

**XXVIII ESCUELA VENEZOLANA DE MATEMÁTICAS
EMALCA-VENEZUELA 2015**

**INTRODUCCIÓN AL ANÁLISIS DE SERIES
DE TIEMPO CON APLICACIONES
A LA ECONOMETRÍA Y FINANZAS**

**Abelardo Monsalve
y Pedro Harmath**

MÉRIDA, VENEZUELA, 30 de agosto al 4 de septiembre de 2015

XXVIII ESCUELA VENEZOLANA DE MATEMÁTICAS
EMALCA - VENEZUELA 2015

INTRODUCCIÓN AL ANÁLISIS DE SERIES DE TIEMPO
CON APLICACIONES A LA ECONOMETRÍA
Y FINANZAS

Abelardo Monsalve y Pedro Harmath

Universidad Centroccidental "*Lisandro Alvarado*"

amonsalve@ucla.edu.ve
pedro.harmath@ucla.edu.ve

MÉRIDA, 30 DE AGOSTO AL 4 DE SEPTIEMBRE DE 2015

XXVIII ESCUELA VENEZOLANA DE MATEMÁTICAS

La Escuela Venezolana de Matemáticas es una actividad de los postgrados en matemáticas de las instituciones siguientes: Centro de Estudios Avanzados del Instituto Venezolano de Investigaciones Científicas, Facultad de Ciencias de la Universidad Central de Venezuela, Facultad de Ciencias de la Universidad de Los Andes, Universidad Simón Bolívar, Universidad Centroccidental Lisandro Alvarado y Universidad de Oriente, y se realiza bajo el auspicio de la Asociación Matemática Venezolana.

La XXVIII Escuela Venezolana de Matemáticas recibió financiamiento de la Academia de Ciencias Físicas, Matemáticas y Naturales de Venezuela, el Banco Central de Venezuela, el Fondo Nacional de Ciencia, Tecnología e Innovación (FONACIT), el Instituto Venezolano de Investigaciones Científicas (Centro de Estudios Avanzados, Departamento de Matemáticas y Ediciones IVIC), la Universidad de los Andes (CEP, CDCHT, Departamento de Matemáticas de la Facultad de Ciencias, Decanato de Ciencias y Vicerrectorado Administrativo), la Unión Matemática de América Latina y el Caribe (UMALCA) y Centre International de Mathématiques Pures et Appliquées (CIMPA).

2010 Mathematics Subject Classification: 60Gxx, 37M10, 62P05, 91G70

©Ediciones IVIC

Instituto Venezolano de Investigaciones Científicas

Rif: G-20004206-0

Introducción al Análisis de Series de Tiempo con Aplicaciones a la Econometría y Finanzas

Abelardo Monsalve Pedro Harmath

Diseño y edición: Escuela Venezolana de Matemáticas

Depósito legal lfi66020155102240

ISBN 978-980-261-163-8

Caracas, Venezuela

2015

A las instituciones y personas que cada año se esfuerzan por llevar a cabo la Escuela Venezolana de Matemáticas, y en particular su XXVIII edición.

Índice general

Prólogo	VII
1. Econometría Financiera	1
1.1. Rentabilidad y Valoración de Activos Financieros	2
1.1.1. Rentabilidad Simple	4
1.1.2. Rentabilidad Continua	6
1.2. Estadística en las Finanzas	7
1.3. Otras Series Financieras	11
2. Modelos de Series de Tiempo	17
2.1. Modelos Lineales	24
2.1.1. El Proceso Ruido Blanco	24
2.1.2. Modelos Autorregresivos	26
2.1.3. Modelos de Medias Móviles	38
2.1.4. Modelos <i>ARMA</i>	45
2.1.5. Representaciones Alternativas de un Proceso <i>AR-MA</i>	49
3. Modelos No-Estacionarios	53
3.1. No estacionariedad en la Varianza	55
3.2. No estacionariedad en la Media	55
3.3. Test de Raíz Unitaria	58
4. Otros Modelos	61
4.1. Modelos Estacionales	61
4.2. Modelos de Memoria Larga	62
4.3. Modelos de Regresión de Series de Tiempo	64
4.3.1. Estimación consistente de la Matriz de Covarianza	74

5. Modelos de Heterocedasticidad Condicional	77
5.1. Estructura de los Modelos	79
5.2. Modelos <i>ARCH</i>	80
5.3. Modelos <i>GARCH</i>	84
5.4. Modelos <i>EGARCH</i>	89
5.5. Modelos <i>IGARCH</i>	92
5.6. Modelos <i>GARCH-M</i>	93
5.7. Modelos <i>TGARCH</i>	94
5.8. Modelos de Volatilidad Estocástica	95
5.8.1. Extensiones del Modelo <i>SV</i>	96
6. Modelos No Lineales	99
6.1. Modelos No Lineales para la Esperanza Condicional . . .	101
6.2. Modelos <i>TAR</i>	101
6.3. Modelos <i>SETAR</i>	102
6.4. Modelos <i>STAR</i>	103
6.5. Modelos Markov Switching	104
6.6. Métodos No-Paramétricos	105
6.6.1. Regresión por Núcleo	107
6.6.2. Selección del Parámetro de Suavizado	108
6.6.3. Método de Regresión Local Lineal	109
6.6.4. Aplicación a Series de Tiempo	111
6.7. Modelo de Coeficiente Funcional Autorregresivo	113
6.8. Modelo No-Lineal Autorregresivo Aditivo	113
6.9. Modelo No-Lineal de Espacio de Estado	114
6.10. Tests de No-Linealidad	115
6.10.1. Test No-Paramétrico	115
6.10.2. Estadístico de los Residuos al Cuadrado	115
6.10.3. Test Paramétricos	116
6.10.4. El test RESET	116
6.10.5. El test <i>F</i>	117
6.10.6. Test de Umbral	118
7. Modelos en Tiempo Continuo	121
7.1. Movimiento Browniano	122
7.2. Puente Browniano	125
7.3. Movimiento Browniano con Tendencia	125
7.4. Movimiento Browniano Geométrico	126

7.5. Proceso de Itô	126
7.5.1. Lema de Itô	129
7.5.2. Aplicación al Precio de los Activos	130
7.5.3. Estimación de μ y σ	131
7.6. Distribución del Precio de los Activos y la Rentabilidad Continua	133
7.7. Procesos de Difusión con Salto	134
7.8. Modelo de Volatilidad Estocástica	137
7.9. Estimación de los Modelos en Tiempo Continuo	138
7.9.1. Métodos basados en la Función de Verosimilitud	139
7.9.2. Método Generalizado de los Momentos	139
A. Nociones básicas de Probabilidad	145
A.1. Espacio de Probabilidad y Variable Aleatoria	145
A.2. Valor Esperado, Varianza y Momentos	148
A.3. Variable Aleatoria n -dimensional	149
A.4. Independencia	150
A.5. Distribución Condicional	151
A.6. Esperanza Condicional	152
A.7. Tipos de Convergencia	154
B. Aspectos Generales de los Procesos Estocásticos	157
B.1. Procesos Estocásticos	157
B.2. Filtraciones	158
B.3. Momentos, Covarianza e Incrementos de un Proceso Es- tocástico	159
B.4. Variación de un Proceso	159
B.5. Martingalas	160
B.6. Propiedad de Markov	161
C. Elementos de Cálculo Estocástico	163
C.1. Integración Estocástica	163
C.1.1. La Integral de Itô	165
C.1.2. Propiedades de la Integral de Itô	166
C.2. Fórmula de Itô	168
C.3. Ecuaciones diferenciales Estocásticas	169
C.3.1. Ecuaciones de Kolmogorov	173

Prólogo

Diversas son las áreas del conocimiento: finanzas, economía, estadística, probabilidad y matemática aplicada, las que se han integrado en el campo de la econometría financiera. Una gran variedad de modelos estocásticos se han desarrollado con la finalidad de comprender y caracterizar la dinámica de las distintas variables económicas involucradas en la actividad económica y en los mercados financieros. La estadística se ha convertido en una herramienta de uso común, y si se quiere, indispensable en la identificación de los parámetros de los distintos modelos propuestos, en la simulación de sistemas financieros complejos y en la validación de las teorías económicas surgidas a partir de los datos generados por la actividad de los mercados financieros y la economía global.

En la econometría financiera, uno de los principales tópicos de estudio es el relacionado con la rentabilidad esperada y la dinámica que caracteriza el precio de los activos financieros. La globalización de los mercados de capitales ha traído como consecuencia un aumento de la volatilidad de las variables que caracterizan a la economía de los estados y los mercados financieros en todo el mundo. Esto ha despertado el interés, tanto de los profesionales de las finanzas como de los académicos, ya que la caracterización de la dinámica de tales variables permite la determinación de su estructura temporal, la valoración de los precios de una gran variedad de activos financieros, el diseño de estrategias de cobertura e inversión y la evaluación de riesgos. Desde el punto de vista macroeconómico, tiene una relevancia especial en cuanto a la determinación de una adecuada política monetaria y de los distintos canales de transmisión, para las relaciones entre las tasas de interés a corto y largo plazo para la formación de expectativas. Basados en la información actual, se requiere una perspectiva a futuro de las variables asociadas a dichas actividades, así pues, se hace vital el conocimiento de técni-

VIII

cas y herramientas que permitan caracterizar a partir de descripciones estocásticas su evolución en el futuro.

El análisis de series de tiempo, tanto sus aspectos teóricos como empíricos, ha sido por muchos años una parte integral en el estudio de los mercados financieros y de las políticas económicas de los estados; principalmente por el interés de estos últimos en la planificación, coordinación y control de la economía y de los primeros por la búsqueda de condiciones que permitan rentabilizar sus inversiones.

Desde la presentación de los trabajos de George E. P. Box and Gwilym M. Jenkins, "*Time Series Analysis: Forecasting and Control*", el análisis de series ha tenido un desarrollo creciente. Una gran variedad de libros sobre este tópico se han presentado desde entonces. Cada texto se ve influenciado principalmente por la orientación de las series que se discuten en sus contenidos. Una gran parte de la literatura está dirigida a exponer los aspectos teóricos alrededor de las series de tiempo, siendo en muchos casos, rigurosamente desarrollados y descritos, pero dejando de lado al lector la vital implementación de las técnicas expuestas y su comprensión final en ejemplos realistas. Por ello, este curso pretende dar a conocer las técnicas más comunes para entender y llevar a cabo un análisis empírico de las series de tiempo en el contexto de la econometría financiera. Los aspectos teóricos serán expuestos de forma simple y sencilla, y en los casos en que sea necesario, serán especialmente rigurosos sin dejar de lado el objetivo fundamental del curso, la exposición de ejemplos que permita entender como se implementan y la importancia de la información que estos aportan.

El curso se organiza de la siguiente forma:

- El primer capítulo está referido a los conceptos básicos de la econometría financiera, base fundamental de conocimiento necesario para entender los modelos de econométricos y su dinámica.
- Un segundo capítulo en donde se discute el concepto de estacionariedad y se presentan de los modelos lineales: sus características y propiedades más relevantes.
- En el tercer capítulo se expone los modelos estacionarios y las estrategias para el estudio de series con dicha característica.
- El capítulo 4 da a conocer los modelos estacionales y los de regre-

sión de series de tiempo, para culminar con una breve introducción a los modelos de memoria larga.

- El capítulo 5 desarrolla la teoría de los modelos en los que se pretende estudiar el comportamiento de la volatilidad de una serie de tiempo, dichos modelos son mejor conocidos como los modelos de heterocedasticidad condicional.
- En los capítulo 6 y 7 se hace una breve presentación de modelos de series de tiempo no lineales y los modelos continuos, con algunas consideraciones particulares.

Algunos de los modelos expuestos en los capítulos antes mencionados son enriquecidos con aplicaciones a series de datos relativos a activos del mercado financiero español, y las series de las tasas de interés interbancario de la zona Euro.

Abelardo Monsalve y Pedro Harmath
Caracas, Mayo 2015

1

Econometría Financiera

Una gran parte de las series de tiempo que se estudian se producen en el campo de la economía, donde estamos continuamente expuestos a diario a cotizaciones bursátiles o a las cifras mensuales de desempleo, precios del petróleo, entre muchos otros indicadores. El enfoque general, es el dominio del tiempo, motivado por la presunción de que la correlación entre puntos próximos en el tiempo se explica mejor en términos de una dependencia del valor actual con los valores pasados. El análisis de series de tiempo, ha sido durante muchos años una parte fundamental en el estudio de los mercados financieros. Las series de tiempo han adquirido una importancia relevante como herramienta para el análisis teórico y práctico de la valoración de activos en el tiempo. Por ejemplo, el pronóstico de cambios en los precios de activos es un tema de investigación de gran interés. La teoría financiera así como las series de tiempo empíricas asociadas, están caracterizadas por un elemento de *incertidumbre*, característica clave que diferencia el análisis de series de tiempo en finanzas del análisis de series de tiempo tradicional. Por ello, la teoría y los métodos estadísticos desempeñan un papel importante en el análisis de las series de tiempo en finanzas. El objetivo de este capítulo es discutir las teorías básicas del análisis de series de tiempo. Para ello introduciremos algunos modelos eco-

nométricos simples, que son de gran utilidad en el análisis de series de tiempo financieras. Los conceptos desarrollados a través de este capítulo serán tratados de manera breve, es decir, sin profundizar en las teorías económicas subyacentes, pero haciendo énfasis en aquellos que consideremos relevantes para las aplicaciones financieras. Previamente introduciremos algunas nociones básicas sobre precios, rentabilidades y selección de carteras, entendiendo que el conocimiento del contexto general que conlleva al planteamiento de un modelo nos ayuda a comprenderlo y hace más fácil su estudio. Para una discusión exhaustiva de los conceptos básicos del análisis lineal de series de tiempo, véase, Box et al. (1994) (capítulos 2 y 3) y Brockwell and Davis (2002).

1.1. Rentabilidad y Valoración de Activos Financieros

Dentro del análisis financiero, es fundamental conocer la influencia de la información del pasado en la información futura, más específicamente, es determinante saber hasta qué punto el pasado de una serie de precios de un activo proporciona información relevante para predecir su comportamiento futuro.

En el análisis de las series de tiempo en finanzas, son diversos los enfoques considerados. Uno de ellos es el presentado por Working (1934), quien centró la atención en una característica notada previamente en los precios de las acciones, a saber: que se asemejan a las situaciones de acumulación de cambios puramente al azar. Algunas teorías se inclinan por la predecibilidad de precios futuros en función del pasado de la serie. Así, a partir de los años 50, se publicaron numerosos artículos en los que se investigaban los cambios en el precio de activos. Entre los investigadores que estudiaron las variaciones en el precio de los activos se encuentran, Kendall (1953), Osborne (1959) y Fama (1965).

Una de las hipótesis manejadas en series de alta frecuencia, sugiere que los precios de activos financieros (o sus logaritmos) se comportan como un paseo aleatorio, es decir, si P_t denota el precio de un activo en el instante t y $p_t = \ln(P_t)$, se tiene

$$p_t = p_{t-1} + u_t, \quad (1.1)$$

donde u_t es una sucesión de variables independientes con media 0.

Desde un punto de vista estadístico, esta suposición quiere decir, que las variaciones en los precios son variables aleatorias independientes e idénticamente distribuidas, lo cual implicaría que la serie de cambios carece de memoria y la información del pasado no es de utilidad para predecir el futuro eficientemente. El análisis financiero también está interesado en determinar la distribución de las variaciones de los precios. Las hipótesis de independencia y normalidad en la distribución de u_t , implica que los precios (o sus logaritmos) están generados como un movimiento Browniano. Aún cuando la hipótesis de normalidad es bastante clásica en finanzas, es cuestionable. Esto se debe a ciertas características de las series de precios, las cuales serán abordadas más adelante. Asimismo la hipótesis de independencia de la serie de las variaciones del precio de los activos es fuertemente restrictiva. Esto se puede deducir a partir de los fundamentos teóricos de los mercados financieros, desarrollados en los años 60. El hecho de suponer un modelo como (1.1), donde u_t son independientes, elimina la dependencia del tiempo de los momentos condicionales de $p_t - p_{t-1}$ de cualquier orden. Ciertamente, la mayoría de las series financieras presentan un comportamiento errático, en los que se alternan períodos de estabilidad con períodos de turbulencia. Esta dinámica presente en las series financieras, puede ser caracterizada permitiendo que el momento de segundo orden condicional sea una función que dependa del tiempo, en consecuencia, la hipótesis del paseo aleatorio quedaría descartada.

En finanzas se suele considerar, con mucha frecuencia, las series de rentabilidades en lugar de las series de precios. Campbell et al. (1997) dan dos razones respecto de la preferencia entre dichas series. La primera de ellas, esta relacionado con los inversores, quienes afirman que la serie de rentabilidades de un activo es una síntesis completa y libre de escala de la oportunidad de inversión. Y la segunda razón tiene que ver con el hecho de que, las series de rentabilidades presentan, en general, mejores propiedades estadísticas que las series de precios de activos, las cuales suelen mostrar una tendencia a largo plazo como se puede apreciar en la figura (1.1) mientras que a corto plazo o a medio plazo presentan movimientos de crecimiento y decrecimiento. Esto arroja como resultado que al tomar pequeñas muestras de la serie de precios correspondientes a un mismo activo pero tomadas en diferentes períodos estas exhiban medias distintas, con lo cual se dificulta el

análisis estadístico.

1.1.1. Rentabilidad Simple

En la toma de decisiones a la hora de invertir en un activo, tiene especial relevancia el hecho de que dicho activo históricamente genera beneficios. Las utilidades obtenidas a partir de la inversión o las pérdidas dependen tanto de la variación de los precios como de la cantidad de activos que se dispongan. Los inversionistas están interesados principalmente en que los beneficios sean mayores en relación con el tamaño de las inversiones iniciales. Una manera de medir esto es a partir de la rentabilidad. La *rentabilidad* indica la variación, expresada generalmente en términos porcentuales, experimentada por el valor de un activo durante un cierto período de tiempo. Dicha variación puede ser positiva o negativa.

Sea P_t el precio de un activo en el instante t . A continuación se presentan algunas definiciones de gran utilidad en el contexto del análisis financiero. Supondremos que el activo no paga dividendos o, como puede suceder en algunos casos, el precio incluye el pago de dividendos.

Definición 1.1.1. *La rentabilidad simple o discreta de un activo en el período $(t - 1, t)$ se define como*

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}. \quad (1.2)$$

El numerador de la expresión (1.2), es el beneficio durante el período de tenencia (tiempo en que el activo permanece en manos del inversor), si el beneficio es negativo se denomina pérdida. El denominador, P_{t-1} , es el precio inicial de la inversión al comienzo del período de tenencia. Por lo tanto la rentabilidad simple se puede ver como la *tasa de beneficio o rentabilidad relativa*.

Definición 1.1.2. *La rentabilidad bruta se define entonces como*

$$1 + R_t = \frac{P_t}{P_{t-1}}, \quad \text{o bien} \quad P_t = P_{t-1}(1 + R_t). \quad (1.3)$$

La definición (1.1.1) hace referencia a la rentabilidad simple para el período $(t - 1, t)$. En general, para un período $(t - k, t)$, con $k > 1$,

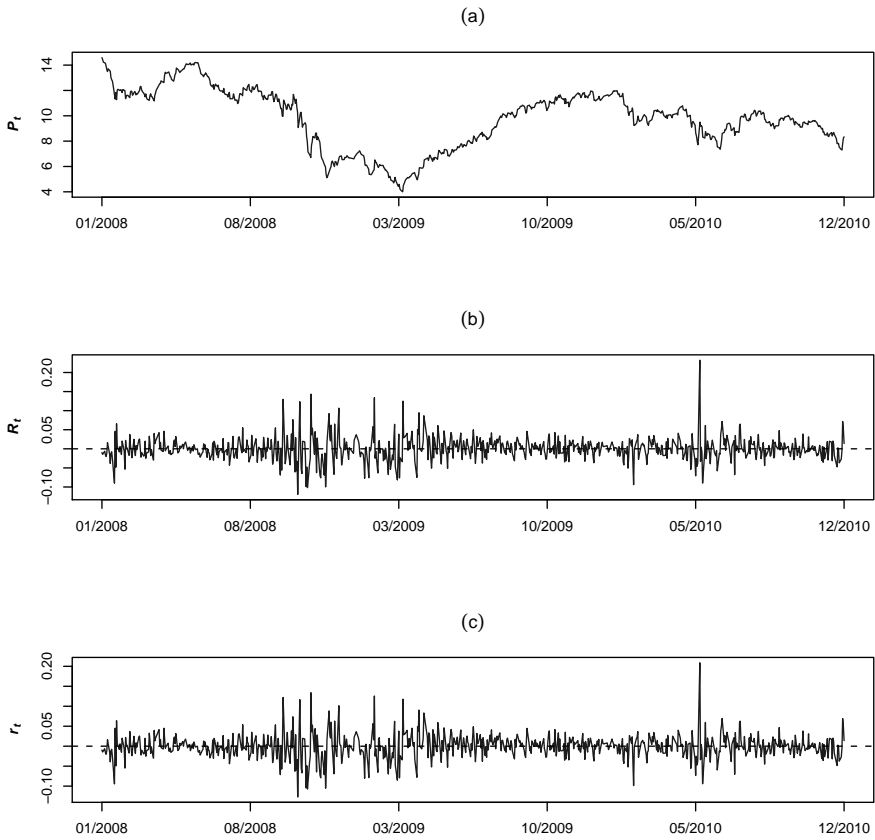


Figura 1.1: Observaciones diarias de los activos del Santander en el IBEX35, en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre de 2010: (a) precio de cierre, (b) rentabilidad simple, (c) rentabilidad continua.

la rentabilidad simple se expresa por

$$R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}}. \quad (1.4)$$

La ecuación (1.4) hace referencia a la rentabilidad simple para k períodos. Así mismo, la rentabilidad bruta para k períodos o bien en el período $(t - k, t)$, se define como el producto de k rentabilidades simples

$$\begin{aligned} 1 + R_t(k) &= \frac{P_t}{P_{t-k}} = \left(\frac{P_t}{P_{t-1}} \right) \left(\frac{P_{t-1}}{P_{t-2}} \right) \cdots \left(\frac{P_{t-k+1}}{P_{t-k}} \right) \\ &= (1 + R_t) \cdots (1 + R_{t-k+1}). \end{aligned} \quad (1.5)$$

Si el período no es dado, entonces se asume que este es de un año. Si los activos se mantienen por k años, entonces la rentabilidad anualizada se define como

$$\begin{aligned} \text{Annualized } (R_t(k)) &= \left[\prod_{j=0}^{k-1} (1 + R_{t-j}) \right]^{1/k} - 1 \\ &= \exp \left[\frac{1}{k} \sum_{j=0}^{k-1} \ln(1 + R_{t-j}) \right] - 1. \end{aligned}$$

La rentabilidad no depende de ningún tipo de unidad monetaria, su unidad es el tiempo t (horas, días, etc.)

1.1.2. Rentabilidad Continua

Las entidades financieras, bancos e inversores hacen uso de la rentabilidad simple para sus aplicaciones empíricas. No obstante, las rentabilidades continuas presentan propiedades estadísticas más atractivas en cuanto a que son más manejables. Por esta razón, en el ámbito académico y de investigación son las rentabilidades continuas las de mayor frecuencia.

Definición 1.1.3. *La rentabilidad continua de un activo en el período $(t - 1, t)$ se define como*

$$r_t = \ln(1 + R_t) = \ln \left(\frac{P_t}{P_{t-1}} \right) = p_t - p_{t-1}, \quad (1.6)$$

donde $p_t = \ln(P_t)$. Para el período $(t - k, t)$,

$$r_t(k) = \ln(1 + R_t(k)) = r_t + r_{t-1} + \dots + r_{t-k+1}.$$

Las figuras (1.1) y (1.2) muestran las series de precios, rentabilidades simple y continua de los activos *Santander*, *Repsol* y *Telefónica* en el mercado de valores de España, concretamente en el IBEX-35 de la Bolsa de Madrid, con frecuencia diaria en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre 2010.

Si tenemos en cuenta el desarrollo de segundo orden de la función logaritmo, se tiene que

$$r_t = \ln(1 + R_t) \approx R_t - \frac{R_t^2}{2},$$

y por lo tanto r_t infravalora el verdadero rendimiento R_t .

Al observar las series de rentabilidades como las de las figuras (1.1) y (1.2) se pueden deducir una serie de características comunes. De forma general, podemos establecer que las series de rentabilidades presentan un comportamiento errático en el sentido de que los valores extremos de rentabilidad se dan con relativa frecuencia, por otro lado, existe agrupamiento de la volatilidad (volatility clustering), es decir, períodos de alta volatilidad suelen ser precedidos por períodos de baja volatilidad de mayor o menor duración. Esto último es especialmente notable en estas series, ya que el período que comprenden los datos incluye los episodios más destacados de la actual crisis, como por ejemplo, el agrupamiento de la volatilidad durante el verano y otoño del 2008, y en la primavera de 2010.

1.2. Estadística en las Finanzas

Para explicar el comportamiento de las series de rentabilidad, se considera a las rentabilidades r_t como variables aleatorias continuas, y además se supone, como es habitual en finanzas, que su distribución es una Normal con media μ_r y varianza σ_r^2 , y que se denota por $(r_t \sim N(\mu_r, \sigma_r^2))$. De los momentos centrales de una variable aleatoria podemos extraer un par de conceptos de gran utilidad. Como es sabido, el tercer momento central mide la simetría de una variable aleatoria con respecto a su

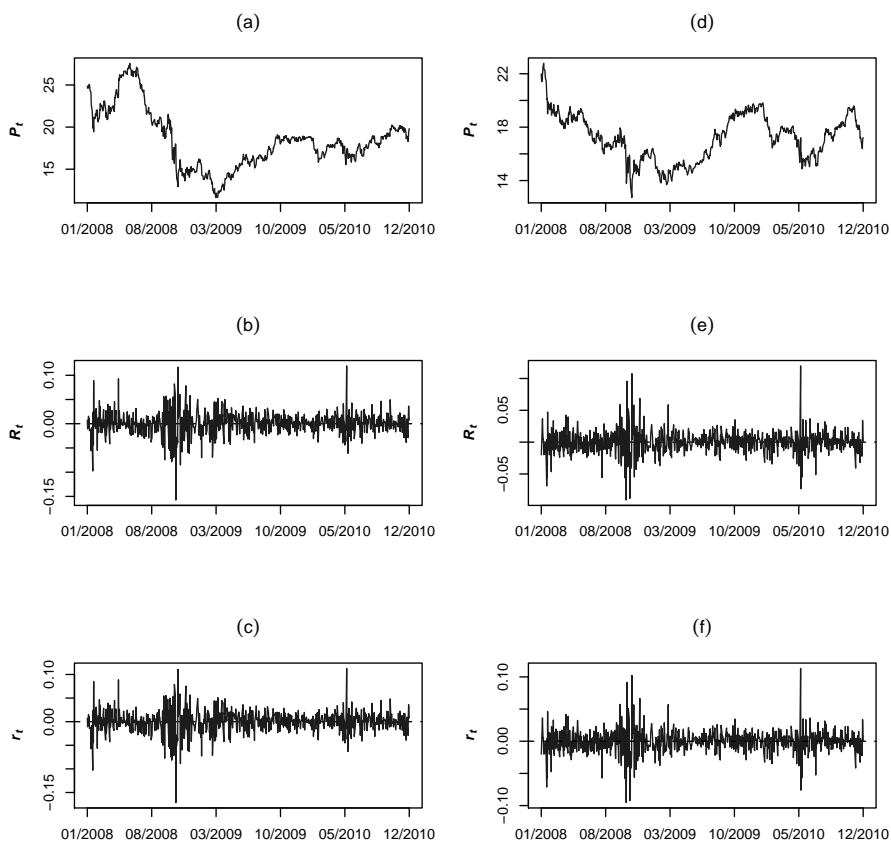


Figura 1.2: Observaciones diarias de los activos de Repsol y Telefónica en el IBEX35, en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre de 2010. Series del activo Repsol: (a) precio de cierre, (b) rentabilidad simple, (c) rentabilidad continua. Series del activo Telefónica: (d) precio de cierre, (e) rentabilidad simple, (f) rentabilidad continua.

valor esperado, mientras que el cuarto momento central mide el comportamiento de las colas de dicha variable, dicho de otra forma, mide la concentración de la información en las colas. En estadística, estos momentos son normalizados y son usados para determinar la asimetría y el comportamiento en las colas de la distribución de la variable en estudio. Así pues, el *coeficiente de asimetría SK (Skewness)* viene dado por

$$SK(x) = \mathbb{E} \left[\frac{(X - \mu_x)^3}{\sigma_x^3} \right], \quad (1.7)$$

y se estima mediante

$$\widehat{SK}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{t=1}^T (x_t - \hat{\mu}_x)^3. \quad (1.8)$$

El *coeficiente de curtosis K (kurtosis)* viene dado por

$$K(x) = \mathbb{E} \left[\frac{(X - \mu_x)^4}{\sigma_x^4} \right], \quad (1.9)$$

y se estima mediante

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{t=1}^T (x_t - \hat{\mu}_x)^4. \quad (1.10)$$

donde σ_x y μ_x son la media y la varianza de la variable en cuestión y $\hat{\mu}_x$ y $\hat{\sigma}_x$ sus respectivos estimadores. En realidad, la cantidad de interés es el *exceso de curtosis*, $K(x) - 3$, puesto que, $K(x) = 3$ para una distribución normal (Ver Apéndice B).

Retomando el análisis, se puede ver que la suposición de normalidad no resulta consistente con las propiedades empíricas que han demostrado las series históricas de rentabilidades, puesto que la tendencia de estas series tienen un exceso de *curtosis* (medida que determina el grado de concentración que presentan los valores en la región central de la distribución) positiva. Como se puede observar en la tabla (1.1), el exceso de curtosis ($\hat{K}(x) - 3$) reflejado en la columna curtosis, evidencia un valor alto en todos los casos, lo cual indica que las series tienen colas pesadas. En otras palabras, valores “grandes” de las rentabilidades (en términos absolutos) ocurren con mayor frecuencia de lo esperado para

una variable con distribución normal. En lo que respecta al coeficiente de asimetría, si la distribución de la variable de interés es simétrica entonces su valor es cero. También se observa que las series de las rentabilidades de los activos *Endesa* y *Repsol* reflejan un coeficiente negativo, poniendo de manifiesto que la cola izquierda de la distribución tiene mayor peso que la derecha, esto quiere decir que, valores negativos de rentabilidades ocurren más a menudo que valores positivos. Lo contrario para el resto de series cuyo coeficiente es positivo.

Activo	Mín	Máx	Media	Mediana	Varianza
Santander	-0,1194	0,2322	-0,0003	0	0,0010
BBVA	-0,1278	0,2203	-0,0006	-0,0010	0,0009
Endesa	-0,2815	0,1288	-0,0006	0,0000	0,0005
Iberdrola	-0,1258	0,1880	-0,0005	0,0008	0,0007
Iberia	-0,1214	0,2643	0,0006	0	0,0010
Inditex	-0,1032	0,1048	0,0008	0	0,0006
Repsol	-0,1575	0,1194	-0,0000	0,0005	0,0006
Telefónica	-0,0910	0,1198	-0,0002	0	0,0003

Activo	Desv	Asimetría	Curtosis
Santander	0,0309	0,8013	6,7419
BBVA	0,0293	0,7468	6,3464
Endesa	0,0234	-2,1358	28,9111
Iberdrola	0,0261	0,8026	9,6047
Iberia	0,0312	1,2481	9,8330
Inditex	0,0237	0,3385	2,5798
Repsol	0,0235	-0,2427	5,6627
Telefónica	0,0184	0,3511	6,9638

Cuadro 1.1: *Estadísticas Descriptivas para las rentabilidades simples de activos del IBEX-35 en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre de 2010.*

1.3. Otras Series Financieras

Es importante destacar que las series financieras no están relacionadas únicamente con la serie de precios de activos financieros. Entre las series financieras que analizaremos en este curso se tienen: las series de tasas de interés y las tasas de cambio. En la sección anterior se han descrito los conceptos de rentabilidades de un activo y conjuntamente se han presentado algunas series de rentabilidades de ciertos activos que serán tomados en cuenta en el desarrollo de esta primera parte de introducción al análisis de las series de tipos de interés financieras.

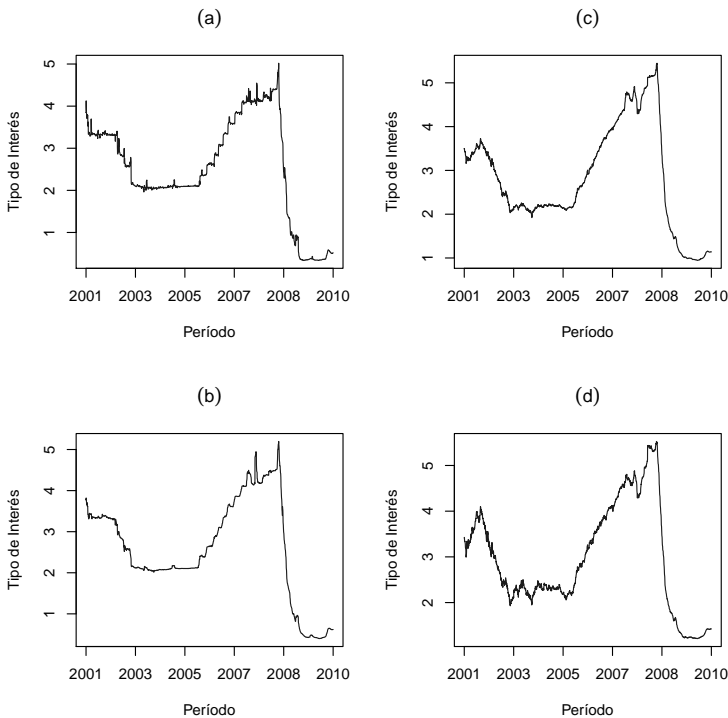


Figura 1.3: Series del *EURIBOR*, período comprendido entre el 15 de Octubre de 2001 al 3 de Diciembre de 2010. Plazo de vencimiento: (a) 1 semana, (b) 1 mes, (c) 6 meses, (d) 1 año (12 meses).

La figura (1.3) muestra las Series del *EURIBOR*, para los distintos pla-

zos de vencimiento, 1 semana, 1 mes, 6 meses, 1 año (12 meses), en el período comprendido entre 15 octubre 2001 al 3 de diciembre de 2010. Como es de esperar, los gráficos muestran que las series de los tipos de interés para distintos plazos se comportan casi de manera similar, sin embargo se puede apreciar que, por ejemplo, la volatilidad es distinta para cada uno de ellas. Para ver esto último, basta con observar las series de cambios o variaciones del tipo de interés en cada plazo en la figura (1.4), en donde se puede apreciar, que las variaciones se producen con distinta intensidad para cada plazo.

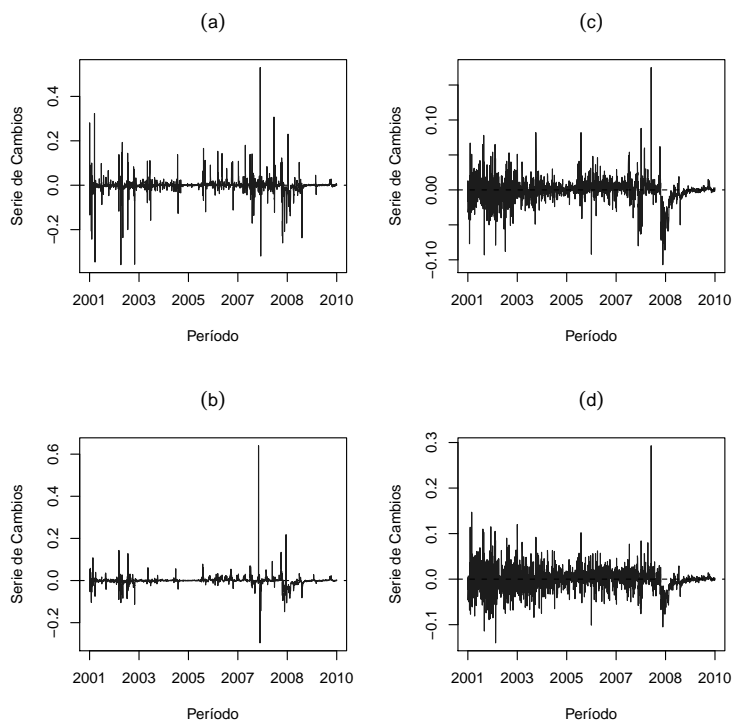


Figura 1.4: Series de las variaciones del tipo de interés del *EURIBOR*, período comprendido entre el 15 de Octubre 2001 al 3 de Diciembre de 2010. Plazo de vencimiento: (a) 1 semana, (b) 1 mes, (c) 6 meses, (d) 1 año (12 meses).

La tabla (1.2) presenta algunos de los estadísticos descriptivos más im-

portantes de las series de tipo de interés antes mencionadas. Por ejemplo, el máximo alcanzado en el *EURIBOR* a un año (12 meses) durante el 2008 fue el día 2 de octubre que consiguió llegar hasta el 5,526 %, durante la crisis. A continuación, su valor comenzó a bajar de forma progresiva como consecuencia de un aumento del crédito y un descenso de los tipos de interés por parte del Banco Central Europeo, llegando a su record histórico mínimo del 1,21 % el día 30 de Marzo de 2010. Otra

Plazo	Mín	Máx	Media	Mediana	Varianza
1 semana	0.34	5.02	2.52	2.37	1.47
1 mes	0.63	5.39	2.73	2.52	1.52
6 meses	0.94	5.45	2.82	2.56	1.43
12 meses	1.21	5.53	2.95	2.64	1.38

Plazo	Desv	Asimetría	Curtosis
1 semana	1.21	-0.31	-0.77
1 mes	1.23	0.13	-0.76
6 meses	1.19	0.29	-0.81
12 meses	1.17	0.37	-0.86

Cuadro 1.2: Estadísticas Descriptivas para las series de tipos de interés del *EURIBOR* en el período comprendido entre el 15 de Octubre de 2001 al 3 de Diciembre 2010.

de las series de gran interés en el contexto del análisis financiero en el ámbito europeo, es la serie de los tipos de cambio del *Dólar-Euro*. La figura (1.5) muestra los gráficos de los tipos de cambio y sus variaciones. En dichos gráficos se observa el impacto de la crisis durante el período de 2008-2009, y otros cambios destacables pero ocasionales a lo largo de la serie. Se puede observar la alta volatilidad para el mismo período. Teniendo en cuenta lo anterior, es claro que, sin ningún tipo de supuestos, las finanzas serían el reino de la incertidumbre. En el instante t , el precio P_t y la rentabilidad R_t no sólo son desconocidas, sino que, desde el punto de vista teórico, no se conocen sus distribuciones de probabilidad. Sin embargo, es posible estimarlas, si se hacen suposiciones acerca de que las rentabilidades en el futuro serán similares a rentabilidades

en el pasado, esta condición es conocida como *estacionariedad*. Con este supuesto, la maquinaria de la inferencia estadística puede ser aplicada y la distribución de probabilidad de P_t puede estimarse a partir de los datos del pasado o bien se puede caracterizar su dinámica a partir de un modelo dado. La siguiente sección discute este tema.

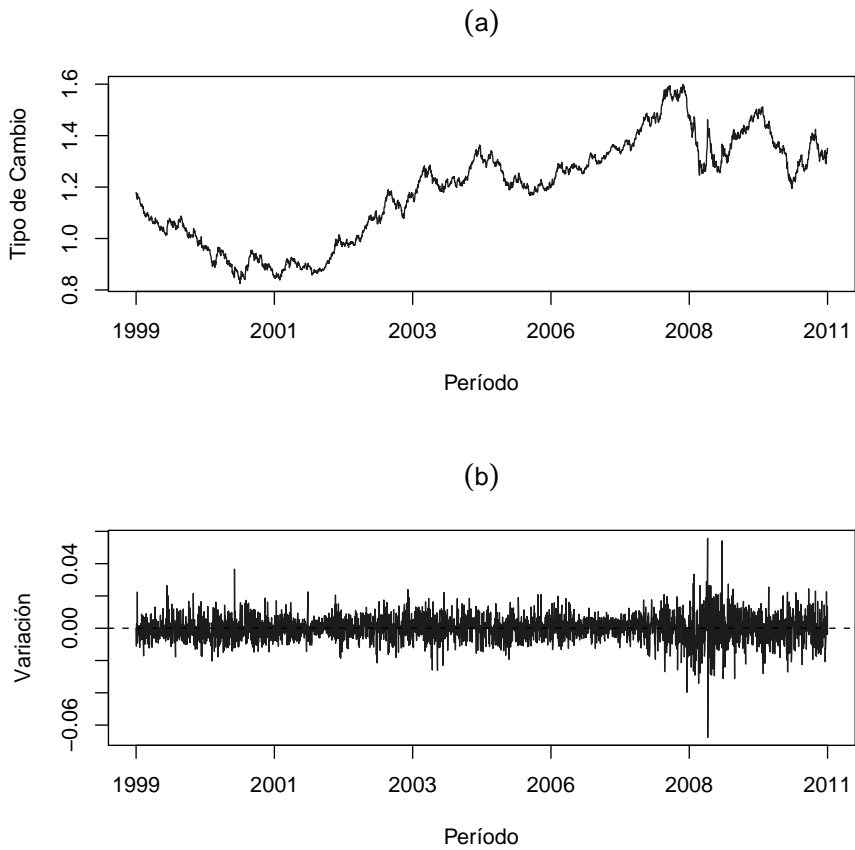


Figura 1.5: Series de los tipos de cambio entre el Dólar Estadounidense y el Euro, período comprendido entre el 4 de Enero de 1999 al 20 de Enero de 2011. (a) Tipos de Cambio, (b) Variaciones de los Tipos de Cambio.

2

Modelos de Series de Tiempo

Cuando se desea analizar una serie de tiempo en finanzas aplicando los diversos métodos estadísticos en la literatura, es de gran utilidad considerar las series observadas (x_1, x_2, \dots, x_T) , como una *realización* particular de un proceso estocástico (ver Apéndice A para detalles acerca de los procesos estocásticos). Dicha realización se suele denotar por $\{x_t\}_1^T$ mientras que, en general, el proceso estocástico será la familia de variables aleatorias $\{x_t\}_{-\infty}^{\infty}$ definidas en un espacio de probabilidad apropiado. Teniendo clara esta relación se establecerán los conceptos relativos a series de tiempo haciendo referencia de manera general a un proceso estocástico, heredando además su notación x_t , siempre que no haya posibilidad de confusión. Por otro lado, en cuanto al proceso estocástico se adoptará, sin pérdida de generalidad, el conjunto de índices habitual de las series de tiempo en el ámbito de las finanzas, $I = (1, T)$ en lugar del más general $I = (-\infty, \infty)$.

En base a estas consideraciones, los procesos estocásticos suelen ser descritos mediante su distribución conjunta de probabilidades, de manera que la relación que existe entre una realización y un proceso estocástico es análoga a la existente entre la muestra y población en el análisis estadístico clásico. Es claro que establecer con precisión por

completo la estructura de la distribución de probabilidad es una tarea bastante ambiciosa, por ello se suele concentrar la atención en los primeros y segundos momentos, es decir, las medias y/o covarianzas. Una suposición habitual es la de normalidad conjunta de la distribución, con lo cual tendríamos una caracterización de las propiedades del proceso estocástico. Sin embargo, tal suposición suele ser poco probable y en ocasiones inapropiada en el contexto financiero. Otra alternativa es suponer que el proceso es *lineal*, es decir, los valores actuales del proceso son generados a partir de una combinación lineal de los valores precedentes del propio proceso y con los valores actuales y precedentes de cualquier otro proceso que este relacionado con el mismo, de manera que se podría captar sus características principales. En cualquier caso, sin embargo, es una tarea difícil por no decir imposible inferir el total de los parámetros desconocidos involucrados, a partir de una realización del proceso (sólo T observaciones). Por ello, es necesario hacer una simplificación en las suposiciones con el objeto de reducir el número de dichos parámetros. Cabe destacar que lo anterior solo es posible bajo la suposición de que el proceso es ergódico, pero al igual que antes, es imposible verificar esta propiedad a partir de una simple realización, por ello, en lo que sigue se asumirá, para todas las series, la propiedad de ergodicidad.

Una manera de simplificar o relajar las suposiciones es estableciendo la *estacionariedad* del proceso. Para algunos autores tales como Box and Jenkins (1976) esta propiedad supone requerir al proceso un estado particular de "equilibrio estadístico". La base del análisis de series de tiempo es la estacionariedad, por ello es importante formalizar dicho concepto. Un proceso estocástico se denomina *estrictamente estacionario* si la distribución conjunta de $(x_{t_1}, \dots, x_{t_m})$ es idéntica a la distribución de $(x_{t_1+k}, \dots, x_{t_m+k})$ donde k es una variación arbitraria en el eje del tiempo y (t_1, \dots, t_m) es una colección de m valores en el eje del tiempo. En otras palabras, la estacionariedad estricta implica invariancia de la distribución de probabilidad ante valores igualmente separados. Sin embargo, en la práctica resulta una condición bastante fuerte y difícil de verificar empíricamente, por lo que, con frecuencia, suele considerarse una versión de estacionariedad más débil. Dicha versión conocida con el nombre de *estacionariedad de segundo orden o débil* se define formalmente de la siguiente manera:

Definición 2.0.1. Un proceso estocástico es estacionario si $\mathbb{E}(x_t^2) < \infty$ para cada t , y

- $\mathbb{E}(x_t) = \mu_x$, (constante), independiente de t ,
- $\text{Cov}(x_t, x_{t+k}) = \gamma_k$, independiente de t para cada k .

La estacionariedad débil implica que en un gráfico de la serie de datos, estos deberían fluctuar respecto de la variable índice, en este caso el tiempo, con variaciones constantes alrededor de un valor fijo.

En resumen, la suposición de estacionariedad estricta implica que la media y la varianza de un proceso son constantes (finitos), entonces el proceso es también débilmente estacionario. El recíproco no es cierto, salvo en el caso en que se considere normalidad del proceso, entonces ambos conceptos son equivalentes. La covarianza entre x_t y x_{t-k}

$$\gamma_k = \text{Cov}(x_t, x_{t-k}) = \mathbb{E}\{(x_t - \mu_x)(x_{t-k} - \mu_x)\},$$

se conoce como la *autocovarianza* del retardo k -ésimo, y es tal que

$$(i.) \gamma_0 = \text{Var}(x_t),$$

$$(ii.) \gamma_{-k} = \gamma_k.$$

En cuanto a la correlación entre x_t y x_{t-k} , cuando la dependencia entre dichos valores es de interés, entonces el concepto de correlación se generaliza al de autocorrelación. El coeficiente de correlación entre x_t y x_{t-k} es conocido con el nombre de *autocorrelación* del retardo k de x_t y se denota comúnmente por ρ_k . Bajo la suposición de estacionariedad débil, es una función que depende sólo del retardo k y viene expresada por:

$$\rho_k = \frac{\text{Cov}(x_t, x_{t-k})}{\{\text{Var}(x_t)\text{Var}(x_{t-k})\}^{1/2}} = \frac{\text{Cov}(x_t, x_{t-k})}{\text{Var}(x_t)} = \frac{\gamma_k}{\gamma_0}, \quad (2.1)$$

con la propiedad $\text{Var}(x_t) = \text{Var}(x_{t-k})$ bajo la suposición de estacionariedad débil. De la definición es claro que,

$$(i.) \rho_0 = 1,$$

$$(ii.) \rho_k = \rho_{-k},$$

$$(iii.) |\rho_k| \leq 1.$$

En este orden de ideas, una serie *no es correlacionada serialmente* si y solamente si $\rho_k = 0$ para todo $k > 0$. El término de correlación serial está relacionado a la correlación de una variable con ella misma en intervalos de tiempo sucesivos. Los analistas técnicos utilizan la correlación serial para determinar qué tan bien los precios en el pasado de un activo pueden predecir los precios futuros.

Cuando las autocorrelaciones ρ_k son consideradas como una función del retardo k entonces se definen como *función de autocorrelación (ACF)*. La ACF juega un papel muy importante en la caracterización de la dependencia entre las observaciones, junto con el proceso de medias $\mu_x = \mathbb{E}(x_t)$ y varianzas $\sigma_x^2 = \gamma_0 = \mathbb{V}ar(x_t)$, los procesos estocásticos estacionarios describen la evolución de x_t .

Dada una serie $\{x_t\}_{t=1}^T$, con $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ la media de la muestra. Entonces la *autocorrelación de la muestra para el retardo k* de x_t se define por

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \quad 0 \leq k < T - 1. \quad (2.2)$$

Bajo algunas condiciones generales, $\hat{\rho}_k$ es un estimador consistente de ρ_k . Por otro lado, si $\{x_t\}$ es una sucesión independiente e idénticamente distribuida que satisface $\mathbb{E}(x_t^2) < \infty$, entonces $\hat{\rho}_k$ es asintóticamente normal con media cero y varianza $1/T$ para un entero k , positivo fijo. En términos generales, si x_t es una serie de tiempo estacionaria débilmente y que además satisface

$$x_t = \mu + \sum_{i=0}^q \eta_i \varepsilon_{t-i}, \quad (2.3)$$

para $\eta_0 = 1$, q un entero no negativo, y $\{\varepsilon_t\}$ es una sucesión de media cero y varianza finita, independientes y con distribución normal, entonces $\hat{\rho}_k$ es asintóticamente normal con media cero y varianza $(1 + 2 \sum_{i=1}^q \rho_i^2)/T$ para $k > q$. Para un análisis detallado de estos tópicos véase Box and Jenkins (1976). En las aplicaciones, suele ser común desear probar de manera conjunta si las autocorrelaciones de una serie x_t son cero. Ljung and Box (1978) proponen un test basado en una modificación del estadístico de *Portmanteau* propuesto por Box and Pierce (1970), para contrastar

$$H_0 : \rho_1 = \dots = \rho_m = 0,$$

contra la alternativa

$$H_a : \rho_i \neq 0, \quad \text{para algún } i \in \{1, \dots, m\}.$$

El estadístico propuesto viene dado por

$$Q(m) = T(T + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T - k}, \quad (2.4)$$

el cual es una modificación del estadístico de *Portmanteau*

$$Q^*(m) = T \sum_{k=1}^m \hat{\rho}_k^2,$$

con la finalidad de mejorar su potencia. Nótese que, El estadístico $Q^*(m)$ es asintóticamente una variable aleatoria chi-cuadrado con m grados de libertad, bajo la suposición de que $\{x_t\}$ es una sucesión independiente e idénticamente distribuida, con ciertas condiciones sobre sus momentos. Entonces, para el estadístico $Q(m)$, el test rechaza H_0 si $Q(m) > \chi_{\alpha}^2$, donde χ_{α}^2 es el $100(1 - \alpha)$ percentil de una distribución chi-cuadrado con m grados de libertad. El valor m se puede seleccionar tomando $m \approx \ln(T)$, siendo T , en este caso, el número de observaciones. Como veremos más adelante, los estadísticos $\{\hat{\rho}_k\}_{k \geq 1}$ definidos en (2.2) caracterizan los modelos lineales de series de tiempo. En la figura (2.1) se muestran las funciones de autocorrelación muestral de las series del precio de los activos, rentabilidad simple y continua, del Banco Santander y del índice del IBEX35. Las ACF del precio del activo, en cada caso indica que las $\hat{\rho}_k$, para $k \geq 1$ son distintas de cero, con un decaimiento lento propio de las series de tiempo no estacionarias. En cuanto a las series de rentabilidades simple y continuas, se puede observar que las ACF son bastante similares entre ellas, este comportamiento se refleja en ambas series.

La tabla (2.2) muestra el estadístico de Ljung-Box aplicado a las series mencionadas, para $m = 5$, obtenidos aplicando el software estadístico **R**. En este caso, el software proporciona el p -valor de $Q(m)$, y a partir de la regla de decisión: H_0 es rechazada si p -valor es menor o igual que α , el nivel de significación. Teniendo en cuenta esto último, es claro que, en el caso de las series de los precios se rechaza la hipótesis del test descrita anteriormente, con lo cual el test confirma que existe correlación significativa par ambas series de precios. En lo que respecta a las

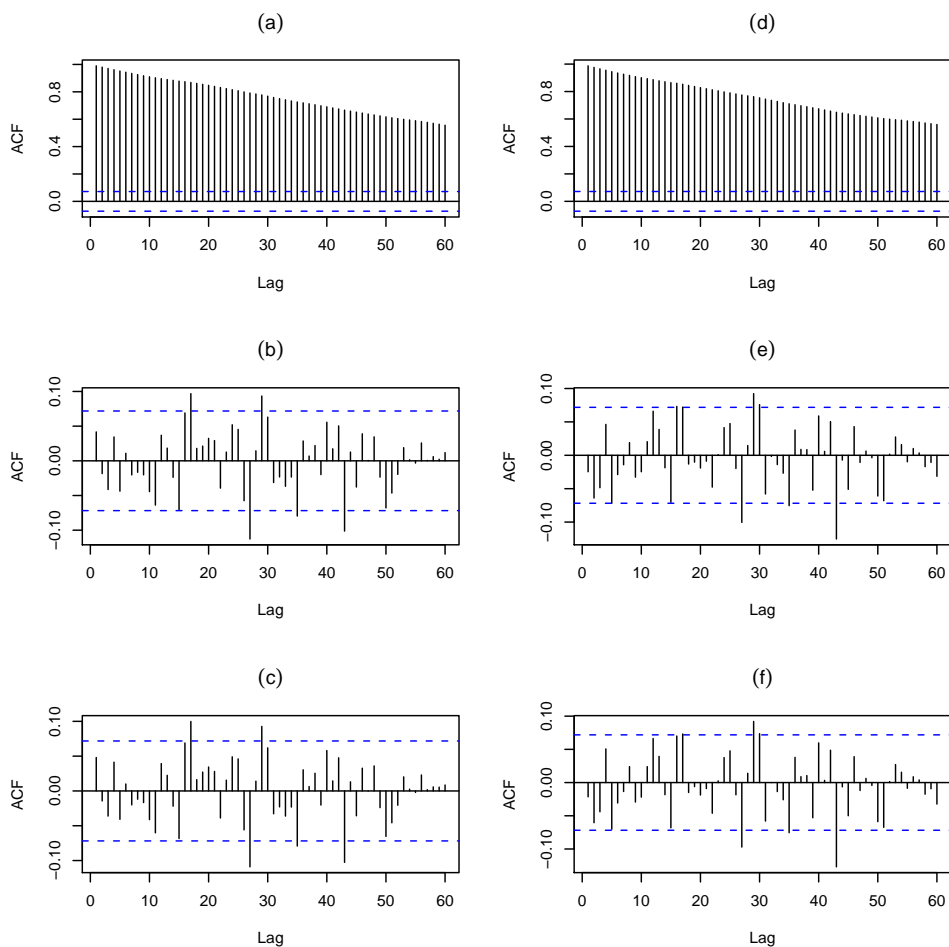


Figura 2.1: Función de Autocorrelación Muestral de las series del Banco Santander e índice del *IBEX35*. Series del Banco Santander : (a) precio del activo, (b) rentabilidad simple, (c) rentabilidad continua. Series el índice *IBEX35*: (d) precio del activo, (e) rentabilidad simple, (f) rentabilidad continua.

Serie	Banco Santander		
	$Q(m)$	m	p -valor
Precio	5516,716	8	$< 2,2e - 16$
Rentabilidad Simple	5,768	8	0,673
Rentabilidad Continua	5,876	8	0,661

Cuadro 2.1: Estadístico de Ljung-Box para la serie del Banco Santander.

Serie	IBEX35		
	$Q(m)$	m	p -valor
Precio	5447,224	8	$< 2,2e - 16$
Rentabilidad Simple	11,834	8	0,159
Rentabilidad Continua	11,452	8	0,177

Cuadro 2.2: Estadístico de Ljung-Box para el índice del Mercado Bursátil Español.

series de rentabilidades, el comportamiento es distinto, en este caso no correlación significativa para ninguna de las series. Estas gráficas son típicas de los paseos aleatorios.

En finanzas suele ser común asumir que las series de los retornos de los activos en estudio son débilmente estacionarias. Por ello, es importante establecer criterios para verificar, empíricamente, dicha suposición. Un resultado de gran utilidad relacionado a las series estacionarias es que cualquier función de una serie de tiempo estacionaria es también una serie de tiempo estacionaria. De manera que, si $\{x_t\}$ es estacionaria entonces $\{y_t\} = \{f(x_t)\}$ es estacionaria para cualquier función $f(\cdot)$. Otro resultado útil a destacar es el de ergodicidad de la serie. Una serie de tiempo $\{x_t\}$ es *ergódica* si los momentos relativos a la muestra convergen en probabilidad a los momentos de la población, es decir,

$$\bar{x} \xrightarrow{p} \mu, \quad \hat{\gamma}_k \xrightarrow{p} \gamma_k, \quad \hat{\rho}_k \xrightarrow{p} \rho_k.$$

En otras palabras, esto quiere decir que las medias calculadas a partir de observaciones pasadas no pueden diferir persistentemente de

la media temporal de acontecimientos futuros, así mismo una relación análoga para las covarianzas es aplicable. Cuando el proceso estocástico es estacionario pero no ergódico la incertidumbre suele caracterizar la dinámica del proceso.

A continuación se abordarán los aspectos más resaltantes del análisis de series de tiempo lineales.

2.1. Modelos Lineales

Los modelos lineales proporcionan un enfoque natural que permite analizar la dinámica de los procesos o series de tiempo en el contexto de las finanzas. En esta sección se discutirán la estructura de dependencia, autocorrelación, modelización y predicción de los modelos lineales teóricos. Para el valor de un determinado activo en un instante dado, el cual se denotará por x_t , los modelos lineales intentan capturar la relación lineal entre dicho valor x_t y la información disponible hasta el instante t . Dicha información puede contener los valores históricos del activo y un conjunto de variables que describen el entorno económico en el cual se determina el precio del activo en cuestión. Así pues, como ya se ha visto en la sección anterior, la correlación juega un papel muy importante en este aspecto, en particular las correlaciones entre la variable de interés y los valores pasados de la misma puesto que representan la herramienta básica en el estudio de los modelos lineales estacionarios. A continuación una descripción de los modelos econométricos básicos en la literatura de los modelos lineales teóricos.

2.1.1. El Proceso Ruido Blanco

Un proceso $\{\varepsilon_t\}$ se denomina *ruido blanco* (*white noise*) de media 0 y varianza σ^2 si satisface

$$\begin{aligned} \mathbb{E}(\varepsilon_t) &= 0, & \text{Var}(\varepsilon_t) &= \sigma_\varepsilon^2 < \infty, \\ \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) &= 0, \text{ para todo } k \neq 0. \end{aligned} \quad (2.5)$$

En particular, una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media 0 y varianza σ_ε^2 representa un caso especial de un proceso de ruido blanco. Este proceso se denota

por $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Si además ε_t se distribuye normalmente, la serie se denomina *ruido blanco gaussiano*.

Teorema de Descomposición de Wold

En el análisis de series de tiempo, un teorema fundamental es el conocido como el *Teorema de Descomposición de Wold* (Wold (1938)), afirma que todo proceso débilmente estacionario, no determinista se puede expresar como una combinación lineal de una sucesión de variables aleatorias no correlacionadas. Dicha representación viene dada por

$$\begin{aligned} x_t - \mu_x &= \varepsilon_t + \eta_1 \varepsilon_{t-1} + \eta_2 \varepsilon_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad \eta_0 = 1, \end{aligned} \quad (2.6)$$

donde $\{\varepsilon_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media cero, es decir, $\{\varepsilon_t\}$ es un proceso ruido blanco. Posteriormente veremos que ε_t denota la nueva información en el instante t del proceso, por lo que suele ser conocida con el nombre de *innovación* en el instante t . Los coeficientes η_i en (2.6) (posiblemente infinitos en número) son conocidos como los η -pesos. Claramente el proceso x_t puede ser escrito como

$$x_t = \mu_x + \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad \eta_0 = 1. \quad (2.7)$$

La estructura de la dinámica del proceso x_t esta gobernada por los coeficientes η_i . Si el número de pesos en (2.7) es infinito, entonces es necesario suponer la convergencia absoluta de los pesos ($\sum |\eta_i| < \infty$). Esta última condición es equivalente a suponer que x_t es estacionario o bien que x_t es débilmente estacionario, de esta manera se puede obtener su media y varianza fácilmente a partir de la independencia de las variables $\{\varepsilon_t\}$ como

- $\mathbb{E}(x_t) = \mu_x$,
- $\gamma_0 = \mathbb{V}ar(x_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \eta_i^2$, donde σ_ε^2 es la varianza del ε_t ,

- La autocovarianza del proceso x_t para el retardo k viene dada por

$$\gamma_k = \text{Cov}(x_t, x_{t-k}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \eta_i \eta_{i+k},$$

lo que implica que

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \eta_i \eta_{i+k}}{\sum_{i=0}^{\infty} \eta_i^2}, \quad k \geq 0.$$

Diversos modelos realistas son el resultado de selecciones particulares de los pesos en (2.7). Por ejemplo, si se considera $\mu_x = 0$, sin pérdida de generalidad, y se seleccionan, $\eta_i = \phi^i$, entonces la ecuación (2.7) puede expresarse como (después de operar algebraicamente)

$$x_t = \phi_1 x_{t-1} + \varepsilon_t. \quad (2.8)$$

Obteniéndose así el conocido proceso *autorregresivo de primer orden*. En el contexto de la econometría financiera, los modelos lineales de series de tiempo son modelos estadísticos y econométricos usados para describir la estructura de los η -pesos de x_t . Para una serie débilmente estacionaria, $\eta_i \rightarrow 0$ cuando $i \rightarrow \infty$ y, por lo tanto, ρ_k converge a cero cuando k aumenta. En el rendimiento de activos, esto se traduce como es de esperarse, la dependencia del rendimiento actual r_t respecto del valor pasado r_{t-k} disminuye para valores grandes de k .

2.1.2. Modelos Autorregresivos

Una parte de las series de tiempo económicas y financieras suelen ser caracterizadas por los modelos autorregresivos. Entre los principales ejemplos de las finanzas tenemos son valoración de precios y de dividendos, las tasas reales de cambio, tasas de interés y los diferenciales de tipos de interés (spreads).

En el contexto de los modelos lineales un proceso cuyo valor pasado inmediato, es decir, el primer retardo x_{t-1} es estadísticamente significativo en la predicción de x_t suele ser expresado como

$$x_t = \phi_0 + \phi_1 x_{t-1} + \varepsilon_t, \quad (2.9)$$

donde se asume que $\{\varepsilon_t\}$ es un proceso ruido blanco. En la sección anterior se introdujo este proceso, para $\phi_0 = \mu_x = 0$. El proceso (2.9) representa uno de los modelos más usados en la literatura del análisis de series de tiempo y es conocido con el nombre de proceso *autorregresivo de primer orden* denotado por $AR(1)$. Nótese que el modelo $AR(1)$ condicionado a los valores del pasado, es decir, x_{t-1} implica que:

$$\mathbb{E}(x_t|x_{t-1}) = \phi_0 + \phi_1 x_{t-1}, \quad \text{Var}(x_t|x_{t-1}) = \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2.$$

En otras palabras, dado el valor pasado x_{t-1} , el valor presente x_t es una variable centrada en $\phi_0 + \phi_1 x_{t-1}$ con desviación estándar σ_ε .

Podemos pensar entonces en el proceso $AR(1)$ como una suma de dos componentes, una de las cuales se puede determinar a partir de la información del pasado y la otra componente un término aleatorio con una estructura a precisar (por ejemplo, que verifica las propiedades de ruido blanco). Así pues, podemos reescribir el proceso autorregresivo de primer orden como

$$x_t = \mathbb{E}(x_t|x_{t-1}) + \varepsilon_t. \quad (2.10)$$

En general, existen situaciones en las cuales el simple valor x_{t-1} no resulta suficiente para determinar la esperanza condicional de r_t , de manera que se debe pensar en otros modelos más flexibles. Una generalización del modelo (2.10) viene dada por

$$x_t = \mathbb{E}(x_t|\mathcal{F}_{t-1}) + \varepsilon_t, \quad (2.11)$$

donde \mathcal{F}_{t-1} representa la información disponible hasta el instante $t-1$ y ε_t verifica las propiedades de ruido blanco. Se trata, por lo tanto, de buscar modelos adecuados para $\mathbb{E}(x_t|\mathcal{F}_{t-1})$. Nuevamente, los modelos autorregresivos suelen ser una solución simple para múltiples problemas en los que se asume que la esperanza condicional anterior es combinación lineal de los retardos de x_t . Así, el modelo autorregresivo de orden p , el cual generaliza el modelo (2.9), comúnmente denotado por $AR(p)$, se escribe como

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t, \quad (2.12)$$

donde p es un entero positivo y $\{\varepsilon_t\}$ un proceso ruido blanco. Este modelo supone que los p valores pasados x_{t-i} ($i = 1, \dots, p$) determinan de

manera conjunta la esperanza condicional de x_t condicionados por los datos del pasado. El modelo $AR(p)$ puede ser visto como un modelo de regresión lineal múltiple en los que los p retardos representan las variables explicativas.

A continuación se presentan algunas características relevantes del proceso $AR(1)$, y que se generalizan para los modelos $AR(p)$. Para comenzar, se estudiará la condición necesaria y suficiente para que el modelo $AR(1)$ dado por la ecuación (2.9) sea débilmente estacionario. Asumiendo que x_t es débilmente estacionario, se tiene

$$i. \mathbb{E}(x_t) = \mu,$$

$$ii. \text{Var}(x_t) = \gamma_0,$$

$$iii. \text{Cov}(x_t, x_{t-k}) = \gamma_k,$$

donde μ y γ_0 son constantes y γ_k es una función que depende solo del retardo k y no del instante t . Expresiones explícitas para la esperanza, varianza y función de autocovarianza se presentan a continuación:

$$i. \mathbb{E}(x_t) = \mu = \frac{\phi_0}{1 - \phi_1},$$

$$ii. \text{Var}(x_t) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2},$$

$$iii. \text{Cov}(x_t, x_{t-k}) = \gamma_k = \begin{cases} \phi_1 \gamma_1 + \sigma_\varepsilon^2 & \text{si } k = 0, \\ \phi_1 \gamma_{k-1} & \text{para } k > 0. \end{cases}$$

De manera que, una condición necesaria y suficiente para que un modelo $AR(1)$ con ecuación (2.9) sea débilmente estacionario es que $|\phi_1| < 1$. Por otro lado, tomando en cuenta (iii.) la función de autocorrelación ACF de x_t satisface

$$\rho_k = \phi_1 \rho_{k-1}, \quad \text{para } k \geq 0.$$

De esta última ecuación y de la definición de la ACF se puede deducir que, $\rho_k = \phi_1^k$. Esto indica que la ACF de un modelo $AR(1)$ débilmente estacionario tiene un decaimiento exponencial con tasa igual a ϕ_1 . Si $\phi_1 > 0$ el decaimiento es constante. Si por el contrario, $\phi_1 < 0$ entonces el decaimiento es compuesto y se presenta de forma alternante con

tasa ϕ_1^2 . Para tener una idea de esto, consideremos los modelos autorregresivos de orden 1 simulados, para distintos valores de ϕ_1 (ver figura (2.2)). A continuación se muestran las funciones de autocovarianzas de

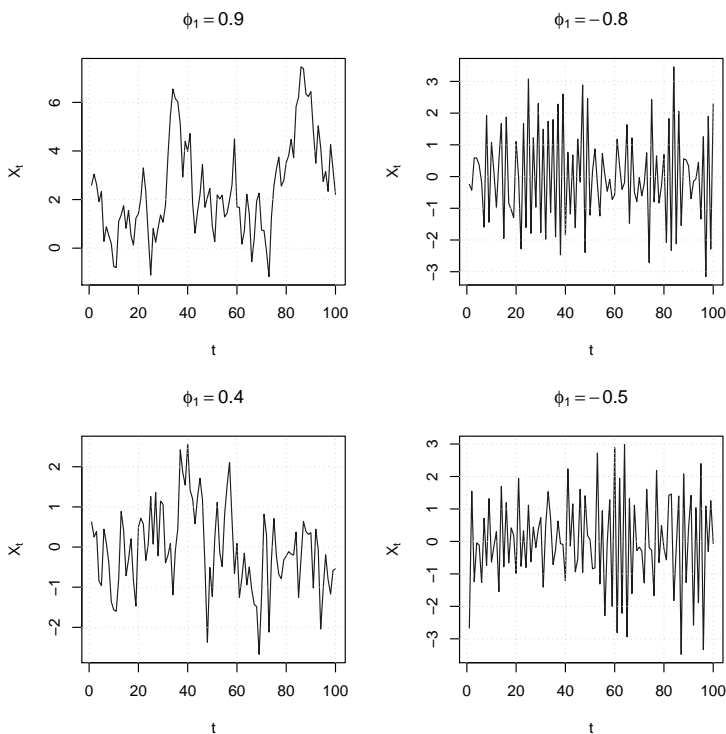


Figura 2.2: Simulaciones de procesos autorregresivos de primer orden, $X_t = \phi_1 X_{t-1} + \varepsilon_t$, para distintos valores de ϕ_1 .

las series simuladas. En la figura (2.3), se muestran las funciones de autocovarianza estimada del lado izquierdo y teórica del lado derecho para las series con parámetros positivos. Se puede observar el comportamiento de sus funciones de autocovarianza estimadas, las cuales tienden a comportarse según lo esperado, un decaimiento constante, salvo ciertas particularidades. Así mismo, en la figura (2.4) se muestran las funciones de autocovarianza estimada del lado izquierdo y teórica del lado derecho para las series con parámetros negativos. Se puede observar el comportamiento de sus funciones de autocovarianza estimadas,

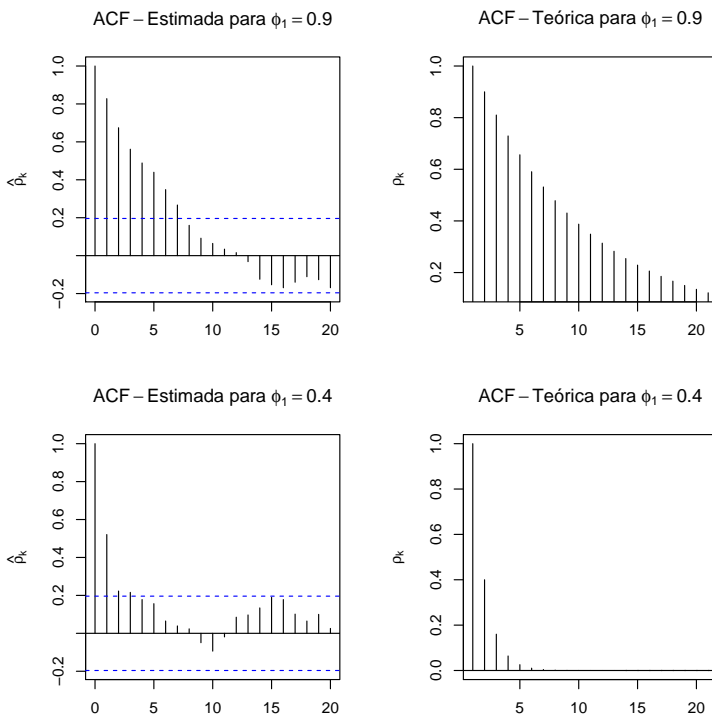


Figura 2.3: Funciones de Autocorrelación (ACF) de procesos autorregresivos de orden 1, $X_t = \phi_1 X_{t-1} + \varepsilon_t$ con parámetro positivo.

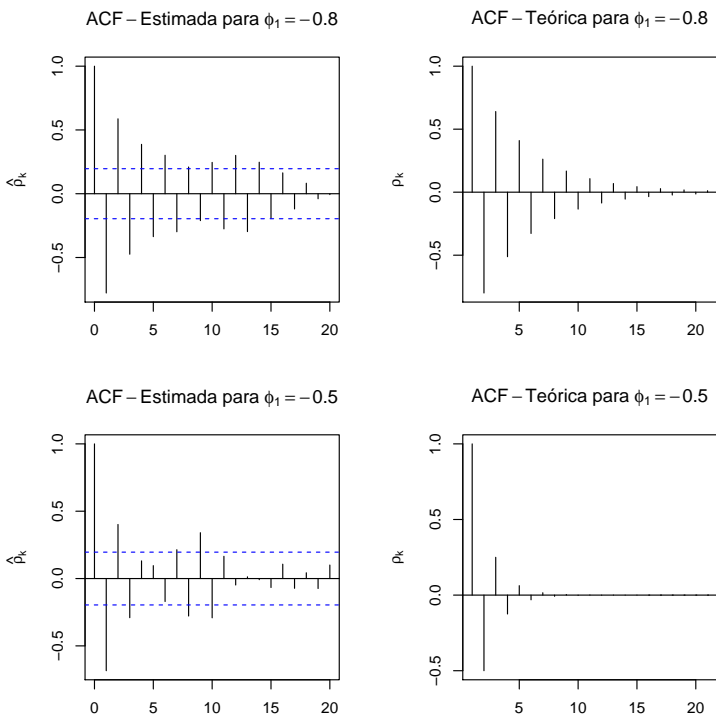


Figura 2.4: Funciones de Autocorrelación (ACF) de procesos autorregresivos de orden 1, $X_t = \phi_1 X_{t-1} + \varepsilon_t$ con parámetro negativo.

las cuales presentan el comportamiento alternante. A continuación, se introduce el *operador de retardos (backshift)* L con el propósito de establecer una notación general y simplificada. El operador L actúa de la siguiente manera

$$Lx_t \equiv x_{t-1}, \quad \text{en general} \quad L^p x_t \equiv x_{t-p}.$$

Esto permite escribir los modelos autorregresivos de una manera concisa. Por ejemplo, usando esta notación el modelo $AR(1)$ se puede escribir como

$$(1 - \phi_1 L)x_t = \varepsilon_t.$$

Aplicando el operador de retardo L , el modelo $AR(p)$ de la ecuación

(2.12) se puede reescribir como

$$(1 - \phi_1 L - \dots - \phi_p L^p)x_t = \varepsilon_t, \quad (2.13)$$

o bien en su forma reducida

$$\phi_p(L)x_t = \varepsilon_t, \quad (2.14)$$

donde

$$\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p,$$

el cual recibe el nombre de *polinomio característico*. Los resultados presentados para el proceso $AR(1)$ se pueden generalizar para el proceso $AR(p)$. Así pues, tenemos que

$$\mathbb{E}(x_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p},$$

siempre que el denominador no se anule. De manera análoga al proceso $AR(1)$, se puede establecer la condición necesaria y suficiente para que el proceso $AR(p)$ sea estacionario. Un proceso autorregresivo de orden p es de este tipo, si las *raíces características* (soluciones) en módulo (valor absoluto) del polinomio $\phi_p(L)$, calculadas a partir de la *ecuación característica* $\phi_p(L) = 0$ son mayores que 1; es decir, están fuera del círculo unitario. Para un proceso estacionario $AR(p)$, la *ACF* satisface la ecuación en diferencias

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)\rho_k = 0, \quad k > 0.$$

Una representación gráfica de la *ACF* de un modelo $AR(p)$ estacionario debería mostrar un efecto que tiende a reducir la amplitud de las oscilaciones (asemejando a un oscilador armónico), en otras palabras, la *ACF* mostraría una mezcla de amortiguación de senos y cosenos con un decaimiento exponencial que dependerá de la naturaleza de las raíces características.

Identificación del orden de un Modelo $AR(p)$

En la práctica, el orden p de un proceso AR es desconocido, por lo que debe ser determinado de manera empírica. Esto se conoce como el *orden*

de determinación del modelo AR , y ha sido estudiado de manera exhaustiva en la literatura del análisis de series de tiempo. Para determinar el valor de p se disponen de dos enfoques generales. El primer enfoque, esta basado en la *función de autocorrelación parcial (PACF)*. Una manera simple para introducir la *PACF* de una serie de tiempo estacionaria, es considerando los modelos autorregresivos con ordenes consecutivos:

$$\begin{aligned}x_t &= \phi_{0,1} + \phi_{1,1}x_{t-1} + \epsilon_{1t}, \\x_t &= \phi_{0,2} + \phi_{1,2}x_{t-1} + \phi_{2,2}x_{t-2} + \epsilon_{2t}, \\x_t &= \phi_{0,3} + \phi_{1,3}x_{t-1} + \phi_{2,3}x_{t-2} + \phi_{3,3}x_{t-3} + \epsilon_{3t}, \\&\vdots \quad \quad \quad \vdots\end{aligned}$$

con $\phi_{0,j}$, $\phi_{i,j}$ y $\{e_{jt}\}$ el término constante, el coeficiente de x_{t-i} y el término de error de un modelo $AR(j)$, respectivamente. Claramente, cada uno de estos modelos están expresados como un modelo de regresión lineal múltiple, lo cuales pueden ser estimados mediante el método de los mínimos cuadrados. El valor estimado $\hat{\phi}_{k,k}$, con $k \leq 1$, de la ecuación k es conocido con el nombre de la *función de autocorrelación parcial de la muestra PACF* para el retardo k -ésimo de x_t . De la definición, se puede extraer que el valor estimado $\hat{\phi}_{k,k}$ muestra la contribución de x_{t-k} para el modelo $AR(k-1)$. Por lo tanto, para un modelo $AR(p)$, la función de autocorrelación parcial de la muestra *PACF* para el retardo p debería ser distinto de cero, mientras que $\hat{\phi}_{j,j}$ debería ser cercano a cero para todo $j > p$. A partir de esta propiedad, se puede determinar el orden p .

Retomando los modelos simulados en la figura (2.2), sus respectivas funciones de autocorrelación parcial (*PACF*) se presentan en la figura (2.5). En cada caso se observa que la *PACF*, tiene su valor máximo para el primer retardo, con lo cual se evidencia lo expresado anteriormente acerca de la contribución de dicho retardo.

El segundo enfoque, esta basado en el uso de algún criterio de información. Existen diversos criterios, entre ellos el más conocido es el *Criterio de Información de Akaike (AIC)* (Akaike (1973)) definido como

$$AIC = \frac{-2}{T} \log(\text{verosimilitud}) + \frac{2}{T}(n^\circ \text{deparámetros}), \quad (2.15)$$

donde la verosimilitud es evaluada en los valores estimados por el método de máxima verosimilitud. El primer término en (2.15) mide la

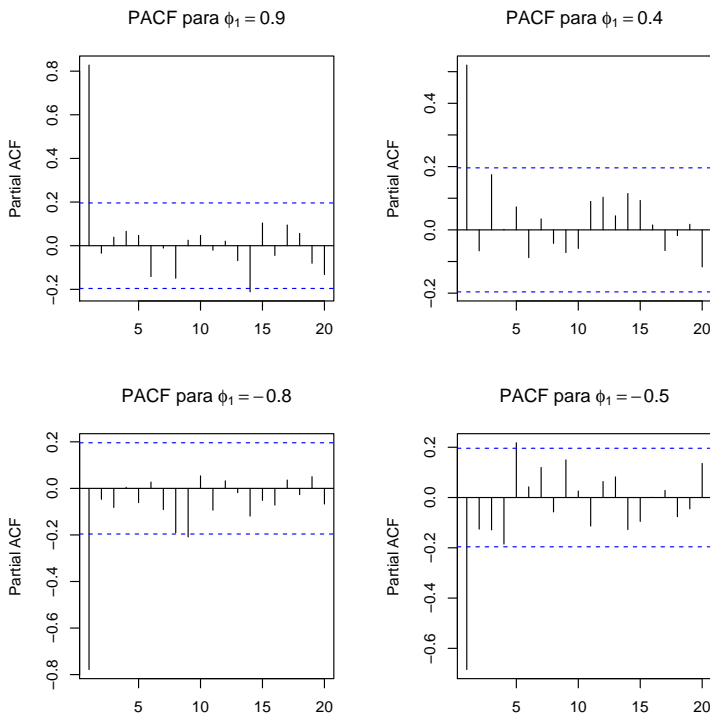


Figura 2.5: Funciones de Autocorrelación parcial (*PACF*) de procesos autorregresivos de orden 1, $X_t = \phi_1 X_{t-1} + \varepsilon_t$, para distintos valores de ϕ_1 .

p	0	1	2	3	4	5
AIC	0	1.88	3.73	5.29	7.07	
$PACF$	0.83	-0.04	0.039	0.07	0.05	

Cuadro 2.3: AIC de la serie simulada $X_t = 0,9X_{t-1} + \varepsilon_t$.

p	1	2	3	4	5
AIC	0	1.78	3.11	5.10	6.73
$PACF$	-0.778	-0.047	-0.082	0.005	-0.061

Cuadro 2.4: AIC de la serie simulada $X_t = -0,8X_{t-1} + \varepsilon_t$.

bondad del ajuste del modelo $AR(k)$ para los datos, mientras que el segundo término se le conoce con el nombre de *función de penalización* del criterio puesto que este penaliza un modelo candidato por el número de parámetros usado. Dependiendo del criterio usado, esta función puede variar. Para seleccionar un modelo AR , en la práctica, se calcula el $AIC(k)$ para $k = 0, \dots, p$, donde p es un entero positivo preespecificado, y entonces se selecciona el orden k como aquel con menor valor de AIC . Como se observa en los cuadro (2.3) y (2.4), los menores valores de AIC son para el retardo 1, en ambos casos. Por tanto, el orden de las series es 1, como era de esperarse.

Estimación de los parámetros de un Modelo $AR(p)$

El método que se usa con frecuencia para estimar los parámetros de un modelo $AR(p)$ es el método de mínimos cuadrados. Teniendo en cuenta el condicionamiento sobre las primeras p observaciones, se tiene la ecuación

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t, \quad t = p + 1, \dots, T,$$

la cual se presenta en la forma de una regresión lineal múltiple y se puede estimar por mínimos cuadrados. El valor estimado de ϕ_i se denota por $\hat{\phi}_i$. El *modelo ajustado* es

$$\hat{x}_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \dots + \hat{\phi}_p x_{t-p},$$

y los residuos asociados al modelo estimado son

$$\hat{\varepsilon}_t = x_t - \hat{x}_t.$$

La serie $\{\hat{\varepsilon}_t\}$ se le conoce con el nombre de *serie de los residuos*, de donde se obtiene que

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=p+1}^T \hat{\varepsilon}_t^2}{T - 2p - 1}.$$

El valor estimado de σ_ε^2 puede cambiar si el método usado es el de la verosimilitud condicional. En este caso, el valor estimado es

$$\tilde{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^2 \frac{(T - 2p - 1)}{T - p}.$$

Verificación del Modelo

Una vez realizado el ajuste o estimación de un modelo es necesario examinar de manera cuidadosa si el modelo considerado es el adecuado. Si es así, entonces la serie de los residuos debería comportarse como un ruido blanco. Una alternativa para verificar esto es usar la *ACF* de los residuos para determinar cuán próximos están los $\hat{\varepsilon}_t$ de un ruido blanco. El estadístico de Ljung y Box (1978) es otra alternativa usada con frecuencia en aplicaciones en las finanzas, este viene dado por

$$Q(m) = T(T + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T - k}. \quad (2.16)$$

La regla de decisión es rechazar la hipótesis nula $H_0 : \rho_1 = \dots = \rho_m = 0$ si $Q(m) > \chi_{\alpha}^2$, donde χ_{α}^2 denota el percentil $100(1 - \alpha)$ -ésimo de una distribución Chi cuadrado con $m - l$ grados de libertad, donde l denota el número de coeficientes usados en el modelo autorregresivo. En cuanto a la selección del valor m , el cual puede afectar la eficacia del test, estudios de simulación sugieren seleccionar $m \approx \ln(T)$. Si el modelo ajustado no resulta ser el adecuado, entonces es necesario redefinirlo. Se puede verificar si algunos de los valores estimados de los coeficientes del modelo *AR* no son significativamente distintos de cero, entonces el modelo debe simplificarse tratando de eliminar los parámetros que no son significativos. Si la *ACF* de los residuos muestra correlación serial (es decir, cuando los términos de error de diferentes períodos de tiempo se correlacionan) entonces el modelo debe ampliarse tomando en cuenta dichas correlaciones.

Bondad del Ajuste

Un estadístico que se usa con frecuencia para medir la *bondad de ajuste* de un modelo estacionario $AR(p)$ con T observaciones $\{x_t | t = 1, \dots, T\}$, es el estadístico *R-cuadrado* (R^2) definido como

$$R^2 = 1 - \frac{\sum_{t=p+1}^T \hat{\varepsilon}_t^2}{\sum_{t=p+1}^T (x_t - \bar{x})^2},$$

donde $\bar{x} = (\sum_{t=p+1}^T x_t) / (T - p)$. Generalmente un valor grande de R^2 indica que el modelo ofrece un buen ajuste a los datos. Para un conjunto de datos dado, es bien sabido que R^2 es una función creciente del número de parámetros usados, pues su valor aumenta cuando se incluyen nuevas variables en el modelo, incluso cuando estas son pocas significativas o tienen poca correlación con la variable dependiente. Para superar esta debilidad, se propone un R^2 -ajustado el cual se define como

$$Adj - R^2 = 1 - \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_x^2},$$

donde $\hat{\sigma}_x^2$ es la varianza de la muestra de x_t . Este nuevo estadístico toma en cuenta el número de parámetros usados en el modelo ajustado.

Predicción

Dentro del análisis de series de tiempo la predicción juega un papel muy importante. Suponga que se tiene un modelo $AR(p)$ y que se está en el instante h , y además se está interesado en la predicción de x_{h+k} , donde $k \geq 1$. El instante h es llamado el *origen de predicción* y el entero positivo k es el *horizonte de predicción*. Sea $\hat{x}_h(k)$ la predicción de x_{h+k} , usando el error cuadrático mínimo de la función de pérdida y \mathcal{F}_h la información disponible en el origen de predicción h . Entonces, la predicción $\hat{x}_h(k)$ se selecciona de manera tal que

$$\mathbb{E}\{(x_{h+k} - \hat{x}_h(k))^2 | \mathcal{F}_h\} \leq \underset{g}{\text{mín}} \mathbb{E}\{(x_{h+k} - g)^2 | \mathcal{F}_h\},$$

donde g una función de \mathcal{F}_h . Se denotará por $\hat{x}_h(k)$ la predicción en el paso k de x_t con origen de predicción h . Así pues, tenemos que

$$x_{h+k} = \phi_0 + \phi_1 x_{h+k-1} + \dots + \phi_p x_{h+k-p} + \varepsilon_{h+k}.$$

La predicción en el paso k basada en el mínimo error cuadrático de la función de pérdida es la esperanza condicional de x_{h+k} condicionada a \mathcal{F}_h , la cual se puede obtener a partir de

$$\hat{x}_h(k) = \mathbb{E}(x_{h+k} | \mathcal{F}_h) = \phi_0 + \sum_{i=1}^p \phi_i \hat{x}_h(k-i),$$

donde se entiende que $\hat{x}_h(i) = x_{h+i}$, si $i \leq 0$. El error de predicción asociado es

$$e_h(k) = x_{h+k} - \hat{x}_h(k).$$

Se puede probar que para un modelo $AR(p)$ estacionario, la predicción $\hat{x}_h(k)$ converge a $\mathbb{E}(x_t)$ cuando $k \rightarrow \infty$, lo cual significa que el punto de predicción a largo plazo se aproxima a su media incondicional. Esta propiedad es conocida en la literatura financiera como la *reversión a la media*. Para un modelo $AR(1)$, la velocidad de reversión a la media se mide por la vida media, la cual se define como $k = \ln(0,5/|\phi_1|)$. La varianza del error de pronóstico entonces, se aproxima a la varianza incondicional de x_t .

2.1.3. Modelos de Medias Móviles

Un modelo alternativo y de gran utilidad en la modelización de series en finanzas es el modelo de medias móviles (*moving-average*) el cual denotaremos por MA por sus siglas en inglés. Tomando en cuenta el modelo dado por la ecuación (2.3) y seleccionando $\eta_0 = 1$, $\eta_1 = -\theta_1$, y $\eta_j = 0$ para todo $j \geq 2$, se obtiene el modelo

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad \mu\text{-constante} \quad (2.17)$$

o bien, aplicando el operador de retardos

$$x_t = \mu + (1 - \theta_1 L)\varepsilon_t, \quad (2.18)$$

conocido con el nombre de *proceso de media-móvil de primer orden* y denotado por $MA(1)$. En el modelo $MA(1)$ presentado en las ecuaciones (2.17) y (2.18), se puede considerar sin pérdida de generalidad $\mu = 0$, obteniéndose así la clásica representación del modelo $MA(1)$

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}. \quad (2.19)$$

En general, un proceso de *medias móviles de orden q* , denotado por $MA(q)$, se puede escribir como

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.20)$$

o de forma equivalente, usando el operador de retardo como

$$x_t = \mu + (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t = \mu + \theta_q(L), \quad (2.21)$$

donde

$$\theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q.$$

De manera análoga a los modelos autorregresivos, primero se establecerán las propiedades para el modelo más simple, $MA(1)$ y posteriormente serán generalizadas al modelo general $MA(q)$.

Los modelos MA son débilmente estacionarios puesto que se definen como una combinación lineal de una sucesión de ruido blanco donde los primeros dos momentos son invariantes respecto del tiempo. Si se considera el proceso $MA(1)$ de la ecuación (2.17), su valor esperado es $\mathbb{E}(x_t) = \mu$, la cual es invariante respecto del tiempo. Así mismo al tomar su varianza se obtiene

$$\text{Var}(x_t) = (1 + \theta_1^2) \sigma_\varepsilon^2,$$

en donde se tiene en cuenta el hecho de que ε_t y ε_{t-1} son no correlacionados. Como se observa, la varianza es también invariante respecto del tiempo. Para el caso general, es decir, un proceso $MA(q)$, se obtiene que:

- El término constante μ es la media del proceso, es decir, se satisface $\mathbb{E}(x_t) = \mu$,
- La varianza de un proceso $MA(q)$ es

$$\text{Var}(x_t) = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2.$$

Sin pérdida de generalidad y por simplicidad se asumirá que $\mu = 0$. Así pues, para un proceso $MA(1)$ la covarianza viene dada por

$$\gamma_1 = -\theta_1 \sigma_\varepsilon^2, \quad \text{y} \quad \gamma_k = 0, \quad \text{para } k > 1.$$

En general, para un modelo $MA(q)$ la covarianza se puede expresar mediante

$$\gamma_k = \begin{cases} (1 + \theta_1^2 + \dots + \theta_q^2)\sigma_\varepsilon^2 & \text{si } k = 0, \\ (-\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_{k+p}\theta_p + \dots + \theta_q\theta_{q-k})\sigma_\varepsilon^2 & \text{para } k = 1, \dots, q, \\ 0 & \text{para } k > q. \end{cases}$$

Tomando en cuenta este resultado, la autocorrelación para el primer retardo de un proceso $MA(1)$ se expresa por

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \quad \rho_k = 0, \quad \text{para } k > 1.$$

En general, para un proceso $MA(q)$, la ACF hasta el orden q es distinta de cero, sin embargo $\rho_k = 0$ para valores de k superiores al orden del proceso, q . La función de autocorrelación viene dada por

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_{k+p}\theta_p + \dots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{para } k = 1, \dots, q, \\ 0 & \text{para } k > q. \end{cases}$$

Como consecuencia de esto, un proceso $MA(q)$ solo tiene relación lineal con sus primeros q retardos y por lo tanto es un *proceso con memoria finita*.

Invertibilidad

Otra de las propiedades deseables de una serie de tiempo es la *invertibilidad*. Un proceso x_t se dice invertible si es posible reconstruir el valor de la innovación ε_t únicamente a partir de observaciones pasadas de x_t . En este sentido, es inmediato comprobar que un proceso AR es siempre invertible. La condición necesaria y suficiente para que el proceso $MA(q)$ sea invertible es que las raíces de la ecuación característica $\theta_q(L) = 0$ estén fuera del círculo unitario. Esta propiedad implica que el proceso se puede escribir en términos de un proceso $AR(\infty)$,

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + \varepsilon_t,$$

donde la serie de los pesos, π_i convergen. Esto es, $\sum |\pi_i| < \infty$. Para ver esto, considere el modelo $MA(1)$

$$x_t = \theta_1(L)\varepsilon_t,$$

donde $\theta_1(L) = 1 - \theta_1 L$. Si se satisface la condición de invertibilidad, entonces la ecuación anterior es equivalente a

$$\theta_1^{-1}(L)x_t = \varepsilon_t,$$

de donde la expansión de $\theta_1^{-1}(L) = 1 - \theta_1 L$ produce

$$(1 + \theta_1 L + \theta_1^2 L^2 + \dots)x_t = \varepsilon_t.$$

Los pesos, $\pi_j = -\theta_1^j$, convergen siempre que $|\theta_1| < 1$, es decir, si el modelo es invertible. Esto último implica la suposición razonable, que el efecto de las observaciones pasadas decrecen con el tiempo. Para el caso general, es decir, un proceso $MA(q)$ los pesos en la representación $AR(\infty)$, esto es $\pi(L)x_t = \varepsilon_t$, vienen dados por $\theta^{-1}(L)$, los cuales se pueden obtener resolviendo la ecuación $\pi(L)\theta(L) = 1$ respecto de los coeficientes L^j . Por la condición de invertibilidad, las raíces de

$$(1 - \theta_1 L - \dots - \theta_q L^q) = (1 - h_1 L) \dots (1 - h_q L) = 0$$

deben satisfacer $|h_1| < 1, \dots, |h_q| < 1$.

Identificación del orden de un Modelo $MA(q)$

La función de autocorrelación ACF es una herramienta de gran utilidad en la identificación del orden de un modelo de medias móviles. La ACF de un $MA(q)$ se anula después del retardo q (es decir: $\rho_k = 0$, para $k > q$). Entonces, se tiene que el proceso puede ser modelizado mediante un proceso de medias móviles de orden q , $MA(q)$. Por otro lado, se puede demostrar que la $PACF$ (función de autocorrelación parcial) no se anula, aunque a partir de un retardo q decaerá de forma rápida. Las expresiones para la $PACF$ de un proceso de medias móviles suelen ser complicadas, sin embargo, en general, son combinaciones de decaimientos exponenciales (para raíces reales de $\theta(L)$) y/o sinusoidales (para raíces complejas de $\theta(L)$). Así pues, la $PACF$ de un proceso MA se comporta de manera análoga a como lo hace la ACF de un proceso AR . Esta dualidad importante entre los procesos AR y MA se puede resumir en:

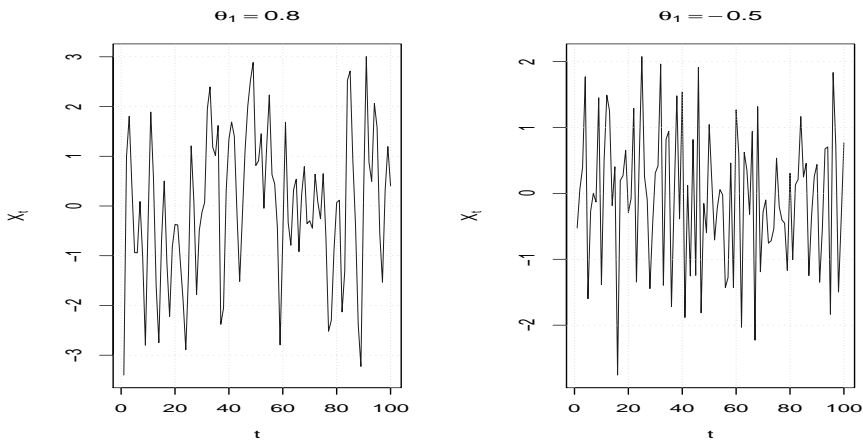


Figura 2.6: Simulaciones de procesos de medias móviles de orden 1, $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$, para distintos valores de θ_1 .

- Mientras la *ACF* de un proceso $AR(p)$ no se anula para ningún valor de p , la *PACF* se anula para retardos superiores a p .
- La *ACF* de un proceso $MA(q)$, por otro lado, se anula después del retardo q , mientras que la *PACF* no se anula para ningún valor de q .

Para ilustrar lo anterior, consideremos un modelo $MA(1)$,

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$

Entonces, para $\theta_1 = 0,8$ y $\theta_1 = -0,5$ considere las series simuladas dadas en la figura (2.6). Sus funciones de autocorrelación estimadas y teóricas se muestran en la figura (2.7). En ella se observa el comportamiento descrito en los párrafos anteriores. De manera que, la función de autocorrelación es una alternativa para la identificación de los modelos de media móviles. Así mismo, en la figura (2.8) se muestran las funciones de autocorrelación parcial estimadas para los procesos simulados.

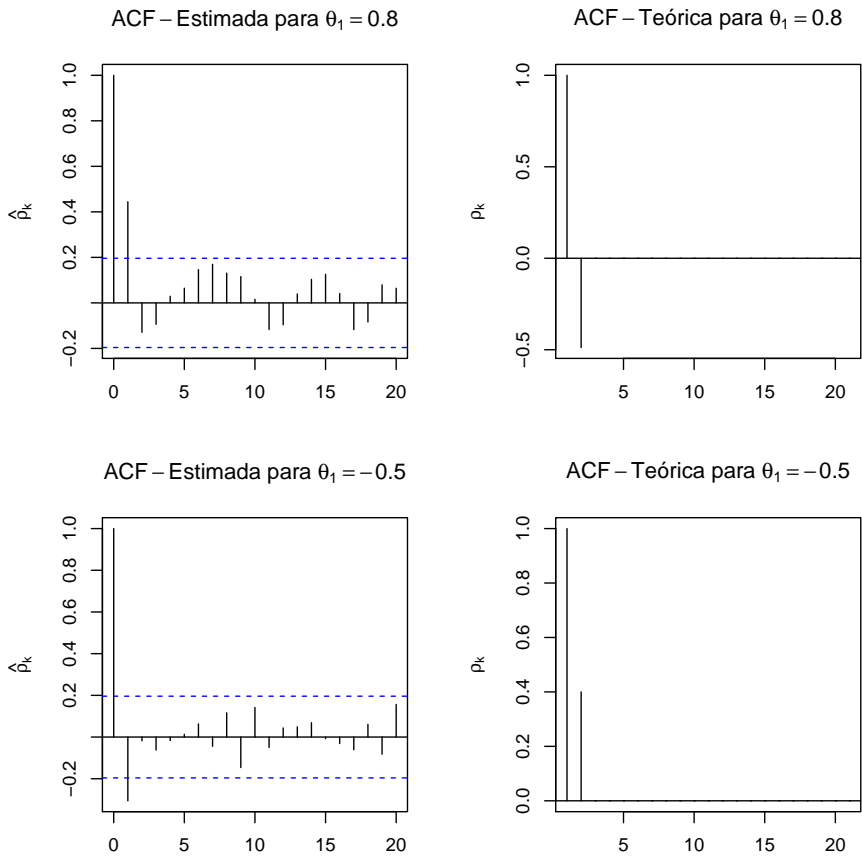


Figura 2.7: Funciones de autocorrelación estimada y teórica de los modelos de medias móviles simulados.

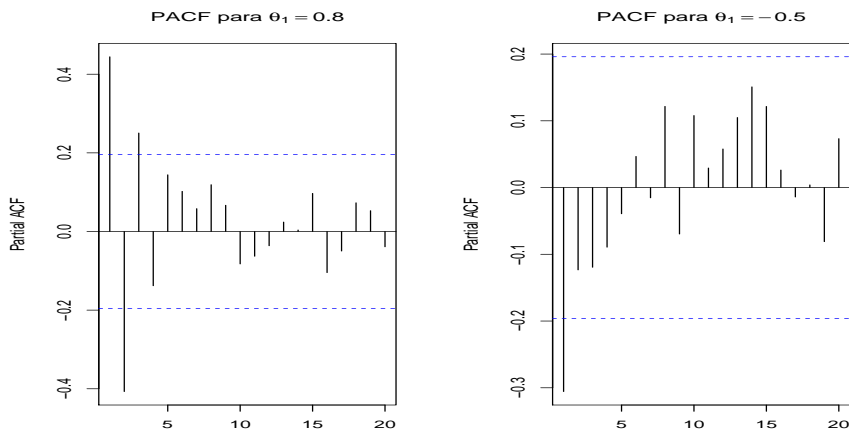


Figura 2.8: Funciones de autocorrelación parcial estimada de los modelos de medias móviles simulados.

Estimación de los parámetros de un Modelo MA

El método de máxima verosimilitud es el usado con más frecuencia en la estimación de un modelo MA. Existen dos maneras de abordar el problema de estimación en lo que respecta a la evaluación de la función de verosimilitud, cuando el método de máxima verosimilitud es considerado en la estimación. La primera de ellas, es asumir que los impulsos o innovaciones $\varepsilon_t = 0$ para $t \leq 0$. De esta manera, los valores necesarios para calcular la verosimilitud se obtienen de manera recursiva a partir del modelo, comenzando con

$$\varepsilon_1 = x_1 - \mu,$$

y

$$\varepsilon_2 = x_2 - \mu + \theta_1 \varepsilon_1.$$

Este enfoque hace referencia al *método de verosimilitud condicional*. Otra manera es considerar el *método de verosimilitud exacta*, es decir, se consideran los valores iniciales ε_t para $t \leq 0$ como parámetros adicionales del modelo y se estiman conjuntamente con el resto de parámetros. Existe preferencia respecto al método de verosimilitud exacta cuando el modelo MA está próximo a ser no invertible, es decir, cuando las raíces de la ecuación característica $\theta_q(L) = 0$ están cercanas al círculo unitario.

En cuanto a la verificación del modelo, una alternativa es usar el estadístico de Ljung y Box (1978), y se procede de manera análoga al caso de la verificación del modelo AR .

Predicción

Las predicciones de un modelo MA se pueden obtener de manera muy sencilla. Puesto que el modelo tiene memoria finita, las predicciones apuntan a la media de la serie rápidamente. Para ver esto, suponga que el origen de predicción es h , y se denota por \mathcal{F}_h a la información disponible en el instante h . Para la predicción a un paso de un proceso $MA(1)$

$$x_{h+1} = \mu + \varepsilon_{h+1} - \theta_1 \varepsilon_h.$$

Tomando la esperanza condicional, se obtiene

$$\begin{aligned}\hat{x}_h(1) &= \mathbb{E}(x_{h+1} | \mathcal{F}_h) = \mu - \theta_1 \varepsilon_h, \\ e_h(1) &= x_{h+1} - \hat{x}_h(1) = \varepsilon_{h+1}.\end{aligned}$$

La varianza del error de la predicción a un paso es $\mathbb{V}ar(e_h(1)) = \sigma_\varepsilon^2$. En la práctica, el valor de ε_h se puede obtener a partir de $\varepsilon_t = x_t - \mu + \theta_1 \varepsilon_{t-1}$ para $2 \leq t \leq h$, con $\varepsilon_0 = 0$, y por lo tanto $\varepsilon_1 = x_1 - \mu$. De manera general, $\hat{x}_h(k) = \mu$ para $k \geq 2$, en otras palabras la media incondicional del modelo. En general, para un proceso $MA(q)$ la predicción a varios pasos tiende a la media después de los primeros q -pasos; y las varianzas de los errores de predicción tienden a la varianza del modelo.

2.1.4. Modelos $ARMA$

Es claro que los modelos considerados en las secciones anteriores, AR y MA , pueden ser de gran utilidad en la modelización de determinadas series de datos en diversos campos de conocimientos. Sin embargo, en la práctica, concretamente en finanzas, puede ser necesario considerar modelos cuyos ordenes conllevan a complicaciones motivado por el gran número de parámetros que se requieren para describir de manera adecuada la estructura dinámica de los datos. Una manera de solventar este inconveniente es considerar un tipo de proceso que conjugue las propiedades de los modelos AR y MA en una expresión compacta y que permita la reducción de parámetros a ser considerados. Dicho

proceso es conocido con el nombre de *proceso autorregresivo de medias móviles* y denotado por *ARMA*. Los modelos *ARMA* se obtienen como combinación de modelos autorregresivos y de medias móviles. Así un proceso *ARMA*(p, q) vendrá definido por la siguiente ecuación,

$$\begin{aligned} x_t &= \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \\ &\quad - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ &= \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \end{aligned} \quad (2.22)$$

donde $\{\varepsilon_t\}$ es un proceso ruido blanco, y p y q son dos enteros no negativos. Entonces, si se aplica el operador de retardo, el proceso se puede expresar de forma equivalente como

$$\phi_p(L)x_t = \phi_0 + \theta_q(L)\varepsilon_t. \quad (2.23)$$

Los polinomios

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p = 1 - \sum_{i=1}^p \phi_i x_{t-i},$$

y

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q = 1 - \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

son los operadores polinomiales de los retardos correspondientes a los modelos *AR* y *MA*, respectivamente.

La condición necesaria y suficiente para que el proceso *ARMA*(p, q) sea estacionario e invertible es que las raíces de $\phi_p(L) = 0$ y $\theta_q(L) = 0$ estén fuera del círculo unitario, respectivamente.

Como antes, se establecen propiedades relativas a un proceso *ARMA*(1, 1) y luego se generalizan a procesos *ARMA*(p, q). Un proceso *ARMA*(1, 1) se expresa mediante la ecuación

$$x_t - \phi_0 - \phi_1 x_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad (2.24)$$

con $\{\varepsilon_t\}$ un ruido blanco. Las propiedades de dicho proceso son una generalización de las presentadas para los procesos *AR*; por supuesto,

con alguna modificación motivado al impacto que genera el componente del proceso MA . Para comenzar, considere la condición de estacionariedad. Se considera entonces la esperanza del proceso en la ecuación (2.24) y se obtiene

$$\mathbb{E}(x_t) = \mu = \frac{\phi_0}{1 - \phi_1},$$

sujeto a la estacionariedad débil del proceso. Como antes, sin pérdida de generalidad se puede asumir que $\phi_0 = 0$, entonces si el proceso x_t es débilmente estacionario,

$$\text{Var}(x_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_\varepsilon^2}{1 - \phi_1^2}.$$

Puesto que la varianza es positiva, es necesario establecer la condición, sobre el parámetro ϕ_1 , $\phi_1^2 < 1$. Así mismo, la autocovarianza para $k = 1$ es

$$\text{Cov}(x_t, x_{t-1}) = \gamma_1 - \phi_1\gamma_0 = -\theta_1\sigma_\varepsilon^2,$$

y para $k > 1$,

$$\text{Cov}(x_t, x_{t-k}) = \gamma_k = \phi_1\gamma_{k-1}.$$

Así pues, la ACF de un proceso estacionario $ARMA(1,1)$ viene dada por

$$\rho_1 = \phi_1 - \frac{\theta_1\sigma_\varepsilon^2}{\gamma_0}, \quad \rho_k = \phi_1\rho_{k-1}, \quad k > 1.$$

De esta manera, la función de autocorrelación ACF de un proceso $ARMA(1,1)$ se comporta de manera similar a la de un proceso $AR(1)$, excepto que el decaimiento exponencial comienza en el segundo retardo. Un comportamiento similar ocurre entre la función de autocorrelación parcial $PACF$ de un proceso $ARMA(1,1)$ y la de un proceso $MA(1)$. Tanto la ACF como la $PACF$ de un $ARMA(1,1)$ no se reducen en ningún retardo finito.

En general, la memoria de un modelo $ARMA(p,q)$ es infinita. Una vez superado el orden de los componentes de promedios móviles, las autocorrelaciones de un modelo $ARMA(p,q)$ se comportan como las de un modelo $AR(1)$; y la $PACF$ del primero mencionado tampoco se anula, como consecuencia de que contiene el modelo $MA(q)$ como un caso especial.

Identificación del orden de un Modelo ARMA

Para los modelos ARMA, se puede dar el caso en que la ACF y la PACF no aporten directamente suficiente información relativa a la determinación de su orden. Tsay and Tiao (1984) proponen un enfoque que utiliza la *función de autocorrelación extendida EACF* para especificar el orden de un proceso ARMA de una manera relativamente simple. Si es posible obtener una estimación coherente del componente autorregresivo AR de un modelo ARMA, entonces se puede inferir el componente MA. A partir de esta última se puede usar la ACF para identificar el orden del componente MA. La salida de la EACF es una tabla, cuyas filas corresponden al orden p del proceso AR y las columnas al orden q del proceso MA. La tabla (2.5) muestra una salida para un modelo ARMA(1, 1). La característica clave de la tabla es que contiene un triángulo de ceros con el vértice superior izquierdo situado en la orden (1,1). Esta es la característica que utilizamos para identificar el orden de un proceso ARMA. En general, para un proceso ARMA(p, q) del modelo, el triángulo de ceros tendrá su vértice superior izquierdo en la posición (p, q). El crite-

Cuadro 2.5: Tabla EACF para un modelo ARMA(1, 1).

AR	MA					
	0	1	2	3	4	5
0	X	X	X	X	X	X
1	X	0	0	0	0	0
2	*	X	0	0	0	0
3	*	*	X	0	0	0
4	*	*	*	X	0	0
5	*	*	*	*	X	0

rio de información discutido en secciones previas puede ser usado para seleccionar el orden de un proceso ARMA. Por lo general, para ciertos valores enteros positivos P y Q especificados previamente, entonces se calcula el estadístico AIC, o el *Criterio de Información Bayesiana de Schwarz (BIC)* para el proceso ARMA(p, q), donde $0 \leq p \leq P$ y $0 \leq q \leq Q$, y se selecciona el modelo con menor AIC (o BIC). Evidentemente, este método requiere el de máxima verosimilitud para la estimación de una gran cantidad de modelos, y en algunos casos puede presentarse problemas de sobreajuste en la estimación.

Una vez especificado el modelo $ARMA(p, q)$, sus parámetros se pueden estimar a partir del método de verosimilitud condicional, o utilizando el de verosimilitud exacta. Además, el estadístico de Ljung-Box de los residuos se puede utilizar para comprobar la adecuación del modelo ajustado.

Predicción

Denotamos por h el origen de predicción, y por \mathcal{F}_h la información disponible hasta dicho instante. La predicción a un paso de x_{h+1} se puede obtener del modelo como

$$\hat{x}_h(1) = \mathbb{E}(x_{h+1}|\mathcal{F}_h) = \phi_0 + \sum_{i=1}^p \phi_i x_{h+1-i} - \sum_{i=1}^q \theta_i \varepsilon_{h+1-i},$$

y el error asociado a la predicción es $e_h(1) = x_{h+1} - \hat{x}_h(1) = \varepsilon_{h+1}$. La varianza del error de predicción a un paso es σ_ε^2 . Para la predicción a k -pasos tenemos

$$\hat{x}_h(k) = \mathbb{E}(x_{h+k}|\mathcal{F}_h) = \phi_0 + \sum_{i=1}^p \phi_i \hat{x}_h(k-i) - \sum_{i=1}^q \theta_i \varepsilon_h(k-i),$$

donde se entiende que $\hat{x}_h(k-i) = x_{h+k-i}$ si $k-i \leq 0$ y $\varepsilon_h(k-i) = 0$ si $k-i > 0$ y $\varepsilon_h(k-i) = \varepsilon_{h+k-i}$ si $k-i \leq 0$. El error de predicción asociado es

$$e_h(k) = x_{h+k} - \hat{x}_h(k).$$

2.1.5. Representaciones Alternativas de un Proceso ARMA

Como hemos visto, un modelo $ARMA$ está caracterizado tanto por su ACF como por su $PACF$ las cuales decaen en el infinito pero sin llegar a anularse a partir de cierto retardo, una característica que lo diferencia de los procesos autorregresivos o medias móviles. Esto se puede ver considerando el modelo (2.23)

$$\phi_p(L)x_t = \phi_0 + \theta_q(L)\varepsilon_t.$$

Ahora bien, dados

$$\frac{\phi(L)}{\theta(L)} = 1 - \pi_1 L - \pi_2 L^2 - \dots \equiv \pi(L), \tag{2.25}$$

y

$$\frac{\theta(L)}{\phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + \dots \equiv \psi(L), \quad (2.26)$$

entonces se tiene lo siguiente:

- Si se cumplen las condiciones de invertibilidad del proceso $ARMA(p, q)$ (es decir, las soluciones del polinomio $\theta(L)$ son mayores en módulo que la unidad), entonces el proceso dado por la ecuación (2.23) se puede escribir, usando el operador definido en (2.25) como

$$\pi(L)x_t = \frac{\phi(L)}{\theta(L)}x_t = \frac{\phi_0}{\theta(L)} + \varepsilon_t = \mu + \varepsilon_t,$$

donde $\mu = \phi_0/\theta(L)$, entonces el proceso $ARMA(p, q)$ queda determinado por

$$x_t = \frac{\phi_0}{1 - \theta_1 - \dots - \theta_q} + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + \varepsilon_t. \quad (2.27)$$

Esta representación corresponde a la de un proceso $AR(\infty)$, y está asociada a la *representación AR* de un proceso $ARMA$. En ella se puede observar la dependencia del valor actual x_t con los valores pasados x_{t-i} , con $i > 0$. Los coeficientes $\{\pi_i\}$ hacen referencia a los pesos del proceso $ARMA$. Otra característica a resaltar es que la contribución de los valores x_{t-i} al valor x_t disminuye a medida que se incrementa el valor i , así, los coeficientes π_i decaen a cero exponencialmente cuando se incrementa i .

- Si se cumplen las condiciones de estacionariedad (es decir, las soluciones del polinomio $\phi(L)$ son mayores en módulo que la unidad), entonces el modelo $ARMA$ se puede expresar, usando el operador definido en (2.26) como

$$x_t = \frac{\phi_0}{\phi(L)} + \frac{\theta(L)}{\phi(L)}\varepsilon_t = \mu + \psi(L)\varepsilon_t,$$

donde $\mu = \phi_0/\phi(L)$, de manera que el proceso $ARMA(p, q)$ se representa por

$$x_t = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \quad (2.28)$$

Esta es la *representación MA* de un proceso *ARMA*, puesto que el proceso obtenido es un $MA(\infty)$. Un aspecto a destacar es que mediante esta representación se muestra de manera explícita el impacto de los valores pasados ε_{t-i} ($i > 0$) en el valor actual x_t . Los coeficientes $\{\psi_i\}$ hacen referencia a la función impulso respuesta de un proceso *ARMA*. Bajo la condición de estacionariedad, los coeficientes ψ_i decaen exponencialmente cuando i es incrementado. Otra característica relevante es el efecto de los ε_{t-i} en los valores x_t , cuyo impacto no es permanente ya que estos tienden a desvanecerse con respecto al tiempo. La representación *MA* es de gran utilidad puesto que proporciona, entre otras cosas, evidencia acerca de la *reversión a la media* (los valores de la serie se aproximan a su media en plazos largos) de una serie de tiempo estacionaria.

3

Modelos No-Estacionarios

Los modelos considerados en la sección anterior están basados en la suposición de que los procesos son estacionarios, lo cual implica que la media, varianza y autocovarianzas del proceso son invariantes bajo traslaciones respecto del conjunto de índices (el tiempo). Es claro que, como hemos visto, esto restringe a la media y varianza a ser constantes y a la autocovarianza a ser una función que depende solo de los retardos. Sin embargo, una gran parte de las series económicas, entre ellas las financieras, son ciertamente *no estacionarias* y, en particular, tienen tendencia a presentar variaciones en sus niveles respecto del tiempo y en algunos casos en la varianza. En finanzas, por ejemplo, las series de tipos de interés, tasas de cambio, o series de precios de un activo son de gran interés. Para una serie de precios de algún activo en particular, la no estacionariedad es debida principalmente al hecho de que no hay un nivel fijo de precios. En las series de tipos de interés interbancario de la zona Euro (*EURIBOR*) por ejemplo, se pueden observar notables variaciones tanto en su nivel como en su varianza. La figura (3.1) muestra 4 series del ámbito financiero; tipos de cambio US-Dólar-Euro comprendida entre el 4 de Enero de 1999 al 20 de Enero de 2011, precio de las acciones Santander *IBEX-36* para el período de negociación 2008-2010, y tipos de interés a plazos de 1 semana y 12 meses, para el período del

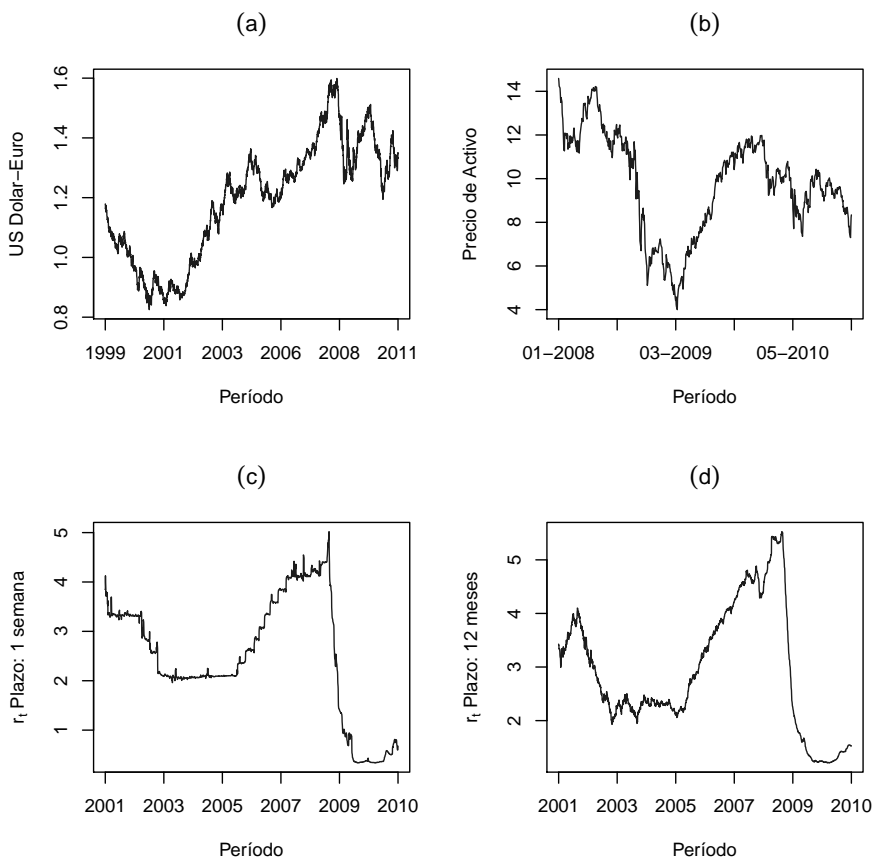


Figura 3.1: Series Financieras: (a) tipos de cambio US-Dólar-Euro, período del 4 de Enero de 1999 al 20 de Enero de 2011, (b) precio de las acciones Santander *IBEX-35* período 2008-2010, (c) y (d) tipos de interés a plazos de 1 semana y 12 meses respectivamente, para el período del 15 de Octubre 2001 al 3 de Diciembre de 2010.

15 de Octubre de 2001 al 3 Diciembre 2010, respectivamente. En todas ellas se observan variaciones notables en media y varianza. A continuación se discuten brevemente estos conceptos.

3.1. No estacionariedad en la Varianza

Cuando una serie se observa a lo largo del tiempo, ocurre con frecuencia que la varianza se ve afectada por una “tendencia”. Para ver esto, en primer lugar se asumirá que una serie de tiempo se puede descomponer en dos términos, un primer término no-estocástico, su media, y un segundo término un error aleatorio

$$x_t = \mu_t + \varepsilon_t, \quad (3.1)$$

y se supondrá además que la varianza de los errores, ε_t , está relacionada de manera funcional a la media μ_t por $\text{Var}(x_t) = \text{Var}(\varepsilon_t) = \mu_t h^2 \sigma^2$, donde h es una función conocida. El objetivo es encontrar una transformación de los datos, digamos $g(x_t)$, la cual se encargue de estabilizar la varianza, en otras palabras, que la varianza de la variable transformada $y_t = g(x_t)$ sea constante. Box and Cox (1964) definieron una manera práctica y general para seleccionar la transformación $g(x_t)$. Esta transformación se define por:

$$g(x_t) = \begin{cases} \frac{x_t^\lambda - 1}{\lambda}; & \lambda \neq 0, \\ \ln(x_t), & \end{cases}$$

en lo que respecta a las series de tiempo en finanzas, el uso del logaritmo es una transformación bastante popular en el contexto financiero, sin embargo, cabe resaltar que no siempre es posible inducir una varianza del todo constante aplicando solo este tipo de transformaciones. Como veremos más adelante, existen propuestas que permiten modelizar de manera más eficiente este tipo de situaciones.

3.2. No estacionariedad en la Media

Considérese nuevamente la ecuación (3.1), y suponga, en este caso, que el término μ_t no es constante. En la literatura de las series de tiempo,

este tipo de modelos ha sido ampliamente estudiado, y existe una amplia variedad de formas para modelizar una serie de tiempo con estas características. Una alternativa es suponer que la media es un término de *tendencia*, el cual viene expresado como un polinomio de orden d en el tiempo y que el término del error ε_t es un estocástico, estacionario, autocorrelacionado y de media cero. Lo cual siempre es posible si se toma en cuenta el resultado que extiende el teorema de descomposición de Wold para procesos no-estacionarios (ver, Cramer (1961)). De esta manera se obtiene que el proceso se puede expresar como

$$x_t = \mu_t + \varepsilon_t = \sum_{j=1}^d \beta_j t^j + \psi(L)\varepsilon_t. \quad (3.2)$$

Puesto que $\mathbb{E}(\varepsilon_t) = \psi(L)E(\varepsilon_t) = 0$, entonces $\mathbb{E}(x_t) = \mathbb{E}(\mu_t) = \sum_{j=1}^d \beta_j t^j$, y dado que los coeficientes β_j permanecen constantes en el tiempo, dicha tendencia se considera un término *determinista*. Así pues, tendencias de este tipo pueden ser removidas aplicando una transformación simple. Considérese $d = 1$ en la ecuación (3.2), entonces tenemos el caso de *tendencia lineal*, se asumirá, por simplicidad, que el componente del error es un ruido blanco, y

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (3.3)$$

Una técnica comúnmente usada en el análisis de series de tiempo es aplicar una cierta transformación, en este caso se consideran la *diferencias de primer orden* de x_t , obteniéndose entonces

$$\begin{aligned} y_t = x_t - x_{t-1} &= (1 - L)x_t = \nabla x_t, \\ &= \beta_1 + \nabla \varepsilon_t, \end{aligned}$$

donde $\nabla = (1 - L)$ es conocido como el *operador de diferencias de primer orden*. De esta manera, el nuevo proceso y_t es generado por un proceso $MA(1)$ estacionario (puesto que $\mathbb{E}(y_t) = \beta_1$, es constante), sin embargo no es invertible.

En general, si la tendencia es determinada por un polinomio de un cierto orden d , y además ε_t está caracterizado por un proceso $ARMA$

$$\phi(L)x_t = \theta(L)\varepsilon_t,$$

entonces se consideran las *diferencias de orden d* ,

$$\nabla^d x_t = (1 - L)^d x_t,$$

y se obtiene el nuevo proceso

$$w_t = \nabla^d x_t = \theta_0 + \frac{\nabla^d \theta(L)}{\phi(L)} \varepsilon_t,$$

donde $\theta_0 = d! \beta_d$. Así, la parte MA del proceso generado por $\nabla^d x_t$ tendrá d raíces iguales a la unidad.

Un caso particular de procesos no-estacionarios, conocido ampliamente en el análisis de series de tiempo, es el *paseo aleatorio*

$$x_t = x_{t-1} + \varepsilon_t, \quad (3.4)$$

donde $\{\varepsilon_t\}$ es un ruido blanco y x_0 es un número real el cual denota el valor inicial. Si x_t representa el logaritmo del precio de un cierto activo en el instante t , entonces x_0 podría ser el precio del activo en su oferta pública inicial. Bajo este modelo, el precio de la acción no se puede predecir y tampoco hay reversión de éste a la media. Claramente (3.4) es un proceso $AR(1)$, con $\phi_1 = 1$. El proceso es no estacionario puesto que la raíz del polinomio asociado es igual a la unidad, además:

$$\text{Var}(x_t) = \text{Var}(x_{t-1}) + \sigma_\varepsilon^2,$$

por lo tanto, no puede ocurrir $\text{Var}(x_t) = \text{Var}(x_{t-1})$. Por otra parte, suponiendo que el proceso comienza en $t = 0$ (con $x_0 = 0$), se tiene que para $h > 0$,

$$\text{Cov}(x_{t+h}, x_t) = t\sigma^2.$$

De la expresión anterior se deduce que la varianza de x_t tiende al infinito a medida que aumenta el valor de t . Así mismo, se puede deducir que la función de autocorrelación del paseo aleatorio se expresa como

$$\rho(x_{t+h}, x_t) = \frac{t}{\sqrt{t(t+h)}}. \quad (3.5)$$

El paseo aleatorio pertenece a una clase de procesos denominados *procesos no estacionarios de raíces unitarias*. Si se incluye un término constante, μ , en la ecuación (3.4) entonces se obtiene el *paseo aleatorio con drift*

$$x_t = \mu + x_{t-1} + \varepsilon_t, \quad (3.6)$$

donde $\mu = \mathbb{E}(x_t - x_{t-1})$, y como antes $\{\varepsilon_t\}$ es un ruido blanco. En finanzas, la serie del logaritmo de los retornos del índice de un mercado tiene un valor medio pequeño y positivo; por ello, la constante μ en el proceso (3.6) es de gran importancia. Si x_t , por ejemplo, es el logaritmo del precio de un activo, esta constante representa la tendencia temporal de x_t . La ecuación (3.6) se puede escribir como

$$\nabla x_t = \mu + \varepsilon_t,$$

de manera que al aplicar las primeras diferencias a x_t se obtiene como resultado un proceso estacionario. Por lo general, no siempre resulta suficiente aplicar solo una diferencia, por lo que se hace necesario aplicar diferencias de primer orden repetidas veces, por ejemplo d , hasta alcanzar la estacionariedad, y la serie así obtenida puede ser en sí misma correlacionada. Si esta correlación es modelizada por un proceso $ARMA(p, q)$ entonces el modelo para la serie original es de la forma

$$\phi(L)\nabla^d x_t = \mu + \theta(B)\varepsilon_t, \quad (3.7)$$

conocido como *proceso autorregresivo integrado de medias móviles de orden p , d y q* , o simplemente $ARIMA(p, d, q)$, y x_t se dice que es *integrado* de orden d y se denota por $I(d)$. Cuando el polinomio AR tiene una sola raíz unitaria, entonces se dice que el modelo $ARIMA$ es un proceso no-estacionario con raíz unitaria. En finanzas, las series de precios son comúnmente no-estacionarias, sin embargo, la diferencia del logaritmo de los precios, $r_t = \ln(p_t) - \ln(p_{t-1})$, es estacionaria. En este caso, la serie de los logaritmos de los precios es no-estacionaria de raíces unitarias y por lo tanto puede ser tratada como un proceso $ARIMA$.

3.3. Test de Raíz Unitaria

En esta sección buscamos probar la presencia de una o más *raíces unitarias* en el polinomio autorregresivo de orden p , $\phi(L)$ en el modelo

$$\phi(L)x_t = \theta_0 + \theta(L)\varepsilon_t,$$

donde x_0 se asume fijo y $\theta_0 = \phi(1)\mu$, con μ la media de x_t . Said and Dickey (1984), modificaron el test de Dickey and Fuller (1979) basado en un proceso autorregresivo no estacionario a un modelo general

$ARMA(p, q)$ cuyos ordenes son desconocidos, este test es conocido como el *test de raíz unitaria ampliado de Dickey-Fuller (ADF)*. Este test contrasta la hipótesis nula de existencia de una raíz unitaria contra la alternativa de que no existen raíces unitarias. Para verificar la existencia de una raíz unitaria en un proceso $AR(p)$, se lleva a cabo el siguiente test:

$$H_0 : \beta = 1 \quad \text{versus} \quad H_a : \beta < 1,$$

aplicando la regresión

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t, \quad (3.8)$$

donde c_t es una función determinista del tiempo t y $\Delta x_j = x_j - x_{j-1}$ es la serie de diferencias de x_t usada para aproximar la estructura $ARMA$ de los errores, y el valor de p se fija de modo que el error e_t sea correlacionado serialmente. El término de error también se supone homoscedástico. En la práctica, c_t puede ser cero o una constante o bien $c_t = \omega_0 + \omega_1 t$. El *ADF* se basa en las estimaciones de mínimos cuadrados de (3.8) y es dado por

$$\text{ADF-test} = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})},$$

donde $\hat{\beta}$ denota el valor estimado por mínimos cuadrados de β . La especificación del retardo, es un punto importante en la aplicación del test *ADF*. Si se selecciona un p demasiado pequeño, entonces la correlación serial restante producirá un sesgo en el test. Si por el contrario, se escoge un p demasiado grande, entonces la potencia del test se verá afectada. Ng and Perron (1995) propusieron el siguiente procedimiento para la selección de p , el cual garantiza estabilidad en el tamaño del test y una pérdida mínima en la potencia. En primer lugar, se establece una cota superior para p denotado por $p_{\text{máx}}$. A continuación, se estima la regresión *ADF* con $p = p_{\text{máx}}$. Si el valor absoluto del estadístico t para el test de la última diferencia tomada es mayor que 1,6 entonces se fija $p = p_{\text{máx}}$, y posteriormente se lleva cabo el test de raíz unitaria. En caso contrario, se reduce $p = p_{\text{máx}} - 1$ y se repite el proceso. Una regla útil para determinar $p_{\text{máx}}$, sugerida por Schwert

(1989), es

$$p_{\text{máx}} = \left[12 \left(\frac{T}{100} \right)^{1/4} \right],$$

donde $[\cdot]$ denota la parte entera. Esta selección de $p_{\text{máx}}$ permite que este aumente cuando el tamaño de la muestra se incrementa.

4

Otros Modelos

4.1. Modelos Estacionales

Algunas series en finanzas tales como los ingresos trimestrales por los activos de una empresa exhiben un cierto comportamiento cíclico o periódico. Tales series son denominadas *series de tiempo estacionales*. En algunas aplicaciones, la estacionalidad es considerada como algo secundario y por ello es removida, dando como resultado una serie de tiempo estacionalmente ajustada y que luego es usada para estudiarla. Este procedimiento de remoción de estacionalidad de una serie de tiempo se conoce como *ajuste estacional*. En otras aplicaciones como la predicción, uno de los principales objetivos del análisis financiero, la estacionalidad es tan importante como otras características de los datos y debe por consiguiente ser considerada en el análisis de la serie. En esta sección se discutirán algunos modelos econométricos que son útiles en la modelización de series de tiempo estacionales.

Para una serie de tiempo estacional y_t con periodicidad s , la *diferenciación estacional* esta referida a

$$\Delta_s y_t = y_t - y_{t-s} = (1 - L^s)y_t.$$

La diferencia convencional $\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$ se conoce como

la *diferenciación usual*. Consideremos ahora el caso especial de la serie de tiempo estacional

$$(1 - L^s)(1 - L)y_t = (1 - \theta L)(1 - \Theta L^s)\varepsilon_t, \quad (4.1)$$

donde s es la periodicidad de la serie, ε_t es un ruido blanco, $|\theta| < 1$, y $|\Theta| < 1$. Este modelo es conocido en la literatura, como el *modelo de la aerolínea* (ver Box et al. (1994), Chapter 9). La parte *AR* del modelo consta simplemente de usuales y estacionales diferencias, mientras que la parte *MA* involucra dos parámetros. Considérese la parte *MA*,

$$w_t = (1 - \theta L)(1 - \Theta L^s)\varepsilon_t = \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-s} + \theta\Theta\varepsilon_{t-s-1},$$

donde $w_t = (1 - L^s)(1 - L)y_t$. El proceso w_t se denomina *modelo estacional multiplicativo MA*. En aplicaciones, un modelo multiplicativo estacional supone que la dinámica de las componentes regular y estacional de la serie son aproximadamente ortogonales.

4.2. Modelos de Memoria Larga

Hasta el momento se ha realizado una exposición de los procesos estacionarios y no estacionarios. Los primeros, tienen función de autocorrelación *ACF* que decae exponencial a cero a medida que se incrementan los retardos. Para los procesos no-estacionarios, en particular los de raíces unitarias, se puede demostrar que su función de autocorrelación muestral converge a 1 para todos los retardos fijos cuando se incrementa el tamaño de la muestra (ver Chan and Wei (1988) y Tiao and Tsay (1983)). Sin embargo, en la literatura de las series de tiempo, existen series cuya *ACF* decae lentamente a cero, a medida que se incrementa el retardo, a una tasa polinómica. Estos procesos se conocen como *procesos de memoria larga*. Un ejemplo de ello son los procesos fraccionalmente diferenciados definidos por:

$$(1 - L)^d x_t = \varepsilon_t, \quad -0,5 < d < 0,5, \quad (4.2)$$

donde $\{\varepsilon_t\}$ es un proceso ruido blanco. Resumimos alguna de las propiedades del modelo (4.2) a continuación (para detalles de este modelo ver Hosking (1981), Granger and Joyeux (2001), y Jun (2001)).

- Si $d < 0,5$, entonces x_t es un proceso débilmente estacionario con representación *MA* infinita

$$x_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}, \quad \text{con} \quad \psi_k = \frac{d(d+1) \dots (k-1+d)}{k!}$$

$$= \frac{(k+d-1)!}{k!(d-1)!}.$$

- Si $d > -0,5$, entonces x_t es invertible con representación *AR* infinita

$$x_t = \sum_{i=1}^{\infty} \pi_i x_{t-i} + \varepsilon_t, \quad \text{con} \quad \pi_k = \frac{-d(1-d) \dots (k-1-d)}{k!}$$

$$= \frac{(k-d-1)!}{k!(-d-1)!}.$$

- Para $-0,5 < d < 0,5$, la *ACF* de x_t es

$$\rho_k = \frac{d(d+1) \dots (k-1+d)}{(1-d)(2-d) \dots (k-d)}, \quad k = 1, 2, \dots$$

En particular, $\rho_1 = d/(1-d)$ y

$$\rho_k \approx \frac{(-d)!}{(d-1)!} k^{2d-1}, \quad \text{cuando} \quad k \rightarrow \infty.$$

- Para $-0,5 < d < 0,5$, la *PACF* de x_t es $\phi_{kk} = d/(k-d)$ para $k = 1, 2, \dots$
- Para $-0,5 < d < 0,5$, la función de densidad espectral $f(\omega)$ de x_t , que es la transformada de Fourier de la *ACF* de x_t , satisface

$$f(\omega) \sim \omega^{-2d}, \quad \omega \rightarrow 0, \tag{4.3}$$

donde $\omega \in [0, 2\pi]$ denota la frecuencia.

El estudio del comportamiento de la *ACF* de x_t es de particular interés para estos procesos con $d < 0,5$. La propiedad $\rho_k \sim k^{2d-1}$, expresa que el decaimiento es a una tasa polinómica, en lugar de una tasa exponencial. Por ello, el proceso es llamado serie de tiempo de memoria larga. Si la serie diferenciada fraccionalmente $(1-B)^d x_t$ obedece a un proceso *ARMA*(p, q) entonces x_t se le conoce con el nombre de proceso *ARFIMA*(p, d, q), el cual generaliza los procesos *ARIMA* permitiendo que el parámetro d sea no entero.

4.3. Modelos de Regresión de Series de Tiempo

Las técnicas de regresión de series de tiempo se aplican con mucha frecuencia en el análisis de datos financieros, en la estimación y en la validación de modelos de precios de activos y rentabilidades. En finanzas, la capacidad de predicción de los rendimientos de los activos haciendo uso de tasas de valoración como *dividendo/precio*, *ganancias/precio* entre otras, generalmente se establecen mediante las técnicas de regresión de series de tiempo, y donde el modelo de regresión resultante es usado para la predicción de las rentabilidades futuras. Los modelos de regresión de series de tiempo también se utilizan para probar la eficiencia informativa de los mercados financieros. Sin embargo, es importante ser cuidadosos a la hora de aplicar este tipo de modelos, puesto que las propiedades de las series de tiempo de los datos puede influir en las propiedades de los valores estimados en la regresión y en la inferencia. En términos generales, estos modelos son apropiados para el análisis de series estacionarias. Para el caso de las series con tendencia no-estacionarias puede ser o no apropiadas, dependiendo entonces de la naturaleza del componente de tendencia.

Consideremos el *modelo de regresión lineal de series de tiempo*

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \varepsilon_t = x_t' \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \dots, T \quad (4.4)$$

donde $x_t = (1, x_{1t}, \dots, x_{kt})'$ es un vector de variables explicativas de dimensión $(k+1) \times 1$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$ un vector de coeficientes dimensión $(k+1) \times 1$, y ε_t es un término de error aleatorio. En su forma matricial el modelo se puede expresar como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4.5)$$

donde $\boldsymbol{\varepsilon}$ y la variable \mathbf{y} son vectores de dimensión $(T \times 1)$ y \mathbf{X} es una matriz de dimensión $(T \times (k+1))$. El modelo (4.4) satisface las suposiciones:

- El modelo está especificado correctamente.
- $\{y_t, x_t\}$ es estacionario y ergódico conjuntamente.
- Las variables regresoras x_t son tales que: $\mathbb{E}[x_{is}\varepsilon_t] = 0$ para todo $s \leq t$ e $i = 1, \dots, k$.

- $\mathbb{E}[x_t x_t'] = \Sigma_{XX}$ es de rango $k + 1$.
- $\{x_t \varepsilon_t\}$ es un proceso no correlacionado con matriz de covarianza $(k + 1) \times (k + 1)$ finita $\mathbb{E}[\varepsilon_t^2 x_t x_t'] = S = \sigma^2 \Sigma_{XX}$.

La segunda suposición descarta los regresores con tendencia, la tercera descarta los regresores endógenos pero permite variables y retardos dependientes. La cuarta suposición evita regresores redundantes o multicolinealidad exacta, y la quinta implica que el término de error es un proceso no correlacionado serialmente con varianza incondicional σ^2 constante. En el modelo de regresión de series de tiempo, los regresores x_t son aleatorios y el error ε_t no se asume distribuido normal. La estimación del modelo se lleva a cabo mediante *mínimos cuadrados ordinarios* y está basada en la suma de los residuos al cuadrado

$$SSR(\beta) = \sum_{t=1}^T (y_t - x_t' \beta)^2 = \sum_{t=1}^T \varepsilon_t^2,$$

y produce el modelo ajustado

$$y_t = x_t' \hat{\beta} + \hat{\varepsilon}_t, \quad t = 1, \dots, T,$$

donde

$$\hat{\beta} = (X'X)^{-1} X'y = \left[\sum_{t=1}^T x_t x_t' \right]^{-1} \sum_{t=1}^T x_t y_t,$$

y $\hat{\varepsilon}_t = y_t - \hat{y}_t = y_t - x_t' \hat{\beta}$. La varianza del error se estima por $\sigma^2 = \hat{\varepsilon}' \hat{\varepsilon} / (T - k - 1)$.

La relación entre dos series de tiempo es de relevante interés en muchas aplicaciones. La estructura de los tipos de interés es un ejemplo en el cual se investiga la evolución en el tiempo de la relación entre tipos de interés con diferentes plazos de vencimiento. Esos ejemplos conllevan a considerar la regresión lineal de la forma:

$$r_{6t} = \beta_0 + \beta_1 r_{3t} + \varepsilon_t,$$

donde r_{6t} y r_{3t} son dos series de tiempo y ε_t denota el error. Si los términos de error $\{\varepsilon_t\}$ constituyen un proceso de ruido blanco, entonces el método de mínimos cuadrados produce estimaciones consistentes. Sin

embargo, en la práctica, es común ver que los errores ε_t están correlacionados serialmente. En este caso, se tiene un modelo de regresión fuertemente influenciado por los términos de perturbación aleatoria, y los valores estimados de β_0 y β_1 por mínimos cuadrados pueden no ser consistentes. Modelo de regresión de este tipo son ampliamente usados en economía y finanzas, y comúnmente son implementados de manera inadecuada, pues la dependencia serial en ε_t es a menudo pasada por alto. Por tal razón, vale la pena estudiarlos cuidadosamente. Para ver esto, se introduce el modelo que considera la relación entre dos series de tipos de interés, con frecuencia diaria, del Mercado Interbancario Europeo (*EURIBOR*):

- r_{3t} : serie de tipo de interés con plazo de vencimiento a 3-meses,
- r_{6t} : serie de tipo de interés con plazo de vencimiento a 6-meses.

Ambas series tienen 2336 observaciones (en porcentajes). El período temporal abarca desde el 15 de Octubre del 2001 hasta el 03 de Diciembre de 2010.

La figura (4.1) muestra el gráfico de estas dos series. En azul (*línea continua*) se observa la serie con plazo de vencimiento a 3-meses, y en rojo (*línea discontinua*) la serie con plazo de vencimiento a 6-meses. La figura (4.2(a)) muestra dos gráficos en los cuales se observa, como era de esperarse, que ambas series de tipos de interés están altamente correlacionadas. Una manera simple para describir la relación entre las dos series de tipos de interés es a partir del modelo lineal $r_{6t} = \beta_0 + \beta_1 r_{3t} + \varepsilon_t$. El modelo que resulta del ajuste es

$$r_{6t} = 0,1953 + 0,9621r_{3t} + \varepsilon_t, \quad \hat{\sigma}_\varepsilon = 0,1, \quad (4.6)$$

con $R^2 = 99,31\%$, donde los errores estándar de los dos coeficientes se muestran en la tabla (4.1).

Cuadro 4.1: Resumen de los datos del ajuste lineal a la serie de los tipos de interés.

	Estimados	Error Estándar	t-valor	$\Pr(> t)$
β_0	0.1953	0.0049	39.61	0.0000
β_1	0.9621	0.0017	578.96	0.0000

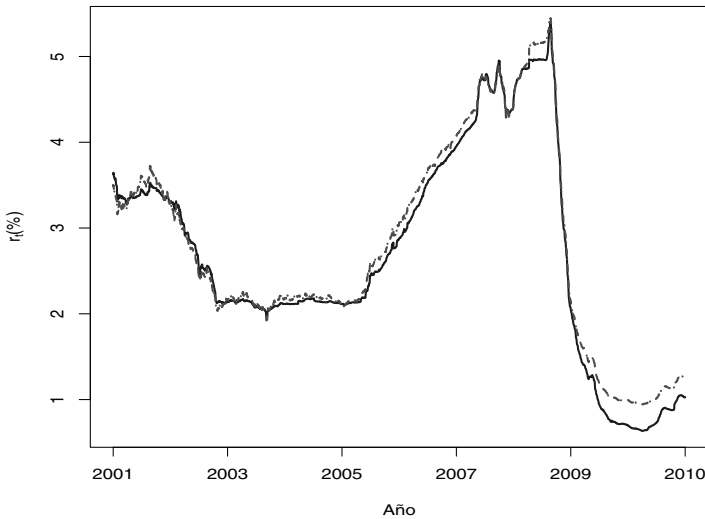


Figura 4.1: Gráficos de las series de tipos de interés del *EURIBOR*, desde el 15 de octubre del 2001 hasta el 03 de Diciembre de 2010. En azul, (línea continua) serie con plazo de vencimiento a 3-meses, y en rojo, (línea discontinua) serie con plazo de vencimiento a 6-meses.

El modelo (4.6) confirma la alta correlación entre las dos series. Sin embargo, al observar el gráfico de los residuos y su función de autocorrelación muestral (*ACF*) en la figura (4.3), se concluye que el modelo ajustado (4.6) es completamente inadecuado.

En particular, la *ACF* de los residuos es excepcionalmente significativa y su decaimiento es bastante lento, mostrando un patrón similar a una serie con raíces unitarias. El comportamiento de los residuos sugiere que existen marcadas diferencias entre los dos tipos de interés.

Tomando en cuenta lo discutido en secciones anteriores, el comportamiento no-estacionario (raíz unitaria) de ambas series de tipos de interés y de los residuos obtenidos a partir de la ecuación (4.6); surge la necesidad de considerar la serie de las diferencias para las series del *EURIBOR*. Así pues, se consideran las nuevas series

- $d_{3t} = r_{3t} - r_{3,t-1} = (1 - L)r_{3t}$ para $t \geq 2$: serie de las variaciones

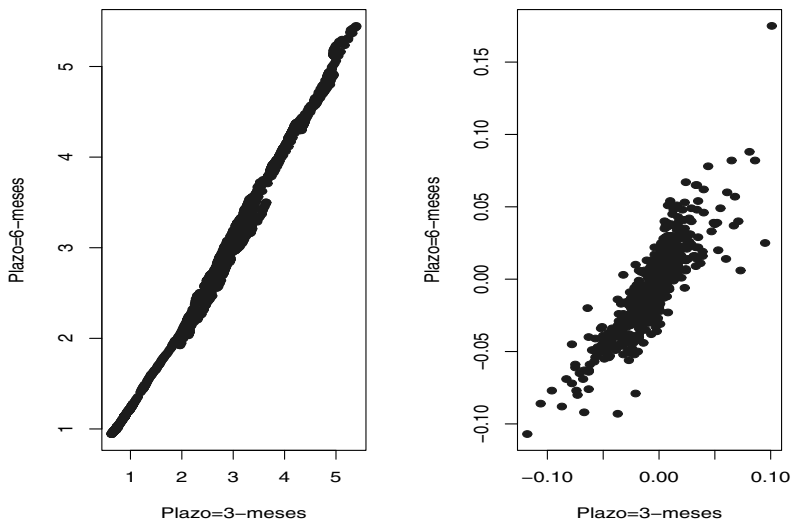


Figura 4.2: Scatterplots de las series de tipos de interés del *EURIBOR*: (a) tipos de interés con vencimiento: 3-meses vs 6-meses, (b) serie de las diferencias de los tipos de interés con vencimiento: 3-meses vs 6-meses.

(cambios) en los tipos de interés con plazo de vencimiento a 3-meses,

- $d_{6t} = r_{6t} - r_{6,t-1} = (1 - L)r_{6t}$ para $t \geq 2$: serie de las variaciones (cambios) en los tipos de interés con plazo de vencimiento a 6-meses,

y el modelo de regresión $d_{6t} = \beta_0 + \beta_1 d_{3t} + \varepsilon_t$. La figura (4.4) muestra los gráficos de las variaciones del tipo de interés para ambos plazos, 3-meses y 6-meses; en ellas se puede observar el impacto de la crisis en el período comprendido entre mediados del 2008 y 2009 el cual genera una distorsión en la serie de las variaciones del tipo de interés y que, por supuesto, incide significativamente en su comportamiento no estacionario. Por otro lado, La figura (4.2(b)) muestra el scatterplot de las series de las variaciones del tipo de interés. Se hace el ajuste del modelo de regresión lineal entre las nuevas series, obtenidas mediante diferenciación, y se tiene el siguiente modelo:

$$d_{6t} = 0,0001 + 0,9845d_{3t} + \varepsilon_t, \quad \hat{\sigma}_\varepsilon = 0,00894, \quad (4.7)$$

con $R^2 = 70,94\%$. Los errores estándar de los coeficientes se muestran en la tabla (4.2). En este nuevo modelo se observa una pequeña disminución de la dependencia lineal entre los tipos de interés, sin embargo, esta sigue siendo alta.

Cuadro 4.2: Resumen de los datos del ajuste lineal a la serie de las variaciones de los tipos de interés.

	Estimados	Error Estándar	t-valor	Pr(> t)
β_0	0.0001	0.0002	0.75	0.4552
β_1	0.9845	0.0130	75.50	0.0000

En la figura (4.5) se muestran los gráficos de la ACF de los residuos obtenidos a partir del modelo (4.7). Para éste, la ACF presenta un mejor comportamiento con respecto a la función de los residuos del modelo inicial estimado. Sin embargo, todavía persisten correlaciones seriales significativas en los residuos para ciertos retardos. Esta correlación serial no es débil; y como consecuencia de ello, se puede considerar algunos de los modelos discutidos en secciones previas para modelizar esta

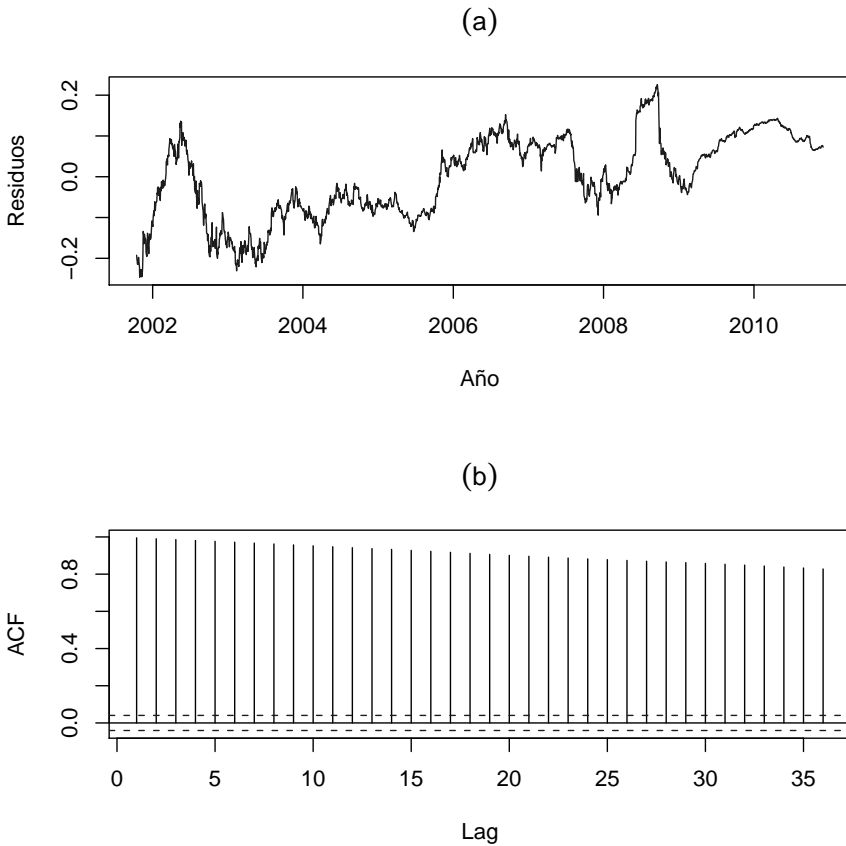


Figura 4.3: Serie de los residuos de la regresión para las dos series de tipos de interés: (a) gráfico de los residuos, (b) función de autocorrelación de la muestra.

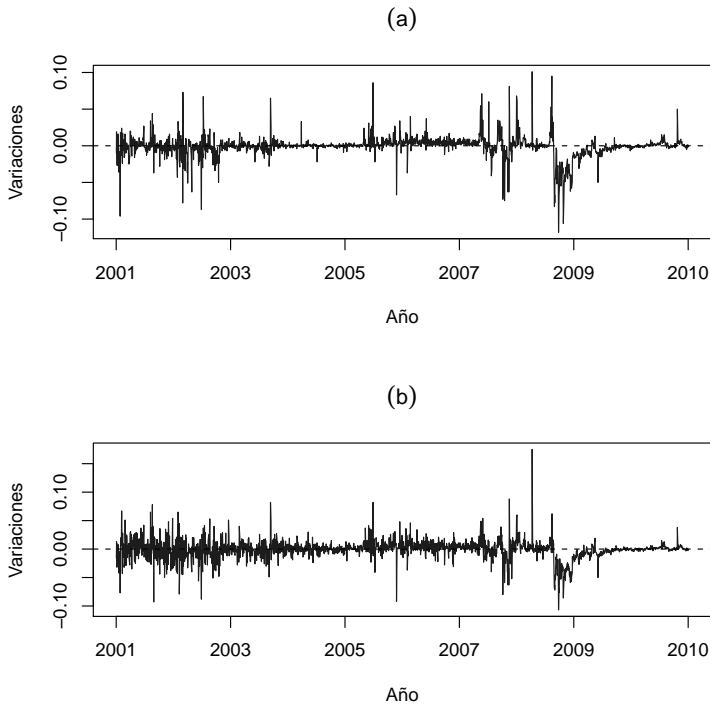


Figura 4.4: Serie de las variaciones (diferencias de primer orden) de los tipos de interés: (a) serie de tipos de interés con vencimiento a 3-meses, (b) serie de tipos de interés con vencimiento a 6-meses.

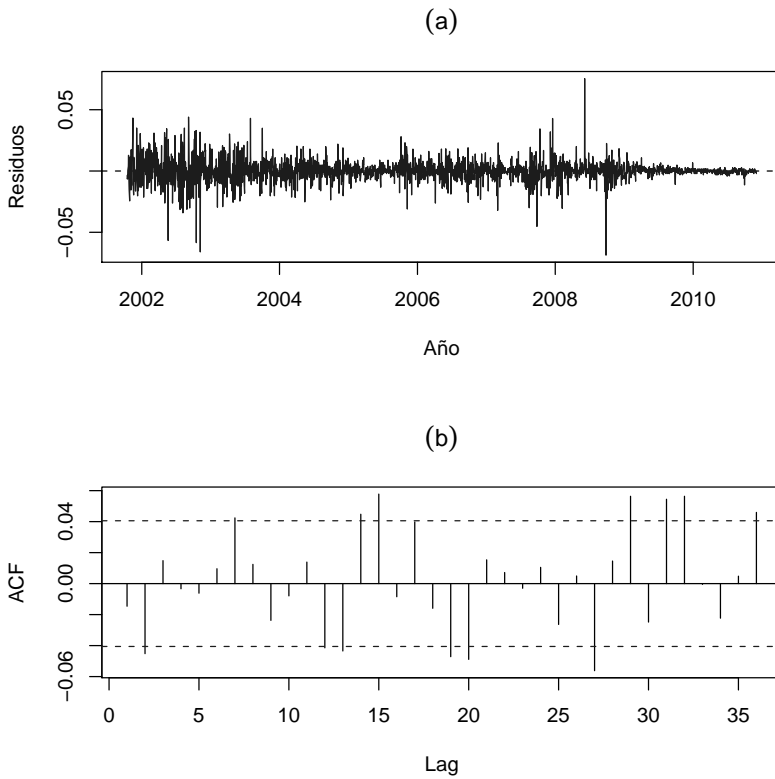


Figura 4.5: Serie de los residuos de la regresión para las variaciones de las dos series de tipos de interés: (a) gráfico de los residuos, (b) función de autocorrelación de la muestra.

dependencia, obteniéndose de esta manera un modelo de regresión lineal para la serie de los errores.

Como se señaló anteriormente, es viable y adecuado discutir una metodología para construir modelos de regresión lineal con series de tiempo como errores. El método en si es relativamente sencillo, basta considerar un modelo de serie de tiempo, como los discutidos en este capítulo para modelizar la serie de los residuos y estimar el modelo completo de manera conjunta. Para ilustrar el procedimiento, considérese el modelo de regresión lineal de la ecuación (4.7). Como la correlación entre los residuos es significativa, se tendrá en cuenta un modelo *ARMA* para estos. Ahora, tomando en cuenta la *ACF* de los residuos mostrada en la figura (4.5) se tendrá en cuenta un modelo *MA*(1) para los residuos y se modificará el modelo de regresión lineal a:

$$d_{6t} = \beta_0 + \beta_1 d_{3t} + \varepsilon_t, \quad \varepsilon_t = \epsilon_t - \theta_1 \epsilon_{t-1}, \quad (4.8)$$

asumiendo que $\{\epsilon_t\}$ es un proceso ruido blanco. El modelo obtenido es un modelo de regresión lineal simple con una serie adicional, la del error. Es claro que, en la práctica, se pueden considerar una gran variedad de modelos de series de tiempo más complejos para ser agregados al modelo de regresión lineal y formar así un modelo general de regresión con la serie de los términos de perturbación aleatoria. Para la serie de tiempo de los tipos de interés del *EURIBOR*, el modelo ajustado, tomando en cuenta la ecuación (4.8), es:

$$d_{6t} = 0,0001 + 0,9844d_{3t} + \varepsilon_t, \quad \varepsilon_t = \epsilon_t - 0,0160\epsilon_{t-1}, \quad \hat{\sigma}_\epsilon = 0,00894, \quad (4.9)$$

con $R^2 = 71,05\%$. Los errores estándar de los parámetros son 0,0002, 0,0129 y 0,0217 respectivamente. Como se puede observar, el modelo resultante no aporta mayor mejoría al modelo previamente estimado. Comparando los distintos modelos ajustados a las series de tipos de interés del *EURIBOR*, podemos establecer algunas observaciones respecto de estos:

- El coeficiente de correlación R^2 es significativamente alto y el coeficiente $\beta_1 = 0,9621$ del modelo ajustado (4.6) pueden conducir a conclusiones erradas, ya que los residuos del modelo muestran una fuerte correlación serial.

- Para la serie de variaciones del tipo de interés, el coeficiente R^2 y el coeficiente estimado \hat{d}_{3t} de los modelos (4.7) y (4.9) son casi similares. En este caso particular, el aporte del modelo $MA(1)$ al general no proporciona una mejora sustancial. Esto no debería sorprender ya que el coeficiente MA estimado no es estadísticamente significativo.
- El análisis demuestra, como ya se ha comentado, que es importante comprobar la dependencia residual de la serie en el análisis de regresión lineal.

Una herramienta eficaz en la determinación de la correlación serial es el estadístico propuesto por Ljung and Box (1978) (discutido en la sección (2.1.2)) el cual se usa en lugar del estadístico de Durbin and Watson (1950) (DW). Esto debido a que este último sólo tiene en cuenta la correlación del primer retardo de la serie. Hay casos en los que la dependencia de la serie de los residuos se presenta en ordenes superiores. Este tipo de situaciones suelen presentarse en las series de tiempo que exhiben un comportamiento estacional.

Observación 4.3.1. Para una serie de residuos ε_t con T observaciones, el estadístico de *Durbin-Watson* se define por:

$$DW = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T \varepsilon_t^2}. \quad (4.10)$$

Un cálculo directo muestra que $DW \approx 2(1 - \hat{\rho}_1)$, donde $\hat{\rho}_1$ es la *ACF* del primer retardo de la serie $\{\varepsilon_t\}$.

4.3.1. Estimación consistente de la Matriz de Covarianza

Bajo la suposición de que las estimaciones (de los coeficientes), mediante el método de mínimos cuadrados ordinarios, son consistentes, los métodos disponibles para obtener una estimación consistente de la matriz de covarianza de los coeficientes son:

- *Heterocedasticidad consistente (HC)* (ver White (1980)),
- *Heterocedasticidad y Autocorrelación consistente (HAC)* (ver Newey and West (1987)).

El estimador de White (1980) es,

$$\text{Cov}(\hat{\beta})_{HC} = \left[\sum_{t=1}^T x_t x_t' \right]^{-1} \left[\sum_{t=1}^T \hat{\varepsilon}_t^2 x_t x_t' \right] \left[\sum_{t=1}^T x_t x_t' \right]^{-1}, \quad (4.11)$$

donde $\hat{\varepsilon}_t = y_t - x_t' \hat{\beta}$ es el residuo estimado en el instante t . El estimador de Newey and West (1987) es

$$\text{Cov}(\hat{\beta})_{HAC} = \left[\sum_{t=1}^T x_t x_t' \right]^{-1} \hat{C}_{HAC} \left[\sum_{t=1}^T x_t x_t' \right]^{-1}, \quad (4.12)$$

donde

$$\hat{C}_{HAC} = \sum_{t=1}^T \hat{\varepsilon}_t^2 x_t x_t' + \sum_{j=1}^l \omega_j \sum_{t=j+1}^T (x_t \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} x_{t-j}' + x_{t-j} \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t x_t'),$$

donde l es un parámetro de truncamiento, el cual se sugiere ser seleccionado como la parte entera de $4(T/100)^{(2/9)}$ y ω_j es la función de pesos de Bartlett definida por

$$\omega_j = 1 - \frac{j}{l+1}.$$

5

Modelos de Heterocedasticidad Condiciona

Antes de comenzar, es necesario precisar el concepto de volatilidad en el contexto del análisis financiero. Se denomina *volatilidad* a la tasa relativa a la que un activo experimenta una drástica disminución o aumento de su precio dentro de un período predeterminado de tiempo. La volatilidad se determina mediante el cálculo de la desviación estándar anualizada de la variación diaria del precio. Si el precio de la acción aumenta y disminuye rápidamente durante cortos períodos de tiempo, entonces se dice que tiene una volatilidad alta. Si el precio se mantiene casi siempre en el mismo valor entonces se dice que tiene volatilidad baja. Los inversores evalúan la volatilidad de las acciones antes de tomar una decisión en, la compra de una oferta de acciones nuevas, la adquisición de acciones adicionales de un activo ya presente en una cartera, o en la venta de acciones que actualmente están en poder del inversionista. La idea detrás de la comprensión del comportamiento de la volatilidad de los activos es organizar las inversiones para obtener el máximo rendimiento con el mínimo de oportunidades de pérdida. En esta sección se discutirán algunos de los modelos estadísticos y econométricos mas importantes para la modelización de la volatilidad de

series de tiempo de rentabilidades de activos. A diferencia del análisis de series de tiempo tradicional, el cual se enfoca principalmente en la modelización del momento condicional de primer orden, los denominados *modelos de heterocedasticidad condicional* buscan captar la dependencia dentro del momento condicional de segundo orden, en otras palabras, el objetivo ahora es modelizar la volatilidad. La incertidumbre o riesgo constituye uno de los temas de investigación principales en el análisis financiero. Como se mencionó, la volatilidad es un factor importante en las finanzas puesto que proporciona un método simple para calcular el valor en riesgo de una situación financiera en la gestión de riesgos. Por otra parte, la modelización de la volatilidad de una serie de tiempo puede mejorar la eficiencia en la estimación de parámetros y la exactitud en los intervalos de predicción. En esta sección se discutirán los modelos univariados de la volatilidad entre los que se incluyen el *modelo autorregresivo de heterocedasticidad condicional (ARCH)* de Engle (1982), el *modelo generalizado ARCH (GARCH)* de Bollerslev (1986), entre otros.

La volatilidad tiene la particularidad de que no es posible su observación directa. Aún cuando esto no es posible, la volatilidad tiene algunas características que pueden ser observadas en las series de rentabilidad de activos entre los que se pueden destacar,

- Agrupamiento de la volatilidad (cluster). En otras palabras, períodos de volatilidades altas y períodos de volatilidades bajas.
- Evolución continua de la volatilidad en el tiempo.
- Las variaciones de la volatilidad se presentan en un rango fijo, es decir, no diverge al infinito. En términos estadísticos, se puede decir que la volatilidad es a menudo estacionaria.
- La volatilidad parece reaccionar de manera diferente a un incremento elevado de los precios o una disminución sustancial de los precios. Este efecto es conocido con el nombre de *apalancamiento* o *efecto palanca*.

Tales propiedades descritas anteriormente juegan un papel importante en el desarrollo de los modelos usados para caracterizar la volatilidad.

5.1. Estructura de los Modelos

Como se discutió en secciones anteriores, más precisamente en la sección de modelos lineales, una serie de tiempo x_t se puede escribir como la suma de dos componentes,

$$x_t = \mu_t + \varepsilon_t = \mathbb{E}(x_t | \mathcal{F}_{t-1}) + \varepsilon_t, \quad (5.1)$$

donde \mathcal{F}_{t-1} representa la información disponible hasta el instante $t - 1$. Comúnmente, \mathcal{F}_{t-1} consiste de todas las funciones lineales del pasado de x_t . Como ya hemos mencionado, el objetivo de los procesos descritos en las secciones previas era la modelización de $\mu_t = \mathbb{E}(x_t | \mathcal{F}_{t-1})$ (el momento condicional de primer orden), por supuesto, bajo la suposición de que ε_t era un proceso de ruido blanco condicionalmente homocedástico, es decir,

$$\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2.$$

Los modelos de heterocedasticidad condicional suponen que el segundo momento condicional depende del tiempo, es decir,

$$\sigma_t^2 = \text{Var}(x_t | \mathcal{F}_{t-1}) = \mathbb{E}((x_t - \mu_t)^2 | \mathcal{F}_{t-1}) = \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = h_t, \quad (5.2)$$

siendo h_t una función no negativa, $h_t = h_t(\mathcal{F}_{t-1})$. A lo largo de esta sección se discutirán algunas de las posibles representaciones de h_t . La manera en que h_t evoluciona respecto del tiempo distinguirá una representación de otra. Ya que el objetivo es el estudio de modelos que permitan caracterizar series financieras, se considerará de forma general que x_t representa la serie de rentabilidades de un activo. Así mismo, haremos referencia de ε_t como la *rentabilidad corregida en media o impulso* del activo.

Los modelos de heterocedasticidad condicional se pueden clasificar en dos categorías generales. La primera categoría, agrupa aquellos modelos que usan una función exacta que rige la evolución de $\sigma_t^2 = h_t$, mientras que una segunda categoría, agrupa aquellos modelos que usan una ecuación estocástica para describir $\sigma_t^2 = h_t$. Los modelos *GARCH* pertenecen al primer grupo, mientras que los modelos de volatilidad estocástica están en la segunda categoría.

5.2. Modelos ARCH

El primer modelo que proporciona un enfoque sistemático para la modelización de la volatilidad es el modelo *Autorregresivo de Heterocedasticidad Condicional* denotado por sus siglas en inglés ARCH (Autorregresive Conditional Heteroscedasticity), introducido por Engle (1982). Un modelo ARCH(p) asume la forma

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2.\end{aligned}\quad (5.3)$$

Donde $\{\vartheta_t\}$ es una sucesión de variables independientes e idénticamente distribuidas con media 0 y varianza 1, $\alpha_0 > 0$ y $\alpha_i \geq 0$, $i \in \{1, \dots, p\}$. La condición de no negatividad sobre los coeficientes α_i garantiza que la varianza condicional h_t sea positiva.

Observación 5.2.1. Algunos autores usan σ_t^2 para denotar la varianza condicional en la ecuación (5.3) en lugar de h_t tal como se ha denotado aquí. Así pues, el modelo ARCH(p) también se puede escribir de la siguiente manera:

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sigma_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2.\end{aligned}$$

Sin embargo, en lo que sigue y por razones prácticas, se usará la primera notación descrita en la ecuación (5.3).

El modelo ARCH(p) se puede escribir como un modelo AR(p) para ε_t^2 . En efecto,

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \eta_t, \quad (5.4)$$

donde $\eta_t \equiv \varepsilon_t^2 - h_t$. Teniendo en cuenta la teoría de los modelos autorregresivos, si las raíces de la ecuación característica del proceso AR están fuera del círculo unitario, entonces el proceso es estacionario y además se puede calcular la varianza incondicional de ε_t , $\text{Var}(\varepsilon_t)$ como

$$\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \dots - \alpha_p},$$

siempre y cuando el denominador sea distinto del origen y positivo. Teniendo en cuenta la expresión (5.3), se puede ver la razón por la cual los

modelos *ARCH* pueden describir el agrupamiento de la volatilidad. El modelo establece que la varianza condicional h_t es una función creciente de ε_{t-i}^2 para $i \in \{1, \dots, p\}$. Por lo tanto, valores grandes de ε_{t-i} (en módulo) dan lugar a valores grandes de h_t . En consecuencia, ε_t también tiende a asumir valores grandes (en módulo).

Además de capturar el agrupamiento de la volatilidad, los modelos *ARCH* también reflejan el exceso de kurtosis típico de las series de rentabilidades. Para estudiar esta y otras propiedades, consideramos por simplicidad el modelo *ARCH*(1), que asume la forma siguiente:

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.\end{aligned}\tag{5.5}$$

Entonces, se tiene que

$$\mathbb{E}(\varepsilon_t) = \mathbb{E}[\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})] = \mathbb{E}(\sqrt{h_t} \mathbb{E}(\vartheta_t)) = 0.$$

Por otra parte, bajo la suposición de estacionariedad, la varianza incondicional de ε_t es

$$\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2) = \frac{\alpha_0}{1 - \alpha_1},$$

siempre y cuando $0 \leq \alpha_1 < 1$. Suponiendo normalidad en ϑ_t ,

$$\mathbb{E}(\varepsilon_t^4 | \mathcal{F}_{t-1}) = 3(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2,$$

y por lo tanto

$$\mathbb{E}(\varepsilon_t^4) = \mathbb{E}[\mathbb{E}(\varepsilon_t^4 | \mathcal{F}_{t-1})] = 3\mathbb{E}(\alpha_0^2 + 2\alpha_0\alpha_1\varepsilon_{t-1}^2 + \alpha_1^2\varepsilon_{t-1}^4).$$

Entonces, si ε_t es estacionario de cuarto orden con $\mu_4 = \mathbb{E}(\varepsilon_t^4)$, tenemos

$$\mu_4 = 3(\alpha_0^2 + 2\alpha_0\alpha_1\text{Var}(\varepsilon_t) + \alpha_1^2\mu_4) = 3\alpha_0^2 \left(1 + 2\frac{\alpha_1}{1 - \alpha_1}\right) + 3\alpha_1^2\mu_4.$$

Consecuentemente,

$$\mu_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}.$$

Imponiendo la condición $0 \leq \alpha_1^2 < \frac{1}{3}$ para que el momento de cuarto orden sea positivo. Por otra parte, la kurtosis incondicional de ε_t es

$$K = \frac{\mathbb{E}(\varepsilon_t^4)}{[\text{Var}(\varepsilon_t)]^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3.$$

En esta última expresión se ve reflejado el exceso de kurtosis de ε_t .

El modelo *ARCH* tiene múltiples propiedades que en cierta forma pueden mejorar la modelización de series financieras, en especial si se quiere modelizar la volatilidad. Sin embargo, este modelo como los ya discutidos presentan limitaciones a la hora de modelizar series de rentabilidades de activos financieros. Es habitual que períodos de rentabilidades negativas sean el prelude de períodos de gran volatilidad. Pues bien, los modelos *ARCH* no tienen la capacidad de captar esta característica debido a que la volatilidad responde igualmente ante impulsos negativos o positivos, pues dependen del cuadrado de los mismos.

Por otro lado, las condiciones para la existencia de momentos de orden mayor, implica colocar restricciones muy estrictas sobre los parámetros del modelo. Como ya se mencionó, para un modelo *ARCH*(1) con momento de cuarto orden finito se exige que $0 \leq \alpha_1^2 < \frac{1}{3}$, de manera que para un modelo *ARCH* de mayor orden las restricciones tienden a complicarse.

Estimación de un Modelo *ARCH*(p)

Los estimadores que con mayor frecuencia se usan para estimar los modelos *ARCH* son los que se derivan de la función de máxima verosimilitud Gaussiana (condicional). Considérese el modelo *ARCH*(p) de la ecuación (5.3). Bajo la hipótesis de normalidad de ϑ_t , se tiene que la distribución de ε_t condicionado a \mathcal{F}_{t-1} , la información disponible hasta el instante $t - 1$ es una distribución normal con media cero y varianza h_t , y la función de máxima verosimilitud adopta la forma:

$$\begin{aligned} f(\varepsilon_1, \dots, \varepsilon_T | \boldsymbol{\alpha}) &= f(\varepsilon_T | \mathcal{F}_{T-1}) f(\varepsilon_{T-1} | \mathcal{F}_{T-2}) \dots f(\varepsilon_{p+1} | \mathcal{F}_p) f(\varepsilon_1, \dots, \varepsilon_p | \boldsymbol{\alpha}) \\ &= \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-\varepsilon_t^2}{2h_t}\right) f(\varepsilon_1, \dots, \varepsilon_p | \boldsymbol{\alpha}), \end{aligned} \quad (5.6)$$

donde $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p)'$ y $f(\varepsilon_1, \dots, \varepsilon_p | \boldsymbol{\alpha})$ es la función de densidad conjunta de $\varepsilon_1, \dots, \varepsilon_p$. Puesto que la forma exacta de esta densidad es

difícil de calcular, se suele considerar la función de verosimilitud condicionada siguiente:

$$f(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \varepsilon_1, \dots, \varepsilon_p) = \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-\varepsilon_t^2}{2h_t}\right). \quad (5.7)$$

Puesto que maximizar la función de verosimilitud (5.7) equivale a maximizar su logaritmo, se considera entonces

$$\ell(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \varepsilon_1, \dots, \varepsilon_p) = - \sum_{t=p+1}^T \frac{1}{2} \left[\ln(2\pi) + \ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right],$$

donde ℓ denota el logaritmo de la función de verosimilitud, es decir,

$$\ell(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \varepsilon_1, \dots, \varepsilon_p) = \ln f(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \varepsilon_1, \dots, \varepsilon_p).$$

Ahora bien, ya que el primer término de la suma no depende de ningún parámetro, los estimadores de máxima verosimilitud condicional de α_0 y $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ se obtienen maximizando

$$\ell(\alpha_0, \boldsymbol{\alpha}) = - \sum_{t=p+1}^T \frac{1}{2} \left[\ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right],$$

donde $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2$ se puede evaluar de forma recursiva.

Como se discutirá más adelante, en ciertas ocasiones y dependiendo de los datos en análisis, será apropiado suponer que ϑ_t sigue una distribución con colas más pesadas que la distribución normal, por ejemplo, una distribución t -Student.

Predicción

De forma general, sean x_1, \dots, x_h observaciones de una serie de tiempo $\{x_t\}$. Se quiere entonces predecir la observación x_{h+k} , para algún horizonte de predicción $k > 0$, a partir de las observaciones disponibles. Se denota por $\hat{x}_h(k)$ el estimador de mínimos cuadrados de $x_h(k)$, esto es

$$\hat{x}_h(k) = \arg \inf_f \mathbb{E}(x_{h+k} - f)^2,$$

donde el ínfimo se considera sobre las funciones medibles de x_1, \dots, x_h . Se puede ver que bajo estas condiciones,

$$\hat{x}_k(h) = \mathbb{E}(x_{h+k} | x_1, \dots, x_h).$$

Teniendo en cuenta este resultado, las predicciones de h_t en los modelos ARCH se obtienen de forma recursiva a partir de la expresión de la volatilidad dada por (5.3), bajo el supuesto de que conocemos la serie hasta el instante t , la predicción de h_{t+1} viene dada por

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 \varepsilon_t^2 + \dots + \alpha_p \varepsilon_{t+1-p}^2.$$

Ahora bien, tomando $\hat{\varepsilon}_t^2(1) = \hat{h}_t(1)$, la predicción a dos pasos es:

$$\begin{aligned} \hat{h}_t(2) &= \alpha_0 + \alpha_1 \hat{\varepsilon}_t^2(1) + \alpha_2 \varepsilon_t^2 + \dots + \alpha_p \varepsilon_{t+2-p}^2, \\ &= \alpha_0 + \alpha_1 \hat{h}_t(1) + \alpha_2 \varepsilon_t^2 + \dots + \alpha_p \varepsilon_{t+2-p}^2. \end{aligned}$$

En general para la predicción a k pasos es,

$$\hat{h}_t(k) = \alpha_0 + \alpha_1 \hat{h}_t(k-1) + \dots + \alpha_p \hat{h}_t(k-p),$$

donde $\hat{h}_t(k-i) = \varepsilon_{t+k-i}^2$ para $i \in \{1, \dots, p\}$ si $k-i \leq 0$.

5.3. Modelos GARCH

Cuando se consideran modelos ARCH para caracterizar el comportamiento dinámico de la volatilidad suele ocurrir que el orden p del modelo sea significativamente alto. Esto conlleva a que las restricciones sobre los parámetros, para garantizar la no negatividad de la varianza y la no estacionariedad del proceso, sean muy fuertes. Bollerslev (1986) propuso como solución alternativa los *modelos generalizados autorregresivos de heterocedasticidad condicional*, y denotados por sus siglas en inglés GARCH, (Generalized ARCH). Para estos modelos, la varianza condicional en un instante depende, no solo de los valores pasados de los impulsos al cuadrado sino también de sus propios retardos. Así, un modelo GARCH(p, q) se expresa mediante la ecuación

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \end{aligned} \quad (5.8)$$

donde ϑ_t es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza 1, independientes de $\{\varepsilon_{t-k}, k \geq 1\}$ para todo t . La no negatividad de la varianza condicional h_t queda garantizada exigiendo que $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$.

Se puede demostrar que el proceso es estrictamente estacionario con $\mathbb{E}(\varepsilon_t^2) < \infty$ si y solamente si $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ (véase Fan and Yao (2003)). En este caso, $\mathbb{E}(\varepsilon_t) = 0$ y la varianza incondicional de ε_t toma el valor

$$\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2) = \frac{\alpha_0}{1 - (\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j)}.$$

En efecto, bajo la suposición de estacionariedad se tiene

$$\begin{aligned} \mathbb{E}(\varepsilon_t^2) &= \mathbb{E}(h_t) = \alpha_0 + \sum_{i=1}^p \alpha_i \mathbb{E}(\varepsilon_{t-i}^2) + \sum_{j=1}^q \beta_j \mathbb{E}(h_{t-j}) \\ &= \alpha_0 + \sum_{i=1}^p \alpha_i \mathbb{E}(\varepsilon_t^2) + \sum_{j=1}^q \beta_j \mathbb{E}(\varepsilon_t^2), \end{aligned}$$

y despejando $\mathbb{E}(\varepsilon_t^2)$ se obtiene la expresión dada para la varianza incondicional.

Por otro lado, el modelo $GARCH(p, q)$ puede representarse alternativamente mediante un modelo $ARMA(m, q)$ para ε_t^2 , siendo $m = \max(p, q)$. Sea $\eta_t = \varepsilon_t^2 - h_t$. Sustituyendo $h_t = \varepsilon_t^2 - \eta_t$ en (5.8) se tiene

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \eta_t - \sum_{j=1}^q \beta_j \eta_{t-j}, \tag{5.9}$$

donde $\alpha_{p+j} = \beta_{q+j} = 0$ para $j \geq 1$.

Observación 5.3.1. Se puede demostrar que el modelo $GARCH$ definido como un modelo $ARMA$ satisface $\mathbb{E}(\eta_t) = 0$ y $Cov(\eta_t, \eta_{t-j}) = 0$ para $j \geq 1$. Sin embargo $\{\eta_j\}$ no es, en general, una sucesión de variables independientes e idénticamente distribuidas.

Similar a las exposiciones anteriores de los modelos de series de tiempo, se considerará, por simplicidad, el estudio de las propiedades para

el modelo $GARCH(1,1)$. Considérese el modelo $GARCH(1,1)$

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}.\end{aligned}\tag{5.10}$$

En primer lugar se verá como se puede reducir el número de retardos de los impulsos al cuadrado al hacer depender la volatilidad de sus valores pasados. El modelo (5.10) se puede escribir

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-2}),$$

y de manera recursiva se obtiene

$$h_t = \alpha_0 \sum_{i=0}^{\infty} \beta_1^i + \alpha_1 \sum_{i=0}^{\infty} \beta_1^{i-1} \varepsilon_{t-i}^2.$$

Esto quiere decir, que el modelo $GARCH(1,1)$ admite una representación $ARCH(\infty)$. De (5.10) se deduce que valores grandes de ε_t^2 o h_{t-1} dan lugar a valores grandes de h_t . Esto significa que grandes valores de ε_{t-1}^2 tienden a ir seguidos de grandes valores de ε_t^2 , dando lugar así al característico agrupamiento de la volatilidad de las series financieras. Por otra parte, si se imponen condiciones de existencia del momento de cuarto orden de ε_t , y suponiendo normalidad en ϑ_t , se tiene

$$\mathbb{E}(\varepsilon_t^4) = 3\mathbb{E}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1})^2.$$

Desarrollando la expresión anterior y despejando μ_4 , se obtiene

$$\mu_4 = \frac{3\alpha_0^2(1 + 2(\alpha_1 + \beta_1))}{(1 - (\alpha_1 + \beta_1))(1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2)}.$$

Por lo tanto, la kurtosis de un proceso $GARCH(1,1)$ será

$$K = \frac{\mathbb{E}(\varepsilon_t^4)}{[\mathbb{E}(\varepsilon_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.\tag{5.11}$$

En consecuencia, $1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 > 0$, de la expresión anterior se deduce que la distribución de las colas de un proceso $GARCH(1,1)$ es más pesada que la de una distribución normal. Por lo tanto, el modelo también es capaz de reflejar el exceso de kurtosis característico de

las series de rentabilidades. Sin embargo, similar a lo que ocurre con los modelos *ARCH*, los modelos *GARCH* responden de igual forma ante los impulsos independientemente de su signo y por esta razón no pueden reflejar los efectos asimétricos de las rentabilidades negativas o positivas.

Estimación de un Modelo *GARCH*

De manera similar al procedimiento realizado para el modelo *ARCH* y bajo la suposición de que ϑ_t sigue una distribución normal, se obtiene una expresión para la función de verosimilitud Gaussiana (condicional) dada por:

$$\ell(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \boldsymbol{\beta}, \varepsilon_1, \dots, \varepsilon_p) = - \sum_{t=p+1}^T \frac{1}{2} \left[cte. + \ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right],$$

pero en este caso $h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$. En esta situación la varianza condicional h_t no se puede expresar en términos de un número finito de observaciones pasadas de ε_t , tal como sucede con los modelos *ARCH*. Como solución a este inconveniente, se sustituye en la función de verosimilitud a h_t por una versión truncada de la misma, \tilde{h}_t . Los estimadores de máxima verosimilitud condicional de α_0 , $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$ y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ se obtienen al maximizar la función

$$\ell(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}) = - \sum_{t=\kappa}^T \frac{1}{2} \left[\ln(\tilde{h}_t) + \frac{\varepsilon_t^2}{\tilde{h}_t} \right],$$

donde κ es un entero ($\kappa > p$).

Hasta ahora se ha asumido que las innovaciones ϑ_t siguen una distribución normal, lo cual implica que la distribución condicional de ε_t es una normal con media cero y varianza h_t . Por otro lado, la distribución incondicional de ε_t en los modelos *GARCH* no es una normal. Por ejemplo, para un modelo *GARCH*(1,1), la kurtosis y la kurtosis incondicional de ε_t es mayor que la kurtosis de una normal.

En aplicaciones prácticas en series de tiempo de alta frecuencia, se suele observar que la kurtosis incondicional de ε_t en un *GARCH*(1,1), dada por (5.11), es menor que la de la serie observada, es decir, el modelo *GARCH*(1,1) con ϑ_t con distribución normal, no tienen la capacidad de

recoger todo el peso de las colas de la distribución incondicional de ε_t . Para solucionar este inconveniente se han desarrollado nuevos modelos, aunque una de las alternativas más comunes consiste en considerar otra distribución para ϑ_t .

Una distribución que comúnmente se considera para ϑ , es la t -Student estandarizada con ν grados de libertad, motivado a sus propiedades relacionadas con las colas pesadas. Sea

$$f(\vartheta_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{\vartheta_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}},$$

con $\nu > 2$, donde $\Gamma(\cdot)$ es la función Gamma.

La función de verosimilitud en este caso se expresa por:

$$f(\varepsilon_{p+1}, \dots, \varepsilon_T | \boldsymbol{\alpha}, \boldsymbol{\beta}, \varepsilon_1, \dots, \varepsilon_p) = \prod_{t=p+1}^T \frac{1}{\sqrt{h_t}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{\vartheta_t^2}{h_t(\nu-2)}\right)^{-\frac{\nu+1}{2}}.$$

Como antes, maximizar f es equivalente a maximizar el $\ln f(\cdot) = \ell(\cdot)$,

$$\ell(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) = - \sum_{t=p+1}^T \frac{1}{2} \left[\ln(h_t) - \ln\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right) + \ln(\pi(\nu-2)) + (\nu+1) \ln\left(1 + \frac{\vartheta_t^2}{h_t(\nu-2)}\right) \right].$$

En este caso, la kurtosis incondicional de ε_t es mayor que la obtenida cuando se asume que la distribución de ϑ_t es normal.

Observación 5.3.2. Los grados de libertad de la distribución t -Student pueden ser especificados a priori o se pueden estimar con el resto de parámetros a partir de la función de verosimilitud.

Predicción

Tal como se discutió para los modelos ARCH, de forma análoga se puede calcular de manera recursiva las predicciones para la varianza condicional en los procesos GARCH. Sea el modelo GARCH(1,1) con ecua-

ción (5.10), la esperanza condicional de h_{t+k} , es teóricamente el estimador óptimo para la predicción de la varianza condicional, y se determina por

$$\hat{h}_t(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}_t^2(k-1) + \beta_1 \hat{h}_t(k-1),$$

donde $\hat{\varepsilon}_t(k-1) = \hat{h}_t(k-1)$ si $k-1 > 0$, mientras que $\hat{\varepsilon}_t(k-1) = \varepsilon_{t+k-1}^2$ y $\hat{h}_t(k-1) = h_{t+k-1}$ si $k-1 \leq 0$. Sustituyendo de forma recursiva,

$$\hat{h}_t(k) = \alpha_0 \sum_{j=0}^{k-2} (\alpha_1 + \beta_1)^j + (\alpha_1 + \beta_1)^{k-1} h_{t+1},$$

donde, en este caso, h_{t+1} se calcula directamente a partir de la serie histórica (conocidos los parámetros del modelo).

Si el proceso es estacionario con $\alpha_1 + \beta_1 < 1$, mediante un simple cálculo, se obtiene

$$\hat{h}_t(k) = \sigma_\varepsilon^2 + (\alpha_1 + \beta_1)^{k-1} (h_{t+1} - \sigma_\varepsilon^2),$$

donde, $\sigma_\varepsilon^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$ es la varianza incondicional de ε_t .

5.4. Modelos EGARCH

Como ya se ha comentado antes, en el caso de las serie de rentabilidades de activos financieros los períodos de gran volatilidad suelen presentarse como respuesta a grandes valores negativos de los impulsos. Esto sugiere que impulsos negativos o positivos pueden tener un impacto asimétrico sobre la varianza condicional. En este sentido, se han encontrado evidencias de que la rentabilidad del precio de un activo presenta correlación negativa con los cambios en la volatilidad, es decir, la volatilidad tiende a aumentar como respuesta a “malas noticias” (grandes rentabilidades menos de lo esperado) y disminuye como respuesta a las “malas noticias” (grandes rentabilidades mayores de lo esperado).

Uno de los modelos en el que impulsos negativos o positivos afectan de diferente manera la varianza condicional es el *modelo exponencial GARCH*, denotado por *EGARCH*, el cual fue desarrollado por Nelson

(1991). La formulación general del modelo es

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ \ln(h_t) &= \alpha_t + \sum_{k=1}^{\infty} \beta_k g(\vartheta_{t-k}), \quad \beta_1 = 1,\end{aligned}\tag{5.12}$$

donde $\{\alpha_t\}_{t=-\infty}^{\infty}$ y $\{\beta_k\}_{k=1}^{\infty}$ son sucesiones de números reales no estocásticos y $g(\vartheta_t)$ es una función que responde al signo y a la magnitud de ϑ_t . La función g considerada en esta sección es

$$g(\vartheta_t) = \theta \vartheta_t + \gamma [|\vartheta_t| - \mathbb{E}|\vartheta_t|].\tag{5.13}$$

Como primera observación a tener en cuenta, es que al trabajar con $\ln(h_t)$ se asegura de forma natural la no negatividad de la varianza condicional, independientemente de los parámetros seleccionados. Como se pudo constatar en los modelos *GARCH*, la condición de no negatividad de la varianza condicional imponía restricciones sobre los coeficientes del modelo y tales restricciones ocasionaban dificultades en la estimación. Por otra parte, $\{g(\vartheta_t)\}$ es una sucesión de variables aleatorias independientes y con idéntica distribución. Los parámetros θ y γ determinan el efecto del signo y de la magnitud de los impulsos sobre la volatilidad. En efecto, para $0 < \vartheta_t < \infty$, se tiene que $g(\vartheta_t)$ es lineal en ϑ_t con pendiente $\theta + \gamma$ y para $-\infty < \vartheta_t < 0$, la función $g(\vartheta_t)$ es lineal en ϑ_t con pendiente $\theta - \gamma$. Por lo tanto, $g(\vartheta_t)$ permite que la varianza condicional responda de forma asimétrica ante caídas o subidas del precio de los activos. Otra consideración a tener en cuenta es que en el modelo *EGARCH*, $\ln(h_t)$ se define como un proceso lineal y por ello las condiciones para garantizar su estacionariedad y ergodicidad son fáciles de obtener. Entonces, en las condiciones anteriores, el proceso $\{\ln(h_t) - \alpha_t\}$ es estacionario si y solamente si $\sum_{k=1}^{\infty} \beta_k^2 < \infty$ (véase Nelson (1991)).

Una expresión más simple para $\ln(h_t)$ se obtiene considerando un proceso *ARMA* en lugar de la representación *MA*(∞) dada en (5.12). Así, un *EGARCH*(p, q) vendría dado por

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ \ln(h_t) &= \alpha_t + \frac{1 + a_1 L + \dots + a_q L^q}{1 - b_1 L - \dots - b_p L^p} g(\vartheta_{t-k}).\end{aligned}\tag{5.14}$$

Atendiendo a la teoría de los modelos *ARMA*, el proceso $\{\ln(h_t) - \alpha_t\}$ será estacionario si las raíces de $1 - b_1L - \dots - b_pL^p = 0$ están fuera del círculo unitario. El siguiente paso es estudiar la no estacionariedad de $\{h_t\}$ y $\{\varepsilon_t\}$. En Nelson (1991) se demuestra que $\{\exp(-\alpha_t)h_t\}$ y $\{\exp(\frac{-\alpha_t}{2})\varepsilon_t\}$ son estrictamente estacionarios y ergódicos si $\sum_{k=1}^{\infty} \beta_k^2 < \infty$, y además se discute acerca de que la estacionariedad estricta en este caso no implica necesariamente la estacionariedad en sentido débil, puesto que los momentos incondicionales pueden no ser finitos dependiendo de la distribución de $\{\vartheta_t\}$. Sin embargo, para determinadas distribuciones como la Normal, $\{h_t\}$ y $\{\varepsilon_t\}$ tienen momentos incondicionales finitos.

Considerando el caso particular, un *EGARCH*(0, 1) con $\alpha_t = \alpha_0$ constante, dado por la ecuación

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ \ln(h_t) &= \alpha_0 + g(\vartheta_{t-1}) - b_1 \ln(h_{t-1}). \end{aligned} \tag{5.15}$$

Sustituyendo $g(\vartheta_{t-1})$ por su valor se tiene

$$\ln(h_t) = a_0 + \theta \vartheta_{t-1} + \gamma [|\vartheta_{t-1}| - \mathbb{E}|\vartheta_{t-1}|] - b_1 \ln(h_{t-1}).$$

Si $\{\vartheta_t\}$ tiene distribución Normal de media cero y varianza 1, entonces $\mathbb{E}|\vartheta_t| = \sqrt{\frac{2}{\pi}}$.

Uno de los métodos utilizados para estudiar la forma en la que impulsos negativos o positivos afectan a la volatilidad en un determinado modelo es el cálculo de la llamada *NIC* (*news impact curve*), que muestra la relación entre el impulso actual ϑ_t y la volatilidad un instante adelante h_{t+1} , manteniendo constante todo el pasado. Según esto, para un modelo *GARCH*(1, 1) como el (5.10) se tiene

$$NIC(\varepsilon_t | h_t = \sigma^2) = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma^2 = A + \alpha_1 \varepsilon_t^2,$$

siendo $A = \alpha_0 + \beta_1 \sigma^2$. Por lo tanto, la *NIC* de un *GARCH*(1, 1) es una función cuadrática centrada en $\varepsilon_t = 0$. Cambios en la varianza condicional h_t únicamente trasladan la curva verticalmente, pero no afectan su forma. Para un *EGARCH*(0, 1) definido en (5.15), se tiene

$$NIC(\varepsilon_t | h_t = \sigma^2) = \begin{cases} A \exp\left(\frac{\theta + \gamma}{\sigma} \varepsilon_t\right), & \text{si } \varepsilon_t > 0, \\ A \exp\left(\frac{\theta - \gamma}{\sigma} \varepsilon_t\right), & \text{si } \varepsilon_t < 0. \end{cases}$$

Siendo $A = \sigma^{2b_1} \exp(a_0 - \gamma\sqrt{2/\pi})$.

5.5. Modelos IGARCH

Cuando se aplica el modelo $GARCH(1, 1)$ a series de tiempo de frecuencia alta, los valores estimados α_1 y β_1 suelen ser tales que su suma es próxima o igual a 1. Esto sugiere la existencia de una raíz unitaria en la ecuación de volatilidad. En esta situación se obtiene un *modelo GARCH integrado*, el cual se denotará por *IGARCH*. El modelo que se obtiene cuando $\alpha_1 + \beta_1 = 1$ se denomina $IGARCH(1, 1)$ y su representación es dada por

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + (1 - \beta_1)\varepsilon_{t-1}^2 + \beta_1 h_{t-1}.\end{aligned}\tag{5.16}$$

donde ϑ_t es definido como antes. En estas condiciones, la varianza incondicional de ε_t no es finita y, por lo tanto, el modelo $IGARCH(1, 1)$ es no estacionario en sentido débil. Sin embargo, el proceso es estrictamente estacionario. Consideremos la ecuación de la volatilidad de un modelo $GARCH(1, 1)$ dada por (5.10), entonces

$$\begin{aligned}h_t &= \alpha_0 + (\alpha_1 \vartheta_{t-1}^2 + \beta_1) h_{t-1} \\ &= \alpha_0 (1 + (\alpha_1 \vartheta_{t-1}^2 + \beta_1)) + (\alpha_1 \vartheta_{t-1}^2 + \beta_1) (\alpha_1 \vartheta_{t-2}^2 + \beta_1) h_{t-2},\end{aligned}$$

y al sustituir de manera recursiva se obtiene

$$h_t = \alpha_0 \left(1 + \sum_{i=1}^{t-1} \prod_{j=1}^i (\alpha_1 \vartheta_{t-j}^2 + \beta_1) \right) + \prod_{i=1}^t (\alpha_1 \vartheta_{t-i}^2 + \beta_1) h_0.$$

Por lo tanto, el efecto de h_0 sobre la volatilidad en el horizonte t viene dado por

$$\prod_{i=1}^t (\alpha_1 \vartheta_{t-i}^2 + \beta_1) h_0.$$

Este efecto puede disiparse rápidamente o persistir. Teniendo en cuenta que las variables ϑ_t son independientes y $\mathbb{E}(\vartheta_t^2) = 1$,

$$\mathbb{E} \left(\prod_{i=1}^t (\alpha_1 \vartheta_{t-i}^2 + \beta_1) h_0 \right) = (\alpha_1 - \beta_1)^t h_0.$$

De donde se deduce que el efecto de h_0 se desvanece asintóticamente si $\alpha_1 + \beta_1 < 1$. Además, como ya se ha comentado, esta condición garantiza la existencia de la varianza incondicional del proceso. Por otro lado, aplicando la ley de los grandes números:

$$\begin{aligned} \prod_{i=1}^t (\alpha_1 \vartheta_{t-i}^2 + \beta_1) h_0 &= \exp \sum_{i=1}^t \ln(\alpha_1 \vartheta_{t-i}^2 + \beta_1) \\ &\approx \exp \{ t \mathbb{E} (\ln(\alpha_1 \vartheta_{t-i}^2 + \beta_1)) \}. \end{aligned}$$

De esta última expresión se deduce que el impacto de h_0 desaparece cuando se verifica

$$\mathbb{E}(\ln(\alpha_1 \vartheta_t^2 + \beta_1)) < 0.$$

Se puede demostrar que la condición anterior garantiza la estacionariedad estricta del proceso $GARCH(1, 1)$. La relación entre la condición de estacionariedad débil y la estacionariedad estricta se obtiene aplicando la desigualdad de Jensen,

$$\mathbb{E}(\ln(\alpha_1 \vartheta_t^2 + \beta_1)) < \ln(\mathbb{E}(\alpha_1 \vartheta_t^2 + \beta_1)) = \ln(\alpha_1 + \beta_1). \quad (5.17)$$

Así pues, si el proceso es débilmente estacionario, también será estrictamente estacionario. La desigualdad (5.17) garantiza la estacionariedad estricta del proceso $IGARCH(1, 1)$, en el que $\alpha_1 + \beta_1 = 1$. Aún así, las distribuciones marginales presentan colas tan pesadas que la varianza condicional no existe.

5.6. Modelos $GARCH-M$

En finanzas, el rendimiento de un activo puede depender de su volatilidad. Más aún, existe una relación directa entre el rendimiento esperado y el riesgo de un activo financiero. Los llamados *modelos GARCH en media*, denotados por $GARCH-M$, han sido diseñados para modelizar este tipo de fenómenos o relaciones. La forma de hacerlo es incluyendo en el modelo de la media condicional una función de la varianza condicional h_t . Así, tenemos el modelo

$$\begin{aligned} x_t &= g(h_t, \xi) + \varepsilon_t, \\ \varepsilon_t &= \vartheta_t \sqrt{h_t}. \end{aligned} \quad (5.18)$$

Donde se asume además que la varianza condicional h_t sigue un modelo *GARCH*. En la práctica, es habitual seleccionar la función $g(h_t, \xi) = \zeta_0 + \zeta_1 h_t$, que se interpreta como una especie de prima de riesgo. En este sentido, los coeficientes ζ_0 y ζ_1 son constantes. Un valor positivo de ζ_1 indica que el rendimiento está relacionado de manera positiva a su volatilidad.

Una formulación simple del modelo *GARCH-M* es el modelo *GARCH(1, 1)* en media o bien *GARCH(1, 1) - M* dado por

$$\begin{aligned} x_t &= \zeta_0 + \zeta_1 h_t + \varepsilon_t, \\ \varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \end{aligned} \tag{5.19}$$

5.7. Modelos *TGARCH*

Al inicio de esta sección se mencionaron algunas características presentes en la volatilidad. Una de ellas era el efecto palanca o apalancamiento. Pues bien, un modelo que con frecuencia se usa para manejar este efecto es el *modelo GARCH con umbral* y que se denota por *T-GARCH* (*Threshold GARCH*) (véanse Glosten et al. (1993), y Zakoian (1994)). Un modelo *T-GARCH(p, q)* asume la forma

$$h_t = \alpha_0 + \sum_{i=1}^p (\alpha_i + \gamma_i \mathbb{1}_{\{\varepsilon_{t-i} < 0\}}) \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \tag{5.20}$$

donde $\mathbb{1}_{\{\varepsilon_{t-i} < 0\}}$ es una variable indicadora para valores negativos de ε_{t-i} , es decir,

$$\mathbb{1}_{\{\varepsilon_{t-i} < 0\}} = \begin{cases} 1, & \text{si } \varepsilon_{t-i} < 0; \\ 0, & \text{si } \varepsilon_{t-i} \geq 0, \end{cases}$$

y α_i , γ_i , y β_j son parámetros no negativos que satisfacen condiciones similares a las de los modelos *GARCH*. El modelo hace uso del valor cero como su *umbral* para separar el impacto de los valores del pasado. Sin embargo, no es el único valor que se puede aplicar, otros valores dependiendo de los datos pueden ser útiles.

5.8. Modelos de Volatilidad Estocástica

Una manera alternativa de caracterizar la evolución de la volatilidad consiste en no suponer que esta es una función determinista de la información pasada, tal y como sucede al modelizar mediante la familia *ARCH-GARCH*. La volatilidad también puede depender de factores no predecibles, tales como decisiones políticas, cambios de estrategia de las empresas, etc., con lo que la volatilidad debería de incluir componentes aleatorios que permitan captar estas características. El *modelo de volatilidad estocástica (SV)*, introducido por Taylor (1994) recoge este hecho suponiendo que la volatilidad depende de una variable no observable que se rige por un proceso estocástico *ARMA* y de una variable aleatoria independiente e idénticamente distribuida. Este modelo asume la forma

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ \alpha_p(L) \ln(h_t) &= \alpha_0 + \eta_t,\end{aligned}\tag{5.21}$$

donde $\alpha_p(L) = (1 - \alpha_1 L - \dots - \alpha_p L^p)$ es el polinomio característico asociado al proceso con todas sus soluciones mayores que 1 en módulo, α_0 es una constante, η_t son variables aleatorias independientes e idénticamente distribuidas $N(0, \sigma_\eta^2)$, $\{\vartheta_t\}$ y $\{\eta_t\}$ son independientes.

Para introducir una mayor persistencia temporal en la volatilidad se puede suponer que el componente observable sigue un modelo *ARIMA* o *ARFIMA* en lugar de un modelo *ARMA*.

Estos modelos permiten una aproximación discreta de los modelos en tiempo continuo, lo que los hace bastante atractivos en el contexto del análisis financiero. Sin embargo, a diferencia de los modelos *ARCH* su estimación resulta muy complicada, ya que no es posible conocer la forma exacta de la función de verosimilitud asociada a la estimación, haciéndose necesario la aplicación de métodos alternativos como estimación casi-máxima verosímil, máxima verosimilitud simulada, inferencia indirecta o el método generalizado de los momentos. Una especificación simple del modelo (5.21) para $p = 1$ es la siguiente:

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ \ln(h_t) &= \alpha_0 + \alpha_1 \ln(h_{t-1}) + \eta_t.\end{aligned}\tag{5.22}$$

Una representación alternativa del modelo (5.22) de volatilidad estocástica es la siguiente:

$$\varepsilon_t = v \exp\left(\frac{1}{2}z_t\right) \vartheta_t, \quad z_t = \theta_1 z_{t-1} + \eta_t, \quad (5.23)$$

donde v es un parámetro de escala, ϑ_t es un proceso independiente e idénticamente distribuido con media cero y varianza uno, z_t es una variable aleatoria independiente de ϑ_t . En el modelo (5.22) la especificación para z_t es autorregresiva, sin embargo esta podría generalizarse a un modelo *ARMA*. De las expresiones anteriores podemos escribir el modelo *SV* en el espacio de estados. Simplemente linealizando el modelo obtenemos una ecuación de estado y una ecuación de transición:

$$\begin{aligned} x_t &= \mu + z_t + \zeta_t, \\ z_t &= \theta_1 z_{t-1} + \eta_t, \end{aligned}$$

donde $x_t = \log(\varepsilon_t^2)$, $\mu = \log(v + \mathbb{E}[\log(\vartheta_t^2)])$ y $\zeta_t = \log(\vartheta_t^2) - \mathbb{E}[\log(\vartheta_t^2)]$ es un ruido blanco no gaussiano de media cero y varianza σ_ζ^2 cuyas propiedades dependen de ϑ_t . Por ejemplo, si ϑ_t tiene distribución normal de media cero y varianza 1, entonces $\log(\vartheta_t^2)$ se distribuye como el logaritmo de una variable X_1^2 de media $\mathbb{E}[\log(\vartheta_t^2)] = -1,26$ y varianza conocida $\sigma_\zeta^2 = 4,93$.

Para estimar los parámetros del modelo *SV*, se puede usar el método de cuasi-verosimilitud mediante el *filtro de Kalman* considerando la ecuación de transición z_t y la ecuación de estados x_t . Una alternativa es aplicar un método de Monte Carlo. Jacquier et al. (1994) proporcionan una comparación de los resultados entre la estimación cuasi-verosimilitud y el método de Monte Carlo (*MCMC*).

5.8.1. Extensiones del Modelo *SV*

A continuación se considera una extensión par el modelo *SV*.

Modelo SV con memoria larga (LMSV):

Esta extensión permite la memoria larga en la volatilidad, aplicando la idea de la diferencia fraccional. Como ya se ha comentado, una serie de tiempo es un proceso de memoria larga si su función de autocorrelación decae hiperbólicamente en lugar de exponencialmente a medida

que se incrementan los retardos. La extensión a los modelos de memoria larga en el estudio de la volatilidad está motivado por el hecho de que la función de autocorrelación de la serie de los valores absolutos al cuadrado de la rentabilidad de un activo decae lentamente, aún si la serie no tiene correlación serial. Una representación simple del modelo de volatilidad estocástica con memoria larga (*LMSV*) es la siguiente

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \sigma^2 \exp(u_t), \\ (1 - L)^d u_t &= \eta_t,\end{aligned}\tag{5.24}$$

donde $\sigma > 0$, ϑ_t y η_t son variables aleatorias normales independientes e idénticamente distribuidas de media cero y varianzas 1 y σ_η^2 , y $0 < d < 0,5$. La propiedad de la memoria larga proviene de la diferencia fraccional $(1 - B)^d$, lo que implica que la *ACF* de u_t decae lentamente de forma hiperbólica, en lugar de forma exponencial, a medida que aumentan los retardos. Para el modelo (5.24), se tiene que

$$\begin{aligned}\ln(\varepsilon_t) &= \ln(\sigma^2) + u_t + \ln(\vartheta_t^2), \\ &= \mu + u_t + e_t,\end{aligned}$$

donde $\mu = \ln(\sigma^2) + \mathbb{E}(\ln(\vartheta_t^2))$ y $e_t = \ln(\vartheta_t^2) + \mathbb{E}(\ln(\vartheta_t^2))$. De esta manera, la serie $\ln(\varepsilon_t)$ es un serie Gaussiana con memoria larga más un ruido blanco no Gaussiano (véase Breidt et al. (1998)). En cuanto a la estimación del modelo de volatilidad estocástica con memoria larga, ésta es complicada, sin embargo el parámetro d de diferencia fraccional se puede estimar mediante el uso de un método de probabilidad de casi-máxima verosimilitud o un método de regresión.

6

Modelos No Lineales

Esta sección está dirigida a estudiar la no linealidad en los datos financieros y los modelos econométricos no lineales que son de gran utilidad en el análisis de series de tiempo financieras. Antes que nada, es necesario precisar lo que entendemos por modelo no lineal. Para ello, considérese una serie de tiempo x_t y un modelo para dicha serie. Matemáticamente, un modelo de serie de tiempo puramente estocástico es una función de una sucesión de variables aleatorias independientes e idénticamente distribuidas, esto es,

$$x_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots). \quad (6.1)$$

Si tenemos en cuenta el modelo,

$$x_t = \mu + \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad (6.2)$$

donde μ es una constante, η_i son números reales con $\eta_0 = 1$, y $\{\varepsilon_t\}$ una sucesión de variables independientes e idénticamente distribuidas (*iid*) con distribución bien definida, entonces en este caso la función f del modelo (6.1) es una función lineal de la sucesión de variables $\{\varepsilon_t\}$. Cualquier no linealidad en la función f tendrá como resultado la

nolinealidad del modelo (6.1). Sin embargo, esta forma de presentar un modelo resulta bastante general, y más si se tiene en consideración el número de parámetros involucrados.

Por ello, el modelo se representará en términos de sus momentos condicionales. Como antes, sea \mathcal{F}_{t-1} la σ -álgebra generada por la información disponible hasta el instante $t - 1$. Generalmente, \mathcal{F}_{t-1} denota la colección de combinaciones lineales de $\{x_{t-1}, x_{t-2}, \dots\}$ y $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$. Entonces, la media y varianza condicional de x_t , dada \mathcal{F}_{t-1} , son

$$\mathbb{E}(x_t | \mathcal{F}_{t-1}) = g(\mathcal{F}_{t-1}) \quad \text{y} \quad \text{Var}(x_t | \mathcal{F}_{t-1}) = h(\mathcal{F}_{t-1}),$$

donde $g(\cdot)$ y $h(\cdot)$ son funciones bien definidas, con $h(\cdot) > 0$.

De esta forma, los modelos que se considerarán se restringen a

$$x_t = g(\mathcal{F}_{t-1}) + \varepsilon_t = g(\mathcal{F}_{t-1}) + \sqrt{h(\mathcal{F}_{t-1})} \vartheta_t, \quad (6.3)$$

donde ϑ_t son los impulsos estandarizados o (innovaciones). Para series x_t lineales, en la ecuación (6.1), $g(\cdot)$ es una función lineal de los elementos de \mathcal{F}_{t-1} y $h(\cdot) = \sigma_\varepsilon^2$. Cuando la función $g(\cdot)$ es no lineal, decimos que el modelo (6.3) es *no lineal en media*. Si $h(\cdot)$ depende del tiempo, decimos que el modelo es *no lineal en varianza*. Según estas consideraciones, los modelos de heterocedasticidad condicional discutidos en secciones previas, son modelos no lineales en varianza, puesto que su varianza condicional σ_t^2 depende del tiempo, excepto para el caso de los modelos *GARCH-M*, en los que μ_t depende de σ_t^2 . Los modelos de volatilidad considerados en la sección anterior también pertenecen a la clase de modelos no lineales en varianza.

De la descomposición de Wold, una serie de tiempo débilmente estacionaria y puramente estocástica se puede expresar como una función lineal de los impulsos. En el caso de las series con volatilidad estacionaria, dichos impulsos no correlacionados son dependientes.

En los siguientes apartados se presentan algunos de los modelos no lineales en media de mayor utilidad en el análisis de series financieras (ver Priestley (1988) y Tong (1990) como referencias para el análisis de modelos no-lineales). Se entenderán por *modelos no lineales en media*, a los modelos no lineales para la esperanza condicional.

6.1. Modelos No Lineales para la Esperanza Condicional

En lo que respecta a las series de precios y rentabilidades de activos financieros, una característica a destacar es que la asimetría en la distribución de los rendimientos podría ser explicada si consideramos un modelo con diferentes *estados* o *regímenes*, de forma que el comportamiento dinámico de la serie depende del régimen en el que se encuentre. En esta sección se consideran modelos en los que los regímenes son generados por procesos estocásticos. Así pues, se pueden considerar dos situaciones. En la primera de ellas, modelos en los que los diferentes estados dependan de una variable observable. Por lo tanto, los regímenes por los que ha pasado la serie en el pasado y en el presente son conocidos con exactitud. Dentro de este grupo se encuentran los modelos *TAR* y *SETAR*. La segunda situación se corresponde con los modelos en los que diferentes estados quedan determinados por un proceso estocástico no observable. En este caso, nada puede asegurarse acerca de la ocurrencia de un determinado régimen, y por lo tanto sólo es posible la asignación de probabilidades a los distintos estados. Los modelos *Markov-Switching* pertenecen a este grupo. En cualquiera de los casos, se asume que el comportamiento de la serie en cada estado queda bien determinado por un modelo tipo *AR*, donde los parámetros autorregresivos dependen del régimen.

6.2. Modelos *TAR*

Este modelo está motivado por diversas características que comúnmente son observadas en la práctica, tales como la asimetría en la disminución y el aumento de la estructura de un proceso. Este modelo usa modelos lineales a trozos para obtener una mejor aproximación de la media condicional. Sin embargo, en comparación con el modelo lineal tradicional a trozos que permite los cambios en el modelo se produzcan en el espacio "tiempo", el *modelo Autorregresivo con Umbral*, denotado por *TAR* (*Threshold Autorregressive*), usa el espacio umbral para mejorar la aproximación lineal. Como ya se ha mencionado, en los modelos *TAR* se asume que el régimen queda determinado por una variable observable q_t . Para un modelo *TAR* con 2 estados se fija un valor c el cual se

conoce como *valor umbral*. Así pues, dependiendo de si $q_t \leq c$ o $q_t > c$, nos situaremos en uno u otro estado. Entonces, suponiendo un $AR(1)$ para cada régimen, se tiene

$$x_t = \begin{cases} \phi_{0,1} + \phi_{1,1}x_{t-1} + \varepsilon_t, & \text{si } q_{t-1} < c, \\ \phi_{0,2} + \phi_{1,2}x_{t-1} + \varepsilon_t, & \text{si } q_{t-1} \geq c, \end{cases} \quad (6.4)$$

donde se asume que $\{\varepsilon_t\}$ son variables aleatorias independientes e idénticamente distribuidas verificando la propiedades de un ruido blanco. La condición necesaria y suficiente para que el modelo (6.4) se geométricamente ergódico es que

$$\phi_{1,1} < 1, \quad \phi_{1,2} < 1 \quad \text{y} \quad \phi_{1,1}\phi_{1,2} < 1.$$

Siendo $\phi_{1,i}$ los coeficientes del modelo $AR(1)$ del régimen i (ver Petrucci and Woolford (1984) y Chen and Tsay (1991)).

6.3. Modelos *SETAR*

Los *auto-excitables/flexibles* denotados por *SETAR* son un caso particular de los modelos *TAR* en los que la variable observable que determina los distintos estados, q_t , es un retardo de la propia serie, es decir, $q_t = x_{t-d}$. Por ejemplo, si se elige $d = 1$ y un $AR(1)$ para cada estado, el modelo *SETAR* toma la forma

$$x_t = \begin{cases} \phi_{0,1} + \phi_{1,1}x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} < \gamma, \\ \phi_{0,2} + \phi_{1,2}x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} \geq \gamma. \end{cases} \quad (6.5)$$

Los modelos *SETAR* dan lugar a patrones de comportamiento muy diversos, dependiendo de los parámetros seleccionados. Un ejemplo de modelo *TAR* con 2 regímenes o estados y un modelo $AR(1)$ para cada régimen es,

$$x_t = \begin{cases} -1,5x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} < 0, \\ 0,5x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} \geq 0, \end{cases} \quad (6.6)$$

donde las ε_t son variables aleatorias independientes e idénticamente distribuidas normal estándar. En este caso, la variable umbral es x_{t-1} y el umbral es 0.

Es posible formular modelos *SETAR* de forma que, en cada régimen, la serie se rija por un modelo $AR(p)$, con $p > 1$. Un proceso x_t se dice que sigue un modelo *SETAR* con variable umbral x_{t-d} si satisface

$$x_t = \phi_{0,j} + \phi_{1,j}x_{t-1} - \dots - \phi_{p,j}x_{t-p} + \varepsilon_{t,j}, \quad \gamma_{j-1} \leq x_{t-d} < \gamma_j, \quad (6.7)$$

donde k y d son enteros positivos, $j = 1, \dots, k$, los γ_i son números reales tales que $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{k-1} < \gamma_k = \infty$, j determina el régimen, y $\{\varepsilon_{t,j}\}$ es una sucesión de variables aleatorias *iid* con media 0 y varianza σ_j^2 y son mutuamente independientes para distintos valores de j . El parámetro d es el *parámetro de retardo* y γ_j son los *valores umbrales*. Obtener propiedades para este tipo de modelos no suele ser una tarea fácil, sin embargo se pueden encontrar alguna de ellas en Tong (1990), Chan (1993), y Chan and Tsay (1998). El modelo (6.7) se puede generalizar usando una variable z_t medible respecto a \mathcal{F}_{t-1} . En este caso, se requiere que z_t sea estacionaria con función de distribución continua sobre un conjunto compacto de la recta real y z_{t-d} conocida en el instante t . Estos modelos son conocidos como *modelos de bucle abierto TAR*.

6.4. Modelos *STAR*

En los modelos *SETAR* la media condicional no es continua. Los umbrales $\{\gamma_j\}$ son los puntos de discontinuidad de la función de media condicional μ_t . Para solventar esto, se proponen los *modelos suavizados TAR* y se denotan por *STAR* (ver Chan and Tong (1986) y Teräsvirta (1994)). Una serie de tiempo x_t sigue un modelo *STAR*(p) con dos regímenes si satisface:

$$x_t = c_0 + \sum_{i=1}^p \phi_{0,i}x_{t-i} + F\left(\frac{x_{t-d} - \Delta}{s}\right) \left(c_1 + \sum_{i=1}^p \phi_{1,i}x_{t-i}\right) + \varepsilon_t, \quad (6.8)$$

donde d es el parámetro de retardo, Δ y s son los parámetros que representan la localización y la escala del modelo transición, y $F(\cdot)$ es una función suavizado de transición. En las aplicaciones, con frecuencia $F(\cdot)$ asume una de las tres formas: logística, exponencial, o función de distribución acumulada. De la ecuación (6.8), la media condicional de

un modelo *STAR* es una combinación lineal ponderada entre las ecuaciones:

$$\begin{aligned}\mu_{1t} &= c_0 + \sum_{i=1}^p \phi_{0,i} x_{t-i}, \\ \mu_{2t} &= (c_0 + c_1) + \sum_{i=1}^p (\phi_{0,i} + \phi_{1,i}) x_{t-i}.\end{aligned}$$

Los pesos son determinados en el continuo mediante $F(x_{t-d} - \Delta/s)$. Un pre-requisito para que el modelo *STAR* sea estacionario es que todos los ceros de ambos polinomios *AR* estén fuera del círculo unitario. Una ventaja del modelo *STAR* respecto del modelo *TAR* es que la función de media condicional es diferenciable. Sin embargo, la estimación de los parámetros de escala y localización es complicada, resultando en estimaciones cuyos errores estándar llegan a ser muy elevados.

6.5. Modelos Markov Switching

En los modelos *SETAR* la transición está determinada por una particular variable de los retardos, en otras palabras, se aplica un enfoque determinista para regir la transición del modelo. Hamilton (1989) propone el *modelo autorregresivo Markov-Switching*, denotado por *MSA* a partir del uso de probabilidades de cambio (*switching*) en series no-lineales (ver Tong (1983)) haciendo énfasis entre las transiciones aperiódicas de los distintos estados. El modelo *MSA*, se distingue del modelo *SETAR* porque usa una cadena de Markov oculta para regir la transición de una función de media condicional a otra. En consecuencia, este modelo está basado en un esquema estocástico. Así pues, en los modelos de Markov Switching el estado que ocurre en el instante t no se conoce, ya que el mismo está determinado por un proceso no observable, s_t . Tal proceso no observable es un proceso de Markov con dos estados. Así, una serie x_t sigue un modelo *MSA* si satisface

$$x_t = \begin{cases} \phi_{0,1} + \sum_{i=1}^p \phi_{1,i} x_{t-i} + \varepsilon_{1,t}, & \text{si } s_t = 1, \\ \phi_{0,2} + \sum_{i=1}^p \phi_{2,i} x_{t-i} + \varepsilon_{2,t}, & \text{si } s_t = 2. \end{cases} \quad (6.9)$$

Las $\{\varepsilon_{j,t}\}$ para $j \in \{1, 2\}$ son sucesiones de variables aleatorias *iid* con media cero y varianza finita e independientes una de la otra, y $p \geq 1$.

El modelo queda completamente determinado al definir las probabilidades de transición,

$$\begin{aligned} P(s_t = 1 | s_{t-1} = 1) &= p_{11}, & P(s_t = 2 | s_{t-1} = 1) &= p_{12}, \\ P(s_t = 1 | s_{t-1} = 2) &= p_{21}, & P(s_t = 2 | s_{t-1} = 2) &= p_{22}, \end{aligned}$$

donde p_{ij} denota la probabilidad de que la cadena de Markov cambie del estado i en el instante $t - 1$ al estado j en el instante t . Este modelo puede ser extendido al caso de m estados al permitir que la cadena de Markov tome m valores distintos.

En cuanto a la predicción, las predicciones se presentan como una combinación lineal de las predicciones generadas por los submodelos de los estados individuales. La estimación de un modelo *MSA* presenta mayor dificultad que la de otros modelos no-lineales, puesto que los estados no son directamente observados. Entre los autores que han estudiado métodos de estimación para el modelo *MSA* podemos destacar a Hamilton (1990) que usa algoritmos *EM*, un método estadístico que itera entre el cálculo de esperanzas y maximizaciones; y McCulloch and Tsay (1994) quienes consideran el método de Monte Carlo por Cadenas de Markov. En el último caso los autores consideran una versión generalizada del modelo (6.9) tomando como función de transición de probabilidades p_{ij} , la logística o probit, como funciones de alguna de las variables explicativas disponibles en el instante $t - 1$.

6.6. Métodos No-Paramétricos

En las aplicaciones financieras no siempre se dispone de información suficiente acerca de la estructura funcional entre la variable explicativa y la variable respuesta. Una manera de abordar este tipo de problema es mediante un enfoque no-paramétrico. Claro está, los métodos y técnicas no-paramétricas agregan diversas dificultades, entre ellas, el coste computacional de implementar, por ejemplo, las técnicas de suavizado, o bien el sobreajuste, por su alta dependencia de los datos. En la actualidad el primer inconveniente citado ha dejado, en algunos casos, de ser un problema, toda vez que la tecnología en el campo computacional se ha desarrollado de manera vertiginosa en los últimos años. Así pues,

sumado a la intensa investigación en la búsqueda de nuevos métodos eficientes y de fácil implementación en el contexto no-paramétrico, hacen posible considerar este tipo de enfoque.

El objetivo de esta sección es presentar algunos métodos no-paramétricos para aplicaciones financieras y algunos modelos no lineales que hacen uso de los métodos no-paramétricos y técnicas de suavizado. Los métodos no-paramétricos que se estudiarán en esta sección comprenden la regresión por núcleo, y la estimación de mínimos cuadrados locales.

La base fundamental de los métodos no-paramétricos radica en las *técnicas de suavizado*. Para entender este concepto, considere dos series financieras X_t e Y_t , las cuales están relacionados mediante el modelo

$$Y_t = m(X_t) + \varepsilon_t, \quad (6.10)$$

donde $\mu(\cdot)$ es una función suave desconocida y $\{\varepsilon_t\}$ es una serie de ruido blanco (obsérvese que X_t pudiera ser Y_{t-1}, Y_{t-2}, \dots , etc). El objetivo es estimar la función $m(\cdot)$ a partir de los datos. Por simplicidad, se considerará un problema sencillo, esto es, el problema de estimar $m(X_t) = \mathbb{E}(Y_t | X_t = x)$ para x , un valor particular dado. Se asumirá que se tienen y_1, y_2, \dots, y_T observaciones independientes en $X = x$. Entonces se tiene que

$$y_t = m(x) + \varepsilon_t, \quad t = 1, \dots, T.$$

Tomando el promedio de los datos, se obtiene que

$$\bar{y} = \frac{\sum_{t=1}^T y_t}{T} = m(x) + \frac{\sum_{t=1}^T \varepsilon_t}{T}.$$

Teniendo en cuenta la ley de los grandes números, el promedio de los impulsos ε_t converge a su valor esperado, es decir, al valor cero cuando T es incrementado. Por lo tanto, el promedio $\bar{y} = \left(\sum_{t=1}^T y_t\right) / T$ es una estimación consistente de $m(x)$. Sin embargo, en las series financieras no se disponen de observaciones repetidas y los datos observados son $\{(y_t, x_t)\}$ para $t = 1, \dots, T$. Si la función $m(\cdot)$ es suave, el valor de Y_t para $X_t \approx x$ da una buena aproximación de $m(x)$, mientras que si es lejano da como resultado una aproximación pobre. Así pues, un estimador natural que tiene en cuenta este tipo de situaciones es el promedio

ponderado:

$$\hat{m}(x) = \frac{1}{T} \sum_{t=1}^T \omega_t(x) y_t, \quad (6.11)$$

con pesos $\omega_t(x)$ de mayor valor para aquellos y_t cuyo x_t es más próxima a x , y por supuesto de menor valor en caso contrario. La aproximación $\hat{m}(x)$ es conocida como *promedio ponderado local* donde los pesos están determinados por, la distancia entre x_t y x , y por la asignación de los pesos en función de la distancia considerada. Como es lógico, existen diversas formas de determinar la distancia entre x_t y x , así mismo esto influirá en la manera de asignar los pesos. En las próximas secciones se discutirán algunas consideraciones que permitirán seleccionar los pesos de forma apropiada.

6.6.1. Regresión por Núcleo

En las técnicas de suavizado un método no-paramétrico de uso frecuente es la *regresión por núcleo* o *regresión kernel*. En este caso, los pesos son determinados mediante una función denominada *núcleo*, la cual es generalmente una función de densidad de probabilidades denotado por $K(\cdot)$ y que satisface

$$K(x) \geq 0, \quad \int K(u) du = 1.$$

Esta función suele ser reescalada mediante un parámetro $h > 0$, denominado *ancho de banda* o *parámetro de suavizado*, así pues la función núcleo $K(\cdot)$ queda definida por

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \quad \int K_h(u) du = 1. \quad (6.12)$$

De esta manera, la función de pesos se puede definir por

$$\omega_t(x) = \frac{K_h(x - x_t)}{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t)}, \quad (6.13)$$

Si se consideran los pesos en (6.13) para el estimador en (6.11) se obtiene el estimador por núcleo de Nadaraya-Watson (ver Nadaraya (1964) y

Watson (1964):

$$\hat{m}(x) = \frac{1}{T} \sum_{t=1}^T \omega_t(x) y_t = \frac{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t) y_t}{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t)}. \quad (6.14)$$

En lo que respecta a la selección de la función núcleo, existen diversos tipos. Sin embargo los más comunes, tanto en el contexto teórico como el aplicado son el *núcleo Gaussiano*

$$K_h(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{x^2}{2h^2}\right), \quad (6.15)$$

y el *núcleo de Epanechnikov* (Epanechnikov (1969))

$$K_h(x) = \frac{0,75}{h} \left(1 - \frac{x^2}{h^2}\right) \mathbb{1}_{\{|x| \leq h\}}. \quad (6.16)$$

El parámetro de suavizado h suele ser un valor cuya variación produce algunos efectos sobre la estimación. Por ejemplo, para valores muy pequeños, es decir $h \rightarrow 0$, entonces se tiene que $\hat{m}(x) \approx y_t$, lo cual quiere decir que el \hat{m} reproduce los datos, puesto que los pesos serán significativos solo para aquellos valores muy cercanos al dato. Por el contrario, si h toma valores muy grandes, es decir $h \rightarrow \infty$, entonces $\hat{m}(x) \rightarrow \bar{y}$, en cuyo caso se obtiene una curva sobresuavizada, la media muestral de los datos. Teniendo en cuenta esto último, resulta crucial la selección del parámetro h .

6.6.2. Selección del Parámetro de Suavizado

Para la elección de h existen diversos métodos en la literatura (ver por ejemplo Härdle (1990) y Fan and Yao (2003)). Entre los métodos que particularmente se aplican en la práctica está el llamado *Método de Validación Cruzada* que consiste en seleccionar el valor de h que minimiza la expresión

$$\min_h CV(h) = \min_h \frac{1}{T} \sum_{j=1}^T (y_j - \hat{m}_{h,j}(x_j))^2 W(x_j), \quad (6.17)$$

donde:

- $\hat{m}_{h,j}$ es el estimador de la función m descartando el dato j -ésimo, es decir

$$\hat{m}_{h,j}(x_j) = \frac{1}{T-1} \sum_{t \neq j} \omega_t(x) y_t,$$

- W es una función de pesos tales que $\sum_{j=1}^n W_j(x_j) = T$,
- $CV(h)$ es conocida con el nombre de función de validación cruzada,

$$CV(h) = \frac{1}{T} \sum_{j=1}^T (y_j - \hat{m}_{h,j}(x_j))^2 W(x_j).$$

Un método de uso frecuente es el denominado *Método Plug-in* el cual está basado en la expansión asintótica del error medio cuadrático integrado *MISE* para funciones de suavizado

$$MISE = \mathbb{E} \left(\int_{-\infty}^{\infty} [\hat{m}(x) - m(x)]^2 dx \right), \tag{6.18}$$

donde, $m(\cdot)$ es la función desconocida. En este caso se obtiene, el *parámetro de suavizado óptimo* siendo este el valor de h que minimize el *MISE*, por supuesto, bajo ciertas condiciones de regularidad. Como se puede apreciar, tal procedimiento requiere de ciertas cantidades desconocidas, las cuales, deben ser estimadas también aplicando algún procedimiento preliminar. Una referencia para la elección del ancho de banda es la ofrecida en Fan and Yao (2003)

$$\hat{h}_{opt} = \begin{cases} 1,06sT^{-1/5}, & \text{para el núcleo Gaussiano,} \\ 2,34sT^{-1/5}, & \text{para el núcleo de Epanechnikov,} \end{cases}$$

donde s es el error estándar de la muestra de la variable independiente, la cual se asume estacionaria. Para una revisión de estos métodos y otros, ver Fan and Yao (2003) y las referencias incluidas en el mismo.

6.6.3. Método de Regresión Local Lineal

El *método de regresión lineal local* es un enfoque de ajuste de curvas a datos mediante técnicas de suavizado en los que el ajuste en x se realiza utilizando únicamente observaciones en un entorno de x . Para ello, se

utiliza una familia paramétrica al igual que en un ajuste de regresión global pero solamente se realiza el ajuste localmente. En la práctica se realizan ciertas suposiciones sobre la función de regresión $m(\cdot)$ tales como la existencia y continuidad de la segunda derivada de en x , para x en el soporte de $m(\cdot)$.

Sea $\{(x_t, y_t)\}_{t=1}^T$ que denota el conjunto de datos disponibles, entonces para estimar $m(x)$ consideramos la expansión local lineal

$$m(\tilde{x}) \approx m(x) + m'(x)(\tilde{x} - x) \equiv \alpha + \beta(\tilde{x} - x),$$

en un entorno de x . El objetivo del método de regresión local lineal es determinar α y β tales que minimicen la función

$$L(\alpha, \beta) = \sum_{t=1}^T (y_t - \alpha - \beta(x_t - x))^2 K_h(x - t - x), \quad (6.19)$$

donde $K_h(\cdot)$ es una función núcleo y h es el ancho de banda. Como resultado se obtiene que \hat{a} es el estimador de $m(x)$ y \hat{b} es el estimador de $m'(x)$. Es claro que, la ecuación (6.19) representa un problema de mínimos cuadrados ponderados, de donde se obtienen expresiones explícitas como solución para a y b . Tomando las derivadas parciales de $L(a, b)$ respecto de a y b se obtiene

$$\hat{a} = \frac{\sum_{t=1}^T \omega_t y_t}{\sum_{t=1}^T \omega_t}, \quad (6.20)$$

donde ω_t se define como

$$\omega_t = K_h(x - x_t)[s_{T,2} - (x - x_t)s_{T,1}].$$

Dado que en la práctica existe la posibilidad de que el denominador en (6.20) se anule, se suele usar

$$\hat{m}(x) = \frac{\sum_{t=1}^T \omega_t y_t}{\sum_{t=1}^T \omega_t + 1/T^2}. \quad (6.21)$$

En cuanto a la selección del ancho de banda, se pueden usar los métodos antes descritos.

6.6.4. Aplicación a Series de Tiempo

Las técnicas anteriores fueron planteadas desde un punto de vista general. Sin embargo, en el análisis de series de tiempo, las variables explicativas son frecuentemente retardos de los valores de la serie. Para el caso de una sola variable explicativa (un retardo), el modelo se expresa por

$$x_t = m(x_{t-1}) + \varepsilon_t. \quad (6.22)$$

Entonces los métodos de regresión por núcleo y regresión local lineal discutidos en las secciones anteriores pueden aplicarse. Cuando hay múltiples variables explicativas (más que un retardo), se hace necesario hacer algunas modificaciones en los métodos. En cuanto a la función núcleo, se considera una expresión multivariante del mismo, por ejemplo, para el núcleo Gaussiano su versión multivariante con una matriz de covarianza preespecificada, se expresa por

$$K_h(x) = \frac{1}{(h\sqrt{2\pi})^p |\Sigma|^{1/2}} \exp\left(-\frac{1}{2h^2} x' \Sigma^{-1} x\right), \quad (6.23)$$

donde p es el número de variables explicativas (o retardos) y Σ es una matriz definida positiva.

Ejemplo 6.6.1. Para ilustrar la aplicación de las técnicas de suavizado en finanzas, se considera la serie de tipo de interés interbancario, *EU-RIBOR*, con plazo a 3-meses, comprendiendo el período del 11 de Julio de 2006 al 8 de Mayo de 2008. El modelo que se propone es

$$y_t = \mu(x_{t-1})dt + \sigma(x_{t-1})dW_t,$$

donde x_t es el tipo de interés con plazo de vencimiento 3-meses, $y_t = x_t - x_{t-1}$, W_t es un movimiento Browniano estándar, y $\mu(\cdot)$ y $\sigma(\cdot)$ son funciones suaves de x_{t-1} . En el modelo considerado, $\mu(x_{t-1}) = \mathbb{E}(y_t|x_{t-1})$. Usaremos el estimador considerado en las secciones previas. Así mismo, por simplicidad usaremos $|y_t|$ como una aproximación (un proxy) de la volatilidad de y_t . La figura (6.1) muestra los gráficos de los tipos de interés para el período en estudio y las variaciones del mismo. Se ha tenido en cuenta este período por contener las observaciones previas al período de crisis, reflejando además el incremento sostenido

del tipo de interés hasta alcanzar el valor máximo y consecuentemente agrupa el período de inestabilidad de los tipos de interés más importante.

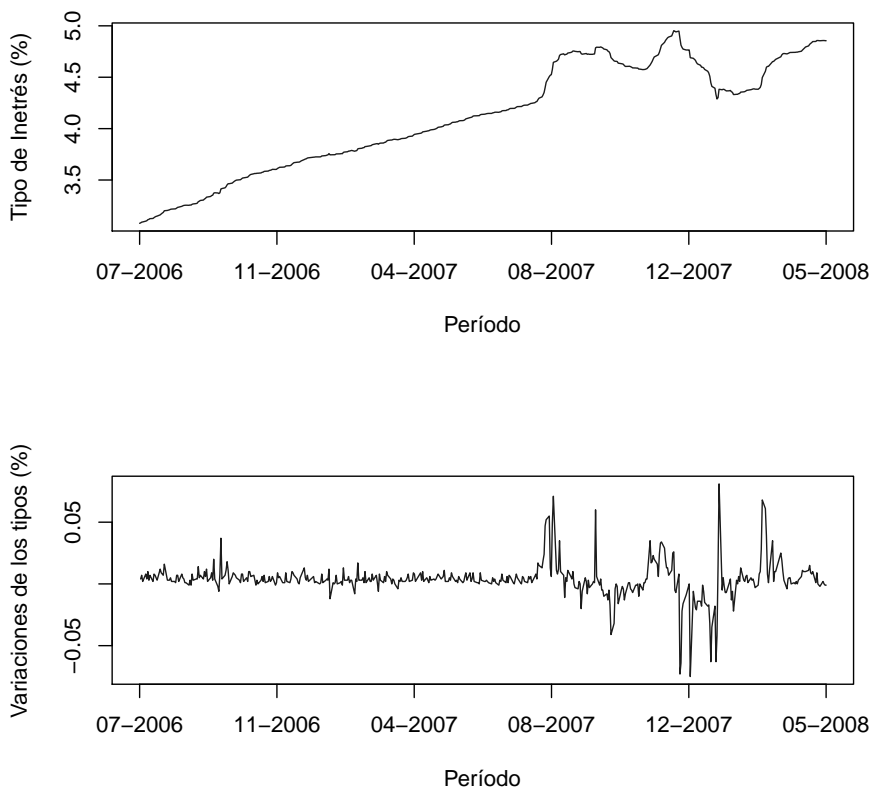


Figura 6.1: Serie de tipo de interés interbancario, *EURIBOR*, con plazo a 3-meses, comprendiendo el período del 11 de Julio de 2006 al 8 de Mayo de 2008. (a) Tipo de interés; (b) Variación del Tipo de Interés.

Los gráficos de la figura (6.2) muestran las estimaciones mediante las técnicas de suavizado de las funciones $\mu(\cdot)$ y $\sigma(\cdot)$. Para el período descrito, se observa en la figura (6.2(a)) el scatterplot entre y_t y x_{t-1} además de la estimación $\hat{\mu}(x_{t-1})$ la cual se observa que es casi nula, sin embargo

al hacer una revisión para una escala más fina (6.2(b)) se observa que hay una tendencia. En cuanto a la volatilidad, se muestra en la (6.2(c)) el scatterplot de $|y_t|$ contra x_{t-1} y la estimación $\hat{\sigma}(x_{t-1})$. El gráfico confirma que la volatilidad es mayor para tipos de interés elevados. Así mismo, la figura (6.2(d)) muestra la estimación $\hat{\sigma}(x_{t-1})$ en una escala refinada. Este ejemplo evidencia el potencial de las técnicas de suavizado en la caracterización de la dinámica de las series financieras.

6.7. Modelo de Coeficiente Funcional Autorregresivo

En el análisis de series de tiempo, los métodos no-paramétricos suelen ser una herramienta preliminar que permite dar una idea acerca del modelo no-lineal paramétrico más apropiado para un conjunto de datos. Chen and Tsay (1993a) proponen el *modelo de coeficiente funcional autorregresivo* denotado por *FAR* expresado por

$$x_t = f_1(X_{t-1})x_{t-1} + \dots + f_p(X_{t-1})x_{t-p} + \varepsilon_t, \quad (6.24)$$

donde $X_{t-1} = (x_{t-1}, \dots, x_{t-k})'$ es un vector de retardos de x_t , el cual puede incluir otras variables explicativas en el instante $t - 1$. Además se asume que las funciones $f_i(\cdot)$ son continuas y dos veces diferenciables, casi seguramente con respecto a sus argumentos. Muchos de los modelos no-lineales discutidos en las secciones anteriores son casos particulares del modelo *FAR*. La estimación de este modelo se puede llevar a cabo aplicando regresión por núcleo o bien regresión local lineal, especialmente para los casos en los que la dimensión de X_{t-1} es pequeña (véase Cai et al. (2000)).

6.8. Modelo No-Lineal Autorregresivo Aditivo

En la aplicación de los métodos no paramétricos para series de tiempo no-lineales es importante tener en cuenta la dimensionalidad, puesto que esta, genera grandes dificultades, sobre todo en la estimación del modelo y especialmente cuando el número de datos no es suficientemente grande. Para superar este inconveniente, una alternativa es considerar un modelo aditivo el cual reduce la dimensionalidad del suavi-

zado. Dada una serie de tiempo x_t , se dice que esta sigue un *modelo no-lineal autorregresivo aditivo* y que se denota por *NAAR* si,

$$x_t = f_0(t) + \sum_{i=1}^p f_i(x_{t-i}) + \varepsilon_t, \quad (6.25)$$

donde $f_i(\cdot)$ son funciones continuas, casi seguramente. Como se puede observar en la definición del modelo, las funciones f_i tienen un único argumento, por lo que el problema de dimensionalidad queda resuelto. Las funciones f_i son estimadas no-paramétricamente de forma iterativa, (véase Chen and Tsay (1993b)). Chen et al. (1995) consideran que las estadísticas de prueba para verificar la hipótesis de aditividad son bastante restrictivas y deben ser examinadas con cuidado en su aplicación.

6.9. Modelo No-Lineal de Espacio de Estado

Las técnicas de Monte Carlo se emplean para controlar la evolución no lineal de la ecuación de estado de transición. A partir de los avances recientes en métodos basados en estas técnicas, Gelfand and Smith (1990), Carlin et al. (1992) proponen un método de Monte Carlo para la modelización no lineal del espacio de estado. El modelo considerado es

$$S_t = f(S_{t-1}) + \eta_t, \quad x_t = g_t(S_t) + v_t, \quad (6.26)$$

donde

- S_t es el vector de estados,
- $f_t(\cdot)$ y $g_t(\cdot)$ son funciones conocidas que dependen de ciertos parámetros desconocidos,
- η_t es una sucesión de vectores aleatorios multivariados *iid* de media cero y matriz de covarianza Σ_η definida no negativa,
- v_t es una sucesión de variables aleatorias *iid* de media cero y varianza σ_v^2 ,
- $\{\eta_t\}$ es independiente de $\{v_t\}$.

La suposición del conocimiento de $f_t(\cdot)$ y $g_t(\cdot)$ en (6.26) puede dificultar el uso práctico del modelo. Una manera de solucionar esto es haciendo uso de métodos no paramétricos tales como los considerados en los modelo *FAR* y *NAAR* para especificar f_t y g_t como un procedimiento previo al uso del modelo no lineal de espacio de estado.

Otros métodos numéricos de suavizado para el análisis de series de tiempo no lineal son los considerados en el trabajo de Kitagawa (1998), y las referencias incluidas en dicho trabajo.

6.10. Tests de No-Linealidad

Como es lógico, es importante determinar la no-linealidad de las series en estudio. En esta sección se discuten algunos tests existentes en la literatura en ese sentido. Dichos tests incluye incluyen tanto enfoques paramétricos como no paramétricos.

6.10.1. Test No-Paramétrico

Es claro que, bajo la hipótesis nula de linealidad, los residuos de un modelo lineal deberían ser independientes. Así pues, cualquier violación de esta premisa indica que el modelo es no apropiado, incluyendo la suposición de linealidad. Esta es la idea básica detrás de un test de no linealidad.

6.10.2. Estadístico de los Residuos al Cuadrado

Uno de los más populares y de sencilla aplicación es el estadístico de los residuos al cuadrado. El estadístico de *Ljung-Box* fue aplicado por McLeod and Li (1983) al cuadrado de los residuos de un modelo *ARMA*(p, q) para determinar si el modelo es apropiado. El estadístico es

$$Q(m) = T(T+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2(\varepsilon_t^2)}{T-i},$$

donde T es el tamaño de la muestra, m es un número adecuado de autocorrelaciones a ser usadas en el test, ε_t denota los residuos de la serie, y $\hat{\rho}_i^2(\varepsilon_t^2)$ es el i -ésimo retardo. Si el modelo lineal es adecuado

$Q(m)$ es asintóticamente una variable aleatoria con distribución chi-cuadrado con $m - p - q$ grados de libertad. La hipótesis nula del test es

$$H_0 : \beta_1 = \dots = \beta_m = 0,$$

donde β_i es el coeficiente de ε_{t-i}^2 en la regresión lineal $\varepsilon_t^2 = \beta_0 + \beta_1\varepsilon_{t-1}^2 + \dots + \beta_m\varepsilon_{t-m}^2 + \epsilon_t$ para $t = m + 1, \dots, T$. Ya que el estadístico es calculado de los residuos, el número de grados de libertad es $m - p - q$.

6.10.3. Test Paramétricos

Entre los test paramétricos tenemos:

6.10.4. El test RESET

Este test propuesto por Ramsey (1969) considera un test de especificación para el análisis de regresión lineal por mínimos cuadrados. El test es conocido como el test *RESET* y se puede aplicar de manera sencilla para los modelos lineales *AR*. Sea el modelo lineal *AR*(p)

$$x_t = X'_{t-1}\boldsymbol{\phi} + \varepsilon_t, \quad (6.27)$$

donde $X_{t-1} = (1, x_{t-1}, \dots, x_{t-p})'$ y $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)'$. Para llevar a cabo este test se siguen los siguientes pasos:

1. Se estima $\boldsymbol{\phi}$ por mínimos cuadrados, se calculan los residuos $\varepsilon_t = x_t - \hat{x}_t$, donde $\hat{x}_t = X'_{t-1}\hat{\boldsymbol{\phi}}$, y la $SSR_0 = \sum_{t=p+1}^T \hat{\varepsilon}_t^2$, con T el tamaño de la muestra.
2. Se considera el modelo de regresión lineal

$$\hat{\varepsilon}_t = X'_{t-1}\boldsymbol{\alpha}_1 + M'_{t-1}\boldsymbol{\alpha}_2 + v_t, \quad (6.28)$$

donde $M_{t-1} = (\hat{x}_t^2, \dots, \hat{x}_t^{s+1})'$ para algún $s \geq 1$, se calculan los residuos

$$\hat{v}_t = \hat{\varepsilon}_t - X'_{t-1}\hat{\boldsymbol{\alpha}}_1 - M'_{t-1}\hat{\boldsymbol{\alpha}}_2$$

y $SSR_1 = \sum_{t=p+1}^T \hat{v}_t^2$.

3. La idea básica del test *RESET* es que si el modelo $AR(p)$ en (6.27) es adecuado entonces α_1 y α_2 en el modelo (6.28) deberían ser cero. Esto último se puede llevar a cabo mediante el estadístico F

$$F = \frac{(SSR_0 - SSR_1)/g}{SSR_1/(T - p - g)}, \quad \text{con } g = s + p + 1, \quad (6.29)$$

para el cual, la linealidad y la suposición de normalidad, tiene distribución F con grados de libertad g y $T - p - g$.

Keenan (1985) propuso un test de no-linealidad para una serie de tiempo que solo usa \hat{x}_t^2 al modificar el segundo paso del test *RESET* para evitar la multicolinealidad entre \hat{x}_t^2 y X_{t-1} . Concretamente, se procede, con la regresión lineal (6.28), de la siguiente forma

1. Se remueve la dependencia lineal de \hat{x}_t^2 en X_{t-1} mediante un ajuste del tipo

$$\hat{x}_t^2 = X'_{t-1}\beta + v_t,$$

del cual se obtienen los residuos $\hat{v}_t = \hat{x}_t^2 - X'_{t-1}\hat{\beta}$.

2. Se considera la regresión

$$\hat{\epsilon}_t = \hat{v}_t\alpha + v_t,$$

del cual se obtiene $SSR_1 = \sum_{t=p+1}^T (\hat{\epsilon}_t - \hat{v}_t\hat{\alpha})^2 = \sum_{t=p+1}^T \hat{v}_t^2$ para probar la hipótesis nula $\alpha = 0$.

6.10.5. El test F

Con el propósito de mejorar la potencia del test de Keenan y el test *RESET*, Tsay (1986) considera $M_{t-1} = \text{vech}(X_{t-1}X'_{t-1})$, siendo $\text{vech}(\cdot)$ el vector *half-stacking* de $X_{t-1}X'_{t-1}$ que considera solo los elementos de la diagonal y bajo la diagonal. La dimensión de M_{t-1} es $p(p+1)/2$ para un modelo $AR(p)$. En la práctica, este es un test que considera un estadístico F parcial para probar la hipótesis $\alpha = 0$ en

$$x_t = X'_{t-1}\phi + M'_{t-1}\alpha + \epsilon_t,$$

donde ϵ_t denota el error. Bajo la suposición de que x_t es un proceso lineal $AR(p)$, el estadístico parcial F sigue una distribución F con g y $T - p - g - 1$ grados de libertad donde $g = p(p+1)/2$.

6.10.6. Test de Umbral

Este test es aplicado a modelos *SETAR*. Para presentar este test, se considera el caso de un modelo *SETAR* con dos regímenes y con variable umbral x_{t-d} . La hipótesis nula H_0 : x_t sigue un modelo lineal $AR(p)$

$$x_t = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t, \quad (6.30)$$

y la hipótesis alternativa H_a : x_t sigue un modelo *SETAR*

$$x_t = \begin{cases} \phi_{01} + \sum_{i=1}^p \phi_{1i} x_{t-i} + \varepsilon_{1t}, & \text{si } x_{t-d} < r_1, \\ \phi_{02} + \sum_{i=1}^p \phi_{2i} x_{t-i} + \varepsilon_{2t}, & \text{si } x_{t-d} \geq r_1, \end{cases} \quad (6.31)$$

donde r_1 es el valor umbral. Para una realización $\{x_t\}_{t=1}^T$ y bajo la suposición de normalidad se considera el logaritmo de la razón de verosimilitudes

$$l(r_1) = l_1(r_1; \hat{\boldsymbol{\phi}}_1, \hat{\sigma}_1^2, \hat{\boldsymbol{\phi}}_2, \hat{\sigma}_2^2) - l_0(\hat{\boldsymbol{\phi}}, \hat{\sigma}_\varepsilon^2)$$

donde l_1 y l_0 son los logaritmos de las funciones de verosimilitud evaluados en los estimados mediante máxima verosimilitud bajo la hipótesis alternativa y bajo la hipótesis nula, respectivamente. La función l es una función del valor umbral, el cual es desconocido y se denomina *parámetro de ruido* bajo la hipótesis nula. Por esta razón, la distribución asintótica de este estadístico es distinta a la distribución del estadístico de razón de verosimilitud usual. Los valores críticos son obtenidos por simulación al considerar $l_{\max} = \sup_{v < r_1 < u} l(r_1)$ para u, v cotas del valor umbral. Una modificación de este test es considerada en Tsay (1989) el cual toma en cuenta una reorganización del modelo autorregresivo.

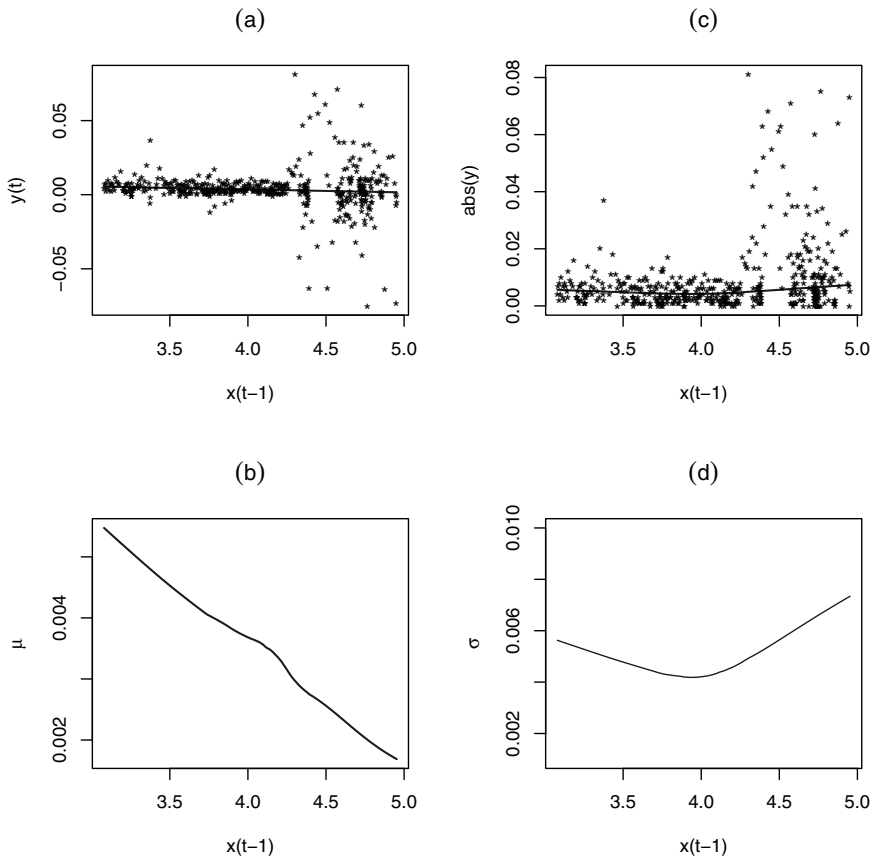


Figura 6.2: Estimación de la media y volatilidad condicional de las series de tipos de interés interbancario, *EURIBOR*, mediante técnicas de suavizado: (a) y_t vs. x_t , con $y_t = x_t - x_{t-1}$, y x_t es el tipo de interés; (b) estimación de $\mu(x_{t-1})$; (c) $|y_t|$ vs. x_{t-1} ; y (d) estimación de $\sigma(x_{t-1})$.

7

Modelos en Tiempo Continuo

El precio de un activo financiero evoluciona con el tiempo formando un proceso estocástico, los precios observados son simplemente una realización del proceso estocástico subyacente. Los procesos discutidos en las secciones anteriores pertenecen a la clase de los procesos estocásticos en tiempo discreto. Por ejemplo, el precio de cierre de un cierto activo se caracteriza mediante un proceso estocástico en tiempo discreto. En este caso particular, los cambios en el precio se presentan al cierre del día de negociación. Sin embargo, durante un mismo día el precio de un activo puede variar significativamente y dicho valor no necesariamente estará relacionado con el precio diario observado. Por ello, es conveniente considerar modelos más generales, que permitan caracterizar tal comportamiento. Es claro que, los modelos en tiempo continuo son los más apropiados para modelizar la dinámica del precio de activos o cualquier otra variable de interés en el contexto de la econometría financiera. En esta sección se hará una breve exposición de los procesos estocásticos en tiempo continuo y de las distintas caracterizaciones en el contexto financiero desarrolladas en los últimos años.

Merton (1990) y Duffie (1995) proporcionan estudios relativos a los modelos en tiempo continuo para datos financieros. En esta sección se pre-

presenta una breve descripción intuitiva y no técnica de dichos modelos. Una minuciosa descripción de estos puede ser encontrada en Kutoyants (2004), Karatzas and Shreve (1991), Kloden and Platen (1992).

A continuación expondremos los aspectos más resaltantes de los procesos en tiempo continuo.

7.1. Movimiento Browniano

Un proceso fundamental a la hora de describir la evolución estocástica de la series financieras es el conocido como *movimiento Browniano* o *Proceso de Wiener*. En los modelos en tiempo discreto, se asume que los impulsos constituyen un proceso de ruido blanco, el cual no es predecible. En los modelos en tiempo continuo, los impulsos tienen su contraparte en los incrementos de un proceso de Wiener. Este proceso estocástico es conocido con el nombre de movimiento Browniano en honor al escocés Robert Brown, biólogo y botánico que descubrió éste fenómeno en 1828. Él observó que pequeñas partículas de polen se desplazaban en movimientos aleatorios e irregulares sin razón aparente. En 1900 Louis Bachelier en su tesis doctoral *la teoría de la especulación*, utilizó el movimiento Browniano en su teoría matemática como modelo del movimiento de los precios de los títulos. Albert Einstein en 1905 hizo una descripción matemática del fenómeno, obteniendo a partir de esta las ecuaciones del movimiento Browniano. Así mismo, Norbert Wiener en 1931 estudia el fundamento matemático empresarial del movimiento Browniano, por ello frecuentemente también es conocido como proceso de Wiener. Indistintamente se hará referencia a movimiento Browniano o Proceso de Wiener, sin que ello sea motivo de confusión.

Hay diversas maneras de caracterizar y definir el proceso de Wiener, $\{W_t\} \equiv \{W(t), t \geq 0\}$, y una de ellas es la siguiente. El proceso de Wiener es un proceso Gaussiano (ver Apéndice A) con trayectorias continuas y con incrementos independientes, tal que $W(0) = 0$ con probabilidad 1, $\mathbb{E}(W(t)) = 0$, y $\text{Var}(W(t) - W(s)) = t - s$, para todo $0 \leq s \leq t$.

Definición 7.1.1. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, entonces el movimiento Browniano o proceso de Wiener es un proceso estocástico $\{W_t\} = \{W_t, t \geq 0\}$ en tiempo continuo tal que satisfice

1. $W_0 = 0$ casi seguro,
2. Las trayectorias $t \mapsto W_t$ son (casi seguramente) continuas.
3. Para cualquier sucesión finita de tiempos $t_0 < t_1 < \dots < t_n$, los incrementos

$$W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$$

son estacionarios e independientes.

4. Para cualesquiera instantes $0 \leq s < t$, los incrementos $W_t - W_s$ tiene distribución normal con media cero y varianza $(t - s)$.

Se deduce de esta definición que $W_t - W_s$ tiene la misma distribución que

$$W_{t-s} - W_0 = W_{t-s},$$

normal con media cero y varianza $(t - s)$. Por tanto, la varianza es proporcional a la longitud del intervalo $[s, t]$. Una interpretación intuitiva de esto es: cuanto más grande el intervalo, mayor son las fluctuaciones del proceso en dicho intervalo. Debe quedar claro que, la identidad $W_t - W_s \stackrel{\mathcal{D}}{=} W_{t-s}$ en términos de distribución, en general no implica identidad a lo largo de la trayectoria: $W_t(\omega) - W_s(\omega) = W_{t-s}(\omega)$.

Por otro lado, es inmediato de la definición (7.1.1) que el movimiento Browniano tiene función de media:

$$\mu_W(t) = \mathbb{E}(W_t) = 0, \quad t \geq 0,$$

y puesto que los incrementos son independientes, su función de covarianza

$$\text{Cov}(W_t, W_s) = \mathbb{E}(W_t W_s) = s, \quad 0 \leq s < t.$$

Otra forma de describir el proceso de Wiener, es la que considera las variaciones del proceso $\Delta W_t = W_{t+\Delta t} - W_t$, (en intervalos de longitud Δt , con $\Delta t \rightarrow 0$), asociados con incrementos de tamaño Δt en el tiempo. Así, el proceso de Wiener satisface:

- i. $\Delta W_t = \epsilon \sqrt{\Delta t}$, donde ϵ es una variable aleatoria normal estándar;
- y

ii. ΔW_t es independiente de W_s para $s \geq t$.

De la condición (i.), se deduce que ΔW_t tiene distribución normal con media cero y varianza Δt , y que se denotará por $\Delta W_t \sim N(0, \Delta t)$. La segunda condición (ii.) es la propiedad de Markov la cual expresa que, condicionado al valor presente W_t , cualquier información del pasado del proceso, W_s con $s < t$, es irrelevante para un valor particular en el futuro W_{t+u} con $u > 0$. De la definición (7.1.1), se asume que el proceso comienza en $t = 0$ con valor inicial W_0 , el cual es fijo y casi seguramente igual a cero. Entonces $W_t - W_0$ se puede representar como una suma de pequeños incrementos, esto es, para $T = t/\Delta t$, donde Δt es un incremento positivo se tiene que

$$W_t - W_0 = W_{T\Delta t} - W_0 = \sum_{i=1}^T \Delta W_i = \sum_{i=1}^T \epsilon_i \sqrt{\Delta t},$$

donde $\Delta W_i = W_{i\Delta t} - W_{(i-1)\Delta t}$. Puesto que los ϵ_i son independientes y de la definición (7.1.1), entonces

$$\mathbb{E}(W_t - W_0) = 0, \quad \text{Var}(W_t - W_0) = T\Delta t = t.$$

Lo cual concuerda con lo mencionado en los párrafos anteriores. Una conclusión que se puede extraer de esto último es que, la varianza de un proceso de Wiener o bien Movimiento Browniano se incrementa de forma lineal con la longitud del intervalo de tiempo.

Los procesos gaussianos son caracterizados por su esperanza y función de covarianza, entonces una definición alternativa a la dada en (7.1.1) es

Definición 7.1.2. *El movimiento Browniano es un proceso gaussiano con media*

$$\mu_W(t) = \mathbb{E}(W_t) = 0$$

y función de covarianza

$$\gamma_W(t, s) = \text{Cov}(W_t, W_s) = \min(t, s).$$

Del movimiento Browniano se pueden deducir varios procesos estocásticos gaussianos y no-gaussianos de importancia práctica. Como antes, $\{W_t, t \geq 0\}$ denota el movimiento Browniano, y los procesos que se derivan de este se presentan a continuación.

7.2. Puente Browniano

Considérese el proceso

$$x_t = W_t - tW_1, \quad 0 \leq t \leq 1. \quad (7.1)$$

Es claro que,

$$x_0 = W_0 - 0W_1 = 0, \quad \text{y} \quad x_1 = W_1 - 1W_1 = 0.$$

Por esta razón, el proceso es llamado *puente Browniano estándar*. Una simple inspección de las trayectorias de este proceso puede ratificar este hecho. Se puede demostrar que las distribuciones finito-dimensionales de x son gaussianas y por tanto el puente Browniano x es un proceso Gaussiano con función de media y covarianzas dadas por

$$\mu_x(t) = \mathbb{E}(x_t) = 0, \quad \text{y} \quad \gamma_x(t, s) = \text{Cov}(x_t, x_s) = \min(t, s) - ts, \quad s, t \in [0, 1].$$

El puente Browniano aparece como el proceso límite de la función de distribución empírica normalizada de una muestra de variables aleatorias uniformes $U(0, 1)$ independientes e idénticamente distribuidas. Este es un resultado fundamental de la estadística no-paramétrica; es la base de numerosos contrastes de bondad de ajuste en estadística.

7.3. Movimiento Browniano con Tendencia

Se introduce una variante del movimiento Browniano, en la que su función esperanza depende del tiempo. Este tipo de proceso es conocido con el nombre de *movimiento Browniano con tendencia* y se expresa por

$$x_t = \mu t + \sigma W_t, \quad t \geq 0 \quad (7.2)$$

donde σ es una constante no negativa y $\mu \in \mathbb{R}$. Evidentemente, este es un proceso Gaussiano con funciones de media y covarianza

$$\mu_x(t) = \mathbb{E}(x_t) = \mu t, \quad \text{y}$$

$$\gamma_x(t, s) = \text{Cov}(x_t, x_s) = \sigma^2 [\min(t, s) - ts], \quad s, t \geq 0.$$

La función $\mu_x(t) = \mu t$ (*función determinista del proceso*), esencialmente determina la forma característica de las trayectorias del proceso.

7.4. Movimiento Browniano Geométrico

A partir del trabajo de Bachelier (1900) acerca de que los precios de los activos con riesgo (índices de títulos, tipos de cambio, precio de activos, etc.) se pueden caracterizar mediante el movimiento Browniano, los procesos estocásticos en tiempo continuo demostraron ser una herramienta útil en las aplicaciones financieras. Sin embargo, el movimiento Browniano como proceso gaussiano puede tomar valores negativos, una propiedad poco deseable en finanzas y menos aún para un precio. Black and Scholes (1973) y Merton (1973) introducen otros procesos estocásticos como modelos para los precios especulativos en la valoración de opciones Europeas. El proceso sugerido por ellos viene dado por

$$x_t = e^{(\mu t + \sigma W_t)}, \quad t \geq 0. \quad (7.3)$$

y es conocido con el nombre de *Movimiento Browniano Geométrico*, denotado por *GBM*. El proceso x_t no es Gaussiano, y sus funciones de media y covarianza son

$$\mu_x(t) = \mathbb{E}(x_t) = e^{(\mu + \frac{1}{2}\sigma^2)t},$$

y

$$\gamma_x(t, s) = \text{Cov}(x_t, x_s) = \left(e^{(\mu + 0.5\sigma^2)(t+s)} \right) (e^{\sigma^2 t} - 1).$$

En particular, el movimiento Browniano Geométrico tiene varianza

$$\sigma_x^2(t) = (e^{\sigma^2 t} - 1) \left(e^{(2\mu + \sigma^2)t} \right).$$

En la literatura, μ_x y σ_x son conocidos como el *drift* y la *volatilidad* del proceso x_t , respectivamente.

7.5. Proceso de Itô

De la ecuación (7.3), se deduce que los parámetros del movimiento Browniano Geométrico denominados drift y volatilidad son invariantes respecto del tiempo. Si se consideran modelos que permitan que dichos parámetros sean funciones que dependan del proceso estocástico x_t , entonces se obtienen los *procesos de Itô*. Concretamente, un proceso x_t , es un proceso de Itô si satisface dada por

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t, \quad x_0, \quad t \geq 0, \quad (7.4)$$

donde W_t es un proceso de Wiener, y x_0 es el valor inicial de x_t en el instante 0, el cual es independiente de W_t . La ecuación (7.4) es conocida como la *ecuación diferencial estocástica* con funciones no aleatorias *drift*, $\mu(x_t, t)$ y *difusión o volatilidad*, $\sigma(x_t, t)$, por ello, también se le conoce con el nombre de *Proceso de Difusión*. El proceso x_t es un proceso estocástico continuo de gran importancia en las finanzas y también se puede representar en su forma integral como

$$x_t = x_0 + \int_0^t \mu(x_u, u) du + \int_0^t \sigma(x_u, u) dW_u, \quad (7.5)$$

La segunda integral de (7.5) es una integral estocástica (ver Apéndice C). El proceso de Wiener es un proceso especial de Itô ya que este satisface la ecuación diferencial estocástica (7.4), donde $\mu(x_t, t) = 0$ y $\sigma(x_t, t) = 1$.

Un caso particular del proceso (7.4) es el denominado *Proceso Difusión Homogéneo* definido como solución de la ecuación diferencial estocástica (SDE),

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \quad x_0, \quad 0 \leq t \leq T. \quad (7.6)$$

En este caso las funciones $\mu(\cdot)$ and $\sigma^2(\cdot)$ no dependen del tiempo, por ello se denomina homogéneo. Como antes, x_t se puede representar en su forma integral como

$$x_t = x_0 + \int_0^t \mu(x_u) du + \int_0^t \sigma(x_u) dW_u, \quad 0 \leq t \leq T.$$

Este proceso puede ser estudiado aplicando dos tipos de enfoque. El primero, aplicando un enfoque paramétrico, en el que se considera al proceso x_t como una solución de

$$dx_t = \mu(x_t; \theta)dt + \sigma(x_t; \theta)dW_t, \quad x_0 \quad 0 \leq t \leq T \quad (7.7)$$

y se tiene en cuenta el conocimiento de las funciones $\mu(x) = \mu(x, \theta)$, y $\sigma(x) = \sigma(x, \theta)$ con $x \in \mathbb{R}$ y θ un parámetro desconocido tal que $\theta \in \Theta \subset \mathbb{R}^d$ el cual se puede estimar por algún método apropiado. O bien un enfoque no-paramétrico en el que las funciones $\mu(x)$ y $\sigma(x)$, con $x \in \mathbb{R}$ son funciones desconocidas. En el contexto paramétrico,

los modelos de difusión se han aplicado para modelizar los tipos de interés, por ejemplo, Vasicek (1977) propone el siguiente modelo

$$dx_t = \kappa(\mu - x_t)dt + \sigma dW_t, \quad \alpha > 0, \quad \beta > 0. \quad (7.8)$$

para el precio de los derivados del tipo de interés. El modelo de Vasicek asume que la volatilidad σ es constante, lo cual es una suposición que no suele coincidir con las observaciones empíricas. Por ejemplo, la volatilidad tienden a agruparse y las observaciones más grandes se asocian con las volatilidades más grandes. Un modelo más realista sería tener en cuenta la volatilidad no constante. Cox et al. (1985) obtienen el modelo *CIR* dado por la ecuación diferencial estocástica

$$dx_t = \kappa(\mu - x_t)dt + \sigma x_t^{1/2} dW_t. \quad (7.9)$$

en donde la volatilidad es una función del tipo de interés. En este caso se asume que $2\mu\kappa \geq \sigma^2$, por lo cual, (7.9) admite una solución no-negativa. Estos y otros modelos serán considerados y sus detalles en el siguiente capítulo, referido a los modelos de tipos de interés.

La existencia de las soluciones de la ecuación diferencial estocástica presentada en (7.4) están sujetas a ciertas condiciones de regularidad, similar a lo que ocurre en las ecuaciones diferenciales ordinarias (ver Apéndice C). Por lo general, no siempre es posible obtener soluciones explícitas, si estas existen, para la ecuación (7.4), salvo casos particulares. En cuyo caso, similar a lo que se hace en el análisis de ecuaciones diferenciales ordinarias, hay una variedad de métodos para aproximar las soluciones a partir de la discretización de los procesos. Entre ellos podemos destacar el *método de Euler* por ser el más sencillo. Dicho método es una extensión del método del mismo nombre para ecuaciones diferenciales ordinarias. Así pues, si consideramos el paso de discretización Δt y $t_i = t_0 + i\Delta t$, la aproximación de Euler del proceso (7.4) viene dada por

$$x_{t_{i+1}} = x_{t_i} + \mu(x_{t_i}, t_i)\Delta t + \sigma(x_{t_i}, t_i)\Delta W_{t_i} \quad (7.10)$$

donde x_{t_0} es la condición inicial. $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$, son los incrementos de un movimiento Browniano. En cuanto a los métodos de discretización, estos pueden ser caracterizados por su orden de convergencia. Un esquema de discretización y^δ con paso de discretización de tamaño

$\delta = \Delta t$ se dice *fuertemente convergente* a x en el instante T con orden $\gamma > 0$ si existe una constante positiva C que no depende de δ y tal que

$$\mathbb{E}(|x_T - y^\delta(T)|) \geq C\delta^\gamma. \quad (7.11)$$

La expresión (7.11) mide cuan próximas están las trayectorias en el instante T . Por otra parte, y^δ se dice *débilmente convergente* con orden $\beta > 0$ a x en el instante T si para cada función g suficientemente regular, existe una constante positiva C que no depende de δ y tal que

$$|\mathbb{E}(g(x_T)) - \mathbb{E}(g(y^\delta(T)))| \geq C\delta^\beta. \quad (7.12)$$

Por lo tanto, la convergencia débil de orden β implica la convergencia de los momentos de todos los ordenes (tomando $g(x) = |x|^q$). Bajo condiciones apropiadas de suavidad y crecimiento sobre las funciones drift y difusión, el método de euler es fuertemente convergente de orden 0,5 y débilmente convergente de orden 1.

Algunos detalles teóricos de estos procesos se describen en el Apéndice C. En cuanto a los diversos modelos aplicados en la teoría financiera, el próximo capítulo hace una descripción más precisa de estos procesos como herramientas para la modelización de las series financieras, concretamente los tipos de interés. Para una discusión más precisa y amplia de los modelos en tiempo continuo y procesos de difusión, ver por ejemplo, Shreve (2004), Karatzas and Shreve (1991), Iacus (2008), Kutoyants (2004).

7.5.1. Lema de Itô

En finanzas, cuando se aplica un modelo en tiempo continuo, es común asumir que el precio de un activo es un proceso de Itô. Por ello es importante tener conocimiento de las herramientas que nos proporciona el cálculo estocástico de Itô. Evidentemente una ecuación diferencial estocástica se distingue de una ecuación diferencial ordinaria por el término aleatorio que involucra el proceso de Wiener. Por esta razón, una extensión natural del calculo diferencial es posible a través del lema de Itô.

Lema 7.5.1. *Supóngase que x_t es un proceso estocástico en tiempo continuo tal que satisface la ecuación diferencial estocástica*

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t, \quad (7.13)$$

donde W_t es un proceso de Wiener. Adicionalmente, $G(x_t, t)$ es una función diferenciable de x_t y t . Entonces

$$dG = \left\{ \frac{\partial G}{\partial x} \mu(x_t, t) + \frac{\partial G}{\partial t} + \frac{1}{2} \frac{\partial^2 G}{\partial x^2} \sigma^2(x_t, t) \right\} + \frac{\partial G}{\partial x} \sigma(x_t, t) dW_t. \quad (7.14)$$

Para ilustrar cómo se aplica este lema, considere como función $G(W_t, t) = W_t^2$, el cuadrado del proceso de Wiener. En este caso tenemos que $\mu(W_t, t) = 0$ y $\sigma(W_t, t) = 1$ y

$$\frac{\partial G}{\partial W_t} = 2W_t, \quad \frac{\partial G}{\partial t} = 0, \quad \frac{\partial^2 G}{\partial W_t^2} = 2.$$

Por lo tanto,

$$dW_t^2 = \left(2W_t \times 0 + 0 + \frac{1}{2} \times 2 \times 1 \right) dt + 2W_t dW_t = dt + 2W_t dW_t.$$

7.5.2. Aplicación al Precio de los Activos

Una aplicación a los precios de un activo, puede ser descrita de la siguiente forma. Sea P_t el precio de un activo en el instante t , el cual es continuo en $[0, \infty)$. Con frecuencia en la literatura financiera se suele asumir que P_t se rige por un proceso estocástico de Itô

$$dP_t = \mu P_t dt + \sigma P_t dW_t, \quad (7.15)$$

donde μ y σ son constantes. Haciendo uso de la notación básica de los procesos de Itô en (7.4), se tiene que $\mu(P_t, t) = \mu P_t$ y $\sigma(P_t, t) = \sigma P_t$. Este proceso está asociado al movimiento Browniano Geométrico discutido previamente. Al aplicar el lema de Itô se obtiene el modelo en tiempo continuo para el logaritmo del precio de los activos P_t . sea $G(P_t, t) = \ln(P_t)$ el logaritmo del precio del activo en cuestión. Entonces aplicando el lema, se obtiene que

$$d \ln(P_t) = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dW_t.$$

Como se puede observar, el logaritmo del precio sigue un movimiento Browniano con drift $(\mu - \sigma^2/2)$ y varianza σ^2 si el precio es un movimiento geométrico Browniano. En consecuencia, las variaciones en el

logaritmo de los precios (logaritmo de las rentabilidades) entre el precio actual en el instante t y algún instante futuro T tienen distribución normal con media $(\mu - \sigma^2/2)(T - t)$ y varianza $\sigma^2(T - t)$. Si el intervalo de tiempo $T - t = \Delta$ es fijo y los incrementos en logaritmo del precio están equiespaciados, entonces la serie de incrementos es un proceso Gaussiano con media $(\mu - \sigma^2/2)\Delta$ y varianza $\sigma^2\Delta$.

7.5.3. Estimación de μ y σ

Los parámetros μ y σ del movimiento Browniano se pueden estimar de diversas maneras, más adelante discutiremos este tema. Pero en este caso, se ha supuesto que el precio sigue un movimiento browniano geométrico. En este caso, una de las formas de estimar los parámetros, más sencilla y práctica es la siguiente. Supóngase que se tienen $n + 1$ observaciones del precio P_t , es decir $\{P_0, P_1, \dots, P_n\}$, igualmente espaciadas en un intervalo, esto es con paso Δ (por ejemplo, observaciones diarias, semanales o mensuales). Sea ahora, $r_t = \ln(P_t) - \ln(P_{t-1})$, para $t = 1, \dots, n$ la rentabilidad continua, entonces si se tienen en cuenta los resultados en la sección anterior para P_t siguiendo un movimiento Browniano geométrico, se tiene que r_t tiene distribución normal con media $\mu_r = (\mu - \sigma^2/2)\Delta$ y varianza $\sigma_r^2 = \sigma^2\Delta$. Sea ahora \bar{r} y s_r la media de la muestra y la desviación estándar de la muestra. Ambos estimadores son consistentes, $\bar{r} \rightarrow \mu_r$ y $s_r \rightarrow \sigma_r$ cuando $n \rightarrow \infty$. Entonces, σ se estima a partir de

$$\hat{\sigma} = \frac{s_r}{\sqrt{\Delta}}.$$

Teniendo en cuenta que $\mu_r = \bar{r}$, se puede estimar μ por

$$\hat{\mu} = \frac{\bar{r}}{\Delta} + \frac{s_r^2}{2\Delta}.$$

Ejemplo 7.5.1. Considérense las rentabilidades continuas (*logaritmo de las rentabilidades*) con frecuencia diaria del activo del IBEX35, Banco Santander, en el período del 2 de Enero 2009 al 30 de diciembre de 2009. La figura (7.1) muestra la evolución de la rentabilidad continua del activo durante dicho período, compuesto de 252 observaciones y la función de autocorrelación de dicha rentabilidad. Claramente la función de autocorrelación evidencia que las rentabilidades no están correlacionadas,

al menos hasta el retardo 15. El estadístico de Ljung-Box arroja como resultado $Q(15) = 14,7195$, y un $p_{value} = 0,4718$, con lo cual se ratifica el hecho anterior.

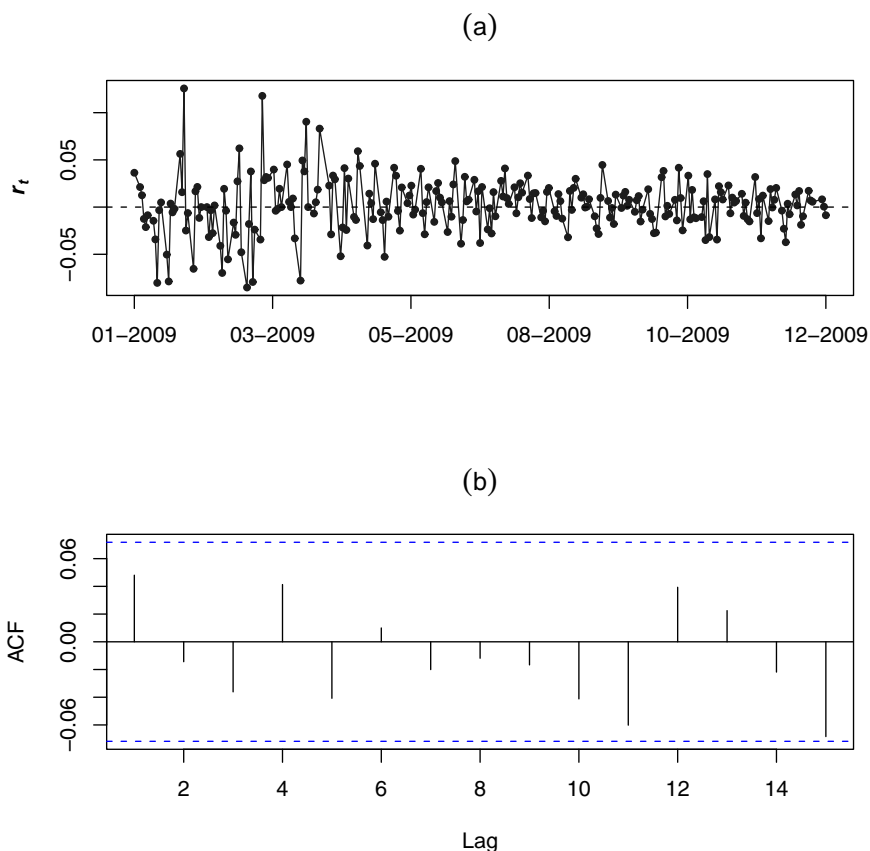


Figura 7.1: Activo: Banco Santander, período 2 de Enero 2008 al 3 de Diciembre 2010; (a) Rentabilidades continuas (b) Función de autocorrelación muestral.

Si se asume que el precio del activo Banco Santander durante el período de 2 de Enero 2009 al 30 de diciembre de 2009, se rige por el movimiento Browniano geométrico *GBM* de la ecuación (7.3), entonces a partir de las rentabilidades continuas se pueden estimar los parámetros de dicho

proceso. De las rentabilidades obtenemos que:

$$\bar{x} = 0,002114736, \quad \text{and} \quad s_x = 0,02874117$$

El número de días de negociación es 252, entonces se establece $\Delta = 1/252$, de donde se obtiene que

$$\hat{\sigma} = \frac{s_x}{\sqrt{\Delta}} = 0,4562518, \quad \text{and} \quad \hat{\mu} = \frac{\bar{x}}{\Delta} + \frac{\hat{\sigma}^2}{2} = 0,6369963$$

7.6. Distribución del Precio de los Activos y la Rentabilidad Continua

Como ya se ha comentado en la sección anterior, el logaritmo de la rentabilidad sigue un proceso Browniano Geométrico, si se el precio del activo asociado satisface la ecuación

$$dP_t = \mu P_t dt + \sigma P_t dW_t.$$

Por lo tanto, las variaciones del logaritmo del precio en el intervalo $(t, T]$ tienen distribución normal

$$\ln(P_T) - \ln(P_t) \sim N(\mu_{P_t}, \sigma_{P_t}^2) \quad (7.16)$$

con $\mu_{P_t} = (\mu - \frac{\sigma^2}{2})(T - t)$ y $\sigma_{P_t}^2 = \sigma^2(T - t)$. Teniendo en cuenta este resultado, se tiene que condicionado al precio P_t en el instante t , el logaritmo de P_T para $T > t$ tiene distribución normal

$$\ln(P_T) \sim N(\ln(P_t) + \mu_{P_t}, \sigma_{P_t}^2). \quad (7.17)$$

Del cálculo básico de probabilidades, la media y varianza (condicionada a P_t) de P_T es

$$\begin{aligned} \mathbb{E}(P_T) &= P_t e^{\mu(T-t)} \\ \text{Var}(P_T) &= P_t^2 e^{2\mu(T-t)} (e^{\sigma^2(T-t)} - 1) \end{aligned}$$

Esto ratifica que la tasa esperada de rentabilidad del activo es μ . Si ahora, se considera r la tasa continua de rentabilidad anual desde el instante t al instante T . Entonces

$$P_T = P_t e^{r(T-t)},$$

donde T y t son medidos en años. Por lo tanto,

$$r = \frac{1}{T-1} \ln \left(\frac{P_t}{P_1} \right).$$

De la ecuación (7.16) se tiene que

$$\ln \left(\frac{P_t}{P_1} \right) \sim N(\mu_{P_t}, \sigma_{P_t}^2).$$

Por consiguiente, la distribución de la tasa continua anual de la rentabilidad es

$$r \sim N \left(\mu - \frac{\sigma^2}{2}, \frac{\sigma^2}{T-t} \right).$$

7.7. Procesos de Difusión con Salto

Mandelbrot (1963) estudió las variaciones en los precios del algodón encontrando ciertas discrepancias en cuanto a la teoría establecida hasta el momento, es decir, la suposición de que las rentabilidades de los activo basadas en logaritmos son variables aleatorias Gaussianas independientes. En primer lugar Mandelbrot se encontró con que, los histogramas de las variaciones del precio tienen puntos máximos (picos) que superan a los de una distribución gaussiana, y en segundo lugar las colas de las distribuciones de dichas variaciones son tan extraordinariamente largas que puede asumirse que el segundo momento es infinito. Asimismo, Mandelbrot sugiere que un buen modelo alternativo para los cambios del precio del algodón es la *distribución estable* (véase Apéndice B) con un índice de 1.7, un enfoque pionero en la modelización de datos financieros basado en los *procesos de Lévy*. El proceso de Lévy es un proceso estocástico cuyas trayectorias muestrales son continuas por la derecha y con incrementos independientes y estacionarios. Casos especiales de procesos de Lévy incluyen movimientos brownianos, procesos de Poisson, y los procesos estables entre otros. Los dos últimos procesos son procesos de tipo salto.

Por otra parte, los estudios empíricos han encontrado que el modelo de difusión basado en el movimiento browniano no es suficiente para explicar algunas características de la rentabilidad de los activos y los precios de sus derivados (ver Bakshi et al. (1997), y las referencias allí citadas). Entre los modelos (modelos en tiempo continuo) alternativos que

se proponen en la literatura se encuentran los modelos de difusión con saltos y los modelo de volatilidad, véase Merton (1976) y Duffie (1995). La idea básica es suponer que hay dos tipos de aleatoriedad que rigen el precio de las acciones: la primera inducida por un movimiento Browniano generando las trayectorias continuas y pequeños movimientos, mientras que la segunda genera los saltos grandes e infrecuentes que representan los impulsos repentinos. En particular, Merton (1976) supone que el precio de los activos siguen el modelo de difusión con salto

$$\frac{dP_t}{P_t} = (\mu - \lambda\kappa)dt + \sigma dW_t + J_t dN_t, \quad (7.18)$$

donde N_t es un proceso de Poisson que caracteriza la ocurrencia de los saltos con intensidad λ , J_t es una variable independiente que determina el tamaño del salto, si este ocurre en el instante t . Se asume que los saltos son independientes e idénticamente distribuidos, y $\kappa = \mathbb{E}(J_t)$. La ecuación (7.18) admite la siguiente solución basada en la fórmula de Doléans-Dade,

$$\frac{P_t}{P_0} = \left(e^{(\mu - \frac{\sigma^2}{2} - \lambda\kappa)t + \sigma W_t} \right) \prod_{i=1}^{N_t} J_{T_i}, \quad (7.19)$$

con la convención $\prod_{i=1}^0 = 1$, donde T_i , $1 \geq i \geq N_t$, denota los instantes en los que ocurren los saltos. Una selección natural de J_t es una variable aleatoria lognormal obteniéndose que P_t/P_0 tiene distribución lognormal.

En los últimos años, el modelo (7.18) ha sido estudiado extensivamente, especificando las distintas estructuras del drift, la difusión, y los componentes de salto. Por ejemplo, suponiendo que la magnitud de los saltos son dependientes, Oldfield et al. (1977) proponen un modelo de salto de difusión autorregresivo. Ball and Torous (1999) reemplazan el proceso de Poisson por un proceso de salto Bernoulli y sostienen que este proceso puede tener mejores propiedades, desde un punto de vista computacional así como también desde un punto de vista empírico y teórico. Ramezani and Zeng (2007) utilizan una distribución doble-exponencial asimétrica para el logaritmo de la variable asociada al tamaño del salto, $\log(J_t)$, con lo cual demuestran que el modelo resultante tiene la capacidad de capturar características relativas a la asimetría leptocúrtica y que además se puede destacar "*La sonrisa de la volatilidad*" una característica frecuentemente observada en los datos financieros.

Kou (2002), propone un modelo en el que el precio sigue la ecuación diferencial estocástica

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t + d \left(\sum_{i=1}^{N_t} (J_i - 1) \right), \quad (7.20)$$

donde, como antes, W_t es un proceso de Wiener, N_t es un proceso de Poisson con tasa λ y $\{J_i\}$ es una sucesión de variables aleatorias no-negativas independientes e idénticamente distribuidas tal que $X = \ln(J)$ tiene distribución doble-exponencial con función de densidad de probabilidades

$$f_X(x) = \frac{1}{2\eta} e^{-|x-\theta|/\eta}, \quad 0 < \eta < 1. \quad (7.21)$$

Teniendo en cuenta la solución (7.19), el modelo (7.20) tiene similar solución

$$P_t = P_0 \left(e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} \right) \prod_{i=1}^{N_t} J_i. \quad (7.22)$$

A partir de esta solución se obtiene que la rentabilidad del activo subyacente para un incremento Δt suficientemente pequeño es

$$\frac{P_{t+\Delta t} - P_t}{P_t} = \exp \left(\left(\mu - \frac{1}{2}\sigma^2 \right) \Delta t + \sigma (W_{t+\Delta t} - W_t) + \sum_{i=N_t+1}^{N_{t+\Delta t}} X_i \right) - 1,$$

donde $X_i = \ln(J_i)$. Si se considera la expansión de Taylor de e^x para aproximar la rentabilidad anterior se obtiene que, teniendo en cuenta $(\Delta W_t)^2 \approx \Delta t$

$$\begin{aligned} \frac{P_{t+\Delta t} - P_t}{P_t} &\approx \left(\mu - \frac{1}{2}\sigma^2 \right) \Delta t + \sigma \Delta W_t + \sum_{i=N_t+1}^{N_{t+\Delta t}} X_i + \frac{1}{2}\sigma^2 (\Delta W_t)^2 \\ &\approx \mu \Delta t + \sigma \epsilon \sqrt{\Delta t} + \sum_{i=N_t+1}^{N_{t+\Delta t}} X_i, \end{aligned}$$

donde $\Delta W_t = W_{t+\Delta t} - W_t$ y ϵ son variables aleatorias normales. Bajo la suposición de que la ocurrencia de los saltos se rige por un proceso Poisson, la probabilidad de que suceda un salto en el intervalo

$(t, t + \Delta t]$ es $\lambda\Delta t$. Para Δt pequeño, ignorando los saltos múltiples, se tiene que

$$\sum_{i=N_t+1}^{N_{t+\Delta t}} X_i \approx \begin{cases} X_{N_t+1} & \text{con probabilidad } \lambda\Delta t, \\ 0 & \text{con probabilidad } 1 - \lambda\Delta t. \end{cases}$$

Combinando los resultados anteriores, entonces la rentabilidad simple de un activo se distribuye aproximadamente como

$$\frac{P_{t+\Delta t} - P_t}{P_t} \approx \mu\Delta t + \sigma\epsilon\sqrt{\Delta t} + I \times X, \tag{7.23}$$

donde I es una variable aleatoria Bernoulli con $\mathbb{P}\{I = 1\} = \lambda\Delta t$, $\mathbb{P}\{I = 0\} = 1 - \lambda\Delta t$, y X es una variable aleatoria doble-exponencial.

7.8. Modelo de Volatilidad Estocástica

En secciones previas se estudió el modelo de volatilidad SV para modelos en tiempo discreto. En esta sección se considera el modelo de volatilidad estocástica en tiempo continuo. El trabajo de Hull and White (1987) es uno de los primeros en estudiar los modelos de volatilidad en tiempo continuo y que denotaremos por CSV . El modelo considerado por ellos es

$$\begin{aligned} dP_t &= \mu P_t dt + \sigma_t P_t dW_{1t} \\ &\text{y} \\ d\sigma_t^2 &= \beta\sigma_t^2 dt + v\sigma_t^2 dW_{2t} \end{aligned} \tag{7.24}$$

donde $\{W_{1t}\}$ y $\{W_{2t}\}$ son dos movimientos Brownianos estándar cuyos incrementos tienen correlación ρ . La volatilidad $\{\sigma_t^2\}$ es un movimiento Browniano Geométrico. Posteriormente, otros modelos de volatilidad estocástica se han considerado, entre los que se pueden destacar, el modelo de Scott (1987) quien introduce el proceso de Ornstein-Uhlenbeck para describir la volatilidad $d\sigma_t = \kappa(\mu - \sigma_t)dt + v dW_{2t}$. Melino and Turnbull (1990) asumen un modelo $CKLS$,

$$dP_t = \kappa(\mu - P_t)dt + \sigma_t P_t^\gamma dW_t$$

para los tipos de cambio enotao por P_t , entre el Dólar Canadiense y el Dólar Estadounidense, con el proceso de volatilidad $\ln(\sigma_t)$ un proceso de Ornstein-Uhlenbeck.

Barndorff-Nielsen and Shephard (2001) introduce una clase de modelos de volatilidad estocástica basados en procesos de Lévy mediante la suposición de que el logaritmo del precio de un activo se rige por

$$\begin{aligned} d \log(P_t) &= (\mu + \beta \sigma_t^2) dt + \sigma_t dW_t \\ &\text{y} \\ d\sigma_t^2 &= -\lambda \sigma_t^2 dt + dZ_{\lambda t} \end{aligned} \tag{7.25}$$

donde $\lambda > 0$ y $\{Z_t\}$ es un proceso de Lévy con incrementos independientes y estacionarios. La volatilidad $\{\sigma_t^2\}$ puede exhibir saltos generados por el proceso de Lévy.

7.9. Estimación de los Modelos en Tiempo Continuo

En esta sección se discuten de manera breve los distintos métodos de estimación de los modelos en tiempo continuo. Se tendrá en cuenta entonces, el problema de la estimación de la ecuación diferencial estocástica a partir de un conjunto de observaciones discretas, teniendo en cuenta lo discutido en la sección dedicada al proceso de Itô. Este problema ha tomado una especial relevancia en los últimos años dada la importancia que han tomado los modelos en tiempo continuo. Diversos son los métodos disponibles en la literatura para la estimación de un proceso de difusión. Como ya se ha comentado antes, dos son los enfoques considerados en el análisis de los modelos en tiempo continuo. El paramétrico y el no-paramétrico. En el primer caso, cuando se dispone de información a suficiente acerca del modelo subyacente, por ejemplo, el modelo pertenece a una familia paramétrica $\{\mathcal{M}_\theta, \theta \in \Theta\}$, donde \mathcal{M}_θ es una forma paramétrica conocida con θ un parámetro desconocido, en este caso el objetivo principal es la estimación de θ . A continuación se describen algunos métodos paramétricos.

7.9.1. Métodos basados en la Función de Verosimilitud

El método de máxima verosimilitud es el método natural si la forma paramétrica del modelo que genera un conjunto de observaciones $\{X_i\}_{0 \leq i \leq n}$ dado, es conocida. Suponiendo que $\{X_i\}_{i \geq 0}$ forman una cadena de Markov estacionaria con densidad invariante $\pi(x, \theta)$ y función de densidad de transición $p(x|y; \theta)$, entonces el logaritmo de la función de verosimilitud es

$$\begin{aligned} \ell(X_0, \dots, X_n; \theta) &= \sum_{i=1}^n \log(p(X_i|X_{i-1}; \theta)) + \log(\pi(X_0; \theta)) \\ &\approx \sum_{i=1}^n \log(p(X_i|X_{i-1}; \theta)) \end{aligned} \quad (7.26)$$

El estimador de máxima verosimilitud *MLE* se obtiene considerando el máximo de la función (7.26)

$$\hat{\theta} = \arg \max_{\theta} \ell(X_0, \dots, X_n; \theta).$$

Se pueden obtener formas explícitas de la función de densidad de transición para algunos modelos, tales como el de Vasicek (7.8) o el modelo *CIR*, sin embargo esto no ocurre con todos los modelos, los cuales no admiten fórmulas cerradas o explícitas para la función de transición. Por ello, es de gran importancia discretizar el proceso aplicando algún método de discretización, por ejemplo, el método de Euler (7.10). Un método alternativo en este sentido es el método presentado en Aït-Sahalia (1999, 2002) donde la función de verosimilitud es aproximada a partir de una sucesión de funciones de verosimilitud basadas en polinomios de Hermite. Para más detalles de este y otros métodos ver Prakasa-Rao (1999) e Iacus (2008).

7.9.2. Método Generalizado de los Momentos

Otro método alternativo es el *método generalizado de los momentos*, *GMM*, (ver Hansen (1982)). Es un método paramétrico muy popular en finanzas. Es básicamente una generalización del método de los momentos que esta basado en la correspondencia de los momentos teóricos y los momentos muestrales. El método se describe a continuación

Se asume que se tiene un proceso estacionario $\{X_t\}_{t \geq 0}$ cuyo mecanismo generador de datos involucra un parámetro θ de dimensión $d \times 1$. Entonces se define una función $u_i = u(X_i, \theta)$ de $r \times 1$ valores, con $r \geq d$, tal que

$$\mathbb{E}(u_i) = \mu, \quad \text{para todo } i,$$

y

$$\text{Cov}(u_i, u_{i+j}) = \mathbb{E}\{(u_i - \mu)(u_{i+j} - \mu)^T\} = S_j, \quad \text{para todo } i, j.$$

Se asume además que la siguiente condición sobre los momentos (o condición de ortogonalidad) se cumple si θ_0 es el valor verdadero del parámetro θ :

$$\mathbb{E}\{u(X_i, \theta)\} = 0 \quad \text{si y solo si} \quad \theta = \theta_0. \quad (7.27)$$

Generalmente, la función $u(\theta)$ es la diferencia entre el k -ésimo momento exacto y X_i^k para alguna potencia de k . Sea

$$g_n(\theta) = \frac{1}{n} \sum_{i=1}^n u(X_i, \theta)$$

la contraparte muestral de la condición (7.27). Se espera que $\mathbb{E}\{g_n(\theta_0)\} = 0$. Se considera además que se satisface la ley fuerte de los grandes números,

$$g_n(\theta) \xrightarrow{\text{c.s.}} \mathbb{E}\{u(X_i, \theta)\}$$

uniformemente en θ . Así pues, el parámetro estimado $\hat{\theta}$, es la solución del problema de optimización

$$\hat{\theta} = \arg \min_{\theta} Q(\theta) = \arg \min_{\theta} g_n(\theta)^T W g_n(\theta), \quad (7.28)$$

donde W es una matriz de pesos dada y que además es definida positiva. Una selección particular de esta matriz W es S^{-1} , donde $S = \mathbb{E}\{uu^T\}$ es la matriz de covarianza a largo plazo. Es claro que en este método es importante la selección de la matriz W , sin embargo, no es el objetivo de este trabajo discutir ese tema. Para más detalles, ver Conley et al. (1997) y las referencias incluidas en dicho trabajo.

Ciertamente los métodos antes descritos no son los únicos que se aplican para la estimación de los parámetros de los distintos modelos descritos en este capítulo, sin embargo no es del interés de esta memoria

abordar la diversidad de estos. Cabe destacar, como se verá más adelante, que los métodos de estimación deben satisfacer condiciones del tipo \sqrt{n} -consistente. En tal sentido, el método de máxima verosimilitud introducido por Aït-Sahalia (2002) satisface dicho requerimiento. Tal método será descrito con cierto detalle posteriormente en el capítulo referido a los modelos de tipos de interés.

Respecto a los métodos no paramétricos, un conjunto de referencias que pueden servir como apoyo son Gao and King (2004), Arapis and Gao (2006) e Iacus (2008).

Apéndices

A

Nociones básicas de Probabilidad

A.1. Espacio de Probabilidad y Variable Aleatoria

Un *espacio de probabilidad* es un terna $(\Omega, \mathcal{F}, \mathbb{P})$ donde Ω es el *espacio muestral* de todos los posibles resultados de un experimento aleatorio; \mathcal{F} es una σ -*álgebra de conjuntos*, es decir, \mathcal{F} es una colección de conjuntos que satisface:

- (i.) El conjunto \emptyset pertenece a \mathcal{F}
- (ii.) Si $A \in \mathcal{F}$, entonces su complemento $A^c \in \mathcal{F}$
- (iii.) Si $A_1, A_2, \dots, \in \mathcal{F}$ entonces

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

en el contexto de los espacios de Probabilidad, los elementos $A \in \mathcal{F}$, pertenecientes a la σ -álgebra son denominados *eventos*, por ello, \mathcal{F} suele ser llamada también la σ -*álgebra de eventos*. En este orden de ideas se define ω como los elementos en Ω , así pues un evento $A \subset \Omega$, es una colección de puntos muestrales. \mathbb{P} es una *medida de probabilidad* en (Ω, \mathcal{F}) , es decir, \mathbb{P} es una función definida en Ω tal que:

- (i.) $0 \leq \mathbb{P}(A) \leq 1$ para todo $A \in \mathcal{F}$,
- (ii.) $\mathbb{P}(\Omega) = 1$ y $\mathbb{P}(\emptyset) = 0$,
- (iii.) Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Definición A.1.1. Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, se define como variable aleatoria a una función medible X tal que

$$\begin{aligned} X : \Omega &\longmapsto \mathbb{R} \\ \omega &\longmapsto X(\omega) = x \end{aligned}$$

El concepto de función medible El término de función medible, es referido a la posibilidad de calcular probabilidades relativas a la variable aleatoria X . En el ámbito de los números reales, se denota por $\mathcal{B}(\mathbb{R})$ la σ -álgebra de conjuntos de Borel (la σ -álgebra generada por los conjuntos abiertos de \mathbb{R}) y dada X^{-1} la función inversa de X , entonces se entenderá X por una *función medible* si para todo conjunto A perteneciente a la σ -álgebra de Borel, $\mathcal{B}(\mathbb{R})$, existe un conjunto B en \mathcal{F} tal que

$$X^{-1}(A) = B;$$

de tal manera que siempre es posible medir el conjunto de valores asumidos por X utilizando la medida de probabilidad \mathbb{P} en el espacio Ω ,

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(\{\omega \in \Omega : \omega \in X^{-1}(A)\}),$$

para $A \in \mathcal{B}(\mathbb{R})$ y $B \in \mathcal{F}$.

Definición A.1.2. La función

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X(\omega) \in (-\infty, x]) = P(X \leq x),$$

para $x \in \mathbb{R}$, es conocida como la función de distribución acumulada o simplemente función de distribución respecto de la variable X .

La definición anterior nos da la posibilidad de definir la probabilidad de que x pertenezca a un intervalo $(a, b]$, es decir,

$$P(X(\omega) \in (a, b]) = F_X(b) - F_X(a), \quad a < .b$$

Puesto que se ha asumido que X es una función medible, el evento $\{\omega \in \Omega : X(\omega) \in B\}$ está también en \mathcal{F} , de manera que su medida de probabilidad \mathbb{P} está bien definida y por tanto F_X . La función F_X es una función no decreciente tal que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1,$$

y es además una función continua por la derecha.

Si F es absolutamente continua, su derivada

$$\frac{d}{dx} F_X(x) = f_X(x)$$

donde f_X es llamada una *función de densidad*, la cual es una función no negativa Lebesgue Integrable cuya integral en la recta real es igual a uno. Dicho de una manera práctica, Si $F_X(x)$ es la probabilidad de que una variable aleatoria tome valores menores o iguales a x , la cantidad $f_X(x)dx$ puede ser interpretada como la probabilidad de que la variable aleatoria tome valores en el intervalo infinitesimal $[x, x + dx)$. En este caso,

$$F_X(x) = \int_{-\infty}^x f_X(u)du, \quad x \in \mathbb{R}$$

donde f_X es la función de densidad tal que $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$ y $\int_{-\infty}^{\infty} f(x)dx = 1$.

Si la variable aleatoria toma valores solamente en un conjunto numerable, entonces se dice que es discreta y su densidad en el punto x se define por $P(X = x)$. En el caso de que X sea una variable continua, $P(X = x) = 0$ para todo x ,

$$\lim_{h \rightarrow 0} F_X(x + h) = F_X(x), \text{ para todo } x.$$

Definición A.1.3. Para un valor dado de probabilidad p , se define como el p -ésimo cuantil de la variable X , al menor número real x_p tal que $p \leq F_X(x_p)$, más concretamente,

$$x_p = \inf_x \{x : p \leq F_X(x)\}.$$

A.2. Valor Esperado, Varianza y Momentos

Definición A.2.1. La media o valor esperado de una variable aleatoria continua X con distribución F_X se define como

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x dF_X(x)$$

siempre que la integral sea finita ($X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$).

Si X tiene una densidad, entonces

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx$$

y la integral es la integral de Riemann; en otro caso las integrales en los casos $d\mathbb{P}$ o dF_X se puede pensar como integrales en el sentido abstracto. Si Ω es numerable, el valor esperado se define por

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega)$$

o de manera equivalente, cuando X es una variable aleatoria discreta, entonces

$$\mathbb{E}(X) = \sum_{s \in I} x \mathbb{P}(X = x)$$

donde I es el conjunto de los posibles valores de X .

Definición A.2.2. La varianza de una variable aleatoria continua X se define por

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \int_{\Omega} (X(\omega) - \mathbb{E}(X))^2 d\mathbb{P}(\omega),$$

por supuesto siempre que la integral sea finita ($X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$).

Definición A.2.3. Se define el k -ésimo momento de una variable aleatoria continua X con distribución F_X como

$$\mathbb{E}(X^k) = \int_{\Omega} X^k(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x^k dF_X(x)$$

siempre que la integral sea finita.

En general, para cualquier función medible $g(\cdot)$ se define

$$\mathbb{E}(g(X)) = \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega)$$

siempre que la integral sea finita (o bien $g(X)$ sea una función integrable).

A.3. Variable Aleatoria n -dimensional

Se define $\mathbf{X} = (X_1, X_2, \dots, X_n)$ como una *variable aleatoria* n -dimensional (*vector aleatorio*) si sus componentes X_i son variables aleatorias unidimensionales.

La función de distribución de la variable \mathbf{X} y que se denota por $F_{\mathbf{X}}$ se define como

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Análogamente a las variables aleatorias se pueden introducir los conceptos antes descritos para vectores aleatorios. En lo que respecta a esta memoria, los vectores aleatorios o variables aleatorias n -dimensionales continuas con una densidad serán el objeto de estudio y por tanto se enfocará la atención en estos conceptos.

Si la función de distribución de un vector aleatorio \mathbf{X} tiene función de densidad $f_{\mathbf{X}}$, entonces la función de distribución se puede representar por

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(u_1, \dots, u_n) du_1 \dots du_n,$$

con $(x_1, \dots, x_n) \in \mathbb{R}^n$, donde la densidad es una función que satisface $f_{\mathbf{X}}(\mathbf{x}) \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$ y

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, \dots, u_n) du_1 \dots du_n = 1.$$

Definición A.3.1. Se define la función de distribución marginal de la variable aleatoria X_i como

$$F_{X_i}(x_i) = F_{\mathbf{X}_i}(\infty, \dots, x_i, \dots, \infty) \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, \dots, x_i, \dots, u_n) du_1 \dots du_n,$$

Además si el vector \mathbf{X} tiene función de densidad $f_{\mathbf{X}}$ entonces se define la función de densidad marginal

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, \dots, u_n) du_2 \dots du_n,$$

La esperanza de una variable aleatoria n -dimensional tiene una forma similar a la esperanza de una variable aleatoria unidimensional. Los valores $X(\omega)$ se concentran alrededor de ésta.

Definición A.3.2. La esperanza o valor esperado de un vector aleatorio \mathbf{X} viene dado por

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n)).$$

Definición A.3.3. La matriz de covarianzas de \mathbf{X} denotada por $\sigma_{\mathbf{X}}$, se define por

$$\Sigma_{\mathbf{X}} = (\text{Cov}(X_i, X_j) : i, j = 1, \dots, n)$$

donde $\text{Cov}(X_i, X_j) = \mathbb{E}\{(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\}$ es a covarianza de X_i y X_j . Si $i = j$, entonces $\text{Cov}(X_i, X_j) = \sigma_{X_i}^2$

Ejemplo A.3.1. Una variable aleatoria n -dimensional Normal o Gaussiana tiene una distribución Gaussiana con función de densidad

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)\Sigma(\mathbf{x} - \mu)^T\right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

con parámetros $\mu \in \mathbb{R}^n$, y Σ . La cantidad Σ es una matriz $(n \times n)$ definida positiva simétrica, Σ^{-1} es su inversa y $\det\Sigma$, su determinante. Se escribe como $N(\mu, \Sigma)$ la distribución de una variable aleatoria Gaussiana n -dimensional \mathbf{X} .

A.4. Independencia

Definición A.4.1. Dos variables aleatorias X e Y son independientes si

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

para cualesquiera dos conjuntos A y B en \mathbb{R}

Alternativamente se puede definir la independencia mediante las funciones de distribución y las funciones de densidad. Las variables aleatorias X_1 y X_2 son independientes si y sólo si:

$$F_{X_1 X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2), \quad x_1, x_2 \in \mathbb{R}.$$

Si (X_1, X_2) tiene función de densidad f_{X_1, X_2} con funciones de densidad marginales f_{X_1} y f_{X_2} . Entonces las variables aleatorias X_1 y X_2 son independientes si y sólo si:

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2), \quad x_1, x_2 \in \mathbb{R}.$$

Una consecuencia importante de la independencia de variables aleatorias es la propiedad siguiente:

Si X_1, \dots, X_n son variables aleatorias independientes, entonces para cualesquiera funciones reales g_1, \dots, g_n

$$E[g_1(X_1), \dots, g_n(X_n)] = E[g_1(X_1)] \dots E[g_n(X_n)]$$

siempre que las esperanzas consideradas estén bien definidas.

Definición A.4.2. Se define la correlación de X_1 y X_2 como

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}.$$

Como se puede ver, es el resultado de estandarizar la covarianza de las variables X_1 y X_2 , por lo que su valor está comprendido entre -1 y 1 .

En particular, se puede concluir que las variables aleatorias independientes X_1 y X_2 son no correlacionadas, esto es, $\text{Corr}(X_1, X_2) = 0$. En general, el recíproco no es cierto, es decir, las variables aleatorias no correlacionadas no son necesariamente independientes.

A.5. Distribución Condicional

Sean dos variables aleatorias X_1 e X_2 , con funciones de distribución F_{X_1} y F_{X_2} y con función de distribución conjunta F_{X_1, X_2} , a continuación se establecen los conceptos de distribución condicional.

Definición A.5.1. La función de distribución condicional de X_1 dado X_2 se expresa por

$$F_{X_1|X_2}(x_1|x_2) = \frac{P(X \leq x_1, X_2 \leq x_2)}{P(X_2 \leq x_2)} = \frac{F_{X_1, X_2}(x_1, x_2)}{F_{X_2}(x_2)}$$

Si las funciones de densidad de las variables X_1 e X_2 existen entonces, la función de densidad condicional es

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

Las variables aleatorias X_1 e X_2 son independientes si y sólo si

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1)$$

A.6. Esperanza Condicional

Sean A y B dos eventos definidos en un espacio de probabilidades $(\Omega, \mathcal{F}, \mathbb{P})$. La esperanza condicional de A dado B se define como

$$P(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{para } \mathbb{P}(B) > 0. \quad (\text{A.1})$$

De la misma forma, tal y como se hizo para variables aleatorias, se define la distribución condicional de la variable aleatoria X con respecto al evento B como

$$F_X(x|B) = \frac{\mathbb{P}(X \leq x \cap B)}{\mathbb{P}(B)}, \quad x \in \mathbb{R} \quad \text{y para } \mathbb{P}(B) > 0. \quad (\text{A.2})$$

Entonces se puede definir la esperanza condicional de una variable aleatoria dado un evento particular.

Definición A.6.1. Se define como la esperanza condicional de la variable aleatoria X dado el evento B , respecto de la distribución condicional, como

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}, \quad (\text{A.3})$$

donde $\mathbb{1}_B$ es la función indicadora del conjunto B ,

$$\mathbb{1}_B(\omega) = \begin{cases} 1 & \text{si } \omega \in B \\ 0 & \text{si } \omega \notin B \end{cases}$$

Para variables aleatorias discretas, la esperanza condicional toma la forma

$$\mathbb{E}(X|B) = \sum_i x_i \frac{\mathbb{P}(\{\omega : X(\omega) = x_i \cap B\})}{\mathbb{P}(B)} = \sum_i x_i \mathbb{P}(X = x_i|B).$$

Para una variable aleatoria con función de densidad f_X , se tiene que

$$\mathbb{E}(X|B) = \frac{1}{P(B)} \int_{\mathbb{R}} x \mathbb{1}_B(x) f_X(x) dx = \frac{1}{P(B)} \int_B x f_X(x) dx.$$

Ahora bien, sea Y una variable aleatoria discreta que toma distintos valores en los conjuntos A_i , es decir,

$$A_i = A_i(\omega) = \{\omega : Y(\omega) = y_i\}, i = 1, 2, \dots$$

, tal que $P(A_i)$ es positiva para todo i . Sea $\mathbb{E}|X| < \infty$. Entonces se define una nueva variable aleatoria Z como

$$Z(\omega) = \mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|A_i(\omega)) = \mathbb{E}(X|Y(\omega) = y_i), \quad \omega \in A_i$$

Para cada $\omega \in A_i$ fijo, la esperanza condicional $\mathbb{E}(X|Y)$ coincide con $\mathbb{E}(X|A_i)$, pero, en su conjunto, es una variable aleatoria en sí misma porque depende de los eventos generados por Y .

Si en lugar de A_i se considera una σ -álgebra de eventos (por ejemplo, la σ -álgebra generada por la variable aleatoria Y), entonces para una variable aleatoria X tal que $\mathbb{E}|X| < \infty$, se define como la *esperanza condicional* de X con respecto a la σ -álgebra \mathcal{F} a la variable aleatoria Z si,

- Z es \mathcal{F} -medible y
- Z es tal que $\mathbb{E}(Z\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A)$ para cada $A \in \mathcal{F}$.

La esperanza condicional es única y se denota por $Z = \mathbb{E}(X|\mathcal{F})$. Con esta notación, la equivalencia anterior se puede escribir como

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(X\mathbb{1}_A) \quad \text{para cada } A \in \mathcal{F}. \tag{A.4}$$

Sean X, X_1, X_2 variables aleatorias y a, b dos constantes, entonces

$$\begin{aligned} \mathbb{E}(aX_1 + bX_2|\mathcal{F}) &= a\mathbb{E}(X_1|\mathcal{F}) + b\mathbb{E}(X_2|\mathcal{F}), \\ \mathbb{E}(X|\mathcal{F}_0) &= \mathbb{E}(X), \end{aligned}$$

si $\mathcal{F}_0 = \{\Omega, \emptyset\}$. Más aún, si Y es \mathcal{F} -medible, entonces

$$\mathbb{E}(YX|\mathcal{F}) = Y\mathbb{E}(X|\mathcal{F}),$$

y seleccionando $X = 1$, se obtiene que

$$a\mathbb{E}(Y|\mathcal{F}) = Y.$$

Finalmente, si se selecciona $A = \Omega$ en (A.4) entonces se obtiene que

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})] = \mathbb{E}(X).$$

Si X es independiente de \mathcal{F} , entonces $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$ y, en particular, si X y Y son independientes, se tiene que $\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y)) = \mathbb{E}(X)$, donde $\sigma(Y)$ es la σ -álgebra generada por la variable aleatoria Y .

A.7. Tipos de Convergencia

Sea $\{F_n\}_{n \in \mathbb{N}}$ una sucesión de funciones de distribución para la sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$. Si se asume que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \tag{A.5}$$

para todo $x \in \mathbb{R}$ tal que $F(\cdot)$ es continua en x . F es la función de distribución de alguna variable aleatoria X , entonces la sucesión X_n se dice que *converge en distribución* a la variable aleatoria X , lo cual se denota por

$$X_n \xrightarrow{d} X. \tag{A.6}$$

Tal convergencia esta relacionada con el comportamiento probabilístico de las variables aleatorias en algún intervalo $(-\infty, x]$, $x \in \mathbb{R}$.

Un sucesión de variables aleatorias X_n se dice que *converge en probabilidad* a una variable aleatoria X si, para cualquier $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \tag{A.7}$$

y se denota por

$$X_n \xrightarrow{p} X \tag{A.8}$$

y es una convergencia puntual de las probabilidades. Esta convergencia implica la convergencia en la distribución. En ciertas ocasiones suele usarse la notación

$$p - \lim_{n \rightarrow \infty} |X_n - X| \geq \epsilon = 0, \quad (\text{A.9})$$

para la convergencia en probabilidad. Un tipo de convergencia más fuerte se define como la probabilidad del límite

$$P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1. \quad (\text{A.10})$$

En este caso se dice que X_n converge a X casi seguramente y se denota por

$$X_n \xrightarrow{c.s.} X. \quad (\text{A.11})$$

Esta convergencia implica la convergencia en probabilidad.

Por otro lado, una sucesión de variables aleatorias X_n se dice que *converge en media de orden r* a una variable aleatoria X si

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^r) = 0, \quad r \geq 1. \quad (\text{A.12})$$

La convergencia en media de orden r implica la convergencia en probabilidad gracias a la desigualdad de Chebyshev, y si X_n converge a X en media de orden r , entonces también converge en media de orden s , para todo $r > s \geq 1$. LA convergencia en media de orden 2 o conocida como *convergencia en media cuadrática* es un caso particular de este tipo de convergencia para $r = 2$ y es de gran interés en estadística.

B

Aspectos Generales de los Procesos Estocásticos

B.1. Procesos Estocásticos

Definición B.1.1. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidades. Un proceso estocástico con valores reales es una familia de variables aleatorias $\{X_\gamma, \gamma \in \Gamma\}$ (Γ subconjunto de índices) definidas en $\Omega \times \Gamma$ tomando valores en \mathbb{R} .

De esta manera, las variables aleatorias de la familia (medibles para todo $t \in \Gamma$) son funciones de la forma

$$X(\omega, \gamma) : \Omega \times \Gamma \longrightarrow \mathbb{R}.$$

Para $\Gamma = \mathbb{N}$, tenemos un proceso en *tiempo discreto* y para $\Gamma \subset \mathbb{R}$ tenemos un proceso en *tiempo continuo*. En lo que respecta a este trabajo, es de interés principal considerar como subconjunto de índices, $\Gamma = (0, +\infty]$.

- Para un valor fijo de ω , por ejemplo $\tilde{\omega}$,

$$\{X(\tilde{\omega}, t), t \geq 0\}$$

se le conoce como la *trayectoria* del proceso (la serie de tiempo), la cual representa una posible evolución del proceso.

- Para un valor fijo de t , por ejemplo, \tilde{t} , el conjunto de valores

$$\{X(\omega, \tilde{t}), \omega \in \Omega\}$$

representa una variable aleatoria a valores reales. Por ejemplo, para el precio de un activo en el instante t , el rango de $X(\omega, \tilde{t})$ es el conjunto de números reales no-negativos.

Por simplicidad, se denotará por $\{X_t, t \geq 0\} \equiv \{X_t\}$ al proceso en tiempo continuo, entendiendo que para t , X_t es una variable aleatoria.

B.2. Filtraciones

Definición B.2.1. *Considérese un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Una filtración $\{\mathcal{F}_t, t \geq 0\}$ es una familia creciente de sub- σ -álgebras de \mathcal{F} indexadas por $t \geq 0$; es decir, para cada $s, t \geq 0$ tal que $s < t$, se tiene $\mathcal{F}_s \subset \mathcal{F}_t$ con $\mathcal{F}_0 = \{\Omega, \emptyset\}$.*

Para cada proceso $\{X_t, t \geq 0\}$ y para cada t , se puede asociar una σ -álgebra denotada por $\mathcal{F}_t = \sigma(X_s; 0 \leq s \leq t)$, y que además es la σ -álgebra generada por el proceso X hasta el instante t ; es decir, la σ -álgebra más pequeña (minimal) de \mathcal{F} que hace a $X(s, \omega)$ medible para cada $0 \leq s \leq t$. Dicha σ -álgebra es el conjunto de subconjuntos más pequeño de Ω que hace posible asignar probabilidades a todos los eventos relacionados con el proceso X hasta el instante t .

Definición B.2.2. *Dado un proceso estocástico $\{X_t, t \geq 0\}$ y una filtration $\{\mathcal{F}_t, t \geq 0\}$ (no necesariamente la que genera X), el proceso X se denomina adaptado a $\{\mathcal{F}_t, t \geq 0\}$ (\mathcal{F}_t -adaptado) si para cada $t \geq 0$ $X(t)$ es \mathcal{F}_t -medible.*

En otras palabras, $X = \{X_t, t \geq 0\}$ es \mathcal{F}_t -adaptado cuando el valor de X_t en el tiempo t solo depende de la información contenida en la trayectoria hasta el instante t .

B.3. Momentos, Covarianza e Incrementos de un Proceso Estocástico

El valor esperado y varianza de un proceso estocástico son definidos por

$$\mathbb{E}(X_t) = \int_{\Omega} X(\omega, t) d\mathbb{P}(\omega), \quad t \in [0, T],$$

y

$$\text{Var}(X_t) = \mathbb{E} [X_t - \mathbb{E}(X_t)]^2 \quad t \in [0, T].$$

El k -ésimo momento de $X_t, k \geq 1$, se define, para $t \in [0, T]$, como $\mathbb{E}(X_t^k)$. Esas cantidades están bien definidas cuando las correspondientes integrales son finitas. La *función de covarianza* del proceso para dos instantes de tiempo t y s se define como

$$\gamma(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}\{(X_t - \mathbb{E}(X_t))(X_s - \mathbb{E}(X_s))\}.$$

La cantidad $X_t - X_s$ es llamada el proceso de *incrementos* desde s a t , con $s < t$.

Tales cantidades son útiles en la descripción de los procesos estocásticos que son introducidos para modelizar la evolución de algunos ciertos impulsos “*shocks*” estocásticos.

B.4. Variación de un Proceso

Sea $\Pi_n = \Pi_n([0, t]) = \{0 = t_0 < t_1 < \dots < t_i < \dots < t_n = t\}$ cualquier partición de un intervalo $[0, t]$ en n intervalos y denotada por

$$\|\Pi_n\| = \max_{j=0, \dots, n-1} (t_{j+1} - t_j)$$

el máximo tamaño de paso de discretización de la partición Π_n .

Definición B.4.1. La variación del proceso X se define como

$$V_t(X) = p - \lim_{\|\Pi_n\|} \sum_{k=0}^{n-1} |X_{t_{k+1}} - X_{t_k}|. \tag{B.1}$$

Si X es diferenciable, entonces $V_t(X) = \int_0^t |X'(u)| du$. Si $V_t(X) < \infty$, entonces se dice que X es de *variación acotada* en $[0, t]$. Si lo anterior es cierto para todo $t \geq 0$, entonces se dice que X tiene variación acotada.

Definición B.4.2. La variación cuadrática de un proceso X , denotada por $[X, X]_t$ se define como

$$[X, X]_t = p - \lim_{\|\Pi_n\|} \sum_{k=0}^{n-1} |X_{t_{k+1}} - X_{t_k}|^2. \quad (\text{B.2})$$

Para procesos estocásticos con trayectorias continua, el límite existe, y en dicho caso la notación usualmente adoptada es $\langle X, X \rangle_t$, y se puede definir alternativamente como

$$\langle X, X \rangle_t = p - \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left(X_{t \wedge k/2^n} - X_{t \wedge (k-1)/2^n} \right)^2,$$

donde $a \wedge b = \min(a, b)$. Si X es continuo y tiene variación cuadrática finita, entonces su total variación es necesariamente infinita. Además se debe tener en cuenta que $V_t(X)$ y $[X, X]_t$ son también procesos estocásticos.

B.5. Martingalas

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una filtración $\{\mathcal{F}_t, t \geq 0\}$, entonces se define el como un *espacio de probabilidad filtrado* a la tupla $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$.

Definición B.5.1. Sea $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. un espacio de probabilidad filtrado. Un proceso X_t con $t \in \Gamma$, $\Gamma \subset \mathbb{R}$ un conjunto de índices, es una martingala relativo a la filtración $\{\mathcal{F}_t, t \geq 0\}$, siempre que:

- (i.) X_t es adaptado a la filtración $\{\mathcal{F}_t, t \geq 0\}$
- (ii.) X_t es integrable, es decir, $\mathbb{E}|X_t| < \infty$,
- (iii.) Para cualesquiera s y t con $s < t$

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s, \quad \text{casi seguramente.}$$

Un concepto de gran utilidad en la teoría de los procesos estocásticos es el relativo a procesos integrables. A continuación se introducen algunas definiciones acerca de este tema. Para apreciar estos conceptos en el contexto de los procesos estocásticos es importante recordar que una variable aleatoria es integrable si $\mathbb{E}|X| < \infty$.

Definición B.5.2. Una variable aleatoria X es cuadrado integrable si $\mathbb{E}(X^2) < \infty$. Un proceso X_t en el intervalo $[0, T]$, donde T puede ser infinito, es cuadrado integrable si

$$\sup_{t \in [0, T]} \mathbb{E}(X_t^2) < \infty,$$

en otras palabras, si sus segundos momentos son acotados.

Definición B.5.3. Un proceso X_t , $0 \leq t \leq T$ es llamado uniformemente integrable si

$$\mathbb{E}(|X_t| \mathbb{1}_{\{|X_t| > n\}})$$

converge a zero cuando $n \rightarrow \infty$ uniformemente en t .

B.6. Propiedad de Markov

La propiedad de Markov establece que si se conoce el estado actual de un proceso estocástico, entonces el comportamiento futuro de dicho proceso es independiente de su pasado. El proceso X_t tiene la *propiedad de Markov* si la distribución condicional del proceso X_{t+s} dado el proceso en el instante $X_t = x$, no depende de los valores pasados (pero esta depende del valor actual). Sea \mathcal{F}_t la σ -álgebra generada por el proceso hasta el tiempo t .

Definición B.6.1. X es un proceso de Markov si para cualquier t y $s > 0$,

$$\mathbb{P}(X_{t+s} \leq y | \mathcal{F}_t) = \mathbb{P}(X_{t+s} \leq y | X_t) \quad \text{c.s.}$$

Definición B.6.2. La función de transición de probabilidad de un proceso X se define como

$$\mathbb{P}(y, t, x, s) = \mathbb{P}(X_t \leq y | X_s = x)$$

la función de distribución condicional del proceso en el instante t , dado que este está en el punto x en el instante $s < t$.

La propiedad de Markov implica una expresión bastante útil en términos de la esperanza condicionada bajo la σ -álgebra de eventos. Esta relación es válida tanto para un proceso en tiempo discreto como en tiempo continuo. Se considerará el caso continuo dada su importancia en este trabajo. Sea un proceso estocástico separable $X = \{X_t, t \geq 0\}$

definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, adaptado a una filtración $\{\mathcal{F}_t, t \geq 0\}$. Entonces para cada $t \geq 0$ se define \mathcal{F}_t^+ la sub- σ -álgebra de \mathcal{F} generada por la totalidad de los subconjuntos de la forma

$$A = \{\omega \in \Omega : X_{s_i}(\omega) \in L_i \text{ for } i = 1, 2, \dots, n\} \quad (\text{B.3})$$

de Ω para cualquier $t \leq s_1 < s_2 < \dots < s_n$ y $L_1, L_2, \dots, L_n \in \mathcal{L}$ where $n = 1, 2, 3, \dots$. En consecuencia, X_s es \mathcal{F}_t^+ : \mathcal{L} -medible para cada $s \geq t$ y, en efecto, \mathcal{A}_t^+ es la menor σ -álgebra con esta propiedad. Esta es la colección de todos los eventos detectable por el proceso estocástico en el instante futuro si t es considerado como el presente. La propiedad de Markov implica que

$$\mathbb{E}(Y|\mathcal{F}_t) = \mathbb{E}(Y|X_t), \quad \text{con probabilidad 1} \quad (\text{B.4})$$

para todo $t \geq 0$ y para todo $\mathcal{F}_t^+ : \mathcal{L}$ -medible Y .

C

Elementos de Cálculo Estocástico

C.1. Integración Estocástica

De manera intuitiva, un proceso estocástico X es una difusión si su dinámica local se puede aproximar mediante la ecuación en diferencias del siguiente tipo:

$$X_{t+\Delta t} - X_t = \mu(t, X_t)\Delta t + \sigma(t, X_t)Z_t \quad (\text{C.1})$$

donde Z_t es un término de perturbación con distribución normal, el cual es independiente de todo lo que haya sucedido hasta el instante t , y μ y σ son dos funciones deterministas. La interpretación de (C.1) es que, en el intervalo de tiempo $[t, t + \Delta t]$, el proceso X es caracterizado por dos términos: Uno, la función $\mu(t, X_t)$ que determina localmente la velocidad de forma determinista, y un segundo término de perturbación de tipo gaussiano el cual se ve amplificado por la función $\sigma(t, X_t)$. La función μ es denominada el término de *drift* del proceso, mientras que σ es denominada el término de *difusión*.

En la modelización del término de perturbación Gaussiano se suele usar el proceso de Wiener.

Definición C.1.1. Un proceso estocástico W es llamado un proceso de Wiener si satisface las siguientes condiciones:

1. $W(0) = 0$.
2. El proceso tiene incrementos independientes, es decir, si $r < s \leq t < u$ entonces $W_u - W_t$ y $W_s - W_r$ son variables aleatorias independientes.
3. Para $s < t$ la variable aleatoria $W_t - W_s$ tiene distribución Gaussiana $N(0, t - s)$.
4. W tiene trayectorias continuas.

De esta manera, el proceso de Wiener se puede usar para expresar

$$X_{t+\Delta t} - X_t = \mu(t, X_t)\Delta t + \sigma(t, X_t)\Delta W_t \quad (\text{C.2})$$

donde ΔW_t se define por

$$\Delta W_t = W_{t+\Delta t} - W_t.$$

Ahora bien, dividiendo por Δt la ecuación (C.2) y haciendo $\Delta t \rightarrow 0$ se obtiene

$$\dot{X}_t = \lim_{\Delta t \rightarrow 0} \frac{X_{t+\Delta t} - X_t}{\Delta t} = \mu(t, X_t) + \sigma(t, X_t)v_t, \quad (\text{C.3})$$

$$X_0 = x_0 \quad (\text{C.4})$$

donde se ha agregado una condición inicial y además

$$v_t = \frac{dW_t}{dt}$$

es la derivada respecto del tiempo del proceso de Wiener. Al parecer el problema esta prácticamente resuelto, bastaría con resolver la ecuación diferencial ordinaria. Sin embargo, lo anterior no es posible, puesto que el proceso v no está bien definido. Se puede demostrar que con probabilidad 1, que las trayectorias del proceso de Wiener son no diferenciables para todo punto.

Una forma de hacer una descripción más precisa de (C.2) es hacer tender $\Delta \rightarrow 0$ sin necesidad de dividir la ecuación por Δt . Formalmente se obtiene entonces la expresión

$$\begin{cases} dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \\ X_0 = x_0 \end{cases} \quad (\text{C.5})$$

y ahora es natural interpretar (C.5) como una versión abreviada de la siguiente ecuación integral

$$X_t = x_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s. \tag{C.6}$$

En la ecuación (C.6) se puede interpretar la primera integral del lado derecho como una simple integral de Riemann. La interpretación natural de la segunda integral es asumirla como una integral de Riemann-Stieltjes para cada trayectoria, pero desafortunadamente esto no es posible ya que se puede demostrar que las trayectorias de W son localmente de variación no acotada. Así, la segunda integral denominada integral estocástica no se puede definir de forma simple.

C.1.1. La Integral de Itô

La ecuación (C.6) introduce la *integral estocástica* o *integral de Itô*

$$I(X) = \int_0^T X_u dW_u$$

con respecto al movimiento Browniano. Esta integral resulta bastante sencilla de definir para *procesos simples* (procesos constantes a trozos). Sin embargo, para procesos genéricos requiere algunas consideraciones. En esta sección daremos algunas detalles de la construcción de la integral estocástica, sin llegar a profundizar en los mismos.

Dada una función genérica g (integrando)

$$g : [0, T] \times \Omega \rightarrow \mathbb{R}$$

$I(g)$ se define como el límite de la sucesión de integrales $I(g^{(n)})$ donde $g^{(n)}$ es llamado un *proceso simple*, definido como

$$g^{(n)}(t, \omega) = g(t_j, \omega), \quad t_j \leq t < t_{j+1},$$

con $t_j \in \Pi_n([0, 1])$ (ver Apéndice (B)) y tal que $\|\Pi_n\| \rightarrow 0$ cuando $n \rightarrow \infty$. Se puede demostrar que $g^{(n)}$ converge a g en media cuadrática.

Entonces $I(g^{(n)})$ se define como

$$\begin{aligned} I(g^{(n)}) &= \sum_{j=0}^{n-1} g^{(n)}(t_j) \{W_{t_{j+1}} - W_{t_j}\} \\ &= \sum_{j=0}^{n-1} g(t_j) \{W_{t_{j+1}} - W_{t_j}\}. \end{aligned} \tag{C.7}$$

La ecuación no converge en el sentido habitual cuando W no tiene variación finita. Por el contrario, si se considera la convergencia en media cuadrática, el límite existe. En efecto, para cada n , se tiene que

$$\mathbb{E} \left\{ I(g^{(n)}) \right\}^2 = \sum_{j=0}^{n-1} \mathbb{E} (g(t_j))^2 (t_{j+1} - t_j),$$

de donde se obtiene que

$$I(g^{(n)}) \rightarrow I(g)$$

en media cuadrática, con límite único.

Sea $\{X_t, 0 \leq t \leq T\}$ un proceso estocástico adaptado a la filtración generada por el proceso de Wiener, y tal que

$$\int_0^T \mathbb{E}((X_s)^2) ds < \infty.$$

La integral estocástica del proceso X se define como

$$I_t(X) = \int_0^t X_s dW_s = \lim_{\|\Pi_n\| \rightarrow 0} \sum_{i=0}^{n-1} X_{t_i} (W_{t_{i+1}} - W_{t_i}), \tag{C.8}$$

donde la convergencia es en media cuadrática y $t_i \in \Pi_n$.

C.1.2. Propiedades de la Integral de Itô

- Si X es Itô integrable, entonces

$$\mathbb{E} \left(\int_0^T X_s dW_s \right) = 0$$

y

$$\text{Var} \left(\int_0^T X_s dW_s \right) = \int_0^T \mathbb{E}(X_s)^2 dW_s = 0, \quad (\text{Isometría de It\^o})$$

- Si X e Y son dos procesos integrables según It\^o y α y β son dos constantes entonces la propiedad de *linealidad* se expresa por

$$\int_0^T (\alpha X_s + \beta Y_s) dW_s = \alpha \int_0^T X_s dW_s + \beta \int_0^T Y_s dW_s.$$

- A partir de la expresión en el ítem anterior

$$\int_0^T \alpha dW_s = \alpha \int_0^T dW_s = \alpha W_T$$

- Se puede probar entonces que

$$\int_0^T W_s dW_s = \frac{1}{2} W^2(T) - \frac{1}{2} T.$$

Para ello, sea $t_j \in \Pi_n([0, T])$ una partición de $[0, T]$, y sea

$$X_t^n = \sum_{i=0}^{n-1} W_{t_i} \mathbb{1}_{(t_i, t_{i+1}]}(t).$$

Entonces para cualquier n , X_t^n es un proceso simple adaptado. Por continuidad de W_t , $\lim_{n \rightarrow \infty} X_t^n = W_t$ casi seguramente cuando $\max_i(t_{i+1} - t_i) \rightarrow 0$. La integral de It\^o (estocástica) de X_t^n está dada por

$$\int_0^T X_t^n dW_t = \sum_{i=0}^{n-1} W_{t_i} (W_{t_{i+1}} - W_{t_i}).$$

Operando algebraicamente (sumando y restando $W_{t_{i+1}}^2$) se obtiene

$$W_{t_i} (W_{t_{i+1}} - W_{t_i}) = \frac{1}{2} \left(W_{t_{i+1}}^2 - W_{t_i}^2 - (W_{t_{i+1}} - W_{t_i})^2 \right)$$

y

$$\begin{aligned} \int_0^T X_t^n dW_t &= \frac{1}{2} \sum_{i=0}^{n-1} \left(W_{t_{i+1}}^2 - W_{t_i}^2 \right) - \frac{1}{2} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 \\ &= \frac{1}{2} W_T^2 - \frac{1}{2} W_0^2 - \frac{1}{2} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2, \end{aligned}$$

puesto que el primer sumando de la derecha es una suma telescópica. En cuanto a la segunda formula, se puede demostrar que la variación cuadrática de un proceso de Wiener (movimiento Browniano) converge en probabilidad a T . Por lo tanto, la integral $\int_0^T X_n^t dW_t$ converge en probabilidad a al límite J

$$\int_0^T W_t dW_t = J = \lim_{n \rightarrow \infty} \int_0^T X_n^t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T.$$

Definición C.1.2. *Un proceso de Itô $\{X_t, 0 \leq t \leq T\}$ es un proceso estocástico que puede ser escrita de la forma siguiente*

$$X_t = X_0 + \int_0^t g(s)ds + \int_0^t h(s)dW_s,$$

donde $g(t, \omega)$ y $h(t, \omega)$ son dos funciones aleatorias adaptados y progresivamente medibles tales que:

$$\mathbb{P} \left\{ \int_0^T |g(t, \omega)| dt < \infty \right\} = 1 \quad \text{y} \quad \mathbb{P} \left\{ \int_0^T |h(t, \omega)|^2 dt < \infty \right\} = 1.$$

C.2. Fórmula de Itô

Una herramienta fundamental en el cálculo estocástico es la *fórmula de Itô*. Esta formula se puede interpretar como la versión estocástica de la expansión de Taylor hasta el orden 2 de $g(X)$, donde X es un proceso de Itô o bien un proceso de difusión. El lema de Itô establece que si $g(t, x)$ es una función dos veces diferenciable tanto en t como en x , entonces

$$g(t, X_t) = g(0, X_0) + \int_0^t g_t(u, X_u)du + \int_0^t g_x(u, X_u)dX_u + \frac{1}{2!} \int_0^t g_{xx}(u, X_u)(dX_u)^2,$$

donde

$$g_t(t, x) = \frac{\partial g}{\partial t}(t, x), \quad g_x(t, x) = \frac{\partial g}{\partial x}(t, x) \quad \text{y} \quad g_{xx}(t, x) = \frac{\partial^2 g}{\partial x^2}(t, x)$$

o, en su forma diferencial

$$dg(t, X_t) = g_t(t, X_t)dt + g_x(t, X_t)dX_t + \frac{1}{2}g_{xx}(t, X_t)(dX_t)^2.$$

Si X_t es un proceso de Wiener (movimiento Browniano), entonces

$$g(t, W_t) = g(0, 0) + \int_0^t \left(g_t(u, W_u) + \frac{1}{2!} \int_0^t g_{xx}(u, W_u) \right) du + \int_0^t g_x(u, W_u) dW_u$$

Supóngase $g(t, x) = g(x) = x^2$, entonces la fórmula de Itô aplicada a $g(W_t)$ es

$$W_t^2 = 0^2 + \int_0^t 2W_s dW_s + \frac{1}{2} \int_0^t 2 ds$$

por lo tanto,

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t.$$

Observación C.2.1. En la fórmula de Itô suelen aparecer término de la forma $(dX_t)^2$ que por lo general no son fáciles de interpretar sin el conocimiento de algunos aspectos del cálculo estocástico. Desde el punto de vista de las aplicaciones se establecen las siguientes condiciones: $(dt dW_t)$ Y $(dt)^2$ son expresiones de orden $O(dt)$, lo cual significa que después de desarrollar el término $(dX_t)^2$, todos los términos en la fórmula para el cual la parte diferencial es bien $(dt dW_t)$ o $(dt)^2$ se pueden despreciar. Además, términos del orden $(dW_t)^2$ se comportan como dt debido a las propiedades del proceso de Wiener. Así, la parte diferencial $(dW_t)^2$ se puede reemplazar por dt .

C.3. Ecuaciones diferenciales Estocásticas

En la definición (C.1.2) se introdujo de manera formal el proceso de Itô. Tales procesos serán considerados en esta sección como los procesos, más concretamente los denominados *Procesos de Difusión* definidos como solución de la ecuación diferencial estocástica, y que se denotará por (*SDE*),

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{C.9}$$

con *condición inicial* X_0 . Las funciones $\mu(\cdot)$ and $\sigma^2(\cdot)$, son denominadas el *coeficiente de tendencia* o *drift* y el *coeficiente de difusión* respectivamente, y W_t es un proceso de Wiener estándar definido en una base estocástica

(espacio de probabilidades filtrado) $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. La condición inicial puede ser aleatoria o no. Si es aleatoria, por ejemplo, $X_0 = Z$, esta sería independiente de la σ -álgebra generada por W y satisfaciendo la condición $\mathbb{E}|Z|^2 < \infty$. La ecuación (C.9) se puede representar en su forma integral mediante

$$X_t = X_0 + \int_0^t \mu(u, X_u) du + \int_0^t \sigma(u, X_u) dW_u, \quad (\text{C.10})$$

Como X_t es un proceso de Itô se asume que se satisface la siguiente condición

$$\mathbb{P} \left\{ \int_0^T \sup_{|x| \leq R} (|\mu(t, x)| + \sigma^2(t, x)) dt < \infty, \right\} = 1,$$

para todo $T, R \in [0, \infty)$.

En lo que respecta a la existencia de la solución de la ecuación diferencial estocástica, se deben considerar algunas suposiciones.

\mathcal{LG} .— (Condición Lipschitz Global) Para todo $x, y \in \mathbb{R}$ y $t \in [0, T]$, existe una constante $K < \infty$ tal que

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y| \quad (\text{C.11})$$

\mathcal{CL} .— (Condición Crecimiento Lineal) Para todo $x, y \in \mathbb{R}$ y $t \in [0, T]$, existe una constante $C < \infty$ tal que

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq C(1 + |x|). \quad (\text{C.12})$$

La condición \mathcal{CL} controla el comportamiento de la solución evitando que esta explote en un espacio de tiempo finito.

Teorema C.3.1. *Bajo las suposiciones \mathcal{LG} y \mathcal{CL} , la ecuación diferencial estocástica (C.9) tiene una única solución fuerte, continua y adaptada tal que*

$$\mathbb{E} \left\{ \int_0^T |X_t|^2 dt \right\} < \infty.$$

Además, si $\mathbb{P}\{|X_0| < \infty\} = 1$ se satisface y $\mathbb{E}X_0^{2m} < \infty$, entonces

$$\mathbb{E}X_t^{2m} \leq (1 + \mathbb{E}X_0^{2m})e^{c_m t} - 1,$$

para c_m alguna constante positiva.

El teorema anterior establece la existencia de una solución fuerte, lo cual implica la unicidad de las trayectorias. Es posible también, obtener *soluciones débiles*, por supuesto, bajo ciertas suposiciones. Desde el punto de vista de la inferencia estadística, las condiciones para las soluciones débiles suelen ser suficientes dado que implican que cualesquiera dos soluciones $X^{(1)}$ y $X^{(2)}$ aún cuando no necesariamente son idénticas sus trayectorias, sus distribuciones si lo son. Las soluciones fuertes implican que éstas también son soluciones débiles.

En lo que sigue y motivado a la importancia de la misma se enfocará el estudio en la ecuación diferencial estocástica (C.9) en su forma homogénea:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad (\text{C.13})$$

y

$$X_t = X_0 + \int_0^t \mu(X_u)du + \int_0^t \sigma(X_u)dW_u, \quad (\text{C.14})$$

En diversas situaciones se pueden obtener una versión local de la condición \mathcal{LG} , la cual resulta ser menos restrictiva y puede resultar suficiente.

\mathcal{LL} .— (*Condición Lipschitz Local*) Para cualquier $N < \infty$, $|x|, |y| \leq N$, existe una constante $L_N > 0$ tal que

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq L_N|x - y| \quad (\text{C.15})$$

y

$$2x\mu(x) + \sigma^2(x) \leq B(1 + x^2). \quad (\text{C.16})$$

Teorema C.3.2. *Bajo la suposición \mathcal{LL} y $\mathbb{P}\{|X_0| < \infty\} = 1$, la ecuación diferencial estocástica (C.9) tiene una única solución fuerte, continua con probabilidad 1.*

Respecto de la existencia de una *solución débil* para a ecuación diferencial estocástica se debe asumir la siguiente condición:

\mathcal{ES} .— Sea $\mu(\cdot)$ localmente acotada, $\sigma^2(\cdot)$ continua y positiva, y para algún A se satisface la siguiente condición:

$$x\mu(x) + \sigma^2(x) \leq A(1 + x^2). \quad (\text{C.17})$$

Teorema C.3.3. *Bajo la suposición \mathcal{ES} , entonces la ecuación diferencial estocástica (C.9) tiene una única solución débil.*

Para más detalles referentes a la existencia y unicidad de soluciones de la ecuación diferencial estocástica ver, Karatzas and Shreve (1991), Kutoyants (2004), Shreve (2004).

Los procesos de difusión poseen la propiedad de Markov

$$\mathbb{P}(X_{t+s} \leq y | \mathcal{F}_t) = \mathbb{P}(X_{t+s} \leq y | X_t) \quad c.s.$$

para cualquier t y $s > 0$. Por otro lado, el proceso de difusión puede ser o no ergódico. La *propiedad ergódica* implica que para cualquier función medible $h(\cdot)$, el siguiente resultado se satisface con probabilidad 1:

$$\frac{1}{T} \int_0^T h(X_t) dt \rightarrow \int_{-\infty}^{\infty} h(x) \pi(x) dx = \mathbb{E}(h(\xi)),$$

donde $\pi(\cdot)$ es llamada la *densidad estacionaria o invariante* del proceso de difusión y ξ es alguna variable aleatoria con $\pi(\cdot)$ como densidad.

Si la distribución estacionaria de un proceso de difusión existe, entonces ésta se puede expresar en términos de la *medida de escala y medida de rapidez* definidas por

$$s(x) = \exp \left\{ -2 \int_{x_0}^x \frac{b(y)}{\sigma^2(y)} dy \right\} \quad (C.18)$$

y

$$m(x) = \frac{1}{\sigma^2(x)s(x)}. \quad (C.19)$$

respectivamente. En particular, la densidad de la distribución invariante $\pi(\cdot)$ es proporcional, a la medida de rapidez, es decir,

$$\pi(x) = \frac{m(x)}{M}, \quad (C.20)$$

donde $M = \int m(x) dx$. Si las funciones $m(\cdot)$ y $s(\cdot)$ son tales que

$$\int_0^x s(y) dy \rightarrow \pm\infty \quad \text{and} \quad \int_{-\infty}^{\infty} m(y) dy < \infty$$

cuando $x \rightarrow \pm\infty$. Bajo estas condiciones entonces el proceso X es ergódico y tiene una función de distribución invariante.

C.3.1. Ecuaciones de Kolmogorov

Como ya se ha comentado, Un proceso que satisface la propiedad de Markov es llamado un proceso de *proceso de Markov* y su probabilidad de transición se escribe como

$$P(t, y, |s, B) = P\{X_t \in B | X_s = x\}, \tag{C.21}$$

donde $s < t$. Para s, x fijos y $t, P(t, x|s, \cdot)$ es una medida de probabilidad bajo la σ -álgebra \mathcal{B} de subconjuntos de Borel de \mathbb{R} . A partir de la propiedad de Markov de un proceso de difusión, se puede definir la *función de densidad de transición*. En efecto, la densidad de transición respecto de un subconjunto B es tal que

$$P(t, y|s, B) = \int_B p(t, y|s, x) dx$$

para todo $B \in \mathcal{B}$. Por conveniencia se define $P(s, x|s, B) = \mathbb{1}_B(x)$ para $t = s$, donde $\mathbb{1}_B$ es la función indicadora del conjunto B . Para un proceso de Markov en tiempo continuo su densidad de transición $p(t, y|s, x)$ depende sólo de la diferencia de los instantes de tiempo $t - s$ más que de los valores de s y t . De esta manera la densidad de transición del valor x en el instante s al valor y en el instante t se denotada por $p(t, y|s, x)$ o, cuando sea conveniente, como $p(t - s, y|x)$. Como ejemplos de procesos de Markov tenemos, el proceso de Wiener estándar con densidad de transición,

$$p(t, y|s, x) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2(t-s)}\right) \tag{C.22}$$

Las densidades de transición de un proceso de difusión satisfacen las *ecuaciones de Kolmogorov* para $p = p(t, y|s, x)$

Ecuaciones Forward de Kolmogorov

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y} \{ \mu(y)p \} - \frac{1}{2} \frac{\partial^2}{\partial y^2} \{ \sigma^2(y)p \} = 0, \quad (s, x) \tag{C.23}$$

y las *Ecuaciones Backward de Kolmogorov*

$$\frac{\partial p}{\partial s} + \frac{\partial}{\partial x} \{ \mu(x)p \} + \frac{1}{2} \sigma^2(y) \frac{\partial^2 p}{\partial x^2} = 0, \tag{C.24}$$

Sea $t \rightarrow \infty$ en la ecuación forward de Kolmogorov (C.23) es posible obtener la ecuación tiempo-independiente de *Fokker-Planck* que en este caso es la ecuación diferencial

$$\frac{d}{dx} \{ \mu(x) \pi(x) \} - \frac{1}{2} \frac{d^2}{dx^2} \{ \sigma^2(x) \pi(x) \} = 0 \quad (\text{C.25})$$

donde $\pi(x)$ es la densidad estacionaria. De esta última ecuación se pueden obtener despejando algebraicamente e integrando las relaciones entre el drift y la difusión.

Bibliografía

- Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal Finance*, 54(4):1361–1395.
- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. 267-281, In B.N Petrov and F. Csaki, Akademia Kiado, Budapes.
- Arapis, M. and Gao, J. (2006). Empirical comparisons in short-term interest rate models using nonparametric methods. *Journal of Financial Econometrics*, 4(2):310–345.
- Bachelier, L. (1900). Théorie de la speculation. *Reprinted in Cootner*, pages 17–78.
- Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52:2003–2049.
- Ball, C. A. and Torous, W. N. (1999). The stochastic volatility of short-term interest rates: Some international evidence. *Journal of Finance*, 54:2339–2359.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal Of The Royal Statistical Society Series B*, 63(2):167–241.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*(26):211–243.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, revised edition.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ, 3rd edition edition.
- Box, G. E. P. and Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65:1509–1526.
- Breidt, F. J., Crato, N., and De Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348–54.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, second edition edition.
- Bryson, A. E. and Ho, Y. C. (1969). *Applied Optimal Control*. Blaisdell, New York.
- Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95:941–956.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association*, 87:493–500.
- Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a continuous autoregressive model. *The Annals of Statistics*, 21:520–533.

- Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:179–190.
- Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the conditional least squares estimator of a continuous tar model. *Biometrika*, 85:413–426.
- Chan, N. H. and Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, 16:367–401.
- Chen, R., Liu, J., and Tsay, R. S. (1995). Additivity tests for nonlinear autoregressive models. *Biometrika*, 82:369–383.
- Chen, R. and Tsay, R. S. (1991). On the ergodicity of tar(1) processes. *Annals of Applied Probability*, 1:613–634.
- Chen, R. and Tsay, R. S. (1993a). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88:298–308.
- Chen, R. and Tsay, R. S. (1993b). Nonlinear additive arx models. *Journal of the American Statistical Association*, 88:955–967.
- Chen, T. and Chen, H. (1995). Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6:911–917.
- Cheng, B. and Titterton, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science*, 9:2–54.
- Conley, T. G., Hansen, L. P., Luttmer, E. G. J., and Scheinkman, A. (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies*, 10:525–577.
- Cox, J., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.
- Cramer, H. (1961). On some classes of non-stationary processes. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pages 57–78.

- Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.
- Duffie, D. (1995). *Dynamic Asset Pricing Theory*. Princeton University Press., Princeton, NJ:, 2nd edition.
- Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):pp. 409–428.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflations. *Econometrica*, (50):987–1007.
- Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1):pp. 34–105.
- Fan, J. and Yao, Q. W. (2003). *Nonlinear Time Series*. Springer, New York.
- Gao, J. and King, M. (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory*, 20(5):844–882.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance*, 48:1179–1801.
- Granger, C. W. J. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Gottingen.
- Granger, C. W. J. and Joyeux, R. (2001). *An introduction to long-memory time series models and fractional differencing*, pages 321–337. Harvard University Press, Cambridge, MA, USA.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

- Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45:30–70.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston.
- Hornik, K. (1993). Some new results on neural network approximation. *Neural Networks*, 6:1609–1072.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.
- Iacus, S. M. (2008). *Simulation and Inference for stochastic Differential Equations*. Springer.
- Jacquier, E., Polson, N. G., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics*, 12:371–417.
- Jun, S. C. (2001). A note on fractional differences based on a linear combination between forward and backward differences. *Computers & Mathematics with Applications*, 41(3-4):373–378.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer, New York, second edition.
- Keenan, D. M. (1985). Atukey non-additivity-type test for time series nonlinearity. *Biometrika*, 72:39–44.
- Kendall, M. G. (1953). The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):pp. 11–34.
- Kitagawa, G. (1998). A self-organizing state space model. *Journal of the American Statistical Association*, 93:1203–1215.
- Kloden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Applications of mathematics , Springer, Berlin.

- Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48:1086–1101.
- Kutoyants, Y. A. (2004). *Statistical Inference for ergodic diffusion processes*. Springer, London.
- Liu, J. and Brockwell, P. J. (1988). On the general bilinear time-series model. *Journal of Applied Probability*, 25:553–564.
- Ljung, G. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 66:67–72.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4):394–419.
- McCulloch, R. E. and Tsay, R. S. (1994). Statistical inference of macroeconomic time series via markov switching models. *Journal of Time Series Analysis*, 15:523–539.
- McLeod, A. I. and Li, W. K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4:260–273.
- Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2):239–265.
- Merton, R. C. (1973). The theory of rational option pricing. *Bell Journal of Economics*, 4(1):141–183.
- Merton, R. C. (1976). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics*, 5:125–144.
- Merton, R. C. (1990). *Continuous-Time Finance*. Blackwell Publishing, Oxford, London.
- Nadaraya, E. A. (1964). On estimating regression. *Theory and Probability Application*, 10:186–190.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370.

- Newey, W. and West, K. (1987). A simple positive semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:863–898.
- Ng, S. and Perron, P. (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90:268–281.
- Oldfield, G. S., Rogalski, R. J., and Jarrow, R. A. (1977). An autoregressive jump process for common stock returns. *Journal of Financial Economics*, 5:389–418.
- Osborne, M. F. M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2):145–173.
- Petrucelli, J. and Woolford, S. W. (1984). A threshold ar(1) model. *Journal of Applied Probability*, 21:270–286.
- Prakasa-Rao, B. L. S. (1999). *Statistical inference for diffusion type processes*. Oxford University Press, New York.
- Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London., London.
- Ramezani, C. A. and Zeng, Y. (2007). Maximum likelihood estimation of the double exponential jump-diffusion process. *Annals of Finance*, 3:487–507.
- Ramsey, J. B. (1969). State-dependent models: a general approach to nonlinear time series analysis. *Journal of the Royal Statistical Society, Series B*(31):350–371.
- Said, S. E. and Dickey, D. (1984). Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika*, 71:599–607.
- Schwert, W. (1989). Test for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics*, 7:147–159.
- Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22(04):419–438.

- Shreve, S. E. (2004). *Stochastic Calculus for Finance II, Continuous-Time Models*. Springer, New York.
- Suba Rao, T. and Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models, Lecture Notes in Statistics*. Springer-Verlag, New York.
- Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4:183–204.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.
- Tiao, G. C. and Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in arma models. *Annals of Statistics*, 11:856–871.
- Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*,. Lecture Notes in Statistics. Springer-Verlag, New York.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford, UK.
- Tsay, R. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, 79:84–96.
- Tsay, R. S. (1986). Nonlinearity test for time series. *Biometrika*, 73:461–466.
- Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84:231–240.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhya*, 26(Series A):359–372.

- White, H. (1980). A heteroscedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48:827–838.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Stockholm:Almqvist and Wksell.
- Working, H. (1934). A random-difference series for use in the analysis of time series. *Journal of the American Statistical Association*, 29.
- Zakoian, J. M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control*, 18:931–955.

Asociación Matemática Venezolana

Presidente: Pedro Berrizbeitia

Consejo Directivo Nacional

Pedro Berrizbeitia
Capítulo Capital

Alexander Carrasco
Capítulo de Centro Occidente

Oswaldo Araujo
Capítulo de Los Andes

Said Kas-Danouche
Capítulo de Oriente

Oswaldo Larreal
Capítulo Zuliano

La Asociación Matemática Venezolana fue fundada en 1990 como una organización civil sin fines de lucro cuya finalidad es trabajar por el desarrollo de las matemáticas en Venezuela.

Asociación Matemática Venezolana
Apartado 47.898, Caracas 1041-A, Venezuela
<http://amv.ivic.gob.ve>

Consejo Directivo
Instituto Venezolano de Investigaciones Científicas (IVIC)

Director
Eloy Sira

Subdirector
Alexander Briceño

Representantes del Ministerio del Poder Popular para Educación
Universitaria, Ciencia y Tecnología

Guillermo Barreto
Juan Luis Cabrera
Jesús Manzanilla

Gerencia General
Martha Velásquez

Comisión Editorial
Eloy Sira (coordinador)
Lucía Antillano
Horacio Biord
Jesús Eloy Conde
María Teresa Curcio
Rafael Gassón
Pamela Navarro
Héctor Suárez
Erika Wagner



Gobierno **Bolivariano**
de Venezuela

Ministerio del Poder Popular
para Educación **Universitaria**, Ciencia y Tecnología

