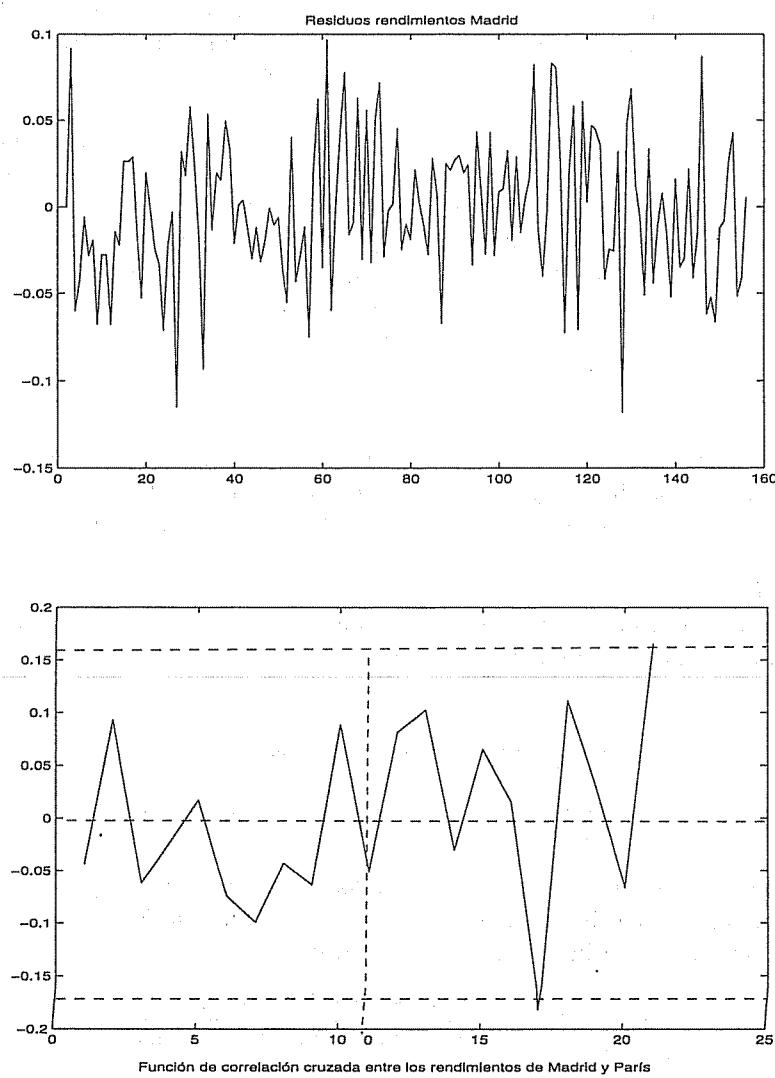
$$g_t = 0.06mv_t - 0.03\nabla mv_t + 0.03\nabla mv_{t-1} + \\ (1.4) \quad (-1.2) \quad (0.9) \\ + 0.002\nabla mv_{t-2} + (1 - 0.63B)(1 - 0.31B^{12})a_t, \\ (0.1)$$

al ir eliminando sucesivamente los retardos, la relación contemporánea va aumentando. El análisis de los residuos muestra estructura, lo que lleva a reformular su estructura hasta obtener el modelo:

$$g_t = 0.053mv_t + \frac{(1 - 0.64B^{12} - 0.14B^{24})}{(1 + 0.86B + 0.52B^2 + 0.28B^3)}a_t,$$

que establece una relación significativa e instantánea y una estructura bastante complicada para el modelo del ruido.

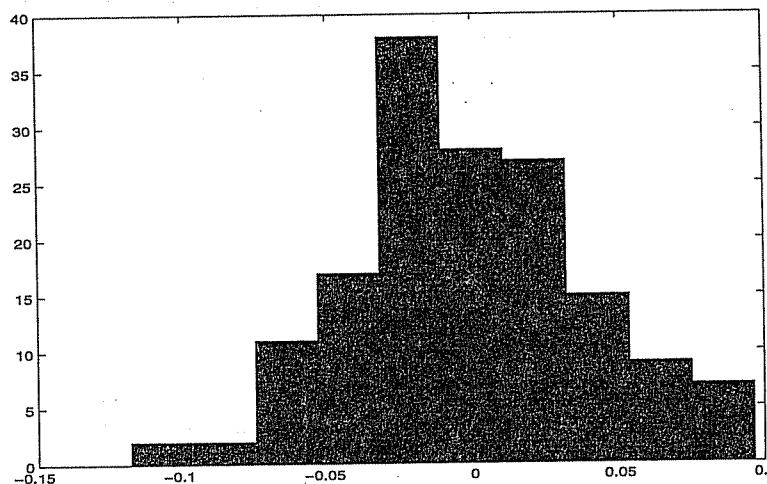
**Figura 17.2 Residuos del modelo de los rendimientos de la Bolsa de Madrid y su función de correlación cruzada con los rendimientos de París**



## 17.5. Relaciones espurias entre variables dinámicas

La metodología presentada para construir modelos de regresión entre variables temporales evita el problema que se conoce como regresión espuria y que consiste en que al relacionar con un modelo de regresión

Figura 17.3 Histograma de los residuos



clásico series temporales independientes podemos encontrar fácilmente relaciones significativas. Vamos a ilustrar este problema.

#### 17.5.1. Relaciones sin dinámica

Consideremos el caso más simple, aunque poco frecuente con series temporales, de un proceso de perturbación,  $n_t = a_t$ , que sea ruido blanco normal con  $E(a_t^2) = \sigma_a^2$  y, por tanto, sin dinámica, y una relación simple del tipo:

$$y_t = \beta_0 + \beta_1 x_t + a_t,$$

donde la serie  $x_t$  es estacionaria. Los parámetros,  $\beta_0$ ,  $\beta_1$  y  $\sigma^2$ , pueden estimarse utilizando la normalidad de las distribuciones condicionadas por máxima verosimilitud, que equivale a mínimos cuadrados. La estimación de la pendiente es:

$$\hat{\beta}_1 = \frac{\hat{\gamma}_{xy}(0)}{s_x^2}, \quad (17.20)$$

donde  $\hat{\gamma}_{xy}(0) = \sum_{t=1}^T (y_t - \bar{y})(x_t - \bar{x})/T$  es la covarianza cruzada muestral de orden cero definida en la sección anterior y  $s_x^2 = \hat{\gamma}_x(0) = \sum_{t=1}^T (x_t - \bar{x})^2/T$ , la varianza de la serie estacionaria  $x_t$ . La varianza de la estimación de la pendiente es:

$$Var(\hat{\beta}_1) = \frac{\sigma_a^2}{Ts_x^2}. \quad (17.21)$$

Observemos que esta varianza depende del cociente entre la varianza del ruido,  $\sigma^2$ , y la varianza de la señal, que es la varianza del proceso estacionario  $x_t$  dada por  $s_x^2$ . Cuanto mayor sea la magnitud de la señal con relación al ruido, mejor estimaremos la pendiente.

Para contrastar que la relación encontrada es significativa necesitamos un estimador de  $\sigma^2$ . El estimador centrado habitual es

$$\hat{s}_R^2 = \frac{\sum \hat{a}_t^2}{T - 2}$$

y el contraste de que el parámetro  $\beta_1$  es distinto de cero utiliza el estadístico:

$$t = \frac{\hat{\beta}_1 s_x \sqrt{T}}{\hat{s}_R} \quad (17.22)$$

que si  $\beta_1$  es cero sigue una t de Student con  $T - 2$  grados de libertad. Rechazaremos que no existe relación cuando este estadístico esté significativamente alejado de cero.

### 17.5.2. Perturbaciones autocorreladas

Supongamos ahora que la perturbación tiene estructura dinámica, es decir, el proceso  $n_t$  de (17.8) es un proceso estacionario general, y la relación es:

$$y_t = \beta_0 + \beta_1 x_t + n_t.$$

Se demuestra en el apéndice 17.1, que si estimamos este modelo como en el caso anterior, suponiendo que la perturbación es ruido blanco, tendremos un estimador muy ineficiente y sesgado a encontrar relaciones entre procesos independientes. La razón es que la varianza del estimador de la pendiente (17.20) puede ser mucho mayor que la calculada por la fórmula (17.21), por lo que si utilizamos el estadístico (17.22) para el contraste podemos concluir que existe una relación inexistente en realidad.

En el apéndice 17.1 se demuestra que cuando relacionamos variables estacionarias y la perturbación tiene estructura dinámica, la varianza del estimador de mínimos cuadrados (17.20) de la pendiente de la recta es:

$$var(\hat{\beta}_1) = \frac{\sigma_n^2}{T s_x^2} \left[ 1 + 2 \sum_{t=1}^{T-1} (T-t) \rho_n(t) E(r_x(t)) \right] \quad (17.23)$$

donde  $\sigma_n^2$  es la varianza de la perturbación y  $\rho_n(t)$  y  $r_x(t)$  son los coeficientes de autocorrelación de orden  $t$  de la perturbación y de la variable  $x$ , respectivamente. De esta ecuación obtenemos las conclusiones siguientes:

## 17. Regresión dinámica entre variables estacionarias

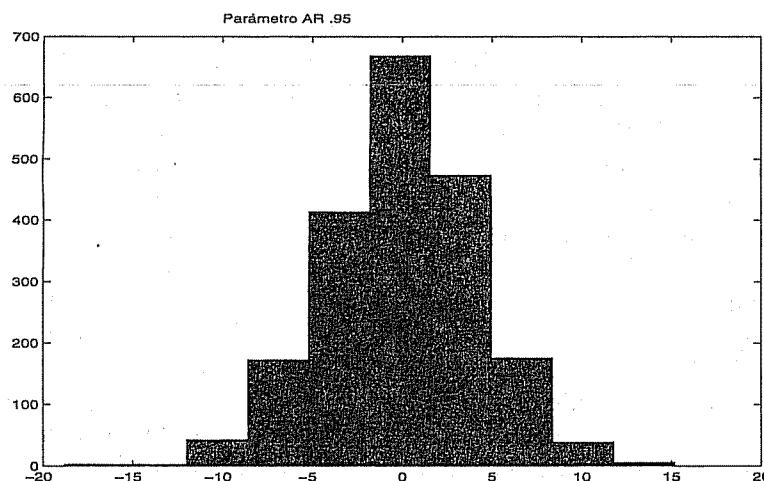
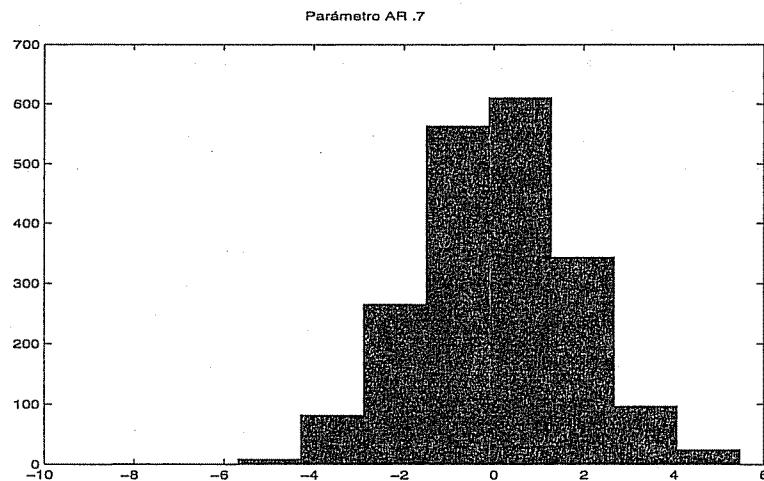
1. Si la perturbación es ruido blanco, sustituyendo en (17.23)  $\rho_n(t)$  por cero y haciendo  $\sigma_n^2 = \sigma_a^2$ , se obtiene la expresión (17.21) de la sección anterior, y el estimador es eficiente.
2. Si tanto la variable explicativa,  $x_t$ , como la perturbación,  $n_t$ , tienen autocorrelación positiva, la varianza del estimador de la pendiente puede ser mucho mayor que la calculada con la fórmula (17.21) bajo la hipótesis de independencia. En efecto, al calcular la varianza de los residuos tendremos una estimación de  $\sigma_n^2$ , la varianza del proceso estacionario, pero no tendremos en cuenta el resto de los términos, subestimando la varianza.
3. Un caso especial aparece cuando la variable  $x_t$  es ruido blanco incomparable con la perturbación  $n_t$ . Entonces, sustituyendo en (17.23)  $E(r_x(t))$  por cero tenemos que  $\text{var}(\hat{\beta}_1) = \sigma_n^2/Ts_x^2$ . En este caso, estimamos bien la varianza del estimador, por lo que evitamos el riesgo de la regresión espuria, pero utilizamos un estimador ineficiente, que podríamos mejorar estimando por mínimos cuadrados generalizados.

La implicación de este resultado es que al relacionar series temporales con autocorrelación, como es esperable que la perturbación,  $n_t$ , también la tenga, no debemos estimar una regresión clásica y utilizar las fórmulas habituales, ya que podemos fácilmente encontrar una relación significativa entre las dos variables independientes. En efecto, el estimador (17.20) tendrá una varianza que puede ser mucho mayor que la dada por la fórmula (17.21), y al contrastar la relación mediante el estadístico  $t$  con (17.22), como la varianza está muy subestimada, podemos concluir fácilmente que existe relación entre dos variables aunque ésta no exista.

Como ilustración la figura 17.4 muestra la distribución del estadístico  $t$  (17.22) obtenida simulando 2000 procesos independientes AR(1) con el mismo valor del parámetro y tamaño muestral igual a 100. Se observa que cuando el parámetro es .7 el estadístico  $t$  será mayor que 2 en valor absoluto aproximadamente el 25 % de las veces, y en estos casos concluiremos que existe relación entre las variables, aunque éstas sean independientes. Si el parámetro es .95, este porcentaje de error aumenta al 63 %, es decir, si hacemos una regresión entre dos variables estacionales mutuamente independientes, pero que tienen alta autocorrelación, es muy probable que encontraremos una falsa relación entre ellas. Este es el efecto de regresión espuria.

Estos resultados sugieren que, como no podemos saber a priori si existe o no autocorrelación en la perturbación, y la experiencia indica que al relacionar series temporales es altamente probable que la tenga, siempre que la variable explicativa,  $x_t$ , tenga dinámica y no sea un proceso

Figura 17.4 Distribución del estadístico t obtenido al regresar dos procesos estacionarios AR independientes. En el primer caso el coeficiente de ambos es .7 y en el segundo .95



de ruido blanco, debemos estimar el modelo permitiendo que la perturbación pueda estar autocorrelada. Si la perturbación estuviese incorrelada porque la variable  $x_t$  es capaz de explicar toda la dinámica de  $y_t$ , detectaremos esta situación por los coeficientes estimados no significativos en el modelo del ruido. Sin embargo, si existe autocorrelación en la

perturbación y no la tenemos en cuenta, los estimadores de los coeficientes  $v_i$  serán muy poco precisos y con desviaciones típicas estimadas erróneas, con lo que podemos fácilmente llegar a conclusiones equivocadas respecto a la relación entre las variables.

### 17.6. El modelo con varias variables explicativas

La metodología que hemos descrito en las secciones anteriores se generaliza sin dificultad para  $p$  variables explicativas. El modelo será entonces:

$$y_t = \frac{w_1(B) B^{b_1}}{\delta_1(B)} x_{1t} + \dots + \frac{w_p(B) B^{b_p}}{\delta_p(B)} x_{pt} + n_t. \quad (17.24)$$

Con varias series explicativas que pueden estar correladas entre sí en lugar de intentar identificar la función de transferencia para cada una conviene pasar a estimar directamente el modelo conjunto y reformularlo a la vista del resultado. Un procedimiento simple es tomar como modelo inicial para la perturbación el modelo univariante para la serie  $y_t$  y suponer estructuras simples iniciales para todas las variables, del tipo:

$$\frac{w_0 + w_1 B}{1 - \delta B}$$

En función de los coeficientes significativos encontrados iremos reformulando el modelo hasta obtener la estructura adecuada.

Cuando las variables explicativas tengan alta autocorrelación, o estén altamente correladas entre sí, no conviene estimar un modelo inicial del tipo

$$y_t = \sum v_{1i} x_{1,t-i} + \dots + \sum v_{pi} x_{p,t-i} + n_t \quad (17.25)$$

ya que existirá multicolinealidad muy alta y los coeficientes se estimarán muy inefficientemente. Es mejor reescribir las funciones de transferencia separando la ganancia de los efectos a corto plazo como vimos para una variable. Por ejemplo, podemos estimar como modelo inicial:

$$y_t = \sum_{i=1}^p \beta_{0i} x_{it} + \sum_{i=1}^p \beta_{1i} \nabla x_{i,t} + n_t$$

que separa los efectos a largo plazo, medidos por  $\beta_{0i}$ , de los a corto plazo, dados por  $\beta_{1i}$ . Para las  $p_1 \leq p$  variables donde se observen efectos significativos a corto plazo podemos aumentar la dinámica a corto plazo añadiendo un término en la variable  $\nabla^2 x_{i,t}$  y así sucesivamente hasta

**Tabla 17.2 Relación entre los rendimientos de la Bolsa de Madrid y otras**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RP	0.436987	0.099646	4.385390	0.0000
RP(-1)	0.004300	0.098648	0.043593	0.9653
RP(-2)	-0.021136	0.093130	-0.226953	0.8208
RF	0.190305	0.095117	2.000752	0.0473
RF(-1)	-0.002702	0.091994	-0.029374	0.9766
RF(-2)	0.002812	0.091787	0.030634	0.9756
RL	0.189117	0.128291	1.474134	0.1427
RL(-1)	0.131758	0.123089	1.070432	0.2863
RL(-2)	0.187006	0.123600	1.512996	0.1325
RNY	0.129813	0.119005	1.090818	0.2772
RNY(-1)	0.107637	0.120227	0.895278	0.3722
RNY(-2)	-0.104204	0.120855	-0.862224	0.3900

obtener el modelo:

$$y_t = \sum_{i=1}^p \beta_{0i} x_{it} + \sum_{j=1}^{p_1} \beta_{1j} \nabla x_{j,t} + \dots + \sum \beta_{im} \nabla^i x_{m,t} + n_t$$

La estimación, selección y diagnosis de estos modelos se realiza con los mismos principios que hemos visto en las secciones anteriores.

### Ejemplo 17.6

Vamos a estudiar la relación entre los rendimientos en Madrid ( $rm_t$ ) y los de las Bolsas de París, ( $rp_t$ ), Frankfurt, ( $rf_t$ ), Londres, ( $rl_t$ ), y Nueva York, ( $rnyt$ ). Como los rendimientos están poco autocorrelados utilizaremos la formulación simple (17.25). Poniendo dos retardos en cada variable la estimación se presenta en la tabla 17.2.

Se observa que hay relación instantánea con todas ellas, significativa con la de París y Frankfurt y no significativa con Londres y Nueva York. Sin embargo, con estas dos últimas parecen existir retardos, mientras que con las dos primeras las relaciones son más instantáneas. Eliminando coeficientes no significativos y seleccionando el modelo con mejor valor del BIC, se obtiene finalmente el modelo de la tabla 17.3:

$$rm_t = .38rp_t + .27rf_t + .22rl_t + .20rl_{t-1} + a_t.$$

Este modelo indica que la relación más fuerte de la Bolsa de Madrid es

## 17. Regresión dinámica entre variables estacionarias

---

**Tabla 17.3** Modelo final para la relación entre la Bolsa de Madrid y otras

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RP	0.387055	0.097053	3.988067	0.0001
RF	0.275974	0.089375	3.087838	0.0024
RL	0.222654	0.106378	2.093035	0.0380
RL(-1)	0.202525	0.076332	2.653217	0.0088

con la de Londres, seguida de París y Frankfurt. La relación con la Bolsa de Londres es instantánea y con un retardo, por lo que la Bolsa de Londres en un mes dado proporciona algo de información para prever la de Madrid en el mes siguiente.

---

### 17.7. Predicción

Las predicciones de un modelo de regresión dinámica se calculan con los mismos principios que en el caso univariante. Se demuestra que la previsión que minimiza el error cuadrático medio de los errores es la esperanza condicionada a la información disponible. En consecuencia:

$$\hat{y}_t(k) = E[y_{t+k} | y_t, \dots, y_1; x_t, \dots, x_1].$$

Las previsiones se obtienen recursivamente como en el caso univariante. Escribiendo el modelo como:

$$y_t = v(B)x_t + \psi(B)a_t, \quad (17.26)$$

tenemos que, para horizonte  $k$ :

$$\begin{aligned} y_{t+k} = & (v_0x_{t+k} + v_1x_{t+k-1} + \dots + v_mx_{t+k-m}) + \\ & + a_{t+k} + \psi_1a_{t+k-1} + \dots + \psi_ha_{t+k-h}, \end{aligned} \quad (17.27)$$

y tomando esperanzas condicionadas a la información en  $t$  y suponiendo que conocemos los parámetros y, por tanto, las perturbaciones  $a_t$  hasta el origen de la predicción:

$$\hat{y}_t(k) = v_0\hat{x}_t^U(k) + \dots + v_kx_t + \dots + \psi_ka_t + \dots + \psi_{k+h}a_{t-h}, \quad (17.28)$$

donde las previsiones  $\hat{x}_t^U(k)$ , son las de esta variable dada su historia y se obtendrán con el modelo univariante para esta serie  $x_t$ . Vemos que, en general, la predicción con un modelo de regresión dinámica requiere cono-

cer no sólo la función de transferencia sino también el modelo univariante para  $x_t$ .

Para calcular la varianza de la predicción, supongamos que  $x_t$  sigue el modelo:

$$x_t = \psi_x(B)\alpha_t, \quad (17.29)$$

sustituyendo (17.29) en (17.26):

$$y_t = v^*(B)\alpha_t + \psi(B)a_t,$$

donde  $v^*(B) = v(B)\psi_x(B)$ . La predicción puede entonces escribirse también como:

$$\hat{y}_t(k) = v_k^*\alpha_k + v_{k+1}^*\alpha_{k-1} + \dots + \psi_k a_t + \psi_{k+1} a_{t-1} + \dots$$

y su varianza será:

$$Var(\hat{y}_t(k)) = E(y_{t+k} - \hat{y}_t(k))^2 = \sigma_\alpha^2 \sum_{j=0}^{k-1} v_j^{*2} + \sigma^2 \sum_{j=0}^{k-1} \psi_j^2.$$

En este análisis hemos supuesto los parámetros del modelo conocidos. Podemos incorporar su incertidumbre y tener en cuenta la asociada a la elección del modelo de forma similar a lo estudiado en el capítulo 11.

---

### Ejemplo 17.7

Calcular las predicciones para uno y dos pasos y sus varianzas en el modelo de regresión dinámica:

$$\begin{aligned} y_t &= \frac{\omega_0}{1-\delta B} x_t + (1-\theta B) a_t \\ \nabla x_t &= u_t, \end{aligned}$$

donde  $u_t$  y  $a_t$  son ruidos blancos incorrelados con varianzas  $\sigma_u^2$  y  $\sigma_a^2$ .

Escribiendo el modelo en forma lineal multiplicando por el denominador de la función de transferencia, tenemos que:

$$y_t = \delta y_{t-1} + \omega_0 x_t + (1 - (\theta + \delta) B + \theta \delta B^2) a_t$$

como según el modelo univariante de  $x$ ,  $\hat{x}_t(k) = x_t$  para todo  $k$ , la predicción será:

## 17. Regresión dinámica entre variables estacionarias

---

$$\begin{aligned}\hat{y}_t(1) &= \delta y_t + \omega_0 x_t - (\theta + \delta) a_t + \theta \delta a_{t-1} \\ \hat{y}_t(2) &= \delta \hat{y}_t(1) + \omega_0 x_t + \theta \delta a_t \\ \hat{y}_t(k) &= \delta \hat{y}_t(k-1) + \omega_0 x_t \quad k > 2.\end{aligned}$$

Para calcular las varianzas de estas predicciones, escribiendo:

$$v^*(B) = \frac{\omega_0}{(1-\delta B)} \frac{1}{(1-B)},$$

los coeficientes  $v^*$  resultan al igualar potencias de  $B$  en la relación:

$$v^*(B)(1-B)(1-\delta B) = v^*(B)(1-(\delta+1)B+\delta B^2) = \omega_0.$$

El coeficiente de  $B^0$  en ambos miembros debe ser igual, y deducimos  $v_0^* = \omega_0$ . Para el de  $B$ :

$$\begin{aligned}v_1^* - (\delta + 1)v_0^* &= 0 \\ v_1^* &= \omega_0(1 + \delta),\end{aligned}$$

y para  $B^2$ ,  $v_2^* - (\delta + 1)v_1^* + \delta v_0^* = 0$ , es decir,  $v_2^* = \omega_0(1 + \delta + \delta^2)$ .

Una vez determinados los coeficientes  $v_j^*$ , la varianza será:

$$\begin{aligned}Var(\hat{y}_t(1)) &= \sigma_u^2 v_0^{*2} + \sigma_a^2 \\ Var(\hat{y}_t(2)) &= \sigma_u^2 (v_0^{*2} + v_1^{*2}) + \sigma_a^2 (1 + \theta^2).\end{aligned}$$


---

### Ejercicios 17

17.1 Dada la relación  $y_t = 2x_{t-1} + a_t$  donde  $x_t = .7x_{t-1} + u_t$  y  $u_t$  y  $a_t$  son ruidos blancos normales independientes de varianza unidad, se pide: (1) comprobar que el modelo univariante de  $y_t$  debe ser ARMA(1,1) (sugerencia, repase las reglas de adición de procesos estacionarios que estudiamos en la sección 5.7), (2) calcule la función de covarianzas cruzadas entre los dos procesos estacionarios.

17.2 Demostrar que si tenemos un proceso  $x_t$  con autocovarianzas  $\gamma_k$  y autocorrelaciones  $\rho_k$ : (1) la covarianza entre  $x_t$  y  $\nabla x_t$  es  $\gamma_0 - \gamma_1$ , (2) la correlación entre  $x_t$  y  $\nabla x_t$  es  $(1/2(1 - \rho_1))^{1/2}$ .

17.3 Demostrar que en el ejercicio anterior la autocorrelación de primer orden del proceso  $\nabla x_t$  es  $\rho_1(\nabla x_t) = (2\rho_1 - \rho_2 - 1)/(2(1 - \rho_1))$ .

- 17.4 Demostrar que si tenemos un proceso  $x_t$  con autocovarianzas  $\gamma_k$  y autcorrelaciones  $\rho_k$ : (1) la covarianza entre  $x_t$  y  $\nabla^2 x_t$  es  $\gamma_0 - 2\gamma_1 + \gamma_2$ . (2) la correlación entre  $x_t$  y  $\nabla^2 x_t$  es  $(1 - 2\rho_1 + \rho_2)/(6 - 8\rho_1 + 2\rho_2)$ .
- 17.5 Utilizar los ejercicios 17.2 y 17.3 para justificar que si una serie temporal  $x_t$  tiene autocorrelaciones  $\rho_1 = .9$ ,  $\rho_2 = .7$ , que definen la siguiente matriz de correlaciones entre las variables  $x_t, x_{t-1}, x_{t-2}$ :

$$\begin{bmatrix} 1 & .9 & .7 \\ .9 & 1 & .9 \\ .7 & .9 & 1 \end{bmatrix}$$

la matriz de correlaciones entre las variables  $x_t, \nabla x_t$  y  $\nabla^2 x_t$  es:

$$\begin{bmatrix} 1 & .22 & -.5 \\ .22 & 1 & .5 \\ -.5 & .5 & 1 \end{bmatrix}.$$


---

### Apéndice 17.1: Varianza de la estimación de mínimos cuadrados con autocorrelación

Supongamos el modelo  $y_t = \beta_0 + \beta_1 x_t + n_t$ . Sumando para todos los datos y dividiendo por  $T$ , tenemos que:  $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{n}$ , y restando estas dos ecuaciones, podemos escribir:

$$y_t = \bar{y} + \beta_1(x_t - \bar{x}) + n_t - \bar{n}.$$

Vamos a ver las consecuencias de estimar los parámetros por mínimos cuadrados cuando la perturbación no es ruido blanco. El estimador de la pendiente será:

$$\hat{\beta}_1 = \sum \frac{(x_t - \bar{x})}{Ts_x^2} (y_t - \bar{y})$$

y sustituyendo en esta ecuación  $y_t = \bar{y} + \beta_1(x_t - \bar{x}) + n_t - \bar{n}$ , se obtiene que

$$\hat{\beta}_1 = \beta_1 + \sum \frac{(x_t - \bar{x})n_t}{Ts_x^2}.$$

Supongamos que  $n_t$  es independiente de  $x_t$  y tiene esperanza cero. Tomando esperanzas en esta expresión comprobamos que  $E(\hat{\beta}_1) = \beta_1$  y el estimador es centrado. Para calcular su varianza, escribamos:

$$(\hat{\beta}_1 - \beta_1)^2 = \left( \sum \frac{(x_t - \bar{x})n_t}{Ts_x^2} \right)^2.$$

---

## 17. Regresión dinámica entre variables estacionarias

---

Vamos ahora a tomar primero esperanzas en esta ecuación condicionando a  $Ts_x^2$ . Entonces:

$$\begin{aligned} E((\hat{\beta}_1 - \beta_1)^2 | Ts_x^2) &= \frac{1}{Ts_x^2} E\left(\sum_{t=1}^T \frac{(x_t - \bar{x})^2}{Ts_x^2} n_t^2 + \right. \\ &\quad \left. + 2 \sum_{t=1}^{T-1} \frac{(x_t - \bar{x})(x_{t+1} - \bar{x})}{Ts_x^2} n_t n_{t+1} + \dots + \right. \\ &\quad \left. + 2 \sum_{t=1}^{T-k} \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2} n_t n_{t+k} + \dots + 2 \frac{(x_1 - \bar{x})(x_T - \bar{x})}{Ts_x^2} n_1 n_T\right). \end{aligned}$$

Esta ecuación indica que esta esperanza condicionada depende de toda la dinámica de ambas variables. Llamando  $r_x(k) = \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) / Ts_x^2$  a los coeficientes de autocorrelación muestral de la variable estacionaria  $x_t$  y  $\rho_n(k)$  a los coeficientes de autocorrelación teóricos del proceso  $n_t$ , como:

$$\begin{aligned} E\left[\sum_{t=1}^{T-k} \frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2} n_t n_{t+k}\right] &= \\ = \sum_{t=1}^{T-k} E\left[\frac{(x_t - \bar{x})(x_{t+k} - \bar{x})}{Ts_x^2}\right] E(n_t n_{t+k}) &= \sigma_n^2 \rho_n(k) E(r_x(k)), \end{aligned}$$

tenemos que,

$$\begin{aligned} E((\hat{\beta}_1 - \beta_1)^2 | Ts_x^2) &= \frac{\sigma_n^2}{Ts_x^2} [1 + 2(T-1)\rho_n(1)E(r_x(1)) + \\ &\quad + \dots + 2\rho_n(T-1)E(r_x(T-1))]. \end{aligned}$$

Tomando esperanzas ahora respecto a  $Ts_x^2$  y utilizando que  $E(y) = E_x E_{y|x}(y)$ , obtenemos finalmente que:

$$\begin{aligned} var(\hat{\beta}_1) &= \frac{\sigma_n^2}{Ts_x^2} [1 + 2(T-1)\rho_n(1)E(r_{xx}(1)) + \\ &\quad + \dots + 2\rho_n(T-1)E(r_{xx}(T-1))]. \end{aligned}$$

Por ejemplo, supongamos para simplificar que ambos procesos siguen procesos AR(1) independientes:

$$x_t = \phi x_{t-1} + \epsilon_t$$

y

$$n_t = \rho n_{t-1} + \zeta_t.$$

Entonces  $\rho_n(k) = \rho^k$  y tomando aproximadamente  $E(r_x(k)) \approx \phi^k$  tenemos que

$$var(\hat{\beta}_1) = \frac{\sigma_n^2}{Ts_x^2} \left[ 1 + 2 \sum_{k=1}^{T-1} (T-k) \rho^k \phi^k \right].$$

Los términos del sumatorio forman una suma progresión geométrica,  $a + a^2 + \dots + a^j$ , con  $a = \rho\phi$ , y puede comprobarse que:

$$\sum_{k=1}^{T-1} (T-k)a^k = \frac{1}{(1-a)^2} [a(1-a) - (1-a^T)]$$

y despreciando para simplificar el término  $\rho^T \phi^T$ , resulta que:

$$var(\hat{\beta}_1) = \frac{\sigma_n^2}{Ts_x^2} \left[ \frac{1+\rho\phi}{1-\rho\phi} \right].$$

Por ejemplo, si  $\rho = \phi = .9$  la varianza real del estimador de mínimos cuadrados es 9.5 veces la calculada sin tener en cuenta la autocorrelación. El estimador es muy ineficiente y podemos fácilmente encontrar relaciones que no existen.

## Apéndice 17.2: Mínimos cuadrados generalizados

Para explicar el procedimiento escribiremos el modelo de regresión lineal en su forma general

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}$$

donde  $\mathbf{Y}$  es un vector de dimensión  $T \times 1$  de la variable dependiente,  $\mathbf{X}$  es una matriz  $T \times p$  de variables explicativas y  $\boldsymbol{\beta}$  es un vector  $p \times 1$  de parámetros. La primera columna de  $\mathbf{X}$  contiene un vector de unos y las siguientes los valores de las variables dependientes, que incluyen valores actuales y retardados de cada una de las variables. Hemos visto en el apéndice 2.1 que el estimador mínimo cuadrático es:

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

Sin embargo, con series temporales es esperable que las perturbaciones,  $\mathbf{N}$ , no sean ruido blanco y tengan una estructura de dependencia definida por su matriz de varianzas covarianzas,

$$E[\mathbf{N}\mathbf{N}'] = \sigma^2 \mathbf{G} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}$$

17. Regresión dinámica entre variables estacionarias

---

Supongamos que la matriz  $\mathbf{G}$  fuese conocida. Entonces podríamos transformar el problema en otro con nuevas perturbaciones incorreladas y aplicar después mínimos cuadrados a la ecuación transformada. Toda matriz definida positiva admite una representación como  $\mathbf{G} = \mathbf{L}\mathbf{L}'$ , donde la matriz  $\mathbf{L}$  es triangular, que se conoce como descomposición de Cholesky, y podemos transformar el modelo mediante:

$$\mathbf{L}^{-1}\mathbf{Y} = \mathbf{L}^{-1}\mathbf{X}\beta + \mathbf{L}^{-1}\mathbf{N}$$

y llamando  $\mathbf{Y}_G = \mathbf{L}^{-1}\mathbf{Y}$  y  $\mathbf{X}_G = \mathbf{L}^{-1}\mathbf{X}$  a las variables transformadas y  $\mathbf{E} = \mathbf{L}^{-1}\mathbf{N}$  al nuevo vector de perturbaciones, que tendrá matriz de varianzas covarianzas:

$$E[\mathbf{EE}'] = E[\mathbf{L}^{-1}\mathbf{NN}\mathbf{L}^{-1}'] = \mathbf{L}^{-1}E[\mathbf{NN}]\mathbf{L}^{-1'} = \sigma^2\mathbf{I}$$

y tendrá perturbaciones incorreladas. Aplicando mínimos cuadrados a las variables  $\mathbf{Y}_G$ ,  $\mathbf{X}_G$ , el estimador obtenido para  $\beta$ , será:

$$\hat{\beta}_G = (\mathbf{X}'_G \mathbf{X}_G)^{-1} \mathbf{X}'_G \mathbf{Y}_G$$

y sustituyendo las variables transformadas por las originales:

$$\hat{\beta}_G = (\mathbf{X}' \mathbf{G}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{G}^{-1} \mathbf{Y} \quad (17.30)$$

que llamaremos *estimador de mínimos cuadrados generalizados* o estimador *MCG*.

Por ejemplo, supongamos que el ruido sigue el proceso ARMA

$$\mathbf{N} = \Psi \mathbf{A} \quad (17.31)$$

donde

$$\Psi = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \psi_1 & 1 & \dots & 0 & 0 \\ \dots & 0 & \dots & 1 & 0 \\ \psi_T & \psi_k & \dots & \psi_1 & 1 \end{bmatrix}$$

es la matriz triangular de coeficientes de la representación MA( $\infty$ ) del proceso y  $\mathbf{A} = (a_1, \dots, a_T)'$  es un proceso de ruido blanco. Entonces:

$$\Pi \mathbf{N} = \mathbf{A}$$

donde  $\Pi = \Psi^{-1}$  es la matriz autorregresiva. La matriz de covarianzas de  $\mathbf{N}$  será

$$E[\mathbf{NN}'] = \Psi \Psi' \sigma^2$$

y la inversa de  $\Psi^{-1}$  es  $\Pi$ . Por tanto, la transformación requerida para un proceso AR(1) utiliza su matriz de coeficientes AR que tendrá unos en la diagonal, el coeficiente  $\phi$  en la siguiente subdiagonal y el resto serán ceros. Es

dicir:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -\phi & 1 & \dots & 0 & 0 \\ \dots & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -\phi & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}.$$

El análisis anterior es aproximado para los primeros valores, ya que la ecuación (17.31) no genera exactamente el proceso, sino sólo después de los primeros retardos. Por ejemplo, para un MA(1) todos los  $\psi_j$  son cero para  $j \geq 2$  y la ecuación que resulta para  $u_1$  es  $u_1 = a_1$  y sólo a partir de  $j = 2$  obtenemos  $u_2 = a_2 + \psi a_1$ . Por tanto, la matriz de varianzas y covarianzas teórica de un MA(1) no es exactamente  $\Psi \Psi' \sigma^2$ , pero sí aproximadamente, si despreciamos la primera observación.

### Apéndice 17.3: Estimación de la función de transferencia mediante preblanqueo

Supongamos dos procesos relacionados por la ecuación:

$$y_t = v(B)x_t + n_t \quad (17.32)$$

donde  $x_t$  sigue el proceso ARMA:

$$\phi_x(B)x_t = \theta_x(B)\alpha_t,$$

siendo  $\alpha_t$  un proceso de ruido blanco con varianza  $\sigma_\alpha^2$ . Se verifica que:

$$\alpha_t = \theta_x^{-1}(B)\phi_x(B)x_t = \psi^{-1}(B)x_t.$$

Si aplicamos  $\psi^{-1}(B)$  a los dos miembros de la relación (17.32), tendremos que:

$$\psi^{-1}(B)y_t = v(B)\psi^{-1}(B)x_t + \psi^{-1}(B)n_t$$

y llamando  $\beta_t = \psi^{-1}(B)y_t$ , y  $\varepsilon_t = \psi^{-1}(B)n_t$ :

$$\beta_t = v(B)\alpha_t + \varepsilon_t \quad (17.33)$$

resultando que existe la misma función de transferencia entre  $\beta_t$  y  $\alpha_t$  que entre  $y_t$  y  $x_t$ . La diferencia principal es que en este último caso la variable explicativa,  $\alpha_t$ , es ruido blanco. En consecuencia, multiplicando ambos miembros de (17.33) por  $\alpha_{t-k}$  y tomando esperanzas:

$$E[\beta_t \alpha_{t-k}] = v_k \sigma_\alpha^2 = \gamma_{\alpha\beta}(k),$$

ya que  $\varepsilon_t$  está incorrelado con  $\alpha_{t-k}$ , al estarlo, por hipótesis,  $n_t$  y  $x_t$ . Entonces,

## 17. Regresión dinámica entre variables estacionarias

---

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2},$$

y llamando  $\rho_{\alpha\beta}(k)$  a la función de correlación cruzada entre ambas series,  $\gamma_{\alpha\beta}(k) = \rho_{\alpha\beta}(k)\sigma_\alpha\sigma_\beta$ , se obtiene finalmente:

$$v_k = \rho_{\alpha\beta}(k) \frac{\sigma_\beta}{\sigma_\alpha} \quad (17.34)$$

Para obtener una estimación preliminar de  $\hat{v}_k$  se sustituyen los coeficientes de correlación teóricos  $\rho_{\alpha\beta}(k)$  por los muestrales,  $r_{\alpha\beta}(k)$ , y las varianzas teóricas por las estimadas. Con ello resulta:

$$\hat{v}_k = r_{\alpha\beta}(k) \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha},$$

y aunque esta estimación no es eficiente, la representación de  $\hat{v}_k$  en función de  $k$  puede servirnos para obtener una identificación inicial de los órdenes  $(m, a, b)$ .

Este procedimiento se conoce como el *método del prelanqueado* de las series y es debido a Box y Jenkins. Tiene el inconveniente de que no puede extenderse cuando existe más de un regresor.

### Apéndice 17.4: Descomposición de la función de transferencia

Dado un operador  $v_k(B) = v_0 + v_1B + \dots + v_kB^k$  de orden finito  $k$ , vamos a demostrar que siempre puede escribirse como:

$$v_k(B) = v(1) + \sum_{i=0}^{k-1} \alpha_i \nabla B^i,$$

y también como:

$$v_k(B) = v(1) + \sum_{i=1}^k \beta_i \nabla^i.$$

La demostración de esta propiedad es inmediata igualando términos en los dos miembros. Por ejemplo para  $k = 2$ , tenemos para la primera ecuación:

$$v_0 + v_1B + v_2B^2 = (v_0 + v_1 + v_2) + \alpha_1(1 - B) + \alpha_2B(1 - B)$$

de donde resulta, igualando las potencias más altas, en  $B^2$ :

$$\alpha_2 = -v_2.$$

Igualando ahora las potencias de  $B$  en ambos miembros, tenemos  $v_1 = -\alpha_1 + \alpha_2$ , que conduce a:

$$\alpha_1 = -v_1 - v_2.$$

Apliquemos el mismo método en el segundo caso para obtener los coeficientes  $\beta$ . Entonces:

$$v_0 + v_1B + v_2B^2 = (v_0 + v_1 + v_2) + \beta_1(1 - B) + \beta_2(1 - 2B + B^2).$$

Es inmediato que ahora  $\beta_2 = v_2$ , y como tenemos  $v_1$  en el primer miembro y  $-\beta_1 - 2\beta_2$  en el segundo, resulta:

$$\beta_1 = -(v_1 + 2v_2).$$

Puede demostrarse, por inducción que, el resultado general es:

$$\beta_i = \frac{(-1)^i}{i!} v_k^{(i)}(1),$$

donde  $v_k^{(i)}(1)$  es la derivada  $i$ -ésima de  $v_k(B)$  respecto a  $B$  particularizada luego para  $B = 1$ .

Por ejemplo, si  $k = 2$  y  $i = 1$ , entonces  $v_2^{(1)}(B) = v_1 + 2v_2B$ , y también  $v_2^{(1)}(1) = v_1 + 2v_2$ . Sustituyendo estas expresiones se obtiene que  $\beta_1 = (-1)(v_1 + 2v_2) = -(v_1 + 2v_2)$ , como obtuvimos directamente. La fórmula es también válida para  $i = 0$  y entonces  $\beta_0 = v_k^{(0)}(1) = v_k(1)$ .

## Apéndice 17.5: Programas para la estimación de funciones de transferencia

Las funciones de transferencia estudiadas pueden estimarse con toda generalidad con el programa SCA y con ciertas restricciones con EViews. Los demás programas comentados no permiten esta opción. La estimación con EViews de un modelo de regresión dinámica con perturbación ARMA requiere que la función de transferencia sea de la forma:  $y_t = v_0 + \sum_{i=0}^k v_i x_{t-i} + n_t$ , sin estructura en el denominador de la función de transferencia. Para identificar el modelo suponiendo estructura AR( $p$ ) en la perturbación se utiliza el comando *Quick*: en ecuación poner estimar por LS-NLS la ecuación,

$$y c x(0 to -k) ar(1) ar(2) \dots ar(p)$$

Si queremos estimar una función de transferencia con términos en el denominador hay que transformar el modelo. Por ejemplo, para estimar la ecuación:

$$y_t = \frac{w_0 + w_1B + w_2B^2}{1 - \delta B} x_t + \frac{1}{1 - \phi B} a_t$$

## 17. Regresión dinámica entre variables estacionarias

escribiremos el modelo multiplicando todo por  $(1 - \delta B)$ , lo que implica introducir como regresor el valor retardado de la variable, y modificar la ecuación de la perturbación, que pasará a ser un ARMA(1,1), del tipo:

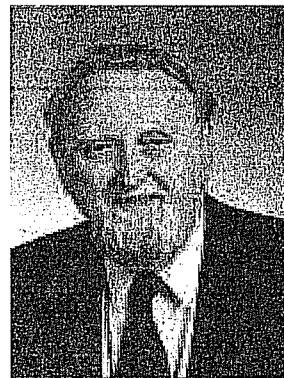
$$y_t = \delta y_{t-1} + (w_0 + w_1 B + w_2 B^2)x_t + \frac{1 - \delta B}{1 - \phi B} a_t$$

El modelo a estimar con EViews será:

y      y(-1)      x(0 to -2)      ar(1)      ma(1).

---

## 18. Regresión entre variables integradas. Cointegración



Clive W. J. Granger (1934)

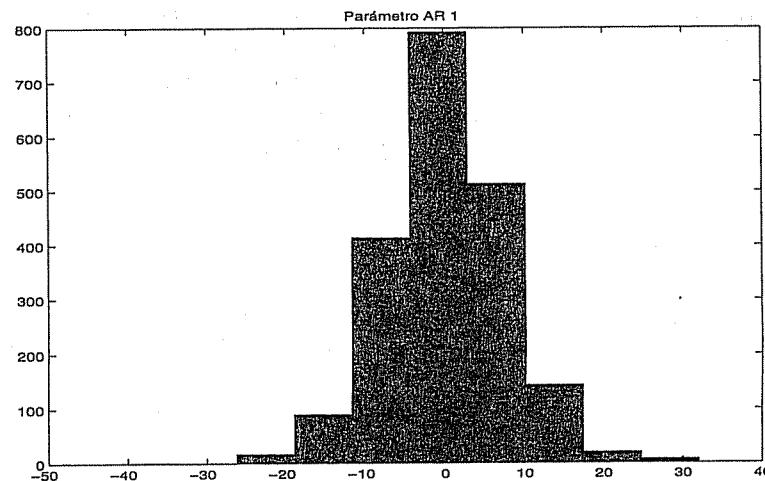
Estadístico y económetra británico. Se doctoró en estadística en la universidad de Nottingham en 1959, donde enseñó, tras un paréntesis como visitante en Princeton, hasta trasladarse a Estados Unidos en 1974. Desde entonces ha sido profesor del Departamento de Economía en la Universidad de California en San Diego. Desarrolló con Engle el concepto de cointegración y ha hecho contribuciones muy destacadas a la metodología y la práctica de la predicción con series temporales. Recibió en 2003 el Premio Nobel de Economía.

### 18.1. Introducción

En este capítulo vamos a estudiar cómo relacionar variables no estacionarias. Con estas variables no podemos confiar en los procedimientos tradicionales, ya que la inferencia con variables no estacionarias requiere, en general, distribuciones no estándar, como vimos en el capítulo 9 al construir un contraste de raíces unitarias. En consecuencia, el riesgo que vimos en el capítulo anterior de obtener regresiones espurias por la autocorrelación en la perturbación es especialmente importante en este caso.

Para ilustrar el problema de la regresión espuria al relacionar dos series independientes no estacionarias, la figura 18.1 muestra el histograma de los valores del estadístico  $t = \hat{\beta}/\hat{s}(\hat{\beta})$  cuando se aplica a regresiones entre dos paseos aleatorios independientes. Este histograma muestra que el 75 % de las veces encontraremos un valor de este estadístico  $t$  mayor que dos, con lo que concluiremos que la pendiente de la ecuación de

**Figura 18.1 Histograma de los valores del estadístico t para contrastar la presencia de relación lineal entre dos paseos aleatorios independientes**



regresión simple que relaciona ambas variables es distinta de cero. Por tanto, utilizando este estadístico encontraremos con alta probabilidad una relación entre las dos variables que no existe.

Una posible solución para evitar estos problemas es diferenciar las series para convertirlas en estacionarias y después aplicar los métodos que vimos en el capítulo anterior para procesos estacionarios. Sin embargo, esta alternativa, que fue popular hace unos años, tiene también sus inconvenientes. Es posible que aunque las dos series sean no estacionarias una de ellas explique totalmente el comportamiento no estacionario de la otra. Si las series están relacionadas por:

$$y_t = \beta_0 + \beta_1 x_t + n_t, \quad (18.1)$$

donde  $n_t$  es estacionario, la variable  $x_t$  es capaz de explicar totalmente el comportamiento no estacionario de  $y_t$ . Decimos entonces que las variables están cointegradas. Si para evitar los problemas de regresión espuria comentados antes diferenciamos las variables, el modelo para las variables estacionarias  $\nabla y_t$  y  $\nabla x_t$  es:

$$\nabla y_t = \beta_1 \nabla x_t + \nabla n_t \quad (18.2)$$

y hemos complicado una relación que era muy simple con consecuencias negativas. En primer lugar, la estimación de la pendiente en este modelo

será menos eficiente que en el modelo (18.1), ya que hemos reducido la variabilidad del regresor al diferenciarlo y, previsiblemente, hemos también aumentado la variabilidad del ruido. En segundo lugar, hemos introducido un término de media móvil no invertible en la perturbación, con lo que la aproximación AR no será válida y se nos complicará mucho la formulación de un modelo para la perturbación.

Por ejemplo, supongamos para simplificar que en (18.1)  $n_t = a_t$  es ruido blanco y que  $\nabla x_t = u_t$ . Al pasar del modelo (18.1) al (18.2) hemos disminuido la varianza del regresor, ya que  $\nabla x_t = u_t$  tendrá una varianza mucho menor que  $x_t$ , que es no estacionaria, y además hemos aumentado la varianza del ruido, ya que si  $\text{var}(a_t) = \sigma^2$ , entonces  $\text{var}(\nabla a_t) = 2\sigma^2$ . Por tanto, el modelo diferenciado es mucho más ineficiente que el que utiliza los niveles de las variables originales.

En este capítulo vamos a proponer una estrategia para construir modelos de regresión que tenga en cuenta estos posibles problemas.

## 18.2. Cointegración

Recordemos que una serie temporal es integrada de orden  $d$ , y lo representaremos por  $I(d)$ , si es necesario aplicar  $d$  diferencias para trasformarla en estacionaria, o  $I(0)$ . Diremos de series  $I(d)$  integradas,  $x_t, y_t$ , que están *cointegradas* si existe una combinación lineal entre ellas que es de orden de integración menor de  $d$ . Es decir, si podemos construir una serie  $n_t^*$ :

$$n_t^* = \alpha_1 y_t + \alpha_2 x_t,$$

que es  $I(d_1)$ , donde  $d_1 < d$ . A la combinación  $(\alpha_1, \alpha_2)$  se la denomina la *relación de cointegración*. Esta relación no es única, ya que cualquier relación del tipo  $(c\alpha_1, c\alpha_2)$  es también de cointegración para cualquier  $c \neq 0$ .

Cuando existe cointegración, una de las dos variables explica parte de la tendencia de la otra. Dividiendo la ecuación anterior por  $\alpha_1$  y reordenando términos, podemos escribir:

$$y_t = \beta x_t + n_t \quad (18.3)$$

donde  $\beta = -\alpha_2/\alpha_1$  y  $n_t = n_t^*/\alpha_1$  es un proceso  $I(d_1)$ . En la ecuación (18.3) diremos que la relación de cointegración es  $(1, -\beta)$  y estimar la relación de cointegración equivale a estimar la relación a largo plazo entre las variables dada por  $\beta$ .

Un caso importante de cointegración en la práctica es cuando las dos series son  $I(1)$ , pero existe una combinación lineal que es estacionaria. Esta combinación será la que encontraremos al realizar una regresión entre

---

## 18. Regresión entre variables integradas. Cointegración

---

las variables por mínimos cuadrados, ya que la suma de cuadrados de los residuos será siempre más pequeña cuando el residuo sea estacionario que cuando no lo sea. Cuando exista cointegración al construir el modelo de función de transferencia encontraremos que la variable independiente o exógena explicará totalmente la no estacionaridad de la variable dependiente, y la perturbación seguirá un proceso estacionario.

Estas ideas se generalizan para cualquier número de series: dadas las series  $x_{1t}, \dots, x_{pt}$  que son  $I(d)$  diremos que están cointegradas si existe una combinación lineal entre ellas que es  $I(d_1)$ , donde  $d_1 < d$ .

Es importante tener en cuenta que la formulación del modelo de función de transferencia debe respetar las reglas de adición de procesos estocásticos. Como:

$$\alpha_1 I(0) + \alpha_2 I(0) = I(0), \quad (18.4)$$

cualquier combinación lineal de dos procesos estacionarios es forzosamente un proceso estacionario: su media será constante y las covarianzas sólo dependerán del retardo. Además, podemos determinar la estructura del proceso suma con las reglas de adición de procesos estacionarios que vimos en la sección 5.7.

Si uno de los dos procesos es no estacionario y el otro es estacionario, la suma tendrá que ser forzosamente no estacionaria. Esta propiedad implica que en la formulación del modelo de transferencia si  $x_t$  es estacionario e  $y_t$  no lo es, entonces el ruido  $n_t$  debe ser no estacionario.

Cuando sumamos procesos no estacionario, su suma verifica:

$$\alpha_1 I(d) + \alpha_2 I(d) = I(h \leq d). \quad (18.5)$$

En efecto, tomando  $d$  diferencias los dos sumandos son estacionarios, luego, según (18.5), lo será su suma. En consecuencia  $h \leq d$ . Cuando  $h = d$ , diremos que no existe cointegración, mientras que si  $h < d$  diremos que los dos procesos están cointegrados.

### 18.2.1. Estimación de la relación entre variables cointegradas

La varianza de un proceso estacionario está acotada, mientras que, como hemos visto, la varianza de un paseo aleatorio crece sin límite. Podemos generalizar este resultado indicando que, si eliminamos el problema de las distintas unidades de medida (estandarizando los procesos, o tomando logaritmos) la varianza muestral que calculamos en una serie generada por un proceso integrado es esperable que sea mucho mayor que la varianza muestral calculada con una serie generada por un proceso estacionario (véase el apéndice 18.1 para un análisis de esta propiedad). Supongamos que relacionamos dos series temporales  $I(1)$  que están coin-

tegradas, y siguen la ecuación (18.3). Si estimamos esta ecuación de regresión por mínimos cuadrados, puede demostrarse, generalizando los resultados del capítulo anterior, que la varianza del estimador de la pendiente,  $\beta$ , depende de la relación entre la varianza del ruido, que será la de un proceso estacionario,  $\beta$ , y la varianza muestral del proceso no estacionario  $x_t$ . Como el numerador de esta varianza, la del ruido, está acotado pero el denominador, la varianza muestral de  $x_t$ , no y crece con el tamaño muestral, la varianza del estimador de  $\beta$  será muy pequeña y podemos estimar este parámetro incluso mejor que si  $x_t$  fuese un proceso estacionario. Puede demostrarse que si aumenta el tamaño muestral, el estimador de mínimos cuadrados converge al verdadero valor del parámetro más rápidamente de lo que lo haría en una relación entre variables estacionarias. Como vimos en el apéndice 17.1, la varianza del estimador de la pendiente en el caso estacionario es:

$$var(\hat{\beta}) = \frac{\sigma^2}{Ts_x^2} f,$$

donde  $f$  es una función de los coeficientes de autocorrelación y lo mismo que  $\sigma^2$  y  $s_x^2$ , cuando  $T \rightarrow \infty$  tiende a una constante. Por tanto, la varianza tiende a cero con  $T$ . Intuitivamente, si pensamos en una fórmula similar cuando la serie  $x_t$  es no estacionaria, entonces su varianza muestral,  $s_x^2$ , tiende a infinito con  $T$ , con lo que la varianza de la pendiente será menor que en el caso estacionario. Por esta razón se dice que el estimador del parámetro  $\hat{\beta}$  en la regresión entre dos variables no estacionarias es *superconsistente*. Este resultado sugiere investigar si las dos variables están cointegradas estimando directamente por regresión su relación en niveles.

Un problema importante al realizar inferencias con variables no estacionarias es que las distribuciones de los estimadores no son en general normales y los estadísticos habituales para contrastar si un coeficiente es cero no son válidos. Esto complica la construcción de estos modelos y aconseja reparametrizarlos para utilizar los resultados de variables estacionarias.

El modelo puede construirse en tres etapas como sigue:

1. Estimar por mínimos cuadrados una regresión entre las dos variables:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}x_t.$$

2. Obtener los residuos de esta ecuación:

$$\hat{n}_t = y_t - \hat{\beta}_0 - \hat{\beta}x_t,$$

y comprobar si son estacionarios. Si lo son, las variables están coin-

## 18. Regresión entre variables integradas. Cointegración

---

tigradas. Si no lo son, las variables no están cointegradas y debemos diferenciarlas para construir la relación.

3a. Construir un modelo para la serie estacionaria  $\hat{n}_t$ .

Para interpretar este procedimiento, supongamos que la verdadera relación entre dos variables  $I(1)$  es:

$$y_t = \beta_0 + v(B)x_t + n_t \quad (18.6)$$

y existe cointegración de manera que el ruido  $n_t$  sea estacionario. Vamos a comprobar que si sustituimos  $v(B)$  por una constante la perturbación de esta ecuación sólo es estacionaria si  $g$  es la ganancia, dada por:

$$g = v(1) = \sum_{i=1}^m v_i.$$

En efecto, como:

$$\hat{n}_t^* = y_t - \beta_0 - gx_t = v(B)x_t + n_t - v(1)x_t = n_t - \sum_{i=1}^m v_i(x_t - x_{t-i})$$

y las diferencias  $(x_t - x_{t-i})$  son estacionarias si  $x_t$  es  $I(1)$ , las perturbaciones  $\hat{n}_t^*$  de esta ecuación serán estacionarias. Supongamos ahora que  $g \neq \sum_{i=1}^m v_i$  no es la ganancia: el segundo miembro contendrá términos en  $x_t$  en lugar de en diferencias de  $x_t$ , y no podrá ser estacionario.

La conclusión práctica de este resultado es que si estimamos por mínimos cuadrados una regresión entre  $y_t$  y  $x_t$  la pendiente estimará la ganancia, o efecto a largo plazo y el ruido será estacionario. Esto justifica que llamemos a la relación de cointegración obtenida por regresión la relación a largo plazo entre las variables.

Un inconveniente de este método es que aunque obtenemos una estimación rápida de la ganancia de la relación, no estimamos bien el dinámica de la función de transferencia. En efecto, si el modelo estimado para el ruido es  $\phi(B)\hat{n}_t = \hat{a}_t$  y ponemos juntos los resultados de 1 y 3a, el modelo será:

$$\hat{y}_t = \hat{\beta}_0 + \hat{g}x_t + \phi(B)^{-1}\hat{a}_t,$$

que puede ser muy distinto de (18.6). Por tanto, la ecuación obtenida debe considerarse una primera aproximación para reformular y mejorar.

Una alternativa al paso 3a es utilizar la descomposición del operador  $v(B)$  por sus efectos a corto y largo plazo, que estudiamos en el capítulo

anterior. Si escribimos la relación utilizando la descomposición:

$$y_t = gx_{t-1} + \sum_{i=0}^{k-1} \gamma_i \nabla x_{t-i} + n_t,$$

y restamos  $y_{t-1}$  en ambos términos, obtenemos el modelo:

$$\nabla y_t = -(y_{t-1} - gx_{t-1}) + \sum_{i=0}^q \omega_i \nabla x_{t-i} + n_t,$$

donde todos los términos son estacionarios. Esto sugiere sustituir el paso 3a por la estimación del denominado *modelo de corrección de error*:

3b. Estimar el modelo:

$$\nabla y_t = \alpha(y_{t-1} - gx_{t-1}) + \sum_{i=0}^q \omega_i \nabla x_{t-i} + \sum_{i=1}^p \phi_i \nabla y_{t-i} + a_t, \quad (18.7)$$

donde  $p$  y  $q$  se eligen para que los residuos  $a_t$  sean ruido blanco. Este modelo incluye la relación de cointegración estimada en el paso 1 como un regresor y todos sus componentes son estacionarios.

La ventaja de esta formulación es que permite obtener mejor la dinámica de la relación y puede estimarse por mínimos cuadrados, ya que todos los términos involucrados son estacionarios. Su inconveniente es que al no permitir términos de media móvil en el ruido ni en el denominador de la función de transferencia puede requerir valores de  $p$  y  $q$  innecesariamente altos.

La tercera alternativa es en lugar de comenzar estimando la relación de cointegración formular directamente el modelo que vimos en el capítulo anterior:

$$y_t = gx_t + \sum_{i=0}^{k-1} \alpha_i \nabla x_{t-i} + n_t \quad (18.8)$$

donde se supone un modelo estacionario para el ruido. Si las variables están cointegradas, el coeficiente en  $x_t$  estimará la relación de cointegración,  $y_t - gx_t$  será una variable estacionaria y los coeficientes de las variables  $\nabla x_{t-i}$  describen cómo las variaciones de  $x_t$  se comportan con relación a la ecuación de equilibrio a largo plazo.

### 18.3. Construcción de modelos dinámicos. Caso general

Supongamos que queremos relacionar dos procesos  $x_t, y_t$  que son integrados, del mismo o distinto orden. Si la variable  $y_t$  tiene estacionalidad con periodo  $s$  y suponemos que  $x_t$  influye sobre  $y_t$  pero no al contrario, el modelo general que los relaciona es:

$$y_t = \frac{w(B)\Omega(B^s)}{\delta(B)\Delta(B^s)}x_{t-b} + \frac{\theta(B)\Theta(B^s)}{\phi(B),\Phi(B^s)}a_t$$

donde  $w(B)\Omega(B^s)$  son los operadores de media móvil de la función de transferencia regular y estacional respectivamente y  $\delta(B)$  y  $\Delta(B^s)$  los denominadores. El modelo de la perturbación contiene los operadores AR  $\phi(B)$  y  $\Phi(B^s)$ , que pueden tener raíces unitarias y los operadores MA  $\theta(B)$  y  $\Theta(B^s)$ , que son invertibles. El caso más frecuente es que únicamente sea estacional el proceso de perturbación o de inercia, con lo que el modelo se simplifica a:

$$y_t = \frac{w(B)}{\delta(B)}x_{t-b} + \frac{\theta(B)\Theta(B^s)}{\phi(B),\Phi(B^s)}a_t \quad (18.9)$$

Para establecer una metodología para construir esta ecuación consideremos los dos casos siguientes.

#### 18.3.1. Variables con mismo orden de integración

Supongamos que las dos variables tienen el mismo orden de integración. Entonces, la perturbación tendrá como máximo también ese orden de cointegración. Supongamos para simplificar que ambas series son  $I(1)$ . Escribiremos el modelo como (18.8) de manera que el número de retardos utilizados sea bajo, menor que tres, y el modelo de la perturbación  $n_t$  se toma como la parte estacionaria del modelo ARMA para  $y_t$ .

A continuación estimamos este modelo. La estimación puede hacerse por mínimos cuadrados generalizados como vimos en el capítulo anterior. Una vez estimado el modelo y eliminado los coeficientes que no sean significativos calculamos la serie de perturbación  $\hat{n}_t$  mediante diferencia y construimos un modelo univariante para esta serie. Nos encontraremos en uno de los dos casos siguientes:

1. El proceso de perturbación obtenido es estacionario. Entonces, concluimos que las variables están cointegradas. Sustituyendo el modelo univariante identificado para  $n_t$  y reformulando la ecuación obtenida, volveremos a estimarla para obtener estimadores más eficientes de

los parámetros. Después de eliminar los coeficientes que no sean significativos pasaremos a realizar la diagnosis sobre los residuos del modelo como en el capítulo anterior.

2. El proceso  $n_t$  es no estacionario. Entonces, las variables no están cointegradas y deberíamos reestimar el modelo permitiendo un ruido no estacionario. Supongamos que se ha estimado el modelo (18.8) y que el modelo obtenido para  $n_t$  es  $\nabla n_t = \psi_n(B)a_t$ . Tenemos entonces la ecuación siguiente entre variables estacionarias:

$$\nabla y_t = g \nabla x_t + \sum_{i=0}^{k-1} \alpha_i \nabla^2 x_{t-i} + \psi_n(B)a_t,$$

con lo que estaremos en el caso de regresión entre variables estacionarias estudiado en el capítulo anterior.

### 18.3.2. Distinto orden de cointegración

Supongamos que las variables que queremos relacionar tienen distinto orden de cointegración. Tendremos entonces que formular el modelo de manera que se respeten las reglas de adición de procesos estocásticos. Por ejemplo, si la respuesta  $y_t$  es estacionaria y  $x_t$  es no estacionario, forzosamente la perturbación debe de ser no estacionaria. El proceso de construcción del modelo es por lo demás similar.

### 18.3.3. Varios regresores

Este modelo se generaliza sin dificultad para cualquier número de variables explicativas. Con  $p$  variables explicativas el modelo general, utilizando la formulación de función de transferencia, será:

$$y_t = \sum_{i=1}^p \frac{w_i(B)}{\delta_i(B)} x_{i,t} + \frac{\theta(B)\Theta(B^s)}{\phi(B), \Phi(B^s)} a_t,$$

y separando los efectos a corto de a largo e incluyendo la primera diferencia de las variables para distintos retardos:

$$y_t = \sum_{i=1}^p g_i x_{i,t} + \sum_{i=1}^p \sum_{j=0}^{k-1} \alpha_{ij} \nabla x_{i,t-j} + n_t.$$

La construcción del modelo se hace con los mismos principios. Supongamos para simplificar que todas las variables son  $I(1)$ . Es posible que exista cointegración y que el ruido sea estacionario o que no lo sea.

## 18. Regresión entre variables integradas. Cointegración

Comenzaremos permitiendo la presencia de cointegración y supondremos que el ruido es estacionario. Escribiendo las funciones de transferencia separando los efectos a corto y largo plazo estimaremos el modelo y comprobaremos si la perturbación resultante es o no estacionaria. Si lo es, las variables están cointegradas y las variables explicativas son capaces de explicar la tendencia o componente no estacionario de la respuesta. Si la perturbación no es estacionaria, estimaremos el modelo en diferencias de las variables.

### Ejemplo 18.1

Vamos a estudiar la relación entre el índice de la Bolsa de Madrid,  $IB_t$ , y el índice de producción industrial,  $IPI_t$  en el periodo enero de 1988 a diciembre del 2000. Los datos para estas series están en los ficheros indicebolsa98-00.dat y ipies88h00.dat. Los modelos univariantes para estas series se construyeron en el capítulo 16. Como la serie de bolsa no tiene estacionalidad, no puede explicar la del IPI y el ruido tendrá que tener estructura estacional. Vamos a investigar la posible presencia de cointegración entre estas dos variables en la parte regular estimando un modelo que relacione las variables en niveles. El primer modelo que pasamos a estimar es:

$$IPI_t = \sum_{i=0}^k v_i IB_{t-i} + \frac{1}{(1 - \phi B)(1 - \Phi B^{12})} a_t,$$

y tomando  $k = 12$  los resultados obtenidos con el programa EViews aparecen en la tabla 18.1.

Esta estimación permite concluir que: (1) se detectan coeficientes significativos en los retardos 1,2 y 6 meses; (2) la serie del índice de bolsa parece explicar la tendencia del IPI, ya que el término AR regular parece estacionario, pero no la estacionalidad que se refleja en el alto coeficiente para el AR estacional.

Hemos repetido esta estimación con un programa que permite MV exacta, como el SCA, y el valor obtenido para el parámetro AR estacional es  $\Phi = .9999$ , indicando, sin ninguna duda, la necesidad de tomar una diferencia estacional.

Vamos a estimar un modelo que separa los efectos a corto y a largo plazo, del tipo:

$$IPI_t = c + gIB_t + \sum_{i=0}^{k-1} \alpha_i \nabla IB_{t-i} + \frac{1}{(1 - \phi B)(1 - \Phi B^{12})} a_t.$$

Se obtiene el modelo que aparece en la tabla 18.2.

Los coeficientes significativos en la ganancia confirman que existe relación positiva, y también son significativos los de las diferencias 0 y -2 y -5. Los resultados son pues similares con ambos modelos ya que incluir el AR estacional,

Tabla 18.1 Estimación con EViews para el primer modelo

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	91.04032	23.05406	3.948992	0.0001
BOL	0.001698	0.009708	0.174896	0.8615
BOL(-1)	0.024400	0.010912	2.235988	0.0273
BOL(-2)	-0.025248	0.011024	-2.290257	0.0238
BOL(-3)	0.018945	0.010883	1.740823	0.0844
BOL(-4)	0.004047	0.011030	0.366945	0.7143
BOL(-5)	-0.014104	0.011014	-1.280564	0.2029
BOL(-6)	0.035663	0.011623	3.068272	0.0027
BOL(-7)	-0.023455	0.011394	-2.058535	0.0418
BOL(-8)	0.001331	0.011217	0.118657	0.9058
BOL(-9)	0.014358	0.011138	1.289036	0.2000
BOL(-10)	-0.011857	0.011327	-1.046790	0.2974
BOL(-11)	-0.000132	0.012002	-0.010970	0.9913
BOL(-12)	0.006473	0.011105	0.582870	0.5611
AR(1)	0.413770	0.085367	4.846957	0.0000
SAR(12)	0.960982	0.028733	33.44554	0.0000

Tabla 18.2 Estimación con EViews para el segundo modelo

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	90.55505	23.41477	3.867432	0.0002
BOL	0.032981	0.007827	4.213907	0.0001
DBOL	-0.035102	0.012163	-2.886055	0.0047
DBOL(-1)	-0.006306	0.011061	-0.570084	0.5698
DBOL(-2)	-0.030220	0.011596	-2.606123	0.0104
DBOL(-3)	-0.012826	0.011928	-1.075275	0.2845
DBOL(-4)	-0.008515	0.011969	-0.711435	0.4783
DBOL(-5)	-0.023638	0.012135	-1.947948	0.0539
DBOL(-6)	0.013789	0.012266	1.124151	0.2633
DBOL(-7)	-0.012676	0.012254	-1.034413	0.3032
DBOL(-8)	-0.009608	0.012048	-0.797440	0.4269
DBOL(-9)	0.003302	0.012055	0.273954	0.7846
DBOL(-10)	-0.008605	0.012057	-0.713706	0.4769
DBOL(-11)	-0.009977	0.012015	-0.830361	0.4081
DBOL(-12)	-0.010378	0.012266	-0.846105	0.3993
AR(1)	0.414525	0.086275	4.804676	0.0000
SAR(12)	0.961053	0.028748	33.43007	0.0000

## 18. Regresión entre variables integradas. Cointegración

que se estima muy próximo a uno, es como aplicar una diferencia estacional en la ecuación, que introducirá correlaciones entre los regresores en ambas ecuaciones, por lo que no tiene ventajas claras en este caso separar los efectos a corto y a largo. Eliminando los términos no significativos, imponiendo la diferencia estacional y reestimando este último modelo, el resultado es, dejando sólo los coeficientes significativos:

$$IPI_t = .034IB_t - .029\nabla IB_t - .032\nabla IB_{t-2} + \frac{(1 - .88B^{12})}{(1 - .37B)\nabla_{12}} a_t.$$

Si comparamos este modelo con el estimado con la formulación de función de transferencia habitual:

$$IPI_t = .027IB_{t-1} - .027IB_{t-2} + .033IB_{t-3} + \frac{(1 - .88B^{12})}{(1 - .37B)\nabla_{12}} a_t,$$

vemos que las dos ecuaciones son muy parecidas, ya que la primera puede escribirse como:

$$\begin{aligned} .034IB_t - .029IB_t + .029IB_t - .032IB_{t-2} + .032IB_{t-3} &\simeq \\ &\simeq .029IB_t - .032IB_{t-2} + .032IB_{t-3}, \end{aligned}$$

que es similar a la segunda. Para contrastar formalmente si el ruido de la función de transferencia es estacionario, lo calculamos mediante:

$$n_t = IPI_t - \sum_{i=0}^L \hat{v}_i IB_{t-i},$$

y investigamos si es estacionario mediante un contraste de Dickey-Fuller. El resultado tomando 24 retardos es un valor del test de -2.34, que es, sin duda, significativo. Aceptamos, por lo tanto, la necesidad de tomar una diferencia regular y otra estacional en la relación, y pasaremos a estimar modelos con dos diferencias. El modelo obtenido con la estructura:

$$IPI_t = \sum_{i=0}^L v_i IB_{t-i} + \frac{(1 - \Theta B^{12})}{\nabla_{12}\nabla(1 - \phi_1 B - \phi_2 B^2)} a_t,$$

resulta en valores significativos que comienzan en el primer retardo, que indica que la subida de la bolsa en este mes tiene un efecto positivo el mes siguiente, y luego hay otros dos valores significativos. Esto sugiere estimar el modelo:

$$IPI_t = (w_0 + w_1 B + w_2 B^2)B_{t-1} + \frac{(1 - \Theta B^{12})}{\nabla_{12}\nabla(1 - \phi_1 B - \phi_2 B^2)} a_t$$

y los resultados de la estimación se indican en la tabla 16.6. Observemos que los coeficientes de  $w_1$  y  $w_2$  son casi idénticos y de sentido opuesto.

Tabla 18.3 Estimación del modelo de transferencia para el IPI

	$w_0$	$w_1$	$w_2$	$\Theta$	$\phi_1$	$\phi_2$	$\sigma$
coefic.	.021	-.030	0302	.917	-.86	-.56	0.0308
t	3.02	-3.22	4.01	18.94	-11.97	-7.75	

Utilizando los coeficientes de la tabla 16.6 podemos escribir la función de transferencia como:

$$.021 - .03B + 03B^2 = .021 - .03B\nabla$$

que implica la siguiente relación entre las variables

$$IPI_t = .021B_{t-1} - .03\nabla B_{t-2} + \frac{(1 - .91B^{12})}{\nabla_{12}\nabla(1 + .86B + .56B^2)}a_t.$$

Esta ecuación indica que el efecto de una subida en el índice de la Bolsa tiene un efecto que depende de lo que ocurra en los meses siguientes. Una subida de 1 punto en el índice de Bolsa en un mes produce un aumento en el mes siguiente, seguido de una caída y una recuperación del mismo tamaño. Aparentemente hay una sobrereacción a los cambios, que introduce un efecto transitorio que depende de las variaciones del índice. Los residuos del modelo no muestran estructura, aunque aparecen dos valores atípicos e 3/1991 y 4/1997.

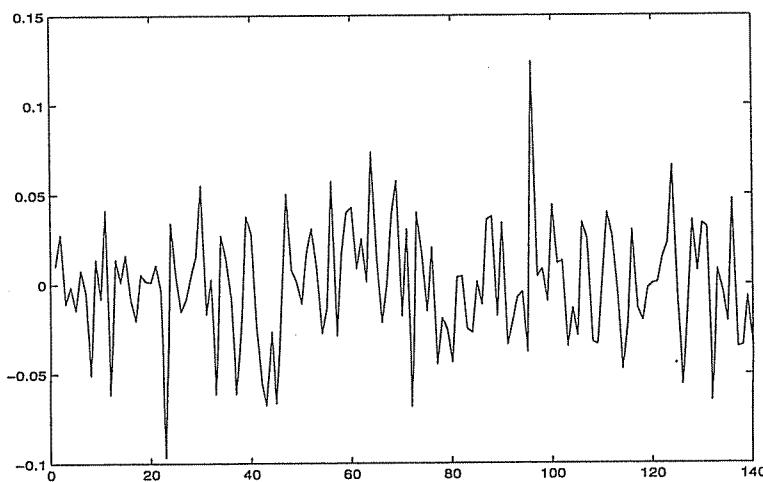
El modelo anterior podría construirse para las variables en logaritmos. Aunque la serie del IPI no muestra necesidad de logaritmos, ya que la variabilidad no parece estar ligada al nivel en el periodo considerado, es esperable que las relaciones entre estas dos variables no dependan de la base de la medida, es decir, de los valores absolutos del índice, que son arbitrarios, y que dependan de los cambios relativos. Esto lleva a estimar el modelo anterior con las variables en logaritmos, con el resultado siguiente proporcionado por EViews:

$$\begin{aligned} \log IPI_t = & .072 \log IB_{t-1} - .088\nabla \log B_{t-2} + \\ & + \frac{(1 - .66B^{12})}{\nabla_{12}\nabla(1 + .84B + .55B^2)}a_t, \end{aligned}$$

que tiene una interpretación similar al modelo anterior. Ahora los coeficientes representan cambios proporcionales y un aumento del índice de Bolsa en un 1 % produce aumentos del IPI de .072 %. Hay, además, un efecto que depende de los rendimientos de la Bolsa y que es negativo: al aumentar los rendimientos en la Bolsa dos meses después se produce una caída en el índice de producción industrial. El gráfico de los residuos  $a_t$  de este modelo se indica en la figura 18.2 y las funciones muestrales de autocorrelación simple y parcial en la figura 18.3.

En el gráfico 18.2 de los residuos se observan algunos valores extremos que

Figura 18.2 Residuos del modelo del IPI y la Bolsa en logaritmos



pueden ser debidos a otras variables distintas de la Bolsa. Para identificarlos, buscamos atípicos en los residuos con el programa SCA obteniendo el modelo final siguiente

$$\begin{aligned} \log IPI_t = & .12 \log IB_{t-1} - .07 \log IB_{t-2} - .039 I_t^{39} + \\ & + .048 I_t^{112} - .031 S_t^{59} + \frac{(1 - .40B^{12})}{\nabla_{12}\nabla(1 + .85B + .52B^2)} a_t. \end{aligned}$$

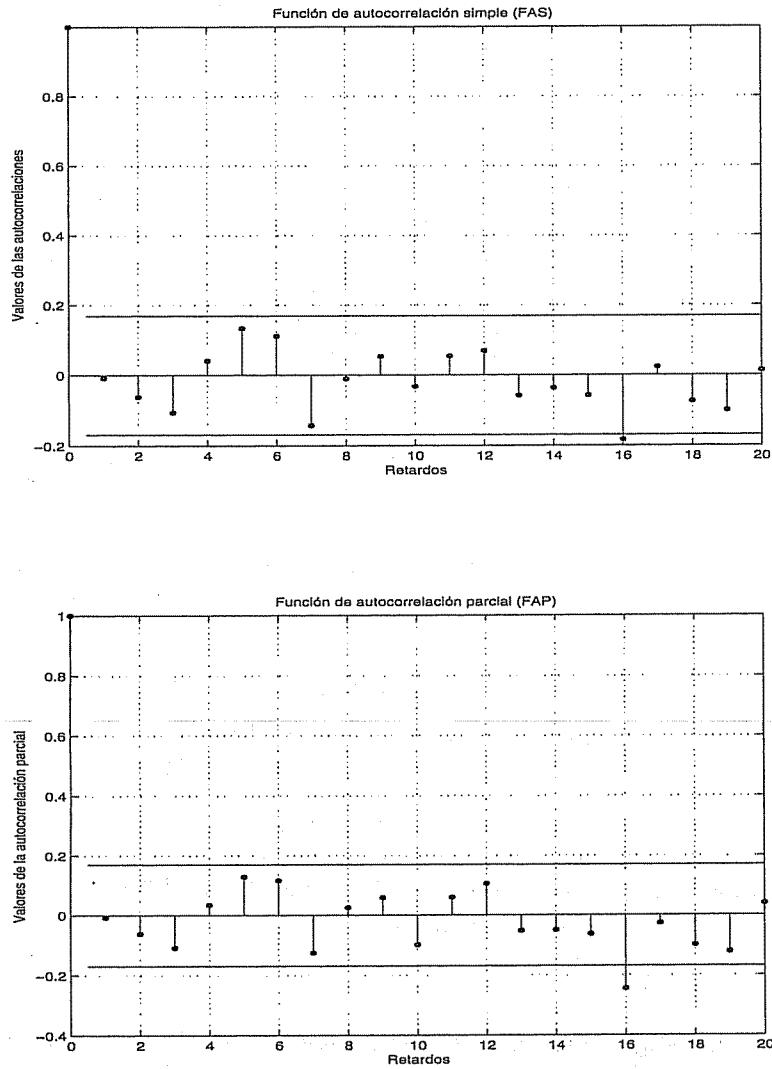
Este modelo indica de nuevo el efecto de sobreajuste en las subidas de la Bolsa sobre el índice de producción industrial y encuentra sucesos atípicos en las observaciones 39 (marzo de 1991), 59 (noviembre de 1992), que supone un cambio de nivel y 112 (abril de 1997).

### Ejemplo 18.2

Vamos a investigar la relación entre el consumo de gasolina y la matriculación de vehículos en el periodo 1980/1999. Buscaremos la relación entre las variables en logaritmos, ya que esperamos efectos proporcionales. Las estacionalidades de las series son diferentes, por lo que no es esperable que una explique totalmente la otra, y al estimar un modelo entre las variables en niveles poniendo un AR(1)<sub>12</sub> estacional el parámetro estacional se estima muy próximo a uno, por lo que tomaremos diferencias estacionales. El modelo estimado a continuación es:

$$\log G_t = 0.100 \log M_t + \sum_{i=1}^3 v_i \log M_{t-i} + \frac{(1 - .88B^{12})}{\nabla_{12}(1 - 0.94B)} a_t, \quad (4.7)$$

Figura 18.3 Correlogramas simple y parcial de los residuos del IPI



donde los coeficientes retardados no aparecen significativos. El alto valor del coeficiente AR sugiere que el ruido es no estacionario y al obtenerlo el gráfico muestra que es no estacionario sin duda, por lo que no es necesario hacer el test. Si estimamos el modelo reparametrizando la función de transferencia, el resultado es similar: sólo el coeficiente inicial es significativo y el coeficiente del AR es muy grande. Tomamos una diferencia regular y estamos en la situación analizada en el ejemplo 17.5 del capítulo anterior.

## 18.4. Lecturas complementarias

Un libro con muchos ejemplos de construcción de funciones de transferencia es Pankratz (1991), aunque no considera la posibilidad de cointegración formalmente. Existe una abundante literatura en econometría sobre la relación de variables cointegradas. En español, recomendamos Espasa y Cancelo (1993) y Novales (1993). En inglés Banerjee et al. (1993), Hamilton (1994), Gourieroux y Monfort (1997) y, para los lectores más avanzados, Davidson (2000).

### Apéndice 18.1: Propiedades de procesos integrados

Vamos a comprobar que si  $x_t$  es un proceso integrado y calculamos su varianza muestral, ésta crece sin límite con  $T$  pero que  $\sum x_t^2/T^2$  converge a una variable aleatoria bien definida. Sea

$$x_t = a_t + a_{t-1} + \dots + a_1,$$

un paseo aleatorio de media cero que suponemos comienza en  $t = 1$ . Entonces:

$$\sum_{t=1}^T x_t^2 = \sum_{t=1}^T (a_t + a_{t-1} + \dots + a_1)^2 = Ta_1^2 + (T-1)a_2^2 + \dots + a_T^2 + DP,$$

donde  $DP$  incluye todos los dobles productos. Tomando esperanzas, como los dobles productos se anulan:

$$E\left(\sum_{t=1}^T x_t^2\right) = \sigma^2(1 + 2 + \dots + T) = \frac{T(T+1)}{2}\sigma^2,$$

es decir:

$$E\left(\frac{1}{T^2} \sum_{t=1}^T x_t^2\right) = \frac{\sigma^2}{2} + \frac{1}{2T}\sigma^2.$$

Este resultado nos indica que para  $T$  grande la variable  $y_T = \sum_{t=1}^T x_t^2/T^2$  tiene una media finita,  $\sigma^2/2$ . Puede comprobarse que esta variable tiene varianza  $\sigma^2/3$  y que su distribución puede calcularse y está bien definida. Véase Hamilton (1994).

## Apéndice 18.2: Estimación con EVViews

Como ilustración vamos a indicar cómo realizar las estimaciones del ejemplo 18.1 con este programa. Comenzaremos creando un workfile (*File → New → Workfile*) en el periodo 1/1988 a 12/2000 (*Monthly, Start date: 1988:1, End date: 2001:12*). Se importan los datos con *File → Import → Read Text-Lotus-Excel*, desde los ficheros. Llamando ipi a la serie del ipi, la serie con una diferencia regular y otra estacional, se obtiene con: *Genr → ddipi=d(ipi,1;12)*. Para ver el gráfico: *View→Line Graph*, y el correlograma: *View→Correlogram*.

Vamos a estimar un modelo ARIMA(2,1,0)  $\times$  (0,1,1)<sub>12</sub> con un atípico aditivo en el dato 4/1997. Se crea la variable ficticia para el atípico aditivo: *Genr → ao1=0; Genr → ao1=1, sample:1997:4 1997:4*, y se estima el modelo:

*Quick → Estimate Equation :*  
*d(ipi,1,12) ar(1) ar(2) sma(12) d(ao1,1,12)*

Se obtiene el gráfico de los residuos mediante: *View→Actual, Fitted, Residual→Actual, Fitted, Residual Graph* y su correlograma con: *View→Residual tests→Correlogram Q statistics*. Para realizar predicciones con este modelo para el año 2001, se hace:

*Forecast → IPI, Forecast name : IPIF, S.E. : se,*  
*Sample range to forecast : 2001 : 1 2001 : 12*

De esta manera tenemos guardada la predicción junto con los errores estándar. Con ellos podemos obtener los intervalos de confianza para la predicción como sigue: *Genr → icinf=ipif-2\*se; Genr → icsup=ipif+2\*se*. Las series icinf e icsup forman los extremos inferiores y superiores de los intervalos de confianza.

Pasamos a estimar el modelo de la función de transferencia. Supondremos una estructura simple AR, AR(12) para el ruido de la ecuación y estimamos:

*Quick → Estimate Equation → ipi c bolsa(0 to -10) ar(1) sar(12)*

Como el parámetro del AR estacional es muy próximo a uno, tomaremos una diferencia estacional. Para ello, definimos una nueva serie: *Genr → dbolsa=d(bolsa,0,12)* y estimamos el modelo, donde hemos puesto una estructura simple para el ruido:

*Procs → Especify/Estimate →*  
*d(ipi,0,12) c dbolsa(0 to -10) ar(1) sma(12)*

Comprobamos que la constante ya no es significativa, con lo que la eliminamos de la estimación. Como el ruido de esta ecuación es no estacionario pasamos a añadir una diferencia regular a la relación. Para ello definimos:

## 18. Regresión entre variables integradas. Cointegración

*Genr → ddbolsa=d(bolsa,1,12), y estimamos el modelo Procs → Especify/Estimate → d(ipi,1,12) ddbolsa(0 to -10) ar(1) sma(12).*

Si queremos separar los efectos a corto y a largo podemos definir: *Genr → dddbolsa=d(ddbolsa)* y estimar el modelo *Procs → Especify/Estimate → d(ipi,1,12) ddbolsa dddbolsa(0 to -10) ar(1) sma(12)*. El modelo final estimado es *Procs → Especify/Estimate → d(ipi,1,12) ddbolsa(-1 to -3) ar(1) ar(2) sma(12)*. Para obtener predicciones de este modelo escribiremos como antes *Forecast → IPI, Forecastname:IPIF2, S.E.:se2, Sample range to forecast: 2001:1 2001:12*. Si comparamos las series se y se2 para los valores del año 2001, comprobamos que el modelo de regresión dinámica tiene menor error estándar en la predicción. Podemos obtener los intervalos de confianza para la predicción como en el caso univariante: *Genr → icinf2=ipif2-2\*se2; Genr → icsup2 = ipif2 + 2\*se2*. Las series icinf2 e icsup2 forman los extremos inferiores y superiores de los intervalos de confianza.

---

# 19. Modelos multivariantes



Maurice K. Quenouille (1924-1973)

Estadístico inglés. Fue profesor en Cambridge, Oxford, Imperial College y Southampton. Pensador de gran originalidad, inventó el procedimiento para reducir el sesgo de un estimador conocido como jacknife y fue pionero en el análisis de series temporales multivariantes. Murió trágicamente en un accidente en Portugal.

## 19.1. Introducción

En este capítulo vamos a estudiar cómo modelar conjuntamente las relaciones entre los componentes de un vector de series. Veremos que la metodología a utilizar es una generalización de la univariante y que podemos utilizar herramientas similares a las estudiadas para identificar el modelo, estimarlo y realizar contrastes diagnósticos.

Este capítulo es una introducción breve a un tema muy amplio y su objetivo es ofrecer al lector unas herramientas simples y útiles en la práctica para explorar la estructura de dependencia de un conjunto de series e introducirle en un campo de rápido crecimiento en la actualidad.

## 19.2. Procesos vectoriales estacionarios y sus autocorrelaciones

### 19.2.1. Procesos vectoriales estacionarios

Dado un vector de  $k$  series temporales  $Z_t = (z_{1t}, \dots, z_{kt})'$  definiremos su esperanza como el vector obtenido tomando las esperanzas en cada uno de los componentes. Diremos que este vector de series, o este proceso vectorial, es estacionario si el vector de medias:

$$E(\mathbf{Z}_t) = \boldsymbol{\mu}$$

es constante, y sus matrices de autocovarianzas, definidas por:

$$\begin{aligned}\Gamma(m) &= E[(\mathbf{Z}_{t-m} - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})'] = \\ &= E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+m} - \boldsymbol{\mu})'] = \{\gamma_{ij}(m)\},\end{aligned}$$

sólo dependen del retardo. Hemos llamado  $\gamma_{ij}(m)$  al elemento  $ij$  de la matriz  $\Gamma(m)$  de autocovarianzas de orden  $m$ , que se obtiene de:

$$\begin{aligned}\gamma_{ij}(m) &= E[(z_{it} - \mu_i)(z_{jt+m} - \mu_j)] \neq \\ &E[(z_{jt} - \mu_j)(z_{it+m} - \mu_i)] = \gamma_{ji}(m).\end{aligned}$$

Por ejemplo, para un vector bivariante,  $\mathbf{Z}_t = (z_{1t}, z_{2t})'$  las matrices  $\Gamma(m)$  son cuadradas de orden dos y contienen en la diagonal la autocovarianza de orden  $m$  de cada componente y, fuera de la diagonal, las covarianzas cruzadas de orden  $m$  que estudiamos en el capítulo 17. El término  $\gamma_{12}(m)$  de esta matriz mide el efecto lineal de la primera componente en  $t$  sobre la segunda en  $t + m$ . Análogamente,  $\gamma_{21}(m)$  mide el efecto lineal de la segunda componente sobre la primera con retardo  $m$ . La covarianza cruzada  $\gamma_{ij}(m)$  representa como afecta el componente  $z_i$  hoy al componente  $z_j$  en  $m$  periodos posteriores, es decir mide la relación que va de la variable  $z_{it}$  a la  $z_{jt+m}$ , mientras que  $\gamma_{ji}(m)$  mide la relación de causalidad del componente  $j$  sobre el  $i$ , que puede ser muy diferente. Concluimos que las matrices  $\Gamma(m)$  incluyen las covarianzas cruzadas entre los componentes del vector y no son, en general, simétricas para  $m \neq 0$ . En particular, la matriz  $\Gamma(0)$  sí es simétrica, ya que representa las varianzas de los componentes y las covarianzas instantáneas, y será además definida no negativa. Para las restantes matrices sus términos verifican, como vimos en el capítulo 17, que:

$$\gamma_{ij}(m) = \gamma_{ji}(-m),$$

por lo que sólo es necesario calcular las matrices  $\Gamma(m)$  para retardos positivos, ya que existe la relación:

$$\Gamma(-m) = \Gamma'(\eta m).$$

Se define el proceso de ruido blanco vectorial  $\mathbf{a}_t$  como un proceso de media cero de variables incorreladas con matriz de varianzas y covarianzas instantáneas,  $\Sigma$ , definida positiva. Observemos que esta definición permite correlaciones contemporáneas entre los componentes del vector de ruido blanco, pero que no pueden existir ni autocorrelaciones

en un componente, ni correlaciones cruzadas entre dos que no sean instantáneas. Si, además, las variables  $a_t$  siguen un proceso normal multivariante, la incorrelación garantiza la independencia de las variables y hablaremos de un proceso de ruido blanco normal.

### 19.2.2. Matrices de autocorrelación simple y parcial

Las matrices de autocorrelación del proceso vectorial se definen por:

$$\mathbf{R}(m) = \{\rho_{ij}(m)\},$$

donde los términos de esta matriz vienen dados por:

$$\rho_{ij}(m) = \frac{\gamma_{ij}(m)}{\sqrt{\gamma_{ii}(0) \gamma_{jj}(0)}}.$$

Las matrices de autocorrelación generalizan las autocorrelaciones al caso vectorial y veremos que juegan un papel similar en la construcción del modelo. De la misma forma, podemos introducir las matrices de autocorrelación parcial del proceso. Definimos la matriz de autocorrelación parcial de orden  $m$  de un proceso estacionario vectorial:

$$\tilde{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu},$$

de esperanza cero, como la matriz  $\mathbf{P}(m)$  obtenida ajustando por mínimos cuadrados el modelo lineal *multivariante*:

$$\tilde{\mathbf{Z}}_t = \Phi_1 \tilde{\mathbf{Z}}_{t-1} + \dots + \Phi_{m-1} \tilde{\mathbf{Z}}_{t-m-1} + \mathbf{P}(m) \tilde{\mathbf{Z}}_{t-m} + \mathbf{u}_t. \quad (19.1)$$

En este modelo cada serie se explica por una regresión sobre sus  $m$  valores pasados y los  $m$  pasados de las otras series. Como en el caso escalar,  $\mathbf{P}(m)$  representa la relación lineal existente entre el vector  $\tilde{\mathbf{Z}}_t$  y el vector  $\tilde{\mathbf{Z}}_{t-m}$  cuando eliminamos de ambos el efecto de las variables explicativas intermedias,  $\tilde{\mathbf{Z}}_{t-1}, \dots, \tilde{\mathbf{Z}}_{t-m-1}$ . La forma de (19.1) es la de un modelo lineal de regresión multivariante. Por ejemplo, tomemos  $m = 1$ . La primera fila de la matriz  $\mathbf{P}(1)$ , que denotaremos por  $\mathbf{p}'_1(1)$ , se obtiene estimando por mínimos cuadrados la regresión múltiple:

$$\tilde{z}_{1t} = \mathbf{p}'_1(1) \tilde{\mathbf{Z}}_{t-1} + u_{1t},$$

que conduce al resultado:

$$\mathbf{p}_1(1) = E(\tilde{\mathbf{Z}}_{t-1} \tilde{\mathbf{Z}}'_{t-1})^{-1} E(\tilde{\mathbf{Z}}_{t-1} \tilde{z}_{1t}) = \boldsymbol{\Gamma}(0)^{-1} \boldsymbol{\gamma}_1(1),$$

donde  $\boldsymbol{\gamma}_1(1)$  es el vector de covarianzas cruzadas entre el primer compo-

nente y todos los demás retardados un periodo, y coincide con la primera columna de la matriz  $\Gamma(1)$ . En consecuencia, para todas las columnas:

$$\mathbf{P}(1)' = \boldsymbol{\Gamma}(0)^{-1} \boldsymbol{\Gamma}(1),$$

que es similar a la ecuación para estimar el parámetro en un AR(1), pero sustituyendo las varianzas y autocovarianzas de orden uno por las matrices de autocovarianzas.

### 19.2.3. Estimación de las matrices simples y parciales

Los elementos de las matrices de autocorrelación se estiman de la forma ya estudiada, mediante:

$$\hat{\rho}_{ij}(m) = \frac{\sum_{t=1}^{T-m} (z_{it} - \bar{z}_i)(z_{jt+m} - \bar{z}_j)}{\sqrt{\sum_{t=1}^T (z_{it} - \bar{z}_i)^2 \sum_{t=1}^T (z_{jt} - \bar{z}_j)^2}},$$

donde  $\bar{z}_i = T^{-1} \sum_{t=1}^T z_{it}$ . Como en el caso univariante, se demuestra que si tenemos un proceso de ruido blanco vectorial, los coeficientes estimados  $\hat{\rho}_{ij}(m)$ , para  $m > 0$ , tienen una distribución asintóticamente normal, con media cero y varianza  $T^{-1}$ . Por lo tanto, consideraremos no significativamente distintos de cero aquellos coeficientes menores, en valor absoluto, que  $2/\sqrt{T}$ .

Las matrices de correlación parcial se obtienen estimando por mínimos cuadrados las ecuaciones (19.1), donde  $\tilde{\mathbf{Z}}_t = \mathbf{Z}_t - \bar{\mathbf{Z}}$ , se obtiene sustituyendo las medias por sus estimaciones muestrales. Llamando  $\hat{\mathbf{P}}(m)$  a las estimaciones de estas matrices, podemos construir un contraste de que las matrices poblacionales son iguales a cero como sigue. Se comienza con  $\mathbf{P}(1)$  y se contrasta si es igual de cero. Si no lo es, se contrasta que  $\mathbf{P}(2)$  es igual a cero, condicionada a que  $\mathbf{P}(1)$  es distinta de cero, y así sucesivamente. Vamos a estudiar este contraste. Sea:

$$\mathbf{e}_{t,m} = \tilde{\mathbf{Z}}_t - \hat{\Phi}_1 \tilde{\mathbf{Z}}_{t-1} - \dots - \hat{\Phi}_{m-1} \tilde{\mathbf{Z}}_{t-m+1} - \hat{\mathbf{P}}(m) \tilde{\mathbf{Z}}_{t-m},$$

los residuos del ajuste al estimar el coeficiente de autocorrelación parcial de orden  $m$ , y

$$\mathbf{S}(m) = \sum_{t=m+1}^T \mathbf{e}_{t,m} \mathbf{e}_{t,m}',$$

la matriz de suma de cuadrados de estos residuos. Se demuestra que el contraste de razón de verosimilitud de que la matriz de correlación parcial de orden  $m$  es nula, condicionada a que esta matriz de orden  $m-1$  no lo sea, depende de la matriz de varianzas de los residuos estimados

mediante:

$$\Lambda = \frac{|\mathbf{S}(m)|}{|\mathbf{S}(m-1)|},$$

donde  $\mathbf{S}(m)$  es la matriz de sumas de cuadrados de los residuos al ajustar un modelo de regresión multivariante con  $m$  regresores y  $\mathbf{S}(m-1)$  la obtenida si el orden máximo incluido es  $m-1$ . El contraste analiza si la reducción en la suma de cuadrados de los residuos es significativa. Bajo la hipótesis de que  $\mathbf{P}(m)$  es cero, la distribución de este estadístico puede aproximarse a una  $\chi^2$  central mediante la transformación:

$$\chi^2 = - \left( T - \frac{1}{2} - mk \right) \log \Lambda,$$

que se distribuye asintóticamente como una  $\chi^2$  con  $k^2$  grados de libertad, el número de términos en la matriz  $\widehat{\mathbf{P}}(m)$ , lo que nos permite hacer un contraste de que esta matriz es cero.

### 19.3. Modelos ARMA vectoriales

La generalización multivariante del teorema de Wold nos indica que todo proceso vectorial estacionario sin componentes deterministas admite una representación:

$$\tilde{\mathbf{Z}}_t = \Psi(\mathbf{B}) \mathbf{a}_t = \sum_{i=1}^{\infty} \Psi_i \mathbf{a}_{t-i}, \quad (19.2)$$

donde  $\tilde{\mathbf{Z}}_t = \mathbf{Z}_t - \mu$ , siendo  $\mu$  el vector de medias de  $\mathbf{Z}_t$ ,  $B$  el operador de retardo, que aplicado a vectores se define por  $B \mathbf{Z}_t = \mathbf{Z}_{t-1}$ ,  $\Psi_i$  son matrices cuadradas  $k \times k$ , con  $\Psi_0 = \mathbf{I}$ , la matriz unidad, y  $\mathbf{a}_t$  un vector de ruido blanco de media cero y matriz de covarianzas  $\Sigma$  definida positiva.

Análogamente al caso univariante, los modelos ARMA vectoriales surgen como representaciones escuetas del operador infinito  $\Psi(\mathbf{B})$ . Supondremos que esta matriz puede aproximarse por el producto de las matrices finitas:

$$\tilde{\mathbf{Z}}_t = \Phi_p(B)^{-1} \Theta_q(B) \mathbf{a}_t \quad (19.3)$$

donde:

$$\Phi_p(B) = \Phi_0 - \Phi_1 B - \dots - \Phi_p B^p$$

es un operador en  $B$  y los términos  $\Phi_i$  matrices cuadradas de orden  $k$  que contienen los coeficientes autoregresivos del proceso. Se demuestra que las raíces de la ecuación  $|\Phi_p(B)| = 0$  deben de estar fuera del círculo unidad para que el proceso sea estacionario. El operador de media móvil

es:

$$\Theta_q(B) = \Theta_0 - \Theta_1 B - \dots - \Theta_q B^q$$

y los ceros de la ecuación  $|\Theta_q(B)| = 0$  deben estar fuera del círculo unidad para que el proceso sea invertible. Con esto se obtiene la representación ARMA multivariante:

$$\Phi_p(B) \tilde{Z}_t = \Theta_q(B) a_t. \quad (19.4)$$

Las matrices de un modelo ARMA vectorial deben cumplir determinadas condiciones para que el modelo sea identificable, es decir, exista una equivalencia uno a uno entre la representación (19.4) y la función de covarianzas del proceso. Por ejemplo los procesos:

$$\tilde{Z}_t = a_t - \Theta_1 a_{t-1}$$

y:

$$\tilde{Z}_t = M u_t - \Theta_1^* u_{t-1}$$

donde  $a_t = M u_t$ ,  $\Theta_1^* = \Theta_1 M^{-1}$  son indistinguibles para cualquier matriz  $M$  no singular. Es necesario introducir condiciones para definir la estructura ARMA vectorial de manera única. Las condiciones necesarias y suficientes han sido obtenidas por Hannan (1969). En particular, el proceso es identificable si:

- a.  $\Phi_0 = \Theta_0 = I$  (matriz unidad) y  $\Sigma$  cualquiera
- b.  $\Phi_0$  cualquiera,  $\Theta_0 = I$  y  $\Sigma$  diagonal
- c.  $\Theta_0$  cualquiera,  $\Phi_0 = I$  y  $\Sigma$  diagonal
- d.  $\Sigma = I$  y  $\Phi_0$  y  $\Theta_0$  cualquiera.

Estas formulaciones difieren en dónde establecen la correlación contemporánea: en el caso (a) se pone en  $\Sigma$ ; en el caso (b) en la matriz  $\Phi_0$ ; en el caso (c) en la matriz  $\Theta_0$ ; y en el caso (d) se permite en ambas matrices  $\Phi_0$  y  $\Theta_0$ , pero se exige una condición de normalización,  $\Sigma = I$ .

En resumen, si permitimos generalidad en la matriz de covarianzas del ruido,  $\Sigma$ , entonces las matrices  $\Phi_0$  y  $\Theta_0$  deben ser unitarias, mientras que si restringimos la matriz de covarianzas a ser diagonal, una de las matrices  $\Phi_0$  o  $\Theta_0$  puede ser cualquiera. En adelante supondremos, por analogía con el caso univariante, que  $\Phi_0 = \Theta_0 = I$  y que la matriz  $\Sigma$  es definida positiva. A continuación estudiamos algunos casos particulares de estos procesos.

### 19.3.1. Proceso VAR(1)

Se define el proceso VAR(1), o autorregresivo vectorial de orden uno, mediante:

$$Z_t - \mu = \Phi(Z_{t-1} - \mu) + a_t.$$

En adelante, para simplificar la exposición, supondremos  $k = 2$ , y  $\mu = 0$ . Entonces, podemos escribir el proceso como:

$$\left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} B \right) \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

Esta ecuación matricial da lugar a las dos ecuaciones:

$$z_{1t} = \phi_{11}z_{1t-1} + \phi_{12}z_{2t-1} + a_{1t}$$

y

$$z_{2t} = \phi_{21}z_{1t-1} + \phi_{22}z_{2t-1} + a_{2t}.$$

La condición de estacionaridad es que los ceros de la ecuación:

$$|I - \Phi B| = 0$$

estén fuera del círculo unidad. Como:

$$|I - \Phi B| = -B |\Phi - B^{-1}I| = -B |\Phi - \lambda I| = 0$$

con  $\lambda = B^{-1}$ , tenemos que el proceso será estacionario si los valores propios de la matriz  $\Phi$  están dentro del círculo unidad. En particular, si  $\Phi$  es diagonal, tenemos dos procesos univariantes AR(1) con ruidos correlados y el proceso será estacionario si  $|\phi_{ii}| < 1$ , que es la misma condición que en el caso univariante.

El modelo VAR(1) determina los modelos univariantes de las series. Escribiendo:

$$Z_t = (I - \Phi B)^{-1} a_t,$$

y utilizando que la inversa de una matriz se calcula obteniendo la matriz adjunta y dividiendo por el determinante de la matriz, tenemos que, para  $k = 2$ :

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \frac{1}{a(B)} \begin{bmatrix} 1 - \phi_{22}B & \phi_{12}B \\ \phi_{21}B & 1 - \phi_{11}B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

donde  $a(B) = ((1 - \phi_{11}B)(1 - \phi_{22}B) - \phi_{12}\phi_{21}B^2)$  que conduce a la siguiente ecuación para, por ejemplo,  $z_{1t}$ :

$$((1 - \phi_{11}B)(1 - \phi_{22}B) - \phi_{12}\phi_{21}B^2)z_{1t} = (1 - \phi_{22}B)a_{1t} + \phi_{12}a_{2t-1}.$$

Vemos que cada componente tiene la misma estructura AR, que será un polinomio de orden igual al número de componentes del vector: AR(2) en este caso y un AR( $k$ ) en general. La parte MA de cada componente será la suma de un ruido blanco y tantos procesos MA como orden tenga el proceso menos uno. Como la suma de procesos MA es un nuevo proceso MA de orden igual al máximo de los órdenes de los sumandos, en el caso bivariante la parte de la izquierda será MA(1) y los componentes del proceso bivariante serán ARMA(2,1). Este es el orden máximo del proceso, porque es posible que exista cancelación entre los operadores y el modelo univariante sea de orden menor.

En el caso en que uno de los coeficientes  $\phi_{ij}, i \neq j$ , sea cero, por ejemplo el  $\phi_{12} = 0$ , y  $\Sigma$  sea diagonal, entonces no existe relación desde el componente dos hacia el uno y tenemos la situación estudiada en los capítulos anteriores de *causalidad unidireccional*, o *exogeneidad*. La variable  $z_{1t}$  se denomina entonces exógena y la variable  $z_{2t}$  es la dependiente o endógena, y el modelo VAR(1) puede escribirse como:

$$(1 - \phi_{11}B)z_{1t} = a_{1t} \quad (19.5)$$

$$z_{2t} = \frac{\phi_{21}}{1 - \phi_{22}B}z_{1t-1} + \frac{a_{2t}}{1 - \phi_{22}B}. \quad (19.6)$$

Vemos que el modelo VAR proporciona tanto el modelo de regresión dinámica entre las dos variables como el modelo univariante para la variable exógena.

Las funciones de autocovarianzas de un VAR(1) se obtienen con:

$$\begin{aligned} \Gamma(1) &= E[(Z_t - \mu)(Z_{t+1} - \mu)'] = \\ &= E[(Z_t - \mu)(\Phi(Z_t - \mu) + a_t)'] = \Gamma(0)\Phi' \end{aligned}$$

y aplicando la definición se obtiene:

$$\Gamma(m) = \Gamma(m-1)\Phi',$$

que es la generalización del resultado univariante. De la definición de matrices de correlación parcial es inmediato que este proceso sólo tendrá la matriz de orden uno distinta de cero.

### Ejemplo 19.1

Supongamos el modelo VAR(1) bivariante donde  $\phi_{12} = 0$  pero la matriz  $\Sigma$  no es diagonal. Entonces, las ecuaciones (19.5) y (19.6) no forman un modelo de función de transferencia, porque los ruidos de las dos ecuaciones están correlados. Para obtener la expresión de la función de transferencia, partimos de la ecuación de regresión que relaciona ambos ruidos:

$$a_{2t} = \beta a_{1t} + u_t,$$

y sustituyendo primero  $a_{2t}$  por esta expresión en la ecuación (19.6) y segundo  $a_{1t}$  por su expresión (19.5), se obtiene:

$$z_{2t} = \frac{\phi_{21}}{1 - \phi_{22}B} z_{1t-1} + \frac{\beta (1 - \phi_{11}B) z_{1t}}{(1 - \phi_{22}B)} + \frac{u_t}{1 - \phi_{22}B}$$

que puede escribirse como:

$$z_{2t} = \frac{\nu_0 + \nu_1 B}{1 - \phi_{22}B} z_{1t} + \frac{u_t}{1 - \phi_{22}B}$$

donde  $\nu_0 = \beta$  y  $\nu_1 = \phi_{21} - \phi_{11}$ . Como los ruidos de esta ecuación y la (19.5) están incorrelados, estas dos ecuaciones definen la función de transferencia.

---

#### 19.3.2. Procesos VAR(p)

Los resultados anteriores se generalizan para un VAR(p). El modelo es:

$$\mathbf{Z}_t - \boldsymbol{\mu} = \Phi_1(\mathbf{Z}_{t-1} - \boldsymbol{\mu}) + \dots + \Phi_p(\mathbf{Z}_{t-p} - \boldsymbol{\mu}) + \mathbf{a}_t$$

y diremos que es estacionario si las raíces de la ecuación  $|\Phi(B)| = 0$  están fuera del círculo unidad. Tomando la traspuesta de esta ecuación y multiplicando por la derecha por  $\mathbf{Z}_{t-m} - \boldsymbol{\mu}$ , se obtiene:

$$\begin{aligned} (\mathbf{Z}_{t-m} - \boldsymbol{\mu})(\mathbf{Z}_t - \boldsymbol{\mu})' &= (\mathbf{Z}_{t-m} - \boldsymbol{\mu})(\mathbf{Z}_{t-1} - \boldsymbol{\mu})'\Phi_1' + \dots + \\ &\quad + (\mathbf{Z}_{t-m} - \boldsymbol{\mu})(\mathbf{Z}_{t-p} - \boldsymbol{\mu})'\Phi_p' + (\mathbf{Z}_{t-m} - \boldsymbol{\mu})\mathbf{a}_t', \end{aligned}$$

y tomando esperanzas, concluimos que las matrices de correlación verifican la relación:

$$\Gamma(m) = \sum_{i=1}^p \Gamma(m-i) \Phi_i' \quad m = 1, 2, 3, \dots \quad (19.7)$$

donde  $\Gamma(m) = -\Gamma'(-m)$ .

Por ejemplo, para un AR(2), tendremos:

$$\Gamma(m) = \Gamma(m-1)\Phi'_1 + \Gamma(m-2)\Phi'_2 \quad m = 1, 2, 3, \dots$$

Para un proceso AR general las matrices  $\Gamma(m)$  serán distintas de cero para cualquier retardo, y su relación de dependencia temporal es la generalización natural de la que estudiamos para los procesos AR univariantes. Análogamente, es inmediato comprobar que un proceso VAR( $p$ ) tendrá las  $p$  primeras matrices de correlación parcial distintas de cero y el resto serán nulas.

### 19.3.3. Procesos VMA( $q$ )

La expresión de un modelo de media móvil vectorial es:

$$\mathbf{Z}_t = (\mathbf{I} - \Theta_1 B - \dots - \Theta_q B^q) \mathbf{a}_t.$$

Por ejemplo, si  $k = 2$  y  $q = 1$ , el modelo MA(1) bivariante es:

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 1 - \theta_{11}B & -\theta_{12}B \\ -\theta_{21}B & 1 - \theta_{22}B \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

y cada serie seguirá un modelo univariante que es MA(1), ya que será la suma de un ruido blanco y un proceso MA(1). La condición de invertibilidad es que los ceros de la ecuación  $|\mathbf{I} - \Theta B|$  estén fuera del círculo unidad, y es fácil comprobar que esta condición equivale para el VMA(1) a que los autovalores de la matriz  $\Theta$  estén dentro del círculo unidad.

Las matrices de autocovarianzas se calculan con:

$$\begin{aligned} \Gamma(m) &= E[(\mathbf{Z}_t - \mu)(\mathbf{Z}_{t+m} - \mu)'] = \\ &= E[(\mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1} - \dots - \Theta_q \mathbf{a}_{t-q}) \times \\ &\quad \times (\mathbf{a}'_{t+m} - \Theta'_1 \mathbf{a}'_{t+m-1} - \dots - \Theta'_q \mathbf{a}'_{t+m-q} \Theta'_q)] \end{aligned}$$

que conducirá al resultado:

$$\begin{aligned} \Gamma(m) &= -\Sigma \Theta'_m + \Theta_1 \Sigma \Theta'_{m+1} + \dots + \Theta_{q-m} \Sigma \Theta'_q \quad m \leq q \\ \Gamma(m) &= 0 \quad m > q. \end{aligned} \tag{19.8}$$

Por lo tanto, un proceso MA multivariante tendrá únicamente  $q$  matrices de autocovarianzas distintas de cero. De nuevo encontramos el paralelismo entre los modelos univariantes y los multivariantes, que va a darse también para las matrices de autocorrelación parcial. Por ejemplo, consideremos el caso más simple MA(1). En la hipótesis de que  $\mathbf{Z}_t$  es

invertible, la serie seguirá el proceso:

$$(\mathbf{I} - \Theta B)^{-1} \mathbf{Z}_t = \mathbf{a}_t,$$

y utilizando la serie de Neumann del operador inverso, dada por:

$$(\mathbf{I} - \Theta B)^{-1} = \mathbf{I} + \Theta B + \Theta^2 B^2 + \dots,$$

el VMA(1) puede también escribirse:

$$\mathbf{Z}_t = -\Theta \mathbf{Z}_{t-1} - \Theta^2 \mathbf{Z}_{t-2} - \Theta^3 \mathbf{Z}_{t-3} - \dots + \mathbf{a}_t.$$

Este es un proceso VAR( $\infty$ ) y hemos generalizado el resultado de que un proceso MA equivale a un AR infinito. En consecuencia, sus matrices de autocorrelación parcial,  $\mathbf{P}(m)$  serán distintas de cero para cualquier retardo.

#### 19.3.4. Procesos VARMA(p,q)

Hemos visto que las propiedades de los procesos ARMA univariantes se generalizan inmediatamente para sus representaciones multivariantes. Un proceso vectorial VAR( $p$ ) tiene únicamente  $p$  matrices de autocorrelación parcial distintas de cero y un proceso VMA( $q$ ) tiene  $q$  matrices de autocorrelación simple distintas de cero. En el caso de procesos mixtos VARMA estas matrices, igual que en el caso univariante, no tienen un corte en el orden del proceso.

Vamos a obtener la función de covarianzas para el proceso ARMA general:

$$\mathbf{Z}_t = \Phi_1 \mathbf{Z}_{t-1} + \dots + \Phi_p \mathbf{Z}_{t-p} + \mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1} - \dots - \Theta_q \mathbf{a}_{t-q}, \quad (19.9)$$

donde  $\mathbf{a}_t$  es un proceso de ruido blanco con matriz de covarianzas  $\Sigma$ . Transponiendo esta ecuación, premultiplicando el resultado por  $\mathbf{Z}_{t-m}$  y tomando esperanzas, tenemos que:

$$\Gamma(m) = \sum_{i=1}^p \Gamma(m-i) \Phi'_i - \sum_{j=0}^q E(\mathbf{z}_{t-m} \mathbf{a}'_{t-j} \Theta'_j) \quad \text{con } \Theta_0 = -\mathbf{I}. \quad (19.10)$$

Para calcular el segundo miembro, observemos que, al ser el proceso estacionario, puede escribirse:

$$\mathbf{Z}_t = \Phi(B)^{-1} \Theta(B) \mathbf{a}_t = \Psi(B) \mathbf{a}_t \quad (19.11)$$

que implica:

$$E(\mathbf{Z}_{t-m}\mathbf{a}'_{t-j}) = E(\Psi(B)\mathbf{a}_{t-m}\mathbf{a}'_{t-j}) = \begin{cases} 0 & j < m \\ \Sigma & j = m \\ \Psi_{j-m}\Sigma & j > m \end{cases}. \quad (19.12)$$

Sustituyendo en la expresión (19.10), obtenemos que:

$$\begin{aligned} \Gamma(m) &= \sum_{i=1}^p \Gamma(m-i)\Phi'_i - \sum_{j=1}^q \Psi_{j-m}\Sigma\Theta'_j, \quad 0 \leq m \leq q \\ \Gamma(m) &= \sum_{i=1}^p \Gamma(m-i)\Phi'_i, \quad m > q, \end{aligned}$$

que generaliza los resultados anteriores para modelos VAR y VMA.

#### 19.4. Construcción de modelos VARMA para series estacionarias

Dado un vector de  $k$  series estacionarias  $\mathbf{Z}_t$ , se quiere construir un modelo VARMA que describa su estructura de dependencias dinámica. Supondremos que, a priori, no especificamos que algunas variables son endógenas y otras exógenas y que se permite, en consecuencia, la existencia de re-alimentación entre todas ellas. El método que se propone a continuación es similar al caso univariante y consta de las etapas siguientes:

1. Identificar la estructura del modelo, entendiendo por ello el orden del modelo VARMA estacionario.
2. Estimar los parámetros mediante máxima verosimilitud y seleccionar el mejor modelo entre los estimados.
3. Realizar contrastes diagnósticos para comprobar si el modelo seleccionado no tiene deficiencias detectables.

##### 19.4.1. Identificación

El objetivo de esta etapa es determinar un conjunto de modelos posibles que pasaremos después a estimar. Cuando tengamos muchas series, la identificación requiere, en primer lugar, comprobar que la dimensión del sistema es realmente  $k$ , es decir, que tenemos  $k$  series independientes y que ninguna de ellas se ha construido como combinación lineal de las demás. Por ejemplo, si tenemos un sistema trivariante con ingresos,  $z_{1t}$ ,

costes,  $z_{2t}$ , y beneficios,  $z_{3t} = z_{1t} - z_{2t}$ , no tenemos realmente tres series temporales, sino dos.

Supongamos que existen  $h$  relaciones exactas contemporáneas entre las series:

$$(Z_t - \bar{Z})' c_i = 0,$$

entonces tendremos únicamente  $k - h$  componentes independientes. Esta situación exacta es poco frecuente con pocas series, pero con muchas series podemos encontrarnos combinaciones lineales que son casi constantes o con muy pequeña variabilidad. Para contrastar este punto, comenzaremos calculando la matriz  $\hat{\Gamma}(0)$ . Como:

$$\hat{\Gamma}(0) = \frac{1}{T} \sum (Z_t - \bar{Z})(Z_t - \bar{Z})'$$

tendremos que:

$$\hat{\Gamma}(0) c_i = \frac{1}{T} \sum (Z_t - \bar{Z})(Z_t - \bar{Z})' c_i = 0,$$

lo que implica que la matriz  $\hat{\Gamma}(0)$  debe tener  $k - h$  autovalores iguales a cero. Si la relación es  $(Z_t - \bar{Z})' c_i = u_i$ , para  $h$  series  $u_i$  que tienen media cero y una varianza muy pequeña, tendremos que la matriz  $\hat{\Gamma}(0)$  tendrá  $h$  autovalores muy pequeños. En lugar de modelar el vector de  $k$  series podemos tomar los vectores y valores propios de la matriz  $\hat{\Gamma}(0)$  y modelar las  $k - h$  combinaciones lineales que tienen alta varianza. Esto equivale a realizar componentes principales de los datos y quedarnos con las  $k - h$  componentes de alta varianza. Un aspecto a tener en cuenta es que cuando los componentes de las series no van en las mismas dimensiones conviene estandarizar desde el principio cada serie dividiéndola por su desviación típica para evitar problemas de escala. Esto equivale a obtener los componentes de la matriz de correlaciones estimada, en lugar de la de covarianzas.

Para determinar la estructura de matrices  $\Phi(B)$  y  $\Theta(B)$  de un vector de series estacionarias utilizamos las matrices estimadas de correlación simple y parcial. Si el orden  $m$  de las matrices de autocorrelación simple es tal que para  $h > m$  las matrices  $\hat{R}(h)$  tiene todos los elementos nulos y  $|\hat{R}(h)| = 0$  es alto (mayor que 3 o 4), esto indica un proceso AR. Si el orden  $h$  es bajo (1, 2 o 3), la hipótesis más simple es que el proceso es MA( $h$ ). Por lo tanto, la función de correlación simple nos permitirá diferenciar, en una primera instancia, entre procesos AR y MA. Las matrices de correlación parcial tienen la propiedad opuesta. Si hay unas pocas distintas de cero, lo que contrastaremos con el test de verosimilitudes, esto

indica un proceso AR, mientras que matrices con coeficientes distintos de cero hasta retardos elevados sugieren un proceso MA.

Una información adicional importante para este análisis es el estudio de las matrices de autocorrelación simple de los residuos estimados en cada etapa, que pueden indicar la existencia de un modelo mixto. Por ejemplo, supongamos que el verdadero modelo es ARMA (1,1):

$$(\mathbf{I} - \Phi B) \mathbf{Z}_t = (\mathbf{I} - \Theta B) a_t,$$

y erróneamente especificamos un AR(1), dado por:

$$(\mathbf{I} - \Phi^* B) \mathbf{Z}_t = \varepsilon_t,$$

entonces, los residuos  $\varepsilon_t$  tendrán estructura multivariante, indicando la posible existencia de un modelo mixto.

#### 19.4.2. Estimación

Una vez determinados los posibles órdenes  $p$  y  $q$  de las partes AR y MA, pasaremos a estimar el modelo:

$$(\mathbf{I} - \Phi_1 B - \dots - \Phi_p B^p) \mathbf{Z}_t = (\mathbf{I} - \Theta_1 B - \dots - \Theta_q B^q) a_t.$$

Sea  $m = (p, q)$ . La función de verosimilitud condicionada a los valores iniciales del proceso  $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ , de las innovaciones  $a_{m+1}, \dots, a_T$ , es:

$$f_c(a_{m+1}, \dots, a_T) = \prod_{t=m+1}^T \frac{1}{|\Sigma|^{1/2} 2\pi^{k/2}} \exp \left\{ -\frac{1}{2} (a_t' \Sigma^{-1} a_t) \right\},$$

dado que  $a_t$  es  $N_k(0, \Sigma)$ . Tomando logaritmos y llamando  $\beta$  al vector de parámetros que incluye todos los AR y MA podemos escribir:

$$\log f_c(\beta) = -\frac{T-m}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=m+1}^T (a_t' \Sigma^{-1} a_t).$$

Utilizando que un escalar es igual a su traza, y  $tr(ABC) = tr(BCA) = tr(CAB)$ , tendremos que:

$$\begin{aligned} \text{tr} \sum_{t=m+1}^T a_t' \Sigma^{-1} a_t &= \sum_{t=m+1}^T \text{tr}(a_t' \Sigma^{-1} a_t) = \\ &= \text{tr}(\Sigma^{-1} \sum_{t=m+1}^T a_t a_t') = \text{tr}(\Sigma^{-1} S(\beta)). \end{aligned}$$

El procedimiento operativo para maximizar esta función es utilizar un algoritmo no lineal y seguir los mismos principios que en el caso univariante. La matriz de covarianzas estimada vendrá dada por:

$$\widehat{\Sigma} = \frac{1}{T-p} \sum_{t=m+1}^T \widehat{\mathbf{a}}_t \widehat{\mathbf{a}}_t'$$

Para decidir entre varios modelos estimados utilizaremos el criterio BIC. Si ajustamos un VAR( $p$ ), el número de parámetros estimados es  $pk^2$  y si ajustamos un VARMA( $p, q$ ) es  $(p+q)k^2$ , con lo que en este caso el criterio es:

$$BIC = \log |\widehat{\Sigma}| + \frac{1}{T}(p+q)k^2 \log T.$$

En las experiencias de Monte Carlo el criterio BIC suele dar mejores resultados que el AIC, que tiende a sobreparametrizar. El criterio AIC viene dado por:

$$AIC = \log |\widehat{\Sigma}| + \frac{2}{T}(p+q)k^2.$$

#### 19.4.3. Diagnosis

Los diagnósticos del modelo están diseñados para identificar posibles errores en la formulación del mismo y consisten en el análisis de los residuos estimados  $\widehat{\mathbf{a}}_t$  para comprobar si siguen un proceso de ruido blanco, lo que implica:

1. Cada componente  $\widehat{a}_{it}$  debe ser ruido blanco. Esto lo comprobaremos analizando las fas y fap de cada componente y utilizando los contrastes estudiados en el caso univariante.
2. Conjuntamente el vector  $\widehat{\mathbf{a}}_t$  sólo debe tener correlación contemporánea, lo que implica que tanto las matrices de autocorrelación  $\mathbf{R}_{at}(l)$ , como las parciales  $\mathbf{P}_{at}(l)$ , deben tener todos los elementos no significativamente distintos de cero para  $l \geq 1$ . Además, los determinantes de estas matrices deben de ser cero para  $l \geq 1$ .

Si observamos que  $\widehat{\mathbf{a}}_t$  tiene estructura dinámica, del tipo:

$$\widehat{\mathbf{a}}_t = \Psi(B) \mathbf{a}_t,$$

reformularemos el modelo mediante:

$$\Phi(B) \mathbf{Z}_t = \Theta(B) \mathbf{a}_t = \Theta(B) \Psi(B) \mathbf{a}_t,$$

y volveremos a la etapa de estimación.

### 19.5. Cointegración y factores comunes

Un proceso vectorial se denomina  $I(0)$  si es estacionario y  $I(1)$  si no lo es, pero al diferenciar todas las series el vector  $\nabla Z_t$  es  $I(0)$ . Generalizando, diremos que el vector de series es  $I(2)$  si es no estacionario, pero el vector  $\nabla^2 Z_t$  es  $I(0)$ .

Supongamos que un vector de series es  $I(1)$ . Podríamos diferenciar todas las series y construir un modelo VARMA para  $\nabla Z_t$ . Esto equivale a suponer que el modelo para el vector de series tiene una parte AR vectorial que puede escribirse:

$$\varphi(B) = \nabla \Phi_p(B)$$

pero esto implica que podemos sacar factor común en todos los términos de la matriz  $\varphi(B)$  un término  $(1 - B)$ , y la ecuación  $|\varphi(B)| = 0$  puede escribirse  $\nabla^k |\Phi_p(B)| = 0$  y tendremos  $k$  raíces iguales a la unidad. Es frecuente en muchos conjuntos de series que el número de raíces unidad sea mucho menor que el número de series. En estos casos, si diferenciamos todas las series, obtendremos un modelo muy complejo para las series diferenciadas, como hemos visto en un caso particular en el capítulo 18 y se ilustra en general en el apéndice 19.1. Supongamos, por ejemplo, que la falta de estacionariedad es debida al primer componente, que influye sobre los restantes y los hace no estacionarios. Sin embargo, para representar esta relación sólo necesitamos una raíz unitaria. Por ejemplo, en el sistema:

$$\begin{aligned} \nabla z_{1t} &= a_{1t}, \\ z_{2t} - \beta z_{1t} &= a_{2t}, \end{aligned} \tag{19.13}$$

el primer componente es no estacionario y al influir sobre el segundo hace que éste también lo sea, pero existe una combinación lineal entre las series, la dada por (19.13), que es estacionaria. Si diferenciamos los dos componentes tendremos un modelo mucho más complejo.

Cuando tenemos muchas series pueden existir varias relaciones de cointegración entre ellas, o combinaciones lineales estacionarias. Estas combinaciones lineales no son únicas, porque si  $\alpha'_1 Z_t$  es estacionaria y  $\alpha'_2 Z_t$  es también estacionaria, es inmediato que la nueva combinación lineal:

$$(a\alpha'_1 + b\alpha'_2) Z_t$$

será también estacionaria. Cuando tenemos más de una relación de cointegración, o más de una combinación lineal estacionaria, tendremos que hablar del espacio de las combinaciones lineales estacionarias, ya que éstas no están definidas de manera única. Si tenemos  $k$  series y podemos

construir  $k - r$  combinaciones lineales independientes que conducen a procesos estacionarios, sólo necesitamos  $r$  raíces unidad o términos no estacionarios en la relación.

Cuando  $k$  es grande, es frecuente que  $r$  sea pequeño y podamos representar la dinámica no estacionaria con unas pocas raíces unidad. Parece conveniente, en este caso, en lugar de pensar en el espacio de alta dimensión,  $k - r$ , que engloba las relaciones de cointegración, pensar en el espacio de dimensión  $r$  donde están los componentes no estacionarios. Podemos representar el conjunto de series como dirigido por  $r$  factores no estacionarios, que crean la no estacionariedad del conjunto. Es posible que estos factores sean algunos de los componentes de las series, pero en general serán combinaciones lineales de los componentes. Por ejemplo, si consideramos la demanda para un grupo grande de productos, los caudales de un conjunto de ríos o las series de mortalidad en un país por distintas edades, es frecuente que el componente no estacionario de estas series esté generado por unos pocos factores comunes no estacionarios, que se estiman combinando todos los componentes. De esta manera, nos concentraremos en el espacio de dimensión  $r$  de no estacionariedad. Un modelo general para describir esta situación es:

$$\mathbf{Z}_t = \Lambda \mathbf{f}_t + \mathbf{n}_t, \quad (19.14)$$

donde  $\mathbf{f}_t$  es un vector de  $r < k$  series no observadas, los factores, que están incorreladas entre sí y que son no estacionarias, y  $\mathbf{n}_t$  es un proceso vectorial estacionario. Supongamos que los factores son no estacionarios y siguen el proceso

$$\nabla \mathbf{f}_t = \mathbf{u}_t, \quad (19.15)$$

es decir, forman un proceso paseo aleatorio vectorial. Por ejemplo, la situación bivariante anterior (19.13) puede modelizarse con un único factor y  $\Lambda' = (1, \beta)$ ,  $\mathbf{n}_t = (\epsilon_{1t}, a_{2t})$ , donde  $\epsilon_{1t}$  tiene una varianza y muy pequeña, y  $\mathbf{u}_t = a_{1t}$ .

Cuando existen factores comunes, como en el modelo (19.14) y (19.15), el vector de series  $\mathbf{Z}_t$  estará cointegrado y existirán  $k - r$  combinaciones lineales de las series que son estacionarias. En efecto, como la matriz  $\Lambda$  es  $k \times r$ , existirá otra matriz  $\Lambda^\perp$  de dimensiones  $k \times (k - r)$  cuyas columnas son ortogonales a las de  $\Lambda$ . Sean  $\mathbf{v}_i$  para  $i = 1, \dots, k - r$  las columnas de  $\Lambda^\perp$ . Entonces, las combinaciones lineales:

$$\mathbf{v}'_i \mathbf{Z}_t = \mathbf{v}'_i \Lambda \mathbf{f}_t + \mathbf{v}'_i \mathbf{n}_t = \mathbf{v}'_i \mathbf{n}_t$$

serán estacionarias, ya que, por construcción  $\mathbf{v}'_i \Lambda = \mathbf{0}$ . Tendremos  $r$  factores comunes, o  $k - r$  relaciones de cointegración, entre las series. Puede demostrarse (véase Escrivano y Peña, 1994) que lo contrario es cierto: si

existen  $k - r$  relaciones de cointegración, el modelo para las series puede escribirse como generado por  $r$  factores comunes.

### 19.6. Especificación del modelo para variables no estacionarias

Supondremos ahora un vector de  $k$  series  $Z_t$ , en general no estacionarias. Vamos a estudiar la especificación del modelo en este caso. La diagnosis se realiza igual que en el caso estacionario.

#### 19.6.1. El modelo de corrección de error

Una forma simple de detectar el número de raíces unitarias podría ser estimar un VAR(1) en las series en niveles y buscar donde aparecen aproximadamente raíces unidad. Fijando a uno las raíces próximas a la unidad, podríamos reestimar el modelo con esta restricción y determinar el resto de la estructura ARMA como hemos visto para variables estacionarias.

Un inconveniente de este método es que al existir raíces unitarias no podemos fiarnos de las desviaciones típicas de los coeficientes estimados por mínimos cuadrados. El problema es similar al visto en el capítulo 9 al estudiar los contrastes de raíces unitarias, y en el 18, al estudiar la regresión entre variables no estacionarias. Para evitar este inconveniente, podríamos intentar estimar un modelo con variables estacionarias, donde este problema no exista, y acudir a la idea de separar las relaciones a corto y largo plazo. Esta idea conduce al *modelo de corrección de error*, que es una generalización del modelo estudiado en el contraste de Dickey-Fuller aumentado. En el caso vectorial tiene la forma:

$$\nabla Z_t = \Pi Z_{t-1} + \sum_{i=1}^p \beta_i \nabla Z_{t-i} + a_t. \quad (19.16)$$

En este modelo si las variables  $Z_t$  son  $I(1)$ , siguen un proceso VAR( $p$ ) y no están cointegradas, la matriz  $\Pi$  será cero y los coeficientes  $\beta_i$  son los coeficientes AR del VAR( $p$ ) en diferencias. Sin embargo, si las variables están cointegradas y existen  $k - r$  relaciones de cointegración, puede demostrarse que la matriz  $\Pi$  tiene rango  $k - r$ , y puede escribirse  $\Pi = AB$ , donde  $A$  es  $k \times (k - r)$  y  $B$  es  $(k - r) \times k$ . La matriz  $B$  contiene las relaciones de cointegración y, por tanto,  $BZ_{t-1}$  es un vector de series estacionarias.

El procedimiento de determinar el número de raíces unitarias es estimar el modelo (19.16) y estudiar el rango de la matriz estimada  $\hat{\Pi}$ . Si

el rango es aproximadamente cero, es decir la matriz no es significativamente distinta de cero, eliminamos esta matriz y estimamos un VAR( $p$ ) en diferencias de las variables. Si el rango es  $h < k$ , podemos obtener su descomposición en valores singulares y escribir:

$$\widehat{\Pi} = \mathbf{UDV},$$

donde  $\mathbf{U}$  es  $k \times h$  e incluye los vectores propios no nulos de  $\widehat{\Pi}'\widehat{\Pi}$ ,  $\mathbf{D}$  es diagonal  $h \times h$  e incluye las raíces cuadradas de los valores propios no nulos de  $\widehat{\Pi}'\widehat{\Pi}$  y  $\mathbf{V}$  es  $h \times k$  y contiene los vectores propios ligados a valores propios no nulos de la matriz  $\widehat{\Pi}'\widehat{\Pi}$ . Tomando  $\widehat{\mathbf{A}} = \mathbf{UD}^{1/2}$  y  $\widehat{\mathbf{B}} = \mathbf{D}^{1/2}\mathbf{V}$  obtenemos las relaciones de cointegración en la matriz  $\widehat{\mathbf{B}}$ .

### 19.6.2. El modelo factorial

Un procedimiento alternativo, que tiene ventajas con muchas series y muchas relaciones de cointegración entre ellas, es estimar un modelo factorial. Puede demostrarse (véase Peña y Poncela, 2005) que si las series tienen  $r$  factores comunes no estacionarios, todas las matrices de autocorrelación estimadas  $\widehat{\Gamma}(m)$  tienen:

1. Aproximadamente rango  $r$ .
2. Sus vectores propios asociados a los  $r$  mayores valores propios definen las combinaciones lineales no estacionarias.
3. Sus vectores propios asociados a los  $k - r$  menores valores propios definen las combinaciones estacionarias o las relaciones de cointegración.

De esta manera, descomponemos el problema en modelar  $r$  variables integradas, que convertiremos en estacionarias tomando diferencias, y  $k - r$  variables estacionarias.

La construcción de modelos factoriales puede verse en Peña y Poncela (2005).

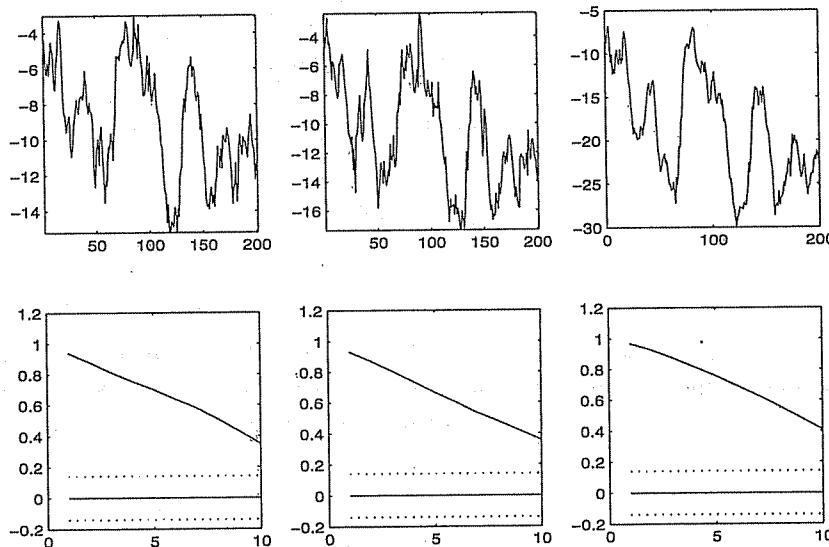
---

### Ejemplo 19.2

Hemos generado tres series temporales mediante un modelo VAR(1) con matriz de parámetros:

$$\Phi = \begin{bmatrix} 1 & 0 & 0 \\ .6 & .5 & 0 \\ .4 & .2 & .7 \end{bmatrix}$$

Figura 19.1 Tres series que siguen un VAR(1) con una única raíz unitaria



y matriz  $\Sigma = \mathbf{I}$ . En este modelo la primera serie sigue un paseo aleatorio y es  $I(1)$ . La segunda viene dada por:

$$z_{2t} = .5z_{2t-1} + .6z_{1t-1} + a_{2t}$$

y como es función lineal de la serie  $z_{1t-1}$ , que es no estacionaria, será también no estacionaria,  $I(1)$ . La tercera, sigue el modelo

$$z_{3t} = .7z_{3t-1} + .2z_{2t-1} + .4z_{1t-1} + a_{3t}$$

y será, como las anteriores también  $I(1)$ . El gráfico de las tres series y sus funciones de autocorrelación univariante se presenta en la figura 19.1.

La estimación por mínimos cuadrados de un AR(1) para estas series conduce a la matriz:

$$\hat{\Phi} = \begin{bmatrix} 1.004 & .034 & -0.023 \\ 0.497 & 0.554 & 0.015 \\ 0.377 & 0.162 & 0.729 \end{bmatrix}$$

Se observa que la estimación detecta con precisión la raíz unitaria debido al coeficiente 1 en  $\phi_{11}$  y estima razonablemente el resto de los términos. Este modelo indica que existe una única raíz unitaria, y que la falta de estacionariedad que se observa en las tres series es debida a la primera.

## Análisis de series temporales

---

Supongamos que estimamos con estas tres series un modelo de corrección de error. El resultado es:

$$\nabla Z_t = \begin{bmatrix} .004 & 0.034 & -0.023 \\ .497 & -0.445 & 0.015 \\ 0.377 & 0.162 & -0.270 \end{bmatrix} Z_{t-1} + a_t$$

que tiene una interpretación similar al modelo anterior. La primera serie sigue un paseo aleatorio y las otras los modelos:

$$z_{2t} = .545 z_{2t-1} + .497 z_{1t-1} + a_{2t}$$

y

$$z_{3t} = .73 z_{3t-1} + .16 z_{2t-1} + .377 z_{1t-1} + a_{3t}$$

que son similares a los estimados con la estructura VAR. Además, la matriz estimada del modelo de corrección de error nos proporciona las relaciones de cointegración. La descomposición en valores singulares de esta matriz es

$$\begin{aligned} & \begin{bmatrix} .004 & 0.034 & -0.023 \\ .497 & -0.445 & 0.015 \\ 0.377 & 0.162 & -0.270 \end{bmatrix} \\ &= \begin{bmatrix} .01 & 0.09 & 1 \\ -.09 & -0.40 & .04 \\ -.40 & 0.91 & .07 \end{bmatrix} \begin{bmatrix} .7 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -.86 & .33 & .38 \\ .48 & .74 & .45 \\ .13 & -.58 & .80 \end{bmatrix} \end{aligned}$$

Vemos que la matriz tiene rango dos, por lo que existen dos relaciones de cointegración o un factor común. La matriz de corrección de error puede escribirse como:

$$\begin{aligned} & \begin{bmatrix} .004 & 0.034 & -0.023 \\ .497 & -0.445 & 0.015 \\ 0.377 & 0.162 & -0.270 \end{bmatrix} \approx \\ & \approx \begin{bmatrix} .01 & 0.06 \\ -.77 & -0.26 \\ -.33 & 0.60 \end{bmatrix} \begin{bmatrix} -.72 & .27 & .32 \\ .32 & .49 & .30 \end{bmatrix} \end{aligned}$$

que muestra las dos relaciones de cointegración.

---

### 19.7. Lecturas complementarias

Wei (1990) y Shumway y Stoffer (2000) son buenas introducciones a los modelos multivariantes. Peña, Tiao y Tsay (2001) incluye también una introducción al tema y varios ejemplos. Textos más avanzados para profundizar en el campo de los modelos VARMA multivariantes son Lutkepohl (1993) y Reinsel (1997). La estimación MV de modelos vectoriales ha sido estudiada, entre otros, por Luceño (1994) y Mauricio (1995). Para los modelos factoriales y la reducción de dimensionalidad véanse Box y Tiao (1977), Peña y Box (1987), Stock y Watson (1988), Tiao y Tsay (1989), Ahn y Reinsel (1988, 1990), Reinsel y Velu (1998) y Peña y Poncela (2004, 2005).

### Apéndice 19.1: Los riesgos de la sobrediferenciación en series multivariantes

Para concretar, supongamos el caso simple bivariante de un AR(1):

$$\begin{bmatrix} 1 - \phi_{11}B & -\phi_{12}B \\ -\phi_{21}B & 1 - \phi_{22}B \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

$$|\Phi(B)| = 1 - (\phi_{11} + \phi_{22})B + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21})B^2.$$

Si esta ecuación tiene una raíz  $B = 1$ , el proceso será no estacionario pero cointegrado. Si tiene dos raíces  $B = 1$ , el proceso será integrado puro de orden uno sin cointegración. Imponiendo la condición  $B = 1$  en el determinante, tenemos que :

$$1 - (\phi_{11} + \phi_{22}) + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21}) = 0$$

es decir:

$$(1 - \phi_{11}) = \phi_{22}(1 - \phi_{11}) + \phi_{12}\phi_{21}.$$

Si alguno de los términos  $\phi_{12}$  o  $\phi_{21}$  son nulos –lo que implica que no hay realimentación–, entonces el determinante tendrá una raíz unitaria si  $\phi_{11} = 1$ , o bien  $\phi_{11} = 1$ , o bien ambos son la unidad. Consideremos los tres casos posibles.

1.  $\phi_{12} = 0; \phi_{11} = 1, \phi_{22} < 1$ . El modelo es:

$$\begin{bmatrix} \nabla & 0 \\ -\phi_{21}B & 1 - \phi_{22}B \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

que conduce a que el primer componente siga un paseo aleatorio, y el

segundo:

$$(1 - \phi_{22}B) z_{2t} = \phi_{21}z_{1t-1} + a_{2t} = \nabla^{-1}\phi_{21}a_{1t-1} + a_{2t},$$

que puede escribirse, multiplicando por  $\nabla$ , como:

$$\nabla(1 - \phi_{22}B) Z_{2t} = (1 - \theta B) \varepsilon_t.$$

La conclusión es que en este caso las dos series son integradas. Si hubiésemos tomado diferencias, el modelo para las series diferenciadas sería:

$$\begin{bmatrix} 1 & 0 \\ \frac{-\phi_{21}B}{\nabla} & \frac{1-\phi_{22}B}{\nabla} \end{bmatrix} \begin{bmatrix} \nabla z_{1t} \\ \nabla z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

y tendría una estructura muy compleja, lo que conduciría a aproximarla por términos MA en el lado derecho o a estimar AR de orden muy elevado.

2. Si  $\phi_{12} = 0$ ;  $\phi_{11} = 1$ ;  $\phi_{22} = 1$ , el modelo sería:

$$\begin{bmatrix} \nabla & 0 \\ -\phi_{21}B & \nabla \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

y si hubiésemos tomado una diferencia:

$$\begin{bmatrix} 1 & 0 \\ \frac{-\phi_{21}B}{\nabla} & 1 \end{bmatrix} \begin{bmatrix} \nabla z_{1t} \\ \nabla z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

y de nuevo se nos complicaría el proceso de identificación.

3. Si ninguno de los términos  $\phi_{ij}$  es nulo, el problema puede ser aún más complicado.

En resumen, si las series univariantes son no estacionarias, esto puede ser debido a la presencia de una única raíz unitaria en  $|\Phi(B)|$ . Si diferenciamos todas las series, esto nos puede llevar a complicar innecesariamente el análisis.

---

## Bibliografía

- Abraham, B., y Ledolter, J. (1983): *Statistical Methods for forecasting*. Wiley.
- Ahn, S. K. (1997): "Inference of vector autoregressive models with cointegration and scalar components", *Journal of American Statistical Association*, 92, 350-356.
- , y Reinsel, G. C. (1988): "Nested Reduced-Rank Autoregressive Models for Multiple Time Series", *Journal of American Statistical Association*, 83, 849-856.
- , y — (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models", *Journal of American Statistical Association*, 85, 813-823.
- Akaike, H. (1969): "Fitting autoregressive models for prediction", *Annals of the Institute of Statistical Mathematics*, 21, 243-247.
- (1973): "Information theory and an extension of the maximum likelihood principle", *Proceedings of the 2nd International Symposium on Information Theory* (Akadémiai Kiadó, Budapest), 267-281.
- (1974): "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, 19, 203-217.
- (1974): "Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes", *Annals of the Institute of Statistical Mathematics*, 26, 363-387.

- Alonso, A., Peña, D., y J. Romo, J. (2004a): "Introducing Model Uncertainty in Time Series Bootstrap", *Statistica Sinica*, 14, 155-174.
- , —, y — (2004b): "Resampling Time Series by Missing Values Techniques", *The Annals of the Institute of Statistical Mathematics*, 55, 765-796.
- Anderson, B. O., y Moore, J. B. (1979): *Optimal Filtering*. Prentice-Hall.
- Anderson, T. W. (1971): *The Statistical Analysis of Time Series*. Wiley.
- Andrews, W. K., y Ploberger, W. (1994): "Optimal tests when the nuisance parameter is present only under the alternative", *Econometrica*, 62, 1383-1414.
- Aoki, M. (1990): *State Space Modeling of Time Series*. Springer-Verlag.
- Aznar, A., García Ferrer, A., y Martín Arroyo, A. (1994): *Ejercicios de Econometría II*. Pirámide.
- , y Salvador, M. (2002): "Selecting the rank of the cointegration space and the form of using an information criterion", *Econometric Theory*, 18, 926-947.
- , y Trivez, F. J. (1993): *Métodos de Predicción en Economía*, (dos vol.). Ariel Economía.
- Bai, J., y Perron, P. (1998): "Estimating and Testing Linear Models with Multiple Structural Changes", *Econometrica*, 66, 47-78.
- , y — (2003): "Critical Values for multiple Structural Change Tests", *Econometrics Journal*, 6, 72-78.
- Banarjee, A., Dolado, J., Galbraith, J. W., y Hendry, D. (1993). *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.
- Bartlett, M. S. (1946): "On the theoretical specification of sampling properties of autocorrelated time series", *Journal of The Royal Statistical Society (Suppl.)*, 7, 211.
- (1978): *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications* (3<sup>a</sup>ed.). Cambridge University Press.
- Basak, G. K., Chan, N. H., y Palma, W. (2001): "The approximation of long-memory processes by an ARMA model", *Journal of Forecasting*, 20, 367- 389.
- Beran, J. (1994): *Statistics for Long-Memory Processes*. Chapman and Hall.
- , Bhansali, R. J., y Ocker, D. (1998): "On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes", *Biometrika*, 85, 921-934.
- Bhansali, R. J. (1999): "Autoregressive model selection for multistep prediction", *Journal of Statistical Planning and Inference*, 78, 295-305.

## Bibliografía

---

- Bianco, A. M., García Ben, M., Martínez, E. J., y Yohai, V. J. (2001) "Outlier Detection in Regression Models with ARIMA Errors using Robust estimates", *Journal of Forecasting*, 20, 565-579.
- Bloomfield, P. (2000): *Fourier Analysis of Time Series: An Introduction*. Wiley.
- Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.
- , Chou, R. Y., y Kroner, K. F. (1992): "ARCH modeling in finance", *Journal of Econometrics*, 52, 5-59.
- , Engle, R. F., y Nelson, D. B. (1994): "ARCH model", *Handbook of Econometrics IV*, ed. Engle, R. F., y McFadden, 2959-3038, D. C. Amsterdam: Elsvier Science.
- Box, G. E. P., Hillmer, S., y Tiao, G. C. (1978): "Analysis and Modeling of Seasonal Time Series", en *Seasonal Analysis of Economic Time Series*, A. Zellner (editor). Bureau of the Census. USA.
- , y Jenkins, G. M. (1976): *Time Series Analysis: Forecasting and Control*. Holden Day.
- , — y Reinsel, G. (1994): *Time Series Analysis: Forecasting and Control*. Prentice Hall.
- , y Luceño, A. (1999): *Statistical control by monitoring and feedback adjustment*. Wiley.
- , y Newbold, P. (1970): "Some comments on a paper by Coen, Gomme and Kendall", *Journal of The Royal Statistical Society A*, 134, 229.
- , y Tiao, G. C. (1975): "Intervention analysis with applications to economic and environmental problems", *Journal of American Statistical Association*, 70, 70-99.
- , y — (1976): "Comparison of Forecast and Actuality", *Applied Statistics*, 25, 3, 195-200.
- , y — (1977): "A canonical analysis of multiple time series", *Biometrika*, 64, 355-365.
- Brillinger, D. R (1969): "Asymptotic properties of spectral estimates of second order", *Biometrika*, 56, 375-390.
- (1975): *Time Series: Data Analysis and Theory*. Holden Day.
- y Krishnaiah, P. R. (1983): *Time Series in the Frequency Domain*. (Handbook of Statistics, vol. 3). North-Holland.

Análisis de series temporales

---

- Brock, W., Dechert, D., Scheinkman, J., y LeBaron, B. (1996): "A Test for Independence Based on the Correlation Dimension", *Econometric Reviews*, 15, 197-235.
- , Hsieh, D., y LeBaron, B. (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. MIT Press: Cambridge, MA.
- Brockwell, P. J., y Davis, R. A. (1987): *Time Series: Theory and Methods*. Springer-Verlag.
- , y — (1996): *Introduction to Time Series Analysis and Forecasting*. Springer-Verlag.
- Caridad, J. M. (1998): *Econometría: modelos econométricos y series temporales*. Reverté.
- Carnero, A., Peña, D., y Ruiz, E. (2001): "Outliers and Conditional Autoregressive Heteroscedasticity in Time Series", *Estadística*, 53, 143-213.
- , — y — (2004a): "Persistence and Kurtosis in GARCH and Stochastic Volatility models", *Journal of Financial Econometrics*, 2, 319-342.
- , — y — (2004b): "Spurious and hidden volatility", *Working Paper*, Universidad Carlos III de Madrid.
- Casals, J., Jerez, M., y Sotoca, S. (2002): "An exact Multivariate Model-Based Structural decomposition", *Journal of American Statistical Association*, 97, 533-564.
- Chan, N. H. (2002): *Time Series, Applications to Finance*. Wiley.
- Chen, H. F. (1985): *Recursive estimation and Control for Stochastic Systems*. Wiley.
- Clements, M. P., y Hendry, D. F. (1998): *Forecasting Economic Time Series*. Cambridge University Press.
- Cleveland, W. P., y Tiao, G. C. (1976): "Descomposition of Seasonal Time Series. A model for the Census X-II Program", *Journal of American Statistical Association*, 71, 581-587.
- Davidson, J. (2000): *Econometric Theory*. Blackwell Publishers.
- Davis, M. H. A. (1977): *Linear Estimation and Stochastic Control*. Chapman and Hall.
- Diebold, F. (1999): *Elementos de Pronósticos*, International Thomson Editores.
- , y Kilian, L. (2001): "Measuring Predictability: Theory and Macroeconomic Applications" , *Journal of Applied Econometrics*, 16, 657-669.

## Bibliografía

- Dickey, D. A., y Fuller, W. A. (1981): "Likelihood ratio statistics for autoregressive processes with a unit root", *Econometrica*, 49, 1057-1072.
- Diks, C. (1999): *Nonlinear Time Series Analysis*. Word Scientific.
- Dolado, J. J. (1989): "Asymptotic properties of OLS and GLS", *Econometric Theory*, 5, 177-180.
- Durbin, J. (1970): "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica*, 38, 410.
- , y Koopman, S. J. (2000): *Time Series Analysis By State Space Models*. Oxford University Press.
- , y Watson, G. S. (1950): "Testing for serial correlation in least squares regression I", *Biometrika*, 37, 409-428.
- , y — (1951): "Testing for serial correlation in least squares regression II", *Biometrika*, 38, 159-178.
- Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations", *Econometrica*, 50, 987-1007.
- (1995): *Arch Selected Readings*. Oxford University Press.
- , y Granger, C. W. J. (1987): "Co-integration and error correction: Representation, estimation and testing", *Econometrica*, 55, 251-276.
- Enders, W. (1995): *Applied Econometric Time Series*. Wiley.
- Escribano, A., y Peña, D. (1994): "Cointegration and common factors", *Journal of Time Series Analysis*, 15, 577-586.
- Espasa, A., (1978): *Estimación y selección de modelos econométricos dinámicos*, Banco de España. Servicio de Estudios. Estudios Económicos, 11.
- , y Cancelo, J. R. (1993): *Métodos cuantitativos para el análisis de la coyuntura económica*. Alianza Editorial.
- , y Peña, D. (1995): "The decomposition of forecasts in seasonal ARIMA models", *Journal of Forecasting*, 14, 565-583.
- Evans, M. E. (2003). *Practical Business Forecasting*, Blackwell.
- Fan, J., y Yao, Q. (2003) *Nonlinear time series: Nonparametric and Parametric Methods*, Springer, Nueva York.
- Flores, R., y Novales, A. (1997): "A general test for univariate seasonality", *Journal of Time Series Analysis*, 18, 29-48.
- Franses, P. H. (1998): *Time Series Models for Business and Economics Forecasting*. Cambridge University Press.

- , y Paap, R. (2004): *Periodic Time Series Models*. Oxford University Press.
- , y van Dijk, D. (2000): *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press.
- Fuller, W. A. (1994): *Introduction to Statistical Time Series*. Wiley (2<sup>a</sup> edición).
- Galeano, P., y Peña, D. (2005a): “Model Selection criteria and quadratic discrimination in time series”, *Working Paper*, Universidad Carlos III de Madrid.
- , y — (2005b): “A Note on Prediction and Interpolation Errors in Time Series”. *Statistics and Probability Letters*, 73, 71-78.
- , — y Tsay, R. S. (2004): “Outlier Detection in Multivariate Time Series Via Projection Pursuit”, *Working Paper*, Universidad Carlos III de Madrid.
- García-Ferrer, A., Highfield, R. A., Palm, F., y Zellner, A. (1987): “Macroeconomic forecasting using pooled international data”, *Journal of Business and Economic Statistics*, 5, 53-67.
- , y Poncela, P. (2002): “Forecasting European GNP data through common factors and other methods”, *International Journal of Forecasting*, 21, 225-244.
- García-Jurado, I., González-Manteiga, W., Prada-Sánchez, J. M., Febrero-Bande, M., y Cao, R. (1995): “Predicting using Box-Jenkins, nonparametric and bootstrap techniques”, *Technometrics*, 37, 303-310.
- Gómez, V., y Maravall, A. (1994): “Estimation, prediction, and interpolation for nonstationary series with the Kalman filter”, *Journal of American Statistical Association*, 89, 611-624.
- , y — (1996): “Programs SEATS and TRAMO: Instructions for the user”, *Working Paper nº 9628*, Banco de España.
- , y — (2001): “Seasonal adjustment and signal extraction in Economic Time Series”, en Peña, D., Tiao, G. C., y Tsay, R. S., *A Course in Time Series*, 202-246.
- , Peña, D., y Maravall, A. (1999): “Missing Observations In ARIMA Models: Skipping Strategy versus Additive Outlier Approach”, *The Journal of Econometrics*, 88, 341-363.
- Gonzalo, J. (1994): “Five alternative methods of estimating long-run equilibrium realtionships”, *The Journal of Econometrics*, 60, 203-233.
- Gini, J. (1982): *The Making of Statisticians*, Springer-Verlag.
- Gooijer, J. G., Abraham, B., Gould, A., y Robinson, L. (1985): “Methods for determining the order of an ARMA process: A survey”, *International Statistical Review*, 53, 3, 301-329.
- Gouriéroux, C. (1997): *Arch Models with Financial Applications*, Springer.

## Bibliografía

---

- , y Monfort A. (1997): *Time Series and Dynamic Models*, Cambridge University Press.
- Granger, C. W. J. (1969): "Investigating causal relations by econometrics models and cross-spectral methods", *Econometrica*, 37, 424-438.
- (1980): "Long memory relationships and the aggregation of dynamic variables", *Journal of Econometrics*, 14, 227-238.
- (1989): *Forecasting in Business and Economics*. Academic Press.
- , y Andersen, A. P. (1980): *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht.
- , y Hatanaka, M. (1964): *Spectral Analysis of Economic Time Series*. Princeton University Press.
- , y Morris, M. (1976): "Time Series modeling and interpretation", *Journal of The Royal Statistical Society A*, 38, 246-257.
- , y Newbold, T. (1977): *Forecasting Economic Time Series*. Academic Press.
- , y T. Teräsvirta. (1992), *Modeling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- Grenander, U., y Rosenblatt, M. (1975): *Statistical Analysis of Stationary Time Series*. Wiley.
- Guerrero, V. M. (1990): "Temporal Disaggregation of Time Series. An ARIMA-based Approach". *International Statistical Review*, 58, 29-46.
- (1991): "ARIMA forecasts with restrictions derived from a structural change". *International Journal of Forecasting*, 7, 339-347.
- , y Peña, D. (2000): "Linear Combination of Information in Time Series with Linear Restrictions", *Journal of Forecasting*, 19, 103-122.
- , y — (2003): "Combining Multiple Time Series Predictors: a Useful Inferential Procedure", *Journal of Statistical Planning and Inference*, 116, 249-276.
- , — y Poncela, P. (1998): "Measuring Intervention Effects on Multiple Time Series subject to Linear Restrictions", *The Journal of Business and Economic Statistics*, 16, 489-497.
- Hamilton, J. D. (1994) *Time Series Analysis*. Princeton University Press.
- , y Watts, O. (1978): "Interpreting partial autocorrelation functions of seasonal time series models", *Biometrika*, 65, 1, 135-140.
- Hannan, E. J. (1980): *Multiple Time Series*. Wiley.
- , y Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley.
- , y Quinn, B. G. (1979): "The determination of the order of an autoregression", *Journal of The Royal Statistical Society B*, 41, 190-195.

## Análisis de series temporales

---

- Harvey, A. C. (1989): *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
- , Ruiz, E., y Shephard, N. (1994): "Multivariate stochastic variance models", *Review of Economic Studies*, 61, 247-264.
- Holt, C. C. (1957): *Forecasting trends and seasonals by exponentially weighted moving averages*, O. N. R. Memorandum, nº 52, Carnegie Institute of Technology.
- Hu, Y., y Chou, R. (2004): "On the Peña-Box model", *Journal of Time Series Analysis*, 25, 811-830.
- Inclán, C., y Tiao, G. C. (1994): "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance", *Journal of the American Statistical Association*, 89, 913-923.
- Jacquier, E., Polson, N. G., y Rossi, P. (1994): "Bayesian analysis of stochastic volatility models" (with discussion), *Journal of Business & Economic Statistics*, 12, 371-417.
- Johansen, S. (1988): "Statistical Analysis of Cointegration vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.
- (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian Vector Autoregressive models", *Econometrica*, 59, 1551-1580.
- (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Justel, A., Peña, D., y Tsay, R. S. (2001): "Detection of Outlier Patches in Autoregressive Time Series", *Statistica Sinica*, 11, 651-673.
- Kaiser, R., y Maravall, A. (2001): *Measuring Business Cycles in Economic Time Series*. Springer.
- Kalman, R. E. (1960): "A new approach to linear filtering and prediction problems", *Journal of Basic Engineering*, 83-D, 95-108.
- Kendall, M. G. y Ord, J. K. (1990): *Time Series* (3<sup>a</sup> ed.), Arnold, Sevenoaks.
- Koopmans, L. H. (1995): *The Spectral Analysis of Time Series*. Academic Press.
- Koreisha, S., y Pukkila, T. M. (1990): "A generalized least-squares approach for estimation of autoregressive moving average models", *Journal of Time Series Analysis*, 11, 139-151.
- Ledolter, J., y Abraham, B. (1981): "Parsimony and its importance in time series forecasting", *Technometrics*, 23, 4, 411-414.
- Li, W. K. (2004): *Diagnostic Checks in Time Series*. Chapman and Hall.

## Bibliografía

---

- Ljung, G., y Box, G. E. P. (1978): "On a measure of lack of fit in time series models", *Biometrika*, 65, 297-303.
- Luceño, A. (1993): "A Fast Algorithm for the Repeated Evaluation of the Likelihood of a General Linear Process for Long Series", *Journal of American Statistical Association*, 88, 229-236
- (1994): "A Fast Algorithm for the exact Likelihood of stationary and partially nonstationary vector autoregressive moving average processes", *Biometrika*, 81, 555-565
- (1998): "Detecting possible non-consecutive outliers in industrial time series", *Journal of The Royal Statistical Society B*, 60, 295-310.
- Lutkepohl, H. (1993): *Introduction to Multiple Times Series Analysis*, Springer-Verlag.
- Maravall, A. (1981): "A note on identification of multivariate time series series models", *Journal of Econometrics*, 16, 237-247.
- (1983): "An application of non linear time series forecasting", *Journal of Business and Economic Statistics*, 1, 1, 66-74.
- , y Pierce, D. (1987): "A Prototypical Seasonal Adjustment Model", *Journal of Time Series Analysis*, 8, 177-193.
- Marquard, D. W. (1963): "An algorithm for least squares estimation of non-linear parameters", *Journal of the Society of Industrial and Applied Mathematics*, 11, 431.
- Mauricio, J. A. (1995): "Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models", *Journal of American Statistical Association*, 90, 282-291.
- (2002): "An algorithm for the exact likelihood of a stationary vector ARMA models", *Journal of Time Series Analysis*, 23, 473-486.
- McCulloch, R. E., y Tsay, R. S. (1994): "Bayesian analysis of autoregressive time series via the Gibbs sampler", *Journal of Time Series Analysis*, 15, 235-250.
- McQuarrie, A. D. R., y Tsai, C. L. (1998): *Regression and Time Series Model Selection*, Singapore, World-Scientific.
- McLeod, A. I., y Li, W. K. (1983): "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.
- McDonald, G., y Galarneau, D. I. (1975): "A Monte Carlo evaluation of some ridge-type estimators", *Journal of American Statistical Association*, 70, 380,

407-416.

- McGarty, T. (1974): *Stochastic Systems and State Estimation*. Wiley.
- Melino, A., y Turnbull, S. M. (1990): "Pricing foreign currency options with stochastic volatility", *Journal of Econometrics*, 45, 239-265.
- Miller, R. B. (1980): "Comment on Robust estimation of autoregressive models", en *Directions in Time Series*, eds. D. R. Brillinger y G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics.
- Mills, T. C. (1999): *The econometric modelling of financial time series*. Cambridge University Press.
- Monti, A. C. (1994): "A Proposal for Residual Autocorrelation Test in Linear Models", *Biometrika*, 81, 776-780.
- Muth, J. F. (1960): "Optimal properties of exponentially weighted forecast", *Journal of American Statistical Association*, 55, 299-306.
- Nelson, D. B. (1991): "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica*, 59, 347-370.
- Nicholls, D. F., y Quinn, B. G. (1982): *Random Coefficient Autoregressive Models: An Introduction*. (Lecture Notes in Statistics no. 11), Springer-Verlag.
- Nieto, F. H., y Martínez, J. (1996). "A recursive approach for estimating missing observations in a univariate time series", *Communication in Statistics (Theory and Methods)*, 25, 9, 2101-2116.
- Novales, A. (1993): *Econometría*, McGraw Hill.
- , y Flores, R. (1997): "Forecasting with periodic models: a comparison with time invariant coefficients", *International Journal of Forecasting*, 13, 393-405.
- Pandit, S. M., y Wu, S. M. (1983): *Time Series and System Analysis with Applications*. Wiley.
- Pankratz, A. (1983): *Forecasting with univariate Box-Jenkins Models*. Wiley.
- (1991): *Forecasting with Dynamic Regression Models*. Wiley.
- Peña, D. (1984): "The autocorrelation function of seasonal ARMA models", *Journal of Time Series Analysis*, 54, 269-272.
- (1986): "Measuring the importance of outliers in ARIMA models", *New Perspectives in Theoretical and Applied Statistics*, (M. L. Puri et al., editores). Wiley.
- (1990): "Influential observations in time series", *Journal of Business and Economic Statistics*, 8, 2, 235-241.
- : "Measuring Influence in Dynamic Regression models", *Technometrics*, 33, 1, 93-102.

## Bibliografía

---

- (1994): "Second Generation Time Series models: A comment to some advances in nonlinear and adaptive modeling in time series", *Journal of Forecasting*, 13, 133-140.
- (1995): "Forecasting Growth with Time Series", *Journal of Forecasting*, 14, 97-105.
- (2001): *Regresión y Diseño de experimentos*. Alianza Editorial.
- , y Box, G. E. P. (1987): "Analysing a simplifying structure in time series", *Journal of American Statistical Association*, 82, 399, 836-843.
- , y Guttman, I. (1988). "Robust Recursive estimation using Kalman Filtering and its applications", en *Bayesian Analysis of Time Series and Dynamic Models*, J. Spall (ed.) Marcel Dekker.
- , y — (1989). "Optimal collapsing of mixture distributions in robust recursive estimation", *Communications in Statistics*, 18, 3, 817-834.
- , y Maravall, A. (1991): "Interpolation, Outliers and Inverse autocorrelation Function", *Communications in Statistics (Theory and Methods)*, 10, 3175-3186.
- , y Poncela, P. (2004): "Forecasting with nonstationary dynamic factor models", *Journal of Econometrics*, 119, 291-321.
- , y — (2005): "Nonstationary dynamic factor analysis", *Journal of Statistical Planning and Inference* (en prensa).
- , y Rodríguez, J. (2002): "A Powerful Portmanteau Test of Lack of Fit for Time Series", *Journal of the American Statistical Association*, 97, 601-610.
- , y — (2005a): "The Log of the Determinant of the Autocorrelation Matrix for Testing Goodness of Fit in Time Series", *Journal of Statistical Planning and Inference* (en prensa).
- , y — (2005b): "Detecting non linearity in time series by model selection criteria", *International Journal of Forecasting* (en prensa).
- , y Sánchez, I. (2005): "Multifold Predictive Validation in Time Series Models", *The Journal of American Statistical Association*, 469, 135-146.
- , y Sánchez-Albornoz, N. (1983): *Dependencia dinámica entre precios agrícolas. Un estudio empírico*. Servicio de Estudios Banco de España.
- , y — (1984): "Wheat Prices in Spain 1857-1890: an application of the Box-Jenkins methodology", *Journal of European Economic history*, 13, 2, 353-373.
- , y Tiao, G. C. (1991): "A note on likelihood estimation of missing values in time series", *The American Statistician*, 45, 212-214.
- , — y Tsay, R. S. (2001): *A Course in Time Series Analysis*. Wiley.

## Análisis de series temporales

---

- Pole, A., West, M., y Harrison, J. (1994): *Applied Bayesian Forecasting and Time Series Analysis*. Chapman and Hall.
- Pourahmadi, M. (2001): *Foundation of Time Series Analysis and Prediction Theory*. Wiley.
- Priestley, M. B. (1983): *Spectral Analysis and Time Series*. Academic Press.
- (1988): *Non-linear and non-stationary Time Series Analysis*. Academic Press.
- Pukkila, T., Koreisha, S., y Kallinen, A. (1990): "The identification of ARMA models", *Biometrika*, 77, 537-548.
- Pulido, A. (1983): *Modelos Econométricos*. Pirámide.
- Quenouille, M. H. (1957): *The Analysis of Multiple Time-Series*. Griffin.
- Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*. Springer Verlag. Nueva York.
- , y Ahn, S. K. (1992): "Vector Autoregressive models with unit roots and reduced rank structure: estimation, Likelihood ratio tests and forecasting", *Journal of Time Series Analysis*, 13, 353-375.
- , y Velu, R. P. (1998): *Multivariate Reduced - Rank Regression*. Springer.
- Robinson, P. M. (1973): "Generalized canonical analysis for time series", *Journal of Multivariate Analysis*, 3, 141-160.
- Rodríguez, J., y Ruiz E. (2005): "A powerful test for conditional heteroscedasticity", *Statistica Sinica* (en prensa).
- Sánchez, I., y Peña, D. (2001): "Properties of Predictors In Overdifferenced Nearly Nonstationary Autoregression", *The Journal of Time Series Analysis*, 22, 45-66.
- Sánchez, M. J.; y Peña, D. (2003): "The Identification of Multiple Outliers in ARIMA models", *Communications in Statistics, Theory and Methods*, 32, 6, 1265-1287.
- Shumway, R. H., y Stoffer, D. S. (2000): *Time Series Analysis and its Applications*. Springer, Nueva York.
- Schwarz, G. (1978): "Estimating the dimension of a model", *The Annals of Statistics*, 6, 461-464.
- Shibata, R. (1980): "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process", *The Annals of Statistics*, 8, 147-164.
- Subba Rao, T., y Gabr, M. M. (1980): "A Test For Linearity of Stationary Time Series", *Journal of Time Series Analysis*, 1, 145-158.

## Bibliografía

---

- Stock, J. H., y Watson, M. W. (1988): "Testing for common trends", *Journal of the American Statistical Association*, 83, 1097-1107.
- Terdik, G. (1999): *Bilinear Stochastic Models and Related Problems of Non-linear Time Series Analysis: A Frequency Domain Approach*, Lecture Notes in Statistics, No. 142. Springer-Verlag: Nueva York.
- Tiao, G. C. y Box, G. E. P. (1981): "Modeling Multiple Time Series with Applications", *Journal of American Statistical Association*, 76, 376, 802-816.
- , y Tsay, R. S. (1983): "Multiple time series modeling and extended sample cross-correlations", *Journal of Business and Economic Statistics*, 1, 43-56.
- , y — (1989): "Model specification in multivariate time series", *Journal of The Royal Statistical Society B*, 51, 157-213.
- , y — (1994): "Some Advances in Nonlinear and Adaptive Modeling in Time Series (with discussion)", *Journal of Forecasting*, 13, 109-140.
- Tong, H. (1983): *Thresholds models in nonlinear times series analysis*. Springer-Verlag
- (1990): *Non linear Time series. A Dynamical System Approach*. Oxford Science Publications.
- (2001): "A personal journay through time series en Biometrika", *Biometrika*, 88, 195-218.
- Tosjtheim, D., y Paulsen, J. (1983): "Bias os some commonly-used time series estimates", *Biometrika*, 70, 389-399.
- Treadway, A., García-Pardo García-Lorenzana, J., y Carabajo, A. (1978): *Efectos sobre la economía española de una devaluación de la peseta*. Fundación Ramón Areces.
- Trivez, J. (1995): "Level Shifts, temporary changes and forecasting", *Journal of Forecasting*, 14, 543-550.
- Tsay, R. S. (1986): "Nonlinearity Tests for Time Series", *Biometrika*, 73, 461-466.
- (1987): "Conditional heteroskedastic time series models", *Journal of the American Statistical Association*, 82, 590-604.
- (1989): "Testing and Modelling Threshold Autoregressive Processes", *Journal of the American Statistical Association*, 84, 231-240.
- (1991): "Nonlinear Time Series Analysis: Diagnostics and Modelling", *Statistica Sinica*, 1, 431-451.
- (2000): "Time Series and Forecasting: Brief History and Future Research", *Journal of the American Statistical Association*, 95, 638-643.

## Análisis de series temporales

---

- (2002): *Analysis of Financial Time Series*. Wiley.
- , y Tiao, G. C. (1984): “Consistent estimates of autoregressive parameters and extended sample autocorrelation function of stationary and nonstationary ARMA models”, *Journal of the American Statistical Association*, 79, 84-96.
- , y — (1990): “Asymptotic properties of multivariate nonstationary processes with applications to autoregressions”, *The Annals of Statistics*, 18, 220-250.
- , Peña, D., y Pankratz, A. (2000): “Outliers in Multivariate time Series”, *Biometrika*, 87, 789-804.
- Uriel, E. (1984): *Serie Temporales. Modelo ARIMA*. Paraninfo.
- Valderrama, M. J., y Ruiz Molina, J. C. (1996): *Filtrado de Kalman*. EUB, Barcelona.
- Velu, R. P., Reinsel, G. C., y Wichern, D. W. (1986): “Reduced rank models for multiple time series”, *Biometrika*, 73, 105-118.
- Vandaele, W. (1983): *Applied Time Series and Box-Jenkins Models*. Academic Press.
- Wei, W. W. S. (1990): *Time Series Analysis*. Addison-Wesley.
- West, M., y Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag.
- Whittle, P. (1963): “On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix”, *Biometrika*, 48, 85-94.
- (1984): *Prediction and Regulation by Linear Least Squares Methods*. (2a ed.). Basil Blackwell Puhlisher.
- Wichern, D. W. (1973): “The behavior of the sample autocorrelation function for an integrated moving average process”, *Biometrika*, 60, 235-239.
- Winters, P. R. (1960): “Forecasting sales by exponentially weighted moving averages”, *Management Science*, 6, 324-342.
- Wold, H. O. (1938): *A Study in the Analysis of Stationary Time Series*, Almquist and Wicksell, Uppsala.
- Yaglom, A. M. (1962): *An introduction to the theory of stationary random functions*. Dover Books.
- Young, P. (1984): *Recursive Estimation and Time Series Analysis*. Springer-Verlag.
- Zellner, A. (1979): “Statistical Analysis of Econometric Models” (con discusión), *Journal of the American Statistical Association*, 74, 628-643.

Bibliografía

—, y Palm, F. (1974): "Time Series Analysis and Simultaneous Equation Econometric Models", *Journal of Econometrics*, 2, 17-54.

# Índice analítico

- Abraham, 36, 73, 190  
Ahn, 569  
Akaike, 32, 33, 298, 299, 304, 313  
algoritmo  
    de Hannan y Rissanen, 286  
algoritmos  
    de optimización no lineal, 305  
alisado  
    doble, 56  
    exponencial simple, 176  
    simple, 52  
Alonso, 16, 334  
análisis  
    en el dominio frecuencial, 101  
    en el dominio temporal, 101  
Andersen, 36, 405  
Anderson, 36, 303, 404, 423  
Andrews, 393  
Aoki, 33, 36  
atípico  
    aditivo, 367  
    innovativo, 371  
autocorrelación  
    parcial, 210  
Aznar, 16  
Bai, 393  
Banerjee, 545  
Bartlett, 32, 35, 78, 96, 496  
Beran, 185, 190  
Bollerslev, 427  
Box, 16, 18, 32–34, 36, 189, 200,  
    234, 240, 243, 247, 251, 252,  
    272–274, 291, 315–317, 338,  
    342, 366, 417, 463, 469, 481,  
    503, 527, 569  
Brillinger, 36, 73  
Brock, 420  
Brockwell, 36, 73, 291, 303  
cambio  
    de nivel, 377  
    transitorio, 380  
Cancelo, 243, 484, 545  
Carnero, 16, 393, 455

- Chan, 40
- Chang, 392
- Chen, 36, 392
- Cholette, 243
- ciclos, 61
- Clements, 36, 243
- coeficiente
  - de asimetría, 87
  - de curtosis, 87
- coeficientes
  - de autocorrelación
    - parcial, 128, 139
  - de correlación, 84
  - de correlación inversa, 359
  - estacionales, 59
- cointegración, 532, 563
- componente
  - estacional, 58, 231
  - permanente, 225
  - transitorio, 225, 231
- contraste
  - BDS, 420
  - con el criterio BIC, 317
  - de estabilidad, 322
  - de homocedasticidad, 319
  - de Ljung-Box, 315
  - de normalidad, 319
  - del determinante, 316
- contrastos
  - de autocorrelación, 314
- correlograma, 95
- Cox, 251
- criterio
  - AIC, 299, 309
  - BIC, 300; 310
- curtosis, 433, 442
- Dagum, 243
- Davidson, 545
- Davies, 73, 291
- Davis, 36, 303
- Deistler, 34, 36
- del Hoyo, 16
- dependencia
  - estacional, 200
  - regular, 200
- Dickey, 33, 257, 259–263; 273–275, 565
- Diebold, 243
- diferencia
  - de martingalas, 91, 103
  - estacional, 198
- diferenciación
  - estacional, 255
  - regular, 253
- Diggle, 484
- distribuciones
  - finito-dimensionales, 79
- Durbin, 36, 303, 488
- ecuación
  - característica, 117, 123
  - de estado, 289
  - de observación, 288
  - de predicción final, 221
  - en diferencias, 133
- ecuaciones
  - de Yule-Walker, 124
- efecto
  - rampa, 380
- efectos de intervención, 26
- Enders, 36
- Engle, 33, 36; 425, 426, 530
- ergodicidad, 94
- escalón, 377
- Escribano, 16, 564
- Espasa, 16, 243, 484; 545
- espectro, 100
- esperanza
  - condicionada, 216
- estacionalidad, 22, 61, 195, 325
- estacionariedad
  - en sentido débil, 87
  - sentido estricto, 87
- estimación
  - MV
    - condicional, 285
    - exacta, 287

## Análisis de series temporales

---

- EViews, 35, 39, 106, 140, 192, 213, 275, 312, 338, 365, 397, 458
- factores, 117, 123  
comunes, 27, 563
- familia  
de Box-Cox, 417
- Fang, 36, 423
- filtro  
de Kalman, 288, 292
- Flores, 16
- Frances, 36, 212, 423
- frecuencia, 63  
angular, 63  
de Nyquist, 70
- Fried, 16
- Fuller, 33, 36, 96, 257, 259–263, 273–275, 303, 565
- función  
de autocorrelación, 85, 88  
muestral, 95  
cruzada, 494  
inversa, 162  
parcial, 126, 128, 144  
simple, 113, 144
- de covarianzas  
cruzadas, 491
- de medias, 82
- de respuesta a impulsos, 342, 352, 497
- de respuesta a un escalón, 347
- de transferencia, 29, 351, 352, 497
- de varianzas, 84
- de verosimilitud, 278
- función de respuesta al impulso, 344
- Gómez, 35, 212, 295, 364, 506
- Galeano, 17, 304, 392, 393
- ganancia, 345, 347
- García Ferrer, 16
- Gauss, 35
- Gini, 35
- Gonzalo, 16
- Gourieroux, 36, 455, 545
- Granger, 16, 33, 36, 73, 160, 188, 242, 243, 404, 405, 423, 425, 530
- Grenander, 363
- Guerrero, 243
- Hamilton, 36, 189, 210, 545
- Hannan, 34, 36, 286, 308, 553
- Harrison, 32, 33, 36, 243, 291, 303
- Harvey, 33, 36, 291, 295, 303, 427, 453
- Hatanaka, 36, 73
- Hendry, 36, 243
- heterocedasticidad  
condicional, 426
- Holt, 32, 44, 52, 56, 58, 73, 74, 177
- interacción, 205
- interpretación  
de las predicciones, 224
- intervalos  
de predicción, 331  
mediante remuestreo, 332
- Jacquier, 427, 453
- Jenkins, 18, 32, 33, 36, 189, 200, 234, 243, 247, 272, 291, 463, 503, 527
- Jerez, 16
- Justel, 16
- Kaiser, 16
- Kalman, 32, 277, 278, 288, 292, 295, 296, 303–305, 364, 506
- Kilian, 243
- Koopman, 36, 303
- Koreisha, 287
- Ledolter, 36, 73, 190
- Li, 419, 446
- Ljung, 315–317, 338, 469, 481
- Luceño, 16, 304, 569
- Lutkepohl, 36, 569
- mínimos cuadrados, 46, 66, 74

## Índice analítico

- generalizados, 524  
Métodos de alisado, 52  
Manzanedo, 16  
Maravall, 16, 35, 212, 295, 364, 419, 506  
Marquard, 308  
Matlab, 35, 42, 107, 141, 193, 275, 461  
matlab, 339  
matriz  
    de autocorrelación, 88  
    de autocorrelación simple y parcial, 550  
    de covarianzas, 88  
    de Toeplitz, 88  
Mauricio, 304, 569  
McLeod, 419, 446  
media móvil, 60  
medida  
    de predecibilidad, 240  
Minitab, 34, 35, 37, 106, 274, 312, 338  
Mira, 16  
modelo  
    ARCH, 429  
    ARMA  
        vectorial, 552  
    de corrección de error, 262, 565  
    en el espacio de los estados, 288  
    estacional, 202  
    factorial, 566  
    GARCH, 439  
    lineal multivariante, 550  
    multivariante, 29  
    regular, 202  
    univariante, 27  
modelos  
    autorregresivos  
        por umbrales, 409  
    de intervención, 356  
    estructurales, 291  
    periódicos, 212  
    STAR, 414  
Monfort, 36, 73, 545  
Moore, 303  
Morris, 160  
Muth, 32  
número complejo, 75  
Nelson, 455  
Novales, 16, 545  
operador  
    diferencia, 132  
    inverso, 133  
    retardo, 131  
Palm, 34  
Pandit, 484  
Pankratz, 36, 272, 484, 545  
paseo  
    aleatorio, 172  
Peña, 36, 51, 204, 243, 255, 291, 298, 301, 304, 316, 334, 364, 392, 393, 413, 419, 422, 446, 455, 564, 566, 569  
periodograma, 68, 100  
periodos  
    básicos, 69  
    de Fourier, 69  
Perron, 393  
Ploberger, 393  
polinomio  
    característico, 119  
Polson, 427, 453  
Poncela, 16, 566, 569  
Pourahmadi, 36  
Prat, 16  
predicción  
    adaptación, 239  
    con incertidumbre, 336  
    mediante promedio de modelos, 334  
Priestley, 33, 36, 73  
Prieto, 16  
proceso  
    autorregresivo, 110  
    de markov, 172  
    de media móvil, 143

- de memoria larga, 184
- de perturbación o de inercia, 497
- de ruido blanco, 90
  - estricto, 91
  - normal, 91
  - ergódico, 93
  - estacional, 230
  - estacionario, 86
    - combinación de, 89
    - suma, 158
  - estocástico, 79
  - fraccional, 184
  - integrado, 167, 179, 181
  - invertible, 143
  - lineal
    - general, 152
    - martingala, 172
    - VAR, 554
    - VARMA, 558
    - vectorial
      - estacionario, 548
- procesos
  - autorregresivos
    - por umbrales, 409
    - bilineales, 404
  - promedio Bayesiano de modelos, 335
  - Pukkila, 287
  - Quinn, 455
  - R, 35
  - raíces
    - unitarias, 256
  - Redondas, 16
  - regresión dinámica, 29
  - Reinsel, 36, 189, 243, 291, 569
  - relación de cointegración, 532
  - relación señal-ruido, 495
  - relaciones
    - espurias, 512
  - Rissanen, 286
  - Rodríguez, 16, 316, 413, 419, 422, 446, 455
  - Romera, 16
  - Romo, 16, 332, 334
- Rosenblatt, 363
- Rossi, 428, 453
- ruido
  - blanco
    - estricto, 403
- Ruiz, 16, 332, 393, 427, 446, 453, 455
- S-plus, 34
- Sánchez, C., 35
- Sánchez, I., 16, 255, 298, 301, 304
- Sánchez, M. J., 16, 392, 393
- SCA, 35, 312
- Schwarz, 300, 304
- SEATS, 35
- serie
  - desestacionalizada, 60
  - estacional, 18
  - estacionaria, 18
  - no estacionaria, 18
- SETAR, 410
- Shephard, 427, 453
- Shumway, 36, 303, 569
- sistema
  - estable, 353
- SPSS, 34, 35, 38, 106, 274, 312, 338
- Statgraphics, 34, 35, 37
- Stock, 569
- Stoffer, 36, 304, 569
- Stram, 243
- TAR, 409
- tendencia, 19, 45, 181, 323
  - determinista, 44, 49
  - lineal, 45, 51
- teorema
  - de Wold, 151
- Terceiro, 16
- Tiao, 16, 33, 34, 36, 240, 243, 272, 291, 331, 342, 366, 392, 413, 419, 455, 569
- Tong, 33, 36, 399, 409, 412, 423
- TRAMO, 35
- Treadway, 16

## Índice analítico

---

- Tsay, 16, 33, 36, 243, 272, 291, 392,  
393, 413, 418, 419, 423, 455,  
484, 569  
TSW, 35, 41, 275, 312, 339, 364,  
389, 396  
validación cruzada, 298  
valores  
ausentes, 358, 362  
van Dijk, 423  
variable  
de umbral, 410  
escalón, 342  
impulso, 342  
Velilla, 16  
Velu, 36, 569  
volatilidad  
estocástica, 453  
Watson, 488, 569  
Watts, 36, 210  
Wei, 36, 73, 243, 569  
West, 33, 36, 243, 291, 303  
Whittle, 32  
Wichern, 349  
Winters, 32, 44  
Wold, 31, 35, 142, 149, 151, 152,  
163, 167, 185, 402–404, 552,  
590  
Wu, 484  
Xu, 331  
Yao, 36, 423  
Yohai, 16  
Zellner, 34

