



TESTS DE BONDAD DE AJUSTE PARA  
MODELOS DE TIPOS DE INTERÉS:

*Un Enfoque basado en Procesos Empíricos*

por

Abelardo Enrique Monsalve Cobis

Directores:

Dr. Wenceslao González Manteiga y Dr. Manuel Febrero Bande

Departamento de Estadística e Investigación Operativa  
Universidad de Santiago de Compostela  
Santiago de Compostela, España

2011







Tests de Bondad de Ajuste para  
Modelos de Tipos de Interés:  
*Un Enfoque basado en Procesos Empíricos.*

*(“Goodness of Fit Tests for Interest Rate Models: An Approach based on Empirical Processes”)*

por

Abelardo E. Monsalve Cobis

*Tesis presentada para optar al grado Doctor en Estadística e  
Investigación Operativa en la Universidad de Santiago de  
Compostela*



Departamento de Estadística e Investigación Operativa,  
Universidad de Santiago de Compostela,  
Santiago de Compostela, España.

Directores:

Dr. Wenceslao González Manteiga Dr. Manuel Febrero Bande









# Resumen

## **Tests de Bondad de Ajuste para Modelos de Tipos de Interés: *Un Enfoque basado en Procesos Empíricos.***

*(“Goodness of Fit Tests for Interest Rate Models: An Approach based on Empirical Processes”)*

A.E. Monsalve Cobis

*Departamento de Estadística e Investigación Operativa,  
Universidad de Santiago de Compostela,  
Santiago de Compostela, España.*

Tesis: (Doctorado en Estadística e IO)

2011

En la literatura financiera en los últimos años la dinámica del tipo de interés a corto plazo ha recibido una especial atención debido a que ésta, constituye el principal factor utilizado en la modelización de la estructura temporal de los tipos de interés, herramienta a su vez indispensable en la valoración y gestión de los activos financieros de renta fija, de sus derivados y de las carteras que los integran. La modelización de la dinámica del tipo de interés a corto plazo puede abordarse mediante modelos en tiempo discreto o continuo; utilizando uno o varios factores; y considerando argumentos de arbitraje o de equilibrio. A partir de los trabajos de [Merton \(1973\)](#), los modelos en tiempo continuo han sido los más utilizados en la modelización de los tipos de interés. [Merton \(1973\)](#) presenta un proceso estocástico para el tipo de interés a corto plazo que consiste en un Movimiento Browniano con drift constante. Más tarde, en [Vasicek \(1977\)](#) se reduce la probabilidad de obtener tipos de interés negativos, considerando el drift como un proceso de reversión a la media; y [Cox et al. \(1985\)](#) garantizan dicha positividad, haciendo incrementar la volatilidad al incrementar el nivel de los tipos, es decir, reflejando el llamado efecto nivel. El modelo empírico de [Chan et al. \(1992\)](#) anida en una sola expresión a estos modelos, siendo un caso particular de ellos los modelos de volatilidad constante de [Vasicek \(1977\)](#) y [Merton \(1973\)](#). La modelización de la volatilidad de los tipos de interés a corto plazo en el contexto descrito puede agruparse en tres grandes grupos. En primer lugar se encuentran los modelos Nivel, los cuales especifican la volatilidad del tipo de interés a corto plazo únicamente en función de su nivel; en segundo lugar los modelos generalizados de heteroscedasticidad condicional autorregresiva o modelos *GARCH*, en los que la volatilidad es función de su propio pasado y del impacto de las innovaciones del tipo de interés a corto plazo; y por último, los modelos Mixtos que combinan el efecto nivel y el efecto *GARCH*. Desafortunadamente no existe consenso unánime sobre la mejor manera de modelizar el tipo de interés a corto plazo y, en particular, sobre la modelización concreta de su volatilidad, pieza fundamental en el análisis de las series de tipos de interés. Esta tesis tiene como objetivo estudiar alternativas en cuanto a la metodología para la selección de modelos de tipos de interés, mediante la construcción de test de bondad de ajuste que permitan contrastar los modelos existentes en la literatura y determinar cual es el más adecuado para ajustar las series de tipos de interés,

teniendo en cuenta, en cada caso particular, las características del entorno subyacente. Así pues, se propone una prueba novedosa para la bondad de ajuste de la forma paramétrica de las funciones de deriva y volatilidad de los modelos de tipos de interés. La prueba esta basada en el proceso empírico para determinados residuos. Concretamente, el proceso empírico determinado por los residuos se construye a partir de un estimador de la función de regresión integrada para la función deriva y mediante la función de varianza condicional integrada para la función de volatilidad. La distribuciones de los procesos obtenidos son aproximadas aplicando técnicas bootstrap. El test se aplica a modelos clásicos de la econometría financiera y se ilustra su desempeño en un conjunto de datos reales, a decir, el conjunto de datos del *EURIBOR*.

# Abstract

**Goodness of Fit Tests for  
Interest Rate Models:  
*An Approach based on Empirical Processes***

A.E. Monsalve Cobis

*Department of Statistics and Operative Research,  
University of Santiago de Compostela,  
Santiago de Compostela, Spain.*

Dissertation: (Doctorate in Statistics and OR)

2011

A new test for the goodness of fit of parametric forms of the drift and volatility functions of interest rate models is proposed. The test is based on a marked empirical process of the residuals. More specifically, a marked empirical process is constructed using estimators of the integrated regression function and the integrated conditional variance function for the drift function and the volatility function, respectively. Distributions of these processes are approximated using bootstrap techniques. This test is then applied to simulated classical financial models and is illustrated in an empirical application to a *EURIBOR* data set.

# Agradecimientos

Mi más sincero agradecimiento a los profesores Wenceslao González y Manuel Febrero, por brindarme la oportunidad de trabajar bajo su dirección, por todas las sugerencias y por su apoyo constante en el desarrollo de este trabajo.

A la Universidad de Santiago de Compostela, y en especial al Departamento de Estadística e Investigación Operativa de la Facultad de Matemáticas, como lugar de enseñanza y desarrollo profesional durante estos últimos años. Así mismo, quiero agradecer el soporte académico logístico y económico brindado por el Proyecto de Investigación: Metodologías y aplicaciones en estadística semiparamétrica, funcional y espacio-temporal *MTM2008-03010* del Ministerio de Ciencia e Innovación.

A la Universidad Centroccidental "*Lisandro Alvarado*" a la cual pertenezco, por darme la oportunidad y el respaldo financiero para llevar a cabo mis estudios de Doctorado.

A mis compañeros de estudio y trabajo , Maria, Adela, Fran, Mikel, Edoardo, Manuel con quienes he compartido las vicisitudes propias del ámbito de la investigación en estadística.

A Elisa, por su valiosa y constante colaboración en el aspecto informático.

A todos los profesores que integran el programa interuniversitario de Doctorado y Master en Técnicas Estadísticas por sus valiosos consejos y enseñanzas.

A mi Esposa, por su constante apoyo y comprensión en el desarrollo de esta memoria.

A todos aquellos a quienes haya podido olvidar mil *Gracias*

# Índice general

<b>Resumen</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Agradecimientos</b>	<b>viii</b>
<b>Índice general</b>	<b>ix</b>
<b>Introducción</b>	<b>1</b>
<b>I Econometría Financiera: Conceptos Básicos</b>	<b>5</b>
<b>1 Modelización de las Series de Tiempo Financieras</b>	<b>6</b>
1.1 Rentabilidad y Valoración de Activos Financieros . . . . .	7
1.1.1 Rentabilidad de un Activo . . . . .	8
1.1.2 Otras Series Financieras . . . . .	12
1.2 Estacionariedad . . . . .	15
1.3 Modelos Lineales . . . . .	19
1.3.1 El Proceso Ruido Blanco . . . . .	19
1.3.2 Modelos Autorregresivos . . . . .	20
1.3.3 Modelos de Medias Móviles . . . . .	25
1.3.4 Modelos <i>ARMA</i> . . . . .	28
1.3.5 Representaciones Alternativas de un proceso <i>ARMA</i> . . . . .	30
1.4 Modelos No-Estacionarios . . . . .	31
1.4.1 No estacionariedad en la Varianza . . . . .	32
1.4.2 No estacionariedad en la Media . . . . .	33
1.4.3 Test de Raíz Unitaria . . . . .	35
1.5 Modelos Estacionales . . . . .	35
1.6 Modelos de Regresión de Series de Tiempo . . . . .	36
1.7 Modelos de Memoria Larga . . . . .	43
1.8 Modelos de Heteroscedásticidad Condicional . . . . .	44
1.8.1 Estructura de los Modelos . . . . .	45
1.8.2 El modelo <i>ARCH</i> . . . . .	45
1.8.3 Modelos <i>GARCH</i> . . . . .	48
1.8.4 Modelos <i>EGARCH</i> . . . . .	51
1.8.5 Modelos <i>IGARCH</i> . . . . .	53
1.8.6 Modelos <i>GARCH-M</i> . . . . .	54
1.8.7 Modelo <i>TGARCH</i> . . . . .	55
1.8.8 Modelo de Volatilidad Estocástica . . . . .	55

1.9	Modelos No Lineales . . . . .	57
1.9.1	Modelos no Lineales para la Esperanza Condicional . . . . .	58
1.9.2	Modelo Bilineal . . . . .	58
1.9.3	Modelos <i>TAR</i> . . . . .	58
1.9.4	Modelos <i>SETAR</i> . . . . .	59
1.9.5	Modelos <i>STAR</i> . . . . .	60
1.9.6	Modelos Markov Switching . . . . .	60
1.9.7	Métodos No-Paramétricos . . . . .	61
1.9.8	Modelo de Coeficiente Funcional Autorregresivo . . . . .	66
1.9.9	Modelo No-Lineal Autorregresivo Aditivo . . . . .	66
1.9.10	Modelo No-Lineal de Espacio de Estado . . . . .	67
1.9.11	Redes Neuronales . . . . .	67
1.9.12	Tests de No-Linealidad . . . . .	69
1.9.13	Predicción . . . . .	71
1.10	Modelos en Tiempo Continuo . . . . .	72
1.10.1	Movimiento Browniano . . . . .	72
1.10.2	Proceso de $It\hat{o}$ . . . . .	75
1.10.3	Distribución del Precio de los Activos y la Rentabilidad Continua . . . . .	78
1.10.4	Procesos de Difusión con Salto . . . . .	80
1.10.5	Modelo de Volatilidad Estocástica . . . . .	81
1.10.6	Estimación de los Modelos en Tiempo Continuo . . . . .	82
<b>2</b>	<b>Modelización de los Tipo de Interés</b> . . . . .	<b>85</b>
2.1	Bonos y Tipos de Interés . . . . .	86
2.1.1	Bonos de Cupón Cero . . . . .	86
2.1.2	Tipos de Interés al Contado . . . . .	87
2.1.3	Tipos de Interés a Plazo . . . . .	90
2.2	Modelos en tiempo Continuo . . . . .	92
2.2.1	Modelos de Difusión Homogéneos . . . . .	93
2.2.2	Modelo de Merton . . . . .	96
2.2.3	Modelo de Black-Scholes o Geométrico Browniano <i>GBM</i> . . . . .	96
2.2.4	Modelo de Vasicek ( <i>VAS</i> ) . . . . .	97
2.2.5	Modelo de Dothan . . . . .	98
2.2.6	Modelo de Cox-Ingersoll-Ross ( <i>CIR</i> ) . . . . .	99
2.2.7	Modelo de Chan-Karolyi-Longstaff-Sanders ( <i>CKLS</i> ) . . . . .	100
2.2.8	Modelo de Ait Sahalia . . . . .	101
2.2.9	Modelo de Ahn y Gao . . . . .	101
2.2.10	Estimación Paramétrica del Modelo de Difusión Homogéneo . . . . .	101
2.2.11	Estimación No-Paramétrica . . . . .	107
2.2.12	Modelos de Difusión no homogéneos respecto del tiempo . . . . .	110
2.2.13	Metodología de Estimación . . . . .	111
2.3	Modelos Alternativos para el Tipo de Interés . . . . .	114
2.3.1	Modelos <i>GARCH</i> para tipos de Interés . . . . .	114
2.3.2	Modelo de Volatilidad Estocástica . . . . .	116
2.3.3	Modelo Markov-Switching . . . . .	116
2.3.4	Modelo de Volatilidad Estocástica con Markov-Switching . . . . .	116
2.3.5	Modelos de Difusión con Saltos . . . . .	117
2.4	Series de Tipos de Interés <i>EURIBOR</i> . . . . .	118

<b>II</b>	<b>Métodos de Validación</b>	<b>124</b>
<b>3</b>	<b>Test de Bondad de Ajuste</b>	<b>125</b>
3.1	Una breve revisión de los Modelos de Validación . . . . .	126
3.2	Test de Bondad de Ajuste basado en Procesos Empíricos . . . . .	128
3.2.1	Test para la función drift . . . . .	128
3.2.2	Test para la función de volatilidad . . . . .	131
3.3	Estudio de simulación . . . . .	132
3.4	Aplicaciones a los datos del <i>EURIBOR</i> . . . . .	142
<b>4</b>	<b>Aspectos Teóricos del Test de Bondad de Ajuste para la función Drift</b>	<b>146</b>
4.1	Estudio de la Convergencia . . . . .	149
4.2	Demostraciones . . . . .	150
4.3	Consistencia del Proceso Bootstrap . . . . .	153
<b>III</b>	<b>Otros Desarrollos Notables</b>	<b>154</b>
<b>5</b>	<b>Desarrollos en Tests de Bondad de Ajuste basado en Procesos Empíricos</b>	<b>155</b>
5.1	Tests de Bondad de Ajuste para la Comparación de Curvas de tipos de Interés . . .	155
5.1.1	Primer Enfoque: Comparación de la Distribución de los Residuos . . . . .	157
5.1.2	Segundo Enfoque: A partir de un particular Proceso Empírico Marcado . .	159
5.1.3	Tercer Enfoque: Test basado en Procesos Empíricos . . . . .	160
5.1.4	Estudio de Simulación . . . . .	161
5.2	Una Extensión del test de bondad de ajuste para procesos de difusión con saltos. .	162
5.3	La Herramienta Gráfica: SiZer . . . . .	167
5.3.1	Sizer y el Test de bondad de ajuste para la función drift. . . . .	169
5.3.2	Sizer y el Test de bondad de ajuste para la función volatilidad . . . . .	176
5.4	Líneas Futuras de Investigación . . . . .	179
<b>IV</b>	<b>Apéndices</b>	<b>180</b>
<b>A</b>	<b>Nociones básicas de Probabilidad</b>	<b>181</b>
A.1	Espacio de Probabilidad y Variable Aleatoria . . . . .	181
A.2	Valor Esperado, Varianza y Momentos . . . . .	183
A.3	Variable Aleatoria n-dimensional . . . . .	183
A.4	Independencia . . . . .	184
A.5	Distribución Condicional . . . . .	185
A.6	Esperanza Condicional . . . . .	185
A.7	Tipos de Convergencia . . . . .	187
<b>B</b>	<b>Aspectos Generales de los Procesos Estocásticos</b>	<b>189</b>
B.1	Procesos Estocásticos . . . . .	189
B.2	Filtraciones . . . . .	190
B.3	Momentos, Covarianza e Incrementos de un Proceso Estocástico . . . . .	190
B.4	Variación de un proceso . . . . .	190
B.5	Martingalas . . . . .	191
B.6	Propiedad de Markov . . . . .	192

<b>C Elementos de Cálculo Estocástico</b>	<b>193</b>
C.1 Integración Estocástica . . . . .	193
C.1.1 La Integral de Itô . . . . .	194
C.1.2 Propiedades de la Integral de Itô . . . . .	195
C.2 Fórmula de Itô . . . . .	196
C.3 Ecuaciones diferenciales Estocásticas . . . . .	197
C.3.1 Ecuaciones de Kolmogorov . . . . .	199
<b>Bibliografía</b>	<b>201</b>



# Introducción

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 financiera. 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 de la volatilidad del precio de los activos financieros y el mercado de bonos. Ambos temas de estudios tienen una fuerte conexión con el precio de los activos, regulación de títulos de valor y de gestión de carteras, entre otros. En el mercado financiero, los inversionistas tienen como objetivo la obtención de beneficios, es decir, la adquisición de bienes y capitales en la medida en que la rentabilidad esperada del costo de una unidad de inversión, supere el costo del dinero o tipo de interés. Así pues, las variaciones del tipo de interés se presentan con mucha frecuencia, afectando de manera directa o indirecta a las inversiones y a la economía en general. Por otro lado, la globalización de los mercados de capitales ha traído como consecuencia un aumento de la volatilidad de los tipos de interés en todo el mundo, lo cual ha despertado el interés, tanto de los profesionales de las finanzas como de los académicos, ya que la caracterización de la dinámica de los tipos de interés 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 los tipos a corto y largo plazo y para la formación de expectativas. Basados en la información actual, los tipos de interés en el futuro son desconocidos: así pues, un modelo estará orientado a caracterizar dicha incertidumbre. Por esta razón, algunos autores consideran descripciones probabilísticas para caracterizar su evolución en el futuro.

Tanto el precio de los activos financieros como los tipos de interés son series que varían de manera aleatoria con respecto al tiempo. La forma en que se presentan estas variaciones puede ser modelizada mediante procesos estocásticos. Existen dos tipos de procesos estocásticos mediante los cuales podemos modelizar dichas series. Un primer grupo, los llamados procesos estocásticos en tiempo discreto, en los cuales la serie en estudio varía en un conjunto discreto del tiempo. Este tipo de procesos son los ampliamente conocidos en el análisis clásico de series de tiempo, como por ejemplo los modelos lineales, los modelos no-lineales, los modelos de heteroscedasticidad condicional, modelos de volatilidad estocástica, modelos no estacionarios de raíz unitaria, etc. Y un segundo grupo, el que comprende los modelos en tiempo continuo, en los que la variación se presenta en un conjunto continuo, aún si el precio es observado sólo en instantes discretos del tiempo. Por ejemplo, en el caso de los tipos de interés un enfoque que suele aplicarse es suponer

que la serie se compone de dos términos, el primero de ellos, un término determinista, llamado el drift (la deriva) del proceso, y un segundo término, la parte aleatoria llamada la volatilidad del proceso, tanto si se considera un modelo en tiempo discreto (serie de tiempo) o modelo en tiempo continuo.

En las finanzas, los modelos de tiempo continuo, en particular, los procesos de difusión, han adquirido una importancia relevante y se han usado con frecuencia en la caracterización de la dinámica de las principales variables económicas, tales como tipos de cambio, valoración de activos y tipos de interés. En los últimos años la investigación en este contexto ha tenido un desarrollo y un crecimiento importante, teniendo como punto de partida el trabajo de Merton (1973), quien propuso un modelo del tipo de interés como un proceso estocástico que se empleó para la valoración de opciones. Posteriormente, se comenzó a trabajar aplicando argumentos de arbitraje similares a los trabajos de Black y Scholes (1973) para modelizar la estructura temporal de los tipos de interés como puede observarse en los trabajos de Vasicek (1977) y Brennan y Schwartz (1979). Basados en estas ideas, se propusieron modelos que perfeccionaban las consideraciones iniciales, entre los que destacan, Cox et al. (1980), Cox et al. (1985), Chan et al. (1992), entre muchos otros.

Como se puede apreciar, existe una extensa literatura relativa a la modelización de los tipos de interés. Debido al dinamismo de las actividades financieras, los modelos están continuamente sometidos a modificaciones para mejorar su rendimiento y para probar su eficacia en el ajuste de los precios actuales. En cuanto a la selección de un modelo que resulte adecuado para un determinado conjunto de datos, no hay un consenso general que permita establecer un criterio para comparar el desempeño de los distintos modelos propuestos en la literatura, más aún, un criterio para decidir qué o cuál modelo es el mejor para cada conjunto particular de datos. Una de las razones es la enorme variabilidad de las condiciones del mercado en cada instante de tiempo. Ciertamente, una especificación correcta de un modelo que describa el comportamiento probabilístico de los tipos de interés a corto plazo determina la estructura temporal asociada. Por otro lado, una especificación incorrecta del modelo asociado conlleva a un análisis inadecuado de los datos y graves errores en la valoración de los tipos de interés. Por ello, teniendo en cuenta los diversos modelos planteados en la literatura, los investigadores y profesionales de las finanzas se centra en la búsqueda de herramientas que permitan determinar, cuál modelo es el apropiado para caracterizar las regularidades empíricas de los tipos de interés. Los modelos paramétricos suelen resultar de gran interés entre los profesionales de las finanzas, dado que este tipo de modelos les permite interpretar las observaciones en términos de los parámetros. En este sentido, cabe preguntarse entonces, si es apropiado utilizar un modelo paramétrico para un determinado conjunto de datos financieros.

Entre algunas de las propuestas presentadas recientemente por diversos autores en relación al problema de selección de modelos, se encuentran los trabajos de Aït-Sahalia (1996b), Gao y King (2004), Hong y Li (2005), Chen et al. (2008), quienes proponen pruebas para contrastar la especificación paramétrica de un proceso de difusión a partir de las funciones de densidad marginales o de transición del proceso; Corradi y White (1999), presentan un test asintóticamente normal para la función de difusión; Dette y von Lieres und Wilkau (2003), proponen un test para la forma paramétrica de la función de volatilidad basados en el proceso estocástico de la volatilidad integrada; Li (2007), propone un test no-paramétrico para la especificación paramétrica de la función de difusión en un proceso de difusión basado en el error cuadrático entre la función de difusión estimada, no paraméricamente y la difusión propuesta como hipótesis nula; Arapis y Gao (2006), Gao y Casas (2008) proponen test para probar la forma paramétrica del drift y la volatilidad basados en técnicas de suavizado, Fan y Zhang (2003), Fan et al. (2003), proponen test simultáneos para la especificación de la función drift y difusión, basados en el test de razón de verosimilitudes.

Las propuestas antes citadas, en su mayoría, suelen estar basadas en técnicas de suavizado, en

cuanto a la estimación del modelo o el test propuesto. Ciertamente esto supone un inconveniente adicional, debido a la selección del parámetro de suavizado lo cual, además, puede repercutir en la potencia del test. El objetivo de este trabajo es presentar un test para contrastar la bondad del ajuste para la forma paramétrica de la función drift y la función de volatilidad de los modelos de tipos de interés basados en procesos empíricos, véase [Stute \(1997\)](#) y [Stute et al. \(1998\)](#) para test de bondad de ajuste en modelos de regresión, y [Koul y Stute \(1999\)](#) para test de bondad de ajuste para series de tiempo. En lo referido a trabajos orientados a contrastar la bondad del ajuste para modelos de difusión basados en procesos empíricos la literatura resulta escasa, exceptuando, por supuesto los trabajos de [Lee y Wee \(2008\)](#), quienes proponen un test basado en el proceso empírico de los residuos del modelo de difusión, y [Negri y Nishiyama \(2009\)](#), [Negri y Nishiyama \(2010\)](#) and [Masuda et al. \(2010\)](#), quienes proponen un test de bondad de ajuste para modelos de difusión ergódicos. Las propuestas anteriores, sólo consideran testear la hipótesis nula simple.

La propuesta presentada en este documento sugiere un test de bondad de ajuste sencillo y eficiente y que además permite considerar el problema de contrastar tanto la hipótesis nula simple como la compuesta relativa a la forma paramétrica de las funciones de drift y la volatilidad. En el primer caso, el test está basado en las ideas presentadas en [Stute \(1997\)](#), considerando la función de regresión integrada de las variaciones del proceso que caracteriza a los tipos de interés, de donde se obtiene un proceso empírico basado en residuos. En el segundo caso, el test para la función de volatilidad, se toma en cuenta la función de varianza condicional integrada, obteniéndose un proceso empírico que luego será comparado con la hipótesis de una forma paramétrica de la función de la volatilidad. Las distribuciones de los dos estadísticos asociados a tales procesos son aproximadas aplicando técnicas bootstrap.

En un contexto general, el problema de comparación de dos o más grupos de variables puede llevarse a cabo mediante la comparación de sus funciones de regresión. En lo que respecta a modelos de interés no existen propuestas relativas a tests de bondad de ajuste para comparación de curvas de tipos de interés, y menos aún basadas en procesos empíricos. Por ello, en este trabajo se proponen tres tests de bondad de ajuste para la comparación de curvas de tipos de interés. Tales tests son adaptaciones de los tests para comparación de curvas propuestos en el contexto de modelos de regresión. Por ejemplo, la comparación de curvas de regresión en [Pardo-Fernández et al. \(2007\)](#), y [Neumeyer y Dette \(2003\)](#) y para la comparación de series de tiempo no lineales en [Dette y Weißbach \(2009\)](#). En este caso particular, el test considera la comparación de las funciones drift.

Esta memoria tiene como objetivo fundamental, en primer lugar, dar una visión general acerca de la modelización de las series de tipos de interés y en segundo lugar presentar una propuesta novel en cuanto a tests de bondad de ajuste para modelos de tipos de interés basados en procesos empíricos, concretamente, para los modelos de difusión usados en la caracterización de la dinámica de los tipos de interés. La memoria comprende tres partes divididas en dos capítulos cada una de ellas, excepto la última parte que consta de un solo capítulo. La memoria se organiza de la siguiente forma:

- I. La primera parte está referida a los conceptos básicos de la econometría financiera, base fundamental de conocimiento necesario para entender los modelos de tipos de interés y su dinámica que, posteriormente, serán objeto de estudio en esta memoria. Esta parte consta de dos capítulos,
  - Un primer capítulo referido a la discusión de los distintos modelos del análisis de series de tiempo para series financieras, entre los que destacan, los modelos estacionarios, modelos lineales, modelos no lineales, modelos de heteroscedasticidad condicional, modelos de volatilidad estocástica, y los modelos en tiempo continuo.

- Un segundo capítulo en donde se expone la teoría de los tipos de interés, sus aspectos más relevantes y los distintos modelos comúnmente usados en su modelización.

Los modelos expuestos en los capítulos antes mencionados son enriquecidos con aplicaciones a series de datos relativos a activos del mercado financiero español, series relacionadas con la economía en España y las series de los tipos de interés interbancario de la zona Euro, es decir, series del *EURIBOR*.

II. La segunda parte comprende la exposición de la propuesta del test de bondad de ajuste. Esta parte se compone de dos capítulos en los que se describe el test y los aspectos teóricos relacionados con el mismo. Así pues,

- Un tercer capítulo, en donde se presenta en detalle los tests de bondad de ajuste basados en procesos empíricos para las funciones drift y volatilidad. Así mismo se presenta sendos estudios de simulación para verificar el desempeño de dichos tests. En este mismo capítulo se considera una aplicación empírica a las series de tipos de interés interbancario del *EURIBOR*.
- Un cuarto capítulo considera los aspectos teóricos relativos al test de bondad de ajuste para el drift. Se determina de manera teórica la consistencia del test.

III. La tercer y última parte esta dedicada a las extensiones de los test de bondad de ajuste.

- Un quinto capítulo, en el que se presentan diversos tests para la comparación de curvas de tipos de interés, acompañados de sus respectivos estudios de simulación para determinar el nivel y potencia de los mismos. Así mismo se presenta una aplicación gráfica para estudiar el comportamiento para una familia de parámetros involucrados en la comparación.

Es importante destacar que las propuestas aquí presentadas pueden ser generalizadas a modelos más complejos, todo ello motivado a las propiedades que inducen los procesos empíricos en cuanto a la simplicidad que estos aportan tanto desde el punto de vista teórico como el práctico para la implementación de los métodos. Por otro lado, esta memoria es solo el comienzo en la línea de test basados en procesos empíricos para modelos de tipos de interés. Quedan abiertos problemas relativos a considerar los estadísticos para distintos esquemas de discretización de los modelos de tipos de interés, en los que el parámetro  $\Delta$  varía respecto del tiempo, y su efecto en la potencia de los tests. Así mismo, las consideraciones teóricas relativas a éste. Por otro lado, es posible desarrollar un test, en esta línea, que aborde procesos de difusión con saltos y/o procesos de volatilidad estocástica, lo cual resultaría provechoso en lo que respecta al problema de modelización de la volatilidad de los procesos, toda vez que como se mostrará en los resultados obtenidos al aplicar el test a las series del *EURIBOR*, los modelos son ampliamente rechazados en cuanto a la modelización de la volatilidad. Otro tema que resulta importante es el de la calibración de los test. Como se observará, salvo en el caso de los modelos de tipos de interés con función de volatilidad constante, para los cuales el procedimiento bootstrap está justificado teóricamente, se presenta un desarrollo empírico de las técnicas bootstrap para los casos más generales, es decir, funciones de volatilidad que dependen del tipo de interés, con el fin de aproximar eficientemente las distribuciones de los tests, esto se ve reflejado en los diversos estudios de simulación.

I

# Econometría Financiera: Conceptos Básicos

# 1

## Modelización de las Series de Tiempo Financieras

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, 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 econométricos 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 y 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.

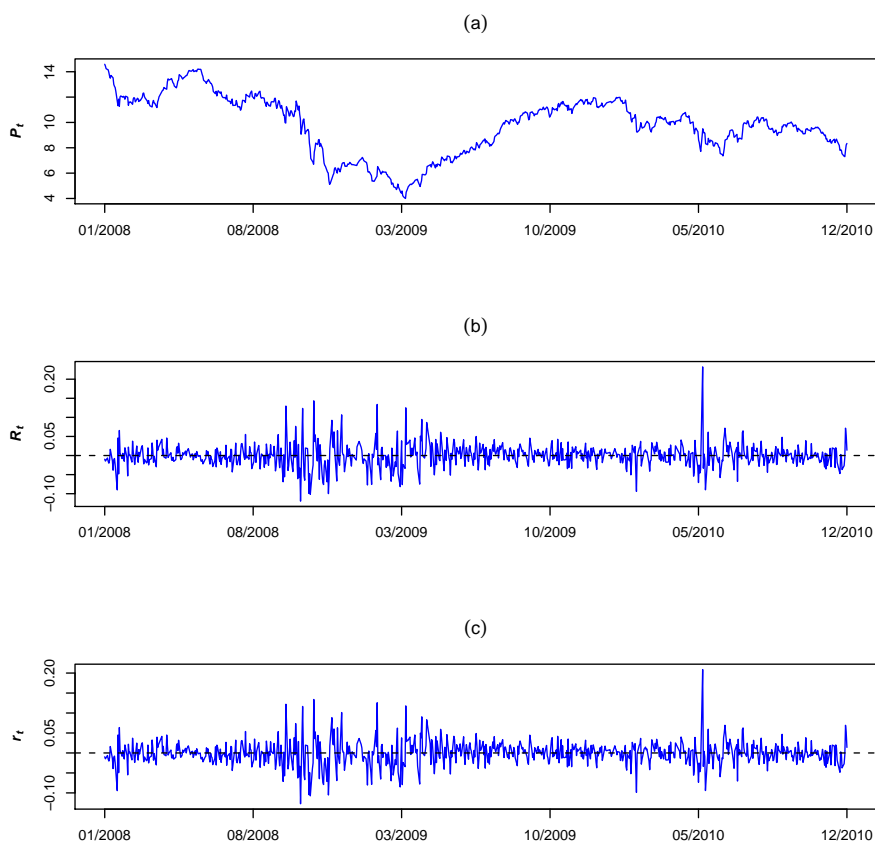
En el análisis de las series de tiempo en finanzas, son diversos los enfoques considerados. Uno de ellos es el presentado por Working (1934), quien centró la atención en una característica notada previamente en los precios de las acciones, a saber: que se asemejan a las situaciones de acumulación de 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.



**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 2010, (a) precio de cierre (b) rentabilidad simple (c) rentabilidad continua.

### 1.1.1 Rentabilidad de un Activo

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.** 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 (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 denomi-



nador,  $P_{t-1}$ , 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 1.2.** 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) hace referencia a la rentabilidad simple para el período  $(t-1, t)$ . En general, para un período  $(t-k, t)$ , con  $k > 1$  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

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

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

**Definición 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  $(t-k, t)$ ,

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

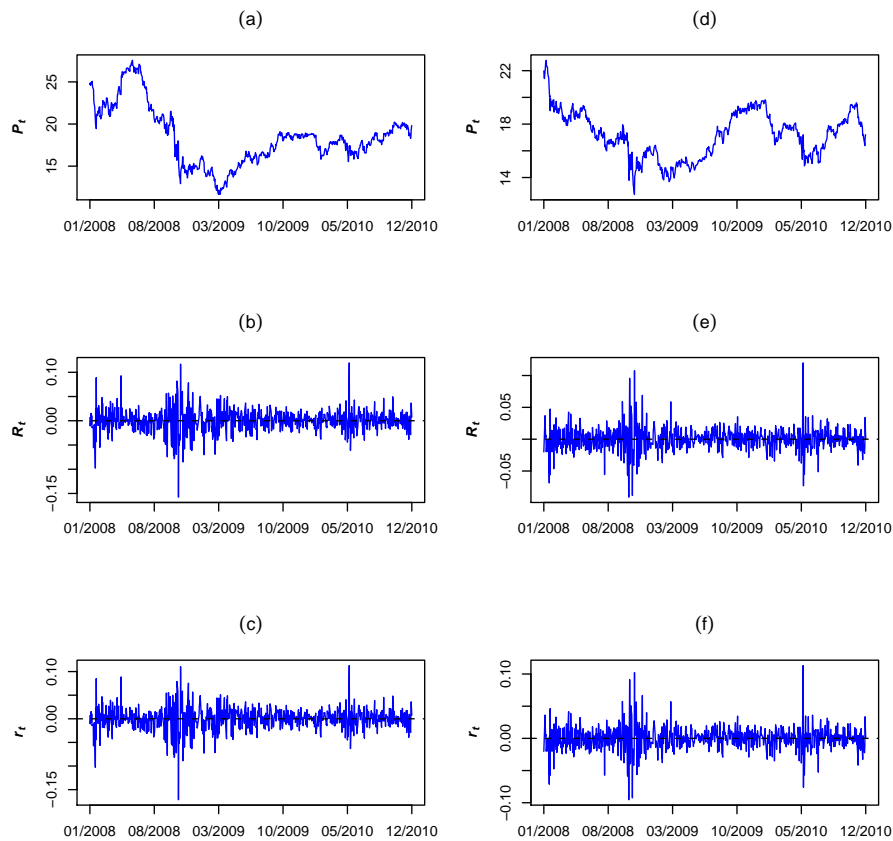
Las figuras (1.1) y (1.2) muestran las series de precios, rentabilidades simple y continua de los activos *Santander*, *Repsol* y *Telefónica* en el mercado de valores de España, concretamente en el IBEX-35 de la Bolsa de Madrid, con frecuencia diaria en el período comprendido entre el 2 de Enero de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

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. 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



**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 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

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.

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  $\mu_r$  y varianza  $\sigma_r^2$ , 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 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

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

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

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

donde  $\sigma_x$  y  $\mu_x$  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 las tablas (1.1) y (1.2), 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 los 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. En las (1.1) y (1.2) 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.

**Tabla 1.1:** Estadísticas Descriptivas para las rentabilidades simples de activos del *IBEX-35* de la Bolsa de Madrid, observaciones diarias, en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre 2010.

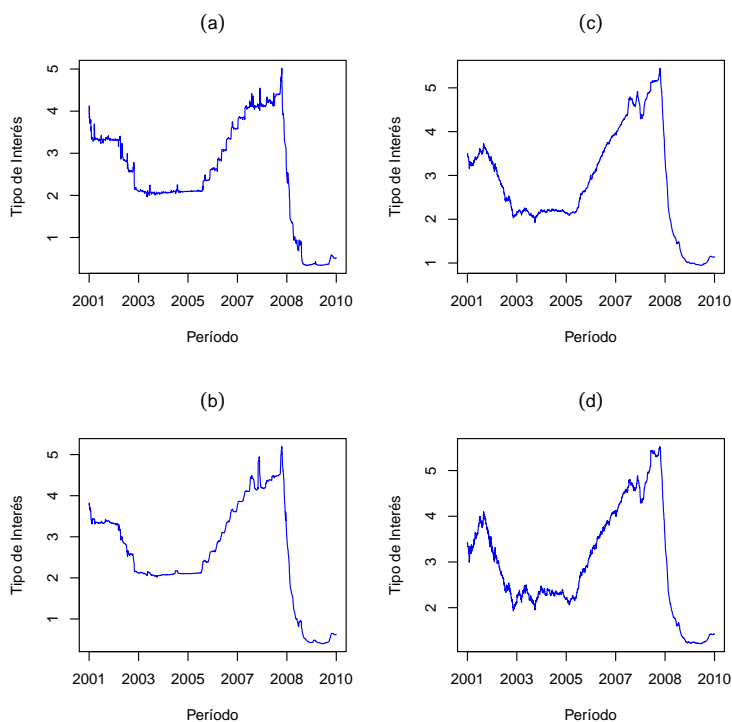
Activo	Mín	Máx	Media	Mediana	Varianza	Desv	Asimetría	Curtosis
Santander	-0.1194	0.2322	-0.0003	0	0.0010	0.0309	0.8013	6.7419
BBVA	-0.1278	0.2203	-0.0006	-0.0010	0.0009	0.0293	0.7468	6.3464
Endesa	-0.2815	0.1288	-0.0006	0.0000	0.0005	0.0234	-2.1358	28.9111
Iberdrola	-0.1258	0.1880	-0.0005	0.0008	0.0007	0.0261	0.8026	9.6047
Iberia	-0.1214	0.2643	0.0006	0	0.0010	0.0312	1.2481	9.8330
Inditex	-0.1032	0.1048	0.0008	0	0.0006	0.0237	0.3385	2.5798
Repsol	-0.1575	0.1194	-0.0000	0.0005	0.0006	0.0235	-0.2427	5.6627
Telefónica	-0.0910	0.1198	-0.0002	0	0.0003	0.0184	0.3511	6.9638

**Tabla 1.2:** Estadísticas Descriptivas para las rentabilidades continuas de activos del *IBEX-35* de la Bolsa de Madrid, observaciones diarias, en el período comprendido entre el 2 de Enero de 2008 y 3 de Diciembre 2010.

Activo	Mín	Máx	Media	Mediana	Varianza	Desv	Asimetría	Curtosis
Santander	-0.1271	0.2088	-0.0008	0	0.0010	0.0306	0.4586	5.2875
BBVA	-0.1367	0.1991	-0.0010	-0.0010	0.0009	0.0291	0.4327	5.0356
Endesa	-0.3306	0.1214	-0.0010	0	0.0006	0.0243	-3.3107	46.2193
Iberdrola	-0.1344	0.1723	-0.0008	0.0008	0.0007	0.0259	0.4007	8.1180
Iberia	-0.1294	0.2345	0.0001	0	0.0009	0.0307	0.8268	7.1239
Inditex	-0.1089	0.0997	0.0006	0	0.0006	0.0236	0.1811	2.4678
Repsol	-0.1714	0.1128	-0.0003	0.0005	0.0006	0.0236	-0.5206	6.3318
Telefónica	-0.0954	0.1131	-0.0003	0	0.0003	0.0183	0.1143	6.4805

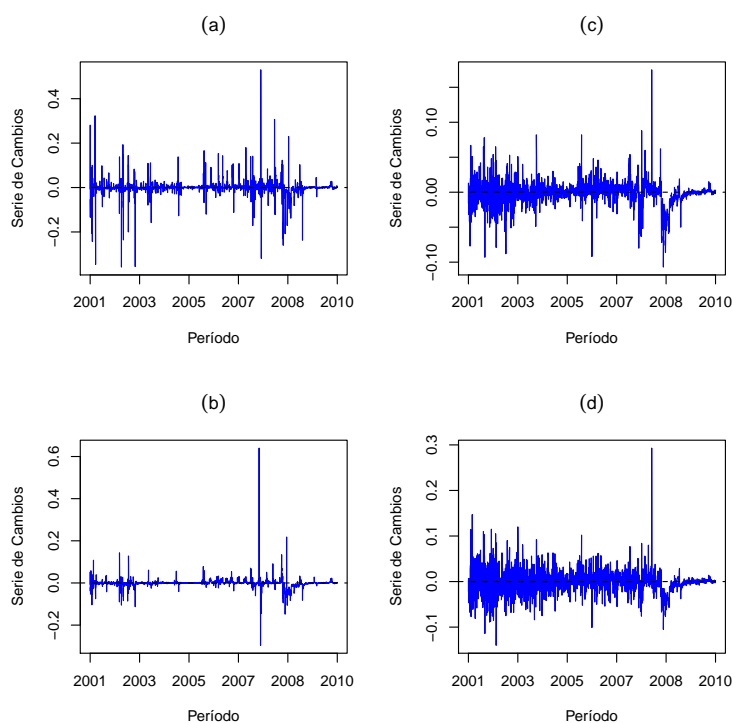
### 1.1.2 Otras Series Financieras

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. Sin embargo, y como se describirá con detalle en el siguiente capítulo, son las series de tipos de interés las que tomarán especial importancia en el desarrollo de la propuesta presentada en esta memoria, concretamente, “*Test de Bondad de Ajuste para Modelos de Tipos de Interés*”, y que se desarrollará en el capítulo 3.



**Figura 1.3:** Series del *EURIBOR*, período comprendido entre 15 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 figura (1.3) muestra las Series del *EURIBOR*, para los distintos plazos 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.



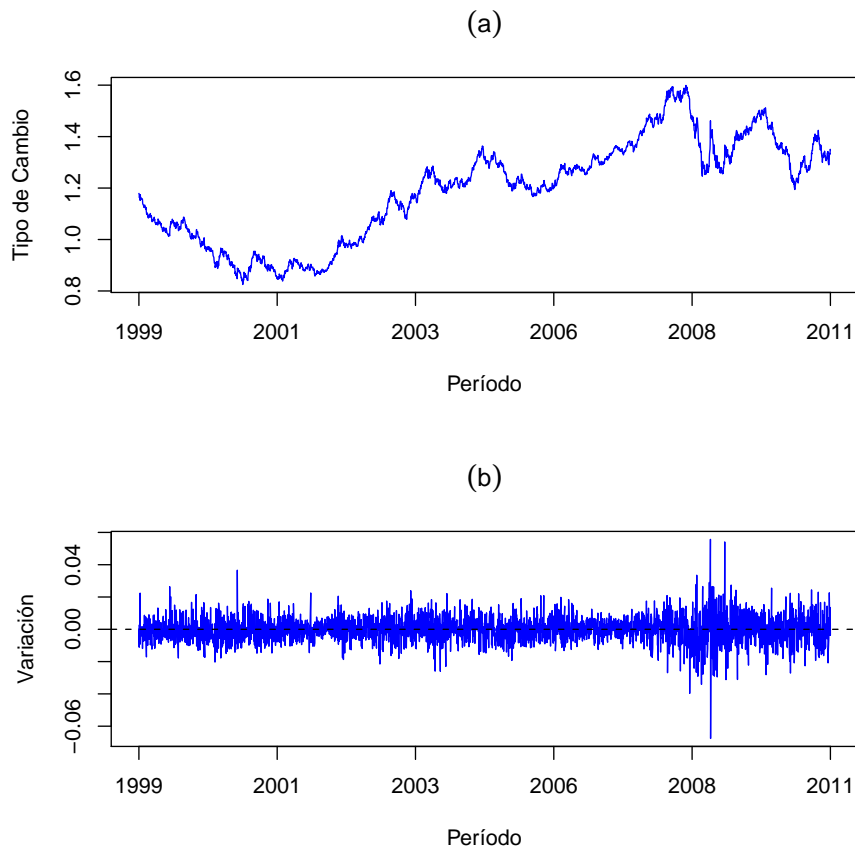
**Figura 1.4:** Series de las variaciones del tipo de interés del *EURIBOR*, período comprendido entre 15 octubre 2001 al 3 de diciembre de 2010. Plazo de vencimiento: (a) 1 semana; (b) 1 mes; (c) 6 meses (d) 1 año (12 meses)

La tabla (1.3) presenta algunos de los estadísticos descriptivos más importantes 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 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.

**Tabla 1.3:** 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.

Plazo	Mín	Máx	Media	Mediana	Varianza	Desv	Asimetría	Curtosis
1 semana	0.34	5.02	2.52	2.37	1.47	1.21	-0.31	-0.77
1 mes	0.63	5.39	2.73	2.52	1.52	1.23	0.13	-0.76
6 meses	0.94	5.45	2.82	2.56	1.43	1.19	0.29	-0.81
12 meses	1.21	5.53	2.95	2.64	1.38	1.17	0.37	-0.86



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

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.

## 1.2 Estacionariedad

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 y 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  $(x_{t_1}, \dots, x_{t_k})$  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 1.4.** Un proceso estocástico se denomina estacionario si  $\mathbb{E}(x_t^2) < \infty$  for each  $t$ , and

- $\mathbb{E}(x_t) = \mu_x$ , (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.)  $\gamma_0 = \text{Var}(x_t)$ , y

(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  $\gamma_k$ , y que bajo la suposición de estacionariedad débil es una función que depende sólo del retardo  $k$  y viene expresada por:

$$\rho_k = \frac{\text{Cov}(x_t, x_{t-k})}{\{\text{Var}(x_t)\text{Var}(x_{t-k})\}^{1/2}} = \frac{\text{Cov}(x_t, x_{t-k})}{\text{Var}(x_t)} = \frac{\gamma_k}{\gamma_0}, \quad (1.11)$$

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.)  $\rho = 1$ ,

(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  $\rho_k$  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  $\sigma_x^2 = \gamma_0 = \text{Var}(x_t)$ , 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

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, \quad 0 \leq k < T - 1 \quad (1.12)$$

Bajo algunas condiciones generales,  $\hat{\rho}_k$  es un estimador consistente de  $\rho_k$ . 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

$$x_t = \mu + \sum_{i=0}^q \eta_i \varepsilon_{t-i} \quad (1.13)$$



para  $\eta_0 = 1$ ,  $q$  un entero no negativo, y  $\{\varepsilon_j\}$  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 y Jenkins \(1976\)](#). En las aplicaciones, suele ser común desear probar de manera conjunta si las autocorrelaciones de una serie  $x_t$  son cero. [Ljung y Box \(1978\)](#) proponen un test basado en una modificación del estadístico de *Portmanteau* propuesto por [Box y 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

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}. \quad (1.14)$$

el cual es una modificación del estadístico de *Portmanteau*

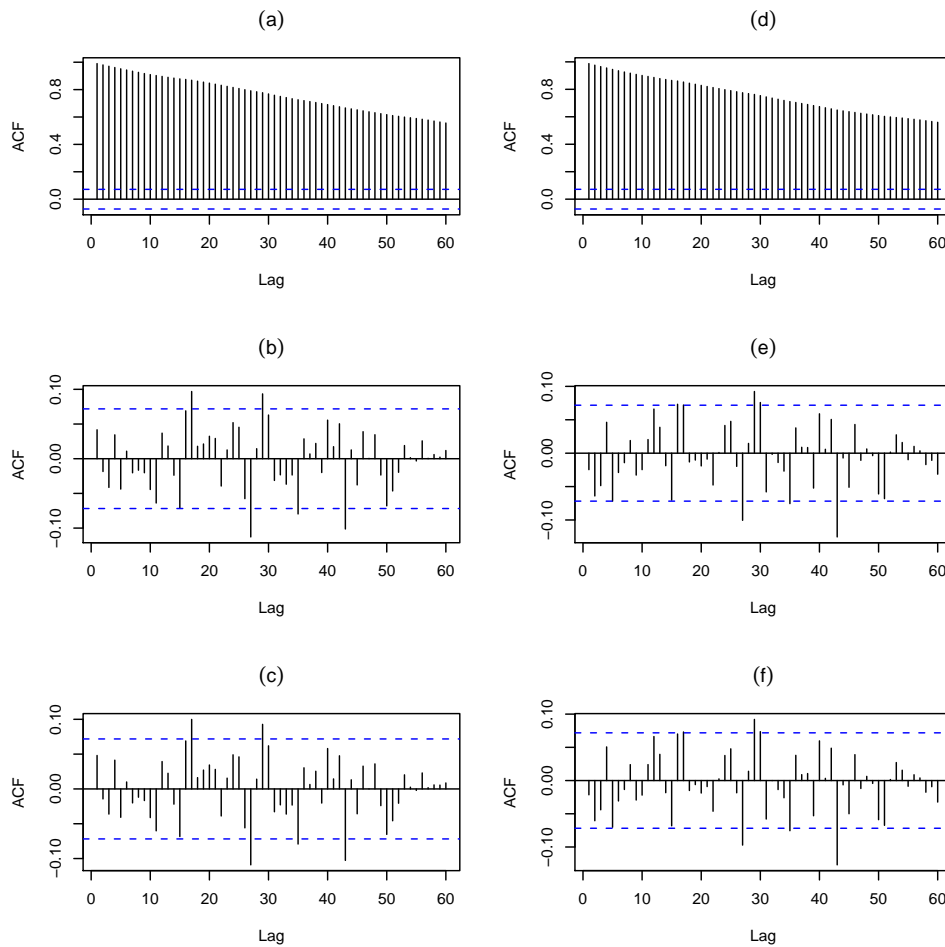
$$Q^*(m) = T \sum_{k=1}^m \hat{\rho}_k^2,$$

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  $\chi_\alpha^2$  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 (1.12) caracterizan los modelos lineales de series de tiempo. En la figura (1.6) 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.

**Tabla 1.4:** Estadístico de Ljung-Box para las series del Banco Santander e índice del *IBEX35*.

Serie	Banco Santander			IBEX35		
	$Q(m)$	$m$	$p$ -valor	$Q(m)$	$m$	$p$ -valor
Precio	5516.716	8	$< 2.2e - 16$	5447.224	8	$< 2.2e - 16$
Rentabilidad Simple	5.768	8	0.673	11.834	8	0.159
Rentabilidad Continua	5.876	8	0.661	11.452	8	0.177

La tabla (1.4) 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  $\alpha$ , 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 series



**Figura 1.6:** 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

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.

## 1.3 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.

### 1.3.1 El Proceso Ruido Blanco

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

$$\mathbb{E}(\varepsilon_t) = 0, \quad \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty, \quad \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0, \text{ para todo } k \neq 0 \quad (1.15)$$

En particular, una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media 0 y varianza  $\sigma_\varepsilon^2$  representa un caso especial de un proceso de ruido blanco, y que se denota por  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ . Si además  $\varepsilon_t$  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

$$\begin{aligned} x_t - \mu_x &= \varepsilon_t + \eta_1 \varepsilon_{t-1} + \eta_2 \varepsilon_{t-2} + \dots \\ &= \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad \eta_0 = 1, \end{aligned} \quad (1.16)$$

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  $\varepsilon_t$  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  $\eta_k$  en (1.16) (posiblemente infinitos en número) son conocidos como los  $\eta$ -pesos. Claramente el proceso  $x_t$  puede ser escrito como

$$x_t = \mu_x + \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad \eta_0 = 1. \quad (1.17)$$

La estructura de la dinámica del proceso  $x_t$  esta gobernada por los coeficientes  $\eta_i$ . Si el número de pesos en (1.17) 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

- $\mathbb{E}(x_t) = \mu_x$
- $\gamma_0 = \text{Var}(x_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \eta_i^2$ , donde  $\sigma_\varepsilon^2$  es la varianza del  $\varepsilon_t$ .
- 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 (1.17). Por ejemplo, si se considera  $\mu_x = 0$ , sin pérdida de generalidad, y se seleccionan,  $\eta_i = \phi^i$ , entonces la ecuación (1.17) puede expresarse como (después de operar algebraicamente)

$$x_t = \phi x_{t-1} + \varepsilon_t. \quad (1.18)$$

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  $\eta$ -pesos de  $x_t$ . Para una serie débilmente estacionaria,  $\eta_i \rightarrow 0$  cuando  $i \rightarrow \infty$  y, por lo tanto,  $\rho_k$  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  $k$ .

### 1.3.2 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

$$x_t = \phi_0 + \phi_1 x_{t-1} + \varepsilon_t. \quad (1.19)$$

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 (1.19) 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  $\sigma_\varepsilon$ .

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

$$x_t = \mathbb{E}(x_t|x_{t-1}) + \varepsilon_t. \quad (1.20)$$

En general, existen situaciones en las cuales el simple valor  $x_{t-1}$  no resulta suficiente para determinar la esperanza condicional de  $x_t$ , de manera que se debe pensar en otros modelos más flexibles. Una generalización del modelo (1.20) viene dada por

$$x_t = \mathbb{E}(x_t|\mathcal{F}_{t-1}) + \varepsilon_t, \quad (1.21)$$

donde  $\mathcal{F}_{t-1}$  representa la información disponible hasta el instante  $t-1$  y  $\varepsilon_t$  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 (1.19), comúnmente denotado por  $AR(p)$ , se escribe como

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t, \quad (1.22)$$

donde  $p$  es un entero positivo y  $\{\varepsilon\}$  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 el modelo  $AR(1)$  dado por la ecuación (1.19). Asumiendo que el  $x_t$  es débilmente estacionario, se tiene que

- i.  $\mathbb{E}(x_t) = \mu$
- ii.  $\text{Var}(x_t) = \gamma_0$
- iii.  $\text{Cov}(x_t, x_{t-k}) = \gamma_k$ ,

donde  $\mu$  y  $\gamma_0$  son constantes y  $\gamma_k$  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.  $\text{Cov}(x_t, x_{t-k}) = \gamma_k = \begin{cases} \phi_1 \gamma_1 + \sigma_\varepsilon^2 & k = 0, \\ \phi_1 \gamma_{k-1} & k > 0. \end{cases}$

De manera que, una condición necesaria y suficiente para que un modelo  $AR(1)$  con ecuación (1.19) 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  $\phi_1$ . Si  $\phi > 0$  el decaimiento es constante. Si por el contrario,  $\phi_1 < 0$  entonces el decaimiento es compuesto y se presenta de forma alternante con tasa  $\phi_1^2$ .

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 (1.22) se puede reescribir como

$$(1 - \phi_1 L - \dots - \phi_p L^p)x_t = \varepsilon_t, \quad (1.23)$$

o bien en su forma reducida

$$\phi_p(L)x_t = \varepsilon_t, \quad (1.24)$$

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, y es que las raíces de la *ecuación característica*  $\phi_p(L) = 0$ , sean menor que uno en modulo, o dicho de otra forma, que las *raíces características* del polinomio característico estén fuera del círculo unidad. Para un proceso estacionario  $AR(p)$ , la *ACF* satisface la ecuación en diferencias

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

Una representación gráfica de la *ACF* de un modelo  $AR(p)$  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 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 e 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:

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

con  $\phi_{0,j}$ ,  $\phi_{i,j}$  y  $\{e_{jt}\}$  son, el termino constante, el coeficiente de  $r_{t-i}$  y el termino error de un modelo  $AR(j)$ , respectivamente. Claramente, cada uno de estos modelos están expresados como un modelo de regresión lineal multiple, 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  $r_t$ . De la definición, se puede extraer que el valor estimado  $\hat{\phi}_{k,k}$  muestra la contribución de  $r_{t-k}$  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  $p$ .

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

$$AIC = \frac{-2}{T} \log(\text{verosimilitud}) + \frac{2}{T}(n^\circ \text{de parámetros}) \quad (1.25)$$

donde la verosimilitud es evaluada en los valores estimados por el método de máxima verosimilitud. El primer término en (1.25) mide la 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 practica, 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  $AIC$ .

### 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

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

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

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

y los residuos asociados al modelo estimado son

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

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

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

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

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

### 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, éste viene dado por

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}. \quad (1.26)$$

La regla de decisión es rechazar la hipótesis nula  $H_0 : \rho_1 = \dots = \rho_m = 0$  si  $Q(m) > \chi_{\alpha}^2$ , donde  $\chi_{\alpha}^2$  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  $\{x_t | t = 1, \dots, T\}$ , el estadístico *R-square* ( $R^2$ ) es definido como

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

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 decreciente del número de parámetros usados. 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  $x_t$ . Esta nuevo estadístico toma en cuenta el número de parámetros usados en el modelo ajustado. Por supuesto los valores de este estadístico no están entre 0 y 1.

### Predicción

Dentro del análisis de series de tiempo la predicción juega un papel muy importante. Supóngase 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

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

donde  $g$  una función de  $\mathcal{F}_h$ . Se denotara 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  $\hat{x}_h(k)$  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 condicional. 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  $x_t$ .

### 1.3.3 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 (1.19) y seleccionando  $\eta_1 = -\theta$  y  $\eta_j = 0$ ,  $j \geq 2$  se obtiene el siguiente modelo

$$x_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, \quad \mu\text{-constante} \quad (1.27)$$

o bien, aplicando el operador de retardos

$$x_t = \mu + (1 - \theta L)\varepsilon_t, \quad (1.28)$$

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 (1.27) y (1.28) se puede considerar sin pérdida de generalidad  $\mu_x = 0$  obteniéndose así la clásica representación del modelo  $MA(1)$

$$x_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad (1.29)$$

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

$$x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}, \quad (1.30)$$

o de forma equivalente, usando el operador de retardo, como

$$x_t = \mu + (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)\varepsilon_t = \theta_q(L)\varepsilon_t, \quad (1.31)$$

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 (1.27), 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  $\varepsilon_t$  y  $\varepsilon_{t-1}$  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  $\mu$  es la media del proceso, es decir,  $\mathbb{E}(x_t) = \mu$ , y
- 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

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

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  $ACF$  para el retardo  $q$  es distinta de cero, sin embargo  $\rho = 0$  para retardos  $k$  superiores al orden del proceso,  $q$

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

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  $\varepsilon_t$  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 unidad. Esta propiedad implica que el proceso se puede escribir en términos un proceso  $AR(\infty)$

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

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

$$x_t = \theta(L)\varepsilon_t.$$

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

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

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

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

Los pesos  $\pi_j = -\theta^j$  convergen siempre que  $|\theta| < 1$ , 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  $\pi = \theta^{-1}(L)$ , 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

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, se anula ( $\rho_k = 0$  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 demostrarse que la  $PACF$  (función de autocorrelación parcial) no se anula, aunque a partir de un retardo  $q$  decaerá de forma rápida. Las expresiones para  $PACF$  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 sinusoidal (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:

- 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  $q$ .

### Estimación de un modelo MA

Similar al caso de un proceso autorregresivo, el método de máxima verosimilitud es el método 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  $\varepsilon_t$  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, las raíces de la ecuación característica  $\theta_q(L) = 0$  están cercanas al círculo unidad.

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 aplicando modelos MA

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 la predicción origen es  $h$  y denote por  $F_h$  la información disponible en el instante  $h$ . Para la predicción a un paso de un proceso  $MA(1)$

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

Tomando la esperanza condicional, se obtiene

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

La varianza del error de la predicción a un paso es  $\text{Var}(e_h(1)) = \sigma_\varepsilon^2$ . En la práctica, el valor de  $\varepsilon_h$  se puede obtener de manera 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  $MA(h)$  la predicción a varios pasos tienden a la media después de los primeros  $q$ -pasos.

### 1.3.4 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,

$$\begin{aligned}x_t &= \phi_0 + \phi_1x_{t-1} + \phi_2x_{t-2} + \dots + \phi_px_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \\ &= \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i},\end{aligned}\tag{1.32}$$

donde  $\{\varepsilon_t\}$  es una serie 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

$$\phi_p(L)x_t = \phi_0 + \theta_q(L)\varepsilon_t.\tag{1.33}$$

Los polinomios

$$\phi(L) = 1 - \phi_1L - \dots - \phi_pL^p = 1 - \sum_{i=1}^p \phi_i x_{t-i},$$

y

$$\theta(L) = 1 - \theta_1L - \dots - \theta_qL^q = 1 - \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

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 unidad, 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

$$x_t - \phi_0 - \phi_1 x_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (1.34)$$

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 (1.34) 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  $\phi_1$ ,  $\phi_1^2 < 1$ . Así mismo, la autocovarianza para  $k = 1$

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

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  $ARMA(1, 1)$

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

De esta manera, la función de autocorrelación  $ACF$  de un proceso  $ARMA(1, 1)$  se comporta como la de un proceso  $AR(1)$  excepto que el decaimiento exponencial comienza en el segundo retardo. En consecuencia, la  $ACF$  de un  $ARMA(1, 1)$  no se reduce en ningún retardo finito.

### Identificación de un proceso $ARMA$

A diferencia de los modelos propuestos anteriormente, en los que la identificación de su orden podía ser llevada a cabo mediante la  $ACF$  o bien la  $PACF$ , en este caso, para los modelos  $ARMA$  estas funciones no aportan información relativa a la determinación de su orden. Tsay y 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 el orden  $q$   $MA$ . La tabla (1.5) 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 criterio 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  $BIC$ ) para el proceso  $ARMA(p, q)$ , donde  $0 \leq p \leq P$  y

**Tabla 1.5:** 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

$0 \leq q \leq Q$ , y se selecciona el modelo con menor *AIC* (o *BIC*). Evidentemente este método requiere el método 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 bien el de verosimilitud exacta. Además, el estadístico de Ljung-Box de los residuos se pueden utilizar para comprobar la adecuación del modelo ajustado.

### Predicción a partir de un proceso *ARMA*

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

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

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

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

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

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

### 1.3.5 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 (1.33)

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

Ahora bien, dados

$$\frac{\phi(L)}{\theta(L)} = 1 - \pi_1 L - \pi_2 L^2 - \dots \equiv \pi(L), \quad (1.35)$$

y

$$\frac{\theta(L)}{\phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + \dots \equiv \psi(L), \quad (1.36)$$

entonces se tiene lo siguiente:

- Si se cumplen las condiciones de invertibilidad del proceso  $ARMA(p, q)$  (es decir, los ceros del polinomio  $\theta(L)$  son mayor en módulo que la unidad), entonces el proceso dado por la ecuación (1.33) se puede escribir, usando el operador definido en (1.35), como

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

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

$$x_t = \frac{\phi_0}{1 - \theta_1 - \dots - \theta_q} + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \dots + \varepsilon. \quad (1.37)$$

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  $x_t$  en 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  $\pi_i$  decaen a cero exponencialmente cuando se incrementa  $i$ .

- Si se cumplen las condiciones de estacionariedad (es decir, los ceros del polinomio  $\phi(L)$  son mayor en módulo que la unidad), entonces el modelo  $ARMA$  se puede expresar, usando el operador definido en (1.36), como

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

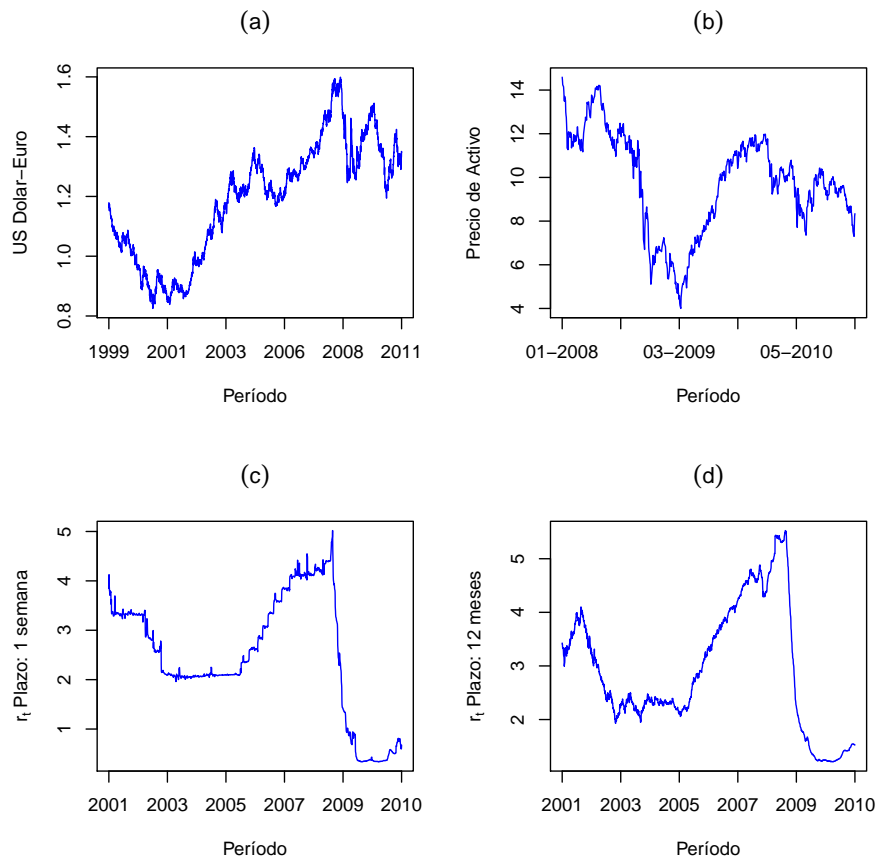
donde  $\mu = \phi_0/\phi(L)$ , de manera que el proceso  $ARMA(p, q)$  se representa por

$$x_t = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p} + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots \quad (1.38)$$

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  $\varepsilon_{t-i}$  ( $i > 0$ ) 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  $\psi_i$  decaen exponencialmente cuando  $i$  es incrementado. Otro característica relevante es el efecto de los  $\varepsilon_{t-i}$  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.

## 1.4 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, variance 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



**Figura 1.7:** 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, plazos 1-semana, 12 meses respectivamente, para el período del 15 de octubre 2001 al 3 Diciembre 2010.

tipos de interés interbancario de la eurozona (*EURIBOR*) por ejemplo, se pueden observar notables variaciones tanto en su nivel como en su varianza. La figura (1.7) 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, plazos 1-semana, 12 meses para el período del 15 de octubre 2001 al 3 Diciembre 2010, respectivamente. En todas ellas se observan variaciones notables en media y varianza. A continuación se discuten brevemente estos conceptos.

### 1.4.1 No estacionariedad en la Varianza

Cuando una serie de tiempo 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

$$x_t = \mu_t + \varepsilon_t, \quad (1.39)$$

y se supondrá además que la varianza de los errores,  $\varepsilon_t$ , esta relacionada de manera funcional a la media  $\mu_t$  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 y Cox \(1964\)](#) definieron una manera práctica y general para seleccionar la transformación  $g(x_t)$ . Esta transformación se define por:

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

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 variancia del todo constante aplicando solo este tipo de transformaciones. Como veremos más adelante, existen modelos que pueden modelizar de manera más eficiente este tipo de situaciones.

### 1.4.2 No estacionariedad en la Media

Considérese nuevamente la ecuación (1.39), y suponga, en este caso, que el término  $\mu_t$  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  $\varepsilon_t$  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

$$x_t = \mu_t + \varepsilon = \sum_{i=1}^d \beta_j t^j + \psi(L)\varepsilon_t. \quad (1.40)$$

Puesto que  $\mathbb{E}(\varepsilon_t) = \psi(L)E(\varepsilon) = 0$ , entonces  $\mathbb{E}(x_t) = \mathbb{E}(\mu_t) = \sum_{i=1}^d \beta_j t^j$ , y dado que los coeficientes  $\beta_j$  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 (1.40), entonces tenemos el caso de *tendencia lineal*, se asumirá, por simplicidad, que el componente del error es un ruido blanco

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (1.41)$$

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  $\varepsilon_t$  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

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

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

$$x_t = x_{t-1} + \varepsilon_t, \quad (1.42)$$

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 (1.42) es un proceso  $AR(1)$ , en este caso  $\phi = 1$ . 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

$$\rho(t+h, t) = \frac{t}{\sqrt{t(t+h)}}. \quad (1.43)$$

El paseo aleatorio pertenece a una clase de procesos denominados *procesos no estacionarios de raíces unitarias*. Si se incluye un término constante,  $\mu$ , en la ecuación (1.42) entonces se obtiene el *paseo aleatorio con drift*

$$x_t = \mu + x_{t-1} + \varepsilon_t, \quad (1.44)$$

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  $\mu$  en el proceso (1.44) es de gran importancia. Si  $x_t$ , por ejemplo, es el logaritmo del precio de un activo, ésta constante representa la tendencia temporal de  $x_t$ . La ecuación (1.44) 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

$$\phi(L)\nabla^d x_t = \mu + \theta(B)\varepsilon_t, \quad (1.45)$$

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 polinomial  $AR$  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  $ARIMA$ .

### 1.4.3 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  $\mu$  la media de  $x_t$ . Said y Dickey (1984), modificaron el test de Dickey y 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 regression

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t, \quad (1.46)$$

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 que homoscedásticos. En la practica,  $c_t$  puede ser cero o una constante o bien  $c_t = \omega + \omega_1 t$ . El *ADF* se basa en las estimaciones de mínimos cuadrados de (1.46) 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  $\beta$ .

La especificación del retardo, es un punto importante en la aplicación del test *ADF*. Si se selecciona un  $p$  demasiado pequeño, entonces la correlación de serie 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 y 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 de la 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

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

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

## 1.5 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 serie de tiempo estacional

$$(1 - L^s)(1 - L)x_t = (1 - \theta L)(1 - \Theta L^s)\varepsilon_t, \quad (1.47)$$

donde  $s$  es la periodicidad de la serie,  $\varepsilon$  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,

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

donde  $y_t = (1 - L^s)(1 - L)$ . El proceso  $y_t$  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.

## 1.6 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*, *las ganancias/precio* etc, 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*

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \varepsilon_t = x_t' \beta + \varepsilon_t, \quad t = 1, \dots, T \quad (1.48)$$

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

$$y = X\beta + \varepsilon \quad (1.49)$$

donde  $\varepsilon$  y la variable  $y$  son vectores de dimensión  $(T \times 1)$  y  $X$  es una matriz de dimensión  $(T \times (k + 1))$ . El modelo (1.48) satisface las suposiciones:

- El modelo esta especificado correctamente.
- $\{y_t, x_t\}$  es estacionario y ergódico conjuntamente.
- Las variables regresores  $x_t$  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 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 termino de error es un proceso no correlacionado serialmente con varianza incondicional  $\sigma^2$  constante. En el modelo de regresión de series de tiempo, los regresores  $x_t$  son aleatorios y el error  $\varepsilon_t$  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

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

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

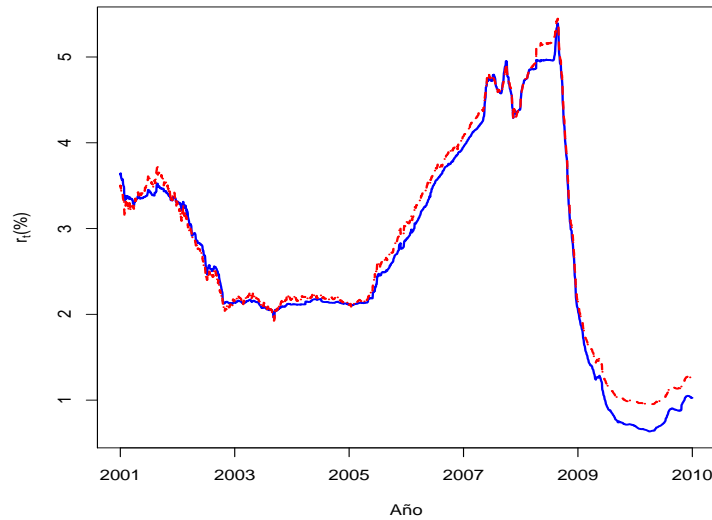
La relación entre dos series de tiempo es de relevante interés en muchas aplicaciones. La estructura de los tipos de interés es un ejemplo en el cual se investiga la evolución en el tiempo de la relación entre tipos de interés con diferentes plazos de vencimiento. Esos ejemplos conllevan a considerar 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  $\varepsilon_t$  denota el error. Si el término de error  $\{\varepsilon_t\}$  constituyen un proceso de ruido blanco entonces el método de mínimos cuadrados produce estimados consistentes. Sin embargo, en la práctica, es común ver que el error  $\varepsilon_t$  es correlacionado serialmente. En este caso, se tiene un modelo de regresión con una serie de tiempo de los errores, y el los valores estimados de  $\beta_0$  y  $\beta_1$  por mínimos cuadrados pueden no ser consistente. Un modelo de regresión con errores como series de tiempo son ampliamente usados en economía y finanzas, pero es uno de los que comúnmente son usados de manera inadecuada en modelos econométricos, porque la dependencia serial en  $\varepsilon_t$  es a menudo pasada por alto. Vale la pena estudiar el modelo 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.



**Figura 1.8:** 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 3-meses, y en rojo, (línea discontinua) serie con plazo de vencimiento 6-meses

La figura (1.8) muestra el gráfico de estas dos series, En azul (línea continua) serie con plazo de vencimiento 3-meses, en rojo (línea discontinua) serie con plazo de vencimiento 6-meses. La figura (1.9(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

$$r_{6t} = 0.1953 + 0.9621r_{3t} + \varepsilon_t, \quad \hat{\sigma}_\varepsilon = 0.1, \quad (1.50)$$

con  $R^2 = 99.31\%$ , donde los errores estándar de los dos coeficientes se muestran en la tabla (1.6)

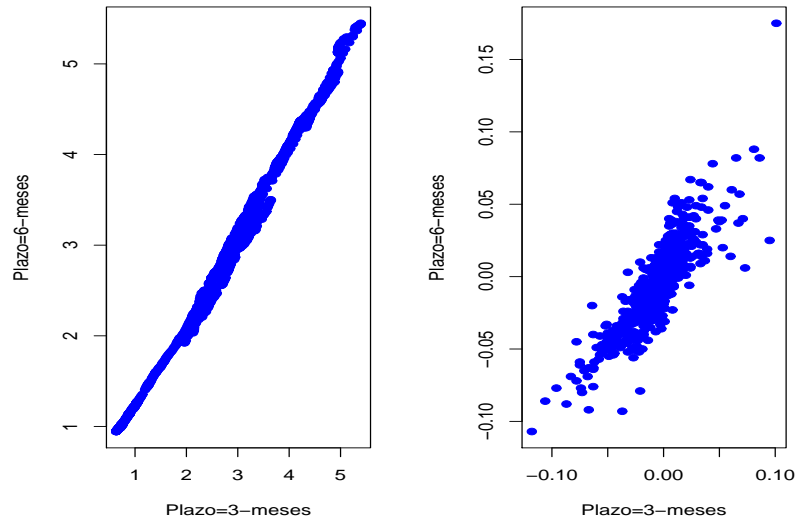
**Tabla 1.6:** Resumen de los datos del ajuste lineal a la serie de los tipos de interés

	Estimados	Error Estándar	t-valor	Pr(> t )
$\beta_0$	0.1953	0.0049	39.61	0.0000
$\beta_1$	0.9621	0.0017	578.96	0.0000

El modelo (1.50) 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 (1.10), se concluye que el modelo ajustado (1.50) 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 (1.50) se hace necesario entonces, considerar la serie de las diferencias para las series del EURIBOR. Así pues, se consideran las nuevas series



**Figura 1.9:** 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

- $d_{3t} = r_{3t} - r_{3,t-1} = (1 - L)r_{3t}$  para  $t \geq 2$ : serie de las variaciones (cambios) en los tipos de interés con plazo de vencimiento 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 6-meses;

y el modelo de regresión  $d_{6t} = \beta_0 + \beta_1 d_{3t} + \varepsilon_t$ . La figura (1.11) 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 (1.9(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:

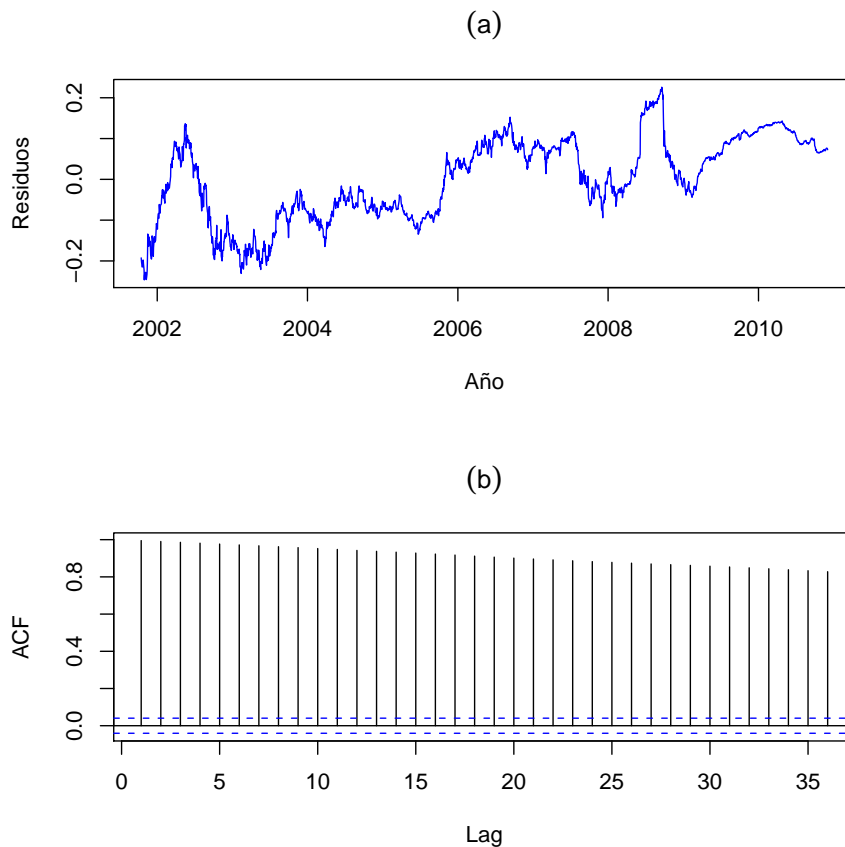
$$d_{6t} = 0.0001 + 0.9845d_{3t} + \varepsilon_t, \quad \hat{\sigma}_\varepsilon = 0.00894, \quad (1.51)$$

con  $R^2 = 70,94\%$ . Los errores estándar de los coeficientes se muestran en la tabla (1.7). En este nuevo modelo se observa una pequeña disminución de la dependencia lineal entre los tipos de interés, sin embargo, ésta sigue siendo alta.

**Tabla 1.7:** Resumen de los datos del ajuste lineal a la serie de las variaciones de los tipos de interés.

	Estimados	Error Estándar	t-valor	$\Pr(> t )$
$\beta_0$	0.0001	0.0002	0.75	0.4552
$\beta_1$	0.9845	0.0130	75.50	0.0000

En la figura (1.12) se muestran los gráficos de la *ACF* de los residuos obtenidos a partir del modelo (1.51). En este caso la *ACF* de los residuos muestra una pequeña serial correlación en los residuos para ciertos retardos. Esta serial correlación es débil y se puede considerar algunos de



**Figura 1.10:** 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

los modelos discutidos en secciones previas para modelizar esta dependencia, obteniéndose de esta manera un modelo de regresión lineal con una serie de tiempo como error.

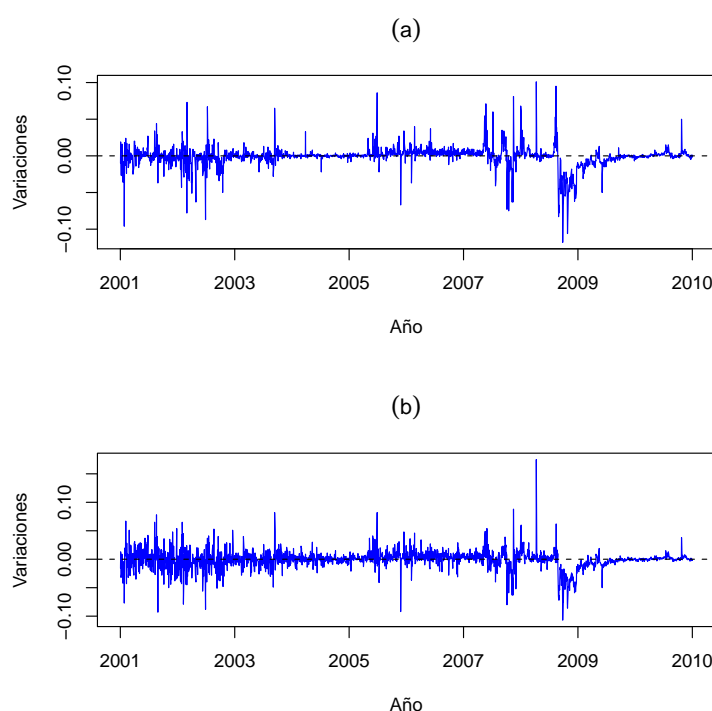
Como se señaló al comienzo de esta sección, el objetivo es discutir una metodología para construir modelos de regresión lineal con series de tiempo como errores. El método en si es relativamente sencillo, basta considerar un modelo de serie de tiempo, como los discutidos en este capítulo para modelizar la serie de los residuos y estimar el modelo completo de manera conjunta. Para ilustrar el procedimiento, considere considerérese el modelo de regresión lineal de la ecuación (1.51). Suponiendo que 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 (1.12) se tendrá en cuenta un modelo *MA*(1) para los residuos y se modificará el modelo de regresión lineal a:

$$d_{6t} = \beta_0 + \beta_1 d_{3t} + \varepsilon_t, \quad \varepsilon_t = \epsilon_t - \theta_1 \epsilon_{t-1}, \quad (1.52)$$

asumiendo  $\{\epsilon_t\}$  es un proceso ruido blanco. El modelo obtenido es un modelo de regresión lineal simple con una serie de tiempo como error. Es claro que, en la practica, 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 series de tiempo como error. Para la serie de tiempo de los tipos de interés del *EURIBOR*, el modelo ajustado, tomando en cuenta la ecuación (1.52), es:

$$d_{6t} = 0.0001 + 0.9844 d_{3t} + \varepsilon_t, \quad \varepsilon_t = \epsilon_t - 0.0160 \epsilon_{t-1}, \quad \hat{\sigma}_\epsilon = 0.00894, \quad (1.53)$$





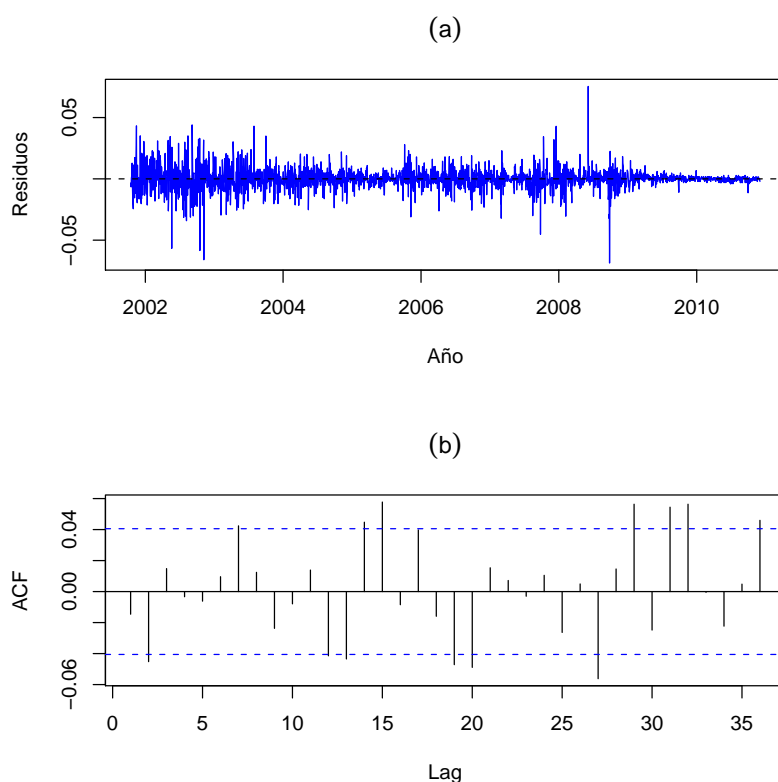
**Figura 1.11:** Serie de las variaciones (diferencias de primer orden) de los tipos de interés. (a) serie de tipos de interés con vencimiento 3-meses (b) serie de tipos de interés con vencimiento 6-meses

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 (1.50) 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 (1.51) y (1.53) son casi similares. En este caso particular, la aportación del modelo  $MA(1)$  a la serie no proporciona una mejora sustancial. Esto no debería sorprender ya que el coeficiente estimado de  $MA$  es pequeño desde el punto de vista numérico, aunque puede ser 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 y Box (1978) (discutido en la sección (1.3.2)) el cual se usa en lugar del estadístico de Durbin y 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.



**Figura 1.12:** 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

*Observación 1.1.* Para una serie de residuos  $\varepsilon_t$  con  $T$  observaciones, el estadístico de *Durbin-Watson* se define por:

$$DW = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T \varepsilon_t^2}. \quad (1.54)$$

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  $\{\varepsilon_t\}$ .

### 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:

- *Heteroscedásticidad consistente (HC)*, (ver, [White \(1980\)](#).)
- *Heteroscedásticidad y Autocorrelación consistente (HAC)* (ver, [Newey y West \(1987\)](#) )

El estimador de [White \(1980\)](#) es,

$$\text{Cov}(\hat{\beta})_{HC} = \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \left[ \frac{T}{T-k} \sum_{t=1}^T \hat{\varepsilon}_t^2 x_t x_t' \right] \left[ \sum_{t=1}^T x_t x_t' \right]^{-1}, \quad (1.55)$$

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

$$\text{Cov}(\hat{\beta})_{HAC} = \left[ \sum_{t=1}^T x_t x_t' \right]^{-1} \hat{C}_{HAC} \left[ \sum_{t=1}^T x_t x_t' \right]^{-1}, \quad (1.56)$$

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  $\omega_j$  es la función de pesos de Bartlett definida por

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

## 1.7 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 y Wei (1988) y Tiao y 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:

$$(1 - L)^d x_t = \varepsilon_t, \quad -0.5 < d < 0.5, \quad (1.57)$$

donde  $\{\varepsilon_t\}$  es un proceso ruido blanco. Resumimos alguna de las propiedades del modelo (1.57) a continuación, (para detalles de este modelo ver Hosking (1981), Granger y Joyeux (2001), y Jun (2001)).

- Si  $d < 0.5$ , entonces  $x_t$  es un proceso débilmente estacionario con representación *MA* infinita

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

- Si  $d > -0.5$ , entonces  $x_t$  es invertible representación *AR* infinita

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

- 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

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

- 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

$$f(\omega) \sim \omega^{-2d}, \quad \omega \rightarrow 0, \quad (1.58)$$

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

En esta sección, el estudio del comportamiento de la ACF de  $x_t$ , es de particular interés para aquellos procesos con  $d < 0.5$ . La propiedad que expresa que  $\rho_k \sim k^{2d-1}$ , cuyo 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.

## 1.8 Modelos de Heteroscedásticidad Condicional

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 acciones que actualmente están en poder del inversionista. La idea detrás de la comprensión del comportamiento de la volatilidad de los activos es el de 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 heteroscedásticidad 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 heteroscedásticidad 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.

### 1.8.1 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,

$$x_t = \mu_t + \varepsilon_t = \mathbb{E}(x_t | \mathcal{F}_{t-1}) + \varepsilon_t \quad (1.59)$$

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  $\varepsilon_t$  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 heteroscedásticidad condicional suponen que el segundo momento condicional depende del tiempo, es decir,

$$\sigma_t^2 = \text{Var}(x_t | \mathcal{F}_{t-1}) = \mathbb{E}((x_t - \mu_t)^2 | \mathcal{F}_{t-1}) = \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = h_t \quad (1.60)$$

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  $\varepsilon_t$  como la *rentabilidad corregida en media* o *impulso* del activo.

Los modelos de heteroscedásticidad condicional se pueden clasificar en dos categorías generales. La primera categoría, la que 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.

### 1.8.2 El modelo ARCH

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

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \end{aligned} \quad (1.61)$$

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  $\alpha_i$  garantiza que la varianza condicional  $h_t$  sea positiva.

*Observación 1.2.* Algunos autores usan  $\sigma_t^2$  para denotar la varianza condicional en la ecuación (1.61) 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 (1.61).

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

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \eta_p, \quad (1.62)$$

donde  $\eta_p \equiv \varepsilon_t^2 - h_t$ . Teniendo en cuenta a teoría de los modelos autorregresivos, si las raíces de la ecuación característica del proceso  $AR$  están fuera del círculo unidad, entonces el proceso es estacionario y además se puede calcular la varianza incondicional de  $\varepsilon_t$  como

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

Teniendo en cuenta la expresión (1.61), 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  $\varepsilon_{t-i}^2$  para  $i \in \{1, \dots, p\}$ . Por lo tanto, valores grandes de  $\varepsilon_{t-i}$  (en módulo) dan lugar a valores grandes de  $h_t$ . En consecuencia,  $\varepsilon_t$  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:

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.\end{aligned} \quad (1.63)$$

Entonces: se tiene que

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

Por otra parte, bajo la suposición de estacionariedad, la varianza incondicional de  $\varepsilon_t$  es

$$\sigma^2 = \mathbb{E}(\varepsilon_t^2) = \frac{\alpha_0}{1 - \alpha_1}.$$

suponiendo normalidad en  $\vartheta_t$

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

y por lo tanto

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

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

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

y, por lo tanto

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

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

$$K = \frac{\mu_4}{\mu^2} \mu_2^2 = 3 \frac{1 - \alpha^2}{1 - 3\alpha_1^2} > 3,$$

donde  $\mu_2$  es el momento de segundo orden no condicional de  $\varepsilon_t$ . En esta ultima expresión se ve reflejado el exceso de kurtosis de  $\varepsilon_t$ .

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 órdenes mayores implican las 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 (1.61) Bajo la hipótesis de normalidad de  $\vartheta_t$ , se tiene que la distribución de  $\varepsilon_t$  condicionado a  $\mathcal{F}_{t-1}$ , la información disponible hasta el instante  $t - 1$  es una normal con media cero y varianza  $h_t$  y la función de máxima verosimilitud adopta la forma:

$$\begin{aligned} f(\varepsilon_1, \dots, \varepsilon_T) &= f(\varepsilon_T | \mathcal{F}_{T-1}) f(\varepsilon_{T-1} | \mathcal{F}_{T-2}) \dots f(\varepsilon_{p+1} | \mathcal{F}_p) f(\varepsilon_1, \dots, \varepsilon_p) \\ &= \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-\varepsilon_t^2}{2h_t}\right) f(\varepsilon_1, \dots, \varepsilon_p) \end{aligned} \quad (1.64)$$

donde  $f(\varepsilon_1, \dots, \varepsilon_p)$  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

$$f(\varepsilon_{p+1}, \dots, \varepsilon_T | \varepsilon_1, \dots, \varepsilon_p) = \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-\varepsilon_t^2}{2h_t}\right). \quad (1.65)$$

Puesto que maximizar la función de verosimilitud (1.65) equivale a maximizar su logaritmo, se considera entonces

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

donde  $\ell$  denota el logaritmo de la función de verosimilitud, es decir,

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

Ahora bien, ya que el primer término de la suma no depende de ningún parámetro, los estimadores de máxima verosimilitud condicional de  $\alpha_0$  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  $\vartheta_t$  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 (1.61), 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 para el segundo retardo es:

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

En general para la predicción a  $k$  retardos,

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

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

### 1.8.3 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 heteroscedásticidad 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

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ h_t &= \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \end{aligned} \quad (1.66)$$



donde  $\vartheta_t$  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  $\alpha_0 \geq 0$ ,  $\beta_j \geq 0$ ,  $i \in \{1, \dots, p\}$ ,  $j \in \{1, \dots, q\}$ .

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

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

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  $\varepsilon_t^2$ , siendo  $m = \max(p, q)$ .

Sea  $\eta_t = \varepsilon_t^2 - h_t$ . Sustituyendo  $h_t = \varepsilon_t - \eta_t$  en (1.66) se tiene

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \eta_t - \sum_{j=1}^q \beta_j \eta_{t-j} \quad (1.67)$$

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

*Observación 1.3.* Se puede demostrar que el modelo  $GARCH$  definido como un modelo  $ARMA$  satisface  $\mathbb{E}(\eta_t) = 0$  y  $\text{Cov}(\eta_t, \eta_{t-j}) = 0$  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  $GARCH(1, 1)$

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \end{aligned} \quad (1.68)$$

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 (1.68) 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 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 (1.68) se deduce que valores grandes de  $\varepsilon_t^2$  o  $h_{t-1}$  dan lugar a valores grandes de  $h_t$ . Esto significa que grandes valores de  $\varepsilon_t - 1^2$  tienden a ir seguidos de grandes valores de  $\varepsilon_t^2$ , 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  $\varepsilon_t$ , y suponiendo normalidad en  $\vartheta_t$ , se tiene

$$E(\varepsilon_t^4) = 3E(\alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1h_{t-1})^2.$$

Desarrollando la expresión anterior y despejando  $\mu_4$ , 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

$$K = \frac{\mu_4}{\mu_2^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3. \quad (1.69)$$

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

Procediendo de manera similar que para el modelo  $ARCH$  y bajo la suposición de que  $\vartheta_t$  sigue una distribución normal, se obtiene una expresión para la función de verosimilitud gaussiana (condicional) dada por

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

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  $\varepsilon_t$ , 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  $\alpha_0$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)'$  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  $\kappa$  es un entero ( $\kappa > p$ ).

Hasta ahora se ha asumido que las innovaciones  $\vartheta_t$  siguen una distribución normal, lo cual implica que la distribución condicional de  $\varepsilon_t$  es una normal con media cero y varianza  $h_t$ . Por otro lado, la distribución incondicional de  $\varepsilon_t$  en los modelos  $GARCH$  no es una normal. Por ejemplo, para un modelo  $GARCH(1, 1)$ , la kurtosis, la kurtosis incondicional de  $\varepsilon_t$  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  $\varepsilon_t$  en un  $GARCH(1, 1)$ , dada por (1.69), es menor que la de la serie observada, es decir, el modelo  $GARCH(1, 1)$  con  $\vartheta_t$  con distribución normal, no tienen la capacidad de recoger todo el peso de las colas de la distribución incondicional de  $\varepsilon_t$ . Para solucionar este inconveniente se han desarrollado nuevos modelos, aunque una de las alternativas más comunes consiste en considerar otra distribución para  $\vartheta_t$ .

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

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

donde  $\Gamma(\cdot)$  es la función Gamma.

La función de verosimilitud en este caso se expresa por

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

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

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

En este caso, la kurtosis incondicional de  $\varepsilon_t$  es mayor que la obtenida cuando se asume que la distribución de  $\vartheta_t$  es normal.

*Observación 1.4.* 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 la predicciones para la varianza condicional en los procesos *GARCH*. Sea el modelo *GARCH*(1, 1) con ecuación (1.68), 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  $k-1 \geq 0$ . 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

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

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

### 1.8.4 Modelos EGARCH

Como ya se ha comentado antes, en el caso de las serie de rentabilidades de activos financieros los periodos 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

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t} \\ \ln(h_t) &= \alpha_t + \sum_{k=1}^{\infty} \beta_k g(\vartheta_{t-k}), \quad \beta_1 = 1\end{aligned}\quad (1.70)$$

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  $\vartheta_t$ . La función  $g$  considerada en esta sección es

$$g(\vartheta_t) = \theta \vartheta_t + \gamma [|\vartheta_t| - \mathbb{E}|\vartheta_t|.] \quad (1.71)$$

Como primer 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  $\theta$  y  $\gamma$  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  $\vartheta_t$  con pendiente  $\theta + \gamma$  y para  $-\infty < \vartheta_t < 0$ , la función  $g(\vartheta_t)$  es lineal en  $\vartheta_t$  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 sólo 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*( $\infty$ ) dada en (1.70). Así, un *EGARCH*( $p, q$ ) vendría dado por

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t} \\ \ln(h_t) &= \alpha_t + \frac{1 + a_1 L + \dots + a_q L^q}{1 - b_1 L + \dots + b_p L^p} g(\vartheta_{t-k}).\end{aligned}\quad (1.72)$$

Atendiendo a la teoría de los modelos *ARMA*, el proceso  $\{\ln(h_t) - \alpha_t\}$  será estacionario si las raíces de  $1 - b_1 L + \dots + b_p L^p = 0$  están fuera del círculo unidad. 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

$$\begin{aligned}\varepsilon_t &= \vartheta_t \sqrt{h_t} \\ \ln(h_t) &= \alpha_0 + g(\vartheta_{t-1}) - b_1 \ln(h_{t-1}).\end{aligned}\quad (1.73)$$

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

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

Se supondrá que  $\{\vartheta_t\}$  tiene distribución Normal de media cero y varianza 1 y 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  $\vartheta_t$  y la volatilidad un instante adelante  $h_{t+1}$ , manteniendo constante todo el pasado. Según esto, para un modelo *GARCH*(1, 1) como el de (1.68) se tendría

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

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  $h_t$  sólo trasladan la curva verticalmente pero no afectan a su forma. Para un *EGARCH*(0, 1) definido en (1.73), se tiene

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

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

### 1.8.5 Modelos *IGARCH*

Cuando se aplica el modelo *GARCH*(1, 1) a series de tiempo de frecuencia alta, los valores estimados  $\alpha_1$  y  $\beta_1$  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

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ h_t &= \alpha_0 + (1 - \beta_1)\varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \end{aligned} \quad (1.74)$$

donde  $\vartheta_t$  es definido como antes. En estas condiciones, la varianza incondicional de  $\varepsilon_t$  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 (1.68), 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  $\vartheta_t$  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

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

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,

$$\mathbb{E}(\ln(\alpha_1 \vartheta_t^2 + \beta_1)) < \ln(\mathbb{E}(\alpha_1 \vartheta_t^2 + \beta_1)) = \ln(\alpha_1 + \beta_1). \quad (1.75)$$

Así pues, si el proceso es débilmente estacionario, también será estrictamente estacionario. La desigualdad (1.75) 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.

### 1.8.6 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

$$\begin{aligned} x_t &= g(h_t, \xi) + \varepsilon_t \\ \varepsilon_t &= \vartheta_t \sqrt{h_t}. \end{aligned} \quad (1.76)$$

Donde se asume además que la varianza condicional  $h_t$  sigue un modelo  $GARCH$ . En la práctica, es habitual seleccionar la función  $g(h_t, \xi) = \xi_0 + \xi_1 h_t$ , que se interpreta como una especie de prima de riesgo. En este sentido, los coeficientes  $\xi_0$  y  $\xi_1$  son constantes. Un valor positivo de  $\xi_1$  indica que el rendimiento está relacionado de manera positiva a su volatilidad.

Una formulación simple del modelo  $GARCH-M$  es el modelo  $GARCH(1, 1)-M$  dado por

$$\begin{aligned} x_t &= \xi_0 + \xi_1 h_t + \varepsilon_t, \\ \varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2. \end{aligned} \quad (1.77)$$

### 1.8.7 Modelo TGARCH

Al inicio de esta sección se mencionaron algunas características de la volatilidad 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

$$h_t = \alpha_0 + \sum_{i=1}^p (\alpha_i + \gamma_i \mathbb{1}_{\{\varepsilon_{t-i} < 0\}}) \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \quad (1.78)$$

donde  $\mathbb{1}_{\{\varepsilon_{t-i} < 0\}}$  es una variable indicador para valores negativos de  $\varepsilon_{t-i}$ , 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  $\alpha_i, \gamma_i$ , y  $\beta_j$  son parámetros no negativos que satisfacen condiciones similares a las de los modelo *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.

### 1.8.8 Modelo 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

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ \alpha_p(L) \ln(h_t) &= \alpha_0 + \eta_t, \end{aligned} \quad (1.79)$$

donde  $\alpha_p(L) = (1 - \alpha_1 L - \dots - \alpha_p L^p)$  es el polinomio característico asociado al proceso con todos sus ceros mayores que 1 en modulo,  $\alpha_0$  es una constante,  $\eta_t$  son variables aleatorias independientes e idénticamente distribuidos  $N(0, \sigma_\eta^2)$ ,  $\{\vartheta_t\}$  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 (1.79) para  $p = 1$  es la siguiente

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t} \\ \ln(h_t) &= \alpha_0 + \alpha_1 \ln(h_{t-1}) + \eta_t. \end{aligned} \quad (1.80)$$

Una representación alternativa del modelo (1.80) de volatilidad estocástica es la siguiente:

$$\varepsilon_t = v \exp\left(\frac{1}{2}z_t\right)\vartheta_t, \quad z_t = \theta_1 z_{t-1} + \eta_t \quad (1.81)$$

donde  $v$  es un parámetro de escala,  $\vartheta_t$  es un proceso independiente e idénticamente distribuido con media cero y varianza uno,  $z_t$  es una variable aleatoria independiente de  $\vartheta_t$ . En el modelo (1.80) 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.

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

donde  $x_t = \log(\varepsilon_t^2)$ ,  $\mu = \log(v + \mathbb{E}[\log(\vartheta_t^2)])$  y  $\xi_t = \log(\vartheta_t^2) - \mathbb{E}[\log(\vartheta_t^2)]$  es un ruido blanco no gaussiano de media cero y varianza  $\sigma_\xi^2$  cuyas propiedades dependen de  $\vartheta_t$ . Por ejemplo, si  $\vartheta_t$  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  $\sigma_\xi^2 = 4.93$ .

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).

### 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

$$\begin{aligned} \varepsilon_t &= \vartheta_t \sqrt{h_t}, \\ h_t &= \sigma^2 \exp(u_t), \\ (1-L)^d u_t &= \eta_t \end{aligned} \quad (1.82)$$

donde  $\sigma > 0$ ,  $\vartheta_t$  y  $\eta_t$  son variables aleatorias normales independientes e idénticamente distribuidas de media cero y varianzas 1 y  $\sigma_\eta^2$ , 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 (1.82), 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 una 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, es que esta es complicada, sin embargo el parámetro  $d$  de diferencia fraccional se puede estimar mediante el uso de un método de probabilidad de cuasi-máxima verosimilitud o un método de regresión.

## 1.9 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,

$$y_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots). \quad (1.83)$$

Si tenemos en cuenta el modelo,

$$x_t = \mu + \sum_{i=0}^{\infty} \eta_i \varepsilon_{t-i}, \quad (1.84)$$

donde  $\mu$  es una constante,  $\eta_i$  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 (1.83) es una función lineal de la sucesión de variables  $\{\varepsilon_t\}$ . Cualquier no linealidad en la función  $f$  tendrá como resultado la no linealidad del modelo (1.83). 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  $\sigma$ -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  $\{a_{t-1}, a_{t-2}, \dots\}$ . Entonces, la media y varianza condicional de  $x_t$ , dada  $\mathcal{F}_{t-1}$ , son

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

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

$$x_t = g(\mathcal{F}_{t-1}) + \varepsilon_t = g(\mathcal{F}_{t-1}) + \sqrt{h(\mathcal{F}_{t-1})} \vartheta_t, \quad (1.85)$$

donde  $\vartheta_t$  son los impulsos estandarizados o (innovaciones). Para series  $x_t$  lineales, en la ecuación (1.83),  $g(\cdot)$  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 es (1.85) 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 heteroscedasticidad condicional discutidos en secciones previas, son modelos no lineales en varianza, puesto que su varianza condicional  $\sigma_t^2$  depende del tiempo, excepto para el caso de los modelos *GARCH-M*, en los que  $\mu_t$  depende de  $\sigma_t^2$ . 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 estocástico 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

el análisis de modelos no-lineales. Se entenderán por *modelos no lineales en media*, a los modelos no lineales para la esperanza condicional.

### 1.9.1 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 lo 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.

### 1.9.2 Modelo Bilineal

[Granger y Andersen \(1978\)](#) introdujeron el *modelo Bilineal* como una extensión natural para la no linealidad. Los autores aplicaron los términos de segundo orden en la expansión de series de Taylor de la función  $f$  del modelo lineal para mejorar la aproximación, obteniendo el modelo definido como

$$x_t = \mu + \sum_{i=1}^p \phi_i x_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^m \sum_{j=1}^s \beta_{ij} x_{t-i} \varepsilon_{t-j} + \varepsilon_t, \quad (1.86)$$

donde  $p, q, m$  y  $s$  son números enteros no negativos. [Suba Rao y Gabr \(1984\)](#) y [Liu y Brockwell \(1988\)](#) estudiaron ampliamente este modelo. A partir de su representación en espacios de estados y usando la ecuación de transición de estados para expresar los estados como un producto de los impulsos del pasado y los vectores de coeficientes aleatorios, es posible obtener condiciones para la estacionariedad del proceso. Una generalización del modelo bilineal con heteroscedasticidad condicional es

$$x_t = \mu + \sum_{i=1}^s \beta_i \varepsilon_{t-i} \varepsilon_t, \quad (1.87)$$

donde  $\{\varepsilon_t\}$  es una serie de ruido blanco. Los momentos condicionales de primer y segundo orden de  $x_t$  son

$$\mathbb{E}(x_t | \mathcal{F}_{t-1}) = \mu, \quad \text{Var}(x_t | \mathcal{F}_{t-1}) = \left( 1 + \sum_{i=1}^s \beta_i \varepsilon_{t-i} \right)^2 \sigma_\varepsilon^2.$$

### 1.9.3 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

$$x_t = \begin{cases} \phi_{0,1} + \phi_{1,1}x_{t-1} + \varepsilon_t, & \text{si } q_{t-1} < c, \\ \phi_{0,2} + \phi_{1,2}x_{t-1} + \varepsilon_t, & \text{si } q_{t-1} \geq c, \end{cases} \quad (1.88)$$

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 (1.88) se geoméricamente 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 [Petruccielli y Woolford \(1984\)](#) and [Chen y Tsay \(1991\)](#).

#### 1.9.4 Modelos *SETAR*

Los *auto-excitables/flexibles* denotado 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

$$x_t = \begin{cases} \phi_{0,1} + \phi_{1,1}x_{t-1} + \varepsilon_t, & \text{si } r_{t-1} < \gamma, \\ \phi_{0,2} + \phi_{1,2}x_{t-1} + \varepsilon_t, & \text{si } r_{t-1} \geq \gamma, \end{cases} \quad (1.89)$$

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,

$$x_t = \begin{cases} -1,5x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} < 0, \\ 0,5x_{t-1} + \varepsilon_t, & \text{si } x_{t-1} \geq 0, \end{cases} \quad (1.90)$$

donde las  $\varepsilon_t$  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 rijan 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

$$x_t = \phi_{0,j} + \phi_{1,j}x_{t-1} - \dots - \phi_{p,j}x_{t-p} + \varepsilon_{t,j}, \quad \gamma_{j-1} \geq x_{t-d} < \gamma_j, \quad (1.91)$$

donde  $k$  y  $d$  son enteros positivos,  $j = 1, \dots, k$ , los  $\gamma_i$  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  $\sigma_j^2$  y son mutuamente independientes para distintos valores de  $j$ . El parámetro  $d$  es el *parámetro de retardo* y  $\gamma_j$  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 \(1993\)](#), [Chan y Tsay \(1998\)](#). El modelo (1.91) 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*.

### 1.9.5 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  $\mu_t$ . Para solventar esto, se proponen los *modelos suavizado TAR* y se denota por *STAR*; Chan y Tong (1986) y Teräsvirta (1994). Una serie de tiempo  $x_t$  sigue un modelo *STAR(p)* con dos regímenes si satisface

$$x_t = c_0 + \sum_{i=1}^p \phi_{0,i} x_{t-i} + F\left(\frac{x_{t-d} - \Delta}{s}\right) \left(c_1 + \sum_{i=1}^p \phi_{1,i} x_{t-i}\right) + \varepsilon_t \quad (1.92)$$

donde  $d$  es el parámetro de retardo,  $\Delta$  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 (1.92), 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 unidad. 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.

### 1.9.6 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

$$x_t = \begin{cases} \phi_{0,1} + \sum_{i=1}^p \phi_{1,i} x_{t-i} + \varepsilon_{1,t}, & \text{si } s_t = 1, \\ \phi_{0,2} + \sum_{i=1}^p \phi_{2,i} x_{t-i} + \varepsilon_{2,t}, & \text{si } s_t = 2. \end{cases} \quad (1.93)$$

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, [Hamilton \(1990\)](#) usa algoritmos *EM*, un método estadístico que itera entre el cálculo de esperanzas y maximizaciones; y [McCulloch y Tsay \(1994\)](#) quienes considera el método de Monte Carlo por Cadenas de Markov. En el último caso los autores consideran una versión generalizada del modelo (1.93) 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  $t - 1$ .

### 1.9.7 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, estimación de mínimos cuadrados locales, y redes neuronales.

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

$$Y_t = m(X_t) + \varepsilon_t \quad (1.94)$$

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  $x$  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  $\varepsilon_t$  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:

$$\hat{m}(x) = \frac{1}{T} \sum_{t=1}^T \omega_t(x) y_t \quad (1.95)$$

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.

### 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

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \quad \int K_h(u) du = 1. \quad (1.96)$$

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

$$\omega_t(x) = \frac{K_h(x - x_t)}{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t)}, \quad (1.97)$$

Si se consideran los pesos en (1.97) para el estimador en (1.95) se obtiene el estimador por núcleo de Nadaraya-Watson (ver [Nadaraya \(1964\)](#) y [Watson \(1964\)](#)):

$$\hat{m}(x) = \frac{1}{T} \sum_{t=1}^T \omega_t(x) y_t = \frac{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t) y_t}{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t)}, \quad (1.98)$$

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*

$$K_h(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{x^2}{2h^2}\right) \quad (1.99)$$

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

$$K_h(x) = \frac{0.75}{h} \left(1 - \frac{x^2}{h^2}\right) \mathbb{1}_{\{|x/h| \leq 1\}}. \quad (1.100)$$

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  $h$ .

### Selección del Parámetro de Suavizado

Para la elección de  $h$  existen diversos métodos en la literatura, ver por ejemplo [Härdle \(1990\)](#) y [Fan y 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

$$\min_h CV(h) = \min_h \frac{1}{T} \sum_{j=1}^T (y_j - \hat{m}_{h,j}(x_j))^2 W(x_j) \quad (1.101)$$

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

$$MISE = \mathbb{E} \left( \int_{-\infty}^{\infty} [\hat{m}(x) - m(x)]^2 dx \right) \quad (1.102)$$

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 y 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 y Yao \(2003\)](#) y las referencias incluidas en el mismo.

### 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 practica se realizan ciertas suposiciones sobre la función de regresión  $m(\cdot)$  tales como: 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  $\alpha$  y  $\beta$  tales que minimicen la función

$$L(\alpha, \beta) = \sum_{t=1}^T (y_t - \alpha - \beta(x_t - x))^2 K_h(x - t - x), \quad (1.103)$$

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 (1.103) 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

$$\hat{a} = \frac{\sum_{t=1}^T \omega_t y_t}{\sum_{t=1}^T \omega_t} \quad (1.104)$$

donde  $\omega_t$  se define como

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

Dado que en la practica existe la posibilidad de que el denominador en *nonpara10LLRest* se anule, se suele usar

$$\hat{m}(x) = \frac{\sum_{t=1}^T \omega_t y_t}{\sum_{t=1}^T \omega_t + 1/T^2}. \quad (1.105)$$

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

### 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

$$x_t = m(x_{t-1}) + \varepsilon_t. \quad (1.106)$$

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

$$K_h(x) = \frac{1}{(h\sqrt{2\pi})^p |\Sigma|^{1/2}} \exp\left(-\frac{1}{2h^2} x' \Sigma^{-1} x\right), \quad (1.107)$$

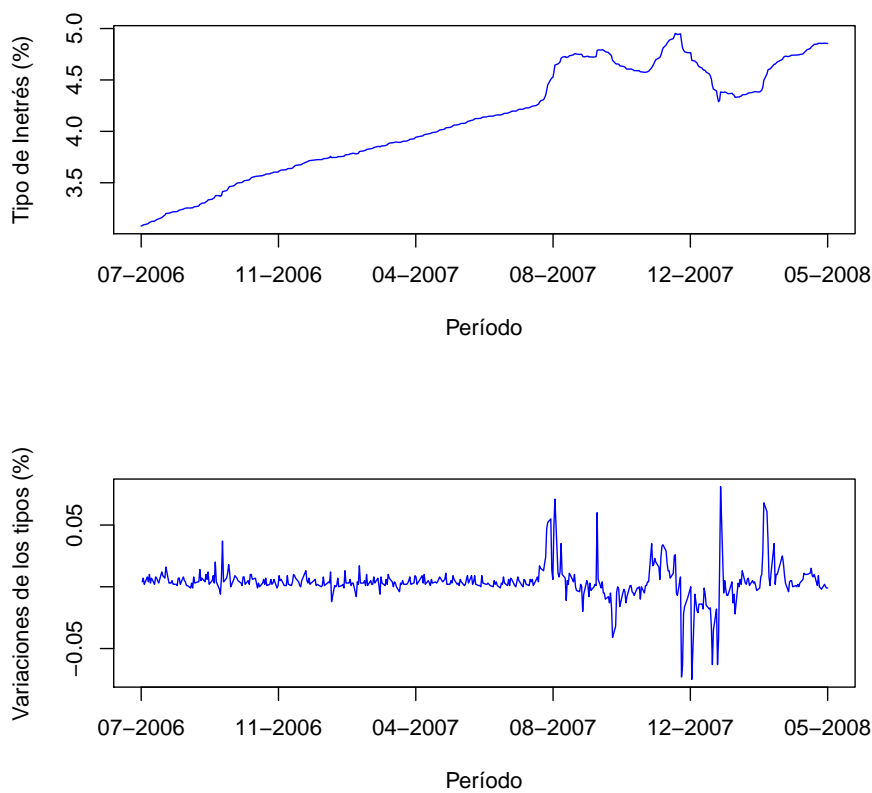
donde  $p$  es el número de variables explicativas (o retardos) y  $\Sigma$  es una matriz definida positiva.

**Ejemplo 1.1.** Para la ilustrar la aplicación de las técnicas de suavizado en finanzas, se considera la 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. El modelo que se propone es

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

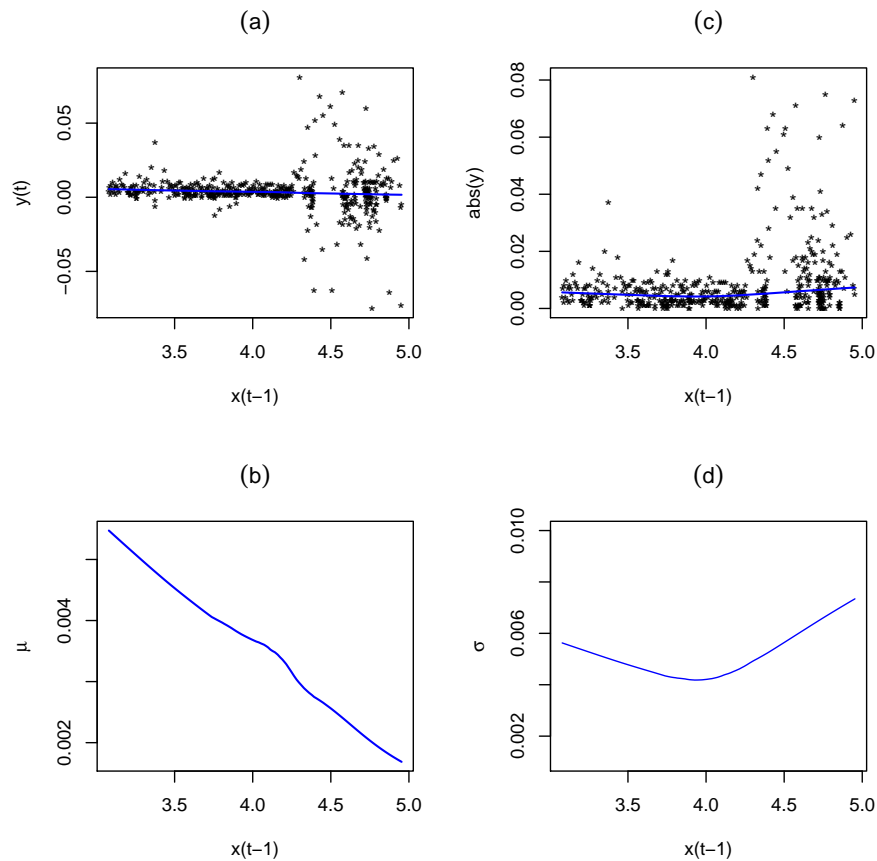


donde  $x_t$  es el tipo de interés con plazo de vencimiento 3-meses,  $y_t = x_t - x_{t-1}$ ,  $W_t$  es un movimiento Browniano estándar, y  $\mu(\cdot)$  y  $\sigma(\cdot)$  son funciones suaves de  $x_{t-1}$ . En el modelo considerado,  $\mu(x_{t-1}) = \mathbb{E}(y_t|x_{t-1})$ . Usaremos el estimador 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 (1.13) 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 1.13:** 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 (1.14) 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 (1.14(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 (1.14(b)) se observa que hay una tendencia. En cuanto a la volatilidad, se muestra en la (1.14(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 (1.14(d)) 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.



**Figura 1.14:** 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  $\sigma(x_{t-1})$

### 1.9.8 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 y Tsay \(1993a\)](#) proponen el *modelo de coeficiente funcional autorregresivo* denotado por *FAR* expresado por

$$x_t = f_1(X_{t-1})x_{t-1} + \dots + f_p(X_{t-1})x_{t-p} + \varepsilon_t, \quad (1.108)$$

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 segura 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  $X_{t-1}$  es pequeña ( véase [Cai et al. \(2000\)](#)).

### 1.9.9 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 suavizado. 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,

$$x_t = f_0(t) + \sum_{i=1}^p f_i(x_{t-i}) + \varepsilon_t, \quad (1.109)$$

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 y Tsay \(1993b\)](#)). [Chen et al. \(1995\)](#) consideran que las estadísticas de prueba para verificar la hipótesis de aditividad, la cual es bastante restrictiva y debe ser examinada con cuidado en su aplicación.

### 1.9.10 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 y 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

$$S_t = f(S_{t-1}) + \eta_t, \quad x_t = g_t(S_t) + v_t, \quad (1.110)$$

donde

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

La suposición del conocimiento de  $f_t(\cdot)$  y  $g_t(\cdot)$  en (1.110) 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.

### 1.9.11 Redes Neuronales

Las *redes neuronales* constituyen una metodología de bastante éxito en el análisis moderno de datos. [Cheng y Titterington \(1994\)](#) proporcionan un enfoque desde el punto de vista estadístico para las redes neuronales. En la búsqueda de métodos para explicar  $Y_t$  a partir de  $x_t$  un vector  $n$ -dimensional con  $n \geq 1$ , denominado *vector de entrada* o *input*, la función  $m$  desconocida, la cual hemos considerado a lo largo de esta sección, podría ser aproximada mediante

$$m \cong f_o \left( \alpha_{0o} + \sum_{j \rightarrow o} \omega_{jo} \eta_j \right) \quad (1.111)$$

donde  $x_t = (x_{t,1}, \dots)$  es el vector de inputs,  $\alpha_{0o}, \alpha_{0j}$  son coeficientes desconocidos,  $\omega_{ij}$  son los pesos desconocidos también,  $f_o, f_j$  son funciones conocidas y preespecificadas, y  $\eta_j$  son los nodos de la capa oculta. Las funciones  $f_j$  se conocen como *funciones de activación*. La figura (1.15) responde a la formula anterior y en ella se presenta la llamada *red neuronal "perceptron" con una capa oculta*, Con un número  $I$  de nodos de la *capa de entrada* que coincide con la dimensión de  $x_t$  y la

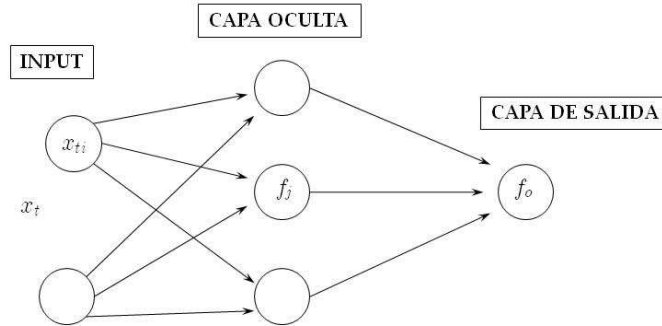


Figura 1.15: Red Neuronal con una capa Oculta (Perceptron)

constante de entrada  $\alpha_{0j}$ , un número  $J$  de nodos definidos por

$$\eta_j = f_j \left( \alpha_{0j} + \sum_{i \rightarrow j} \omega_{ij} x_{ti} \right) \quad (1.112)$$

de la capa oculta y un nodo de la *capa de salida*. Entre las funciones de activación que suelen considerarse están las funciones logísticas,

$$f_j(z) = \frac{\exp(z)}{1 + \exp(z)}, \quad j \geq 1$$

Por ejemplo, el  $j$ -ésimo nodo de la capa oculta de la red neuronal de la figura (1.15), tomando como función de activación la función logística, es

$$\eta_j = \frac{\exp(\alpha_{0j} + \omega_{1j}x_{t1} + \omega_{2j}x_{t2})}{1 + \exp(\alpha_{0j} + \omega_{1j}x_{t1} + \omega_{2j}x_{t2})}, j = 1, \dots$$

La función de activación  $f_o(\cdot)$  puede ser considerada como una función lineal o una función denominada *neurona umbral* (threshold neuron). Si  $f_o$  es una función lineal, entonces  $m$  se aproxima por

$$m \cong \alpha_{0o} + \sum_{j=1}^k \omega_{jo} \eta_j \quad (1.113)$$

donde  $k$  es el número de nodos en la capa oculta. O bien si es una función umbral, es decir,

$$f_o(z) \begin{cases} 1, & \text{si } z > 0, \\ 0, & \text{si } z \leq 0. \end{cases}$$

La base fundamental del éxito de esta metodología es que permiten aproximar cualquier función continua uniformemente en conjuntos compacto, tanto como se quiera, para ello solo basta con aumentar el número de nodos de la capa oculta, propiedad conocida como de aproximación universal, véanse [Hornik \(1993\)](#) y [Chen y Chen \(1995\)](#). La ecuación (1.111) suele ser considerada como una función semiparamétrica puesto que su forma funcional es conocida, pero el número de nodos y los pesos son desconocidos.

Evidentemente dependiendo del problema en estudio, será necesario considerar redes neuronales más complejas.

Si se permiten conexiones directas desde la capa de entrada a la capa de salida, entonces la red se convierte en

$$o_t \cong f_o \left[ \alpha_{0o} + \sum_{i \rightarrow o} \alpha_{io} x_{ti} + \sum_{j \rightarrow o} \omega_{jo} f_j \left( \alpha_{0j} + \sum_{i \rightarrow j} \omega_{ij} x_{ti} \right) \right]$$

donde el primer término de la suma agrupa todos los nodos asociados a los inputs.

La aplicación de esta metodología involucra dos pasos. El primero de ellos es el *entrenamiento* de la red, entendiendo por esto el diseño de la red, y la determinación de los parámetros involucrados. El segundo paso está relacionado con la predicción. Para llevar a cabo esto, en la etapa de entrenamiento, la muestra en estudio es dividida en dos submuestras no superpuestas. La primera submuestra será usada para la estimación de los parámetros del modelo, entonces la segunda muestra es usada para llevar a cabo la predicción y determinar su capacidad predictiva. Este procedimiento usa las ideas de la validación cruzada en la selección de modelos estadísticos, con el objeto de seleccionar la “mejor” red para hacer predicción. En aplicaciones al análisis de las series de tiempo, sean  $\{(x_t, y_t)\}_{t=1}^T$  la muestra de entrenamiento. Para una red dada, sea  $o_t$  dado por (1.111) la salida de la red con input  $x_t$ . Entonces se estiman los parámetros de la red, por ejemplo

$$\text{mín } S^2 = \text{mín } \sum_{t=1}^T (y_t - o_t)^2$$

Para resolver el problema de optimización se pueden utilizar diversos algoritmos numéricos, como por ejemplo el algoritmo de aprendizaje “*Back propagation*” entre otros introducido por Bryson y Ho (1969). Posteriormente la red neuronal es validada mediante la predicción.

### 1.9.12 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 incluyen tanto enfoques paramétricos como no paramétricos.

#### 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.

#### 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 y 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,  $\varepsilon_t$  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  $\beta_i$  es el coeficiente de  $\varepsilon_{t-i}^2$  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  $m - p - q$ .

### Test paramétrico

Entre los test paramétricos tenemos:

#### 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 un modelos lineales AR. Sea el modelo lineal  $AR(p)$

$$x_t = X'_{t-1}\phi + \varepsilon, \quad (1.114)$$

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

1. Se estima  $\phi$  por mínimos cuadrados, se calculan los residuos  $\varepsilon_t = x_t - \hat{x}_t$ , donde  $\hat{x}_t = X'_{t-1}\hat{\phi}$ , 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

$$\hat{\varepsilon}_t = X'_{t-1}\alpha_1 + M'_{t-1}\alpha_2 + v_t, \quad (1.115)$$

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

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

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

3. La idea básica del test *RESET* es que si el modelo  $AR(p)$  en (1.114) es adecuado entonces  $\alpha_1$  y  $\alpha_2$  en el modelo (1.115) deberían ser cero. Esto último se puede llevar a cabo mediante el estadístico  $F$

$$F = \frac{(SSR_0 - SSR_1)/g}{SSR_1/(T - p - g)}, \quad \text{con } g = s + p + 1, \quad (1.116)$$

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 (1.115), de la siguiente forma

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

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

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

2. Se considera la regresión

$$\hat{\varepsilon}_t = \hat{\nu}_t\alpha + v_t,$$

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

**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 practica, 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  $\epsilon_t$  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  $g = p(p+1)/2$ .

**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  $AR(p)$

$$x_t = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t, \quad (1.117)$$

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

$$x_t = \begin{cases} \phi_{01} + \sum_{i=1}^p \phi_{1i} x_{t-i} + \epsilon_{1t}, & \text{si } x_{t-d} < r_1, \\ \phi_{02} + \sum_{i=1}^p \phi_{2i} x_{t-i} + \epsilon_{2t}, & \text{si } x_{t-d} \geq r_1, \end{cases} \quad (1.118)$$

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

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

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  $l_{\text{máx}} = \sup_{v < r_1 < u} l(r_1)$  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.

**1.9.13 Predicción**

En cuanto la predicción en los modelos no-lineales, por lo general, no es posible obtener formulas explícitas como en el caso de los modelos lineales, para un horizonte de predicción mayor que 1. En este caso, suele ser bastante útil aplicar las técnicas bootstrap, concretamente bootstrap paramétricos para determinar las predicciones.

**Bootstrap Paramétrico**

Este método considera calcular las realizaciones  $x_{T+i}$ , con  $i = 1, \dots, \ell$ , donde  $T$  es el origen de la predicción y  $\ell$  es el horizonte de predicción. Para obtener las realizaciones se procede como sigue:

Se generan nuevas impulsos o innovaciones a partir de la distribución de los impulsos del modelo especificado, y se calculan  $x_{T+i}$  usando el modelo, los datos y las predicciones anteriores.

Con esto se obtiene una realización  $x_{T+\ell}$ . Este procedimiento se repite  $M$  veces y la predicción es entonces obtenido como el promedio de las  $M$  replicas

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

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

### 1.10.1 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  $Var(W(t) - W(s)) = t - s$ , para todo  $0 \leq s \leq t$ .

**Definición 1.5.** 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 satisface

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 (1.5) 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  $\Delta t$ , con  $\Delta t \rightarrow 0$ ), asociados con incrementos de tamaño  $\Delta t$  en el tiempo. Así, el proceso de Wiener satisface:

- i.  $\Delta W_t = \epsilon \sqrt{\Delta t}$ , donde  $\epsilon$  es una variable aleatoria normal estándar; y
- ii.  $\Delta W_t$  es independiente de  $W_s$  para  $s \geq t$ .

De la condición (i.), se deduce que  $\Delta W_t$  tiene distribución normal con media cero y varianza  $\Delta t$ , 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 (1.5), 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  $\Delta t$  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  $\epsilon_i$  son independientes y de la definición (1.5), 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 proceso gaussianos son caracterizados por su esperanza y función de covarianza, entonces una definición alternativa a la dada en (1.5) es

**Definición 1.6.** 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.

### Puente Browniano

Considérese el proceso

$$x_t = W_t - tW_1, \quad 0 \leq t \leq 1. \quad (1.119)$$

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.

### 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

$$x_t = \mu t + \sigma W_t, \quad t \geq 0 \quad (1.120)$$

donde  $\sigma$  es una constante no negativa y  $\mu \in \mathbb{R}$ . Evidentemente, este es un proceso Gaussiano con funciones de media y covarianza

$$\mu_x(t) = \mathbb{E}(x_t) = \mu t, \quad \text{y} \quad \gamma_x(t, s) = \text{Cov}(x_t, x_s) = \sigma^2 [\min(t, s) - ts], \quad s, t \geq 0.$$

La función  $\mu_x(t) = \mu t$  (*función determinista del proceso*) esencialmente determina la forma característica de las trayectorias del proceso.

### 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 y 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

$$x_t = e^{(\mu t + \sigma W_t)}, \quad t \geq 0. \quad (1.121)$$

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,  $\mu_x$  y  $\sigma_x$  son conocidos como el *drift* y la *volatilidad* del proceso  $x_t$ , respectivamente.

### 1.10.2 Proceso de Itô

De la ecuación (1.121), 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

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t, \quad x_0, \quad t \geq 0, \quad (1.122)$$

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 (1.122) 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

$$x_t = x_0 + \int_0^t \mu(x_u, u)du + \int_0^t \sigma(x_u, u)dW_u, \quad (1.123)$$

La segunda integral de (1.123) 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 (1.122), donde  $\mu(x_t, t) = 0$  y  $\sigma(x_t, t) = 1$ .

Un caso particular del proceso (1.122) es el denominado *Proceso Difusión Homogéneo* definido como solución de la ecuación diferencial estocástica (SDE),

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \quad x_0, \quad 0 \leq t \leq T. \quad (1.124)$$

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

$$dx_t = \mu(x_t; \theta)dt + \sigma(x_t; \theta)dW_t, \quad x_0 \quad 0 \leq t \leq T \quad (1.125)$$

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  $\theta$  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

$$dx_t = \kappa(\mu - x_t)dt + \sigma dW_t, \quad \kappa > 0, \quad \sigma > 0. \quad (1.126)$$

para el precio de los derivados del tipo de interés. El modelo de Vasicek asume que la volatilidad  $\sigma$  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

$$dx_t = \kappa(\mu - x_t)dt + \sigma x_t^{1/2} dW_t. \quad (1.127)$$

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, (1.127) 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 (1.122) 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 (1.122), 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  $\Delta t$  y  $t_i = t_0 + i\Delta t$ , la aproximación de Euler del proceso (1.122) viene dada por

$$x_{t_{i+1}} = x_{t_i} + \mu(x_{t_i}, t_i)\Delta t + \sigma(x_{t_i}, t_i)\Delta W_{t_i} \quad (1.128)$$

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  $y^\delta$  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  $\delta$  y tal que

$$\mathbb{E}(|x_T - y^\delta(T)|) \leq C\delta^\gamma. \quad (1.129)$$

La expresión (1.129) mide cuan próximas están las trayectorias en el instante  $T$ . Por otra parte,  $y^\delta$  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  $\delta$  y tal que

$$|\mathbb{E}(g(x_T)) - \mathbb{E}(g(y^\delta(T)))| \leq C\delta^\beta. \quad (1.130)$$

Por lo tanto, la convergencia débil de orden  $\beta$  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 y Shreve (1991), Iacus (2008), Kutoyants (2004).

### 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 termino 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 1.2.** Supóngase que  $x_t$  es un proceso estocástico en tiempo continuo tal que satisface la ecuación diferencial estocástica

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t, \quad (1.131)$$

donde  $W_t$  es un proceso de Wiener. Adicionalmente,  $G(x_t, t)$  es una función diferenciable de  $x_t$  y  $t$ . Entonces

$$dG = \left\{ \frac{\partial G}{\partial x} \mu(x_t, t) + \frac{\partial G}{\partial t} + \frac{1}{2} \frac{\partial^2 G}{\partial x^2} \sigma^2(x_t, t) \right\} + \frac{\partial G}{\partial x} \sigma(x_t, t) dW_t. \quad (1.132)$$

Para ilustrar como opera 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,

$$dW_t^2 = \left( 2W_t \times 0 + 0 + \frac{1}{2} \times 2 \times 1 \right) dt + 2W_t dW_t = dt + 2W_t dW_t.$$

### 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

$$dP_t = \mu P_t dt + \sigma P_t dW_t, \quad (1.133)$$

donde  $\mu$  y  $\sigma$  son constantes. Haciendo uso de la notación básica de los procesos de Itô en (1.122), se tiene que  $\mu(P_t, t) = \mu P_t$  y  $\sigma(P_t, t) = \sigma P_t$ . Este proceso esta 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  $\sigma^2$  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  $\sigma^2\Delta$ .

### Estimación de $\mu$ y $\sigma$

Los parámetros  $\mu$  y  $\sigma$  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  $\Delta$  (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,  $\sigma$  se estima a partir de

$$\hat{\sigma} = \frac{s_r}{\sqrt{\Delta}}.$$

Teniendo en cuenta que  $\mu_r = \bar{r}$ , se puede estimar  $\mu$  por

$$\hat{\mu} = \frac{\bar{r}}{\Delta} + \frac{s_r^2}{2\Delta}.$$

**Ejemplo 1.3.** 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 (1.16) 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.

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 (1.121), 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

$$\hat{\sigma} = \frac{s_x}{\sqrt{\Delta}} = 0.4562518, \quad \text{and} \quad \hat{\mu} = \frac{\bar{x}}{\Delta} + \frac{\hat{\sigma}^2}{2} = 0.6369963$$

### 1.10.3 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

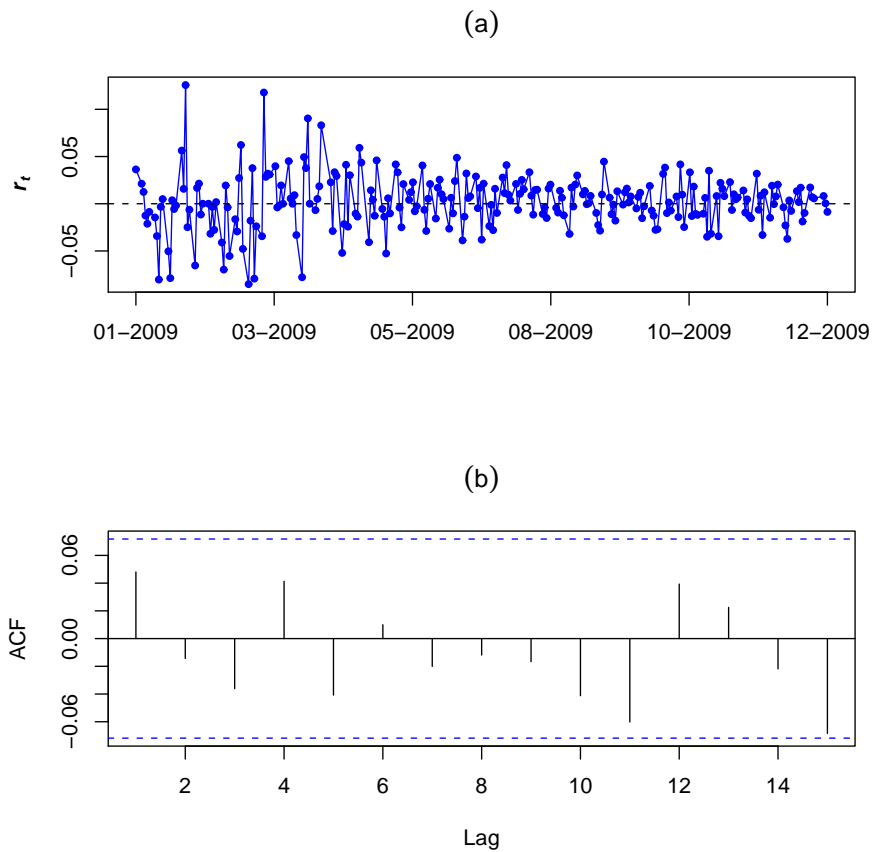
$$dP_t = \mu P_t dt + \sigma P_t dW_t.$$

Por lo tanto, las variaciones del logaritmo del precio en el intervalo  $(t, T]$  tienen distribución normal

$$\ln(P_T) - \ln(P_t) \sim N(\mu_{P_t}, \sigma_{P_t}^2) \quad (1.134)$$

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

$$\ln(P_T) \sim N(\ln(P_t) + \mu_{P_t}, \sigma_{P_t}^2). \quad (1.135)$$



**Figura 1.16:** Activo: Banco Santander, período 2 de Enero 2008 al 3 de Diciembre 2010; (a) Rentabilidades continuas (b) Función de autocorrelación muestral.

Del calculo 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  $\mu$ . 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,

$$r = \frac{1}{T-t} \ln \left( \frac{P_T}{P_t} \right).$$

De la ecuación (1.134) se tiene que

$$\ln \left( \frac{P_T}{P_t} \right) \sim N(\mu_{P_t}, \sigma_{P_t}^2).$$

Por consiguiente, la distribución de la tasa continua anual de la rentabilidad es

$$r \sim N \left( \mu - \frac{\sigma^2}{2}, \frac{\sigma^2}{T-t} \right).$$



### 1.10.4 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

$$\frac{dP_t}{P_t} = (\mu - \lambda\kappa)dt + \sigma dW_t + J_t dN_t, \quad (1.136)$$

donde  $N_t$  es un proceso de Poisson que caracteriza la ocurrencia de los saltos con intensidad  $\lambda$ ,  $J_t$  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 (1.136) admite la siguiente solución basada en la fórmula de Doléans-Dade,

$$\frac{P_t}{P_0} = \left( e^{(\mu - \frac{\sigma^2}{2} - \lambda\kappa)t + \sigma W_t} \right) \prod_{i=1}^{N_t} J_{T_i}, \quad (1.137)$$

con la convención  $\prod_{i=1}^0 = 1$ , donde  $T_i$ ,  $1 \leq i \leq N_t$ , 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 (1.136) 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 y 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 y 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

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t + d \left( \sum_{i=1}^{N_t} (J_i - 1) \right), \quad (1.138)$$

donde, como antes,  $W_t$  es un proceso de Wiener,  $N_t$  es un proceso de Poisson con tasa  $\lambda$  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

$$f_X(x) = \frac{1}{2\eta} e^{-|x-\theta|/\eta}, \quad 0 < \eta < 1. \quad (1.139)$$

Teniendo en cuenta la solución (1.137), el modelo (1.138) tiene similar solución

$$P_t = P_0 \left( e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} \right) \prod_{i=1}^{N_t} J_i. \quad (1.140)$$

A partir de esta solución se obtiene que la rentabilidad del activo subyacente para un incremento  $\Delta t$  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  $\epsilon$  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  $\Delta t$  pequeño, ignorando los saltos multiples, 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

$$\frac{P_{t+\Delta t} - P_t}{P_t} \approx \mu \Delta t + \sigma \epsilon \sqrt{\Delta t} + I \times X, \quad (1.141)$$

con  $I$  es una variable aleatoria Bernoulli con  $\mathbb{P}\{I = 1\} = \lambda \Delta t$  y  $\mathbb{P}\{I = 0\} = 1 - \lambda \Delta t$ , y  $X$  es una variable aleatoria doble-exponencial.

### 1.10.5 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 y 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

$$\begin{aligned} dP_t &= \mu P_t dt + \sigma_t P_t dW_{1t} \\ &\text{y} \\ d\sigma_t^2 &= \beta \sigma_t^2 dt + v \sigma_t^2 dW_{2t} \end{aligned} \quad (1.142)$$

donde  $\{W_{1t}\}$  y  $\{W_{2t}\}$  son dos movimientos Brownianos estándar cuyos incrementos tienen correlación  $\rho$ . 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 y 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 y 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

$$\begin{aligned} d\log(P_t) &= (\mu + \beta \sigma_t^2)dt + \sigma_t dW_t \\ &\text{y} \\ d\sigma_t^2 &= -\lambda \sigma_t^2 dt + dZ_{\lambda t} \end{aligned} \quad (1.143)$$

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.

### 1.10.6 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  $\mathcal{M}_\theta$  es una forma paramétrica conocida con  $\theta$  un parámetro desconocido, en este caso el objetivo principal es la estimación de  $\theta$ . A continuación se describen algunos métodos paramétricos.

#### 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

$$\begin{aligned} \ell(X_0, \dots, X_n; \theta) &= \sum_{i=1}^n \log(p(X_i|X_{i-1}; \theta)) + \log(\pi(X_0; \theta)) \\ &\approx \sum_{i=1}^n \log(p(X_i|X_{i-1}; \theta)) \end{aligned} \quad (1.144)$$

El estimador de máxima verosimilitud *MLE* se obtiene considerando el máximo de la función (1.144)

$$\hat{\theta} = \arg \max_{\theta} \ell(X_0, \dots, X_n; \theta).$$

Se pueden obtener formas explícitas de la función de densidad de transición para algunos modelos, tales como el de Vasicek (1.126) 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 (1.128). Un método alternativo en este sentido es el método presentado en Ait-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).

### 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  $\theta$  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  $\theta_0$  es el valor verdadero del parámetro  $\theta$ :

$$\mathbb{E}\{u(X_i, \theta)\} = 0 \quad \text{si y solo si } \theta = \theta_0. \quad (1.145)$$

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 (1.145). Se espera que  $\mathbb{E}\{g_n(\theta_0)\} = 0$ . Se considera además que se satisface la ley fuerte de los grandes números,

$$g_n(\theta) \xrightarrow{c.s.} \mathbb{E}\{u(X_i, \theta)\}$$

uniformemente en  $\theta$ . Así pues, el parámetro estimado  $\hat{\theta}$ , es la solución del problema de optimización

$$\hat{\theta} = \arg \min_{\theta} Q(\theta) = \arg \min_{\theta} g_n(\theta)^T W g_n(\theta), \quad (1.146)$$

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 y King \(2004\)](#), [Arapis y Gao \(2006\)](#) e [Iacus \(2008\)](#).

# 2

## Modelización de los Tipo de Interés

Todo estudio sobre el dinero, conduce de manera inevitable a la discusión del tipo de interés y la caracterización de la dinámica que los rige. Ciertamente, las inversiones tienen un objetivo bastante preciso, y es el de maximizar los beneficios, esto último guarda una relación íntima con el tipo de interés, puesto que los beneficios o ganancias de una inversión están sujetos al rendimiento de los activos asociados y no menos importante, a que el coste de dichos activos no supere en términos relativos al del tipo de interés. Así pues, en finanzas se hace imprescindible tener un conocimiento amplio de la dinámica que rige el comportamiento de los tipos de interés.

En cuanto a las tasas o tipos de interés, debido a su diversidad, es preciso establecer una distinción. Una primera clasificación natural determina dos clases: los tipos de interés interbancarios y los gubernamentales o públicos. Los *tipos gubernamentales o públicos* se deducen generalmente por los bonos emitidos por los gobiernos. Mientras que los *tipos de interés interbancarios* se establecen a partir del intercambio de depósitos y transacciones de títulos de valor (*swaps*) entre los bancos. Los *tipos de cupón cero* se pueden “extraer”, bien de los bonos en el sector público, bien del mercado o de los subproductos en el mercado interbancario, lo cual da como resultado dos curvas distintas del tipo de interés. Sin embargo, a la hora de analizarlas no es un problema, puesto que, una vez establecido el modelo matemático este es análogo en los dos casos. En esta memoria son de principal importancia los tipos de interés interbancario, no obstante, se tendrán en cuenta algunos conceptos útiles que involucren otras clases de tipos de interés.

Entre los tipos de interés interbancarios conocidos y de gran importancia se encuentran el *LIBOR*, (*London Interbank Offered Rate*), Tasa de Oferta Interbancaria de Londres, y el *EURIBOR*, (*European Interbank Offered Rate*), Tasa de Oferta Interbancaria del Euro. Ambas tasas son fijadas diariamente. En esta memoria se consideran principalmente los tipos del *EURIBOR*, sin embargo, queda claro que las propuestas que se presentarán pueden ser aplicadas a cualquier clase de tipo de interés.

En este capítulo se hará una exposición acerca de los conceptos básicos que giran en torno a los tipos de interés. El primer concepto de gran importancia en el contexto financiero es el de los *bonos*. Seguidamente se establecerá su relación con los tipos de interés para luego exponer los principales aspectos de la modelización de los tipos de interés, entre los cuales destacaremos los modelos de difusión como una propuesta alternativa a las series de tiempo, para la caracteriza-

ción de la dinámica de la estructura temporal. Es importante destacar que para efectos de esta memoria, solo se abordarán los conceptos a título informativo y se hará énfasis en aquellos que se consideren relevantes para el desarrollo de la misma, pues el objetivo final que se persigue en este trabajo es el de proponer un test de bondad de ajuste para los modelos propuestos en la literatura para caracterizar la dinámica de los tipos de interés. Finalmente se presentan una discusión breve acerca de la estimación de los modelos propuestos en la literatura en los que se hace énfasis en el método aplicado en este trabajo.

## 2.1 Bonos y Tipos de Interés

En esta sección se discutirán los conceptos básicos relacionados con el mercado de bonos y tipos de interés a fin de establecer diferencias entre la diversidad de los tipos de interés en la literatura y en el mercado financiero.

### 2.1.1 Bonos de Cupón Cero (Zero-Coupon Bonds)

Un *bono de cupón cero* con vencimiento en  $T$ , llamado también *T-bono*, es un contrato que garantiza al titular, el pago de una *unidad monetaria* en el instante  $T$ , la fecha de su vencimiento. Dicho de otra forma, este tipo de bonos no paga intereses periódicos, sino que los abona al vencimiento junto con el capital. El precio en el instante  $t$  de un bono con vencimiento en  $T$  se denota por  $p(t, T)$ . Se hace referencia a que el pago al vencimiento, conocido como *valor nominal*, es de una unidad monetaria simplemente por una convención computacional.

Los *bonos al portador* (*coupon bond*), denominado así porque cualquier persona que presente el cupón ante el emisor o agente pagador tendrá derecho a cobrar los intereses, dan al propietario un flujo de pagos durante un intervalo de tiempo,  $[0, T]$ . Estos instrumentos tienen la propiedad de que proporcionan al propietario con un flujo de caja determinista, y por esta razón también se les conoce como *instrumentos de renta fija*.

El precio de los bonos de cupón cero son la base de la teoría de los tipos de interés, ya que todos los tipos de interés se pueden definir en términos de dichos precios, tal y como se verá posteriormente. Es importante destacar, sin embargo, que los tipos de interés son los términos corrientemente usados en los mercados financieros (interbancario), mientras que los bonos de cupón cero son instrumentos más bien teóricos que, como tales, no son observables directamente en el mercado.

A continuación se presentan algunas suposiciones generales, necesarias para garantizar la existencia de un mercado de bonos regular y suficientemente abundante.

Se asumirá que:

- Existe un *mercado sin fricciones*<sup>\*</sup> para los T-bonos para cada  $T > 0$
- La relación  $p(t, t) = 1$  se cumple para todo  $t$ .
- Para cada  $t$ -fijo, el precio del bono  $p(t, T)$  es diferenciable con respecto al tiempo de vencimiento  $T$ .

La condición  $p(t, t) = 1$  se hace necesaria con la finalidad de evitar el arbitraje, en otras palabras, para evitar la posibilidad de realizar operaciones que consisten en comprar y vender instrumentos financieros en dos o más mercados para generar una diferencia de precios. Por lo tanto el precio  $p(t, T)$  es un elemento estocástico con dos variables,  $t$  y  $T$ , y que para cada resultado en el espacio muestral subyacente, la dependencia entre estas variables es muy diferente. Para un valor

<sup>\*</sup>Un entorno comercial teórico en el que todos los costos y las restricciones asociadas a las transacciones son inexistentes.

fijo  $t$ , el precio  $p(t, T)$  es una función del vencimiento  $T$ , proporcionando así, el precio para los distintos plazos de vencimiento en el instante  $t$ . El gráfico de esta función es denominado *curva del precio del bono en  $t$  o estructura temporal en  $t$* . Por lo general, esta curva es suave, es decir, para cada  $t$ , la función  $p(t, T)$  es diferenciable con respecto a  $T$ . Por otra parte, si se fija el plazo de vencimiento  $T$ , el precio  $p(t, T)$  es un proceso estocástico escalar. Este proceso da el precio del bono con vencimiento  $T$ , para distintos instantes de tiempo, y la trayectoria es por lo general bastante irregular (comparable a un proceso de Wiener).

En resumen, las suposiciones antes mencionadas conllevan a considerar un mercado de bonos que contiene un número infinito de activos (un tipo de bono por cada plazo de vencimiento). En este sentido, el objetivo de la teoría de los tipos de interés es, en términos generales, el estudio de las relaciones entre los tipos de bonos existentes. Antes de continuar, es preciso definir los conceptos relativos a los tipos de interés.

### 2.1.2 Tipos de Interés al Contado

El concepto de tasa o tipo de interés no es del todo ajeno para el común de las personas. Ciertamente, cuando se ingresa una cierta cantidad de dinero en una cuenta bancaria, es de esperarse que dicha cantidad se incremente en el tiempo, a una *tasa* o *razón* determinada. Así mismo, cuando un inversor coloca una cantidad de capital en un activo o conjunto de activos persigue como objetivo la obtención de beneficios, en otras palabras, que la rentabilidad prevista para el costo de una unidad de inversión sea superior al costo del dinero (o tipo de interés). En este orden de ideas también es cierto que, cuando se facilita una determinada cantidad de dinero en calidad de préstamo, este debe ser recompensado de alguna manera al ser devuelto, puesto que, dicha cantidad de dinero en el futuro próximo o lejano no es del todo equivalente a la cantidad inicialmente facilitada. Evidentemente, para expresar este conocimiento intuitivo se hace necesario introducir algunas definiciones para desarrollar una teoría consistente.

#### Cuenta Bancaria (*money-market account*)

Un primer concepto, el cual es bastante común en nuestra vida diaria, es el de *cuenta bancaria*. Una cuenta bancaria representa una inversión (a nivel local) sin riesgo, donde el beneficio se devenga continuamente a la tasa libre de riesgo vigente en el mercado en cada instante. Así pues, se define  $B(t)$  como el valor de una *cuenta bancaria* en el instante  $t \geq 0$ , para el cual se supone que  $B(0) = 1$  y que evoluciona de acuerdo a la siguiente ecuación diferencial :

$$dB(t) = r_t B(t), \quad B(0) = 1, \quad (2.1)$$

donde  $r_t$  es una función positiva del tiempo. Como una consecuencia,

$$B(t) = \exp \left( \int_0^t r_s ds \right). \quad (2.2)$$

Lo anterior se puede interpretar de la siguiente manera, invertir un importe unitario en el instante 0 produce un rendimiento cuyo valor es dado por (2.2) en el instante  $t$ , donde  $r_t$  representa el *tipo* o *tasa instantánea* en la que se acumula en la cuenta bancaria. Dicha tasa se le conoce usualmente con el nombre de *tipo instantáneo al contado*, o bien como el *tipo a corto plazo*. En efecto, si se considera una expansión de primer orden en  $\Delta t$  se obtiene

$$B(t + \Delta t) = B(t)(1 + r(t)\Delta t), \quad (2.3)$$

De (2.2) se puede decir, de forma equivalente que para cualquier intervalo  $[t, t + \Delta t)$ ,

$$\frac{B(t + \Delta t) - B(t)}{B(t)} = r(t)\Delta t.$$

### Factor de Descuento

Partiendo de la hipótesis inicial de que los tipos de interés se rigen por un proceso estocástico general, se introduce la definición de *factor de descuento estocástico* denotado por  $D(t, T)$  entre dos instantes de tiempo  $t$  y  $T$  es la cantidad en el tiempo  $t$  “equivalente” a una unidad monetaria a pagar en el instante  $T$ , y está dada por

$$D(t, T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r_s ds\right). \quad (2.4)$$

La naturaleza probabilística de  $r_t$  es de gran importancia puesto que esta afecta la naturaleza de la cuenta bancaria.

En aplicaciones  $r$  se supone que es una función determinista del tiempo, dando como resultado que la cuenta bancaria (2.1) y los el factor de descuento (2.4) en cualquier instante en el futuro son funciones deterministas de tiempo. Esto suele ser común cuando se aplican las formulas de Black y Scholes, por ejemplo en el mercado de capitales o divisas. Y es motivado por la suposición de que la variabilidad de los tipos de interés contribuye a que el precio del capital o las opciones del mercado de divisas. Sin embargo, cuando se trata de productos de los tipos de interés, la variabilidad principal que es de importancia son los tipos de interés. Por ello, es necesario modelizar la evolución de los tipos  $r$  en el tiempo a través de un proceso estocástico.

Si los tipos  $r$  son considerados deterministas, entonces  $D$  el factor de descuento también lo es, y entonces  $D(t, T) = p(t, T)$ . Sin embargo, como se ha mencionado anteriormente, si los tipos  $r$  son regidos por un proceso estocástico entonces el factor de descuento  $D(t, T)$  es una cantidad aleatoria en el instante  $t$  el cual depende de la evolución de los tipos de interés  $r$  en el futuro, entre  $t$  y  $T$ . Por el contrario, el precio  $p(t, T)$  de un bono de cupón cero con valor del contrato en  $t$  y con vencimiento en  $T$ , conocido (determinista) en el tiempo  $t$ . El precio de los bonos  $p(t, T)$  y el factor de descuento  $D(t, T)$  están estrechamente vinculados, en el sentido de que  $p(t, T)$  se puede considerar como el valor esperado de la variable aleatoria  $D(t, T)$  bajo una medida probabilidad particular.

### Términos relativos al tiempo

Se considerará ahora, el mercado de bonos descrito en la sección anterior. Se define entonces como el *plazo de vencimiento* “ $T-t$ ” a la cantidad de tiempo (en años) desde el instante presente  $t$  al del vencimiento  $T > t$ . La definición “ $T-t$ ” tiene sentido siempre y cuando un  $t$  y  $T$  son números reales asociados a dos instantes de tiempo. Sin embargo, si  $t$  y  $T$  indican dos fechas expresada en (*día, mes, año*), por ejemplo

$$D_1 = (d_1, m_1, a_1) \quad \text{y} \quad D_2 = (d_2, m_2, a_2),$$

entonces se debe definir la cantidad de tiempo entre la dos fechas en cuanto al número de días entre dichas fechas. Esta elección, sin embargo, no es única, y el mercado evalúa el tiempo entre  $t$  y  $T$  de diferentes maneras. En efecto, el número de días entre la  $D_1$  y  $D_2$  se determina de acuerdo a la convención de mercado de referencia, que establece cómo se han de contar tales días, si se incluyen las vacaciones en dicha cuenta, etc.

Otro término técnico referido al tiempo, es el de *fracción del año*, denotado por  $\tau(t, T)$ , mide el instante seleccionado entre  $t$  y  $T$ , que normalmente hace referencia a la fracción de año entre la fecha  $t$  y  $T$ . Cuando la distancia entre dichos instantes es menor a un día, entonces  $\tau(t, T)$  es interpretado como la diferencia  $T-t$  (en años). La selección particular que se considera para medir el tiempo entre dos fechas es lo que se conoce como *convención del recuento de días*. Para aclarar este concepto considere los siguientes ejemplos:



- Convención *Actual/365*: bajo esta convención el año equivale a 365 días y la fracción del año entre dos fechas es el número *Actual* de días entre dichas fechas dividido por 365. Si se denota por  $D_2 - D_1$  el número *Actual* de días entre dos fechas, por ejemplo,  $D_1 = (04, 01, 2000)$  incluido y  $D_2 = (04, 07, 2000)$  excluido, entonces se tiene que la fracción del año en este caso es

$$\frac{D_2 - D_1}{365} = \frac{182}{365} = 0.49863$$

- Convención *Actual/360*: en este caso se asume que el año comprende 360 días. Entonces la fracción del año en este caso es

$$\frac{D_2 - D_1}{360} = \frac{182}{360} = 0.50556$$

- Convención *30/360*: con esta convención, se asume que los meses tienen 30 días y los años 360. Entonces se tiene que la fracción del año entre  $D_1$  y  $D_2$  esta dada por

$$\frac{\max(30 - d_1, 0) + \min(d_2, 30) + 360 \times (y_2 - y_1) + 30 \times (m_2 - m_1 - 1)}{360}.$$

En el caso de  $D_1$  y  $D_2$  definidos en 1, la fracción de año es

$$\frac{(30 - 4) + 4 + 360 \times 0 + 30 \times 5}{360} = 0.5$$

Las convenciones tienen como objetivo dejar de lado los días festivos, así pues, cada mercado establecerá dependiendo de la zona financiera y de sus necesidades la convención que más convenga. Si  $D_2$  es un festivo, este puede ser reemplazado por el primer día hábil que le sigue, modificando así la evaluación de la fracción del año. Para más detalles acerca de este tema ver, [Miron y Swannell \(1991\)](#)

Teniendo claro estos términos técnicos, y como ya se ha comentado en la sección anterior que, el precio de los bonos de cupón cero son la base de la teoría de los tipos de interés, entonces el siguiente paso es establecer la relación entre bonos y tipos.

### Tipo de interés al contado de capitalización continua

La tasa constante a la que una inversión de  $p(t, T)$  unidades monetarias en el instante  $t$  acumula continuamente para producir un importe unitario de la moneda corriente en el vencimiento  $T$  es conocido como el *tipo de interés al contado de capitalización continua* vigente en el instante  $t$  hasta el vencimiento  $T$ , se denota por  $R(t, T)$  y se expresa mediante la fórmula:

$$R(t, T) := -\frac{\ln(p(t, T))}{\tau(t, T)}. \quad (2.5)$$

Por lo tanto, este tipo de interés es un tipo constante que es consistente con el precio del bono de cupón cero en

$$e^{R(t, T)\tau(t, T)}p(t, T) = 1, \quad (2.6)$$

del cual se puede expresar el precio del bono en términos del tipo de interés compuesto continuamente  $R$ :

$$p(t, T) = e^{-R(t, T)\tau(t, T)}, \quad (2.7)$$

La fracción del año involucrada es por lo general  $\tau(t, T) = T - t$ , la diferencia de los tiempos expresada en años.

### Tipo de interés al contado de capitalización simple

Un tipo de interés alternativo es el *tipo de interés al contado de capitalización simple*, el cual es un tipo constante en el que la inversión se hace para producir un importe de una unidad monetaria en el vencimiento, comenzando desde  $p(t, T)$  unidades monetarias en el instante  $t$ , cuando la acumulación ocurre proporcional al tiempo de inversión. Se denota por,  $L(t, T)$  y se expresa por

$$L(t, T) := \frac{1 - p(t, T)}{\tau(t, T)p(t, T)}. \quad (2.8)$$

El mercado de los tipos del *LIBOR* son tipos de interés de esta clase, lo que motiva la notación designada por la letra  $L$ . los tipos del *LIBOR* están generalmente vinculados a los precios de los bonos de cupón cero mediante la convención de recuento diario *Actual/360*. La definición (2.8) nos conduce a su contraparte (2.6)

$$p(t, T)(1 + L(t, T)\tau(t, T)) = 1, \quad (2.9)$$

por lo tanto el precio del bono se puede expresar en términos de  $L$  como:

$$p(t, T) = \frac{1}{1 + L(t, T)\tau(t, T)}.$$

Es importante destacar que todas las definiciones anteriores de los tipos de interés al contado son equivalentes en intervalos de tiempo infinitesimal. En efecto, se puede demostrar que el tipo a corto plazo se puede obtener como límite de los distintos tipos definidos anteriormente, es decir, para cada  $t$

$$\begin{aligned} r(t) &= \lim_{T \rightarrow t^+} R(t, T) \\ &= \lim_{T \rightarrow t^+} L(t, T) \end{aligned}$$

### 2.1.3 Tipos de Interés a Plazo

El objetivo de esta sección es introducir el concepto de tipo de interés a plazo o futuro. Dichos tipos de interés asumen que en el instante  $t$ , se fijan dos puntos en el tiempo,  $T$  y  $S$ , con  $t < T < S$ . Se establece así, un contrato en el instante  $t$  que permite hacer una inversión de una unidad monetaria en el instante  $T$ , con tasa de rentabilidad determinista, considerado en el instante  $t$  del contrato, en el intervalo  $[T, S]$ . Esto puede llevarse a cabo de la siguiente manera:

1. En el instante  $t$  se efectúa la venta de un bono con plazo de vencimiento  $T$ , un  $T$ -bono, con lo cual obtenemos  $p(t, T)$  unidades monetarias.
2. A partir de estos ingresos se realiza la compra de exactamente  $p(t, T)/p(t, S)$ ,  $S$ -bonos. Así pues, la inversión neta en el instante  $t$  es igual a cero.
3. En el tiempo  $T$  el  $T$ -bono vence, de manera que se está en la obligación de pagar una unidad monetaria.
4. En el tiempo  $S$  los  $S$ -bonos vencen obteniéndose de estos una unidad monetaria por cada una de ellos, en este caso se obtienen  $p(t, T)/p(t, S)$  unidades monetarias.
5. El resultado neto de todo esto es que, basado en el contrato en el instante  $t$ , una inversión de una unidad monetaria en el instante  $T$  tiene un rendimiento de  $p(t, T)/p(t, S)$  unidades monetarias en el instante  $S$ .

6. Por lo tanto, en el instante  $t$ , se ha llevado a cabo un contrato que garantiza una tasa o tipo de interés libre de riesgo en el *intervalo futuro*  $[T, S]$ . Dicho tipo de interés es denominado *tipo de interés futuro* o *tipo de interés a plazo*

Bajo estas premisas, se presentan las siguientes definiciones.

### Tipo de Interés a plazo de capitalización simple

El tipo de interés a plazo de capitalización simple vigente en el instante  $t$  con expiración para  $T > t$  y con vencimiento en  $S > T$  se denota por  $F(t; T, S)$  y se define por

$$F(t; T, S) := \frac{p(t, T) - p(t, S)}{\tau(T, S)p(t, S)}. \quad (2.10)$$

El tipo a plazo (2.10) se puede ver como un estimado del tipo a plazo al contado  $L(T, S)$ , el cual es aleatorio en el instante  $t$ , basado en las condiciones del mercado en el instante  $t$ . En particular, como veremos posteriormente,  $F(t; T, S)$  es el valor esperado de  $L(T, S)$  en el instante  $t$  bajo una medida de probabilidad apropiada.

Cuando el vencimiento del tipo a plazo tiende a su momento de expiración, se obtiene el denominado *tipo a plazo instantáneo*. En efecto, si se considera el límite

$$\begin{aligned} \lim_{S \rightarrow T^+} F(t; T, S) &= - \lim_{S \rightarrow T^+} \frac{1}{p(t, S)} \frac{p(t, S) - p(t, T)}{(S - T)} \\ &= - \frac{1}{p(t, T)} \frac{\partial p(t, T)}{\partial T} \end{aligned} \quad (2.11)$$

$$= - \frac{\partial \ln p(t, T)}{\partial T}, \quad (2.12)$$

donde  $\tau(T, S) = S - T$  cuando  $S$  está lo bastante cerca de  $T$ . Esto conlleva a la siguiente definición.

### Tipo de interés Instantáneo a plazo

Este tipo de interés vigente en el instante  $t$  con vencimiento  $T > t$  es denotado por  $f(t, T)$  y se define por

$$f(t, T) := \lim_{S \rightarrow T^+} F(t; T, S) = - \frac{\partial \ln p(t, T)}{\partial T} \quad (2.13)$$

de donde se obtiene que

$$p(t, T) = \exp \left( - \int_t^T f(t, u) du \right)$$

Intuitivamente, los tipos de interés instantáneo a plazo es un tipo de interés en el instante  $t$  cuyo vencimiento está muy cerca de su expiración  $T$ , esto es

$$f(t, T) \approx F(t; T, T + \Delta t) \quad \text{para } \Delta t \text{ pequeño.}$$

A partir de los tipos a plazo se define el *tipo instantáneo a corto plazo* en el instante  $t$  como

$$r(t) = f(t, t). \quad (2.14)$$

Como se puede ver, los tipos al contado son tipos a plazo en donde el instante de la contratación coincide con el punto inicial del intervalo en el cual el tipo de interés es efectivo, es decir,  $t = T$ .

Los tipos antes descritos, son magnitudes fundamentales en la teoría de los tipos de interés. Una de las formas más generales de expresar la "imparcialidad" de un modelo de tipo de interés

es relacionando las cantidades determinadas en la expresión para la evolución de  $f$ . Por “imparcialidad” se entiende a la ausencia oportunidades de arbitraje, esta ausencia de arbitraje es considerada en Heath et al. (1992). Los autores se basan en los tipos instantáneos a plazo,  $f$ , como magnitudes fundamentales para la modelización.

Ahora bien, la teoría de la modelización de los tipos de interés se basa originalmente en la suposición de una dinámica unidimensional del proceso de tipos de interés instantáneo al contado  $r$ . La modelización directa de dicha dinámica es conveniente puesto que todas las cantidades fundamentales (tipos y bonos) son totalmente definidas, por argumentos de no-arbitraje, como la esperanza de un funcional del proceso  $r$ . En efecto, la existencia de una medida de riesgo neutral implica que el precio libre de arbitraje en el instante  $t$  de un reclamo contingente con pago  $H_t$  en el instante  $T$  es dado por

$$H_t = \mathbb{E}_t(D(t, T)H_T) = \mathbb{E}_t\left(e^{-\int_t^T r(s)ds}H_T\right), \quad (2.15)$$

con  $\mathbb{E}_t$  denotando el instante  $t$ -condicional valor esperado bajo dicha medida. En particular, el precio del bono de cupón cero en el instante  $t$  para el vencimiento  $T$  esta caracterizado por un importe unitario de la moneda en uso en el instante  $T$ , así que  $H_T = 1$  y se obtiene entonces que

$$p(t, T) = \mathbb{E}_t\left(e^{-\int_t^T r(s)ds}\right). \quad (2.16)$$

Teniendo en cuenta esta última expresión, se puede ver que siempre es posible caracterizar la distribución  $e^{-\int_t^T r(s)ds}$  en términos de la dinámica seleccionada para  $r$ , condicionada a la información disponible en el instante  $t$ , entonces se puede determinar el precio de los bonos  $p$ .

## 2.2 Modelos en tiempo Continuo

Las distintas turbulencias de la que son objeto los mercados financieros en las últimas décadas han dado origen a un sin fin de propuestas para la modelización de la estructura temporal de los tipos de interés, que en lugar de considerar el tradicional enfoque determinista ha optado por tener en cuenta que dicha modelización se lleve a cabo en un entorno estocástico. Por tanto, el surgimiento de teorías que consideran modelos dinámicos que estudien las relaciones entre los rendimientos de bonos de cupón cero, libres de riesgo, y a distintos plazos o vencimientos, en un contexto de incertidumbre y en una economía en tiempo continuo, son de vital importancia en la caracterización de los tipos de interés.

En la sección anterior se comentó que, el precio de los bonos de cupón cero son la base de la teoría de los tipos de interés. de manera que, para caracterizar la dinámica de la estructura temporal de los tipos de interés la mayoría de los modelos considera como variable dependiente el precio los bonos de cupón cero,  $p(t, T)$ . Una vez establecido dicho precio se puede deducir el tipo al contado  $R(t, T)$  y en consecuencia los tipos a plazo  $f(t, T)$ .

En los modelos dinámicos de la estructura temporal, se tiene que el precio  $p(t, T)$  en cada instante, depende tanto del vencimiento de la obligación como de un vector de variables de estado

$$x(t) = (x_1(t), \dots, x_k(t)) \quad (2.17)$$

donde  $x_i$ ,  $1 \leq i \leq k$  representa las fuentes de incertidumbre, y por tanto, de riesgo del mercado. En base a esto, se reescriben los términos definidos en la sección anterior teniendo en cuenta la dependencia de los factores de riesgo del modelo, obteniéndose

$$\begin{aligned} p(t, T) &\rightarrow p(t, x(t), T), \\ R(t, T) &\rightarrow R(t, x(t), T), \\ f(t, T) &\rightarrow f(t, x(t), T), \end{aligned}$$

El precio del bono, así como también de las demás funciones, varía de acuerdo al vencimiento, y a los cambios de las variables  $x(t)$  que reflejan el estado del mercado en cada instante, por ello estas variables son denominadas *variables de estado*. Dichas variables de estado denotan los factores de riesgo de los modelos dinámicos de la estructura de los tipos. Desde un punto de vista determinista, la dinámica de una cierta variable que varía respecto del tiempo  $x(t)$ , se puede describir mediante la ecuación diferencial ordinaria de primer orden

$$\frac{dx(t)}{dt} = f(t, x(t)) \quad (2.18)$$

que relaciona a una función  $f$  de la variable de estado con su variación en  $t \in [0, T]$  y con condición inicial  $x(0) = x_0$ . Esta ecuación caracteriza la evolución de la variable  $x(t)$  en  $t$ , a través de  $f(t, x(t))$  que denota la tendencia determinista de la variable o la variación esperada, en cada instante  $t$ . Bajo la hipótesis de incertidumbre, una hipótesis más realista en el contexto financiero, la ecuación anterior debe ser modificada considerando que la dinámica de la variable en cuestión no está del todo explicada simplemente por una tendencia determinista, sino que en cada instante se ve afectada por impulsos (shocks) aleatorios implícitos a la incertidumbre existente en el mercado. De esta manera, en la ecuación (2.18) una manera de introducir dicha incertidumbre, que afecta la dinámica de la variable de estado es a través del proceso estocástico de Wiener, discutido en el capítulo 1, y mediante una función  $\sigma(t, x(t)) > 0$ , dando como resultado el proceso de Itô (discutido previamente en el capítulo 1)

$$dx(t) = f(t, x(t))dt + \sigma(t, x(t))dW_t \quad (2.19)$$

donde  $W_t$  es el un proceso de Wiener. Por tanto, la dinámica de la variable  $x(t)$  queda explicada por dos componentes. Un primer componente de tendencia determinista y un segundo componente estocástico.

Los distintos modelos dinámicos que caracterizan la estructura temporal del tipo de interés se diferencian, entre otras aspectos, en función de las variables explicativas que consideran. En base a esto, los modelos que incorporan información de toda la curva de tipos de interés para describir su dinámica, se denominan *modelos consistentes*. Por otro lado, los modelos que sólo consideran una o varias variables de estado como variables explicativas en la evolución en el tiempo de la curva, son denominados *modelos factoriales*.

Bajo la suposición de ausencia de arbitraje, es posible derivar una expresión para valorar los bonos de cupón cero para cualquier vencimiento, de manera que, se puede obtener la estructura temporal del tipo de interés compatible con el comportamiento de este en un ámbito estocástico.

Como ya se mencionó, múltiples variables de estado o factores pueden estar asociados a la estructura de temporal, sin embargo en esta memoria se consideran únicamente los *modelos unifactoriales*, es decir, modelos que involucran solo una variable de estado: el tipo de interés instantáneo  $r(t) \equiv r_t$ . De manera general, se supondrá que el tipo de interés  $r$  sigue un proceso estocástico Markoviano definido por la siguiente ecuación diferencial (Proceso de Itô)

$$dr_t = \mu(r_t, t)dt + \sigma(r_t, t)dW_t \quad (2.20)$$

donde  $W_t$  es un proceso de Wiener,  $\mu(r_t, t)$  es la tendencia del proceso (drift) o valor esperado de la variación del tipo de interés instantáneo a corto plazo condicionado al tipo de interés por unidad de tiempo, y por supuesto  $\sigma(r_t, t)$  la volatilidad del proceso.

### 2.2.1 Modelos de Difusión Homogéneos

Dada la relevancia que han adquirido los modelos en tiempo en las aplicaciones en el ámbito de las finanzas, concretamente los procesos de difusión en tiempo continuo, los cuales se han

usado con frecuencia para caracterizar la dinámica de variables económicas importantes, tales como tipos de cambio, valoración de activos y de manera especial los tipos de interés. Los modelos de difusión en tiempo continuo homogéneos son uno de los primeros modelos propuestos para caracterizar los tipos de interés a corto plazo, entendiendo que la dinámica asumida para los tipos a corto plazo depende solo de ciertos coeficientes constantes. En los últimos años la investigación en este contexto ha tenido un desarrollo importante, teniendo como punto de partida el trabajo de [Merton \(1973\)](#), quien propuso modelizar el tipo de interés como un proceso estocástico que luego uso en la valoración de opciones. Posteriormente, se profundiza en este tipo de modelización y se comienza a trabajar aplicando argumentos de arbitraje similares a los trabajos de [Black y Scholes \(1973\)](#) para modelizar la estructura temporal de los tipos de interés tal y como se puede observar en los trabajos de [Vasicek \(1977\)](#) y [Brennan y Schwartz \(1979\)](#). Así pues, con el fin de perfeccionar las ideas expuestas en estos primeros modelos, se proponen una serie de modelos más complejos, entre los que destacan, [Cox et al. \(1980\)](#), [Cox et al. \(1985\)](#), [Chan et al. \(1992\)](#), entre otros.

En la teoría de los modelos para tipos de interés se tienen los *modelos endógenos*, es decir, modelos que se caracterizan por utilizar como variable dependiente, para explicar la evolución de la estructura temporal de los tipos de interés, el precio de un bono de cupón cero libre de riesgo de insolvencia. Dicho precio se considera dependiente del vencimiento además de una o más variables de estado, denominadas factores. Estos modelos pueden ser agrupados en dos categorías: los llamados *modelos de no-arbitraje o equilibrio parcial* y los *modelos de equilibrio general*. En los modelos de no-arbitraje se establece la hipótesis de la evolución estocástica del tipo de interés principalmente si, como es el caso de estudio en esta memoria, es la única variable de estado a considerar, en caso contrario, dicha hipótesis se establece para todas las variables de estado involucradas, además de la forma funcional de los precios del riesgo de mercado asociado a dichas variables. En base a esto y suponiendo que en el mercado no existen oportunidades de arbitraje, se obtiene la estructura temporal del tipo de interés y los precios de los distintos activos derivados de los tipos de interés. Como ejemplo de estos modelos se tienen el modelo de [Vasicek \(1977\)](#), [Dothan \(1978\)](#), [Brennan y Schwartz \(1979\)](#), entre otros. Por otro lado, en los modelos de equilibrio general, el enfoque se centra en una descripción de la economía real tomando en cuenta las preferencias de los inversores. De esta manera, y considerando como punto de partida un equilibrio intertemporal del mercado de activos además de utilizar la metodología de la optimización dinámica estocástica en la determinación de la prima de riesgo y el precio de los activos se obtiene una modelización de la estructura temporal de los tipos de interés. Bajo este enfoque, se encuentran los trabajos de [Cox et al. \(1985\)](#), [Longstaff y Schwartz \(1992\)](#), entre otros.

Esta sección pretende presentar las características más importantes de los modelos de difusión en tiempo continuo homogéneos para los tipos de interés a corto plazo, para entender su mecanismo y naturaleza, puesto que, serán base fundamental en el desarrollo de la propuesta que será presentada en el próximo capítulo.

Como ya se ha mencionado en este capítulo, en finanzas el modelo usado con frecuencia para caracterizar la dinámica de los tipos de interés, es el modelo de difusión en tiempo continuo o proceso de Itô dado por la ecuación diferencial estocástica,

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t, \quad (2.21)$$

donde  $W_t$  es un movimiento Browniano, y  $\mu(r_t)$  y  $\sigma(r_t)$  dependen únicamente del nivel o tipo de interés. Como es sabido, en la ecuación (2.21),  $\mu(r_t)$  representa la función drift (deriva) o función de rendimiento instantáneo, y  $\sigma(r_t)$  la *difusión* o función de volatilidad. Dichas funciones satisfacen las siguientes expresiones:

$$\mu(r_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbf{E}(r_{t+\Delta} - r_t | r_t), \quad \text{y} \quad \sigma^2(r_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbf{E}((r_{t+\Delta} - r_t)^2 | r_t).$$

Como se puede observar, en este caso, a diferencia de las anteriores presentaciones de este modelo, se ha suprimido la variable relativa al tiempo  $t$ , debido a que se considerará el proceso en su forma homogénea. Pues bien, para estudiar el modelo (2.21) y observar su comportamiento empírico, existen básicamente dos enfoques. En un primer enfoque, se supone que las funciones tienen un comportamiento conocido y determinado por un parámetro,  $\theta$ , en un espacio de parámetros,  $\Theta \subset \mathbb{R}^d$  con  $d$  un entero positivo, el cual es posible estimar a partir de un estimador apropiado. El modelo (2.21) entonces queda expresado en su versión paramétrica como,

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t. \quad (2.22)$$

En un segundo enfoque, no se presupone ningún comportamiento específico para las funciones  $\mu(r_t)$  y  $\sigma(r_t)$ , y se utilizan técnicas no paramétricas para su estimación, ver por ejemplo Stanton (1997), Arapis y Gao (2006), Gao y Casas (2008), entre otros. En lo que sigue se establecerán algunas particularidades acerca del modelo continuo en su forma paramétrica.

### El modelo paramétrico continuo subyacente

Considérese la ecuación diferencial estocástica unidimensional homogénea respecto del tiempo (2.22) para  $\theta \in \Theta \subset \mathbb{R}^d$  un parámetro multidimensional y  $\theta_0$  el parámetro verdadero a ser estimado. En este caso las funciones

$$\mu : \mathbb{R} \times \Theta \rightarrow \mathbb{R} \quad \text{y} \quad \sigma : \mathbb{R} \times \Theta \rightarrow (0, \infty)$$

se consideran funciones conocidas con  $\theta$  desconocido, tal que la solución de (2.22) existe. En lo que respecta a la solución de la ecuación diferencial estocástica, ciertas suposiciones sobre el modelo son asumidas para garantizar la existencia de la misma, ver Apéndice B. Cuando la distribución estacionaria o invariante denotada por  $\pi_\theta$  existe, esta tiene la siguiente forma

$$\pi_\theta(x) = \frac{1}{M(\theta)\sigma^2(x, \theta)s(x, \theta)}, \quad (2.23)$$

donde

$$s(x, \theta) = \exp \left\{ -2 \int_{x_0}^x \frac{\mu(y, \theta)}{\sigma^2(y, \theta)} \right\}$$

para algún  $x_0 \in I$ , con  $I$  el espacio de estado del proceso, y  $M(\theta)$  una constante de normalización. La función  $s$  denominada la *medida de escala* y  $m(x) = \pi_\theta(x) \cdot M(\theta)$  su *medida de rapidez* (ver detalles en el Apéndice C). La distribución de  $r$  con  $r_0 \sim \pi_\theta$  se denota por  $P_\theta$ , y bajo esta medida,  $r_t \sim \pi_\theta$  para todo  $t$ . Se denotará además por  $p_\theta(t, \cdot | r)$  la *densidad condicional* (o *densidad de transición*) de  $r_t$  dado  $r_0 = r$ . Cuando  $r$  es homogénea respecto del tiempo  $p_\theta(t, \cdot | r)$  esta función representa la densidad de  $r_{s+t}$  condicionada a  $r_s = r$  para todo  $t \geq 0$ . En algunos casos, se usará la notación  $p_{r, \theta}(t, \cdot | r)$ .

A continuación presentamos los rasgos más resaltantes de alguna de las especificaciones paramétricas de los modelos de difusión que serán objeto de estudio en el próximo capítulo de esta memoria, destacando una característica principal en el enfoque desarrollado en esta memoria, y es que para efectos de la propuesta a presentar, el escenario considerado es el de riesgo neutral, es decir, los procesos considerados son definidos bajo una medida de riesgo-neutral. Adicionalmente, siempre que sea posible, y teniendo en cuenta su utilidad dentro de la propuesta que se presentará en esta memoria, cada modelo será descrito en términos de los momentos de primer y segundo orden del proceso asociado, así como también en términos de su densidad estacionaria y de su distribución condicional.



### 2.2.2 Modelo de Merton

Este modelo, desarrollado por [Merton \(1973\)](#), es uno de las primeras propuestas que consideran el uso de un proceso estocástico para modelizar el tipo de interés. [Merton \(1973\)](#) asume que las funciones de volatilidad (difusión) y tendencia o drift son constantes, es decir, que el tipo de interés a corto plazo se comporta como un movimiento browniano con tendencia,

$$dr_t = \alpha dt + \sigma dW_t, \quad r_0 = r_o \quad (2.24)$$

donde  $\sigma > 0$  y  $r_0 = r_o$  la condicional inicial de la ecuación. Esta es una especificación bastante simple que proporciona soluciones analíticas para la estructura temporal del tipo de interés, así como también para algunos activos como las opciones sobre bonos de cupón cero. Sin embargo resulta, muy poco atractiva desde el punto de vista empírico dado que no recoge ninguna dinámica ni para la media ni para la volatilidad de los tipos. Otro inconveniente, es que bajo este modelo el tipo de interés puede tomar valores negativos, lo cual es poco deseable en los tipos de interés.

### 2.2.3 Modelo de Black-Scholes o Geométrico Browniano *GBM*

Este modelo, desarrollado por [Black y Scholes \(1973\)](#), modelo para la valoración de derivados financieros fue publicado en el *Journal of Political Economy* en 1973, conocido en el ámbito financiero como el modelo de *Black-Scholes-Merton*, y aceptado desde entonces, como uno de los modelos matemáticos más influyentes en grandes decisiones financieras a nivel mundial. Para este modelo, el movimiento Browniano geométrico es el proceso básico asociado al movimiento de los precios por ello se suele denotar el modelo por las siglas *GBM*. Este modelo supone que el tipo de interés se caracteriza mediante la solución de la ecuación diferencial estocástica

$$dr_t = \alpha r_t dt + \sigma r_t dW_t, \quad r_0 = r_o \quad (2.25)$$

donde  $\sigma > 0$  y  $r_0 = r_o$  la condicional inicial de la ecuación. En este caso, el parámetro  $\alpha$  es interpretado como el tipo de interés constante y  $\sigma$  se considera como la volatilidad del riesgo de las actividades. En el capítulo 1 de esta memoria, se estudió el *movimiento Browniano Geométrico*, en donde se presentó el proceso solución del mismo, y que en este caso es dado por

$$r_t = r_o e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}. \quad (2.26)$$

Como se puede ver, en este caso se obtiene una solución explícita dada por (2.26) de la ecuación (2.25) con lo cual se puede caracterizar la dinámica del tipo de interés a partir de sus propiedades estadísticas, tales como su media y su varianza condicional

$$\mathbb{E}(r_t | r_0 = r_o) = e^{\mu_{\log r} + \frac{1}{2}\sigma_{\log r}^2} = r_o e^{\alpha t} \quad (2.27)$$

$$\mathbb{V}ar(r_t | r_0 = r_o) = e^{2\mu_{\log r} + \sigma_{\log r}^2} (e^{\sigma_{\log r}^2} - 1) = r_o^2 e^{2\alpha t} (e^{t\sigma^2} - 1) \quad (2.28)$$

teniendo en cuenta que la solución tiene densidad condicional log-normal, denotada por  $p(t, s|r)$ , con media y varianza dadas por

$$\mu_{\log r} = \log(r_o) + \left(\alpha - \frac{1}{2}\sigma^2\right)t, \quad \sigma_{\log r}^2 = \sigma^2 t, \quad (2.29)$$

por lo tanto

$$p(t, s|r_o) = \frac{1}{s\sigma\sqrt{2\pi t}} \exp\left\{-\frac{\log(s) - (\log(r_o) + (\alpha - \frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t}\right\}. \quad (2.30)$$



De esta manera, el modelo queda determinado de forma total a partir de sus propiedades. Esta especificación tiene la ventaja, frente al modelo de Merton, que el tipo de interés nunca toma valores negativos. Para valores de  $\alpha > 0$  el proceso revierte a la media. Además, la especificación de la volatilidad se caracteriza por ser cambiante en el tiempo en función del nivel del tipo de interés: cuanto mas (menos) altos son los tipos de interés mayor (menor) es su volatilidad.

### 2.2.4 Modelo de Vasicek (VAS)

El modelo propuesto por Vasicek (1977) y que se denota por VAS, está basado en el proceso de Ornstein-Uhlenbeck (ver Uhlenbeck y Ornstein (1930)) para modelizar el tipo de interés instantáneo. Considera dicho proceso con coeficientes constantes como la solución de la ecuación

$$dr_t = (\alpha - \beta r_t)dt + \sigma dW_t, \quad r_0 = r_o \quad (2.31)$$

con  $r_0 = r_o$ -condición inicial,  $\alpha, \beta, \sigma$  constantes positivas. La dinámica de este modelo tiene algunas particularidades, que hacen al modelo bastante interesante. Como se puede observar, la ecuación es lineal y se puede demostrar que su solución es explícita, con distribución Gaussiana. Además, bajo la condición de que  $\beta > 0$  se tiene que el proceso revierte a la media y es de volatilidad constante para  $\sigma > 0$ . La propiedad de reversión a la media supone para este modelo que el proceso asociado tiende a oscilar alrededor de un estado de equilibrio. La aportación del proceso de Ornstein-Uhlenbeck, a diferencia del movimiento Browniano geométrico, es que garantiza que este proceso tenga varianza finita para todo  $t \geq 0$ . En finanzas, como es el caso que interesa en esta memoria, suele presentarse este modelo mediante la parametrización

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t, \quad (2.32)$$

donde  $\sigma$  se interpreta, por supuesto, como la volatilidad,  $\mu$  es el valor de equilibrio a largo plazo del proceso, y  $\kappa$  es la velocidad de la reversión. En otras palabras, el proceso tiende a ser empujado hacia un nivel medio de largo plazo  $\mu = \alpha/\beta$  a lo largo de su trayectoria con esfuerzo  $\kappa = \beta$ . La fuerza es proporcional al alejamiento del tipo de interés de su media. La reversión de la media es una metodología matemática de uso común para la inversión de valores. En términos generales la idea es que, valores con precios altos y bajos tienen un carácter temporal, y que estos tenderán a un precio medio en el tiempo. Este proceso es un ejemplo de un proceso Gaussiano que tiene varianza acotada y admite una distribución de probabilidad estacionaria (más generalmente un proceso estacionario, Markoviano, gaussiano y continuo en probabilidad) en contraste con el proceso de Wiener. La varianza (a largo plazo) está dada por

$$\text{Var}(r_t) = \frac{\sigma^2}{2\kappa}.$$

Esto último se deduce del cálculo de Itô, al integrar la ecuación (2.31) se obtiene,

$$r_t = \frac{\alpha}{\beta} + \left(r_o - \frac{\alpha}{\beta}\right) e^{-\beta t} + \sigma \int_0^t e^{\beta(t-u)} dW_u$$

o bien, si se considera la parametrización en el contexto financiero,

$$r_t = \mu + (r_o - \mu) e^{-\kappa t} + \sigma \int_0^t e^{\kappa(t-u)} dW_u.$$

Para cualquier  $t \geq 0$ , el proceso de Ornstein-Uhlenbeck tiene función de densidad condicional o de transición  $p_\theta(t, r_t|r_0)$  Gaussiana con media y varianza dada por

$$\mathbb{E}(r_t|r_0 = r_o) = \frac{\alpha}{\beta} + \left(r_o - \frac{\alpha}{\beta}\right) e^{-\beta t} \quad (2.33)$$

y

$$\text{Var}(r_t | r_0 = r_o) = \frac{\sigma^2 (1 - e^{-2\beta t})}{2\beta}. \quad (2.34)$$

La función de covarianza condicional es dada por

$$\text{Cov}(r_s, r_t | r_0 = r_o) = \frac{\sigma^2}{2\beta} e^{-\beta(s+t)} \left( e^{2\beta(s \wedge t)} - 1 \right).$$

De manera general, para cada  $s \leq t$

$$r_t = \mu + (r_s - \mu) e^{-\kappa(t-s)} + \sigma \int_s^t e^{\kappa(t-u)} dW_u.$$

de manera que  $r_t$  condicionado a  $\mathcal{F}_s$  (la información disponible hasta el instante  $s \leq t$ ) tiene distribución normal con media y varianzas

$$\mathbb{E}(r_t | \mathcal{F}_s) = \frac{\alpha}{\beta} + \left( r_s - \frac{\alpha}{\beta} \right) e^{-\beta(t-s)} \quad (2.35)$$

$$\text{Var}(r_t | \mathcal{F}_s) = \frac{\sigma^2}{2\beta} \left( 1 - e^{-2\beta(t-s)} \right),$$

o bien en su parametrización en el contexto financiero

$$\mathbb{E}(r_t | \mathcal{F}_s) = \mu + (r_s - \mu) e^{-\kappa(t-s)} \quad (2.36)$$

$$\text{Var}(r_t | \mathcal{F}_s) = \frac{\sigma^2}{2\kappa} \left( 1 - e^{-2\kappa(t-s)} \right).$$

Esto implica que, para cada instante  $t$ , la tasa o tipo  $r_t$  puede tomar valores negativos con probabilidad mayor que cero. Esta posibilidad de obtener tipos negativos un inconveniente en la modelización de los tipos de interés a partir del modelo de Vasicek. Sin embargo, en cuanto al análisis de las propiedades del modelo obtenidas a partir de su densidad Gaussiana lo hacen atractivo puesto que en otros casos difícilmente se obtiene esta ventaja. De las ecuaciones (2.35) o (2.36), la tasa o tipo de interés a corto plazo  $r$  revierte a la media, puesto que su valor esperado tiende al valor  $\mu = \alpha/\beta$  cuando  $t$  tiende a infinito (en el largo plazo). Nótese además que el drift del proceso  $r$  es positivo siempre que el tipo a corto plazo está por debajo de  $\mu$  y es negativo en caso contrario, de manera que el tipo de interés tiende a ser presionado, en cada instante, a permanecer cercano en promedio a  $\mu$ . Para finalizar, para  $\beta > 0$ , el proceso es también ergódico, y su ley invariante es la densidad Gaussiana con media  $\alpha/\beta = \mu$  y varianza  $\sigma^2/2\beta = \sigma^2/2\kappa$ , lo cual se denota por

$$r_t \sim N \left( \frac{\alpha}{\beta}, \frac{\sigma^2}{2\beta} \right).$$

### 2.2.5 Modelo de Dothan

Dothan (1978) introdujo una modificación en el modelo de Vasicek la cual consiste en asumir que la volatilidad del tipo de interés es proporcional al tipo de interés, con lo que la volatilidad cambia a lo largo del tiempo,

$$dr_t = \sigma r_t dW_t. \quad (2.37)$$

En este caso el tipo de interés se rige por un proceso Browniano geométrico sin tendencia, dando como resultado que el tipo de interés se distribuya como una variable aleatoria lognormal, y por

tanto, pudiendo tomar valores negativos. el modelo también se conoce como camino aleatorio geométrico o camino aleatorio elástico, y ha sido usado también por Brennan y Schwartz (1979) para la valoración de bonos con opciones incorporadas y en el desarrollo de modelos numéricos de ahorro y de bono rescatable.

Con este modelo se presenta el inconveniente de que al modelizar el comportamiento de los tipos de interés mediante una lognormal, esta, eleva su complejidad, y presenta problemas de estabilidad. Por otro lado, no es posible obtener una solución exacta para el precio de los bonos de cupón cero ni de las opciones.

### 2.2.6 Modelo de Cox-Ingersoll-Ross (CIR)

Otro modelo de gran interés dentro del enfoque de los modelos de equilibrio general es el desarrollado por Cox et al. (1985) por lo que se denota por las iniciales CIR. Dicho modelo inicialmente introducido por Feller (1951) para el estudio del crecimiento de poblaciones y posteriormente Cox, Ingersoll y Ross lo popularizan en las finanzas en la modelización de los tipos de interés a corto plazo. Concretamente, los autores introducen una modificación en el coeficiente de difusión propuesto por Vasicek (1977), al considerar en este, la raíz cuadrada del tipo de interés instantáneo. El modelo resultando se ha convertido en una referencia por su capacidad analítica y porque contrario al modelo propuesto por Vasicek (1977), el tipo de interés no toma valores negativos. La formulación del modelo bajo riesgo neutral viene dada por la ecuación diferencial estocástica

$$dr_t = (\alpha - \beta r_t)dt + \sigma\sqrt{r_t}dW_t, \quad r_0 = r_o > 0, \quad (2.38)$$

para  $\alpha, \beta$  y  $\sigma$  constantes positivas. O bien en su parametrización alternativa

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t, \quad r_0 = r_o > 0, \quad (2.39)$$

con  $\mu, \kappa$  y  $\sigma$  constantes positivas. Bajo la condición,  $2\alpha = 2\kappa\mu > \sigma^2$ , el proceso es estrictamente positivo, con lo cual el proceso no alcanza el origen. La ecuación diferencial estocástica (2.38) tiene solución explícita dad por

$$r_t = \left(r_0 - \frac{\alpha}{\beta}\right)e^{-\beta t} + \sigma e^{-\beta t} \int_0^t e^{\beta u} \sqrt{r_u} dW_u. \quad (2.40)$$

La estacionariedad del proceso se satisface bajo la condición mencionada anteriormente,  $2\alpha = 2\kappa\mu > \sigma^2$ , y su función densidad condicional existe en forma explícita. El proceso

$$Y_t = 2cr_t, \quad \text{con } c = \frac{2\beta}{\sigma^2(1 - e^{-\beta t})},$$

tiene distribución condicional  $Y_t|Y_0 = y_0$ , chi-cuadrado no centrada con  $\nu = 4\alpha/\sigma^2$  grados de libertad y parámetro nocentral  $y_0 e^{-\beta t}$ . Así pues, la densidad condicional del proceso original (es decir, la distribución de  $r_t|r_0 = r_o$ ) se puede obtener a partir de la distribución condicional de  $Y_t$

$$p(t, r|r_o) = ce^{u+v} \left(\frac{u}{v}\right)^{q/2} I_q(2\sqrt{uv}), \quad (2.41)$$

con  $u = cr_o e^{-\beta t}$ ,  $v = cr$ ,  $q = 2\alpha/\sigma^2 - 1$ , y donde

$$I_q(x) = \sum_{j=0}^{\infty} \left(\frac{x}{2}\right)^{2j+q} \frac{1}{j!\Gamma(j+q+1)}, \quad x \in \mathbb{R},$$

y  $\Gamma(\cdot)$  es la función gamma. La media y varianza condicionada de  $r_t$

$$\begin{aligned} \mathbb{E}\{r_t|r_0 = r_o\} &= \frac{\alpha}{\beta} + \left(r_o - \frac{\alpha}{\beta}\right)e^{-\beta t}, \\ \text{Var}\{r_t|r_0 = r_o\} &= r_o \frac{\sigma^2}{\beta} (e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta^2} (1 - e^{-2\beta t}). \end{aligned} \quad (2.42)$$

Y la función de covarianza es

$$\mathbb{C}ov(r_s, r_t | r_0) = r_0 \frac{\sigma^2}{\beta} \left( e^{-\beta t} - e^{-\beta(t+s)} \right) + \frac{\alpha \sigma^2}{2\beta^2} \left( e^{-\beta(t-s)} - e^{-\beta(t+s)} \right).$$

Como antes, para el caso  $s \geq t$ , la media y la varianza condicionada a la información hasta el instante  $s$ , esto es,  $\mathcal{F}_s$  son dadas por

$$\begin{aligned} \mathbb{E} \{ r_t | \mathcal{F}_s \} &= \frac{\alpha}{\beta} + \left( r_s - \frac{\alpha}{\beta} \right) e^{-\beta(t-s)}, \\ \mathbb{V}ar \{ r_t | \mathcal{F}_s \} &= r_s \frac{\sigma^2}{\beta} \left( e^{-\beta(t-s)} - e^{-2\beta(t-s)} \right) + \frac{\alpha \sigma^2}{2\beta^2} \left( 1 - e^{-\beta(t-s)} \right)^2. \end{aligned} \quad (2.43)$$

La distribución estacionaria del proceso  $r_t$  se rige por una ley Gamma con parámetro de forma  $2\alpha/\sigma^2$  y parámetro de escala  $\sigma^2/2\beta$ . De manera que la ley estacionaria tiene media  $\alpha/\beta$  y varianza  $\alpha\sigma^2/(2\beta^2)$ . La función de covarianza en este caso es

$$\mathbb{C}ov(r_s, r_t) = \frac{\alpha \sigma^2}{2\beta^2} e^{-\beta(t-s)}.$$

### 2.2.7 Modelo de Chan-Karolyi-Longstaff-Sanders (CKLS)

El modelo propuesto por Chan et al. (1992) y denotado por CKLS, generaliza el proceso de volatilidad constante con reversión a la media. Los autores en su artículo realizan un estudio empírico sobre el comportamiento de este modelo usado para caracterizar los tipos de interés, afirmando que el modelo es apropiado para describir el comportamiento de los tipos de interés. Posteriormente el modelo es aplicado en la valoración de opciones suponiendo que el precio del riesgo de mercado es cero. El modelo se representa por la ecuación diferencial estocástica

$$dr_t = (\alpha - \beta r_t)dt + \sigma r_t^\gamma dW_t. \quad (2.44)$$

Como se puede apreciar en la ecuación (2.44), este modelo conjuga a todos los modelos antes descritos. En este caso, a diferencia de los modelos desarrollados previamente, este modelo no admite una función de densidad de transición explícita, excepto por supuesto cuando  $\alpha = 0$  o  $\gamma = 1/2$ , en cuyo caso estamos en presencia de los modelos antes descritos. En este modelo el tipo de interés, toma valores en  $(0, +\infty)$  si  $\alpha, \beta > 0$ , y  $\gamma > \frac{1}{2}$ . En todos los casos  $\sigma$  se asume positivo.

Como hemos mencionado, el modelo CKLS incluye en una sola expresión a los modelos antes expuestos. En estos modelos se supone que

$$\mu(r_t) = \alpha + \beta r_t, \quad (2.45)$$

$$\sigma^2(r_t) = \sigma^2 r_t^{2\gamma}, \quad (2.46)$$

con  $\sigma^2$  el parámetro de volatilidad que representa un factor de escala de las variaciones no anticipadas de los tipos de interés, y  $2\gamma$  la sensibilidad de la volatilidad ante las variaciones de los tipos de interés, que se conoce como la elasticidad de la varianza.

La relación entre la volatilidad y el tipo de interés, para este modelo, parece evidente. A medida que el tipo de interés aumenta es de esperar que la volatilidad también aumente, por tanto, este tipo de modelos parece coherente. Sin embargo, en ocasiones, valores muy elevados del parámetro  $\gamma$  no dan lugar a este efecto en determinados períodos de tiempo. Otra crítica que han recibido estos modelos consiste en que restringen a la volatilidad a ser una función que únicamente depende del tipo de interés.

Para este modelo no existe una formula cerrada o bien una expresión explícita de su función de densidad condicional. Ait-Sahalia (1999) propone una aproximación de su densidad condicional basada en polinomios de Hermite, la cual permite aproximar dicha densidad de forma eficiente, y que posteriormente se puede usar en la estimación de los parámetros del modelo.

### 2.2.8 Modelo de Aït Sahalia

Aït-Sahalia (1996a) introduce un modelo que produce una ligera reversión a la media, para aquellos tipos de interés que permanecen en la parte media de su dominio, y una fuerte reversión no lineal a la media en los extremos del dominio. Este modelo satisface la ecuación estocástica diferencial no lineal

$$dr_t = (\alpha_1 r_t^{-1} + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t^\gamma dW_t. \quad (2.47)$$

En general, no es posible obtener la distribución exacta del modelo, sin embargo Aït-Sahalia (1999) propone una aproximación de la densidad de transición mediante una expansión de polinomios de Hermite. Así mismo, el autor propone en Aït-Sahalia (1996a) propone una generalización que incluye un término de difusión más complejo

$$dr_t = (\alpha_1 r_t^{-1} + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sqrt{\beta_0 + \beta_1 r_t + \beta_2 r_t^{\beta_3}} dW_t. \quad (2.48)$$

Se imponen algunas restricciones sobre los coeficientes (ver Aït-Sahalia (1996a)). Similar al caso del modelo CKLS, la densidad condicional de este modelo no es posible obtener de manera explícita. De manera que se aplica la misma aproximación usada para el modelo CKLS y que se describirá más adelante.

### 2.2.9 Modelo de Ahn y Gao

Ahn y Gao (1999), proponen un modelo en el que el tipo de interés se rige por un proceso con tendencia no lineal, basados en las ideas expuestas en Aït-Sahalia (1996a) y Aït-Sahalia (1996b). Este modelo se obtiene a partir de una apropiada transformación del modelo CIR. Su especificación viene expresada mediante la ecuación diferencial estocástica

$$dr_t = r_t(\alpha - (\sigma^2 - \alpha\beta)r_t)dt + \sigma r_t^{\frac{3}{2}} dW_t. \quad (2.49)$$

La distribución condicional de este proceso está relacionada a la del modelo CIR,

$$p(t, r|r_0) = \frac{1}{r^2} p^{CIR} \left( t, \frac{1}{r} \mid \frac{1}{r_0} \right),$$

donde  $p^{CIR}$  es la densidad condicional (2.41).

Evidentemente existen otros modelos de difusión que permiten caracterizar el tipo de interés, sin embargo, se ha tenido en cuenta este pequeño grupo puesto que son los modelos más usados y desde el punto de vista teórico son relativamente sencillos para el análisis de sus propiedades.

Los modelos expuestos en esta sección permiten especificar el tipo de interés como una solución de una ecuación diferencial estocástica y por lo tanto hace posible el uso de la Teoría de los Procesos de Markov, puesto que dichos procesos poseen propiedades Markovianas, lo cual facilita su tratamiento analítico. Por otro lado, el uso de este tipo de ecuaciones en la modelización del tipo de interés puede permitir obtener soluciones analíticas, en la medida de lo posible, que faciliten la valoración de los diversos derivados del tipo de interés del mercado.

### 2.2.10 Estimación Paramétrica del Modelo de Difusión Homogéneo

En esta sección se consideran los problemas asociados a la estimación paramétrica de los procesos de difusión. Existe una extensa literatura que hace referencia a la estimación de las funciones de drift y difusión o volatilidad del modelo (2.22), para una revisión detallada de los distintos métodos véase por ejemplo Kutoyants (2004), Prakasa-Rao (1999) y Iacus (2008).

En el caso continuo, es muy sencillo estimar los parámetros eficientemente. En particular,  $\theta$  (al menos el subconjunto de parámetros correspondiente a la difusión en (2.22) se puede calcular en lugar de estimar, a partir de la variación cuadrática del proceso, ya que para todo  $t \geq 0$ ,

$$\langle X, X \rangle_t = \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} (X_{t \wedge k/2^n} - X_{t \wedge (k-1)/2^n})^2 = \int_0^t \sigma^2(X_s, \theta) ds$$

cuando  $n \rightarrow \infty$  en probabilidad bajo  $P_\theta$ . El resto de los parámetros presentes solo en el drift se pueden estimar usando máxima verosimilitud. En realidad, una vez conocido el coeficiente de difusión, la función de verosimilitud  $X$  es dada por:

$$L_T(\theta) = \exp \left( \int_0^T \frac{\mu(X_s, \theta)}{\sigma^2(X_s)} dX_s - \frac{1}{2} \int_0^T \frac{\mu(X_s, \theta)}{\sigma^2(X_s)} ds \right) \quad (2.50)$$

De manera que  $\theta$  se puede estimar maximizando  $L_T(\theta)$ . Evidentemente, en la práctica, todos los modelos en tiempo continuo son observados en tiempo discreto. Por lo tanto, para la estimación de los parámetros del modelo en tiempo continuo (2.22) se considera una versión discreta del mismo. Entre los esquemas de observación considerados en la literatura se tienen:

- *Esquema de muestras numerosas.*

Bajo este esquema, el tiempo  $\Delta$  entre dos observaciones consecutivas es fijo y el número de observaciones  $n$  creciente. En este caso, la ventana de observación  $[0, n\Delta = T]$  también se incrementa con  $n$ . Este enfoque, considerado el más natural y habitual en el contexto financiero, se requieren suposiciones adicionales sobre el modelo continuo subyacente tales como estacionariedad o ergodicidad.

- *Esquema de altas-frecuencias.*

En este caso,  $\Delta = \Delta_n$  tiende a cero cuando se incrementa  $n$ , y la ventana de observación  $[0, n\Delta_n = T]$  es fija. Bajo este esquema no se hace necesario exigir condiciones de estacionariedad ni de ergodicidad.

- *Diseño de incremento rápido.*

Este enfoque asume que  $\Delta_n$  disminuye-reduce a cero cuando  $n$  aumenta, y la ventana de observación  $[0, n\Delta_n]$  se incrementa con  $n$ , es decir,  $n\Delta_n \rightarrow \infty$ . En este caso, es necesario suponer estacionariedad y ergodicidad. Además, el tamaño del paso de discretización  $\Delta_n$  debería tender a cero a una tasa  $n\Delta_n^k \rightarrow 0$ , con  $k \geq 2$ . Para valores de  $k$  altos, esto no es una restricción severa porque esto significa que  $\Delta_n$  tiende a cero lentamente.

En lo que respecta a esta memoria y para efectos de la propuesta que se desarrolla en un contexto financiero se asumirá el primero de los esquemas antes descrito. Así pues, se tiene que una versión alternativa en tiempo discreto del modelo (2.22) y que se obtiene aplicando el método de Euler (discutido en la sección (1.10.2) del capítulo 1), viene dada por:

$$r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)\Delta^{1/2}\varepsilon_{t_i}, i = 1, \dots, n, \quad (2.51)$$

donde  $\{\varepsilon_{t_i}\}$  son variables aleatorias normales estándar independientes y  $t_i = i\Delta$ , con  $i = 1, \dots, n$  en un intervalo de tiempo  $[0, n\Delta = T]$ , para  $\Delta$  fijo, es decir, las observaciones son equidistantes. Por ejemplo, cuando la unidad de tiempo es un año, y la muestra es seleccionada con frecuencia semanal (mensual) en  $t_i = i/52$ , entonces  $\Delta = 1/52$  ( $\Delta = 1/12$ ) para  $i = 0, \dots, n$ .

Cabe destacar que, el método Euler de aproximación mediante discretización del modelo (2.22) no es el único método disponible para tal fin. Otro de los métodos de uso frecuente en la literatura es el método de Milstein (1978), el cual hace uso del lema de Itô para incrementar la precisión de

la aproximación. La diferencia entre este método y el de Euler, es que este método incluye un término de segundo orden, obteniéndose

$$r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)(W_{t_{i+1}} - W_{t_i}) + \frac{1}{2}\sigma(r_{t_i}, \theta)\sigma_r(r_{t_i}, \theta) [(W_{t_{i+1}} - W_{t_i})^2 - \Delta] \quad (2.52)$$

con  $\sigma_x(r_{t_i}, \theta)$  la derivada parcial de  $\sigma(r, \theta)$  con respecto a  $r$ . Esta aproximación tiene orden de convergencia  $\gamma = 1$ , indiferentemente si es fuerte o débil.

Es claro que si se consideran más términos en la expansión de Taylor en el contexto del cálculo de Itô, entonces el orden (fuerte) de convergencia se incrementa por encima de  $\gamma = 1$ . Kloden et al. (2000) proponen un método de aproximación, denotado por *KPS*, con orden de convergencia fuerte  $\gamma = 1.5$ . Las siglas *KPS* son las iniciales de los autores involucrados en el método, Kloden, Platen y Schurz. Es importante destacar que para efectos del desarrollo de esta memoria solo será considerado el método de Euler como método de aproximación al modelo de difusión en tiempo continuo.

La versión discreta (2.51) del modelo (2.22) es de gran utilidad en la simulación de trayectorias del modelo de difusión, así mismo, dicha aproximación en tiempo discreto permite estimar los parámetros del modelo, como se verá a continuación.

Se denotará por  $\mathcal{F}_n = \sigma\{r_{t_i}, i \leq n\}$  la sigma algebra generada por las  $n$  primeras observaciones y que suele denominarse la *información disponible hasta el instante  $n$* . Claramente  $\mathcal{F}_0$  es referida a la sigma algebra trivial. Entonces la función de verosimilitud del proceso discreto viene dada por

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, r_{t_i} | r_{t_{i-1}}) p_\theta(r_{t_0}) \quad (2.53)$$

donde  $p_\theta(\Delta, r_{t_i} | r_{t_{i-1}})$  denotará la función densidad de transición o densidad condicional asociada al modelo de difusión paramétrico, con parámetro desconocido  $\theta$ . La función de verosimilitud  $L_n(\theta)$  se puede obtener haciendo uso de la propiedad de Markov del proceso  $r_t$  (ver por ejemplo, Arnold (1974)). Sea

$$\begin{aligned} \ell_n(\theta) &= \log L_n(\theta) = \sum_{i=1}^n \ell_i(\theta) + \log(p_\theta(r_{t_0})) \\ &= \sum_{i=1}^n \log p_\theta(\Delta, r_{t_i} | r_{t_{i-1}}) + \log(p_\theta(r_{t_0})). \end{aligned} \quad (2.54)$$

el logaritmo de la función de verosimilitud. Por lo general  $p_\theta(r_{t_0})$  es desconocida, o bajo ciertos supuestos no siempre es fácil de determinar. Si el número de observaciones es suficientemente grande se puede asumir que el peso relativo de  $p_\theta(r_{t_0})$  en la función de verosimilitud decrece, por lo que puede asumirse que  $p_\theta(r_{t_0}) = 1$ .

Si se conoce la forma paramétrica del modelo que genera las observaciones  $\{r_{t_i}\}_i$ , entonces el método a aplicar naturalmente es el de máxima verosimilitud. Así, el estimador de máxima verosimilitud (*MLE*) viene expresado por

$$\hat{\theta} = \arg \max_{\theta} \ell_n(\theta).$$

Como ya se mencionó, mientras que los modelos son descritos en tiempo continuo, la disponibilidad de los datos son siempre en tiempo discreto. Si se ignora esta diferencia, las estimaciones pueden resultar ser inconsistentes. El problema principal entonces es que la verosimilitud para observaciones discretas generadas por el modelo (2.51) no se puede determinar explícitamente para todos los modelos, salvo en algunos casos como por ejemplo los modelos de difusión de Vasicek (VAS)-Ornsntein Uhlenbeck o *CIR*, en general, no se dispone de formulas cerradas para la densidad de transición o de probabilidad condicional, como es el caso del modelo *CKLS*, entre



otros. Aún en el caso del modelo *CIR*, en el que se conoce la distribución asociada al modelo, se presentan inconvenientes, relativos a las funciones que comprende dicha distribución, tales como la evaluación de la función de Bessel o bien la función Gamma, las cuales pueden complicar el proceso de optimización.

Una técnica común que se emplea para estimar el modelo (2.21), es proceder como si las observaciones proviniesen de una distribución gaussiana, de media el drift del modelo y desviación estándar la función de volatilidad, y entonces obtener el estimador de máxima verosimilitud. Esto es, la densidad de transición tiene la siguiente forma

$$p_{r,\theta}(\Delta, y|r) = \frac{1}{\sqrt{2\pi\Delta\sigma^2(r,\theta)}} \exp \left\{ -\frac{1}{2} \frac{(y - r - \mu(r,\theta)\Delta)^2}{\Delta\sigma^2(r,\theta)} \right\} \quad (2.55)$$

Es importante destacar que este método es eficiente si  $\Delta$ , el paso de discretización, es suficientemente pequeño (véase [Fan y Zhang \(2003\)](#)), ([Iacus \(2008\)](#), pág. 122) ya que puede producir sesgos significativos en las estimaciones cuando  $\Delta$  es grande.

Antes de continuar con la exposición, es necesario introducir algunos supuestos básicos relativos a los coeficientes del proceso de difusión, las funciones drift y difusión concretamente.

### Conjunto de Supuestos Básicos

*Supuesto 1.* ( Lipschitz Global ) Existe una constante  $K$  independiente de  $\theta$  tal que, para todo  $r$  y  $s$ ,

$$|\mu(r,\theta) - \mu(s,\theta)| + |\sigma(r,\theta) - \sigma(s,\theta)| \leq K(|r - s|).$$

*Supuesto 2.* ( Crecimiento lineal )

Existe una constante  $K$  independiente de  $\theta$  tal que,

$$|\mu(r,\theta)| + |\sigma(r,\theta)| \leq K(1 + |r|).$$

*Supuesto 3.* ( Positividad del coeficiente de difusión )

Existe una constante  $K$  independiente de  $\theta$  tal que, para todo  $r$ ,

$$\inf_r \sigma^2(r,\theta) > 0$$

*Supuesto 4.* ( Momentos Acotados )

Para todo  $k > 0$ , todos los momentos de orden  $k$  del proceso de difusión existen y son tales que

$$\sup_t \mathbb{E}|r_t|^k < \infty.$$

*Supuesto 5.* ( Suavidad de los coeficientes )

Los dos coeficientes  $\mu$  y  $\sigma$  y sus derivadas en  $\theta$  (eventualmente hasta el orden 3) son suaves en  $r$  y de crecimiento de orden polinomial en  $r$  uniformemente en  $\theta$ .

Este conjunto de supuestos se completa además con condiciones técnicas que aseguren la adecuada tasa de convergencia y la existencia de la información de fisher asociada al método de estimación. Cuando los parámetros del drift y la difusión están separados, la tasa de convergencia es siempre más rápida para los coeficientes de la difusión, y se puede aplicar una aproximación en dos etapas para hallar los parámetros. Por lo general, la tasa de convergencia para el coeficiente de la difusión es  $\sqrt{n}$  y, bajo la condición  $n\Delta_n^3 \rightarrow 0$ , el estimador para los parámetros en el drift tienen tasa de convergencia  $\sqrt{n\Delta_n}$  (en algunos casos se requiere  $n\Delta_n^2 \rightarrow 0$ ).

A objeto de reducir el sesgo de estimación [Aït-Sahalia \(1999\)](#), sugiere una aproximación de la función de densidad de transición mediante polinomios de Hermite, la cual es eficiente en



cualquiera de los casos, conocida o no la distribución del proceso. Aplicando esta aproximación el autor obtiene un estimador de máxima verosimilitud, en el que todos los parámetros, del drift y difusión, son estimados de manera conjunta, ver además [Aït-Sahalia \(2002\)](#), y dicho estimador es “ $\sqrt{n}$ ” consistente, para un proceso de difusión.

En este trabajo, es de suma importancia obtener estimaciones consistentes de los parámetros, de manera que en lo que sigue, se aplicará el método de Máxima Verosimilitud sugerido por [Aït-Sahalia \(2002\)](#), dado que su implementación y el costo computacional son bastante razonables, conforme a los requerimientos necesarios para el desarrollo de la propuesta que se presentará en el próximo capítulo.

### Método de Máxima Verosimilitud de Aït-Sahalia

El método de Estimación por Máxima Verosimilitud propuesto por [Aït-Sahalia \(2002\)](#) para muestras discretas de un proceso de Difusión esta basado en una aproximación de la función de verosimilitud. El autor construye una sucesión de funciones explícitas cerradas mediante polinomios de Hermite, y demuestra que dichas funciones convergen a la verdadera función de verosimilitud. El estimador se obtiene al maximizar la sucesión obtenida, dando como resultado un estimador que converge al verdadero estimador de máxima verosimilitud, el cual comparte las mismas propiedades asintóticas. Tal como es expresado por el autor en su artículo, este método supera a otras aproximaciones en situaciones pertinentes a modelos financieros. Para ver como funciona este método considere el proceso de difusión (2.22) y  $p_{r,\theta}(\Delta, r|r_0)$  la densidad condicional de  $r_{t+\Delta} = r$  dado  $r_t = r_0$  inducida por el modelo (2.22). [Aït-Sahalia \(1999\)](#) construye una sucesión de aproximaciones  $p_{r,\theta}^{(j)}$  de la densidad de transición, de manera que el logaritmo de la verosimilitud

$$\ell_n(\theta) = \frac{1}{n} \sum_{i=1}^n \ln\{p_{r,\theta}(\Delta, r_{i\Delta}|r_{(i-1)\Delta})\} \quad (2.56)$$

sea una sucesión de aproximaciones  $\ell_n^{(j)}$ . Evidencia empírica proporcionada por el autor sugiere que para  $j = 2$  y  $j = 3$  la aproximaciones son adecuadas en el contexto de las finanzas. Ahora bien, para construir la sucesión de aproximaciones se estandariza la difusión de  $r$ , esto es, se aplica una transformación al proceso  $r$  para obtener otro proceso de difusión  $U$

$$U_t = \gamma(r_t; \theta) = \int^{r_t} \frac{ds}{\sigma(s, \theta)}. \quad (2.57)$$

Se puede seleccionar cualquier primitiva de la función  $1/\sigma$ . En el contexto que nos compete se definen los dominios de la difusión  $D_r$  y  $D_U$  como la recta real o el conjunto  $(0, +\infty)$ . Aplicando las formulas de Itô al nuevo proceso  $U$  se tiene la difusión unitaria

$$dU_t = \mu_U(U_t; \theta)dt + dW_t \quad (2.58)$$

con función drift  $\mu_U$  dada por

$$\mu_U(u; \theta) = \frac{\mu(\gamma^{-1}(u; \theta); \theta)}{\sigma(\gamma^{-1}(u; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial r}(\gamma^{-1}(u; \theta); \theta). \quad (2.59)$$

Esta transformación permite determinar la densidad de transición  $p_{r,\theta}$  desde  $p_U$  a través de la fórmula del jacobiano

$$\begin{aligned} p_{r,\theta}(\Delta, r|r_0) &= \frac{\partial}{\partial r} \text{Prob}(r_{t+\Delta} \leq r | r_t = r_0; \theta) \\ &= \frac{p_{U,\theta}(\Delta, \gamma(r; \theta) | \gamma(r_0; \theta))}{\sigma(\gamma(r; \theta))} \end{aligned} \quad (2.60)$$

Se pueden derivar expresiones explícitas de la densidad de transición de la variable  $U$  basadas en expansiones de Hermite. La parte analítica de la expansión de  $p_{U,\theta}$  hasta el orden  $K$  esta dada por:

$$\tilde{p}_{U,\theta}^{(K)}(\Delta, u|u_0) = \Delta^{-1/2} \phi\left(\frac{u-u_0}{\Delta^{1/2}}\right) \exp\left\{\int_{u_0}^u \mu_U(w; \theta) dw\right\} \sum_{k=0}^K c_k(u|u_0; \theta) \frac{\Delta^k}{k!} \quad (2.61)$$

donde  $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$  densidad normal  $N(0, 1)$ . Además  $c_0(u|u_0; \theta) = 1$  y para todo  $j \geq 1$

$$c_j(u|u_0; \theta) = j(u-u_0)^{-j} \left( \int_{u_0}^u (w-u_0)^{j-1} \{\lambda_u(w)c_{j-1}(w|u_0; \theta) + (\partial^2 c_{j-1}(w|u_0; \theta)/\partial w^2)/2\} dw \right) \quad (2.62)$$

donde  $\lambda_u(u; \theta) \equiv -(\mu_{U^2}(u; \theta) + \partial\mu_U(u; \theta)/\partial u)/2$ . De esta manera,

$$\tilde{p}_{r,\theta}^{(K)}(\Delta, r|r_0) \equiv \sigma(r; \theta)^{-1} \tilde{p}_U^{(K)}(\Delta, \gamma(r; \theta)|\gamma(r_0; \theta)) \quad (2.63)$$

la sucesión de funciones explícitas  $\tilde{p}_U^{(K)}$  en la ecuación (2.61) aproxima a  $p_U$ . El objetivo fundamental de la transformación  $r \rightarrow U$  es para controlar el tamaño de las colas de la densidad de transición. Ait-Sahalia (1999) demuestra que

$$\tilde{p}_{r,\theta}^{(K)}(\Delta, r|r_0) \rightarrow p_{r,\theta}(\Delta, r|r_0).$$

Para calcular los estimadores de máxima verosimilitud, El autor maximiza la aproximación del logaritmo de la verosimilitud

$$\ell_n^{(K)} \equiv n^{-1} \sum_{i=1}^n \ln\{p_{r,\theta}^{(K)}(\Delta, r_{i\Delta}|r_{(i-1)\Delta})\} \quad (2.64)$$

Con lo cual, se obtiene un estimador  $\hat{\theta}_n^{(K)}$  para el cual, demuestra el autor es cercano al exacto  $\hat{\theta}_n$ . Para el caso particular en que el proceso admite una distribución estacionaria, la varianza asintótica del estimador de máxima verosimilitud esta dado por la inversa de la matriz de información de Fisher. El estimador propuesto satisface para procesos estacionarios

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1}), \quad (2.65)$$

donde la matriz de información de Fisher es

$$i(\theta) \equiv E\left[\dot{L}(\theta)\dot{L}(\theta)^T\right] = -E\left[\ddot{L}(\theta)\right].$$

con  $L(\theta) = \ln(p_{r,\theta}(\Delta, r_\Delta|r_0))$ ,  $\dot{L}(\theta) = \partial L(\theta)/\partial \theta$  y  $\ddot{L}(\theta) = \partial^2 L(\theta)/\partial \theta \partial \theta^T$

### Método Generalizado de los Momentos

Chan et al. (1992), introducen el modelo (CKLS), y consideran el método generalizado de los momentos GMM para la estimación de los parámetros de dicho modelo

$$\begin{aligned} dr_t &= (\alpha_0 + \beta_0 r_t) dt + \sigma_0 r_t^{\gamma_0} dW_t \\ \theta_0 &= (\alpha_0, \beta_0, \sigma_0, \gamma_0)' \end{aligned} \quad (2.66)$$

La discretización de Euler del modelo (2.66)

$$\begin{aligned} r_{t+\Delta_t} - r_t &= (\alpha_0 + \beta_0 r_t) \Delta_t + \sigma_0 r_t^{\gamma_0} \sqrt{\Delta_t} \mathbf{z}_{t+\Delta_t} \\ E[\mathbf{z}_{t+\Delta_t}] &= 0, \quad E[\mathbf{z}_{t+\Delta_t}^2] = 1 \end{aligned} \quad (2.67)$$

define el error del modelo como

$$\begin{aligned}\varepsilon_{t+\Delta_t} &= a(r_{t+\Delta_t} - r_t, r_t; \alpha_0, \beta_0, \sigma_0, \gamma_0) = (r_{t+\Delta_t} - r_t) - (\alpha_0 + \beta_0 r_t) \Delta_t \\ &= a(\mathbf{z}_{t+\Delta_t}, \theta_0)\end{aligned}\quad (2.68)$$

donde  $\mathbf{z}_{t+\Delta_t} = (r_{t+\Delta_t} - r_t)'$ . Sea  $I_t$  que representa la información disponible en el tiempo  $t$ , el error verdadero satisface  $E(\varepsilon_{t+\Delta_t} | I_t) = 0$ .

Como base para la estimación, [Chan et al. \(1992\)](#) usan  $\mathbf{x}_t = (1, r_t)'$  como vector de variables instrumentales y deducen las cuatro condiciones siguientes:

$$\begin{aligned}E[\varepsilon_{t+\Delta_t} | I_t] &= 0; & E[\varepsilon_{t+\Delta_t}^2 | I_t] &= \sigma_0^2 r_t^{2\gamma_0} \Delta_t \\ E[\varepsilon_{t+\Delta_t} r_t | I_t] &= 0; & E[\varepsilon_{t+\Delta_t}^2 r_t | I_t] &= \sigma_0^2 r_t^{2\gamma_0} \Delta_t \cdot r_t\end{aligned}\quad (2.69)$$

Para  $\alpha$  y  $\beta$  dados se define el residuo no-lineal

$$e_{t+\Delta_t} = (r_{t+\Delta_t} - r_t) - (\alpha + \beta r_t) \Delta_t$$

y, para  $\mathbf{w}_{t+\Delta_t} = (r_{t+\Delta_t} - r_t, r_t, r_t^2)'$ , define el vector de momentos

$$g(\mathbf{w}_{t+\Delta_t}, \theta) = \begin{pmatrix} e_{t+\Delta_t} \\ e_{t+\Delta_t}^2 \end{pmatrix} \otimes \mathbf{x}_t = \begin{pmatrix} e_{t+\Delta_t} \\ e_{t+\Delta_t} r_t \\ e_{t+\Delta_t}^2 - \sigma^2 r_t^{2\gamma} \Delta_t \\ (e_{t+\Delta_t}^2 - \sigma^2 r_t^{2\gamma} \Delta_t) r_t \end{pmatrix}$$

Entonces  $E[g(\mathbf{w}_{t+\Delta_t}, \theta_0)] = 0$  da las ecuaciones de estimación del método. Para la especificación más general, (2.66),  $K = L$  por lo tanto el modelo esta totalmente especificado. Mientras que para el resto de modelos que se derivan de éste  $K > L$ , así que el modelo esta sobreidentificado. Para mayores detalles ver [Chan et al. \(1992\)](#).

### 2.2.11 Estimación No-Paramétrica

Cuando no se establece ninguna especificación o bien no existe información acerca de la forma paramétrica de las funciones de drift y difusión o volatilidad del modelo, los métodos no-paramétricos son de gran utilidad a la hora de analizar las propiedades del proceso y en la determinación de las funciones antes mencionadas. En esta sección se hará una revisión de algunas de las técnicas no-paramétricas (métodos de suavizado) aplicadas en el contexto de los modelos de difusión, principalmente orientados a la determinación de la función de densidad invariante, y las funciones de drift y difusión.

Sea el proceso de difusión ergódico  $r$ , solución de la ecuación diferencial estocástica

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t, \quad (2.70)$$

donde  $\mu(\cdot)$  y  $\sigma(\cdot)$  satisfacen condiciones de regularidad (ver Apéndice C) y además, existe una medida invariante  $\pi(r)$  para el proceso  $r$ . Sea  $K$  una función que satisfice:

- $\int K(u)du = 1$ ,
- $K$  es acotado y diferenciable dos veces en  $\mathbb{R}$ .

Además  $K$  y sus derivadas están definidas en el espacio  $L^2(\mathbb{R})$ . Entonces la función  $K$  es denominada una *función Kernel* o *función Núcleo* de orden  $r > 1$  si existe un entero  $r$  tal que

$$\int_{-\infty}^{\infty} x^i K(x)dx = 0, \quad i = 1, \dots, r-1,$$

y

$$\int_{-\infty}^{\infty} x^r K(x) dx \neq 0, \quad \int_{-\infty}^{\infty} |x|^r K(x) dx < \infty,$$

Se asumirá que  $K$  es de orden 2 y que

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

Nótese además que  $\lim_{h \rightarrow \infty} K_h(u) = \delta(u)$ , para  $\delta$  una función delta de Dirac.

### Estimación de la densidad estacionaria

El *estimador tipo núcleo* de la densidad estacionaria viene dado por

$$\hat{\pi}_n(r) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{r - r_{t_i}}{h_n}\right) = \frac{1}{n} \sum_{i=1}^n K_{h_n}(r - r_{t_i}) \quad (2.71)$$

El núcleo que con mayor frecuencia se usa es el núcleo Gaussiano

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2},$$

sin embargo, cualquier otra función núcleo se puede considerar. Como es común en un enfoque no-paramétrico, las técnicas de suavizado requieren de una selección adecuada del *ancho de banda* o *parámetro de suavizado*  $h_n$ . El ancho de banda  $h_n$  es una sucesión que se contrae/disminuye su valor con  $n$ , esto es,  $h_n \rightarrow 0$  cuando  $n \rightarrow \infty$ . Una selección básica o piloto es  $h_n = c_n n^{-1/5}$ , donde  $c_n$  es  $c$  veces la desviación estándar de los datos dividido por  $\log(n)$  y  $c$  se escoge de tal manera que minimiza el error cuadrático medio integrado (*MISE*) del estimador de la densidad  $\pi(r)$ . Otra selección suele ser, aplicando el método de cross-validación. Bajo la suposición  $\lim_{n \rightarrow \infty} nh_n^{4.5} = 0$  y bajo condiciones de regularidad, el estimador de la densidad estacionaria  $\hat{\pi}_n$  se comporta como en el caso de variables aleatorias independientes, e idénticamente distribuidas. En particular, se tiene que

$$\sqrt{nh_n}(\hat{\pi}_n(r) - \pi(r)) \xrightarrow{d} N(0, V_r)$$

donde

$$V_r = \pi(r) \int_{-\infty}^{+\infty} K^2(u) du.$$

Resultados referidos a la consistencia asintótica, se pueden encontrar en [Boente y Fraiman \(1988\)](#).

### Estimación de las funciones de difusión y drift

En cuanto a la estimación de los coeficientes del proceso de difusión o bien funciones drift y difusión, se tienen dos enfoques. El primero, considera el conocimiento de la función drift, y el segundo el total desconocimiento de ambas funciones, drift y difusión.

El trabajo de [Aït-Sahalia \(1996a\)](#) es uno de los principales referencias en cuanto a un enfoque paramétrico para los modelos de tipos de interés se refiere. De las ecuaciones de Fokker-Planck (ver (2.2) de [Aït-Sahalia \(1996a\)](#)), se puede obtener

$$\frac{d^2}{dr^2}(\sigma^2(r)\pi(r)) = 2\frac{d}{dr}(\mu(r)\pi(r)). \quad (2.72)$$

Integrando y reordenando alternativamente

$$\mu(r) = \frac{1}{2\pi(r)} \frac{d}{dr}[\sigma^2(r)\pi(r)] \quad \sigma^2(r) = \frac{1}{2\pi(r)} \int_0^r \mu(u)\pi(u) du \quad (2.73)$$

bajo la condición  $\pi(0) = 0$ . Como se puede observar, estas ecuaciones permiten estimar, bien, el drift  $\mu(\cdot)$  dada una especificación de la función de difusión  $\sigma(\cdot)^2$  y la densidad marginal o bien, el término de difusión dada una especificación de la función de drift y la densidad marginal estimada. De esta manera, tomando (2.71) como el estimador de la densidad estacionaria, y si  $\mu(\cdot)$  se conoce o tiene forma paramétrica conocida para el cual se disponen estimadores consistentes, entonces se puede estimar la función de difusión por:

$$\hat{\sigma}_n^2(r) = \frac{2}{\hat{\pi}_n(r)} \int_0^r \hat{\mu}(u) \pi_n(u) du = \frac{2}{\hat{\pi}_n(r)} \int_0^r \hat{\mu}(u, \hat{\theta}) \pi_n(u) du \quad (2.74)$$

donde  $\mu(\cdot, \hat{\theta})$  es el estimador de la función drift y  $\hat{\theta}$  es un estimador " $\sqrt{n}$ " consistente de  $\theta$ . Aït-Sahalia (1996a) y Gao (2007), Arapis y Gao (2006) consideran similares alternativas, como es el caso de  $\mu(r, \theta) = \kappa(\mu - r)$ . Evidentemente, en este caso este enfoque no debería ser considerado del topo no-paramétrico, dado que se debe especificar alguna de las funciones de forma paramétrica. En todo caso, este enfoque es denominado *semiparamétrico* y es discutido en detalle en Gao (2007). Las propiedades asintóticas de este estimador se pueden revisar en (§ 4.2, Iacus (2008), pág. 202) y las referencias incluidas en el mismo.

Si se supone que no se dispone de información acerca de la función drift, entonces la estimación de los coeficientes se lleva totalmente de forma no-paramétrica. En algunos casos, si se conoce la función drift, es posible obtener estimaciones consistentes de la difusión, como se verá más adelante. En el primer caso, el enfoque no-paramétrico requiere un esquema de altas frecuencias. Así pues, el estimador propuesto por Florens-Zmirou (1993) para la función de difusión

$$\hat{\sigma}_n^2(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r - r_{t_i}}{h_n}\right) (r_{t_{i+1}} - r_{t_i})^2}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{r - r_{t_i}}{h_n}\right)}. \quad (2.75)$$

En este caso, el estimador requiere la condición  $\Delta_n \rightarrow 0$  en alguna tasa apropiada para obtener la consistencia y la normalidad asintótica. Así mismo, un estimador no paramétrico del drift se puede obtener como sigue:

$$\hat{\mu}_n(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r - r_{t_i}}{h_n}\right) (r_{t_{i+1}} - r_{t_i})}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{r - r_{t_i}}{h_n}\right)}. \quad (2.76)$$

Si se toma en cuenta el trabajo de Stanton (1997), los dos estimadores, no son otros que los estimadores de tipo núcleo de Nadaraya-Watson de las esperanzas condicionales

$$\mu(r) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}[(r_t - r | r_0 = r)],$$

$$\sigma^2(r) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}[(r_t - r)^2 | r_0 = r].$$

Las ecuaciones anteriores corresponden a las medias y varianza condicionales del proceso cuando

$r_0 = r$ . Estas se pueden reescribir, para  $\Delta_n$  fijo, como

$$\mu(r) = \frac{1}{\Delta_n} \mathbb{E} [(r_{t_{i+1}} - r_{t_i}) | r_{t_i} = r] + \frac{o(\Delta_n)}{\Delta_n}, \quad (2.77)$$

$$\sigma^2(r) = \frac{1}{\Delta_n} \mathbb{E} [(r_{t_{i+1}} - r_{t_i})^2 | r_{t_i} = r] + \frac{o(\Delta_n)}{\Delta_n}, \quad (2.78)$$

En cuanto al parámetro de suavizado  $h_n$ , como antes, diversas son las alternativas, sin embargo, [Gao \(2007\)](#) discute algunas consideraciones acerca del ancho de banda, entre las que incluye una bastante genérica, obtenida a partir del método de validación cruzada, dada por

$$h_{cv} = 1.06n^{-1/5} s$$

donde  $s$  es la desviación estándar de la muestra,

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (r_{t_i} - \bar{r})^2}.$$

Para una revisión detallada de este tema ver [Fan y Zhang \(2003\)](#), [Gao \(2007\)](#), [Iacus \(2008\)](#).

### 2.2.12 Modelos de Difusión no homogéneos respecto del tiempo

Los tipos de interés a corto plazo libres de riesgo son fundamentales e importantes en los mercados financieros. Ellos están directamente relacionados con gastos de consumidor, ganancias de empresas, precios activos, inflación y a la economía global. Muchos modelos de tipos de interés a corto plazo han tenido el propósito de explicar la dinámica de la estructura temporal de dichos modelos y otras cuestiones financieras.

Gran parte de los modelos desarrollados, tales como, el movimiento Browniano geométrico ([Black y Scholes \(1973\)](#)) y los modelos expuestos en el capítulo anterior, son simples modelos paramétricos homogéneos respecto del tiempo, que intentan obtener ciertas características relevantes de la dinámica de los movimientos observados.

Las condiciones económicas varían con el transcurrir del tiempo. Por ello, es razonable esperar que tanto los retornos esperados instantáneos como la volatilidad dependan del tiempo y del nivel de precios para una variable de estado dada, tal como el precio de acciones y rendimiento de bonos. Sin embargo, resulta difícil describir, con precisión, como las funciones bivariadas de los retornos esperados y la volatilidad varían respecto del tiempo y del nivel de precios. Una manera de asumir el problema es hacer un planteamiento no paramétrico del mismo. Una visión general del planteamiento no paramétrico en regresión se puede ver en [Hastie y Tibshirani \(1990\)](#), [Härdle \(1990\)](#), [Fan y Gijbels \(1996\)](#), entre otros. Aún así, no se dispone de información suficiente para determinar de manera no paramétrica dichas funciones. Por lo tanto, será crucial imponer alguna forma a las funciones de retornos instantáneos y a la función de volatilidad.

[Fan et al. \(2003\)](#), tratan de determinar la influencia de la variación del tiempo en los retornos instantáneos y la volatilidad, extendiendo el modelo de tipo de interés de [Chan et al. \(1992\)](#) desde dos puntos de vista: primero consideran los coeficientes como funciones suaves respecto del tiempo y segundo permiten una transformación de la variable de estado dentro de la ecuación. El objetivo que persiguen los autores es reducir el sesgo en el modelo evitando una excesiva inflación de la varianza de las funciones estimadas. A continuación una breve explicación del método.

El modelo más genérico, viene dado por:

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dW_t, \quad (2.79)$$

con  $\{W_t, 0 \leq t \leq T\}$  es un movimiento Browniano estándar, y donde el *drift* y la *difusión* del proceso  $r_t$  satisfacen

$$\mu(t, r_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E} [r_{t+\Delta} - r_t \mid r_t] \quad (2.80)$$

y

$$\sigma^2(t, r_t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E} [(r_{t+\Delta} - r_t)^2 \mid r_t] \quad (2.81)$$

respectivamente.

Son casos particulares del modelo (2.79), entre otros, los de Ho y Lee (1986)(HL), Hull y White (1990)(HW), Black et al. (1990)(BDT) y Black y Karasinski (1991)(BK). Las formas respectivas son las siguientes:

$$\text{HW} : dr_t = (\alpha_0(t) + \alpha_1(t)r_t)dt + \sigma(t)r_t^i dW_t, \quad i = 0 \text{ o } 0.5,$$

$$\text{BDT} : dr_t = (\alpha_1(t)r_t + \alpha_2(t)r_t \log r_t)dt + \beta_0(t)r_t dW_t,$$

$$\text{BK} : dr_t = (\alpha_1(t)r_t + \alpha_2(t)r_t \log r_t)dt + \beta_0(t)r_t dW_t, \quad \text{con } \alpha_2(t) = \frac{d \log \beta_0(t)}{dt}.$$

Todos estos últimos pueden condensarse en la expresión:

$$dr_t = (\alpha_0(t) + \alpha_1(t)g(r_t))dt + \beta_0(t)(h(r_t))^{\beta_1(t)} dW_t, \quad (2.82)$$

la cual a su vez representa un caso particular del modelo (2.79). Aquí  $\alpha_0, \alpha_1, g, \beta_0, \beta_1$  representan funciones genéricas.

Una notable distinción entre los modelos (2.79) y (2.82) es la posibilidad de estimar el drift y la difusión en este último. Para el modelo genérico dado por la ecuación (2.79) el drift y la difusión son inestimables, desde el punto de vista teórico, dado que solo es observable una trayectoria en el tiempo y en el espacio de estados.

Una subclase útil del modelo (2.82) es la dada por:

$$dr_t = (\alpha_0(t) + \alpha_1(t)r_t)dt + \beta_0(t)(r_t)^{\beta_1(t)} dW_t, \quad (2.83)$$

extracción de tipo no homogéneo del modelo CKLS y estudiado por Fan et al. (2003). Aquí nos centraremos en este modelo.

### 2.2.13 Metodología de Estimación

En esta sección presentamos la metodología propuesta por Fan et al. (2003) para estimar las funciones de retornos instantáneos (drift) y de volatilidad (difusión) del modelo (2.83). Se asume que los coeficientes del modelo (2.83) son dos veces continuamente diferenciables y que los datos son observados en tiempos discretos, es decir,

$$\{r_{t_i}, i = 1, \dots, n+1\} \quad \text{en los instantes } t_1 < \dots < t_{n+1},$$

igualmente espaciados como suele ser común en muchas aplicaciones. Por ejemplo, si la unidad es el año, los datos semanales se toman en los instantes  $t_i = t_0 + \frac{i}{52}, i = 1, \dots, n+1$ , a partir de un instante inicial  $t_0$ . Sea

$$Y_{t_i} = r_{t_{i+1}} - r_{t_i}, \quad Z_{t_i} = W_{t_{i+1}} - W_{t_i} \quad \text{y} \quad \Delta_{t_i} = t_{i+1} - t_i.$$

debido a las propiedades del movimiento Browniano, las variables  $Z_{t_i}$  son independientes y normalmente distribuidos con media cero y varianza  $\Delta_{t_i}$ , de manera que la versión discretizada del modelo (2.83) se expresa como

$$Y_{t_i} \approx (\alpha_0(t_i) + \alpha_1(t_i)r_t i)\Delta_{t_i} + \beta_0(t_i)r_t i^{\beta_1(t_i)}\sqrt{\Delta_{t_i}}\varepsilon_{t_i}, \quad i = 1, \dots, n \quad (2.84)$$

donde  $\{\varepsilon_{t_i}\}_{i=1}^n$  son i.i.d  $N(0, 1)$ .

Como se señala en Chan et al. (1992) y se demuestra en Stanton (1997), la aproximación discretizada (2.84) es de segundo orden respecto al modelo continuo (2.83) cuando los datos se observan en un corto período de tiempo. En efecto, como apunta Stanton (1997) para datos observados con frecuencia mensual, o bien, de forma más frecuente los errores numéricos de la discretización son extremadamente pequeños comparados con los errores estadísticos.

### Estimación de la función drift

Para la estimación de la función drift  $\mu(t, r_t)$  se considera la versión discretizada (2.84) y se estiman  $\alpha_i(t)$ ,  $i = 0, 1, \dots$  para cada punto  $t$ , siguiendo las ideas de la regresión no-paramétrica lineal local. Es decir,  $\alpha_i(t) \simeq \alpha_i(t_0)$  para  $t$  en un entorno de  $t_0$  y de este modo si  $h$  es el parámetro de ventana y  $K$  la función núcleo (ver Fan y Gijbels (1996)) entonces se minimiza:

$$\psi(a, b) = \sum_{i=1}^n \left( \frac{Y_{t_i}}{\Delta_{t_i}} - a - br_{t_i} \right)^2 K_h(t_i - t_0), \quad (2.85)$$

en donde  $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$ , (Fan et al. (2003), sugieren usar la función núcleo de Epanechnikov,  $K(t) = \frac{3}{4}(1 - t^2)$ ), resultando  $\hat{\alpha}_0(t_0) = \hat{a}$  y  $\hat{\alpha}_1(t_0) = \hat{b}$ . Con lo cual la función de drift queda expresada para cada  $t$  como:

$$\hat{\mu}(t, r_t) = \hat{\alpha}_0(t) + \hat{\alpha}_1(t)r_t.$$

Es de notar que, se ha ignorado la heterocedasticidad en (2.85), sin embargo, esta se puede incorporar minimizando

$$\psi(a, b) = \sum_{i=1}^n \frac{\left( \frac{Y_{t_i}}{\Delta_{t_i}} - a - br_{t_i} \right)^2 K_h(t_i - t_0)}{\hat{\beta}_0^2(t_i)r_t i^{2\hat{\beta}_1(t_i)}}, \quad (2.86)$$

donde  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son obtenidos mediante el procedimiento que describe a continuación.

### Estimación de la función de volatilidad (difusión).

Para la estimación de la función de volatilidad o difusión se considera el estimador

$$\hat{\mu}(t, r_t) = \hat{\alpha}_0(t) + \hat{\alpha}_1(t)r_t,$$

de la función de drift. Entonces se construye el residuo

$$\hat{E}_t = \frac{Y_t - \hat{\mu}(t, r_t)\Delta_t}{\sqrt{\Delta_t}},$$

por (2.84) se tiene que:  $\hat{E}_t \simeq \beta_0(t)r_t^{\beta_1(t)}\varepsilon_t$ . El logaritmo de la verosimilitud condicional de  $\hat{E}_t$  dado  $r_t$ , salvo un término constante, viene dado por

$$-\frac{1}{2} \log \left( \beta_0^2(t)r_t^{2\beta_1(t)} \right) - \frac{\hat{E}_t^2}{2\beta_0^2(t)r_t^{2\beta_1(t)}}.$$



El logaritmo de la verosimilitud local en  $t = t_0$  es por tanto

$$\ell(a, b) = -\frac{1}{2} \sum_{i=1}^n \left( \log(a^2 r_t^{2b}) + \frac{\hat{E}_t^2}{a^2 r_t^{2b}} \right) K_h(t_i - t_0). \quad (2.87)$$

Se maximiza (2.87) y se obtienen los estimadores  $\hat{a} = \hat{\beta}_0(t_0)$  y  $\hat{b} = \hat{\beta}_1(t_0)$ . Por lo tanto la función de volatilidad queda expresada para cada  $t$  como:

$$\sigma^2(t, r_t) = \hat{\beta}_0(t) r_t^{\hat{\beta}_1(t)}.$$

Por otro lado, obsérvese que dado  $\beta_1$  la maximización de  $\ell(a, b)$  se obtiene

$$\hat{a}^2 = \frac{\sum_{i=1}^n K_h(t_i - t_0) \hat{E}_{t_i}^2 |r_{t_i}|^{-2\beta_1}}{\sum_{i=1}^n K_h(t_i - t_0)}.$$

De este modo en cada punto  $t_0$  solo se necesita maximizar una función unidimensional  $\ell(\hat{a}, b)$  respecto de  $b = \beta_1$ .

### Elección del parámetro de suavizado

La elección del parámetro ventana  $h$  es crucial en la estimación no paramétrica tipo núcleo. Se puede elegir por inspección visual equilibrando el sesgo y la varianza o por contrapartida utilizando algún criterio clásico. Eligiendo un núcleo de una sola cara, solo los datos históricos se utilizan para la contribución de los estimadores, y en ese caso un criterio viene dado por el error de predicción medio:

$$APE(h) = \frac{1}{m} \sum_{i=1}^m \frac{(Y_{t_i^*} - \hat{Y}_{t_i^*})^2}{\sigma_{t_i^*}^2},$$

en donde

$$\hat{Y}_{t_i^*} = (\hat{\alpha}_0(t_i^*) + \hat{\alpha}_1(t_i^*)) \Delta_{t_i^*} \quad \text{y} \quad \sigma_{t_i^*} = \hat{\beta}_0(t_i^*) r_{t_i^*}^{\hat{\beta}_1(t_i^*)},$$

con todos los estimadores no paramétricos construidos con núcleos evaluados en los intervalos  $[t_i^* - h, t_i^*]$ ,  $i = 1, \dots, m$  (ver Hart (1994, 1996)).

Para la estimación de la volatilidad, ya que se usa la pseudo verosimilitud local para construir los estimadores no paramétricos, el ancho de banda se escoge maximizando la pseudo verosimilitud de  $\hat{E}_t$  dado  $r_t$ . Más específicamente, para un ancho de banda  $h$  dado la función de pseudo verosimilitud se define como

$$-\frac{1}{2} \sum_{i=1}^n \left( \log(\hat{\beta}_0^2(t_i^*) r_{t_i^*}^{2\hat{\beta}_1(t_i^*)}) + \frac{\hat{E}_{t_i^*}^2}{\hat{\beta}_0^2(t_i^*) r_{t_i^*}^{2\hat{\beta}_1(t_i^*)}} \right)$$

### Aproximación Bootstrap

Formulas para los errores estándar de los estimadores del drift y la difusión son conocidos en el contexto de datos independientes. Ver por ejemplo Fan y Gijbels (1996). Para el caso de datos dependientes, una posibilidad es el uso de bootstrap, por ejemplo del tipo del introducido por Franke et al. (2002). La idea consiste en generar respuestas bootstrap  $Y_{t_i}^*$  en el modelo (2.84) usando los  $r_{t_i}$ , las funciones estimadas y nuevos "shocks" impulsos aleatorios  $\varepsilon_{t_i}^*$ . Con la nueva muestra bootstrap  $\{(r_{t_i}, Y_{t_i}^*)\}_{i=1}^n$  se estiman las funciones de drift y difusión, y esto se replica un número  $B$  de veces, de modo que la variabilidad muestral puede ser analizada.

## 2.3 Modelos Alternativos para el Tipo de Interés

En esta sección se presentarán algunos modelos alternativos para la caracterización del tipo de interés. Tales modelos se consideran en tiempo discreto, y su objetivo es resolver los inconvenientes que se suscitan a la hora de obtener las soluciones de la ecuación diferencial estocásticas, las cuales no siempre es posible obtenerlas de manera explícita. Los modelos que se expondrán a continuación están relacionados con los modelos de series de tiempo expuestos en el primer capítulo, por lo que daremos detalles de estos siempre que sea necesario, puesto que para efectos de esta memoria los modelos que son considerados de principal interés son los de difusión presentados en a sección anterior. En lo que sigue, a menos que se especifique lo contrario, en aras de simplificar la exposición, el modelo de tipos de interés tomado como base para la exposición de los modelos alternativos que se describen a continuación será una versión discretizada del modelo *CKLS*.

### 2.3.1 Modelos *GARCH* para tipos de Interés

En las secciones anteriores hemos visto que, por lo general, para trabajar con los modelos de difusión en la práctica se suele considerar una versión discretizada de estos. Pues teniendo en cuenta dicha forma de proceder, la modelización *GARCH* de los tipos de interés a corto plazo se introduce con el fin de mejorar las deficiencias de los modelos discretizados de un único factor en la caracterización de la volatilidad. Así, [Brenner et al. \(1996\)](#) introduce varias especificaciones *GARCH* para la volatilidad del modelo de difusión discretizado al considerar la varianza del componente gaussiano como un proceso *GARCH* a efectos de capturar el agrupamiento de la volatilidad. Esto puede llevarse a cabo de tres maneras distintas:

- (a) Partiendo de la especificación para el modelo de tipos de interés sin efecto nivel en la volatilidad  $\gamma = 0$ :

$$\Delta r_t = \alpha + \beta r_{t-1} + \sigma_t \varepsilon_t, \quad (2.88)$$

donde  $\{\varepsilon_t\}$  son variables aleatorias independientes e idénticamente distribuidas  $N(0, 1)$ , y  $\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \vartheta_{t-1}^2$ . De este modo, el modelo tiene un residuo  $\vartheta_t$  con media cero y varianza  $\mathbb{E}(\vartheta_t^2) = h_t = \sigma_t^2$ .

- (b) Considerando la presencia del efecto nivel en la volatilidad:

$$\Delta r_t = \alpha + \beta r_{t-1} + \sigma_t \varepsilon_t, \quad (2.89)$$

donde  $\{\varepsilon_t\}$  son variables aleatorias independientes e idénticamente distribuidas  $N(0, 1)$ , y  $\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \vartheta_{t-1}^2 + \beta_3 r_{t-1}^\gamma$ . Así, el modelo tiene un residuo  $\vartheta_t$  con media cero y varianza  $\mathbb{E}(\vartheta_t^2) = h_t = \sigma_t^2$  donde el parámetro  $\beta_3$  mide el impacto del nivel del tipo de interés en la volatilidad.

- (c) Considerando la presencia del efecto nivel en la volatilidad:

$$\Delta r_t = \alpha + \beta r_{t-1} + \sigma_t r_{t-1}^\gamma \varepsilon_t, \quad (2.90)$$

donde  $\{\varepsilon_t\}$  son variables aleatorias independientes e idénticamente distribuidas  $N(0, 1)$ , y  $\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \vartheta_{t-1}^2$ . Bajo esta suposición, el modelo tiene un residuo  $\vartheta_t$  con media cero y varianza  $\mathbb{E}(\vartheta_t^2) = h_t = \sigma_t^2 r_{t-1}^{2\gamma}$ . Con esta especificación *GARCH* la volatilidad condicionada de los tipos de interés se puede expresar como:

$$h_t = \beta_0 r_{t-1}^{2\gamma} + \beta_1 h_{t-1} \left( \frac{r_{t-1}}{r_{t-2}} \right)^{2\gamma} + \beta_2 r_{t-1}^{2\gamma} \vartheta_{t-1}^2. \quad (2.91)$$

Así, el tipo de interés es un componente de la volatilidad condicional de sus variaciones, pero de forma distinta al apartado (b). Ciertamente la especificación  $GARCH(1, 1)$  es el caso más sencillo, sin embargo, la extensión a las distintas especificaciones existentes en la literatura se puede hacer de forma inmediata. Entre las múltiples extensiones que ofrecen los modelos  $GARCH$ , y que fueron discutidos en el capítulo 1 de esta memoria, para caracterizar los tipos de interés, podemos considerar las siguientes:

1. Modelos que permiten asimetrías en la especificación de la volatilidad, de modo que:

$$h_t = \beta_0 r_{t-1}^{2\gamma} + \beta_1 h_{t-1} \left( \frac{r_{t-1}}{r_{t-2}} \right)^{2\gamma} + \beta_2 r_{t-1}^{2\gamma} \vartheta_{t-1}^2 + \beta_3 r_{t-1}^{2\gamma} \xi_{t-1}^2 \quad (2.92)$$

donde  $\xi_t^2 = \min\{\vartheta_t, 0\}$ . En este modelo se asume que el impacto de  $\varepsilon_t^2$  sobre la varianza condicional es distinto cuando  $\varepsilon_t$  es positivo o negativo. En este caso se utiliza el modelo de [Glosten et al. \(1993\)](#) para caracterizar la volatilidad, puesto que, este modelo permite recoger el efecto apalancamiento observable en las series financieras.

2. El Modelo  $EGARCH(1, 1)$  propuesto por [Nelson \(1991\)](#):

$$\log(\sigma_t^2) = \kappa + \phi_1 \log(\sigma_{t-1}^2) + \theta_1 \{|\varepsilon_{t-1}| - \mathbb{E}|\varepsilon_{t-1}| + \Psi\varepsilon_{t-1}\} \quad (2.93)$$

donde  $\{\varepsilon_t\}$  son variables aleatorias independientes e idénticamente distribuidas  $N(0, 1)$  y  $\mathbb{E}|\varepsilon_{t-1}| = \sqrt{2/\pi}$ . De este modo, el modelo tiene como residuo a  $\vartheta$  con media cero y varianza

$$\mathbb{E}(\vartheta_t^2) = h_t = \sigma_t^2 r_{t-1}^{2\gamma},$$

con

$$\sigma_t^2 = (\sigma_{t-1}^2)^{\phi_1} e^{(\kappa + \theta_1 \{|\varepsilon_{t-1}| - \mathbb{E}|\varepsilon_{t-1}| + \Psi\varepsilon_{t-1}\})}.$$

De donde se obtiene que

$$h_t = h_{t-1}^{\phi_1} \left( \frac{r_{t-1}^{2\gamma}}{r_{t-2}^{2\gamma\phi_1}} \right) e^{(\kappa + \theta_1 \{|\varepsilon_{t-1}| - \mathbb{E}|\varepsilon_{t-1}| + \Psi\varepsilon_{t-1}\})}.$$

3. Modelo  $ARCH$  con *umbral*, permite una respuesta simétrica cuyo efecto cuantitativo depende del signo del residuo de cada período:

$$\sum_{j=1}^p \phi_j \sigma_{t-j}^{2\lambda} + \sum_{i=1}^m [\theta_i^+ \mathbb{1}_{\varepsilon_{t-i} > 0} |\varepsilon_{t-i}|^{2\lambda} + \theta_i^- \mathbb{1}_{\varepsilon_{t-i} \leq 0} |\varepsilon_{t-i}|^{2\lambda}], \quad (2.94)$$

donde  $\mathbb{1}$ . es una función indicador. De manera que, el residuo del modelo  $\vartheta_t$  tiene media cero y varianza  $\mathbb{E}(\vartheta_t^2) = h_t = \sigma_{t-1}^{2\gamma}$ . Dependiendo del valor del parámetro  $\lambda$  se obtienen diferentes especificaciones para la volatilidad, por ejemplo, para  $\lambda = 1$  se obtiene el modelo de [Glosten et al. \(1993\)](#).

4. Modelo de [Longstaff y Schwartz \(1992\)](#), utilizan un modelo de dos factores para la media y la volatilidad:

$$\begin{aligned} \Delta r_t &= r_t - r_{t-1} = \alpha + \beta r_{t-1} + \theta h_{t-1} + \varepsilon_t, \\ h_t &= \beta_0 + \beta_1 h_{t-1} + \beta_2 \varepsilon_{t-1}^2 + \beta_3 r_{t-1}, \end{aligned}$$

En realidad, este modelo no es otro que el modelo  $GARCH$  en media en el que se considera una variable exógena en la especificación de la volatilidad. En cuanto a la predicción y la estimación de este tipo de modelos es similar a la de los modelos  $GARCH$  usuales.

### 2.3.2 Modelo de Volatilidad Estocástica

Tal y como se discutió en el capítulo 1 de esta memoria, este modelo permite que el logaritmo de la volatilidad evoluciones a lo largo del tiempo. La aplicación de este modelo a los tipos de interés constituye una aproximación discreta a los modelos de tiempo continuo, concretamente al modelo de difusión descrito en (2.51) usado para caracterizar la dinámica de los tipos de interés. La especificación asociada al modelo *CKLS* es la siguiente:

$$\begin{aligned}\Delta r_t &= r_t - r_{t-1} = \alpha + \beta r_{t-1} + \sigma_t r_{t-1}^\gamma \varepsilon_t, \\ \log(\sigma_t^2) &= a + b \log(\sigma_{t-1}^2) + \eta_t,\end{aligned}\tag{2.95}$$

donde  $\varepsilon_t$  son variables aleatorias normal estándar independientes e idénticamente distribuidas, y  $\eta_t$  son variables aleatorias independientes e idénticamente distribuidas con media cero y varianza  $\sigma_\eta^2$ . Esto permite que la volatilidad este sujeta a impulsos aleatorios o factores no predecibles como decisiones políticas, el sentir de los inversores, cambios de estrategias de las empresas, etc. La estimación del modelo de volatilidad estocástica es una tarea compleja, para más detalles de este procedimiento ver [Harvey et al. \(1994\)](#), quien propone un método que hace uso de una transformación del residuo y en el que aplica el potencial del Filtro de Kalman en la estimación del modelo.

### 2.3.3 Modelo Markov-Switching

En el primer capítulo se estudió este modelo, y se describieron algunas de sus propiedades. Este modelo ha resultado ser de gran utilidad en la modelización de diversas series económicas. Básicamente es un modelo que permite caracterizar los cambios de régimen. En cuanto a las series de tipos de interés [Cai \(1994\)](#), [Ball y Torous \(1998\)](#) y [Ang y Bekaert \(2002\)](#) proponen modelos alternativos de Markov-Switching para el estudio de tales series.

La evolución de los tipos de interés al contado a lo largo del tiempo puede verse afectada por una serie de factores como cambios en las decisiones de política monetaria, el ciclo económico o por las condiciones macroeconómicas en general. Por ello, los modelos de Markov-Switching pueden ser de utilidad en la modelización de dicha dinámica. En este contexto, la introducción del modelo Markov Switching requiere una aproximación discreta a los modelos de difusión utilizados habitualmente. Tomando como punto de partida tal aproximación se introduce la posibilidad que los parámetros del modelo puedan tomar valores diferentes dependiendo del estado en el que se encuentren, determinándose de esta manera la dinámica de la serie de tipos de interés. Por ejemplo, si se consideran solo dos estados, en los que se permite a la volatilidad no condicionada y el valor esperado de la variación del tipo de interés cambien dependiendo del régimen en el que se encuentren, el modelo de difusión discretizado tendría la forma siguiente:

$$\Delta r_t = \alpha_{S_t} + \beta_{S_t} r_{t-1} + \sigma_{S_t} r_{t-1}^\gamma \varepsilon_t\tag{2.96}$$

donde  $\varepsilon_t$  son variables aleatorias normal estándar independientes, el subíndice  $S_t$  es una variable aleatoria, referida al estado o régimen del proceso en el instante  $t$ , que toma el valor 1 si se encuentra en el primer estado o bien 2 si se encuentra en el segundo estado. En este caso el comportamiento de  $S_t$  debe ser recogido por un modelo probabilístico, dado que el régimen es no observable, de manera similar a como se planteó en el primer capítulo, o bien considerando algún modelo probabilístico particular (ver [Hamilton \(1990\)](#)).

### 2.3.4 Modelo de Volatilidad Estocástica con Markov-Switching

[So et al. \(1998\)](#) y [Smith \(2002\)](#), proponen este modelo como una generalización del modelo de volatilidad estocástica y del de Markov-Switching. La especificación es muy similar a la del

modelo de volatilidad estocástica. Para

$$\Delta r_t = \alpha + \beta r_{t-1} + \sigma_{S_t} r_{t-1}^\gamma \varepsilon_t$$

Sea  $e_t = \Delta r_t - \alpha - \beta r_{t-1} = \sigma_{S_t} r_{t-1}^\gamma \varepsilon_t$  y  $y_t = \log(e_t^2) = \log(\sigma_{S_t}^2) + 2\gamma \log(r_{t-1}) + \log(\varepsilon_t^2)$ . entonces el modelo queda expresado por

$$y_t = x_t + 2\gamma \log(r_{t-1}) + \log(\varepsilon_t^2) \quad (2.97)$$

$$x_t = a_{S_t} + b_{S_t} x_{t-1} + \eta_t \quad (2.98)$$

donde  $x_t = \log(\sigma_{S_t}^2)$ , y en este caso los parámetros  $a_{S_t}$  y  $b_{S_t}$  pueden tomar valores distintos dependiendo del régimen en el que se esté. Considerando solo dos estados para la volatilidad, se tiene que si  $a_1 = a_2$ , y  $b_1 = b_2$  se estará en presencia del modelo de volatilidad estocástica y las probabilidades de transición no son identificadas, mientras que si  $\sigma_\eta^2 = 0$ , entonces se está en el Modelo Markov-Switching y el parámetro  $b$  no es identificado.

### 2.3.5 Modelos de Difusión con Saltos

Ciertamente el modelo de difusión con saltos fue introducido en el capítulo 1. En esta sección, se hace una descripción basada en los Das (2002) y Johanes (2004) quienes introducen los saltos en el modelo de difusión con la finalidad de caracterizar la dinámica de los tipos de interés, motivados principalmente por el hecho de que los impulsos “shocks” económicos, intervenciones del gobierno en el mercado, la difusión de noticias, entre otros factores, causan saltos elevados en los tipos de interés, lo que genera discontinuidades en su dinámica.

La especificación más general que suele ser presentada en la literatura del modelo de difusión con saltos, viene dado por la ecuación diferencial estocástica,

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t + J(\lambda, \rho^2)d\pi_t(q_t)$$

Así pues, los tipos de interés evolucionan con una función drift,  $\mu(r_t, \theta)$ , conocida, con  $\theta$  su parámetro, el cuál si es desconocido se puede estimar, y dos términos aleatorios, la función de volatilidad  $\sigma(r_t, \theta)$ , conocida, (con  $W_t$  un movimiento Browniano estándar) y un proceso Poisson  $\pi_t$  con frecuencia de sucesos  $q_t$ , incluyendo un salto aleatorio  $J(\lambda, \rho^2)$  con distribución  $N(\lambda, \rho^2)$ .

Por otra parte, teniendo en cuenta, la especificación propuesta por Das (2002), en base a la versión discreta del modelo CKLS y el modelo (2.3.5), se tiene que la variación del tipo de interés se puede modelizar por

$$\Delta r_t = \alpha + \beta r_t + \sigma r_t^\gamma \varepsilon_t + J(\lambda, \rho^2)\Delta\pi_t(q_t)$$

Así pues, los tipos de interés evolucionan con una función drift con reversión a la media y dos términos aleatorios, la función de volatilidad o difusión  $\sigma(r_t, \theta) = \sigma r_t^\gamma \varepsilon_t$ , con  $\{\varepsilon_t\}$  variables aleatorias independientes con distribución  $N(0, 1)$  e independientes de  $\{r_t\}$ ;  $\pi_t$  una variable aleatoria independiente con distribución de Bernoulli de parámetros  $a$  y  $b$ .

$$q_t = \frac{1}{1 + \exp(-a - br_{t-1})} \quad (2.99)$$

y con tamaño de salto el valor de la variable aleatoria  $J(\lambda, \rho^2)$  con distribución  $N(\lambda, \rho^2)$ . La media condicionada para la variación en el tipo de interés es:

$$\mathbf{E}(\Delta r_t | r_{t-1}) = \alpha + \beta r_{t-1} + q_t \lambda \quad (2.100)$$

mientras que la varianza

$$\mathbf{E}[(\Delta r_t)^2 | r_{t-1}] = \sigma^2 r_{t-1}^2 + q_t (\rho^2 + (1 - q_t)\lambda^2) \quad (2.101)$$

## 2.4 Series de Tipos de Interés *EURIBOR*

En esta sección se presentan los rasgos más importantes de la serie de tipos de interés interbancario que serán estudiadas en el próximo capítulo y a lo largo de esta memoria, las series del *EURIBOR*, (European Interbank Offered Rate, es decir, tipo europeo de oferta interbancaria) para distintos plazos de vencimiento: 1, 2, 3 semanas; 1 mes, 2 meses, ..., 12 meses.

El *EURIBOR*, es un índice de referencia publicado diariamente que indica el tipo de interés promedio al que las entidades financieras intercambian dinero en el mercado interbancario del euro. Se calcula usando los datos de los principales bancos europeos, y su valor mensual es usado como referencia para los préstamos bancarios. El panel de bancos que determina el valor de los tipos de interés se compone exclusivamente de bancos con una excelente valoración crediticia. La elección de los bancos a los que se solicita que remitan sus datos para la determinación del *EURIBOR* se controla por parte de una comisión de supervisión de la Federación de Bancos Europeos. Este índice de referencia varía según el plazo de vencimiento al que se efectúa el préstamo, por ejemplo: con plazo a una semana, a un mes o a un año. Por otro lado, el *EURIBOR* es considerado el tipo básico de referencia para todo tipo de productos (derivados) de renta como *futuros*, *swaps* y en acuerdos sobre tipo de interés *futuros*. Se emplea además con frecuencia como tipo de referencia para el cálculo de hipotecas e imposiciones a plazo. El *EURIBOR* a un año (12 meses) es el que se usa normalmente como referencia para las hipotecas. Este índice es de aplicación exclusiva en los bancos de los Estados miembros de la Unión Europea que forman parte de la denominada *Eurozona*.

En España, el *EURIBOR* sustituyó al *MIBOR* a raíz del traspaso de competencias del Banco de España al Banco Central Europeo. 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. El valor del *EURIBOR* sufrió un importante descenso a lo largo del año 2009, tomando sus valores más bajos desde su creación.

Otro índice de referencia ampliamente conocido es el *LIBOR*, mencionada al inicio de este capítulo, y está relacionado con el mercado interbancario londinense. El *EURIBOR* y *LIBOR* son tasas comparables. Como ya se ha mencionado, el primero es el tipo de interés medio al que una selección de bancos se otorgan préstamos a corto plazo no cubiertos en euros. Mientras que el segundo, es el tipo de interés interbancario medio al que una selección de bancos se otorgan préstamos a corto plazo no cubiertos en el mercado monetario londinense. Ambos se publican para 15 vencimientos. La diferencia radica en que *LIBOR* se calcula para 10 divisas diferentes.

El cálculo y publicación del valor del *EURIBOR* se realiza todos los días laborables por la agencia *Reuters*, de acuerdo al siguiente proceso:

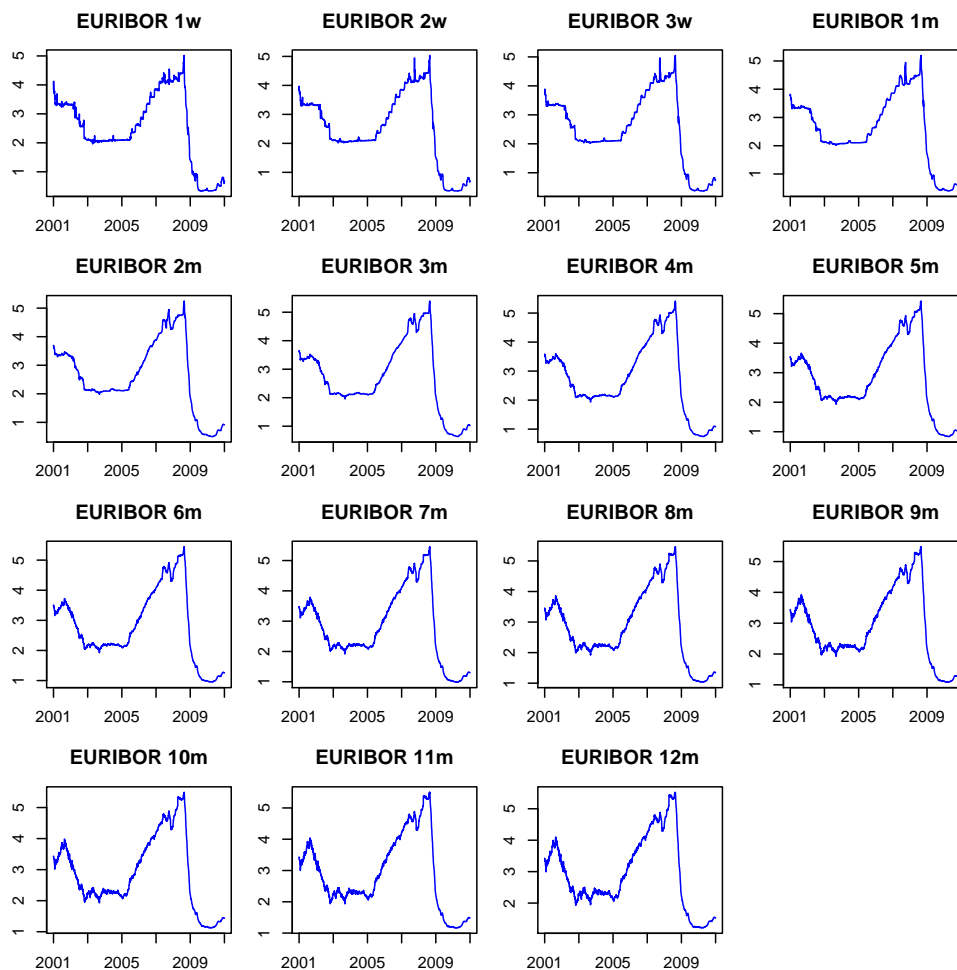
- Cada día hábil del mercado se solicita a cada banco de referencia que envíe sus tipos de interés actuales (a más tardar a las 10:45 a.m.). *Reuters* genera en su sistema una página privada que solamente puede visitar el banco y el personal encargado de realizar el cálculo. En un período que abarca entre las 10:45 a.m. y 11:00 a.m. los bancos tienen oportunidad de corregir sus contribuciones, si es necesario.
- A las 11:00 a.m. *Reuters* realiza el cálculo del nuevo valor del tipo de interés, para lo cuál elimina el 15 % más alto y el 15 % más bajo de los tipos de interés suministrado por las entidades designadas, y realiza la media aritmética del resto de valores.
- Tras el cálculo, *Reuters* publica el tipo de referencia *EURIBOR*.

La figura (2.1) muestra los gráficos del nivel diario de la serie de interés del *EURIBOR* para distintos plazos de vencimiento: 1, 2, 3 semanas; 1 mes, 2 meses, ..., 12 meses en el período tem-

poral que abarca desde el 15/10/2001 hasta el 03/12/2010, con frecuencia diaria. Las series son obtenidas de:

[www.euribor.org/html/content/euribor\\_about.html](http://www.euribor.org/html/content/euribor_about.html),  
EURIBOR® (Euro Interbank Offered Rate).

De los gráficos de la figura (2.1) se puede destacar que, los valores del tipo de interés a corto plazo permanecen largos períodos por encima o debajo de su valor medio de largo plazo, lo cual indica una fuerte persistencia en las series, para cada plazo de vencimiento. Por otro lado, se aprecian cambios abruptos en el tipo de interés, los cuales suelen ser más abruptos para las series con plazo de vencimiento más breves y que tiende a suavizarse a medida que consideramos plazos más amplios. El tabla (2.1) muestra la estructura de correlaciones entre los diferentes plazos conside-



**Figura 2.1:** Evolución diaria del EURIBOR a distinto plazos en el período del 15/10/2001 al 03/12/2010, (valores porcentuales).

rados. En ella se observa una fuerte correlación entre el valor del tipo de interés para los distintos plazos, lo cual sugiere que las variaciones en los tipos están determinados por un número limitado de factores comunes.

Ahora bien, ciertamente el tipo de interés puede aportar información útil, sin embargo son las variaciones las que aportan información relevante del comportamiento y dinámica de los tipos de interés conforme al tiempo. La tabla (2.2) recoge diferentes estadísticos descriptivos que



**Tabla 2.1:** Correlación entre los tipos de interés a diferentes plazos

	1 s.	2 s.	3 s.	1 m.	2 m.	3 m.	4 m.	5 m.	6 m.	7 m.	8 m.	9 m.	10 m.	11 m.	12 m.
1 s.	1.000	0.999	0.997	0.994	0.988	0.980	0.976	0.972	0.967	0.964	0.961	0.958	0.956	0.953	0.951
2 s.	0.999	1.000	0.999	0.997	0.991	0.984	0.980	0.976	0.971	0.969	0.966	0.963	0.961	0.959	0.956
3 s.	0.997	0.999	1.000	0.998	0.994	0.987	0.984	0.980	0.976	0.973	0.970	0.968	0.965	0.963	0.961
1 m.	0.994	0.997	0.998	1.000	0.997	0.992	0.989	0.985	0.981	0.979	0.976	0.973	0.971	0.969	0.966
2 m.	0.988	0.991	0.994	0.997	1.000	0.998	0.996	0.994	0.991	0.989	0.987	0.984	0.982	0.980	0.978
3 m.	0.980	0.984	0.987	0.992	0.998	1.000	0.999	0.998	0.996	0.995	0.993	0.991	0.989	0.988	0.986
4 m.	0.976	0.980	0.984	0.989	0.996	0.999	1.000	0.999	0.998	0.997	0.996	0.994	0.993	0.991	0.990
5 m.	0.972	0.976	0.980	0.985	0.994	0.998	0.999	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.993
6 m.	0.967	0.971	0.976	0.981	0.991	0.996	0.998	0.999	1.000	0.999	0.999	0.998	0.998	0.997	0.996
7 m.	0.964	0.969	0.973	0.979	0.989	0.995	0.997	0.999	0.999	1.000	0.999	0.999	0.998	0.998	0.997
8 m.	0.961	0.966	0.970	0.976	0.987	0.993	0.996	0.998	0.999	0.999	1.000	0.999	0.999	0.999	0.998
9 m.	0.958	0.963	0.968	0.973	0.984	0.991	0.994	0.997	0.998	0.999	0.999	1.000	0.999	0.999	0.999
10 m.	0.956	0.961	0.965	0.971	0.982	0.989	0.993	0.996	0.998	0.998	0.999	0.999	1.000	0.999	0.999
11 m.	0.953	0.959	0.963	0.969	0.980	0.988	0.991	0.994	0.997	0.998	0.999	0.999	0.999	1.000	0.999
12 m.	0.951	0.956	0.961	0.966	0.978	0.986	0.990	0.993	0.996	0.997	0.998	0.999	0.999	0.999	1.000

corroboran que la serie de variaciones en los tipos de interés tiene características muy similares a las series de rentabilidades de otros activos financieros. Se observa entonces que, por ejemplo, que las series de las variaciones son asimétricas, y que dicha asimetría varía respecto del plazo de vencimiento, por otra parte, los valores de curtosis son elevados, indicando que la distribución de probabilidad presenta colas pesadas.

**Tabla 2.2:** Estadísticas Descriptivas para los tipos de interés *EURIBOR*, observaciones diarias, en el período comprendido entre el 15 de Octubre de 2001 y 3 de Diciembre 2010.

Activo	Mín	Máx	Media	Mediana	Varianza	Desv	Asimetría	Curtosis
1 Semana	-0.358	0.530	-0.0014	0	0.0012	0.0354	-0.0635	53.4123
2 Semanas	-0.493	0.800	-0.0013	0	0.0010	0.0310	4.7157	230.0387
3 Semanas	-0.422	0.760	-0.0013	0	0.0007	0.0269	7.4809	306.8073
1 Mes	-0.296	0.640	-0.0013	0	0.0005	0.0225	8.8782	301.8418
2 Meses	-0.187	0.287	-0.0012	0	0.0003	0.0160	0.7089	63.3664
3 Meses	-0.118	0.101	-0.0011	0	0.0002	0.0142	-1.1684	14.3322
4 Meses	-0.114	0.105	-0.0011	0	0.0002	0.0145	-1.0651	11.2880
5 Meses	-0.107	0.139	-0.0010	0	0.0002	0.0155	-0.6468	10.1288
6 Meses	-0.107	0.175	-0.0010	0	0.0003	0.0166	-0.2329	11.1476
7 Meses	-0.108	0.205	-0.0009	0	0.0003	0.0180	0.1519	11.8254
8 Meses	-0.107	0.227	-0.0009	0	0.0004	0.0194	0.3852	12.0498
9 Meses	-0.118	0.246	-0.0009	0	0.0004	0.0209	0.5257	11.8864
10 Meses	-0.129	0.262	-0.0009	0	0.0005	0.0223	0.6484	11.6556
11 Meses	-0.135	0.273	-0.0008	0	0.0006	0.0237	0.6882	10.9686
12 Meses	-0.140	0.293	-0.0008	0	0.0006	0.0248	0.8130	11.5883

La figura (2.2) muestra las variaciones del tipo de interés para los distintos vencimientos, en la que se aprecia el comportamiento dispar de las variaciones del tipo de interés atendiendo a los plazos de vencimiento. Se observa que para plazos cortos, los tipos se comportan prácticamente como un paseo aleatorio, en dichas series también se observan saltos que suelen estar relacionados con decisiones del Banco Central Europeo en cuanto a la política monetaria. Por el contrario, para plazos más largos se observa una volatilidad mayor que varía en el tiempo y se caracteriza por



su persistencia. Por otro lado, se puede observar el efecto de la crisis en las variaciones del tipo de interés. Evidentemente, en los gráficos se observa un punto en el que hay un salto abrupto, y posterior a este la volatilidad tiende a remitir, sin embargo, no ocurre de forma inmediata en todos los casos. Esto puede verse en la serie del cuadrado de las variaciones.

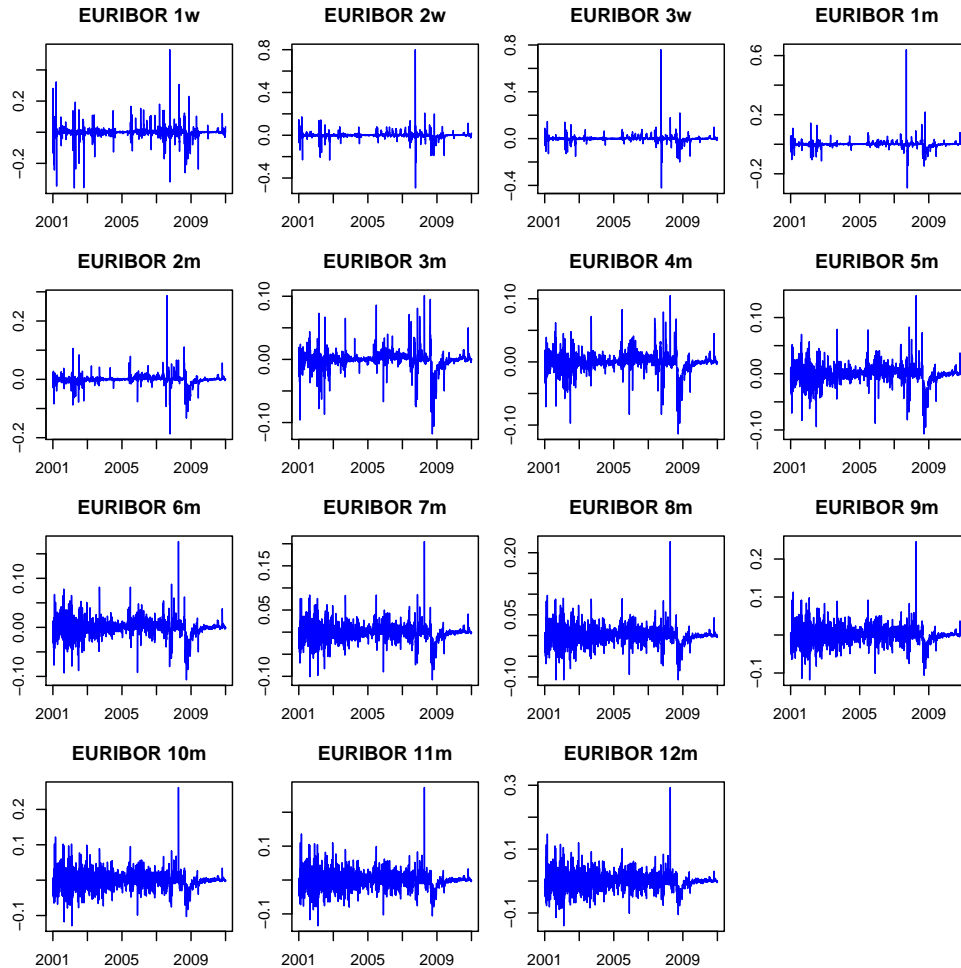


Figura 2.2: Cambios en los tipos de interés,  $r_{t_{i+1}} - r_{t_i}$ , para diferentes plazos.

La figura (2.3) representa la evolución de los cuadrados de la serie,  $(r_{t_{i+1}} - r_{t_i})^2$ ,  $i = 1, \dots, n$ . se observa agrupamiento y persistencia de la volatilidad, y se evidencia heteroscedasticidad. Además, se puede apreciar que la serie de los cuadrados tiene distintos comportamientos dependiendo del plazo de vencimiento. Los gráficos muestran que para plazos cortos se observan saltos discontinuos en las series, por el contrario, para plazos más largos existen claros cambios de régimen con cierta persistencia. Desde el punto de vista de las finanzas, un buen modelo debe recoger todas las características descritas anteriormente a fin de generar un buen ajuste.

A continuación, se estimarán los parámetros de los modelos Vasicek, (CIR) y (CKLS) para las series del EURIBOR correspondientes a los plazos: 1 semana, 1 mes, 3 meses, 6 meses, 9 meses y 12 meses respectivamente para el período comprendido entre el 15/10/2001 al 30/03/2006. El motivo por el cual se escoge este período, es para evitar el período conflictivo pre y post crisis, ya que como es común, el Banco Central Europeo, suele intervenir en estos casos en cuanto a una bajada en los tipos de interés a fin de reactivar la economía.

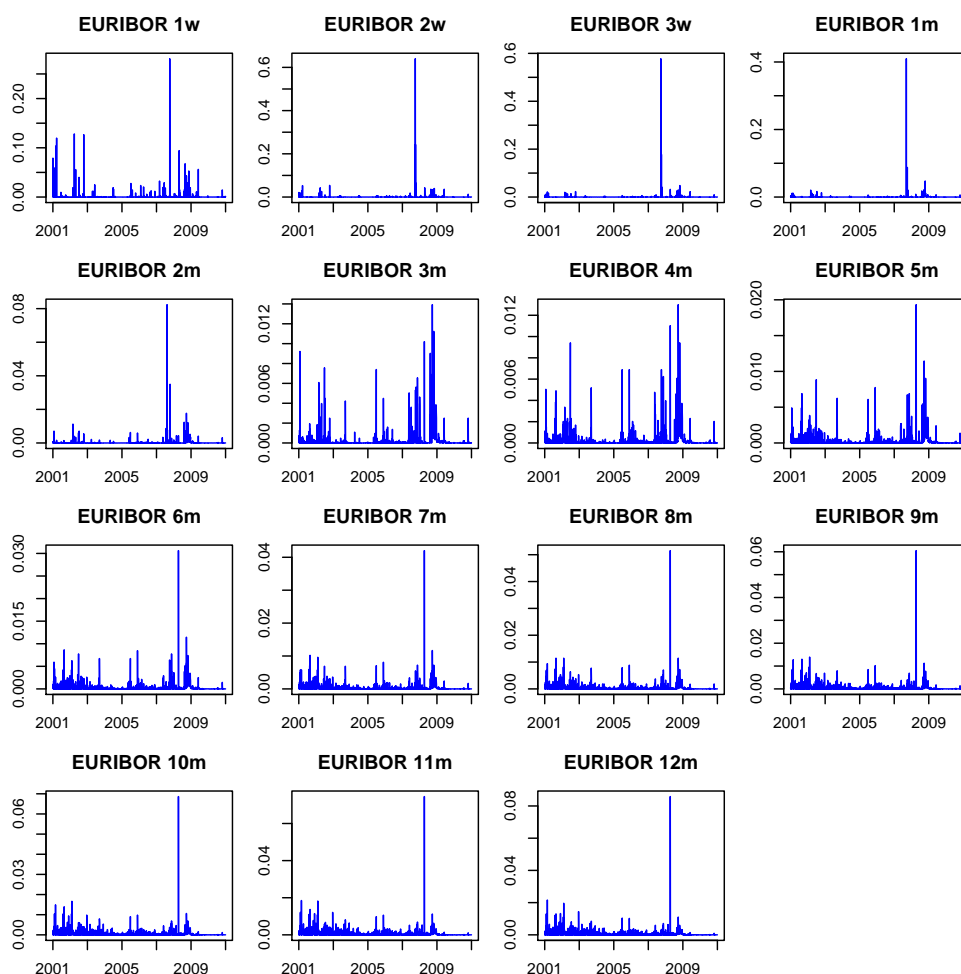


Figura 2.3: Cuadrado de la serie cambios en el tipo de interés.

Se puede observar en las tablas (2.3), (2.4), (2.5) que los valores estimados de los parámetros  $\alpha$ ,  $\beta$  no presentan diferencias significativas, no ocurre así con el parámetro  $\sigma$  el cual sí presenta diferencia significativa entre uno y otro modelo. Por supuesto esto es debido a la capacidad de cada modelo de capturar la dinámica de la serie.

Tabla 2.3: Parámetros estimados para la serie EURIBOR con plazo a 1 semana

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.010223695	-0.004501609	0.033854026	-
CIR	0.009366916	-0.004160321	0.019578015	-
CKLS	0.009137199	-0.004080253	0.002407942	2.628719641

En cuanto a las series para los plazos 6 meses, 9 meses, y 12 meses, se puede ver en las tablas (2.6), (2.7) y (2.8) que los parámetros  $\alpha$  y  $\beta$  comienzan a diferir, sobre todo en el caso del modelo (CKLS), esto según se recoge en, Chan y otros (1992) al particular parámetro  $\gamma$ , el coeficiente de la elasticidad de la volatilidad respecto a  $r$ . Valores de  $\gamma$  superiores a la unidad indican alta sensibilidad de la volatilidad al nivel de los tipos.

**Tabla 2.4:** Parámetros estimados para la serie *EURIBOR* con plazo a 1 mes

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.005839463	-0.002713334	0.0137340924	-
CIR	0.005758077	-0.002680998	0.0081314078	-
CKLS	0.005787660	-0.002695434	0.0009810425	2.693338431

**Tabla 2.5:** Parámetros estimados para la serie *EURIBOR* con plazo a 3 meses

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.004398255	-0.002020095	0.010774603	-
CIR	0.004433385	-0.002033954	0.006483656	-
CKLS	0.004543242	-0.002080191	0.001882727	1.796611391

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.003691582	-0.001622465	0.016007179	-
CIR	0.003783199	-0.001658259	0.009658361	-
CKLS	0.004057763	-0.001771058	0.004276420	1.353563682

**Tabla 2.6:** Parámetros estimados para la serie *EURIBOR* con plazo a 6 meses

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.003877594	-0.001601271	0.02291661	-
CIR	0.004066989	-0.001674124	0.01381373	-
CKLS	0.004531086	-0.001860435	0.00740913	1.147042878

**Tabla 2.7:** Parámetros estimados para la serie *EURIBOR* con plazo a 9 meses

Modelo	Parámetros			
	$\alpha$	$\beta$	$\sigma$	$\gamma$
Vasicek	0.004354022	-0.001704489	0.02850585	-
CIR	0.004707803	-0.001837698	0.01710960	-
CKLS	0.005289640	-0.002065105	0.01054904	0.994670321

**Tabla 2.8:** Parámetros estimados para la serie *EURIBOR* con plazo a 12 meses

## II

# Métodos de Validación

# 3

## Test de Bondad de Ajuste

En el capítulo 1 de esta memoria se presentó una descripción de los modelos en tiempo discreto y en tiempo continuo que suelen ser usados en el estudio de las series de tiempo en el ámbito de la econometría financiera. Posteriormente, el capítulo 2 dedicó su contenido a presentar con cierto detalle los modelos en tiempo continuo, concretamente los procesos de difusión, pero esta vez orientados a la modelización del tipo de interés, tema principal en este trabajo.

La modelización de la dinámica de los tipos de interés constituye actualmente uno de los principales intereses tanto para académicos como para los profesionales de las finanzas puesto que esta permite la determinación de su estructura temporal, y a partir de esta se pueden llevar a cabo la valoración de los precios de activos financieros, la evaluación de riesgos, etc. Los modelos en tiempo continuo o bien procesos de difusión discutidos en el capítulo 2 se presentan como una alternativa, cada vez más popular en las finanzas, para satisfacer tal interés. Por supuesto, dichos modelos deben ser capaces de explicar y/o predecir el patrón de comportamiento de los tipos de interés. En este trabajo se considera una especificación paramétrica (2.22) de los modelos de tipos de interés, principalmente porque dichos modelos tienen la particularidad de que pueden dar una interpretación parsimoniosa del mecanismo que genera los datos del proceso. Sin embargo, tal afirmación sólo es posible siempre que se garantice la especificación correcta del mismo. Cuando se trabaja con modelos paramétricos, siempre existe el riesgo de errores de especificación que conllevan a conclusiones erróneas. Así pues, es de vital importancia validar su adecuación antes de aplicarlo a los datos reales. Teniendo en cuenta esto, suponga que  $\mathcal{M}$  es la característica desconocida de interés detrás del mecanismo que genera los datos. Si se considera el proceso de difusión (2.22) discutido en el capítulo 2, se puede considerar  $\mathcal{M} = (\mu, \sigma)$ . Para una especificación dada  $\mathcal{M}_\theta$  con posible parámetro desconocido  $\theta$ , se quiere probar la hipótesis nula  $H_0 : \mathcal{M} = \mathcal{M}_\theta$ . Este problema a menudo se denomina *test de especificación*, *validación de modelos*, *verificación de modelos*, *bondad de ajuste*, entre otros. Existe una amplia literatura sobre métodos de validación de los distintos modelos discutidos en el capítulo 1, entre los que se pueden destacar: la propuesta presentada por Eubank y Spiegelman (1990) en la que sugieren un test basado en regresión no-paramétrica para probar la adecuación de un modelo lineal paramétrico, el test de razón de pseudo-verosimilitud PLRT propuesto por Azzalini y Bowman (1993), el procedimiento de validación propuesto por Härdle y Mammen (1993) a partir de regresión no-paramétrica de curvas,

un enfoque alternativo al anterior es el introducido por [Hong y White \(1995\)](#) que busca comparar la suma del cuadrado de los residuos del modelo paramétrico con los del modelo no-paramétrico, el test propuesto por [Horowitz y Spokoiny \(2001\)](#) el cuál expone un enfoque en la línea de los dos anteriores, el test de [Fan y Li \(1996\)](#) basado en residuos, el test propuesto por [Zheng \(1996\)](#) para la forma funcional de un modelo de regresión no-lineal, el test de razón de verosimilitud generalizada *GLRT* propuesto por [Fan et al. \(2001\)](#), entre muchos otros. En lo que respecta a los modelos de difusión las propuestas son variadas. Gran parte de estas propuestas, que serán brevemente descritas en este capítulo, son en espíritu adaptaciones de las antes mencionadas.

Este capítulo tiene como objetivo presentar dos tests de bondad de ajuste basados en procesos empíricos, dirigidos a dar una respuesta al problema de especificación de los modelos de difusión, discutidos en el capítulo 2, para caracterizar la dinámica del tipo de interés. En la Sección (3.1) se presenta una breve exposición de los distintos enfoques orientados a resolver el problema antes descrito. En la Sección (3.2) se describen los detalles de la propuesta central de este trabajo de investigación, los tests de bondad de ajuste basados en procesos empíricos para la funciones drift y volatilidad del modelo de tipo de interés. En la sección (3.3) se lleva a cabo un estudio de simulación para mostrar el desempeño de los tests descritos en la sección anterior. Finalmente en la Sección (3.4) se muestra una aplicación a las series de los tipos de interés del *EURIBOR*.

### 3.1 Una breve revisión de los Modelos de Validación

Como ya se ha mencionado, los procesos de difusión son usados cada vez con mayor frecuencia en la modelización de la dinámica de las principales variables económicas. En lo que respecta a esta memoria dicha variable económica no es otra que el tipo de interés. El modelo de difusión en tiempo continuo, base de la propuesta que se desarrollará en éste capítulo no es otro que el presentado en (2.22) y es dado por la ecuación diferencial estocástica

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t.$$

Bajo este tipo de modelo, las propiedades de los tipos de interés son determinados completamente por las funciones drift y volatilidad. En consecuencia, el problema de selección entre los distintos modelos existentes en la literatura se reduce a la elección y/o estimación de dichas funciones. En cuanto a la modelización de los tipos de interés, existe una extensa literatura que se ha desarrollado a lo largo de los últimos años. Sin embargo, la selección del modelo apropiado para un cierto conjunto de datos, es un tema de discusión para el que no hay un consenso. Así pues, el principal objetivo de los investigadores y profesionales de las finanzas se centra en la búsqueda de herramientas que permitan establecer un criterio de selección. Como se ha comentado previamente, entre los profesionales de las finanzas suele haber preferencia por los modelos paramétricos, de manera que cabe preguntarse entonces, si es apropiado utilizar un modelo paramétrico del tipo (2.22) para un determinado conjunto de datos financieros, en particular las series de tipo de interés. Dentro de los estudios recientes que abordan el problema de la especificación de procesos de difusión en modelos econométricos, en particular tipos de interés, tenemos:

- El trabajo pionero presentado por [Aït-Sahalia \(1996b\)](#) quien propone un test para modelos paramétricos mediante la comparación de la densidad paramétrica implícita con la misma densidad estimada no-paramétricamente. El autor destaca que en este caso, el modelo en tiempo continuo no es reemplazado por una aproximación discreta del mismo. Similares propuestas
- [Corradi y White \(1999\)](#) quienes proponen tests de especificación para la varianza de un proceso de difusión, en los que no se requieren del conocimiento de la forma funcional del

modelo bajo la hipótesis nula. Tales tests consideran como hipótesis nula, varianza constante o bien un modelo paramétrico versus una hipótesis alternativa de una forma funcional más general.

- La propuesta presentada por [Dette y von Lieres und Wilkau \(2003\)](#) en la que se introduce un test de especificación para la forma paramétrica de la función de volatilidad en modelos en tiempo continuo en finanzas, el cual no requiere del conocimiento de la forma funcional del modelo. Dicho test está basado en el función de volatilidad integrada.
- [Fan y Zhang \(2003\)](#), [Fan et al. \(2003\)](#), proponen test simultáneos para la especificación de la función drift y difusión, basados en el test de razón de verosimilitudes.
- [Gao y King \(2004\)](#) proponen un test óptimo basado en las funciones de densidad marginales de una clase de procesos de difusión no lineal.
- El test de especificación no-paramétrico desarrollado por [Hong y Li \(2005\)](#) en el que se hace uso de la función de densidad de transición para modelos en tiempo continuo.
- El test no-paramétrico de [Arapis y Gao \(2006\)](#) para la especificación de la linealidad en la función drift.
- [Li \(2007\)](#), propone un test no-paramétrico para la especificación paramétrica de la función de difusión de un proceso de difusión, el cual está basado en el error cuadrático entre la función de difusión estimada, no paraméricamente y la difusión propuesta como hipótesis nula
- Los test propuestos en [Gao y Casas \(2008\)](#) para la especificación de las funciones de drift y volatilidad a partir de una discretización de un modelo en tiempo continuo semi-paramétrico en un contexto de econometría financiera. Dicho tests suponen la elección de un parámetro de suavizado, para el que los autores proponen un procedimiento de selección.
- Las propuestas basadas en el proceso estocástico de la volatilidad integrada presentados por [Dette y Podolskij \(2008\)](#) para la forma paramétrica de la función de volatilidad.
- El test de [Chen et al. \(2008\)](#) para la especificación de un proceso de difusión paramétrico basado en la densidad de transición del proceso estimada mediante técnicas de suavizado.
- La propuesta de [Song y Lee \(2009\)](#) en la que se considera un test para el cambio de parámetro en procesos de difusión observados discretamente. Dicho test es basado en el estimador propuesto por [Kessler \(1997\)](#).
- Los test de bondad de ajuste propuestos por [Kutoyants \(2010\)](#) para procesos de difusión ergódicos basados en la función de distribución empírica y en el estimador *local time* de la densidad invariante. Dichos tests son de distribución asintótica libre a partir de la aplicación de cierta transformación.
- [Aït-Sahalia et al. \(2010\)](#) quienes proponen un test basado en la ecuación de Chapman-Kolmogorov.

Las propuestas antes citadas, en su mayoría, suelen estar basadas en técnicas no-paramétricas, en cuanto a la estimación del modelo o el test propuesto. Ciertamente esto supone un inconveniente adicional, debido a la selección del parámetro de suavizado lo cual además puede repercutir en la potencia del test. Por ello, en la sección siguiente presenta dos propuestas de tests basados en procesos empíricos, ya que este tipo de metodología es eficiente en la evaluación de las propiedades en muestras grandes de los estimadores basados en modelos paramétricos.

## 3.2 Test de Bondad de Ajuste basado en Procesos Empíricos

Esta memoria considera un test de bondad de ajuste basado en procesos empíricos como método de validación. El objetivo es introducir un test para comparar la bondad de ajuste para las formas paramétricas de las funciones drift y la volatilidad en modelos de tipos de interés. Las hipótesis en estudio son

$$H_0 : \mu \in \{\mu(\cdot, \theta) : \theta \in \Theta\} \quad (3.1)$$

para la forma paramétrica de la función drift y

$$H_0 : \sigma \in \{\sigma(\cdot, \theta) : \theta \in \Theta\} \quad (3.2)$$

para la forma paramétrica de la función de volatilidad. Una manera alternativa de abordar este problema es considerar ideas relativas a los tests para los modelos de regresión basados en la función de regresión integrada. Un trabajo que considera este tipo de enfoque es el presentado en [Stute \(1997\)](#) para modelos de regresión. Posteriormente una propuesta similar, pero orientada a series de tiempo se discute en [Koul y Stute \(1999\)](#). En lo que respecta a tests de bondad de ajuste basados en procesos empíricos para modelos de difusión, la literatura disponible es limitada, por supuesto, con la excepción de algunos trabajos recientes: [Lee y Wee \(2008\)](#), quien propone un test basado en el proceso empírico de los residuos del modelo de difusión; [Negri y Nishiyama \(2009\)](#), [Negri y Nishiyama \(2010\)](#) y [Masuda et al. \(2010\)](#), quienes proponen tests de bondad de ajuste basados en procesos marcado por los residuos tanto para observaciones en tiempo continuo como para observaciones en tiempo discreto de un proceso de difusión ergódico. Las propuestas anteriores tienen la particularidad de que solo consideran test para la hipótesis nula simple. En este sentido, la propuesta presentada en [Monsalve-Cobis et al. \(2011\)](#) y que se describe a lo largo de este capítulo, está dirigida a un caso más general, es decir, la hipótesis compuesta de la forma paramétrica de la función drift. Además, en esta sección también se propone un test para la función de difusión o volatilidad basado en procesos empíricos que no ha sido considerada en trabajos anteriores. Dicha propuesta pretende sugerir un test de bondad de ajuste cuya implementación resulta sencilla y eficiente. La idea es reescribir el proceso de difusión como si de un modelo de regresión o de serie de tiempo se tratase. Así pues, el test de bondad de ajuste para la función drift se obtiene a partir de la función de regresión integrada del proceso de variaciones que caracteriza el tipo de interés, dando como resultado un proceso empírico basado en los residuos. El test para la función de volatilidad se construye a partir de la función de varianza condicional integrada, obteniéndose un nuevo proceso empírico a partir del cual se contrastará la hipótesis de la forma paramétrica de la función de volatilidad. Las distribuciones de ambos estadísticos son aproximadas mediante técnicas bootstrap, al estilo de las ideas aplicadas en el contexto de los modelos de regresión presentadas en [Stute et al. \(1998\)](#). Es importante destacar que para la implementación del test solo se requiere de la estimación de los parámetros del proceso (mediante estimadores "root-n" consistentes) y la aplicación de un procedimiento bootstrap, el cual se detallará posteriormente, para la calibración de la distribución del estadístico del test, con lo cual su implementación resulta bastante sencilla como se verá a continuación en las siguientes secciones.

### 3.2.1 Test para la función drift

Antes de presentar el test para la función drift, se hará una breve exposición del test basado en la función de regresión integrada presentado en [Stute \(1997\)](#). En el contexto de los modelos de regresión un procedimiento alternativo, a los métodos basados en técnicas de suavizado, para la construcción de Test de Bondad de Ajuste para la función de regresión, es el basado en la Función de Regresión Integrada. Considérese un vector  $(X, Y)$ , con  $F$  la función de Distribución de  $X$  y la



función de regresión asociada

$$m(x) = \mathbb{E}(Y|X = x).$$

La función de regresión integrada se define como:

$$I(x) = \mathbb{E}(Y \mathbb{1}_{\{X \leq x\}}) = \int_{-\infty}^x m(y) dF(y). \quad (3.3)$$

La función (3.3) se puede estimar empíricamente mediante

$$I_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} Y_i. \quad (3.4)$$

La idea es comparar el estimador de la función de regresión integrada,  $I_n(x)$ , con algún estimador basado en las suposiciones bajo la hipótesis nula. Así pues, si se considera la hipótesis nula

$$H_0 : m = m_0$$

e  $I_0(x)$  un estimador apropiado de (3.3) bajo la hipótesis nula, por ejemplo

$$I_0(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} m_0(X_i),$$

entonces el test se define a partir del proceso empírico determinado por los residuos de la regresión

$$R_n(x) = n^{1/2}(I_n(x) - I_0(x)) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (Y_i - m_0(X_i)), \quad x \in \mathbb{R}.$$

En el caso de una hipótesis nula compuesta,  $H_0 : m \in \{m_\theta : \theta \in \Theta\}$ , el test de bondad de ajuste se define a partir del proceso:

$$R_n(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} (Y_i - m_{\hat{\theta}}(X_i)) \quad x \in \mathbb{R}.$$

con  $\hat{\theta}$  un estimador apropiado del parámetro verdadero  $\theta$ . En cada caso, para contrastar la hipótesis  $H_0$  se aplica algún funcional (por ejemplo, el supremo dando lugar así, al estadístico de Kolmogorov-Smirnov).

En esta sección se propone un método general (véase [Monsalve-Cobis et al. \(2011\)](#)) basado en el proceso empírico determinado por los residuos, en el contexto de los modelos de difusión, para probar la bondad del ajuste de la forma paramétrica de la función drift del modelo de tipo de interés (2.22). En otras palabras, la hipótesis en estudio es

$$H_0 : \mu \in \{\mu(\cdot, \theta) : \theta \in \Theta\}, \quad (3.5)$$

donde  $\mu(\cdot, \theta)$  representa la función drift del modelo (2.22) con  $(\theta \in \Theta \subset \mathbb{R}^p)$ . Para contrastar esta hipótesis, se considera la versión discretizada del modelo (2.22), es decir, el modelo

$$\frac{Y_{t_i}}{\Delta} = \mu(r_{t_i}, \theta) + \sigma(r_{t_i}, \theta) \Delta^{-\frac{1}{2}} \varepsilon_{t_{i+1}}, \quad i = 1, \dots, n \quad (3.6)$$

donde  $Y_{t_i} = r_{t_{i+1}} - r_{t_i}$ , representa las variaciones o diferencias del proceso  $\{r_{t_i}\}$  y  $\{\varepsilon_{t_{i+1}}\}$  son variables aleatorias independientes con distribución  $N(0, 1)$  e independientes del proceso  $\{r_{t_i}\}$ . Reescribiendo el proceso anterior como si de un modelo de serie de tiempo paramétrico se tratara,

$$\frac{Y_{t_i}}{\Delta} = \mu(r_{t_i}, \theta) + \eta_{t_i}, \quad (3.7)$$

donde  $\eta_{t_i} = \sigma(r_{t_i}, \theta) \Delta^{-\frac{1}{2}} \varepsilon_{t_{i+1}}$ , se pueden aplicar ideas relativas a test de bondad de ajuste para modelos de regresión basados en la función de regresión integrada.

Para un estimador apropiado  $\hat{\theta}$  del valor verdadero  $\theta$ , el estadístico del Test de Bondad de Ajuste, para modelos de difusión, estará basado en el proceso

$$R_n^d(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right), \quad x \in \mathbb{R}. \quad (3.8)$$

Considérese un funcional continuo  $\Psi(\cdot)$  para definir el estadístico  $D_n = \Psi(R_n^d)$ . Entonces la hipótesis nula  $H_0$  es rechazada si  $D_n > c_{1-\alpha}$  donde  $c_{1-\alpha}$  satisface:

$$\mathbb{P}\{\Psi(R_n^d) > c_{1-\alpha}\} = \alpha,$$

es decir,  $c_{1-\alpha}$  es un valor crítico para un test de nivel  $\alpha$ . Para determinar el valor de  $c_{1-\alpha}$ , es necesario conocer la distribución del proceso  $R_n^d$ . Una alternativa, es aproximar mediante técnicas bootstrap, la distribución de  $R_n^d$ , ver [Stute et al. \(1998\)](#). El valor crítico,  $c_{1-\alpha}$ , es aproximado por  $c_{1-\alpha}^*$  tal que

$$\mathbb{P}^*\{\Psi(R_n^{d*}) > c_{1-\alpha}^*\} = \alpha,$$

donde  $\mathbb{P}^*$  denota una medida de probabilidad generada por la muestra bootstrap y

$$R_n^{d*}(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i}^* \leq x\}} \left( \frac{Y_{t_i}^*}{\Delta} - \mu(r_{t_i}^*, \hat{\theta}^*) \right), \quad x \in \mathbb{R}. \quad (3.9)$$

con  $\hat{\theta}^*$  es un estimador calculado a partir de la muestra bootstrap,  $\{(r_{t_i}^*, Y_{t_i}^*)\}$ , la cual se definirá más adelante. En la práctica,  $c_{1-\alpha}^*$  se aproxima mediante MonteCarlo

$$c_{1-\alpha}^* = D_n^{*[B(1-\alpha)]}$$

siendo el  $[B(1-\alpha)]$ -ésimo estadístico de orden de las  $B$  replicas bootstrap

$$D_n^{*j} = \Psi(R_n^{d*}), \quad 1 \leq j \leq B.$$

De esta manera, el test queda definido por el estadístico  $D_n = \Psi(R_n^d)$  tal que se rechaza  $H_0$  si  $D_n > c_{1-\alpha}^*$ . Como funcional  $\Psi(\cdot)$  se consideran los criterios de Kolmogorov-Smirnov (KS) y de Cramér-von Mises (CvM); es decir,

$$D_n^{KS} = \sup_x |R_n^d(x)|, \quad y \quad D_n^{CvM} = \int_{\mathbb{R}} (R_n^d(x))^2 F_n(dx) \quad (3.10)$$

donde  $F_n$  es la función de distribución empírica de  $\{r_{t_i}\}_i$ . Según sea el caso, se denota por  $D_n = D_n^{KS}$  o bien  $D_n = D_n^{CvM}$  para indicar si se aplica uno u otro estadístico. Por otro lado, también se estima el  $p$ -valor empírico mediante la expresión

$$\frac{\#\{D_n^{*j} > D_n\}}{B},$$

esto es, la proporción de replicas bootstrap  $D_n^*$  que exceden a  $D_n$ .

Entonces el procedimiento bootstrap para llevar a cabo la aproximación del valor crítico es como sigue:

1. Para cada  $i = 1, 2, \dots, n$  se generan

$$Y_{t_i}^* = \mu(r_{t_i}, \hat{\theta}) \Delta + \sigma(r_{t_i}, \hat{\theta}) \Delta^{1/2} \varepsilon_{t_i}^*, \quad (3.11)$$

la muestra bootstrap  $\{(r_{t_i}^*, Y_{t_i}^*)\}_{i=1}^n$ , donde  $\hat{\theta}$  es un estimador apropiado del proceso, la variable  $r_{t_i}^* = r_{t_i}$  permanece inalterada (diseño fijo), y con  $\{\varepsilon_{t_i}^*\}$ , variables aleatorias independientes con distribución  $N(0, 1)$ . Las variables  $\{\varepsilon_{t_i}^*\}$  son independientes de  $\{r_{t_i}\}$ .

2. A partir de un estimador apropiado, se estima  $\hat{\theta}^*$ , el parámetro bajo la hipótesis nula de la remuestra bootstrap,  $\{(r_{t_i}, Y_{t_i}^*)\}_{i=1}^n$ , obtenida en el paso (1),

3. Se determina

$$R_n^{d*}(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}^*}{\Delta} - \mu(r_{t_i}, \hat{\theta}^*) \right), \quad x \in \mathbb{R}.$$

4. Se calcula  $D_n^* = \Psi(R_n^{d*}(x))$ .

5. Se repiten  $B$  veces los pasos anteriores obteniéndose,  $j = 1, 2, \dots, B$  replicas de  $D_n^{*j}$ .

6. Finalmente se calcula

$$\hat{c}_{1-\alpha}^* = D_n^{*[B(1-\alpha)]}$$

*Observación 3.2.1.* Para el procedimiento descrito anteriormente, los estimadores  $\hat{\theta}$  y  $\hat{\theta}^*$  considerados como apropiados, son aquellos que cumplen con criterios del tipo “root- $n$ ” consistente. Una alternativa razonable es aplicar el estimador de máxima verosimilitud propuesto por [Aït-Sahalia \(2002\)](#). Cualquier otro estimador que cumpla dichos criterios puede ser considerado para la estimación de los parámetros. Por supuesto, la selección de un estimador con las características antes mencionadas esta sujeta a la eficiencia del mismo en cuanto a tiempo computacional.

### 3.2.2 Test para la función de volatilidad

En este caso y de manera análoga al test para la función drift, el objetivo es construir un test de bondad de ajuste para contrastar la hipótesis nula acerca de la forma paramétrica de la función de volatilidad,

$$H_0 : \sigma \in \{\sigma(\cdot, \theta) : \theta \in \Theta\}. \quad (3.12)$$

Para definir el test, considérese la función de varianza condicional integrada

$$\begin{aligned} V_o(x) &= \int_{-\infty}^x \sigma^2(u) dF(u) = \mathbb{E}(\sigma^2(r_t, \theta) \mathbb{1}_{\{r_t \leq x\}}) \\ &= \mathbb{E} \left[ \left( \frac{Y_t}{\Delta} - \mu(r_t, \theta) \right)^2 \mathbb{1}_{\{r_t \leq x\}} \right]. \end{aligned}$$

donde  $F_\theta$  es la función de distribución del proceso  $\{r_t\}$ . Un estimador empírico de  $V_o(x)$  es obtenido mediante la expresión:

$$V_{o_n}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \theta) \right)^2. \quad (3.13)$$

Entonces, para  $\hat{\theta}$  un estimador apropiado del parámetro verdadero  $\theta$ , se define el proceso en el que se basará el test de bondad de ajuste, para un modelo de difusión, como

$$R_n^v(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i} \leq x\}} \left( \left( \frac{Y_{t_i}}{\Delta} - m(r_{t_i}, \hat{\theta}) \right)^2 - \frac{\sigma^2(r_{t_i}, \hat{\theta})}{\Delta} \right), \quad x \in \mathbb{R} \quad (3.14)$$

De manera similar al test para la función drift, se considera un funcional continuo  $\Psi(R_n^v)$  para definir el estadístico  $V_n = \Psi(R_n^v)$ , tal que  $H_0$  es rechazada si  $c_{1-\alpha}$  es un valor crítico que satisface

$$\mathbb{P}(\Psi(R_n^v) > c_{1-\alpha}) = \alpha$$

Para aproximar la distribución del estadístico se aplicaran técnicas bootstrap similares a las aplicadas en la sección anterior. Así pues,  $c_{1-\alpha}$  se aproxima por  $c_{1-\alpha}^*$  tal que

$$\mathbb{P}^* (\Psi(R_n^{v*}) > c_{1-\alpha}^*) = \alpha$$

donde,  $\mathbb{P}^*$  denota la medida de probabilidad generada por la muestra bootstrap y

$$R_n^{v*}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i}^* \leq x\}} \left( \left( \frac{Y_{t_i}^*}{\Delta} - m(r_{t_i}^*, \hat{\theta}^*) \right)^2 - \frac{\sigma^2(r_{t_i}^*, \hat{\theta}^*)}{\Delta} \right) \quad x \in \mathbb{R}, \quad (3.15)$$

con  $\hat{\theta}^*$  un estimador del parámetro  $\theta$  calculado a partir de la muestra bootstrap  $\{(r_{t_i}^*), Y_{t_i}^*\}$  que se genera de igual manera que en el caso del drift, mediante (3.11). El valor  $c_{1-\alpha}^*$  es aproximado mediante MonteCarlo

$$c_{1-\alpha}^* = V_n^{*[B(1-\alpha)]}$$

el  $[B(1-\alpha)]$ -ésimo estadístico de orden de las  $B$  replicas bootstrap

$$V_n^{*j} = \Psi(R_n^{v*}), \quad 1 \leq j \leq B.$$

Por lo tanto, el test queda definido por el estadístico  $V_n = \Psi(R_n^v)$  con región de rechazo de  $H_0$  si  $V_n > c_{1-\alpha}^*$ . Análogo al test para la función drift, se consideran como funcionales  $\Psi(\cdot)$  los criterios de Kolmogorov-Smirnov (KS) y Cramér-von Mises (CvM) respectivamente denotados por,

$$V_n^{KS} = \sup_x |R_n^v(x)|, \quad \text{y} \quad V_n^{CvM} = \int_{\mathbb{R}} (R_n^v(x))^2 F_n(dx) \quad (3.16)$$

donde  $F_n$  es la función de distribución empírica de  $\{r_{t_i}\}_i$ . Así mismo, el  $p$ -valor se estima a partir de la expresión

$$\frac{\#\{V_n^{*j} > V_n\}}{B}$$

esto es, la proporción de replicas bootstrap  $V_n^{*j}$  que exceden a  $V_n$ .

### 3.3 Estudio de simulación

Con el propósito de comprobar el rendimiento del test de bondad de ajuste, en esta sección se presenta un estudio de simulación para determinar el nivel y potencia del test propuesto. Para tal fin, se consideran simulaciones para evaluar el desempeño del test, tanto para contrastar la forma paramétrica de la función drift como la forma paramétrica de la función de volatilidad. En esta sección se consideran dos modelos de difusión del tipo CKLS (modelos con reversión a la media y parámetro de elasticidad en la volatilidad constante) y sus respectivas alternativas. Para el primer modelo se considera como hipótesis alternativa el modelo CKLS con ciertas perturbaciones generadas mediante una función no lineal  $\rho(\cdot)$  en su drift, para el caso de test de la función drift, y en la volatilidad para el caso del test para la función de volatilidad. Para el segundo modelo, como hipótesis alternativa supondremos un modelo de difusión con saltos.

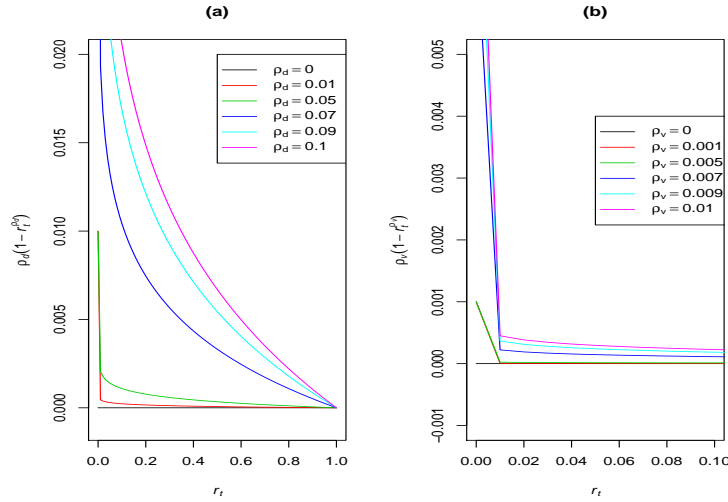
#### Modelo 1.

En el primer caso, el modelo CKLS dado por la ecuación

$$dr_t = (0.0408 - 0.5921r_t)dt + \sqrt{1.6704r_t^{1.4999}} dW_t. \quad (3.17)$$

es considerado como hipótesis nula. Los valores de los parámetros son dados en Chan et al. (1992) basados en el rendimiento de los bonos del tesoro americano con plazo a un mes, y posteriormente usados en Fan et al. (2003). La potencia del test de bondad de ajuste para la función drift se

**Figura 3.1:** Comportamiento de las funciones  $\rho$ . (a) Función de perturbación del drift,  $\rho_d$ ; (b) Función de perturbación de la volatilidad,  $\rho_v$



evaluará en una sucesión de modelos alternativos indexados por el parámetro,  $0 \leq \rho_d \leq 0.1$ , de una función no lineal  $\rho_d(r_t) = \rho_d(1 - r_t^{\rho_d})$ , (ver figura 3.1(a))

$$dr_t = (0.0408 - 0.5921r_t + \rho_d(1 - r_t^{\rho_d})) dt + \sqrt{1.6704r_t^{1.4999}} dW_t. \quad (3.18)$$

Claramente bajo la Hipótesis Nula,  $\rho_d = 0$ . La potencia para el test de bondad de ajuste para la forma paramétrica de la volatilidad se evaluará en una sucesión de modelos alternativos indexados por el parámetro,  $0 \leq \rho_v \leq 0.02$ , de una función no lineal  $\rho_v(r_t) = \rho_v(1 - r_t^{\rho_v})$ , (ver figura 3.1(b))

$$dr_t = (0.0408 - 0.5921r_t)dt + \left( \sqrt{1.6704r_t^{1.4999}} + \rho_v(1 - r_t^{\rho_v}) \right) dW_t. \quad (3.19)$$

Nuevamente, bajo la Hipótesis Nula,  $\rho_v = 0$ . Entonces, para cada valor

$$\rho_d \in \{0, 0.01, 0.05, 0.07, 0.09, 0.1\} \quad \text{y} \quad \rho_v \in \{0, 0.001, 0.005, 0.009, 0.01, 0.02\},$$

se generan series de datos de con frecuencia diaria, es decir  $\Delta = 1/364$ , siguiendo el esquema de Euler y posteriormente se extraen submuestras con frecuencia semanal, es decir  $\Delta = 1/52$ , de los modelos (3.18) y (3.19), de tamaños:  $n = 100, 400, 500, 600$  y  $1000$  para cada caso. Basados en 5000 simulaciones, y para  $B = 1000$  replicas bootstrap, se calcula el porcentaje de rechazos a partir de los respectivos estadísticos (3.10), (3.16) para distintos niveles de significación,  $\alpha = 0.01, 0.05, 0.10$ .

Tal como se aprecia en la tabla (3.1), para  $\rho_d = \rho_v = 0$ , los modelos (3.18) y (3.19), son en realidad el modelo CKLS (3.17), por lo tanto el valor de la potencia bajo la hipótesis nula debería ser aproximadamente el valor nominal del nivel de significación 1 %, 5 %, y 10 %. Como se puede observar en las tablas (3.1) y (3.2), para  $\rho_d = \rho_v = 0$ , los valores estimados se aproximan al nivel nominal de significación. Conforme los valores de  $\rho_d$  y  $\rho_v$  son incrementados, en sus respectivos modelos (3.18) y (3.19), las hipótesis alternativas se desvían de la hipótesis nula por lo que es de esperarse que las tasas de rechazo se incrementen. En efecto, las simulaciones (ver tablas (3.1) y (3.2)) confirman lo anterior, y muestran que el test posee una potencia razonablemente buena. Por ejemplo, el caso de la potencia para el test de bondad del ajuste de la función drift, para  $\rho = 0.07$ , el test es capaz de rechazar aproximadamente 70 % de las veces, para un nivel de significación del 10 %. En el caso de la potencia para el test de bondad de ajuste de la función de volatilidad, para  $\rho_v = 0.009$ , el test es capaz de rechazar cerca del 70 % de las veces, para un nivel de significación del 10 %. Esto sugiere que el test tiene una capacidad de discriminación admisible para la diferenciación del modelo (3.17) respecto de los modelos (3.18) y (3.19).

**Tabla 3.1:** Tasa de rechazos,  $\hat{\alpha}_{KS}$  y  $\hat{\alpha}_{CvM}$ , estimada del modelo (3.18), para diversos valores de  $\rho_d$  y diversos tamaños de muestra,  $n$ , cuando se aplica el estadístico del test para la función drift,  $D_n$ ;  $\hat{p}_{KS}$  y  $\hat{p}_{CvM}$  corresponden a los  $p$ -valores estimados. Los subíndices ( $KS$ ) y ( $CvM$ ) denotan los criterios de Kolmogorov-Smirnov y Crámer von Misses respectivamente.

$\rho_d$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$			$\hat{\alpha}_{CvM}$		
				10 %	5 %	1 %	10 %	5 %	1 %
0	100	0.452	0.477	0.110	0.043	0.008	0.105	0.041	0.007
	400	0.482	0.499	0.091	0.048	0.005	0.083	0.047	0.005
	500	0.494	0.502	0.094	0.049	0.005	0.089	0.047	0.005
	600	0.501	0.513	0.095	0.045	0.007	0.094	0.045	0.008
	1000	0.499	0.507	0.097	0.052	0.008	0.091	0.048	0.016
0.05	100	0.271	0.318	0.297	0.169	0.027	0.281	0.200	0.060
	400	0.263	0.202	0.307	0.198	0.042	0.235	0.216	0.052
	500	0.268	0.200	0.312	0.211	0.054	0.302	0.201	0.061
	600	0.271	0.203	0.309	0.213	0.068	0.332	0.221	0.089
	1000	0.275	0.209	0.301	0.205	0.060	0.341	0.217	0.092
0.07	100	0.160	0.176	0.701	0.515	0.190	0.781	0.617	0.302
	400	0.171	0.160	0.732	0.528	0.187	0.762	0.624	0.298
	500	0.153	0.165	0.714	0.587	0.196	0.754	0.645	0.305
	600	0.134	0.159	0.722	0.562	0.203	0.760	0.633	0.311
	1000	0.101	0.112	0.780	0.526	0.210	0.765	0.660	0.308
0.1	100	0.108	0.121	0.799	0.601	0.208	0.840	0.689	0.311
	400	0.097	0.089	0.831	0.647	0.221	0.791	0.676	0.317
	500	0.058	0.085	0.807	0.716	0.230	0.802	0.680	0.328
	600	0.060	0.076	0.901	0.781	0.343	0.854	0.741	0.421
	1000	0.047	0.030	0.950	0.811	0.380	0.861	0.769	0.453

Un aspecto a destacar en este estudio de simulación es el referido al tamaño de la muestra. Evidentemente en todo estudio de simulación de MonteCarlo el aumento del tamaño de la muestra conlleva a una mejora en las estimaciones. Por ejemplo en el caso de un test, la potencia. Sin embargo, en el estudio presentado anteriormente no se refleja este hecho del todo. En la tabla (3.1) pueden observarse casos en los que un aumento del tamaño de la muestra, por ejemplo  $n = 1000$  para  $\rho_d = 0.07$ , no produce un aumento de la potencia,  $\hat{\alpha}_{KS} = 0.526$  al nivel de significación 5%, si se lo compara con el correspondiente tamaño muestral  $n = 500$  el cual produce un estimado  $\hat{\alpha}_{KS} = 0.587$ . Una razón que explica este fenómeno es el efecto que produce un aumento de la muestra en la ventana de observación. El muestreo considerado en este trabajo es de diseño fijo, es decir, la ventana de observación es  $[0, n\Delta = T]$ , con  $\Delta = \text{constante}$  que se incrementa cuando  $n$  (tamaño de la muestra) aumenta. En otras palabras, al aumentar el tamaño de la muestra no se están incluyendo más observaciones en un mismo período, por el contrario, se está ampliando el horizonte de observación, lo cual implica que en principio no estamos observando el mismo proceso, sino un proceso con un subconjunto de nuevas observaciones. Por ello, la serie pudiera verse afectada por este hecho, así mismo las estimaciones de los parámetros y la potencia, ya que en realidad no es un aumento de la muestra, sobre todo cuando de procesos con reversión a la media se trata.

A fin de verificar el desempeño de los test propuestos en esta memoria, los mismos serán com-

**Tabla 3.2:** Tasa de rechazos,  $\hat{\alpha}_{KS}$  y  $\hat{\alpha}_{CvM}$ , estimada del modelo (3.18), para diversos valores de  $\rho_v$  y diversos tamaños de muestra,  $n$ , cuando se aplica el estadístico del test para la función de volatilidad,  $V_n$ ;  $\hat{p}_{KS}$  y  $\hat{p}_{CvM}$  corresponden a los  $p$ -valores estimados. Los subíndices (KS) y (CvM) denotan los criterios de Kolmogorov-Smirnov y Crámer von Misses respectivamente.

$\rho_v$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$			$\hat{\alpha}_{CvM}$		
				10 %	5 %	1 %	10 %	5 %	1 %
0	100	0.502	0.484	0.083	0.031	0.004	0.078	0.036	0.005
	400	0.511	0.501	0.091	0.039	0.006	0.098	0.046	0.006
	500	0.509	0.502	0.095	0.042	0.007	0.097	0.049	0.011
	600	0.508	0.498	0.105	0.048	0.009	0.096	0.050	0.009
	1000	0.501	0.494	0.102	0.052	0.010	0.098	0.050	0.013
0.005	100	0.411	0.418	0.162	0.082	0.019	0.154	0.070	0.010
	400	0.334	0.317	0.240	0.171	0.062	0.228	0.190	0.056
	500	0.378	0.390	0.287	0.203	0.088	0.233	0.204	0.044
	600	0.365	0.401	0.291	0.219	0.094	0.256	0.201	0.063
	1000	0.360	0.351	0.283	0.220	0.119	0.264	0.240	0.100
0.009	100	0.301	0.321	0.331	0.271	0.254	0.279	0.177	0.155
	400	0.222	0.217	0.487	0.441	0.296	0.567	0.462	0.321
	500	0.210	0.198	0.562	0.462	0.309	0.619	0.495	0.350
	600	0.181	0.171	0.649	0.498	0.321	0.644	0.520	0.387
	1000	0.145	0.101	0.706	0.601	0.511	0.738	0.561	0.523
0.02	100	0.211	0.152	0.410	0.298	0.129	0.509	0.401	0.202
	400	0.055	0.041	0.791	0.798	0.702	0.903	0.900	0.803
	500	0.041	0.038	0.865	0.804	0.711	0.899	0.912	0.799
	600	0.034	0.018	0.890	0.821	0.729	0.912	0.905	0.814
	1000	0.005	0.003	0.989	0.965	0.936	0.979	0.934	0.950

parados a los test desarrollados en Gao y Casas (2008), para testear las formas paramétricas de las funciones drift y volatilidad, los cuales son basados en técnicas de suavizado. Gao y Casas (2008) proponen dos tests tipo núcleo para la especificación del drift y difusión en una versión discretizada de un modelo financiero semiparamétrico en tiempo continuo. El test para la especificación paramétrica del drift es dada por el estadístico

$$L_{2n}(h) = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n p_{ij} \hat{\epsilon}_i \hat{\epsilon}_j}{\hat{\sigma}_{2n}},$$

donde  $p_{ij} = K\left(\frac{r_{t_i} - r_{t_j}}{h}\right)$ ,  $\hat{\epsilon}_i = (Y_{t_i}/\Delta - \mu(r_{t_i}, \hat{\theta}))$ ,  $\hat{\theta}$  es un estimador "root- $n$ " consistente de  $\theta$  y  $\hat{\sigma}_{2n}^2 = 2\hat{\nu}^2 \int K^2(u) du$  con  $\hat{\nu}_2 = \frac{1}{n} \sum_{i=1}^n \sigma^2(r_{t_i}, \hat{\theta})$ . El test para la especificación paramétrica de la difusión es dada por el estadístico

$$L_{1n}(h) = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n \hat{\eta}_i p_{ij} \hat{\eta}_j}{\hat{\sigma}_{1n}},$$

donde  $\hat{\eta}_i = (Y_{t_i}/\Delta - \mu(r_{t_i}, \hat{\theta}))^2 - \sigma^2(r_{t_i}, \hat{\theta})$ ,  $\hat{\theta}$  es un estimador "root- $n$ " consistente de  $\theta$  bajo la hipótesis nula y  $\hat{\sigma}_{1n}^2 = 2\hat{\mu}_2^2 \int K^2(u) du$ , con  $\hat{\mu}_2 = \frac{2}{n} \sum_{i=1}^n \sigma^2(r_{t_i}, \hat{\theta})$ .

Aplicamos dichos test para verificar la forma paramétrica del drift y la difusión del modelo CKLS dado por

$$dr_t = (0.00738 - 0.087r_t) + 0.779r_t^{1.48} dW_t,$$

cuyos parámetros son tomados de Aït-Sahalia (1999). Se seleccionan  $N = 250$  simulaciones, siguiendo las mismas pautas que en el estudio de simulación anterior, y en este caso se considera  $B = 500$  el tamaño de las replicas bootstrap para calcular el tamaño y la potencia del test. Nuevamente, el modelo CKLS es considerado como la hipótesis nula. La potencia de la función drift es evaluada en una serie de modelos alternativos indexados por el parámetro  $0 \leq \rho_d \leq 0.1$ , donde  $\rho_d$  es un parámetro de una función n-lineal  $\rho_d(r_t) = \rho_d(1 - r_t^{\rho_d})$ . Así mismo, la potencia para el test de la forma paramétrica de la volatilidad es evaluado en una serie de alternativas indexadas por el parámetro,  $0 \leq \rho_v \leq 0.02$ , donde  $\rho_v$  es un parámetro de la función no-lineal  $\rho_v(r_t) = \rho_v(1 - r_t^{\rho_v})$ . Para la especificación paramétrica de la función drift, Gao y Casas (2008) aplicaron dos parámetros de suavizado:  $h_{cv}$  (validación cruzada) y  $h_{2test}$  (ancho de banda óptimo) para el estadístico  $L_2(\cdot)$ . En el caso de la especificación paramétrica de la función de difusión, los parámetros de suavizado aplicados por los autores fueron:  $h_{cv}$  (validación cruzada) y  $h_{1test}$  (ancho de banda óptimo) para el estadístico  $L_1(\cdot)$ . Los detalles de como son obtenidos se pueden ver en Gao y Casas (2008). Los resultados correspondientes son presentados en las in Tablas (3.3) and (3.4).

**Tabla 3.3:** Tamaño y Potencia (especificación del la función drift mediante el estadístico  $L_{2n}(\cdot)$ )

$\rho_d$	$n$	$\hat{p}_{h_{cv}}$	$\hat{p}_{h_{2test}}$	$\hat{\alpha}_{h_{cv}}$			$\hat{\alpha}_{h_{2test}}$		
				10%	5%	1%	10%	5%	1%
0	100	0.450	0.487	0.080	0.048	0.008	0.107	0.054	0.012
	400	0.508	0.502	0.092	0.040	0.008	0.115	0.055	0.051
	500	0.465	0.498	0.132	0.064	0.004	0.106	0.052	0.011
	600	0.487	0.472	0.116	0.060	0.008	0.092	0.047	0.009
	1000	0.501	0.490	0.072	0.036	0.012	0.089	0.056	0.006
0.05	100	0.504	0.481	0.111	0.071	0.018	0.132	0.075	0.004
	400	0.506	0.464	0.132	0.059	0.020	0.169	0.071	0.012
	500	0.462	0.451	0.171	0.101	0.043	0.125	0.056	0.011
	600	0.429	0.423	0.181	0.112	0.031	0.130	0.058	0.009
	1000	0.410	0.429	0.197	0.109	0.024	0.152	0.062	0.013
0.1	100	0.303	0.349	0.240	0.152	0.036	0.188	0.076	0.004
	400	0.361	0.402	0.236	0.144	0.032	0.220	0.120	0.020
	500	0.342	0.454	0.232	0.144	0.048	0.148	0.088	0.028
	600	0.344	0.439	0.192	0.108	0.020	0.172	0.072	0.020
	1000	0.365	0.421	0.220	0.148	0.048	0.191	0.079	0.031

Bajo las mismas condiciones, se aplican los test de bondad de ajustes basados en procesos empíricos, propuestos en este trabajo y los resultados son mostrados en las Tablas (3.5) and (3.6). Las tablas (3.5) y (3.6) muestran que el nivel es aproximado de forma satisfactoria en mucho de los casos y además que el test basado en procesos empíricos exhibe un mejor comportamiento en lo que respecta a la potencia en comparación con el método de Gao y Casas (2008). Esto último es de esperarse ya que, como es sabido, la selección del parámetro de suavizado suele afectar la potencia de los test basados en técnicas de suavizado.

Finalmente, los tiempos de ejecución para cada simulación tomando  $B = 500$  replicas bootstrap se muestran en la Tabla (3.7). En ella se muestra que los tiempos de ejecución para los test propuestos en este trabajo, basados en procesos empíricos, son significativamente menores que los empleados por el enfoque propuesto por Gao y Casas (2008).



**Tabla 3.4:** Tamaño y Potencia (especificación de la función drift mediante el estadístico  $L_{1n}(\cdot)$ )

$\rho_v$	$n$	$\hat{p}_{h_{cv}}$	$\hat{p}_{h_{1test}}$	$\hat{\alpha}_{h_{cv}}$			$\hat{\alpha}_{h_{1test}}$		
				10 %	5 %	1 %	10 %	5 %	1 %
0	100	0.495	0.516	0.097	0.048	0.000	0.092	0.060	0.008
	400	0.516	0.493	0.095	0.056	0.012	0.080	0.046	0.014
	500	0.488	0.479	0.096	0.056	0.016	0.089	0.043	0.011
	600	0.464	0.498	0.089	0.060	0.008	0.084	0.047	0.008
	1000	0.482	0.530	0.087	0.076	0.004	0.072	0.042	0.021
0.005	100	0.504	0.470	0.108	0.080	0.028	0.132	0.070	0.074
	400	0.506	0.445	0.152	0.072	0.028	0.169	0.081	0.062
	500	0.462	0.428	0.160	0.104	0.060	0.191	0.105	0.041
	600	0.429	0.411	0.220	0.132	0.056	0.215	0.223	0.021
	1000	0.410	0.345	0.312	0.198	0.074	0.241	0.267	0.073
0.02	100	0.404	0.455	0.264	0.168	0.096	0.200	0.100	0.044
	400	0.264	0.403	0.456	0.404	0.252	0.248	0.144	0.068
	500	0.189	0.378	0.520	0.471	0.345	0.296	0.165	0.097
	600	0.166	0.358	0.688	0.580	0.404	0.320	0.244	0.100
	1000	0.138	0.345	0.693	0.618	0.504	0.396	0.312	0.176

**Tabla 3.5:** Tasa de Rechazos para la forma paramétrica del drift. ( Estadístico  $R_n^d(x)$  )

$\rho_d$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$			$\hat{\alpha}_{CvM}$		
				10 %	5 %	1 %	10 %	5 %	1 %
0	100	0.443	0.459	0.088	0.048	0.005	0.092	0.032	0.008
	400	0.514	0.535	0.112	0.052	0.012	0.108	0.060	0.008
	500	0.507	0.513	0.084	0.048	0.007	0.080	0.044	0.008
	600	0.481	0.506	0.104	0.050	0.009	0.092	0.044	0.007
	1000	0.488	0.492	0.108	0.056	0.014	0.104	0.060	0.009
0.05	100	0.311	0.297	0.274	0.149	0.041	0.283	0.241	0.067
	400	0.302	0.281	0.289	0.163	0.050	0.277	0.218	0.081
	500	0.288	0.277	0.266	0.208	0.047	0.289	0.251	0.053
	600	0.260	0.261	0.251	0.201	0.044	0.268	0.239	0.057
	1000	0.268	0.279	0.309	0.211	0.065	0.296	0.211	0.061
0.1	100	0.094	0.083	0.756	0.520	0.180	0.740	0.605	0.272
	400	0.089	0.081	0.722	0.541	0.208	0.753	0.623	0.311
	500	0.087	0.075	0.767	0.571	0.217	0.760	0.635	0.305
	600	0.096	0.089	0.801	0.565	0.223	0.771	0.624	0.317
	1000	0.106	0.092	0.851	0.500	0.197	0.765	0.601	0.279

**Tabla 3.6:** Tasa de rechazos para la forma paramétrica de la función de volatilidad. (Estadístico  $R_v^d(x)$ )

$\rho_v$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$			$\hat{\alpha}_{CvM}$		
				10 %	5 %	1 %	10 %	5 %	1 %
0	100	0.532	0.533	0.080	0.036	0.004	0.070	0.036	0.003
	400	0.488	0.513	0.096	0.045	0.006	0.100	0.052	0.006
	500	0.496	0.488	0.094	0.042	0.008	0.096	0.040	0.016
	600	0.501	0.487	0.116	0.044	0.012	0.128	0.056	0.028
	1000	0.523	0.537	0.098	0.049	0.012	0.112	0.068	0.008
0.005	100	0.431	0.429	0.149	0.078	0.031	0.121	0.068	0.011
	400	0.361	0.367	0.212	0.135	0.047	0.202	0.155	0.043
	500	0.388	0.363	0.225	0.171	0.049	0.215	0.143	0.050
	600	0.392	0.355	0.217	0.166	0.051	0.222	0.151	0.069
	1000	0.371	0.363	0.278	0.201	0.096	0.278	0.206	0.113
0.02	100	0.318	0.319	0.271	0.171	0.048	0.279	0.151	0.051
	400	0.198	0.185	0.579	0.463	0.350	0.694	0.529	0.336
	500	0.178	0.170	0.681	0.580	0.345	0.727	0.601	0.366
	600	0.172	0.163	0.699	0.578	0.367	0.719	0.611	0.342
	1000	0.122	0.128	0.713	0.643	0.520	0.788	0.694	0.501

**Tabla 3.7:** Tiempos de Ejecución por iteración de los enfoques basados en los procesos  $L_i(\cdot)$ , y  $R_n^{(\cdot)}$ , propuestos por Gao y Casas (2008) and en este trabajo, respectively

$n$	Run-times (minutes)		
	$L_i(h_{cv})$	$L_i(h_{itest})$	$R_n^{(\cdot)}$
100	0.189	0.206	0.098
400	0.559	0.984	0.249
500	0.783	1.871	0.302
600	0.982	2.148	0.357
1000	2.253	4.204	0.595

## Modelo 2.

Los modelos de difusión con salto fueron introducidos por Das (2002) y Johanes (2004) con el fin de caracterizar la dinámica de los tipos de interés. Su utilización esta fundamentada principalmente en el hecho de que los shocks económicos, intervenciones del gobierno en el mercado, la difusión de noticias, etc, causan saltos elevados en los tipos de interés, lo que genera discontinuidades en su dinámica. Se considera el modelo de difusión CKLS con saltos, en base a la especificación propuesta por Das (2002), es decir,

$$dr_t = (\alpha + \beta r_t) dt + \sigma r_t^\gamma dW_t + Z_t d\pi_t(q_t),$$

Así pues, los tipos de interés evolucionan con una función drift  $(\alpha + \beta r_t)$  con reversión a la media  $\beta < 0$ , y dos términos aleatorios, la función de volatilidad ( $W_t$  un movimiento Browniano estándar) o difusión y un proceso Poisson  $\pi_t$  con frecuencia de sucesos  $q_t$ , incluyendo un salto aleatorio  $Z_t$  con distribución  $N(\lambda, \rho^2)$ . Sea la especificación del modelo (discretizado) considera-

da por Das (2002),

$$Y_{t_i} = \alpha + \beta r_{t_i} + \sigma r_{t_i}^\gamma \varepsilon_{t_i} + Z_{t_{i+1}} \Delta \pi_{t_{i+1}}(q_{t_{i+1}}), \quad i = 1, \dots, n, \quad (3.20)$$

donde  $\{\varepsilon_{t_i}\}$  son variables aleatorias independientes con distribución  $N(0, 1)$  e independientes de  $\{r_{t_i}\}$ ;  $\Delta \pi_{t_i}(q_{t_i})$  es el proceso Poisson de incrementos en tiempo discreto, aproximado por una variable aleatoria independiente con distribución de Bernoulli de parámetro

$$q_{t_i} = \frac{1}{1 + \exp(-a - br_{t_{i-1}})}. \quad (3.21)$$

y con tamaño de salto el valor de la variable aleatoria  $Z_{t_i}$  con distribución  $N(\lambda, \rho^2)$ . La media condicionada para la variación en el tipo de interés es:

$$\mathbb{E}(r_{t_{i+1}} - r_{t_i} | r_{t_i}) = \alpha + \beta r_{t_i} + q_{t_{i+1}} \lambda, \quad (3.22)$$

mientras que la varianza

$$\mathbb{E}[(r_{t_{i+1}} - r_{t_i})^2 | r_{t_i}] = \sigma^2 r_{t_i}^2 + q_{t_{i+1}} (\rho^2 + (1 - q_{t_{i+1}}) \lambda^2). \quad (3.23)$$

Para el estudio de simulación se considera el modelo CKLS como hipótesis nula,

$$dr_t = (0.00739344 - 0.0876r_t)dt + 0.7791r_t^{1.48}dW_t. \quad (3.24)$$

cuyos parámetros son tomados de (Aït-Sahalia (1999)), basado en los datos de los fondos federales con frecuencia mensual, entre Enero de 1963 y Diciembre de 1998. Para el estudio de la potencia del test, se generan  $N = 500$ , para  $n = 300, 400, 500$ , observaciones de una sucesión de alternativas indexadas por

$$\lambda \in \{0; 0.0009; 0.0079; 0.019\}, \quad \rho \in \{0, 0.001\}, \quad \text{y} \quad a = b = -0.5$$

del modelo de difusión con saltos (3.3),

$$dr_t = (0.00739344 - 0.0876r_t) dt + 0.7791r_t^{1.48}dW_t + Z_t d\pi_t(q_t). \quad (3.25)$$

Para  $B = 1000$  replicas bootstrap se obtiene los siguientes resultados. Es claro que bajo la hipótesis nula,  $\lambda = \rho = 0$ , el valor de la potencia debería ser aproximadamente el valor nominal del nivel de significación 5%, y 10%. Como se puede observar en las tablas (3.8), los valores estimados se aproximan al nivel nominal de significación, además puede observarse la influencia del tamaño de la muestra en los estimados.

Cuando el valor de  $\lambda$  es incrementado, las hipótesis alternativas se desvían de la hipótesis nula por lo tanto la tasa de rechazo se incrementa. Esto se ve reflejado en las simulaciones presentadas en la tabla (3.8) la cual confirma lo antes mencionado, y muestra que el test posee una potencia razonable. Así mismo, la influencia del parámetro  $\lambda$  en la varianza (3.23) del proceso de difusión con saltos (3.25) produce un conjunto de hipótesis alternativas para la función de la volatilidad, con lo cual, al aplicar el test de bondad de ajuste para la función de volatilidad se obtienen los resultados en la tabla (3.9). En dicha tabla se aprecia un incremento en la potencia respecto del valor  $\lambda$ .

*Observación 3.3.1. (Una aproximación alternativa al Proceso  $R_n^d$  computacionalmente más eficiente)*

En la sección (3.2.1) la distribución del proceso empírico,  $R_n^d(x)$ , fue aproximada mediante técnicas bootstrap. Una manera alternativa de aproximar la distribución del proceso empírico,  $R_n^d(x)$ , es a partir de una función de distribución límite correspondiente a un proceso Gaussiano de media cero. En el contexto de los tests de bondad de ajuste para modelos lineales mixtos generalizados, Pan y Lin (2005), sugieren este tipo de aproximación para métodos basados en sumas

**Tabla 3.8:** Tasa de rechazos,  $\hat{\alpha}_{KS}$  y  $\hat{\alpha}_{CvM}$ , estimada del modelo (3.25), para diversos valores de  $\rho_d$  y diversos tamaños de muestra,  $n$ , cuando se aplica el estadístico del test para la función drift,  $D_n$ ;  $\hat{p}_{KS}$  y  $\hat{p}_{CvM}$  corresponden a los  $p$ -valores estimados. Los subíndices (KS) y (CvM) denotan los criterios de Kolmogorov-Smirnov y Crámer von Misses respectivamente.

$\lambda$	$\rho^2$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$		$\hat{\alpha}_{CvM}$	
					10%	5%	10%	5%
0	0	300	0.470	0.478	0.080	0.042	0.070	0.041
		400	0.489	0.487	0.091	0.047	0.079	0.042
		500	0.509	0.502	0.109	0.052	0.090	0.048
0.0009	0.001	300	0.387	0.433	0.164	0.092	0.140	0.074
		400	0.384	0.434	0.206	0.110	0.148	0.078
		500	0.256	0.285	0.368	0.252	0.350	0.252
0.0079	0.001	300	0.200	0.209	0.488	0.358	0.434	0.292
		400	0.181	0.194	0.502	0.350	0.456	0.300
		500	0.145	0.156	0.578	0.404	0.516	0.376
0.019	0.001	300	0.177	0.163	0.516	0.374	0.518	0.386
		400	0.137	0.124	0.596	0.418	0.600	0.446
		500	0.108	0.092	0.676	0.508	0.624	0.504

acumulativas de residuos sobre covariables o predicciones de la variable respuesta. En este caso, es decir, en el contexto de los modelos de difusión para modelos de tipos de interés, la aproximación puede llevarse a cabo de la siguiente manera. Sea

$$\begin{aligned} R_n^d(x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left[ \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \theta) \right) + \left( \mu(r_{t_i}, \theta) - \mu(r_{t_i}, \hat{\theta}) \right) \right] \end{aligned}$$

el proceso definido en la sección (3.2.1). Bajo la suposición de que las primeras y segundas derivadas de  $\mu(r_{t_i}, \theta)$  son acotadas y considerando la expansión en series de Taylor de

$$\left( \mu(r_{t_i}, \theta) - \mu(r_{t_i}, \hat{\theta}) \right),$$

alrededor de  $\hat{\theta}$ , el proceso  $R_n^d(x)$  es asintóticamente equivalente al proceso

$$W_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \theta) \right) + v^T(x, \theta) n^{1/2} (\hat{\theta} - \theta).$$

con

$$v(x, \theta) = -\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \dot{\mu}(r_{t_i}, \theta),$$

la cual es una función del gradiente de la función drift evaluada en cada observación  $r_{t_i}$ . Bajo la suposición "root-n"-consistente del estimador del parámetro  $\theta$  se tiene que

$$n^{1/2}(\hat{\theta} - \theta) = i^{-1}(\theta) n^{-1/2} V_n(\theta) + o_p(1),$$

donde

**Tabla 3.9:** Tasa de rechazos,  $\hat{\alpha}_{KS}$  y  $\hat{\alpha}_{CvM}$ , estimada del modelo (3.25), para diversos valores de  $\rho_d$  y diversos tamaños de muestra,  $n$ , cuando se aplica el estadístico del test para la función drift,  $V_n$ ;  $\hat{p}_{KS}$  y  $\hat{p}_{CvM}$  corresponden a los  $p$ -valores estimados. Los subíndices (KS) y (CvM) denotan los criterios de Kolmogorov-Smirnov y Crámer von Misses respectivamente.

$\lambda$	$\rho^2$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$		$\hat{\alpha}_{CvM}$	
					10%	5%	10%	5%
0	0	300	0.470	0.478	0.080	0.042	0.070	0.041
		400	0.489	0.487	0.091	0.047	0.079	0.042
		500	0.509	0.502	0.109	0.052	0.09	0.048
0.0009	0.001	300	0.233	0.237	0.458	0.346	0.468	0.338
		400	0.205	0.213	0.516	0.404	0.512	0.404
		500	0.009	0.008	0.978	0.966	0.984	0.970
0.0079	0.001	300	0.037	0.030	0.904	0.858	0.912	0.880
		400	0.015	0.013	0.960	0.934	0.972	0.948
		500	0.017	0.018	0.962	0.948	0.958	0.946
0.01	0.001	300	0.034	0.029	0.906	0.874	0.930	0.890
		400	0.020	0.020	0.952	0.922	0.950	0.926
		500	0.015	0.013	0.960	0.946	0.966	0.948

$$i(\theta) = \lim_{n \rightarrow \infty} \mathcal{I}(\theta), \quad \text{and} \quad \mathcal{I}(\theta) = -\frac{1}{n} \frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta^T}.$$

Finalmente, el proceso es aproximado por el proceso de sumas acumuladas  $\hat{W}_n(x)$  cuyo primer término corresponde al proceso original estimado y cuyo segundo término corresponde a la estructura de covarianza generada por la estimación de los parámetros en el modelo considerado, dada por

$$\hat{W}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left\{ \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right) + v^T(x, \hat{\theta}) \mathcal{I}^{-1}(\hat{\theta}) \dot{\ell}_i(\hat{\theta}) \right\} Z_i, \quad (3.26)$$

donde  $Z_1, Z_2, \dots, Z_n$  son variables aleatorias normales estándar independientes. De manera similar, para al test presentado en a Sección (3.2.1), el estadístico se define usando

$$S_n^{KS} = \sup_x |\hat{W}_n(x)|, \quad \text{or} \quad S_n^{CvM} = \int_{\mathbb{R}} \left( \hat{W}_n(x) \right)^2 F_n(dx),$$

donde  $F_n$  es la distribución empírica de  $\{r_{t_i}\}_i$ . Según sea el caso, denotaremos por  $S_n = S_n^{KS}$  o bien  $S_n = S_n^{CvM}$  para indicar si aplicamos uno u otro estadístico. Entonces rechazamos  $H_0$  si  $S_n > s_{1-\alpha}^*$ . El punto crítico  $s_{1-\alpha}^*$  es estimado mediante MonteCarlo por

$$s_{1-\alpha}^* = S_n^{*[B(1-\alpha)]}.$$

El estimador del  $p$ -valor viene dado por:

$$\frac{\#\{S_n^{*j} > S_n\}}{B},$$

la proporción de replicas bootstraps  $S_n^*$  que exceden  $S_n$ . La ventaja que ofrece esta aproximación, es la de reducir el costo computacional originado por la aproximación del proceso  $R_n^d$ , ya que por cada iteración, requiere de una nueva estimación de los parámetros para cada remuestra

**Tabla 3.10:** Tasa de rechazos,  $\hat{\alpha}_{KS}$  y  $\hat{\alpha}_{CvM}$ , estimada del modelo (3.18), para diversos valores de  $\rho_d$  y diversos tamaños de muestra,  $n$ , cuando se aplica el estadístico del test para la función drift,  $S_n$ ;  $\hat{p}_{KS}$  y  $\hat{p}_{CvM}$  corresponden a los  $p$ -valores estimados. Los subíndices ( $KS$ ) y ( $CvM$ ) denotan los criterios de Kolmogorov-Smirnov y Crámer von Misses respectivamente.

$\rho_d$	$n$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{\alpha}_{KS}$			$\hat{\alpha}_{CvM}$		
				10%	5%	1%	10%	5%	1%
0	100	0.434	0.468	0.146	0.056	0.005	0.089	0.031	0.007
	500	0.491	0.484	0.114	0.054	0.010	0.101	0.052	0.005
	1000	0.497	0.483	0.092	0.050	0.009	0.094	0.048	0.006
0.05	100	0.24	0.24	0.387	0.237	0.039	0.408	0.271	0.075
	500	0.243	0.247	0.398	0.227	0.074	0.40	0.285	0.09
	1000	0.255	0.267	0.348	0.218	0.059	0.355	0.216	0.06
0.07	100	0.138	0.125	0.615	0.442	0.123	0.651	0.499	0.183
	500	0.13	0.116	0.615	0.434	0.17	0.672	0.513	0.235
	1000	0.145	0.136	0.566	0.373	0.138	0.6	0.438	0.149
0.1	100	0.07	0.062	0.794	0.64	0.227	0.827	0.695	0.305
	500	0.077	0.057	0.781	0.609	0.295	0.843	0.727	0.414
	1000	0.085	0.07	0.75	0.583	0.231	0.785	0.661	0.325

bootstrap. Evidentemente, esa ventaja se ve restringida a una buena aproximación de los términos  $v^T(x, \hat{\theta})$ ,  $\mathcal{I}^{-1}(\hat{\theta})$ , y  $\dot{\ell}_i(\hat{\theta})$ . En la tabla (3.10), se presentan los resultados de aplicar este procedimiento alternativo a los datos simulados a partir del modelo (3.18). Como se puede observar, para  $\rho_d = 0$ , es decir el modelo CKLS, los valores estimados se aproximan al nivel del test. Así mismo, cuando se incrementa el valor de  $\rho_d$ , se incrementan los valores estimados de la potencia, similar a los resultados obtenidos mediante la aproximación bootstrap.

Nótese en la tabla (3.10), el efecto producido por el aumento del tamaño de la muestra en el estudio de simulación y la estimación de la potencia, se acentúa. Es evidente que en este caso, las estimaciones de los parámetros del proceso así como la aproximación de la distribución del proceso involucrado en el test son ciertamente más sensibles a cualquier cambio en el proceso simulado. El incremento del tamaño de la muestra, como ya ha sido mencionado antes, no es considerado como un aumento del número de observaciones en el horizonte de observación, por el contrario, esto genera un proceso distinto en otro horizonte de observación, lo cual no genera un beneficio directo en la estimación de los parámetros ni del estadístico del test.

La tabla (3.11) muestra los tiempos de ejecución de ambas aproximaciones,  $D_n$  y  $S_n$  para el caso particular,  $\rho = 0$  (la hipótesis nula), y con nivel de significación  $\alpha = 0.10$ . Las diferencias entre los tiempos de ejecución son significativas entre una y otra aproximación, las cuales se acentúan cuando se incrementa el tamaño de las series.

### 3.4 Aplicaciones a los datos del EURIBOR

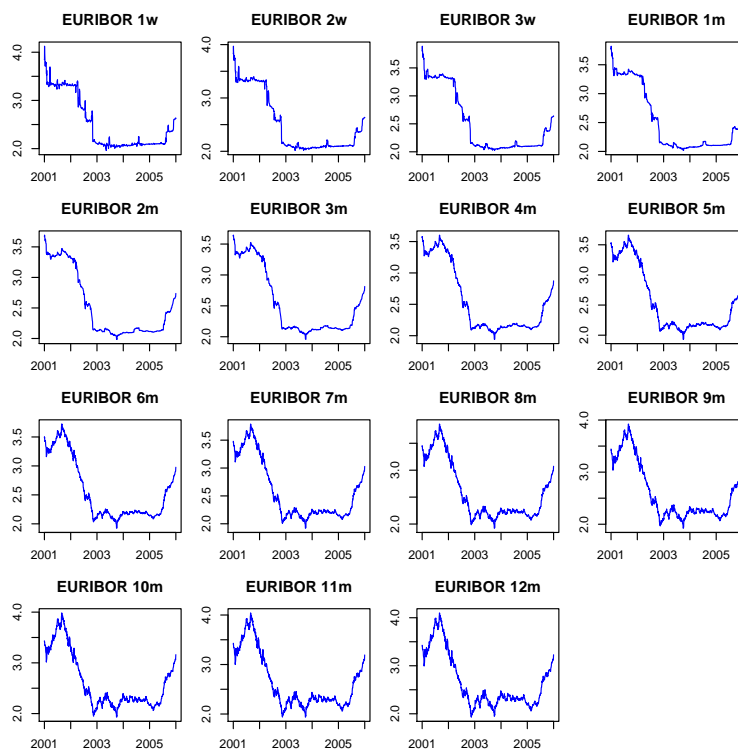
En esta sección se presentan los rasgos más importantes de la serie de tipos de interés del EURIBOR a distintos plazos. El período temporal abarca desde el 15 de octubre del 2001 hasta el 30 de marzo de 2006, con frecuencia diaria. Los plazos considerados son de 1, 2 y 3 semanas y 1, 2, ..., 12 meses. La gráfica (3.2) representa el nivel diario de la serie de interés del EURIBOR para los distintos plazos considerados.

**Tabla 3.11:** Tabla comparativa mostrando los tiempo de ejecución entre las aproximaciones  $\hat{W}_n$  y  $R_n^{d*}$ , de la distribución del proceso  $R_n^d$ ;  $t_{iter}$ , tiempo de ejecución por cada aproximación de  $D_n$  y  $S_n$  correspondiente a una muestra tamaño  $n$ ;  $t_{sim}$ , tiempo de ejecución hasta obtener  $\hat{\alpha}_{KS}$ .

$n$	$D_n$				$S_n$			
	$\hat{p}_{KS}$	$\hat{\alpha}_{KS}$	$t_{iter}(\text{sec})$	$t_{sim}(\text{min})$	$\hat{p}_{KS}$	$\hat{\alpha}_{KS}$	$t_{iter}(\text{sec})$	$t_{sim}(\text{min})$
100	0.477	0.081	10.78	180	0.434	0.146	0.136	2.26
500	0.499	0.110	34.92	582	0.491	0.114	0.362	6.04
1000	0.507	0.107	62.52	1042	0.497	0.092	0.621	10.35

Se aplicará el test de bondad de ajuste planteado en las secciones anteriores a las series de tipos de interés del EURIBOR mencionadas previamente, para probar la bondad del ajuste de los modelos de Vasicek (1977) (VAS), Chan et al. (1992) (CKLS), en cuanto a la forma paramétrica de la función drift y de la función de volatilidad. En la tabla (3.12) se ilustran los valores estimados

**Figura 3.2:** Series del Euribor



de: los  $p$ -valores, los valores críticos  $\hat{c}_{0.10}^*$ , el valor del estadístico en la muestra dada,  $D_n^{KS}$  y  $V_n^{KS}$ . El  $p$ -valor estimado resulta significativo para las series con plazos prolongados más allá de los tres meses. Esto puede traducirse como un indicativo de que el modelo puede ser adecuado para explicar la dinámica de las series del EURIBOR, al menos en cuanto a la función drift se refiere.

Sin embargo, el  $p$ -valor en el caso de la función de volatilidad evidencia lo inadecuado del modelo para explicar la volatilidad de las series. Se observa que, en el caso de las series con plazos menores a los tres meses los estimados del  $p$ -valor, para el test de bondad de ajuste de la función drift, muestran valores relativamente bajos. La tabla (3.13) muestra los resultados obtenidos cuando aplicamos el estadístico  $S_n$  presentado en la observación (3.3.1) a las series del EURIBOR y un

**Tabla 3.12:** *p*-valores estimados para el estudio de la bondad del ajuste del modelo de Vasicek (VAS) y del modelo CKLS

Maturity	Drift function				Volatility function			
	Vasicek		CKLS		Vasicek		CKLS	
	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$	$\hat{p}_{KS}$	$\hat{p}_{CvM}$
1 week	0.033	0.011	0.053	0.138	0	0	0	0
2 weeks	0.212	0.029	0.405	0.239	0	0	0	0
3 weeks	0.109	0.030	0.251	0.168	0	0	0	0
1 month	0.105	0.018	0.254	0.122	0	0	0.002	0.0002
2 months	0.199	0.044	0.265	0.111	0	0	0	0
3 months	0.409	0.169	0.454	0.275	0	0	0	0
4 months	0.710	0.653	0.693	0.717	0	0	0	0.0002
5 months	0.999	0.977	0.999	0.975	0	0	0	0.0004
6 months	0.963	0.985	0.989	0.976	0	0	0	0
7 months	0.928	0.937	0.945	0.913	0	0	0.002	0
8 months	0.923	0.860	0.942	0.834	0	0	0.004	0
9 months	0.897	0.803	0.957	0.768	0	0	0	0
10 months	0.930	0.736	0.966	0.691	0	0	0.001	0.0004
11 months	0.885	0.666	0.909	0.618	0	0	0.001	0
12 months	0.804	0.598	0.918	0.558	0	0	0.001	0

**Tabla 3.13:** Comparativa de las aproximaciones bootstrap de los estadísticos  $D_n$  y  $S_n$  del test de bondad para la función drift, (KS) correspondientes al criterio de Kolmogorov-Smirnov, valores críticos  $\hat{c}_{0.10}^{*KS}$  (10 % significación), *p*-valores y  $t_{iter}$  tiempos de ejecución por iteración, para el modelo CKLS.

Plazo de Vencimiento	$R_n^{d*}$				$\hat{W}_n$			
	$\hat{p}_{KS}$	$c_{0.10}^{KS}$	$D_n^{KS}$	$t_{iter}$ (min)	$\hat{p}_{KS}$	$c_{0.10}^{KS}$	$S_n^{KS}$	$t_{iter}$ (min)
1 semana	0.044	0.028	0.032	1.489	0.072	0.033	0.035	0.094
3 meses	0.422	0.010	0.007	1.102	0.501	0.010	0.007	0.091
6 meses	0.964	0.013	0.006	1.308	0.990	0.014	0.005	0.090
9 meses	0.976	0.018	0.008	1.189	0.968	0.019	0.009	0.090
12 meses	0.888	0.022	0.011	1.136	0.926	0.023	0.011	0.090

modelo CKLS es considerado para explicar la dinámica de los tipos de interés. Como puede observarse, los valores obtenidos en cada caso presentan discrepancias poco significativas en cuanto a los valores de los puntos críticos, estadístico del test y *p*-valores, con lo cual, las conclusiones extraídas por uno u otro método no implican, en la mayoría de los casos, diferencias. Lo que destaca en entre ambos métodos es el tiempo de ejecución, el cual es significativo.

A la vista de los resultados obtenidos, se demuestra claramente que un test de bondad de ajuste basado en procesos empíricos es una herramienta de gran utilidad que puede ser usada para en la discriminación de los diversos modelos de tipos de interés. El test presentado exhibe un desempeño satisfactorio, lo cual se refleja en su tamaño y en la potencia asociada a este. Otro punto a destacar es la implementación del test y el remuestreo bootstrap, los cuales son bastante simples si se lo compara con otros métodos basados en técnicas de suavizado. Cuando el test es aplicado a las series de tipos de interés del EURIBOR, los resultados indican que el test tiene la capacidad de discriminar entre los modelos que intentan modelizar la dinámica de dichas series.



Por ejemplo, para series con plazo de vencimiento mayor a tres meses, el test determina que el modelo *CKLS*, el cual exhibe un drift lineal con reversión de la media, puede ser adecuado. Sin embargo, para períodos inferiores a los tres meses, dichos modelos no son considerados para explicar el comportamiento del tipo de interés. Este último punto es corroborado por la teoría económica, la cual establece que los modelos con plazos de vencimiento cortos tienden a ser más susceptibles a las decisiones de los organismos que regulan el mercado.

# 4

## Aspectos Teóricos del Test de Bondad de Ajuste para la función Drift

El objetivo de este capítulo es estudiar la distribución límite y la consistencia del test de bondad de ajuste para la función drift discutido en el capítulo 3. En éste capítulo se estudiará la convergencia débil del estadístico propuesto para el test bajo la hipótesis nula y se probará que el proceso límite es un proceso Gaussiano continuo donde la estructura de su función de covarianza permite calcular la distribución límite del estadístico en cuestión.

Dado el proceso de difusión homogéneo o también conocido como *Proceso de Itô*, estudiado en el capítulo 1, el cual es definido como la solución de la ecuación diferencial estocástica (SDE),

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t, \quad r_0, \quad 0 \leq t \leq T. \quad (4.1)$$

donde, como ya se ha mencionado con anterioridad, las funciones  $\mu(r)$  y  $\sigma^2(r)$ , con  $r \in \mathbb{R}$ , son funciones no aleatorias conocidas respectivamente como *coeficiente de tendencia* y *coeficiente de difusión* en un contexto teórico o bien como la *función drift* o *función de volatilidad* en el contexto financiero, y  $W_t$  es un proceso de Wiener estándar definido en una base estocástica  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ , es decir, un espacio de probabilidades con una filtración  $\{\mathcal{F}_t\}_{t \geq 0}$ . La ecuación diferencial estocástica anterior se puede expresar en su versión integral como

$$r_t = r_0 + \int_0^t \mu(r_u)du + \int_0^t \sigma(r_u)dW_u, \quad 0 \leq t \leq T.$$

El test descrito en la sección (3.2) estudia la bondad de ajuste para la forma paramétrica de la función drift de un proceso de difusión, el cuál es usado para caracterizar el tipo de interés dado por la ecuación:

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t)dW_t, \quad r_0, \quad 0 \leq t \leq T \quad (4.2)$$

donde

- $\theta$  es un parámetro desconocido de dimensión  $d$ , ( $\theta \in \Theta \subset \mathbb{R}^d$ ).
- $\mu(r_t, \theta)$  - la *función drift*, cuya forma funcional se asume conocida.

- $\sigma(r_t)$  - la *función de volatilidad*, la cual se asumirá, en lo que sigue, como una función constante,  $\sigma(\cdot) \equiv \sigma$ . Si es desconocida, ésta se puede estimar de forma consistente.
- $W_t$  - el proceso de Wiener o (*Movimiento Browniano estándar*) y
- $r_0$  la condición inicial, independiente de  $W_t$ ,  $t \geq 0$ .

Puesto que el proceso  $r_t$  es la solución de la ecuación diferencial (4.2), se hace necesario establecer condiciones que permitan garantizar la existencia de la solución de dicha ecuación. Para ello, las funciones drift y volatilidad deben satisfacer un conjunto de supuestos básicos descritos en la sección (2.2.10) del capítulo (2) para garantizar la existencia de dicha solución. Tales condiciones de regularidad típicas en este tipo de problema (véase Karatzas y Shreve (1991), p. 289) incluyen: el crecimiento lineal, continuidad global Lipschitz, y la suavidad de las funciones drift y volatilidad.

En efecto, bajo los supuestos (1) y (2), la ecuación diferencial (4.2) admite una solución fuerte única con condición inicial  $r_0$  adaptada a la filtración  $(\mathcal{F}_t)_{t \geq 0}$ . Por otra parte, esta solución es un proceso de Markov con función de densidad de transición dada por  $p_\theta(\Delta, r_t | r_s)$ . El supuesto (2) controla el comportamiento de la solución para que  $r_t$  no crezca sin control en un tiempo finito. Adicionalmente se consideran los siguientes supuestos:

*Supuesto 6.*

- (i) Sea  $D = (0, +\infty)$  el espacio de estado del proceso de difusión  $r_t$  que es solución de (4.2). La *medida de escala* se define como

$$s(r, \theta) = \exp \left\{ -2 \int_{\tilde{r}}^r \frac{\mu(u, \theta)}{\sigma^2(u, \theta)} du \right\},$$

donde  $\tilde{r}$  es arbitrario pero fijo en  $D$ , y su selección es irrelevante. Entonces la integral de  $s(r, \theta)$  diverge en las fronteras de  $D$ .

- (ii) Sea

$$m(r, \theta) = \frac{1}{\sigma^2(r, \theta)s(r, \theta)},$$

la *medida de rapidez*. Entonces, la integral de  $m(r, \theta)$  converge en las fronteras de  $D$ .

*Supuesto 7.* Las funciones  $\mu(r, \theta)$  y  $\sigma(r, \theta)$  son, infinitamente diferenciables en  $r$ , y tres veces continuamente diferenciable en  $\theta$ , para todo  $r \in D$  y  $\theta \in \Theta$ .

Bajo el supuesto (6), el proceso  $r_t$  tiene propiedades ergódicas, con densidad invariante

$$f_\theta(r) = \frac{1}{M(\theta)\sigma^2(r, \theta)s(r, \theta)},$$

donde  $M(\theta)$  es una constante de normalización. En lo que sigue, las funciones de densidad y distribución se denotarán por  $f_\theta$  y  $F_\theta$ , respectivamente. Se supondrá además que la condición inicial  $r_0$  es una variable aleatoria con ésta función de distribución y que el proceso observado es estacionario. El supuesto (7) es requerido para establecer resultados asintóticos.

Se asumirá que el proceso  $\{r_t : 0 \leq t \leq T\}$  es observado en  $t_i = i\Delta$ , con  $i = 1, \dots, n$ , para  $\Delta$  fijo, tal y como se describió en el capítulo 3. Aplicando el método de Euler, se obtiene la versión discreta del modelo (4.2)

$$r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(W_{t_{i+1}} - W_{t_i}), \quad i = 1, \dots, n.$$

Sean  $Y_{t_i} = r_{t_{i+1}} - r_{t_i}$  y  $Z_{t_i} = W_{t_{i+1}} - W_{t_i}$ . Entonces  $Z_{t_i} \sim N(0, \Delta)$ , debido a las propiedades del proceso de Wiener  $W_t$ . En lo que sigue, y para facilitar la notación, se denotarán

$$Y_{t_i} = Y_{i\Delta}, \quad r_{t_i} = r_{i\Delta}, \quad \text{y} \quad \varepsilon_{t_i} = \varepsilon_{i\Delta},$$

mediante

$$Y_i, r_i, \text{ and } \varepsilon_i, \text{ respectivamente.}$$

De esta forma, el proceso (4.2) queda expresado en su versión discretizada como

$$Y_i = \mu(r_i, \theta)\Delta + \sigma\sqrt{\Delta}\varepsilon_i, \quad i = 1, \dots, n.$$

donde  $\{\varepsilon_i\}$  son variables aleatorias independientes e idénticamente distribuidas.

Como se mencionó en el capítulo 3, el interés principal del test de bondad de ajuste para la función drift propuesto es contrastar la hipótesis nula

$$H_0 : \mu \in \{\mu(\cdot, \theta) : \theta \in \Theta\}, \quad (4.3)$$

donde  $\mu(\cdot, \theta)$  representa la función drift de la ecuación (4.2).

Si  $\mu$  es un elemento de (4.3), es decir,  $\mu = \mu(\cdot, \theta_0)$  (hipótesis nula simple) para algún parámetro  $\theta_0$  desconocido, entonces el test de bondad de ajuste estará basado en el proceso empírico marcado por los residuos definido por

$$R_n^d(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_{i-1} \leq x\}} \left( \frac{Y_i}{\Delta} - \mu(r_i, \theta_0) \right). \quad (4.4)$$

El proceso  $R_n^d$  está determinado de forma única por las variables  $\{r_i\}$  y por las marcas o pesos dado por los residuos

$$\left( \frac{Y_i}{\Delta} - \mu(r_i, \theta_0) \right), \quad 1 \leq i \leq n.$$

El proceso  $R_n^d(x)$  toma valores en el espacio de Skorokhod,  $D(-\infty, \infty)$  y es tal que

$$R_n^d(-\infty) = 0 \quad \text{y} \quad R_n^d(+\infty) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{Y_i}{\Delta} - \mu(r_i, \theta_0) \right).$$

Por lo tanto,  $R_n^d$  es un proceso en  $D[-\infty, \infty]$ . Como ya se mencionó, bajo los supuestos (1),(2) y (6) se garantiza que el proceso  $r_t$  es ergódico, que la distribución estacionaria  $F_\theta$  del proceso  $r_t$  es continua y además satisface

$$\mathbb{E} \left[ \left( \frac{Y_0}{\Delta} - \mu(r_0, \theta_0) \right) \middle| r_0 \right] = 0 \quad (4.5)$$

y

$$\mathbb{E} \left[ \left( \frac{Y_0}{\Delta} - \mu(r_0) \right)^2 \right] < \infty \quad (4.6)$$

Por otro lado, la media del proceso  $R_n^d(x)$  es

$$\begin{aligned} \mathbb{E} (R_n^d(x)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left( \mathbb{1}_{\{r_i \leq x\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[ \mathbb{E} \left( \mathbb{1}_{\{r_i \leq x\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i \middle| r_i \right) \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[ \mathbb{1}_{\{r_i \leq x\}} \frac{\sigma}{\sqrt{\Delta}} \mathbb{E}(\varepsilon_i | r_i) \right] \\ &= 0 \end{aligned}$$

y función de covarianzas dada por

$$\begin{aligned}
\text{Cov}(R_n^d(x_1), R_n^d(x_2)) &= \frac{1}{n} \text{Cov} \left( \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x_1\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i, \sum_{j=1}^n \mathbb{1}_{\{r_{j-1} \leq x_2\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_j \right) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j=i}^n \text{Cov} \left( \mathbb{1}_{\{r_i \leq x_1\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i, \mathbb{1}_{\{r_{j-1} \leq x_2\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_j \right) \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left( \mathbb{1}_{\{r_i \leq x_1\}} \mathbb{1}_{\{r_{j-1} \leq x_2\}} \frac{\sigma^2}{\Delta} \varepsilon_i \varepsilon_j \right) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left( \mathbb{E} \left[ \mathbb{1}_{\{r_i \leq \min\{x_1, x_2\}\}} \frac{\sigma^2}{\Delta} \varepsilon_i^2 \middle| r_i \right] \right) \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left( \mathbb{1}_{\{r_i \leq \min\{x_1, x_2\}\}} \frac{\sigma^2}{\Delta} \mathbb{E} [\varepsilon_i^2 | r_i] \right) \\
&= \mathbb{E} \left( \mathbb{1}_{\{r_0 \leq \min\{x_1, x_2\}\}} \frac{\sigma^2}{\Delta} \right) = \mathbb{K}(x_1, x_2)
\end{aligned}$$

El teorema (4.1) descrito más adelante, establece la convergencia débil del proceso  $R_n^d$ , bajo la hipótesis nula simple, a un proceso Gaussiano de media cero y función de covarianza  $K(x, y)$ .

Considérese ahora, la hipótesis nula compuesta

$$H_0 : \mu(\cdot) = \mu(\cdot, \theta) \quad \text{para algún } \theta \in \Theta.$$

Sea  $\hat{\theta}$  un estimador tal que, bajo la hipótesis  $H_0$ ,  $\hat{\theta}$  satisface condiciones del tipo  $\sqrt{n}$ -consistente. Entonces se define el proceso

$$R_{n, \hat{\theta}}^d(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{Y_i}{\Delta} - \mu(r_i, \hat{\theta}) \right) \mathbb{1}_{\{r_{i-1} \leq x\}}, \quad x \in \mathbb{R} \quad (4.7)$$

Similar al proceso  $R_n^d$ , el proceso  $R_{n, \hat{\theta}}^d$  es un proceso empírico marcado, con marcas o pesos en  $r_{i-1}$ , dados por los residuos

$$\left( \frac{Y_i}{\Delta} - \mu(r_i; \hat{\theta}) \right).$$

Como antes, el objetivo es establecer la convergencia débil del proceso  $R_{n, \hat{\theta}}^d$ . En la próxima sección el teorema (x) presenta una demostración para obtener dicha convergencia.

## 4.1 Estudio de la Convergencia

A continuación se presentan dos teoremas mediante los cuales se establecen la consistencia de los procesos empíricos  $R_n^d$  y  $R_{n, \theta}^d$  presentados en la sección anterior. El primer teorema, estudia la convergencia del proceso empírico  $R_n^d$  mediante el cual se define el estadístico del test para la forma paramétrica del drift bajo la hipótesis nula simple.

*Teorema 4.1.* Asuma que se satisfacen (4.5) y (4.6). Entonces las distribuciones finito-dimensionales de  $R_n^d$  convergen débilmente a un proceso continuo Gaussiano  $R_\infty$  con función de covarianza dada por

$$\mathbb{K}(x, y) = \mathbb{E} \left( \mathbb{1}_{\{r_0 \leq \min\{x, y\}\}} \frac{\sigma^2}{\Delta} \right)$$

y

$$R_n^d \Rightarrow R_\infty^d \quad \text{en el espacio } D[-\infty, \infty]. \quad (4.8)$$

El segundo teorema estudia el comportamiento de asintótico del proceso  $R_{n,\theta}$  bajo la hipótesis nula compuesta. Para ello, es necesario establecer algunas suposiciones adicionales sobre el estimador  $\hat{\theta}$ . Se asumirán los siguientes supuestos

*Supuesto 8.* El estimador  $\hat{\theta}$  satisface

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, i(\theta)^{-1}),$$

donde  $i(\theta) \equiv \mathbb{E}_\theta \left\{ \dot{\ell}_1(\theta) \dot{\ell}_1(\theta)^T \right\}$  existe y es positiva-definida. La función  $\dot{\ell}_1$  denota la derivada respecto de  $\theta$  de la función dada en (2.54) para  $i = 1$ . La información de Fisher  $i(\theta)$  se puede obtener de la estacionariedad del proceso  $r_t$  la cual se deriva de las suposiciones (1),(2) y (6).

*Supuesto 9.* Existen constantes  $C, k > 0$  tales que para cualquier  $\theta$  y  $r$ ,

$$\begin{aligned} \sup_{\theta' \in N_\theta} \|\dot{\mu}(r, \theta')\| &\leq C(1 + |r|^k), \\ \sup_{\theta' \in N_\theta} \|\ddot{\mu}(r, \theta')\| &\leq C(1 + |r|^k), \end{aligned}$$

donde

$$\dot{\mu} = \frac{\partial \mu}{\partial \theta}, \quad \ddot{\mu} = \frac{\partial^2 \mu}{\partial \theta \partial \theta^T},$$

y  $N_\theta$  es una vecindad abierta de  $\theta$ .

Los supuestos (8) y (9) son necesarios para establecer la convergencia uniforme de ciertos términos relacionados con el proceso empírico en estudio.

*Teorema 4.2.* Asuma que se satisfacen (8) y (9). Entonces, bajo la hipótesis nula compuesta se tiene que, uniformemente en  $x$ ,

$$R_{n,\hat{\theta}}^d(x) = R_n^d(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_i \leq x\}} \dot{\mu}(r_i, \theta) i(\theta)^{-1} n^{-\frac{1}{2}} V_n(\theta) + o_p(1),$$

con  $V_n(\theta) = \sum_{i=1}^n \dot{\ell}_i(\theta)$ , la función score. Entonces

$$R_{n,\hat{\theta}}^d \Rightarrow R_{\infty,\hat{\theta}}^d \quad \text{en distribución en el espacio } D[-\infty, \infty],$$

donde  $R_{\infty,\hat{\theta}}$  es un proceso Gaussiano de media cero y función de covarianza,

$$\begin{aligned} \mathbb{K}_\theta(x_1, x_2) &= \mathbb{K}(x_1, x_2) + \mathbb{E} \left\{ \vartheta(x_1, \theta) i^{-1}(\theta) n \mathbb{E} \left[ \dot{\ell}_1(\theta) \left( \dot{\ell}_1(\theta) \right)^T \right] [i^{-1}(\theta)]^T [\vartheta(x_2, \theta)]^T \right\} \\ &\quad - \frac{\sigma}{\sqrt{\Delta}} \mathbb{M}^T(x_2, \theta) \mathbb{E} \left\{ \varepsilon_1 \mathbf{1}_{\{r_0 \leq x_1\}} i^{-1}(\theta) \ln(p_\theta(\Delta, r_1, r_0)) \right\} \\ &\quad - \frac{\sigma}{\sqrt{\Delta}} \mathbb{M}^T(x_1, \theta) \mathbb{E} \left\{ \varepsilon_1 \mathbf{1}_{\{r_0 \leq x_2\}} i^{-1}(\theta) \ln(p_\theta(\Delta, r_1, r_0)) \right\}, \end{aligned} \quad (4.9)$$

donde

$$\mathbb{M}(z, \theta) = (\mathbb{M}_1(z, \theta), \dots, \mathbb{M}_n(z, \theta))^T, \quad \text{con} \quad \mathbb{M}_j(z, \theta) = \mathbb{E} \left( \dot{\mu}_j(r_0, \theta) \mathbf{1}_{\{r_0 \leq z\}} \right) \quad \text{y}$$

$$\dot{\mu}_j(r_0, \theta) = \frac{\partial \mu(r_0, \theta)}{\partial \theta_j} \quad j = 1, \dots, d.$$

## 4.2 Demostraciones

*Demostración.* (Teorema 4.1)

La prueba de la convergencia débil del proceso  $R_n^d$  al proceso  $R_\infty^d$  consiste de dos partes:

1. Probar la convergencia débil de la distribuciones finito-dimensionales de  $R_n^d$  a  $R_\infty^d$
2. Probar el *tightness* de  $R_n^d$ .

De (4.5) y (4.6) se deduce que el proceso  $R_n^d$  es una martingala cuadrado integrable de media cero para cada  $x \in \mathbb{R}$ . Del Teorema del Límite Central para Martingalas (ver Hall y Heyde (1980), Corolario 3.1.) se obtiene la convergencia de las distribuciones finito-dimensionales del proceso  $R_n^d$ . En consecuencia, la función de covarianza es dada por

$$\mathbb{K}(x, y) = \mathbb{E} \left( \mathbf{1}_{\{r_0 \leq \min\{x, y\}\}} \frac{\sigma^2}{\Delta} \right)$$

Para probar el *tightness* del proceso  $R_n^d$ , se aplicará el criterio de Chentsov (ver Billingsley (1961), Teorema 15.6). Para  $-\infty \leq x_1 < x_2 < x_3 \leq \infty$  fijos

$$\begin{aligned} & \mathbb{E} \left\{ [R_n^d(x_3) - R_n^d(x_2)]^2 [R_n^d(x_2) - R_n^d(x_1)]^2 \right\} = \\ &= \mathbb{E} \left\{ \frac{1}{n^2} \left[ \sum_{i=1}^n \mathbf{1}_{\{x_2 < r_i \leq x_3\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i \right]^2 \left[ \sum_{i=1}^n \mathbf{1}_{\{x_1 < r_i \leq x_2\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i \right]^2 \right\} \\ &= \mathbb{E} \left\{ \frac{1}{n^2} \left[ \sum_{i,j,k,l} U_i U_j V_k V_l \right] \right\}, \end{aligned}$$

donde

$$U_i = \mathbf{1}_{\{x_2 < r_i \leq x_3\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i \quad y \quad V_i = \mathbf{1}_{\{x_1 < r_i \leq x_2\}} \frac{\sigma}{\sqrt{\Delta}} \varepsilon_i.$$

Obsérvese que,

- $\mathbb{E}\{U_i U_j V_k V_l\} = 0$ , si el índice más grande entre  $i, j, k, l$  no coincide con ningún otro.
- $U_i V_i \equiv 0$ , puesto que  $(x_1, x_2] \cap (x_2, x_3] = \emptyset$

Entonces, se obtiene

$$\mathbb{E} \left\{ \frac{1}{n^2} \left[ \sum_{i,j,k,l} U_i U_j V_k V_l \right] \right\} = \frac{1}{n^2} \sum_{i,j < k} \mathbb{E} \{V_i V_j U_k^2\} + \frac{1}{n^2} \sum_{i,j < k} \mathbb{E} \{U_i U_j V_k^2\} \quad (4.10)$$

Los siguientes pasos y argumentos son análogos a los del Lema 3.1 en Koul y Stute (1999). Además, en esta prueba, las constantes  $K$  y  $\tilde{K}$  son genéricas, y pueden variar de expresión en expresión, sin embargo éstas no dependen de  $n$  o de las  $x'_i$ s. Considérese la primera suma en la ecuación (4.10) para  $2 \leq k \leq n$  fijo

$$\begin{aligned} \sum_{i,j < k} \mathbb{E} \{V_i V_j U_k^2\} &= \mathbb{E} \left\{ \left( \sum_{i=1}^{k-1} V_i \right)^2 U_k^2 \right\} \\ &= \mathbb{E} \left\{ \mathbb{E} \left[ \left( \sum_{i=1}^{k-1} V_i \right)^2 U_k^2 \middle| \mathcal{F}_k \right] \right\} \\ &= \mathbb{E} \left\{ \left( \sum_{i=1}^{k-1} V_i \right)^2 \mathbb{E} (U_k^2 | \mathcal{F}_k) \right\} \\ &\leq 2 \mathbb{E} \left\{ \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{E} (U_k^2 | \mathcal{F}_k) \right\} + 2 \mathbb{E} \{V_{k-1}^2 \mathbb{E} (U_k^2 | \mathcal{F}_k)\} \quad (4.11) \end{aligned}$$

La primera ecuación en (4.11)

$$\begin{aligned}
\mathbb{E} \left\{ \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{E} (U_k^2 | \mathcal{F}_k) \right\} &= \mathbb{E} \left\{ \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{1}_{\{x_2 < r_k \leq x_3\}} \frac{\sigma^2}{\Delta} \mathbb{E} (\varepsilon_k^2 | \mathcal{F}_k) \right\} \\
&= \mathbb{E} \left\{ \mathbb{E} \left[ \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{1}_{\{x_2 < r_k \leq x_3\}} \frac{\sigma^2}{\Delta} \middle| \mathcal{F}_{k-1} \right] \right\} \\
&= \mathbb{E} \left\{ \frac{\sigma^2}{\Delta} \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{E} [\mathbb{1}_{\{x_2 < r_k \leq x_3\}} | \mathcal{F}_{k-1}] \right\} \\
&= \mathbb{E} \left\{ \frac{\sigma^2}{\Delta} \left( \sum_{i=1}^{k-2} V_i \right)^2 \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\} \right\} \\
&= \mathbb{E} \left\{ \frac{\sigma^2}{\Delta} \left( \sum_{i=1}^{k-2} V_i \right)^2 \right\} \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\} \\
&\leq \frac{\sigma^4}{\Delta^2} K(k-2) [F_\theta(x_2) - F_\theta(x_1)] \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\} \quad (4.12)
\end{aligned}$$

donde  $\mathbb{P}$  es la probabilidad de transición del proceso  $r_t$ . Por otro lado, el segundo término en (4.11)

$$\begin{aligned}
\mathbb{E} \left\{ (V_{k-1})^2 \mathbb{E} (U_k^2 | \mathcal{F}_k) \right\} &= \mathbb{E} \left\{ V_{k-1}^2 \mathbb{1}_{\{x_2 < r_k \leq x_3\}} \mathbb{E} \left( \frac{\sigma^2}{\Delta} \varepsilon_k^2 \middle| \mathcal{F}_k \right) \right\} \\
&= \frac{\sigma^2}{\Delta} \mathbb{E} \left\{ \mathbb{E} [V_{k-1}^2 \mathbb{1}_{\{x_2 < r_k \leq x_3\}} | \mathcal{F}_{k-1}] \right\} \\
&= \frac{\sigma^2}{\Delta} \mathbb{E} \left\{ \mathbb{1}_{\{x_1 < r_{k-1} \leq x_2\}} \mathbb{E} \left[ \mathbb{1}_{\{x_2 < r_k \leq x_3\}} \frac{\sigma^2}{\Delta} \varepsilon_{k-1}^2 \middle| \mathcal{F}_{k-1} \right] \right\} \\
&\leq K \frac{\sigma^4}{\Delta^2} [F_\theta(x_2) - F_\theta(x_1)] \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\}. \quad (4.13)
\end{aligned}$$

Combinando las expresiones anteriores (4.12) y (4.13) y sumando desde  $k = 2$  hasta  $k = n$ , se obtiene

$$\frac{1}{n^2} \sum_{i,j < k} \mathbb{E} \{V_i V_j U_k^2\} \leq K [F_\theta(x_2) - F_\theta(x_1)] \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\}.$$

obtiene del segundo término en (4.11). En resumen, se puede ver que

$$\begin{aligned}
\mathbb{E} \left\{ \frac{1}{n^2} \left[ \sum_{i,j,k,l} U_i U_j V_k V_l \right] \right\} &\leq K [F_\theta(x_2) - F_\theta(x_1)] \mathbb{P} \{x_2 < r_k \leq x_3 | r_{k-1}\} + \\
&\quad + \tilde{K} [F_\theta(x_3) - F_\theta(x_2)] \mathbb{P} \{x_1 < r_k \leq x_2 | r_{k-1}\} \quad (4.14)
\end{aligned}$$

De la expresión (4.14) se puede ver que las sumas en (4.10) satisfacen el criterio de Chentsov para el *tightness*, esto completa la prueba. Para detalles relevantes ver Billingsley (1961).  $\square$

*Demostración.* (Teorema 4.2)

Considérese el proceso

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x\}} \left( \frac{Y_i}{\Delta} - \mu(r_i, \hat{\theta}) \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x\}} \left( \frac{Y_i}{\Delta} - \mu(r_i, \hat{\theta}) - \mu(r_i, \theta) + \mu(r_i, \theta) \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x\}} \left\{ \left( \frac{Y_i}{\Delta} - \mu(r_i, \theta) \right) - \left( \mu(r_i, \hat{\theta}) - \mu(r_i, \theta) \right) \right\},
\end{aligned}$$



Aplicando el Teorema de Taylor y bajo los supuestos (8) y (9), se deduce que

$$\mu(r_i, \hat{\theta}) - \mu(r_i, \theta) = \dot{\mu}(r_i, \theta)(\hat{\theta} - \theta) + o_{\mathbb{P}}(1)$$

y por lo tanto se obtiene el proceso asintóticamente equivalente denotado por

$$R_{n,\theta}^d(x) = R_n^d(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x\}} \dot{\mu}(r_i, \theta)(\hat{\theta} - \theta).$$

Bajo el supuesto (8), se obtiene

$$R_{n,\theta}^d(x) = R_n^d(x) - \vartheta(x, \theta) i(\theta)^{-1} n^{-\frac{1}{2}} V_n(\theta) + o_{\mathbb{P}}(1) \quad (4.15)$$

con

$$\vartheta(x, \theta) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{r_i \leq x\}} \dot{\mu}(r_i, \theta).$$

Nótese que el proceso  $R_{n,\theta}^d$  en (4.15) consta de dos términos. El primero, corresponde al proceso  $R_n^d$ . Es claro que del teorema (4.1), dicho proceso converge débilmente a un proceso límite, y además es *TeoConsist*. Respecto del segundo término en (4.15), por la Ley de los Grandes Números  $\vartheta(x, \theta)$  converge uniformemente a una función no aleatoria en  $x$ . El *tightness* del segundo término se obtiene de la normalidad asintótica dada por el supuesto (8) y de la convergencia uniforme de  $\vartheta(x, \theta)$ . Por lo tanto, el proceso  $R_{n,\theta}^d$  converge débilmente a un proceso Gaussiano de media cero y función de covarianzas dada por  $K_\theta$ .  $\square$

### 4.3 Consistencia del Proceso Bootstrap

Bajo la hipótesis nula  $H_0$  y la suposición de que la función de difusión es constante  $\sigma(r_t, \theta) = \sigma$ , el proceso definido en (4) es dado por

$$Y_{t_i} = \mu(r_{t_i}, \theta) \Delta + \sigma \sqrt{\Delta} \varepsilon_{t_i}, \quad (4.16)$$

donde  $\{\varepsilon_{t_i}\}$  son variables aleatorias independientes normales estándar e independientes de  $r_{t_i}$ . Por otro lado el proceso obtenido a partir del remuestreo bootstrap es dado por

$$Y_{t_i}^* = \mu(r_{t_i}, \hat{\theta}) \Delta + \sigma \sqrt{\Delta} \varepsilon_{t_i}^*, \quad (4.17)$$

donde  $\{\varepsilon_{t_i}^*\}$  son variables aleatorias independientes normales estándar e independientes de  $r_{t_i}$ .

Bajo las suposiciones 8 and 9, el proceso en (4.17) se puede aproximar por

$$Y_{t_i}^* = \mu(r_{t_i}, \theta) \Delta + \sigma \sqrt{\Delta} \varepsilon_{t_i}^* + O_{\mathbb{P}}(n^{-1/2}). \quad (4.18)$$

En consecuencia, como la remuestra bootstrap  $\{(Y_{t_i}, r_{t_i})\}$  imita el comportamiento de la muestra original  $\{(Y_{t_i}, r_{t_i})\}$ , la convergencia débil del proceso

$$R_n^{d*} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}^*}{\Delta} - \mu(r_{t_i}, \hat{\theta}^*) \right)$$

se obtiene de la convergencia débil del proceso  $R_n^d(x)$ .

### III

## Otros Desarrollos Notables

# 5

## Desarrollos en Tests de Bondad de Ajuste basado en Procesos Empíricos

Este capítulo tiene como objetivo introducir algunos desarrollos del test de bondad de ajuste relativos al: problema de comparación de curvas de tipos de interés, test para un proceso de difusión con saltos en el que el tamaño del salto se rige por una distribución normal, y una aplicación gráfica mediante el método SiZer desarrollado por [Chaudhuri y Marron \(1999\)](#). La Sección 5.1 presenta los tres enfoques para tests de bondad de ajuste para la comparación de curvas de tipos de interés. La Sección 5.2 hace una breve referencia al test de bondad de ajuste basado en procesos empíricos para modelos de difusión con saltos, en el que se considera un tipo particular de dichos modelos. Finalmente la Sección 5.3 presenta una aplicación gráfica del test de bondad de ajuste para los modelos de tipos de interés.

### 5.1 Tests de Bondad de Ajuste para la Comparación de Curvas de tipos de Interés

Uno de los problemas de mayor importancia en la inferencia estadística es el de comparación de dos o más grupos de variables. Esta comparación puede ser llevada a cabo mediante la comparación de las medias, medianas o cualquier otra característica de la variable de interés. En el contexto de modelos de regresión, cuando la variable es acompañada por una covariable  $X$ , el objetivo es comparar las funciones de regresión

$$m_k(x) = \mathbb{E}(Y_k|X = x), \quad k = 1, 2, \dots, n$$

asociadas a cada grupo.

Si la especificación de la función de regresión en los distintos grupos es paramétrica,  $m_k = m_{\theta_k}$ , con  $\theta_k \in \Theta \subset \mathbb{R}^d$ , se tiene el análisis de covarianza clásico. En un contexto no-paramétrico el objetivo es el de comparar las curvas de regresión estimadas de forma no-paramétrica. Sea  $(X_i, Y_j)$  para  $j = 1, \dots, k$  vectores aleatorios independientes tal que satisfacen el siguiente modelo

de regresión no-paramétrico

$$Y_j = m_j(X_j) + \sigma_j(X_j)\varepsilon_j, \quad j = 1, \dots, k, \quad j = 1, \dots, n_i$$

donde

- $\varepsilon_j$  - variables error, con distribución  $F_{\varepsilon_j}$ , independientes de  $X_j$ ,
- $m_j(X_j) = \mathbb{E}(Y_j|X_j)$  - la función de regresión (desconocida).
- $\sigma_j^2(X_j) = \text{Var}(Y_j|X_j)$ - función de varianza condicional.

Supóngase además que las variables  $X_j$  tienen soporte común  $R_X$  y sea  $(X_{ij}, Y_{ij})$ ,  $i = 1, \dots, n_j$ , con  $n = \sum_{j=1}^k n_j$ , una muestra i.i.d. de la distribución de  $(X_j, Y_j)$ , para  $j = 1, \dots, k$ . Entonces el problema se centra en probar la hipótesis nula  $H_0$  de la igualdad de las funciones de regresión versus la alternativa  $H_a$ :

$$H_0 : m_1 = \dots = m_k, \quad \text{vs} \quad H_a : \exists(i, j) \text{ such that } m_i \neq m_j.$$

Este problema ha sido estudiado ampliamente en la literatura y es abordado desde diferentes perspectivas.

- Aplicando técnicas de suavizado:
  - Härdle y Marron (1990),
  - Hall y Hart (1990),
  - Kulasekera (1995)
  - Dette y Neumeyer (2001)
  - Vilar-Fernández y González-Manteiga (2004)

entre muchas otras..

- Aplicando procesos empíricos :
  - Delgado (1993).
  - Neumeyer y Dette (2003).
  - Ferreira y Stute (2004).
  - Pardo-Fernández et al. (2007).
  - Dette y Weißbach (2009).

En esta sección se introduce un test de bondad de ajuste para comparar las curvas de tipos de interés motivado en las ideas de los test de comparación de curvas de regresión basados en procesos empíricos.

Considérese dos modelos de difusión que caracterizan la dinámica de los tipos de interés para dos curvas distintas,

$$\begin{aligned} dr_{t1} &= \mu_1(r_{t1})dt + \sigma_1(r_{t1})dW_{t1} \\ dr_{t2} &= \mu_2(r_{t2})dt + \sigma_2(r_{t2})dW_{t2}, \end{aligned}$$

donde, para  $j = 1, 2$ :

- $W_{tj}$  son procesos de Wiener independientes,
- $\mu_j(r_{tj})$  y  $\sigma_j(r_{tj})$  son las funciones drift y difusión respectivamente, que satisfacen

$$\mu_j(r_{t_j}) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E} (r_{(t+\Delta)_j} - r_{t_j} \mid r_{t_j}),$$

y

$$\sigma_j^2(r_{t_j}) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \mathbb{E} ((r_{(t+\Delta)_j} - r_{t_j})^2 \mid r_{t_j}),$$

Como ya es habitual se aplicará el método de Euler para obtener la versión discreta de los modelos antes descritos,

$$r_{(t_i+1)_j} - r_{t_{ij}} = \mu_j(r_{t_{ij}})\Delta + \sigma_j(r_{t_{ij}})(W_{(t_i+1)_j} - W_{t_{ij}}), i = 0, 1, \dots, n, \quad j = 1, 2,$$

en el intervalo  $[t_0, T]$  para un  $\Delta$  fijo,  $t_i = i\Delta$ . Entonces para  $Y_{ij} = r_{(t_i+1)_j} - r_{t_{ij}}$  se obtienen los modelos de la forma

$$\begin{aligned} Y_{i1} &= \mu_1(r_{i1})\Delta + \sigma_1(r_{i1})\varepsilon_{i1}, \quad i = 1, \dots, n_1 \\ Y_{i2} &= \mu_2(r_{i2})\Delta + \sigma_2(r_{i2})\varepsilon_{i2} \quad i = 1, \dots, n_2, \end{aligned}$$

donde se denotará por  $r_{ij} = r_{t_{ij}}$ , y  $\varepsilon_{ij} = W_{(t_i+1)_j} - W_{t_{ij}} \sim N(0, \Delta)$ . El objetivo principal es contrastar la hipótesis de la igualdad de las funciones drift

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_a : \mu_1 \neq \mu_2$$

Para ello, se consideran tres enfoques basados en procesos empíricos y que se presentan a continuación.

### 5.1.1 Primer Enfoque: Comparación de la Distribución de los Residuos

En esta sección se propone un test de bondad de ajuste basado en procesos empíricos para la comparación de las curvas de tipos de interés mediante la comparación de las funciones drift. El test que se describe a continuación es una adaptación para procesos de difusión del test propuesto en [Pardo-Fernández et al. \(2007\)](#) para la igualdad de curvas de regresión. El procedimiento es el siguiente:

- Se estiman para  $j = 1, 2$ , and the  $i = 1, \dots, n_j$  los errores

$$\varepsilon_{ij} = \frac{Y_{ij} - \hat{\mu}_j(r_{ij})\Delta}{\hat{\sigma}_j(r_{ij})},$$

con

$$\hat{\mu}_j(r) = \frac{1}{\Delta} \sum_{i=1}^{n_j} \omega_{i,j}(r, h) Y_{ij}$$

$$\hat{\sigma}_j^2(r) = \frac{1}{\Delta} \sum_{i=1}^{n_j} \omega_{i,j}(r, h) Y_{ij}^2$$

los estimadores no-paramétricos del drift y la volatilidad respectivamente, (véase [Gao \(2007\)](#))

y

$$\omega_{i,j}(r, h) = \frac{K\left(\frac{r-r_{ij}}{h}\right)}{\sum_{k=1}^{n_j} K\left(\frac{r-r_{kj}}{h}\right)}$$

son los pesos Nadaraya-Watson,  $K$  es la función kernel (en este caso kernel Gaussian),  $h$  es un parámetro de suavizado apropiado y

$$\begin{aligned} \hat{\pi}(r) &= \frac{1}{n_j h} \sum_{i=1}^{n_j} K\left(\frac{r-r_{ij}}{h}\right) \\ &= \frac{1}{\sqrt{2\pi} n_j h} \sum_{i=1}^{n_j} \exp\left\{-\frac{(r-r_{ij})^2}{2h^2}\right\} \end{aligned}$$

el estimador no-paramétrico de la densidad marginal del proceso  $r_t$ .

- Se estiman para  $j = 1, 2$  y para  $i = 1, \dots, n_j$

$$\varepsilon_{ij0} = \frac{Y_{ij} - \hat{\mu}(r_{ij})\Delta}{\hat{\sigma}_j(r_{ij})}$$

con

$$\hat{\mu}(r) = \frac{1}{\Delta} \sum_{j=1}^2 \sum_{i=1}^{n_j} W_{ij}(r, h) Y_{ij}$$

la función de regresión común  $m(\cdot)$  bajo la hipótesis nula, y

$$W_{ij}(r, h) = \frac{K\left(\frac{r-r_{ij}}{h}\right)}{\sum_{l=1}^2 \sum_{k=1}^{n_l} K\left(\frac{r-r_{kl}}{h}\right)}$$

- Se estima para  $j = 1, 2$ ,

$$\hat{F}_{\varepsilon_j}(y) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{1}_{\{\varepsilon_{ij} \leq y\}}, \quad y \quad \hat{F}_{\varepsilon_j0}(y) = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{1}_{\{\varepsilon_{ij0} \leq y\}}$$

la distribución empírica de los errores y la distribución empírica de los errores bajo la hipótesis nula.

- Se aplica el proceso

$$\widehat{\mathbf{W}}(y) = \left( \widehat{W}_1(y), \widehat{W}_2(y) \right)^t, \quad -\infty < y < \infty,$$

donde,

$$\widehat{W}_j(y) = (n_j)^{\frac{1}{2}} \left( \hat{F}_{\varepsilon_j0}(y) - \hat{F}_{\varepsilon_j}(y) \right) \quad \text{for } j = 1, 2.$$

para comparar los dos estimadores de la distribución de los errores en cada población aplicando los funcionales del tipo Kolmogorov-Smirnov y Crámer-von Misses para obtener los correspondientes estadísticos

$$T_{KS} = \sum_{j=1}^k \sup |\widehat{W}_j(y)| \quad \text{y} \quad T_{CM} = \sum_{j=1}^k \int \widehat{W}_j^2(y) d\hat{F}_{\varepsilon_j0}(y).$$

Bajo la hipótesis nula  $H_0$ , ambos estimadores aproximan la distribución del error  $F_{\varepsilon_j}$ . Si la hipótesis nula no es cierta, entonces existirá diferencias significativas en los estimados de las funciones drift. En la practica se suele aproximar la distribución del estadístico aplicando técnicas bootstrap, con lo cual se obtienen los cuantiles para establecer un patrón de decisión del test. Entonces, el procedimiento bootstrap se lleva a cabo de la siguiente forma:

### Bootstrap Procedure

Sea  $b$  el índice del número de replicas bootstrap, es decir,  $b = 1, \dots, B$

- Para  $j = 1, 2$ , sea  $\{\varepsilon_{ij,b}^*\}$  una muestra independiente e idénticamente distribuida de la distribución Normal con varianza variance  $\Delta$ .
- Para  $j = 1, 2$  y  $i = 1, \dots, n_j$ , genere nuevas variables respuesta bajo la hipótesis nula

$$Y_{ij,b}^* = \hat{\mu}(r_{ij})\Delta + \hat{\sigma}_j(r_{ij})\varepsilon_{ij,b}^*$$

- Sea  $T_{KS,b}^*$  y  $T_{CVM,b}^*$  los estadísticos obtenidos a partir de la replicas bootstraps

$$\{(r_{ij}, Y_{ij}^*), i = 1, \dots, n_j\}$$

,  $j = 1, 2$ .

- Entonces

$$T_{KS,((1-\alpha)B)}^* \quad \text{y} \quad T_{CVM,((1-\alpha)B)}^*$$

aproximan los  $(1 - \alpha)$ -cuantiles de la distribución de  $T_{KS}$  y  $T_{CVM}$  bajo la hipótesis nula, respectivamente.

### Selección del parámetro de Suavizado

Por simplicidad el parámetro de suavizado  $h$  será el mismo para todos los valores a estimar,  $\hat{\mu}_j(r)$ ,  $\hat{\sigma}_j(r)$  y  $\hat{\mu}(r)$ , véase section 5 de [Pardo-Fernández et al. \(2007\)](#). La regla para la selección del ancho de banda óptimo puede ser encontrada en [Gao \(2007\)](#).

### 5.1.2 Segundo Enfoque: A partir de un particular Proceso Empírico Marcado

Este enfoque está basado en las ideas de [Dette y Weißbach \(2009\)](#) para la especificación de dos series de tiempo. En el caso de los procesos de difusión para tipos de interés, se formula el problema de comparación de las funciones drift

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_a : \mu_1 \neq \mu_2$$

de la siguiente manera:

Sea

$$R_{n_1 n_2}(x) = \frac{1}{n_1 n_2 h} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} (Y_{i1} - Y_{k2}) K \left( \frac{r_{i1} - r_{k2}}{h} \right) \mathbb{1}_{\{r_{i1} \leq x\}} \mathbb{1}_{\{r_{k2} \leq x\}}$$

el proceso empírico marcado, el cual se usa para probar la hipótesis de igualdad de las funciones drift. Aquí  $K : \mathbb{R} \rightarrow \mathbb{R}$  es un kernel simétrico y  $h$  es el ancho de banda tal que converge a cero 0 con tamaño de muestra creciente. La hipótesis se rechaza para valores significativamente grandes del estadístico basado en el funcional de Kolmogorov-Smirnov

$$T_{n_1, n_2} = \sup_x |R_{n_1, n_2}(x)|.$$

La distribución límite del estadístico se puede aproximar mediante técnicas bootstrap. Se generan observaciones bootstrap a partir de

$$\begin{aligned} Y_{i1}^* &= \hat{\mu}(r_{i1})\Delta + \hat{\sigma}_1 \varepsilon_{i1}^*, \quad t = 2, \dots, n_1 \quad \text{and} \\ Y_{i2}^* &= \hat{\mu}(r_{i2})\Delta + \hat{\sigma}_2 \varepsilon_{i2}^*, \quad t = 2, \dots, n_2 \end{aligned}$$

con  $\{\varepsilon_{ij}^*, b\}$  variables aleatorias independientes normales de media cero y varianza  $\Delta$ , y se calcula el proceso

$$R_{n_1 n_2}^*(x) = \frac{1}{n_1 n_2 h} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} (Y_{i1}^* - Y_{k2}^*) K \left( \frac{r_{i1} - r_{k2}}{h} \right) \mathbb{1}_{\{r_{i1} \leq x\}} \mathbb{1}_{\{r_{k2} \leq x\}}$$

para obtener el estadístico análogo bootstrap basado en el funcional Kolmogorov-Smirnov  $T_{n_1 n_2}^*$ ,

$$T_{n_1 n_2}^* = \sup_x |R_{n_1 n_2}^*(x)|.$$

Las funciones  $\mu$  y  $\sigma_j$  se estiman de acuerdo a los estimadores introducidos en la sección anterior  $\hat{\mu}$  y  $\hat{\sigma}_j$ . Se define

$$\hat{p}^* = \mathbb{P}(T_{n_1 n_2}^* > T_{n_1 n_2} | \mathcal{Y}_{n_1 n_2})$$

como el correspondiente  $p$ -valor bootstrap, donde

$$\mathcal{Y}_{n_1 n_2} = \{Y_{i1}, Y_{i2}, r_{i1}, r_{i2}, i = 1, \dots, n_j, j = 1, 2\}$$

denota la muestra en su totalidad. La hipótesis nula de la igualdad de las funciones drift se rechaza siempre que

$$\hat{p}^* < \alpha$$

donde  $\alpha \in (0, 1)$  denota el nivel nominal.

### Selección del parámetro de suavizado

El ancho de banda se selecciona de acuerdo a la regla a la regla empírica

$$h = \left\{ \frac{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2}{(n_1 + n_2)^2} \right\}^{1/5},$$

donde  $\hat{\sigma}_1$  y  $\hat{\sigma}_2$  son los estimadores de la función de varianza condicional, ver [Fan y Yao \(2003\)](#). Para el remuestreo bootstrap  $\mu$  se estima usando el parámetro de suavizado  $g = h^{5/6}$ .

### 5.1.3 Tercer Enfoque: Test basado en Procesos Empíricos

Finalmente, [Neumeyer y Dette \(2003\)](#) proponen un enfoque basado en un proceso empírico para estudiar el problema de la comparación de dos curvas de regresión desde un punto de vista no-paramétrico. En el contexto de los modelos de difusión el problema el problema de probar la igualdad de las funciones drift

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_a : \mu_1 \neq \mu_2$$

se lleva a cabo a partir de dos procesos empíricos marcados

$$\hat{R}_n^{(1)}(x) = \frac{1}{n} \sum_{i=1}^{n_1} e_{i1} \mathbb{1}_{\{r_{i1} \leq x\}} - \frac{1}{n} \sum_{i=1}^{n_2} e_{i2} \mathbb{1}_{\{r_{i2} \leq x\}} \quad (5.1)$$

y

$$\hat{R}_n^{(2)}(x) = \frac{1}{n} \sum_{i=1}^{n_1} \epsilon_{i1} \mathbb{1}_{\{r_{i1} \leq x\}} - \frac{1}{n} \sum_{i=1}^{n_2} \epsilon_{i2} \mathbb{1}_{\{r_{i2} \leq x\}} \quad (5.2)$$

donde  $x \in \mathbb{R}$ ,  $\mathbb{1}_{\{\cdot\}}$  denotan la función indicador,  $n = n_1 + n_2$  denota el tamaño total de la muestra

$$\mathcal{Y}_{n_1 n_2} = \{Y_{i1}, Y_{i2}, r_{i1}, r_{i2}, i = 1, \dots, n_j, j = 1, 2\},$$

y los residuos  $e_{ij}$  y  $\epsilon_{ij}$  para  $j = 1, 2$  se definen por

$$e_{ij} = \frac{n_{3-j}}{n} (Y_{ij} - \hat{\mu}(r_{ij})) \hat{\pi}(r_{ij}) \hat{\pi}_{3-j}(r_{ij}), \quad (5.3)$$

$$\epsilon_{ij} = \frac{n}{n_i} \frac{(Y_{ij} - \hat{\mu}(r_{ij}))}{\hat{\pi}_j(r_{ij})}. \quad (5.4)$$

Los estimadores  $\hat{\mu}$ ,  $\hat{\pi}$  y  $\hat{\pi}_j$  son los estimadores de Nadaraya-Watson estimators introducidos en los enfoques anteriores.



Por lo tanto, un test para la hipótesis de igualdad de las funciones drift se puede basar en los funcionales de los criterios de Kolmogorov-Smirnov y Cramér von Mises aplicados a los procesos (5.1) y (5.2) tales que  $j = 1, 2$

$$T_{CVM} = \int_0^1 \left( \widehat{R}_n^{(j)} \right)^2(x) dx, \quad (5.5)$$

$$T_{KS} = \sup_x \left| \widehat{R}_n^{(j)}(x) \right| \quad (5.6)$$

### Remuestreo Bootstrap

En esta sección se describe el procedimiento de remuestreo basado en técnicas bootstrap. Sea  $\widehat{\mu}_g(r)$  denotando el estimador de Nadaraya-Watson de la media condicional de la muestra total definida en el primer enfoque usando un parámetro de suavizado  $g > 0$ . Entonces la remuestra bootstrap se obtiene por

$$Y_{ij}^* := \widehat{\mu}_g(r_{ij}) + \widehat{\sigma}_g(r_{ij}) \varepsilon_{ij}^* \quad (5.7)$$

con  $\varepsilon_{ij}$  variables aleatorias independientes con distribución normal de media cero y varianza  $\Delta n$  i.i.d.. Los correspondientes procesos empíricos marcados son dados por

$$\widehat{R}_n^{(1)*}(x) = \frac{1}{n} \sum_{k=1}^2 \sum_{i=1}^{n_k} (-1)^{k-1} \frac{n_{3-k}}{n} (Y_{ik}^* - \widehat{\mu}_h^*(r_{ik})) \widehat{\pi}_h(r_{ik}) \widehat{\pi}_{3-k,h}(r_{kj}) \mathbb{1}_{\{r_{ik} \leq x\}}, \quad (5.8)$$

y

$$\widehat{R}_n^{(2)*}(x) = \frac{1}{n} \sum_{k=1}^2 \sum_{i=1}^{n_k} (-1)^{k-1} (Y_{ik}^* - \widehat{\mu}_h^*(r_{ik})) \frac{1}{\widehat{\pi}_{k,h}(r_{ik})} \frac{n}{n_k} \mathbb{1}_{\{r_{ik} \leq x\}}, \quad (5.9)$$

donde  $(\cdot)^*$  denota el correspondiente elemento definido bajo la distribución bootstrap. Nótese que se usa el parámetro de suavizado  $h$  para calcular los estadísticos de los tests ( y que se indican mediante el índice  $\widehat{\mu}_h^*$  y  $\widehat{\pi}_h$  ) y un parámetro de suavizado  $g$  para calcular los residuos. Sean  $T_{KS}^{(j)*}$   $j = 1, 2$  denotando los estadísticos derivados del funcional de Kolmogorov-Smirnov en (5.5) obtenido a partir de la muestra bootstrap. Entonces la hipótesis de igualdad de las funciones drift se rechaza si  $T_{KS}^{(j)} \geq c_{1-\alpha}^{(j)*}$ , donde  $c_{1-\alpha}^{(j)*}$  denota el valor crítico obtenido a partir de la distribución bootstrap, es decir,

$$\mathbb{P} \left( T_{KS}^{(j)} \geq c_{1-\alpha}^{(j)*} | \mathcal{G}_n \right) = \alpha, \quad j = 1, 2.$$

### Selección del parámetro de suavizado

El ancho de banda se selecciona de acuerdo a la regla a la regla empírica

$$h = \left\{ \frac{n_1 \widehat{\sigma}_1^2 + n_2 \widehat{\sigma}_2^2}{(n_1 + n_2)^2} \right\},$$

donde  $\widehat{\sigma}_1$  y  $\widehat{\sigma}_2$  son los estimadores de la función de varianza condicional, ver Fan y Yao (2003). Para el remuestreo bootstrap  $\mu$  se estima usando el parámetro de suavizado  $g = h^{5/6}$ .

### 5.1.4 Estudio de Simulación

Para estudiar el desempeño de los tests propuestos, a partir del estudio del tamaño o nivel de significación y la potencia de los mismos a partir de la simulación de un modelo el cual se considera como hipótesis nula y un conjunto de perturbaciones sobre este modelo, las cuales se consideran las hipótesis alternativas.

Considérese el modelo CKLS

$$dr_t = (0.00739344 - 0.0876r_t + \rho_d(1 - r_t^{\rho_d}))dt + \sqrt{0.7791}r_t^{1.48}dW_t, \quad (5.10)$$

Es claro que  $\rho_d = 0$ , representa el modelo CKLS bajo la hipótesis nula. Los valores de los parámetros son dados Aït-Sahalia (1999), y han sido objeto de estudio a lo largo de esta memoria. Para un conjunto de  $N = 1000$  simulaciones de tamaños  $(n_1, n_2) = (400, 400), (400, 500)$  y  $(400, 600)$  y basados en  $B = 500$  replicas bootstrap para la obtención de los valores críticos, para distintos anchos de banda  $h \in \{0.009, 0.100, 0.15\}$ , se estima el tamaño del test o nivel de significación para  $\alpha = 0.01, \alpha = 0.05$  and  $\alpha = 0.10$ . Bajo las alternativas, es decir,  $\rho_d \in 0.01, 0.05, 0.09$  se estima la potencia para los distintos niveles de significación  $\alpha = 0.01, \alpha = 0.05$  and  $\alpha = 0.10$ . Los resultados se muestran en las tablas (5.1), (5.2) y (5.3), (5.4). En dichas tablas se puede apreciar, que los tests basados en los tres enfoque aproximan razonablemente bien el nivel o tamaño para los distintos niveles de significación, para el caso  $\rho_d = 0$ . A medida que se incrementa el valor de  $\rho_d$  es claro que el modelo se aleja de la hipótesis nula, por lo que, la potencia tiende a incrementarse. Dicho comportamiento es reflejado por cada uno de los tests, para los distintos enfoques. Sin embargo, es importante resaltar lo observado en la tabla (5.1), en la que se puede apreciar que la potencia crece más rápido que en los resultados mostrados en las tablas (5.2), (5.3) y (5.4) correspondientes a los enfoques 2 y 3, y que además la potencia alcanzada para los distintos valores de  $\rho_d > 0$  es mayor para el test basado en el primer enfoque en comparación con los enfoques 2 y 3. Esto indica que el test basado en los estimadores de la distribución de los errores puede resultar más eficiente.

## 5.2 Una Extensión del test de bondad de ajuste para procesos de difusión con saltos.

El proceso de difusión con saltos se introdujo en el capítulo anterior como un modelo para la hipótesis alternativa en el estudio de la potencia del test de bondad de ajuste. En los últimos años, los modelos de difusión con saltos han adquirido gran relevancia en la modelización financiera, particularmente en la modelización del tipo de interés. En lo que respecta a modelización, un trabajo en este contexto puede encontrarse en Das (2002)) o en lo que respecta a la determinación de existencia de saltos en un conjunto de datos dados el trabajo de Johanes (2004) es una referencia importante. Recientemente, Aït-Sahalia y Jacod (2009)) también propone un test para la presencia de saltos en un conjunto de datos, Aït-Sahalia et al. (2009) proponen un test de especificación para la densidad de transición de un proceso en tiempo continuo con saltos con observaciones discretas, el test está basado en la comparación un estimador no-paramétrico de la densidad de transición. En opinión de los profesionales y académicos que estudian la modelización del tipo de interés, los procesos de difusión con saltos pueden ser una alternativa más apropiada que los modelos de difusión puros. En este sentido, puede ser interesante extender el test propuesto en el capítulo 3 de esta memoria para un modelo de difusión con saltos. Puesto que los modelos de difusión difieren en sus características, dependiendo de la distribución del tamaño del salto, por ejemplo, en esta sección se propone un test de bondad de ajuste para los modelos de difusión con saltos para el caso particular en el que el tamaño del salto es regido por una distribución normal. El caso más general, en el que otras distribuciones gobiernan el comportamiento de dicho tamaño no será discutido en este trabajo.

A continuación se describe los detalles del test. Desde un punto de vista práctico, se considera el modelo de difusión en tiempo discreto que caracteriza el tipo de interés estudiado en Johanes (2004),

$$Y_{t_i} = r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)\varepsilon_{r_{t_i}} + Z_{t_{i+1}}J_{t_{i+1}}, \quad (5.11)$$

**Tabla 5.1:** Proporción de rechazos del modelo (5.10) para el test basado en el estadístico  $T_{KS}$  (Primer Enfoque).

$\rho_d$	$h$	$n_1$	$n_2$	$p$ -valor	$\hat{\alpha}_{KS}$		
					10 %	5 %	1 %
0	0.0090	400	400	0.522	0.108	0.036	0.008
			500	0.475	0.108	0.064	0.004
			600	0.477	0.116	0.048	0.008
	0.0100	400	400	0.519	0.100	0.052	0.008
			500	0.482	0.096	0.052	0.008
			600	0.482	0.120	0.040	0.004
	0.0150	400	400	0.536	0.096	0.044	0.004
			500	0.485	0.092	0.048	0.012
			600	0.494	0.124	0.052	0.008
0.01	0.009	400	400	0.4574	0.120	0.056	0.012
			500	0.4482	0.144	0.092	0.016
			600	0.4322	0.148	0.088	0.028
	0.010	400	400	0.4099	0.188	0.080	0.008
			500	0.4406	0.136	0.052	0.008
			600	0.4690	0.128	0.068	0.016
	0.015	400	400	0.4376	0.160	0.096	0.028
			500	0.4274	0.176	0.120	0.032
			600	0.4492	0.116	0.080	0.020
0.05	0.009	400	400	0.0093	0.984	0.960	0.840
			500	0.0157	0.956	0.924	0.788
			600	0.0095	0.976	0.952	0.860
	0.010	400	400	0.0137	0.964	0.928	0.788
			500	0.0142	0.964	0.932	0.800
			600	0.0153	0.952	0.920	0.784
	0.015	400	400	0.0145	0.964	0.932	0.788
			500	0.0244	0.936	0.892	0.716
			600	0.0271	0.908	0.852	0.684
0.09	0.009	400	400	0.0003	1.000	1.000	0.992
			500	0.0003	1.000	1.000	0.996
			600	0.0001	1.000	1.000	1.000
	0.010	400	400	0.0002	1.000	1.000	0.996
			500	0.0005	1.000	1.000	0.988
			600	0.0002	1.000	1.000	0.996
	0.015	400	400	0.0001	1.000	1.000	1.000
			500	0.0003	1.000	1.000	0.996
			600	0.0009	1.000	0.996	0.976

**Tabla 5.2:** Proporción de rechazos del modelo (5.10) para el test basado en el estadístico  $T_{n_1 n_2}$  (Segundo Enfoque)

$\rho_d$	$h$	$n_1$	$n_2$	$p$ -valor	$\hat{\alpha}_{KS}$		
					10 %	5 %	1 %
0	0.0090	400	400	0.484	0.104	0.060	0.020
			500	0.459	0.120	0.072	0.008
			600	0.472	0.140	0.076	0.016
	0.0100	400	400	0.494	0.108	0.048	0.016
			500	0.462	0.112	0.076	0.004
			600	0.474	0.136	0.076	0.004
	0.0150	400	400	0.520	0.080	0.040	0.008
			500	0.481	0.092	0.056	0.008
			600	0.476	0.116	0.060	0.000
0.01	0.009	400	400	0.4813	0.092	0.056	0.020
			500	0.4765	0.140	0.080	0.012
			600	0.4510	0.160	0.088	0.012
	0.010	400	400	0.4920	0.084	0.024	0.000
			500	0.4543	0.104	0.056	0.008
			600	0.4913	0.144	0.068	0.012
	0.015	400	400	0.5245	0.076	0.020	0.000
			500	0.4673	0.116	0.060	0.008
			600	0.4621	0.084	0.032	0.008
0.05	0.009	400	400	0.2766	0.420	0.308	0.168
			500	0.2050	0.504	0.368	0.180
			600	0.1862	0.552	0.388	0.188
	0.010	400	400	0.3026	0.396	0.304	0.120
			500	0.2153	0.484	0.360	0.188
			600	0.2326	0.496	0.376	0.168
	0.015	400	400	0.2287	0.448	0.340	0.140
			500	0.1966	0.512	0.352	0.176
			600	0.2119	0.488	0.336	0.152
0.09	0.009	400	400	0.2008	0.612	0.524	0.136
			500	0.0639	0.844	0.780	0.612
			600	0.0519	0.852	0.756	0.616
	0.010	400	400	0.1539	0.668	0.608	0.172
			500	0.0748	0.820	0.776	0.612
			600	0.0525	0.860	0.800	0.616
	0.015	400	400	0.1516	0.688	0.620	0.388
			500	0.0559	0.884	0.832	0.648
			600	0.0543	0.872	0.808	0.624

**Tabla 5.3:** Proporción de rechazos del modelo (5.10) para el test basado en el estadístico  $T_{KS}^{(1)}$  (Tercer Enfoque)

$\rho_d$	$h$	$n_1$	$n_2$	$p$ -valor	$\hat{\alpha}_{KS}$		
					10 %	5 %	1 %
0	0.0090	400	400	0.503	0.120	0.052	0.000
			500	0.517	0.096	0.060	0.004
			600	0.485	0.132	0.040	0.004
	0.0100	400	400	0.504	0.108	0.048	0.004
			500	0.519	0.104	0.052	0.004
			600	0.487	0.112	0.052	0.004
	0.0150	400	400	0.520	0.072	0.020	0.000
			500	0.532	0.076	0.032	0.000
			600	0.505	0.084	0.040	0.000
0.01	0.009	400	400	0.4742	0.112	0.072	0.004
			500	0.4794	0.152	0.072	0.004
			600	0.4824	0.160	0.072	0.024
	0.010	400	400	0.4638	0.112	0.020	0.004
			500	0.5013	0.108	0.052	0.008
			600	0.4905	0.116	0.076	0.008
	0.015	400	400	0.5175	0.060	0.036	0.008
			500	0.4951	0.112	0.060	0.012
			600	0.4802	0.084	0.040	0.012
0.05	0.009	400	400	0.2323	0.484	0.336	0.132
			500	0.2130	0.468	0.312	0.136
			600	0.1869	0.504	0.392	0.164
	0.010	400	400	0.2592	0.468	0.364	0.140
			500	0.2095	0.464	0.344	0.128
			600	0.2375	0.484	0.344	0.148
	0.015	400	400	0.2005	0.540	0.368	0.124
			500	0.2048	0.496	0.360	0.160
			600	0.2347	0.480	0.324	0.088
0.09	0.009	400	400	0.0887	0.800	0.704	0.476
			500	0.0642	0.836	0.748	0.500
			600	0.0448	0.900	0.828	0.552
	0.010	400	400	0.0658	0.828	0.768	0.540
			500	0.0624	0.864	0.780	0.512
			600	0.0493	0.888	0.828	0.556
	0.015	400	400	0.0704	0.832	0.748	0.576
			500	0.0622	0.860	0.812	0.536
			600	0.0506	0.876	0.808	0.592

**Tabla 5.4:** Proporción de rechazos del modelo (5.10) para el test basado en el estadístico  $T_{KS}^{(2)}$  (Tercer Enfoque)

$\rho_d$	$h$	$n_1$	$n_2$	$p$ -valor	$\hat{\alpha}_{KS}$		
					10 %	5 %	1 %
0	0.0090	400	400	0.411	0.100	0.044	0.004
			500	0.411	0.108	0.056	0.008
			600	0.427	0.092	0.040	0.008
	0.0100	400	400	0.418	0.120	0.044	0.004
			500	0.422	0.096	0.048	0.008
			600	0.435	0.108	0.020	0.000
	0.0150	400	400	0.449	0.104	0.044	0.004
			500	0.452	0.096	0.048	0.008
			600	0.444	0.112	0.044	0.008
0.01	0.009	400	400	0.3960	0.128	0.032	0.000
			500	0.4024	0.108	0.032	0.008
			600	0.4147	0.100	0.016	0.000
	0.010	400	400	0.3825	0.128	0.076	0.008
			500	0.4091	0.144	0.052	0.000
			600	0.4253	0.064	0.020	0.008
	0.015	400	400	0.4712	0.084	0.028	0.004
			500	0.4822	0.056	0.028	0.004
			600	0.4742	0.060	0.032	0.008
0.05	0.009	400	400	0.2919	0.228	0.140	0.044
			500	0.3302	0.256	0.140	0.032
			600	0.3200	0.280	0.160	0.032
	0.010	400	400	0.3300	0.256	0.152	0.056
			500	0.3321	0.232	0.104	0.036
			600	0.3591	0.196	0.084	0.028
	0.015	400	400	0.3435	0.196	0.116	0.044
			500	0.3892	0.152	0.104	0.028
			600	0.3518	0.192	0.092	0.040
0.09	0.009	400	400	0.2968	0.296	0.172	0.064
			500	0.2904	0.372	0.252	0.096
			600	0.2854	0.364	0.236	0.100
	0.010	400	400	0.2853	0.312	0.184	0.056
			500	0.3008	0.320	0.232	0.084
			600	0.3025	0.368	0.220	0.080
	0.015	400	400	0.3680	0.176	0.120	0.028
			500	0.3240	0.236	0.140	0.044
			600	0.3381	0.224	0.088	0.028

donde  $J_t = 1$  (con probabilidad  $\lambda\Delta$ ) indica que se produce un salto,  $\varepsilon_{t_i}$  son variables aleatorias con distribución normal de media cero y varianza  $\Delta$  y  $Z_{t_i}$  son variables con distribución normal de media cero y varianza  $\sigma_z^2$ . En este modelo, la media condicional es  $\mu(r_{t_i}, \theta)\Delta$ , y a varianza condicional es  $\sigma^2(r_{t_i}, \theta) + \lambda\Delta\sigma_z^2$ . Es claro que, basado solo en los primeros dos momentos, el modelo de difusión con saltos (5.11) es equivalente a un modelo de difusión sin saltos con función drift  $\mu(r_{t_i}, \theta)\Delta$  y función de difusión igual a  $\sigma^2(r_{t_i}, \theta) + \lambda\Delta\sigma_z^2$ . De esta forma, el procedimiento para obtener un test estadístico para este modelo es similar al caso de un proceso de difusión sin saltos, por supuesto, este debe incluir un procedimiento para la estimación consistente de los parámetros asociados a los términos de saltos. El procedimiento para la generación de la muestra artificial bootstrap es como sigue:

1. Para cada  $i = 1, 2, \dots, n$ , se generan

$$Y_{t_i}^* = \mu(r_{t_i}, \hat{\theta})\Delta + (\sigma(r_{t_i}, \hat{\theta})\Delta^{1/2} + \hat{\sigma}_z \tilde{J}_{t_i})\epsilon_{t_i}^*. \quad (5.12)$$

Y la muestra bootstrap  $\{(r_{t_i}^*, Y_{t_i}^*)\}_{i=1}^n$ , donde  $\hat{\theta}$  y  $\hat{\sigma}_z$  son estimadores apropiados del proceso con saltos. La variable  $r_{t_i}^* = r_{t_i}$  permanece inalterada (diseño fijo), y  $\{\epsilon_{t_i}^*\}$  son variables aleatorias independientes con distribución normal estándar  $N(0, 1)$ . Las variables  $\{\epsilon_{t_i}^*\}$  son independientes de  $\{r_{t_i}\}$ .  $\tilde{J}_{t_i} = 1$  con probabilidad  $\hat{\lambda}\Delta$ , donde  $\hat{\lambda}$  es el estimador del parámetro de intensidad del proceso original.

2.  $\hat{\theta}^*$  se estima usando un estimador apropiado de la remuestra bootstrap  $\{(r_{t_i}, Y_{t_i}^*)\}_{i=1}^n$  obtenida en el Paso 1.

3. Se calcula

$$R_n^{J*}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}^*}{\Delta} - \mu(r_{t_i}, \hat{\theta}^*) \right).$$

4. Se determina  $D_n^* = \Psi(R_n^{J*}(x))$ .

5. Se repiten los pasos anteriores  $B$  veces, obteniéndose  $B$  bootstrap replicas de  $D_n^{*j}$ .

6. Finalmente, se calcula  $\hat{c}_{1-\alpha}^* = D_n^{*[B(1-\alpha)]}$ .

El proceso  $R_n^{J*}(x)$  corresponde a la versión bootstrap del proceso  $R_n^J(x)$  dado por

$$R_n^J(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{1}_{\{r_{t_i} \leq x\}} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right).$$

obtenido bajo la hipótesis nula.

### 5.3 La Herramienta Gráfica: SiZer .

El SiZer es un método de visualización que combina ideas relativas a *espacio-escala* al considerar simultáneamente una familia de funciones de suavizado (por ejemplo, ajustes locales lineales) con la inferencia estadística empleada para el análisis exploratorio de datos en presencia de ruido (ver Chaudhuri y Marron (1999)). Dicha metodología aporta una visión inmediata de un tema central en el análisis exploratorio de datos: *¿cuales características observadas en un suavizado de los datos son "realmente" de los datos?*. El análisis SiZer tiene la cualidad de evitar los problemas clásicos de la selección del parámetro de suavizado puesto que esta metodología considera una amplia gama de estos. La idea esencial es que SiZer investiga la derivada de la función de suavizado y

a partir de esta puede ofrecer información relativa a un gran número de pruebas de hipótesis simultáneas.

La Herramienta gráfica SiZer introducida en [Chaudhuri y Marron \(1999\)](#), hace una evaluación visual de la importancia de características tales como “picos” y “valles” en una familia de funciones suaves  $\{\hat{f}_h(x) : h \in [h_{\min}, h_{\max}]\}$  basado en los límites de confianza de la derivada  $f'_h(x)$  en el espacio de escala. El comportamiento en  $x$  y la localización  $h$  se presenta a través del mapa de color SiZer, donde el azul (negro en las versiones donde sólo el blanco y negro están disponibles) indica los lugares donde  $\hat{f}'_h(x)$  es significativamente positiva, rojo (blanco en las versiones en blanco y negro) muestra donde  $\hat{f}'_h(x)$  es claramente negativo, y púrpura (color gris en las versiones en blanco y negro) indica donde  $\hat{f}'_h(x)$  no es significativamente diferente de 0. Los límites de confianza para  $\hat{f}'_h(x)$  son de la forma

$$\hat{f}'_h(x) \pm q \cdot \widehat{SD}(\hat{f}'_h(x)), \quad (5.13)$$

donde  $q$  es un cuantil apropiado, la función  $\hat{f}'_h(x)$  se estima por el método de suavizado lineal local

$$\hat{f}'_h(x) = \arg \min_b \sum_{i=1}^n n [Y_i - (a + b(X_i - x))]^2 K_h(x - X_i)$$

y la desviación estándar se estima como se explica en la Sección 3.1. de [Chaudhuri y Marron \(1999\)](#). Una localización  $(x, h)$  (en el espacio de escala) se llama significativamente creciente, decreciente o no significativo, cuando está por debajo de 0, por encima o dentro de estos límites de confianza. Los candidatos para el cálculo de los cuantiles  $q$  son:

- Cuantiles Gaussianos puntuales:

$$q_1(h) = q_1 = \Phi^{-1}[1 - (\alpha/2)]$$

- Aproximaciones simultáneas sobre cuantiles Gaussianos  $x$ : basados en el “número de bloques independientes”, definido como  $q_2$ .
- Bootstrap simultáneos sobre  $x$ , definido como  $q_3$ .
- Bootstrap simultáneos sobre  $x$  y  $h$ , definidos como  $q_4$ .

[Chaudhuri y Marron \(1999\)](#) sugieren la selección de  $q_2$ ,

$$q_2 = q_2(h) = \Phi^{-1} \left( \frac{1 + (1 - \alpha)^{1/m}}{2} \right) \quad (5.14)$$

donde

$$m(h) = \frac{n}{\text{avg}_x \text{ESS}(x, h)}. \quad (5.15)$$

y

$$\text{ESS}(x, h) = \frac{\sum_{i=1}^n K_h(x - X_i)}{K_h(0)}. \quad (5.16)$$

El rango del ancho de banda  $[h_{\min}, h_{\max}]$  se puede elegir de diversas maneras. [Chaudhuri y Marron \(1999\)](#) consideran usar un rango “suficientemente amplio”, el cual se determina es “una gama muy amplia de alisa”, el cual es determinado principalmente por la estimación de la curva que ajusta los datos. En este caso, se toma  $h_{\min}$  como el menor ancho de banda que produce distorsiones importantes en la implementación de la herramienta computacional para el suavizado,  $h_{\min} = 2 * (\text{bandwidth})$ , y  $h_{\max}$  el rango de los datos. Para más detalles acerca de la metodología SiZer, ver [Fan y Marron \(1994\)](#) y [Chaudhuri y Marron \(1999\)](#).

El objetivo en esta sección es usar SiZer como un test gráfico y constatar si el análisis SiZer se corresponde con la aceptación o rechazo, en contraste con el  $p$ -valor obtenido al implementar los tests discutidos en el capítulo 3 a las series de tipos de interés del EURIBOR que se muestran en la tabla (3.12).



### 5.3.1 Sizer y el Test de bondad de ajuste para la función drift.

En esta sección se propone aplicar el análisis SiZER, a cada serie de datos del tipo de interés a distintos plazos del *EURIBOR* de acuerdo a un conjunto de discretizaciones  $\Delta_1, \Delta_2, \dots, \Delta_d$ . De esta manera se pretende generar un panel de gráficos SiZers en el que se visualiza el impacto del parámetro  $\Delta$ . Para este caso concreto, la metodología SiZER se aplica de forma directa y consistirá entonces en explorar las características subyacentes en el conjunto de residuos del modelo de difusión bajo la hipótesis nula en su versión discreta. En otras palabras, para un modelo de difusión

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t,$$

su versión discretizada, como en el capítulo 3, a partir del método de Euler es dada por

$$Y_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)\sqrt{\Delta}\varepsilon_{t_i},$$

donde  $t_i = i\Delta, i=1, \dots, n$ , en el intervalo  $[0, t] = [0, n\Delta]$  con  $\Delta$  fijo;  $Y_{t_i} = r_{t_{i+1}} - r_{t_i}$ , y  $\varepsilon_{t_i}$  variables aleatorias normales estándar independientes. Entonces los residuos se definen por

$$\hat{e}_{t_i} = \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}),$$

donde  $\hat{\theta}$  es un estimador apropiado del modelo.

Para nuestro caso, la motivación es la misma que se desarrolla e ilustra en distintos problemas tratados en [González-Manteiga et al. \(2008\)](#) y parte del sencillo test para el contraste de modelos de regresión paramétricos propuesto en [Härdle y Mammen \(1993\)](#), en los que se contrasta la hipótesis nula

$$H_0 : m \in \{m_\theta : \theta \in \Theta\}$$

contra la hipótesis alternativa

$$H_1 : m \text{ es suave.}$$

El estadístico considerado es dado por

$$T_n = nh^{D/2} \int \{\hat{m}_h(r) - \mathcal{H}_{h,n}(m_{\hat{\theta}}(r))\}^2 \pi(x) dx$$

donde  $\mathcal{H}_{h,n}$  es el operador de suavizado. El test propuesto compara las dos familias de funciones de suavizado:

$$\mathcal{M}_1 = \{\hat{m}_h | h \in H, x \in I\}, \quad \mathcal{M}_0 = \{\mathcal{H}_{h,n}(\hat{m}_{\hat{\theta}}) | h \in H, x \in I\}$$

En lo que respecta al proceso de difusión discretizado (3.6), vamos a explorar la familia de funciones de suavizado de los residuos indexados por el ancho de banda  $h$ , es decir,

$$\begin{aligned} \mathcal{M}_{\hat{e}_{t_i}} &= \{\mathcal{H}_{h,n}(\hat{e}_{t_i}) : h \in H, i \in \{0, 1, \dots, n\}\} \\ &= \left\{ \mathcal{H}_{h,n} \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right) : h \in H, i \in \{0, 1, \dots, n\} \right\} \end{aligned}$$

De esta forma, tomando como base el conjunto de los residuos, nuestra estrategia es buscar diferencias significativas entre las dos familias usando SiZER. Si la hipótesis nula es cierta, entonces no debe aparecer ninguna característica importante en el gráfico SiZER. La presencia de rasgos significativos en un cierto rango de anchos de banda conllevaría a rechazar la hipótesis nula. Llevaremos a cabo el SiZER para,  $\Delta \in \{1/252, 1/52, 1/12, 1/2, 1\}$ , con el fin de visualizar el impacto de dicho parámetro y compararlo con las conclusiones derivadas de los  $p$ -valores obtenidos al aplicar el test de bondad de ajuste. Los gráficos que se mostraran se han generado usando el software implementado por en línea disponibles Marron en Matlab

([http://www.stat.unc.edu/profesores/marron/marron\\_software.html](http://www.stat.unc.edu/profesores/marron/marron_software.html)).

Los gráficos consisten en tres subgráficos: El primero, muestra la familia de ajustes por suavizado a distintos niveles de la serie de datos, el subgráfico (*Slope SiZer Map*) central muestra el gráfico SiZer de las características de la derivada (es decir, cambios significativos en la señal de la primera derivada), y el subgráfico inferior corresponde a las características de la segunda derivada *Curvature SiZer Map*. Los distintos colores se pueden interpretar como: azul (negro para versiones en blanco y negro) si la derivada es no significativamente positiva; rojo (gris oscuro) si la derivada es significativamente negativa, y purpura (gris claro) si la derivada no es significativamente distinta de cero. Las zonas en gris indican que los valores del parámetro de suavizado demasiado pequeños y que invalidan el cálculo de la curva estimada, es decir, el número de observaciones que caen dentro de cada ventana local es insuficiente para la estimación. Respecto del gráfico *Curvature SiZer Map*, los colores se pueden interpretar como: cian (negro) si la segunda derivada es positiva, naranja (gris oscuro) si la segunda derivada es negativa y verde (gris claro) para derivadas no significativamente distinta de cero.

Las tablas (5.5) y (5.6) muestran los  $p$ -valores estimados obtenidos al aplicar los tests de bondad de ajuste planteados en el capítulo 3 a las series de tipos de interés del *EURIBOR* para diversos valores de  $\Delta$ , para probar la bondad del ajuste de los modelos de Vasicek (1977) (VAS) y Chan et al. (1992) *CKLS*, en cuanto a la forma paramétrica de la función drift y de la función de volatilidad. El período temporal abarca desde el 15 de octubre del 2001 hasta el 30 de marzo de 2006, con frecuencia diaria. Los plazos considerados son de 1 semana y 1 y 12 meses. Es claro que, el valor de  $\Delta$  no influye en el  $p$ -valor de forma significativa.

**Tabla 5.5:**  $p$ -valores para el Test de la forma paramétrica del drift para el Modelo de Vasicek(VAS)

Plazo	Drift					Volatilidad				
	$\Delta$					$\Delta$				
	$\frac{1}{252}$	$\frac{1}{52}$	$\frac{1}{12}$	$\frac{1}{2}$	1	$\frac{1}{252}$	$\frac{1}{52}$	$\frac{1}{12}$	$\frac{1}{2}$	1
1 week	0.026	0.026	0.027	0.028	0.040	0	0	0	0	0
12 months	0.798	0.798	0.790	0.791	0.801	0	0	0	0	0

**Tabla 5.6:**  $p$ -valores para el Test de la forma paramétrica del drift para el Modelo de *CKLS*

Plazo	Drift					Volatilidad				
	$\Delta$					$\Delta$				
	$\frac{1}{252}$	$\frac{1}{52}$	$\frac{1}{12}$	$\frac{1}{2}$	1	$\frac{1}{252}$	$\frac{1}{52}$	$\frac{1}{12}$	$\frac{1}{2}$	1
1 week	0.059	0.061	0.061	0.053	0.060	0	0	0	0	0
12 months	0.909	0.925	0.916	0.906	0.919	0.002	0	0	0	0.002

A la vista de los resultados mostrados en los gráficos al aplicar la metodología SiZer, estos se corresponden con los obtenidos por el test de bondad de ajuste para el drift en las tablas (5.5) y (5.6). Se observa además que la influencia del parámetro  $\Delta$  no es del todo determinante, al menos para el caso de la serie *EURIBOR* con plazo a 12 meses, aún en el caso de 1 semana. Los  $p$ -valores,  $p = 0.026, 0.026, 0.027, 0.028, 0.040$  para  $\Delta = 1/252, 1/52, 1/12, 1/2, 1$  en el caso de las series con plazo a una semana y bajo la hipótesis nula del Modelo de Vasicek, indican que dicha hipótesis solo es rechazada para un nivel inferior al %5, por el contrario, las series con plazo

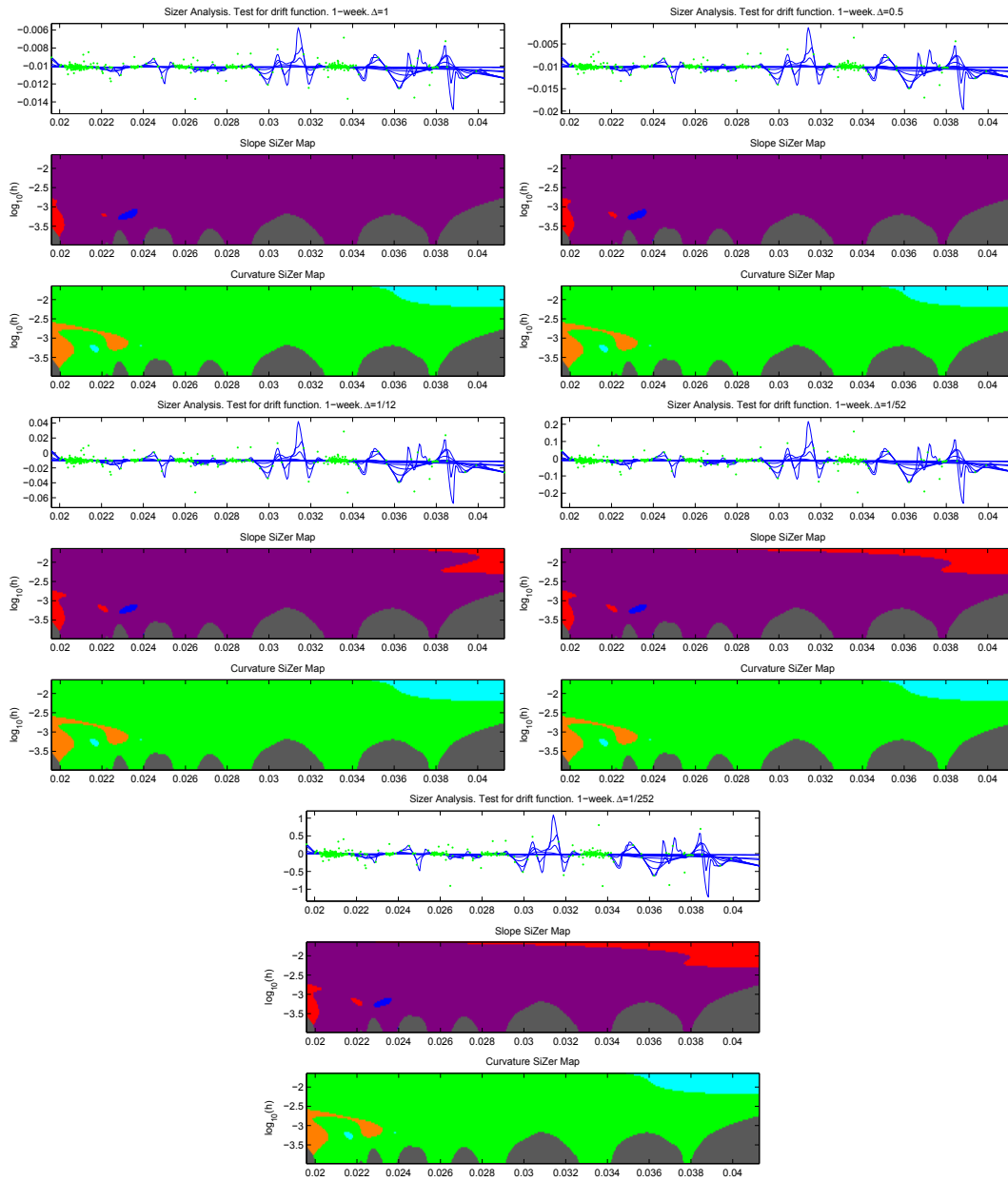


Figura 5.1: Gráficos Sizer para los residuos bajo la hipótesis nula (Modelo de Vasicek) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo una semana

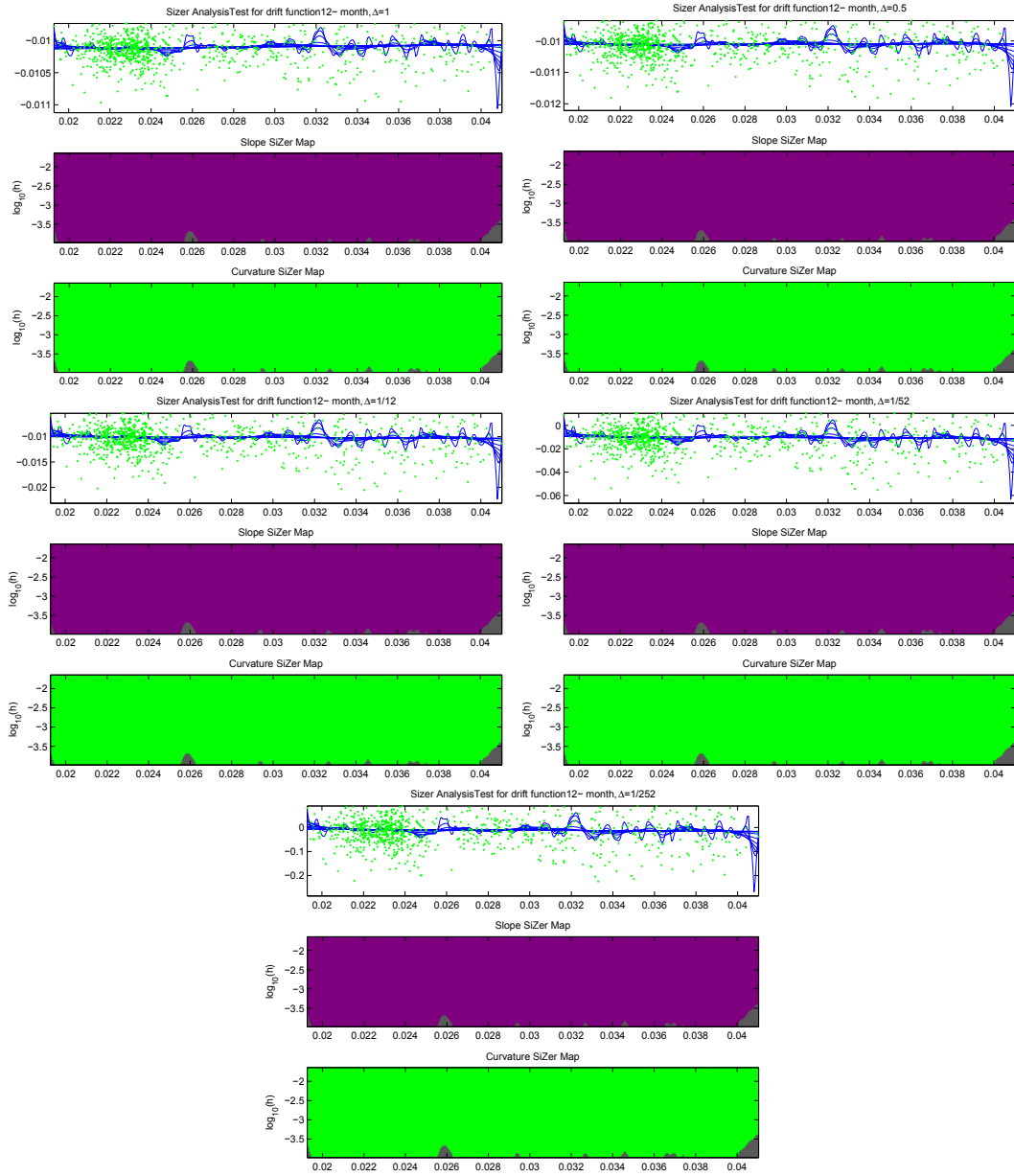


Figura 5.2: Gráficos SiZer para los residuos bajo la hipótesis nula (Modelo de Vasicek) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo 12 meses

superior a un mes, presentan  $p$ -valores que sugieren que dicha hipótesis no debe ser rechazada. Los gráficos , *Slope Sizer Map*, de la figura (5.1) obtenidos por SiZer en el caso de la serie con plazo a una semana y para los distintos  $\Delta$  correspondientes a evidencia características significativas. En este caso dicha significación se presenta como un decrecimiento para valores pequeños del tipo de interés  $r_{t_i}$  y para un rango de ancho de banda también pequeño, y por un decrecimiento y crecimiento para valores intermedios del tipo de interés  $r_{t_i}$  para un rango reducido de ancho de banda, reflejando así la presencia de un posible pico. A medida que disminuimos el parámetro  $\Delta$  se presenta una región de decrecimiento para valores superiores de  $r_{t_i}$  con anchos de banda mayores. Similar comportamiento los gráficos *Curvature SiZer Map*. Lo anterior nos lleva a la conclusión de rechazar la hipótesis de que la función drift se pueda modelizar mediante la función drift de un modelo de Vasicek.

Respecto a la serie de tipo de interés con plazo a 12 meses, en los gráficos de la figura (5.2) no se evidencia significación alguna. Así pues se concluye que no hay evidencia para rechazar la hipótesis nula, coincidiendo así con los resultados obtenidos a partir de los  $p$ -valores calculados aplicando el test. Los gráficos de las figuras (5.3) y (5.4) consideran el modelo CKLS como hipótesis nula. Las conclusiones que se deducen de este son similares a las del modelo de Vasicek.

Similar comportamiento para las series con plazos de vencimiento distintos a los de 1 semana y 12 meses, cuyos gráficos no se muestran por no presentar características relevantes y sus gráficos son bastante similares a los presentados.

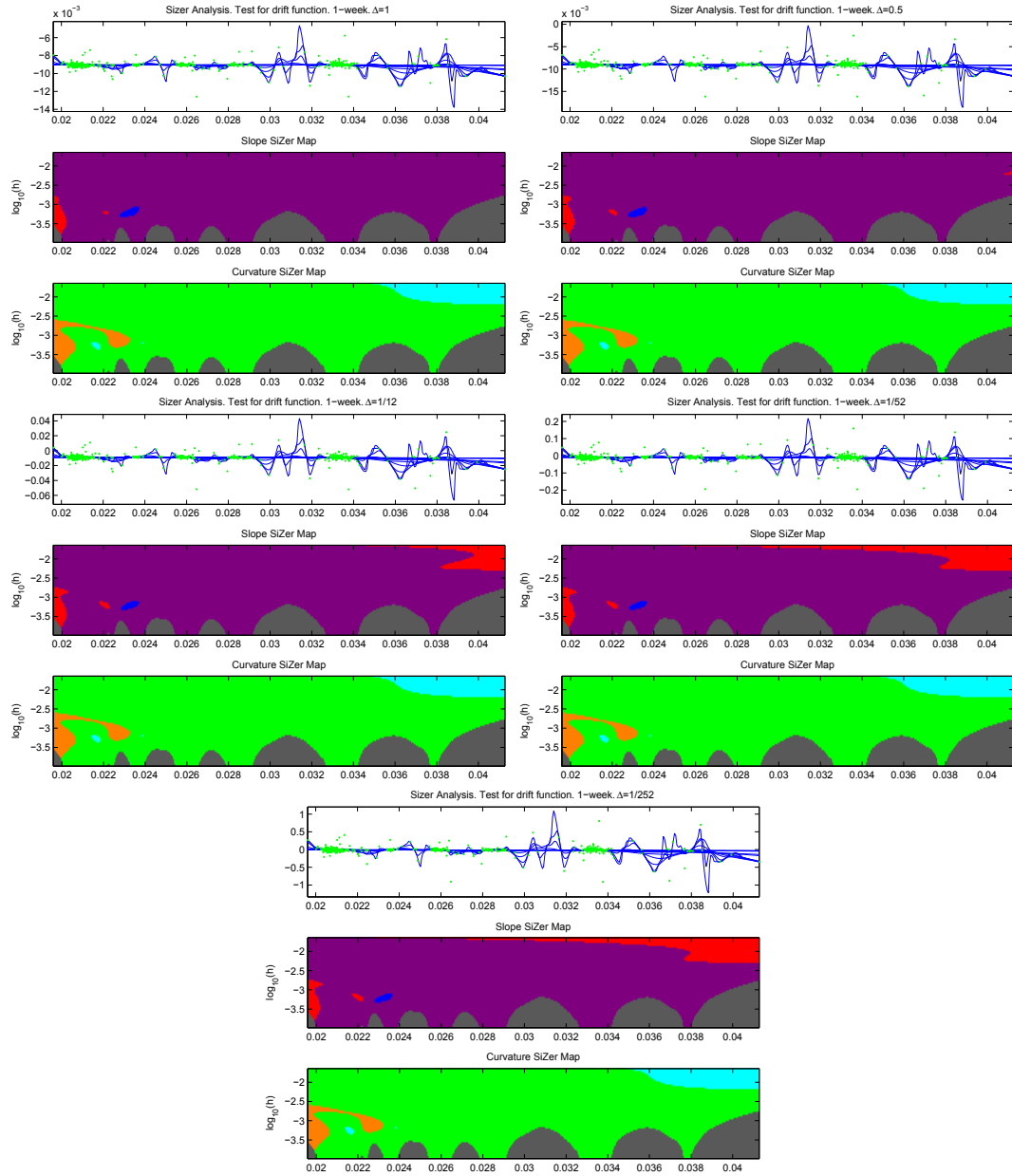


Figura 5.3: Gráficos SiZer para los residuos bajo la hipótesis nula (Modelo de CKLS) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo una semana

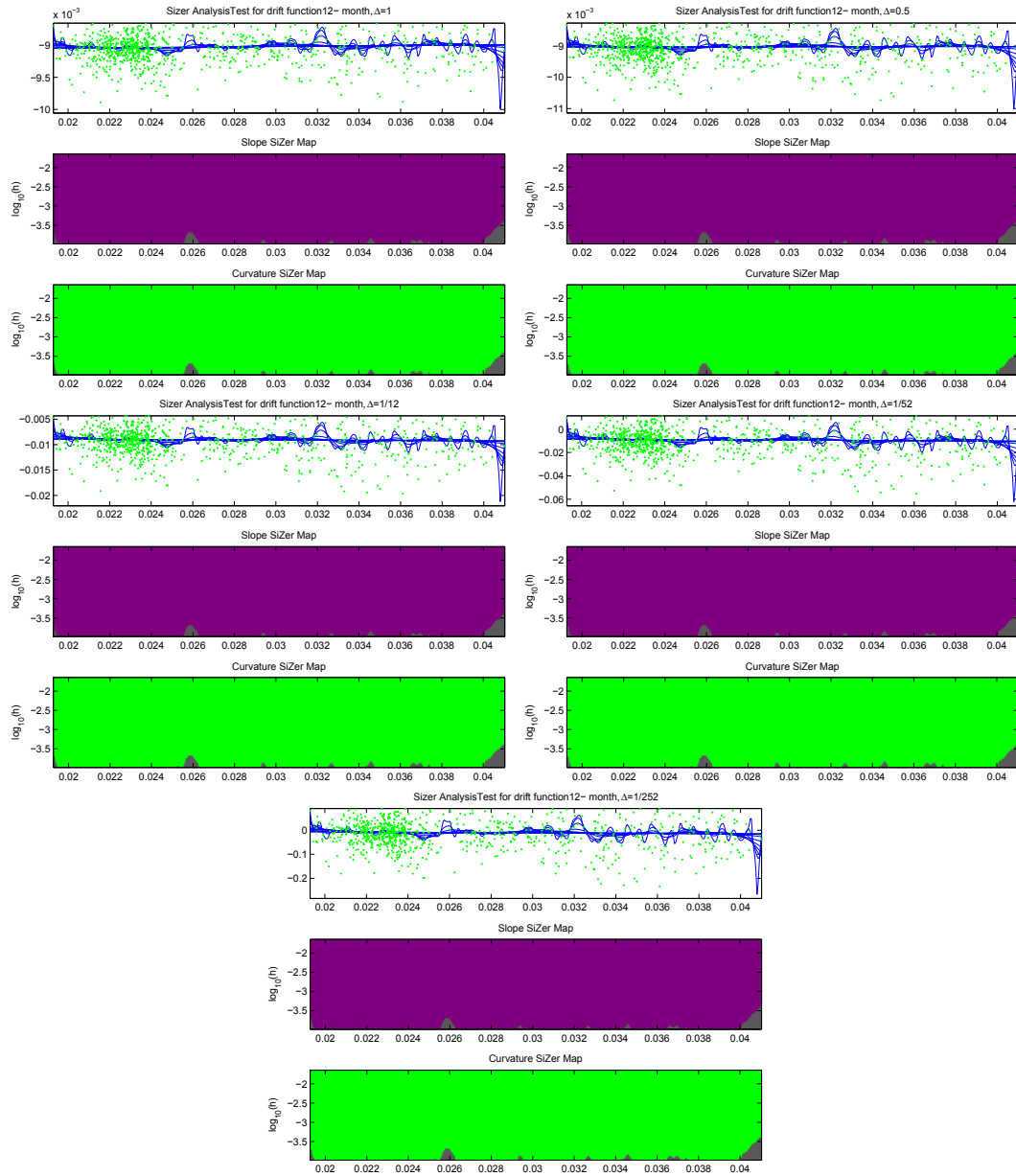


Figura 5.4: Gráficos SiZER para los residuos bajo la hipótesis nula (Modelo de CKLS) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo 12 meses

### 5.3.2 Sizer y el Test de bondad de ajuste para la función volatilidad

En lo que respecta al proceso (3.6), vamos a explorar la familia de suavizado de la diferencia al cuadrado entre los residuos y la volatilidad, indexados por el ancho de banda  $h$ , es decir,

$$\mathcal{M}_{\widehat{ev}_{t_i}} = \{\mathcal{H}_{h,n}(\widehat{ev}_{t_i}) | h \in H\} = \left\{ \mathcal{H}_{h,n} \left[ \left( \frac{Y_{t_i}}{\Delta} - \mu(r_{t_i}, \hat{\theta}) \right)^2 - \frac{\sigma^2(r_{t_i}, \hat{\theta})}{\Delta} \right]^2 | h \in H \right\}$$

Es importante destacar, que este planteamiento es solo exploratorio. De esta forma, nuestra estrategia es buscar diferencias significativas entre las dos familias usando SiZer. Si la hipótesis nula es cierta, entonces no debe aparecer ninguna característica importante en el gráfico SiZer, es decir, la presencia de rasgos significativos en un cierto rango de anchos de banda conlleva a rechazar la hipótesis nula. Llevaremos a cabo el SiZer para,  $\Delta \in \{1/252, 1/52, 1/12, 1/2, 1\}$ , con el fin de visualizar el impacto de dicho parámetro y compararlo con las conclusiones derivadas de los  $p$ -valores obtenidos al aplicar el test para la forma paramétrica de la función de volatilidad que se muestran en



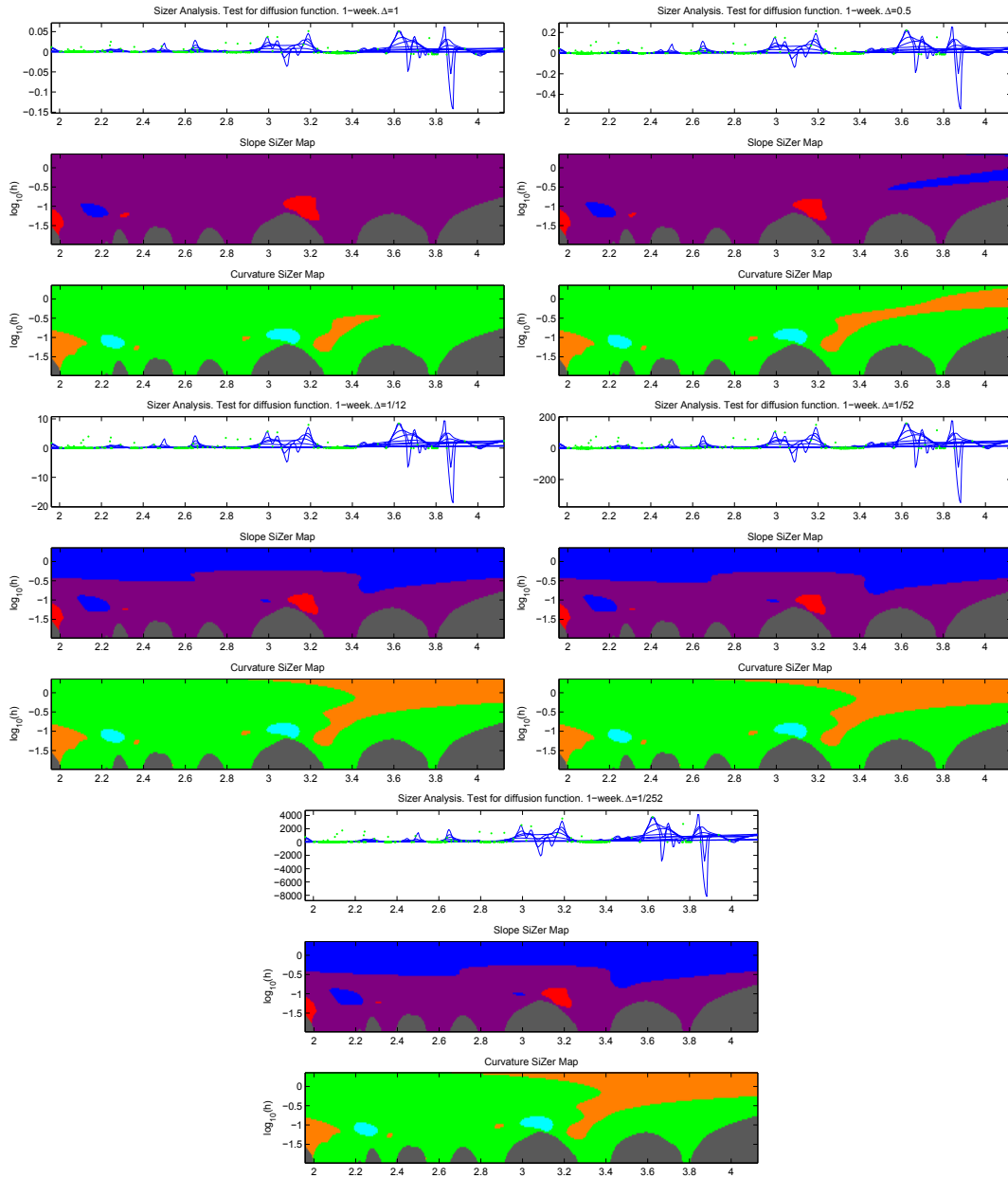


Figura 5.5: Gráficos SiZER para los residuos bajo la hipótesis nula (Modelo de CKLS) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo una semana

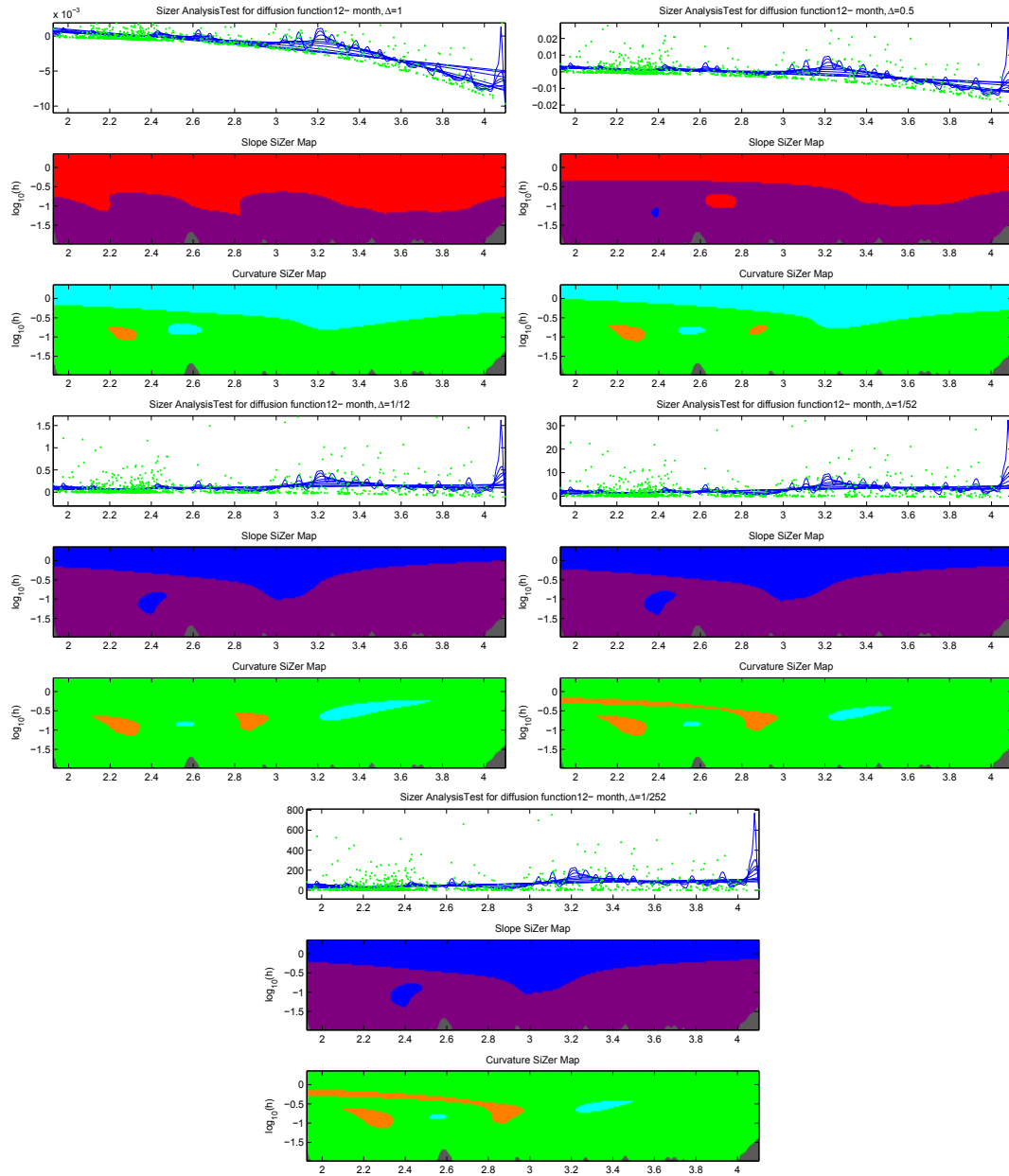


Figura 5.6: Gráficos SiZer para los residuos bajo la hipótesis nula (Modelo de CKLS) versus la covariable  $r_{t_i}$ , el tipo de interés del Euribor con plazo 12 meses

## 5.4 Líneas Futuras de Investigación

El trabajo desarrollado en esta memoria aborda el problema de la bondad de ajuste de los modelos de tipos de interés. El problema ha sido estudiado desde el punto de vista de los procesos empíricos. Aún cuando el desarrollo ha sido amplio, en el sentido de que se han cubierto distintos puntos de interés, tales como: tests para la forma paramétrica de las funciones drift y volatilidad, pruebas de bondad de ajuste para la comparación de curvas, una extensión para un caso particular de procesos de difusión y una aplicación gráfica basada en SiZer. este trabajo abre paso a un número importante de problemas que quedan abiertos: Entre los que se pueden destacar:

- En primer lugar, el problema relativos a considerar los estadísticos para distintos esquemas de discretización de los modelos de tipos de interés,

$$r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta(t) + \sigma(r_{t_i}, \theta)\sqrt{\Delta(t)}\varepsilon_{t_i}$$

en los que el parámetro  $\Delta(t)$  es una función de  $t$  con  $t \in [0, T]$  y su efecto en la potencia de los tests y las consideraciones teóricas relativas a éste.

- El problema de extender el test de bondad de ajuste basado en procesos empíricos que aborde los distintos procesos de difusión con saltos,

$$Y_{t_i} = r_{t_{i+1}} - r_{t_i} = \mu(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)\varepsilon_{r_{t_i}} + Z_{t_{i+1}}J_{t_{i+1}},$$

donde  $J_t = 1$  (con probabilidad  $\lambda\Delta$ ) indica que se produce un salto,  $\varepsilon_{t_i}$  son variables aleatorias con distribución normal de media cero y varianza  $\Delta$  y  $Z_{t_i}$  son variables con distribución  $F$ , donde  $F$  no es la distribución normal.

- Un test de bondad de ajuste basado en proceso empíricos para modelos de volatilidad estocástica, por ejemplo

$$dr_t = \mu r_t dt + \sigma_t r_t dW_{1t}$$

y

$$d\sigma_t^2 = \beta\sigma_t^2 dt + v\sigma_t^2 dW_{2t}$$

lo cual resultaría de gran beneficio en lo que respecta al problema de modelización de la volatilidad de los procesos, ya que, como se ha mostrado en las aplicaciones, los resultados obtenidos al aplicar el test a las series del *EURIBOR*, los modelos son ampliamente rechazados en cuanto a la modelización de la volatilidad.

- La formulación de tests en la línea de las ideas presentadas en esta memoria para modelos de difusión en los que se consideren como variable de estado, además del tipo de interés, otras variables exógenas que permitan modelizar el tipo de interés de forma eficiente.
- Aplicación de la Metodología SiZer como herramienta gráfica en la comparación de curvas de tipos de interés siguiendo las ideas planteadas en [Park y Kang \(2008\)](#). Y una formalización ara el caso del test de bondad de ajuste para la función de volatilidad.

# IV

## 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  $\Omega$  es el *espacio muestral* de todos los posibles resultados de un experimento aleatorio;  $\mathcal{F}$  es una  $\sigma$ -*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  $\sigma$ -álgebra son denominados *eventos*, por ello,  $\mathcal{F}$  suele ser llamada también la  $\sigma$ -*álgebra de eventos*. En este orden de ideas se define  $\omega$  como los elementos en  $\Omega$ , así pues un evento  $A \subset \Omega$ , es una colección de puntos muestrales.  $\mathbb{P}$  es una *medida de probabilidad* en  $(\Omega, \mathcal{F})$ , es decir,  $\mathbb{P}$  es una función definida en  $\Omega$  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.** Dado un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ , se define como *variable aleatoria* a una *función medible*  $X$  tal que

$$\begin{aligned} X &: \Omega \mapsto \mathbb{R} \\ \omega &\mapsto X(\omega) = x \end{aligned}$$

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  $\sigma$ -álgebra de conjuntos de Borel (la  $\sigma$ -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  $\sigma$ -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  $\Omega$ ,

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(\{\omega \in \Omega : \omega \in X^{-1}(A)\}) = P(B),$$

para  $A \in \mathcal{B}(\mathbb{R})$  y  $B \in \mathcal{F}$ .

**Definición A.2.** La función

$$F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X(\omega) \in (-\infty, x]) = P(X \leq x), \quad x \in \mathbb{R}$$

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.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.4.** 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  $\Omega$  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.5.** 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.6.** 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

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(u_1, \dots, u_n) du_1 \dots du_n, \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

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.7.** Se define la función de distribución marginal de la variable aleatoria  $X_i$  como

$$F_{\mathbf{X}_i}(x_i) = F_{\mathbf{X}_i}(\infty, \dots, x_i, \dots, \infty) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(u_1, \dots, x_i, \dots, u_n) du_1 \dots du_n,$$

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.8.** 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.9.** 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.1.** Una variable aleatoria  $n$ -dimensional Normal o Gaussiana tiene una distribución Gaussiana con función de densidad

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{n/2}} \exp \left\{ -\frac{1}{2} (x - \mu) \Sigma (x - \mu)^T \right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$

con parámetros  $\mu \in \mathbb{R}^n$ , y  $\Sigma$ . La cantidad  $\Sigma$  es una matriz  $(n \times n)$  definida positiva simétrica,  $\Sigma^{-1}$  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.10.** 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.11.** 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 (incorreladas), 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.12.** 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

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

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.13.** 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  $\sigma$ -álgebra de eventos (por ejemplo, la  $\sigma$ -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  $\sigma$ -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

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(X\mathbb{1}_A) \quad \text{para cada } A \in \mathcal{F}. \quad (\text{A.4})$$

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  $\sigma$ -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

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (\text{A.5})$$

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

$$X_n \xrightarrow{d} X. \quad (\text{A.6})$$

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  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \quad (\text{A.7})$$

y se denota por

$$X_n \xrightarrow{p} X \quad (\text{A.8})$$

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

$$X_n \xrightarrow{\text{c.s.}} X. \quad (\text{A.11})$$

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.** Sea  $(\Omega, \mathcal{F}, \mathbb{P})$  un espacio de probabilidades. Un *proceso estocástico* con valores reales es una familia de variables aleatorias  $\{X_\gamma, \gamma \in \Gamma\}$  ( $\Gamma$  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  $\omega$ , 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.** Considérese un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . Una *filtración*  $\{\mathcal{F}_t, t \geq 0\}$  es una familia creciente de sub- $\sigma$ -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  $\sigma$ -álgebra denotada por  $\mathcal{F}_t = \sigma(X_s; 0 \leq s \leq t)$ , y que además es la  $\sigma$ -álgebra generada por el proceso  $X$  hasta el instante  $t$ ; es decir, la  $\sigma$ -álgebra más pequeña (minimal) de  $\mathcal{F}$  que hace a  $X(s, \omega)$  medible para cada  $0 \leq s \leq t$ . Dicha  $\sigma$ -álgebra es el conjunto de subconjuntos más pequeño de  $\Omega$  que hace posible asignar probabilidades a todos los eventos relacionados con el proceso  $X$  hasta el instante  $t$ .

**Definición B.3.** 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  $\Pi_n$ .

**Definición B.4.** La *variación* del proceso  $X$  se define como

$$V_t(X) = p - \lim_{\|\Pi_n\| \rightarrow 0} \sum_{k=0}^{n-1} |X_{t_{k+1}} - X_{t_k}|. \quad (\text{B.1})$$

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.5.** 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

$$\langle X, X \rangle_t = p - \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} (X_{t \wedge k/2^n} - X_{t \wedge (k-1)/2^n})^2,$$

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.6.** 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.7.** 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.8.** Un proceso  $X_t, 0 \leq t \leq T$  es llamado *uniformemente integrable* si

$$\mathbb{E}(|X_t| \mathbf{1}_{\{|X_t| > n\}})$$

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  $\sigma$ -álgebra generada por el proceso hasta el tiempo  $t$ .

**Definición B.9.**  $X$  es un *proceso de Markov* si para cualquier  $t$  y  $s > 0$ ,

$$\mathbb{P}(X_{t+s} \leq y | \mathcal{F}_t) = \mathbb{P}(X_{t+s} \leq y | X_t) \quad c.s.$$

**Definición B.10.** 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  $\sigma$ -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- $\sigma$ -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  $\Omega$  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  $\sigma$ -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  $\mu$  y  $\sigma$  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  $\mu$  es denominada el término de *drift* del proceso, mientras que  $\sigma$  es denominada el término de *difusión*.

En la modelización del término de perturbación Gaussiano se suele usar el proceso de Wiener.

**Definición C.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  $\Delta W_t$  se define por

$$\Delta W_t = W_{t+\Delta t} - W_t.$$

Ahora bien, dividiendo por  $\Delta t$  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  $\Delta t$ . 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

$$X_t = x_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \quad (\text{C.6})$$

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

$$\text{Var} \left( \int_0^T X_s dW_s \right) = \int_0^T \mathbb{E}(X_s)^2 dW_s = 0, \quad (\text{Isometría de Itô})$$

- Si  $X$  e  $Y$  son dos procesos integrables según Itô y  $\alpha$  y  $\beta$  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 Itô (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 fórmula, se puede demostrar que la variación cuadrática de un proceso de Wiener (movimiento Browniano) converge en probabilidad a  $T$ . Por lo tanto, la integral  $\int_0^T X_t^n dW_t$  converge en probabilidad a al límite  $J$

$$\int_0^T W_t dW_t = J = \lim_{n \rightarrow \infty} \int_0^T X_t^n dW_t = \frac{1}{2} W_T^2 - \frac{1}{2} T.$$

**Definición C.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 fórmula 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

$$W_t^2 = 0^2 + \int_0^t 2W_s dW_s + \frac{1}{2} \int_0^t 2ds$$

por lo tanto,

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

*Observación C.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.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),

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad (\text{C.9})$$

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  $\sigma$ -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.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 estás 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.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.** 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 y 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, los proceso de difusión puede ser o no ergódicos. 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  $\xi$  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

$$s(x) = \exp \left\{ -2 \int_{x_0}^x \frac{b(y)}{\sigma^2(y)} dy \right\} \quad (\text{C.18})$$

y

$$m(x) = \frac{1}{\sigma^2(x)s(x)}. \quad (\text{C.19})$$

respectivamente. En particular, la densidad de la distribución invariante  $\pi(\cdot)$  es proporcional, a la medida de rapidez, es decir,

$$\pi(x) = \frac{m(x)}{M}, \quad (\text{C.20})$$

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

$$P(t, y, |s, B) = P\{X_t \in B | X_s = x\}, \quad (\text{C.21})$$

donde  $s < t$ . Para  $s, x$  fijos y  $t$ ,  $P(t, x | s, \cdot)$  es una medida de probabilidad bajo la  $\sigma$ -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 denota 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,

$$p(t, y|s, x) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2(t-s)}\right) \quad (\text{C.22})$$

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*

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}\{\mu(y)p\} - \frac{1}{2}\frac{\partial^2}{\partial y^2}\{\sigma^2(y)p\} = 0, \quad (s, x) \quad (\text{C.23})$$

y las *Ecuaciones Backward de Kolmogorov*

$$\frac{\partial p}{\partial s} + \frac{\partial}{\partial x}\{\mu(x)p\} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 p}{\partial x^2} = 0, \quad (\text{C.24})$$

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

- Ahn, D. H. y Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12(4):721–762.
- Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64:527–560.
- Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9(2):385–426.
- Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of 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.
- Aït-Sahalia, Y., Fan, J., y Jiang, J. (2010). Nonparametric test of the markov hypothesis in continuous-time models. *The Annals of Statistics*, 38(5):3129–3163.
- Aït-Sahalia, Y., Fan, J., y Peng, H. (2009). Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association*, 104(487):1102–1116.
- Aït-Sahalia, Y. y Jacod, J. (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37(1):184–222.
- Akaike, H. (1973). Information theory and the extension of the maximum likelihood principle. 267-281, In B.N Petrov and F. Csaki, Akademia Kiado, Budapes.
- Ang, A. y Bekaert, G. (2002). Regime switches in interest rates. *Journal of Business and Economic Statistics*, 20:163–182.
- Arapis, M. y Gao, J. (2006). Empirical comparisons in short-term interest rate models using non-parametric methods. *Journal of Financial Econometrics*, 4(2):310–345.
- Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. John Wiley and Sons, New York.
- Azzalini, A. y Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc., Serie B(55)*:549–557.
- Bachelier, L. (1900). Théorie de la speculation. *Reprinted in Cootner*, pages 17–78.
- Bakshi, G., Cao, C., y Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52:2003–2049.
- Ball, C. A. y Torous, W. N. (1998). Unpublished manuscript. university of california los angeles. *Regimen switches in short-term riskless interest rates*, pages 163–182.

- Ball, C. A. y 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. y 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.
- Billingsley, P. (1961). *Convergence of Probability Measures*. Wiley, New York.
- Black, F., Derman, E., y Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, (46):33–39.
- Black, F. y Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Finan. Analysts' J.*, (47):52–59.
- Black, F. y Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54.
- Boente, G. y Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis*, 25:90–99.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Box, G. E. P. y Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*(26):211–243.
- Box, G. E. P. y Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, revised edition.
- Box, G. E. P., Jenkins, G. M., y Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ, 3rd edition edition.
- Box, G. E. P. y 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., y De Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348–54.
- Brennan, M. y Schwartz, E. (1979). A continuous time approach to the pricing of bonds. *Journal of Banking and Finance*, 3(2):133–155.
- Brenner, R., Harjes, R., y Kroner, K. (1996). Another look at alternative models of short-term interest rate. *Journal of Financial and Quantitative Analysis*, 31:85–107.
- Brockwell, P. J. y Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, second edition edition.
- Bryson, A. E. y Ho, Y. C. (1969). *Applied Optimal Control*. Blaisdell, New York.
- Cai, J. (1994). Journal of business and economic statistics. *A Markov model of switching-regime ARCH*, 12:309–316.
- Cai, Z., Fan, J., y 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., y MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Carlin, B. P., Polson, N. G., y 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. C., Karolyi, G. A., Longstaff, F. A., y Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3):1209–1227.
- 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. y Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:179–190.
- Chan, K. S. y Tsay, R. S. (1998). Limiting properties of the conditional least squares estimator of a continuous tar model. *Biometrika*, 85:413–426.
- Chan, N. H. y Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, 16:367–401.
- Chaudhuri, P. y Marron, J. (1999). Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823.
- Chen, R., Liu, J., y Tsay, R. S. (1995). Additivity tests for nonlinear autoregressive models. *Biometrika*, 82:369–383.
- Chen, R. y Tsay, R. S. (1991). On the ergodicity of tar(1) processes. *Annals of Applied Probability*, 1:613–634.
- Chen, R. y Tsay, R. S. (1993a). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88:298–308.
- Chen, R. y Tsay, R. S. (1993b). Nonlinear additive arx models. *Journal of the American Statistical Association*, 88:955–967.
- Chen, S. X., Gao, J., y Tang, C. Y. (2008). A test for model specification of diffusion processes. *Annals of Statistics*, 36(1):167–198.
- Chen, T. y 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. y 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., y Scheinkman, A. (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies*, 10:525–577.
- Corradi, V. y White, H. (1999). Specification tests for the variance of a diffusion. *Journal of Time Series Analysis*, 20(3):253–270.
- Cox, J., Ingersoll, J. E., y Ross, S. A. (1980). An analysis of variable rate loan contracts. *Journal of Finance*, 35(2):389–403.
- Cox, J., Ingersoll, J. E., y 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.
- Das, S. R. (2002). The surprise element: jumps in interest rates. *Journal of Econometrics*, 106(1):27–65.
- Delgado, M. A. (1993). Testing for the equality of nonparametric regression curves. *Statistics and Probability Letters*, (17):199–204.
- Dette, H. y Neumeier, N. (2001). Nonparametric analysis of covariance. *Annals of Statistics*, 29:1361–1400.
- Dette, H. y Podolskij, M. (2008). Testing the parametric form of the volatility in continuous time diffusion models—a stochastic approach. *Journal of Econometrics*, 143:56–73.
- Dette, H. y von Lieres und Wilkau, C. (2003). On a test for a parametric form of volatility in continuous time financial models. *Finance and Stochastics*, 7(3):363–384.
- Dette, H. y Weißbach, R. (2009). A bootstrap test for the comparison of nonlinear time series. *Computational Statistics and Data Analysis*, 53(4):1339–1349.
- Dickey, D. y Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.
- Dothan, L. U. (1978). On the term structure of interest rates. *Journal of Financial Economics*, (6):59–69.
- Duffie, D. (1995). *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, NJ, 2nd edition.
- Durbin, J. y 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.
- Eubank, R. L. y Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, 85:387–392.
- Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1):pp. 34–105.
- Fan, J. y Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hell.
- Fan, J., Jiang, J., Zhang, C., y Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, 13(4):965–992.
- Fan, J. y Marron, J. S. (1994). Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics*, 3:35–56.
- Fan, J. y Yao, Q. W. (2003). *Nonlinear Time Series*. Springer, New York.
- Fan, J. y Zhang, C. (2003). A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, 98(461):118–134.
- Fan, J., Zhang, C., y Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics*, 29:153–193.

- Fan, Y. y Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64:865–890.
- Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1):173–182.
- Ferreira, E. y Stute, W. (2004). Testing for difference between conditional means in a time series context. *Journal of the American Statistical Association*, 99(465):169–174.
- Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied of Probability*, 30:790–804.
- Franke, J., Kreiss, J. P., y Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*, (85):1–37.
- Gao, J. (2007). *Nonlinear time series: semiparametric and nonparametric methods*. Chapman & Hall/CRC.
- Gao, J. y Casas, I. (2008). Specification testing in discretized diffusion models: Theory and practice. *Journal of Econometrics*, 147(1):131–140.
- Gao, J. y King, M. (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory*, 20(5):844–882.
- Gelfand, A. E. y 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., y 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.
- González-Manteiga, W., Martínez-Miranda, M. D., y Raya-Miranda, R. (2008). Mean reversion and volatility of short-term london interbank offer rates. an empirical comparison of competing models. *Statistics and Computing*, (18):297–312.
- Granger, C. W. J. y Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Gottingen.
- Granger, C. W. J. y Joyeux, R. (2001). *An introduction to long-memory time series models and fractional differencing*, pages 321–337. Harvard University Press, Cambridge, MA, USA.
- Hall, P. y Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, 85:1039–1049.
- Hall, P. y Heyde, C. C. (1980). *Martingale limit Theory and Its Application*. Academic Express, Inc, New York.
- 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.
- Härdle, W. y Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1996–1947.

- Härdle, W. y Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, 18:63–89.
- Hart, J. D. (1994). Automated kernel smoothing of dependent data by using time series crossvalidation. *Journal Royal Statistic Society*, (56):529–542.
- Hart, J. D. (1996). Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistic*, (2-3):115–142.
- Harvey, A., Ruiz, E., y Shephard, N. (1994). Review of economic studies. *Multivariate stochastic variance models*, 61:247–264.
- Hastie, T. J. y Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Heath, D., Jarrow, R., y Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105.
- Ho, T. S. Y. y Lee, S. B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, (41):1011–1029.
- Hong, Y. y Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *The Review of Financial Studies*, 18(1):844–882.
- Hong, Y. y White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica*, 63:1133–1159.
- Hornik, K. (1993). Some new results on neural network approximation. *Neural Networks*, 6:1609–1072.
- Horowitz, J. L. y Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69:599–631.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176.
- Hull, J. y White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.
- Hull, J. y White, A. (1990). Pricing interest rate derivative counties. *Review of Financial Studies*, (3):573–592.
- Iacus, S. M. (2008). *Simulation and Inference for stochastics Differential Equations*. Springer.
- Jacquier, E., Polson, N. G., y Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics*, 12:371–417.
- Johanes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, 59(1):227–260.
- 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. y 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.
- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24:1–19.
- Kitagawa, G. (1998). A self-organizing state space model. *Journal of the American Statistical Association*, 93:1203–1215.
- Kloden, P. y Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Applications of mathematics , Springer, Berlin.
- Kloden, P., Platen, E., y Schurz, H. (2000). *Numerical Solution of SDE through Computer Experiments*. Springer, Berlin.
- Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48:1086–1101.
- Koul, H. L. y Stute, W. (1999). Nonparametric model checks for time series. *Annals of Statistics*, 27(1):204–236.
- Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, 90:1085–1093.
- Kutoyants, Y. A. (2004). *Statistical Inference for ergodic diffusion processes*. Springer, London.
- Kutoyants, Y. A. (2010). On the goodness-of-fit testing for ergodic diffusion processes. *Journal of Nonparametric Statistics*, (4):529–543.
- Lee, S. y Wee, I.-S. (2008). Residual empirical process for diffusion processes. *Journal of the Korean Mathematical Society*, 45(3):683–693.
- Li, F. (2007). Testing the parametric specification of the diffusion function in a diffusion process. *Econometric Theory*, 23:221–250.
- Liu, J. y Brockwell, P. J. (1988). On the general bilinear time-series model. *Journal of Applied Probability*, 25:553–564.
- Ljung, G. y Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 66:67–72.
- Longstaff, F. A. y Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance*, 47(4):1259–82.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4):394–419.
- Masuda, H., Negri, I., y Nishiyama, Y. (2010). Goodness of fit test for ergodic diffusion processes by discrete time observations: an innovation martingale approach. *Journal of Nonparametric Statistics*, to appear:available at <http://www.ism.ac.jp/~nisiyama/list.html>.
- McCulloch, R. E. y 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. y Li, W. K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4:260–273.
- Melino, A. y 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.
- Milstein, G. N. (1978). A method of second-order accuracy integration of stochastic differential equations. *Theory of Probability and its Applications*, 23(2):396–401.
- Miron, P. y Swannell, P. (1991). *Pricing and hedging swaps*. Euromoney Books.
- Monsalve-Cobis, A., González-Manteiga, W., y Febrero-Bande, M. (2011). Goodness of fit test for interest rates models: an approach based on empirical processes. *Computational Statistics & Data Analysis*, To Appear:–.
- Nadaraya, E. A. (1964). On estimating regression. *Theory and Probability Application*, 10:186–190.
- Negri, I. y Nishiyama, Y. (2009). Goodness of fit test for ergodic diffusion processes. *Annals of the Institute of Statistical Mathematics*, 61(4):167–198.
- Negri, I. y Nishiyama, Y. (2010). Goodness of fit test for ergodic diffusion processes by tick time sample scheme. *Statistical Inference for Stochastic Processes*, 13:81–95.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370.
- Neumeyer, N. y Dette, H. (2003). Nonparametric comparison of regression curves : an empirical process approach. *Annals of Statistics*, 31(3):880–920.
- Newey, W. y West, K. (1987). A simple positive semidefinite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:863–898.
- Ng, S. y 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., y 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.
- Pan, Z. y Lin, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics*, 61(4):1000–1009.
- Pardo-Fernández, J. C., Van-Keilegom, I., y González-Manteiga, W. (2007). Testing for the equality of  $k$  regression curves. *Statistica Sinica*, 17.
- Park, C. y Kang, K.-H. (2008). Sizer analysis for the comparison of regression curves. *Computational Statistics and Data Analysis*, 52:3954–3970.
- Petrucelli, J. y 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. y 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. y 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.
- Smith, D. R. (2002). Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, 20:183–197.
- So, M. K., Lam, K., y Li, W. K. (1998). Journal of business and economic statistics. *A stochastic volatility model with Markov switching*, 16:244–253.
- Song, J. y Lee, S. (2009). Test for parameter change in discretely observed diffusion processes. *Statistical Inference for Stochastic Processes*, 12:165–183.
- Stanton, R. (1997). A nonparametric models of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52(5):1973–2002.
- Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25(2):613–641.
- Stute, W., González-Manteiga, W., y Presedo-Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441):141–149.
- Suba Rao, T. y 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. y 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. y 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.
- Uhlenbeck, G. E. y Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, 36(5):823–841.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.
- Vilar-Fernández, J. M. y González-Manteiga, W. (2004). Nonparametric comparison of curves with dependent errors. *Statistics*, 38:81–99.
- 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.
- Zheng, J. X. (1996). A consistent test of a functional form via nonparametric estimation techniques. *J. Econ.*, (75):263–289.