



Universidade de Vigo

Trabajo Fin de Máster

Modelos de tipo de interés: revisión y comparativa de métodos de estimación.

Alejandra María López Pérez

Máster en Técnicas Estadísticas
Curso 2017-2018

Propuesta de Trabajo Fin de Máster

<p>Título en galego: Modelos de tipo de interese: revisión e comparativa de métodos de estimación.</p>
<p>Título en español: Modelos de tipo de interés: revisión y comparativa de métodos de estimación.</p>
<p>English title: Interest rate models: a review and comparison of estimation methods.</p>
<p>Modalidad: Modalidad A</p>
<p>Autor/a: Alejandra María López Pérez, Universidade de Santiago de Compostela</p>
<p>Director/a: Manuel Febrero Bande, Universidade de Santiago de Compostela; Wenceslao González Manteiga, Universidade de Santiago de Compostela</p>
<p>Breve resumen del trabajo:</p> <p>En el ámbito financiero, los modelos en tiempo continuo, en la perspectiva estocástica, se han modelizado mediante procesos de difusión y han adquirido especial relevancia en la caracterización de la dinámica de las principales variables económicas, entre ellas el tipo de interés.</p> <p>En las últimas décadas, han surgido numerosas propuestas para la modelización de la estructura temporal de los tipos de interés, que en lugar de considerar un enfoque determinista optan por llevar a cabo la modelización en un entorno estocástico con los modelos de difusión en tendencia y volatilidad.</p> <p>El objetivo de este trabajo fin de máster es revisar distintos modelos de difusión en tiempo continuo para los tipos de interés y comparar las técnicas de estimación paramétrica y no-paramétrica de las componentes de los procesos asociados.</p>
<p>Recomendaciones: Haber cursado la asignatura de series de tiempo.</p>
<p>Otras observaciones: Este trabajo se perfila como la semilla de una futura codirección de tesis doctoral por parte de los profesores que dirigen el trabajo.</p>

Don Manuel Febrero Bande, Catedrático de Universidad de la Universidade de Santiago de Compostela, y don Wenceslao González Manteiga, Catedrático de Universidad de la Universidade de Santiago de Compostela, informan que el Trabajo Fin de Máster titulado

Modelos de tipo de interés: revisión y comparativa de métodos de estimación.

fue realizado bajo su dirección por doña Alejandra María López Pérez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 26 de enero de 2018.

El director:

Don Manuel Febrero Bande

El director:

Don Wenceslao González Manteiga

La autora:

Doña Alejandra María López Pérez

Índice general

Resumen	ix
1. Introducción	1
2. Modelización de los tipos de interés	3
2.1. Procesos estocásticos y ecuaciones diferenciales estocásticas	3
2.2. Familias paramétricas de procesos estocásticos	5
2.2.1. Modelo de Merton o Aritmético Browniano	6
2.2.2. Modelo de Black-Scholes o Geométrico Browniano	6
2.2.3. Modelo de Vasicek o proceso de Ornstein-Uhlenbeck	7
2.2.4. Modelo de Dothan	8
2.2.5. Modelo de Cox-Ingersoll-Ross	8
2.2.6. Modelo de elasticidad constante de la varianza	9
2.2.7. Modelo de Chan-Karolyi-Longstaff-Sanders	9
2.2.8. Modelo de Aït-Sahalia	10
2.2.9. Modelo Inverse-Feller	11
3. Métodos de estimación	13
3.1. Estimación paramétrica	15
3.1.1. Máxima verosimilitud exacta	16
3.1.2. Máxima verosimilitud aproximada	17
3.1.3. Método de máxima verosimilitud de Aït-Sahalia	18
3.1.4. Método Generalizado de los Momentos	22
3.1.5. Filtro de Kalman	23
3.2. Estimación no paramétrica	26
3.2.1. Estimación de la densidad estacionaria	26
3.2.2. Estimación de las funciones drift y difusión	28
4. Aplicación a datos simulados y reales	31

4.1. Estimación paramétrica	31
4.1.1. Estudio de simulación	34
4.2. Estimación no-paramétrica	37
4.2.1. Estudio de simulación	39
4.3. Procedimiento bootstrap	42
4.3.1. Justificación asintótica del procedimiento	44
4.3.2. Aplicación del procedimiento a la serie simulada	47
4.3.3. Aplicación del procedimiento a datos reales	49
5. Conclusiones	55
A. Código R	57
A.1. Simulación de trayectorias del proceso CIR	57
A.2. Máxima verosimilitud exacta	58
A.3. Máxima verosimilitud aproximada	59
A.4. Método de máxima verosimilitud de Aït-Sahalia	60
A.5. Método Generalizado de los Momentos	61
A.6. Filtro de Kalman	62
Bibliografía	65

Resumen

Resumen

Los modelos de difusión descritos por ecuaciones diferenciales estocásticas son una extensión natural de modelos deterministas con ecuaciones diferenciales ordinarias. Estos modelos han sido aplicados de forma intensiva en el campo de la matemática financiera para el análisis de precios de activos financieros, tipos de cambio, tipos de interés, etc.

Comprender y modelar la estructura temporal de los tipos de interés representa uno de los principales retos de la investigación financiera actual. Los tipos de interés son series que varían de forma aleatoria respecto al tiempo y estas variaciones pueden ser modelizadas mediante procesos estocásticos de dos tipos: procesos estocásticos en tiempo discreto y modelos en tiempo continuo.

En este trabajo se revisan distintos modelos en tiempo continuo, en el contexto estocástico, destinados a explicar la dinámica del tipo de interés. Se presentan diferentes métodos de estimación, paramétricos y no paramétricos, así como su aplicación a datos reales y simulados. Por último, se emplea la metodología bootstrap para aproximar la distribución de los estimadores.

Abstract

Diffusion models described by stochastic differential equations are natural extensions of deterministic models described by ordinary differential equations. These models have been applied intensively in the mathematical finance field for the analysis of assets prices, exchange rates, interest rates, etc.

Understanding and modeling the term structure of interest rates represents one of the main challenges of current financial research. Interest rates are series that change randomly with respect to time and these variations can be modeled by two types of stochastic processes: stochastic processes in discrete time and continuous time models.

In this work, we review different models in continuous time, in the stochastic context, designed to explain the dynamics of the interest rate. Parametric and nonparametric estimation methods are presented, as well as their application to real and simulated data. Lastly, the bootstrap methodology is used to approximate the estimators distribution.

Capítulo 1

Introducción

Los modelos de difusión descritos por ecuaciones diferenciales estocásticas son una extensión natural de modelos deterministas con ecuaciones diferenciales ordinarias. Estos modelos han sido aplicados de forma intensiva en el campo de la matemática financiera para el análisis de precios de activos financieros, tipos de cambio, tipos de interés, etc. Comprender y modelar la estructura temporal de los tipos de interés representa uno de los principales retos de la investigación financiera actual. Los tipos de interés son series que varían de forma aleatoria respecto al tiempo y estas variaciones pueden ser modelizadas mediante procesos estocásticos de dos tipos: procesos estocásticos en tiempo discreto y modelos en tiempo continuo. Respecto a estos últimos, si bien los modelos están formulados en tiempo continuo, los datos son registrados en puntos discretos del tiempo (por ejemplo, cada minuto, diariamente o semanalmente), lo cual complica de forma considerable el análisis estadístico.

La necesidad de desarrollar el análisis de la estructura temporal de los tipos de interés en un entorno estocástico surge como consecuencia de las turbulencias del mercado financiero de la década de los setenta. La aplicación de la metodología de Black y Scholes (1973) a la valoración de activos derivados del tipo de interés no es posible, ya que su valor no depende del precio de ningún activo subyacente. Además, los precios de los bonos cupón cero convergen a su valor nominal al vencimiento, lo que da lugar a que la volatilidad de los bonos disminuya a lo largo del tiempo. Progresivamente fueron surgiendo teorías propias de la estructura de los tipos de interés que se basan en modelos de valoración por ausencia de arbitraje en un entorno estocástico. En los años 70, Merton (1973) modelizó el tipo de interés como un proceso estocástico y lo empleó para la valoración de opciones. Posteriormente, se empezaron a aplicar argumentos de arbitraje similares a los de Black y Scholes (1973) para modelizar la estructura temporal de los tipos de interés, como muestran los trabajos de Vasicek (1977) o de Brennan y Schwartz (1979). Esta idea contribuyó a la construcción de nuevos modelos más perfeccionados, como los de Cox, Ingersoll, y Ross (1985), Chan et al. (1992) o Ahn y Gao (1999). Estos modelos permiten especificar el tipo de interés como solución de una ecuación diferencial estocástica, permitiendo el uso de la Teoría de los Procesos de Markov y facilitando así su tratamiento analítico. Aunque cabe notar que, a medida que el proceso que explica la dinámica del tipo de interés es más realista, se complica y, en ocasiones, no es posible obtener una solución exacta.

En este trabajo se revisan distintos modelos de difusión en tiempo continuo para los tipos de interés, así como las técnicas de estimación paramétrica y no-paramétrica de las componentes de los procesos asociados. Para ello, en el Capítulo 2 se expone la teoría relativa a los procesos estocásticos y ecuaciones diferenciales estocásticas, así como los

distintos modelos utilizados en la modelización de los tipos de interés. En el Capítulo 3 veremos técnicas de estimación paramétricas y no paramétricas de las componentes de los procesos asociados y en el Capítulo 4 compararemos estos métodos de estimación, mediante una aplicación a datos simulados y a datos reales. Por último, en el Capítulo 5 se presentan las conclusiones del trabajo.

Capítulo 2

Modelización de los tipos de interés

El precio de los activos financieros evolucionan en el tiempo formando un proceso estocástico. Existen dos tipos de procesos estocásticos para modelar los precios de un activo: por un lado, los procesos estocásticos en *tiempo discreto*, donde el precio varía en puntos discretos del tiempo; por otro lado, tenemos los procesos estocásticos en *tiempo continuo*, donde el precio varía de manera continua, aunque sólo sea observado en puntos discretos del tiempo. Para ambos procesos, el precio puede ser continuo o discreto. El valor que puede tomar un precio continuo es cualquier número real positivo, pero el precio discreto sólo puede tomar un número finito de valores positivos.

En este trabajo trataremos los precios (tipos de interés) como un proceso estocástico continuo en tiempo continuo, el cual se define en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde Ω es un conjunto, \mathcal{F} es una σ -álgebra de subconjuntos de Ω y \mathbb{P} es una medida de probabilidad, $\mathbb{P}(\Omega) = 1$. En el contexto financiero, cuando utilizamos modelos en tiempo continuo, es habitual asumir que el tipo de interés es un proceso de Itô.

2.1. Procesos estocásticos y ecuaciones diferenciales estocásticas

Las integrales estocásticas y, en particular, la integral de Itô, son una generalización de la integral de Riemann-Stieltjes que fueron introducidas para poder definir de manera rigurosa las ecuaciones diferenciales estocásticas. Las ecuaciones diferenciales estocásticas involucran una variable aleatoria cuyos incrementos en el tiempo son descritos por una parte determinista más una parte estocástica dada por un proceso de Wiener.

Consideremos como ejemplo el valor de una acción en el momento t . Denotado por S_t el precio del *stock* en t y por $\Delta S = S_{t+\Delta t} - S_t$ el incremento de S en un intervalo de tiempo pequeño $[t, t + \Delta t)$, la rentabilidad de la acción viene dada por $\Delta S/S$. Asumiendo que la parte determinista depende del tipo de interés del activo sin riesgo y es, por tanto, proporcional al tiempo en una constante μ , tenemos que la parte determinista es $\mu\Delta t$. Como veremos más adelante, también podemos considerar una función que dependa de t o S_t en lugar de la constante μ . Respecto a la parte estocástica, asumimos que depende de algún ruido aleatorio y de la propia naturaleza volátil del mercado. Denotando por $\Delta W = W_{t+\Delta t} - W_t$ a los incrementos del proceso aleatorio y σ a la volatilidad del mer-

cado, la parte estocástica del modelo viene dada por $\sigma\Delta W$. Notar que también podemos considerar en lugar de la constante σ una función de t y/o S . La hipótesis natural es asumir el comportamiento Gaussiano del ruido, con $\Delta W \sim N(0, 1)$, lo que implica que W es un proceso de Wiener, o movimiento Browniano estandarizado, si, además, suponemos que los incrementos son independientes. Por todo lo anterior, podemos escribir

$$\frac{\Delta S}{S} = \mu\Delta t + \sigma\Delta W.$$

Si consideramos la ecuación diferencial anterior para intervalos infinitesimales, $\Delta t \rightarrow 0$, para obtener una ecuación diferencial estocástica podemos escribir $S'_t/S_t = \mu + \sigma W'_t$, donde $S'_t = \mu S_t + \sigma S_t W'_t$, que también podemos expresar en forma diferencial

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \quad (2.1)$$

Esta ecuación debe ser interpretada como una expresión informal, es un ejemplo de ecuación diferencial estocástica pero la expresión, matemáticamente, no tiene significado, ya que dW_t no es finito y el proceso de Wiener tiene trayectorias continuas pero no diferenciables. La ecuación (2.1) es rigurosamente interpretable en forma de ecuación integral,

$$S_t - S_0 = \mu \int_0^t S(u) du + \sigma \int_0^t S(u) dW(u),$$

donde la última de las integrales es una integral de Itô, mientras que la primera es simplemente una integral de Riemann. Las propiedades y la construcción de la integral estocástica pueden consultarse en Steele (2001).

En este trabajo consideraremos los procesos de difusión solución de la ecuación diferencial estocástica de la forma

$$dr_t = m(r_t, t) dt + \sigma(r_t, t) dW_t, \quad \text{con } r_0 > 0, t \geq 0, \quad (2.2)$$

o escrito en forma de ecuación integral,

$$r_t - r_0 = \int_0^t m(r_u, u) du + \int_0^t \sigma(r_u, u) dW_u.$$

Las funciones $m(\cdot, t)$ y $\sigma(\cdot, t)$ son, respectivamente, el *drift* y la *difusión* o *volatilidad* de la ecuación diferencial estocástica (2.2). Estos coeficientes pueden ser interpretados como medida del crecimiento esperado en el corto plazo y la variabilidad en el corto plazo, de modo que (2.2) facilita un patrón para la construcción de procesos estocásticos que reflejen el comportamiento de variables reales. Notar que el proceso de Wiener es un caso particular del proceso de Itô porque satisface la ecuación (2.2) con $m(r_t, t) = 0$ y $\sigma(r_t, t) = 1$.

Los procesos de difusión poseen la propiedad de Markov y pueden ser ergódicos. La ergodicidad implica que, para cualquier función medible $h(\cdot)$, el resultado

$$\frac{1}{T} \int_0^T h(r_t) dt \rightarrow \int_{-\infty}^{\infty} h(r) \pi(r) dr = \mathbb{E}[h(\xi)],$$

donde $\pi(\cdot)$ es la *densidad estacionaria* del proceso de difusión y ξ es una variable aleatoria con densidad $\pi(\cdot)$, se cumple con probabilidad 1. La distribución estacionaria de los procesos de difusión puede ser expresada en términos de la *medida de escala*,

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

y de la *medida de rapidez*,

$$b(x) = \frac{1}{\sigma^2(x)s(x)}.$$

En particular, la densidad estacionaria $\pi(\cdot)$ es proporcional a la medida de rapidez $b(x)$,

$$\pi(x) = \frac{b(x)}{M},$$

donde $M = \int b(x)dx$ es una constante de normalización.

La naturaleza Markoviana del proceso también nos permite definir la densidad de transición del valor x en el momento s al valor y en el momento t como $p(t, y | s, x)$, o $p(t - s, y | x)$. La densidad de transición satisface la *ecuación "forward" de Kolmogorov*,

$$\frac{\partial}{\partial t} p(t, y | s, x) = -\frac{\partial}{\partial y} [m(y)p(t, y | s, x)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [\sigma^2(y)p(t, y | s, x)],$$

y la *ecuación "backward" de Kolmogorov*,

$$-\frac{\partial}{\partial s} p(t, y | s, x) = m(x) \frac{\partial}{\partial x} p(t, y | s, x) + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} p(t, y | s, x).$$

Si $t \rightarrow -\infty$ en la ecuación "forward" de Kolmogorov, obtenemos

$$\frac{\partial^2}{\partial x^2} [\sigma^2(x)\pi(x)] = 2 \frac{\partial}{\partial x} [m(x)\pi(x)].$$

Esta ecuación establece una relación entre el drift $m(\cdot)$, el coeficiente de difusión $\sigma(\cdot)$ y la densidad estacionaria $\pi(\cdot)$.

2.2. Familias paramétricas de procesos estocásticos

A lo largo de este trabajo nos centraremos en los procesos de difusión homogéneos, donde $\Delta_n = \Delta$, un caso particular de (2.2) y definido como solución de la ecuación diferencial estocástica,

$$dr_t = m(r_t) dt + \sigma(r_t) dW_t, \quad \text{con } r_0 > 0, \quad 0 \leq t \leq T,$$

donde $m(\cdot)$ y $\sigma(\cdot)$ no dependen del tiempo. Representado en su forma integral tenemos

$$r_t - r_0 = \int_0^t m(r_u) du + \int_0^t \sigma(r_u) dW_u, \quad 0 \leq t \leq T.$$

Podemos estudiar este proceso con dos enfoques: no-paramétrico y paramétrico. En el primero, las funciones $m(r)$ y $\sigma(r)$, con $r \in \mathbb{R}$, son desconocidas. En la sección 3 veremos métodos de estimación no-paramétricos para ambas funciones. En el enfoque paramétrico consideramos el proceso r_t como una solución de

$$dr_t = m(r_t; \theta) dt + \sigma(r_t; \theta) dW_t, \quad \text{con } r_0 > 0, \quad 0 \leq t \leq T,$$

y tenemos en cuenta el conocimiento de las funciones $m(r) = m(r, \theta)$ y $\sigma(r) = \sigma(r, \theta)$, con $r \in \mathbb{R}$ y θ un parámetro desconocido tal que $\theta \in \Theta \subset \mathbb{R}^d$. En el contexto paramétrico, existen procesos estocásticos ampliamente utilizados en el ámbito financiero que introduciremos a continuación.

2.2.1. Modelo de Merton o Aritmético Browniano

Desarrollado por Merton (1973), fue una de las primeras propuestas para modelizar el tipo de interés a través de un proceso estocástico. El modelo asume que las funciones drift y difusión son constantes, $m(r_t, \theta) = \alpha$ y $\sigma(r_t, \theta) = \sigma$, de modo que el tipo de interés a corto plazo se comporta como un movimiento Browniano con tendencia, también denominado movimiento Browniano aritmético,

$$dr_t = \alpha dt + \sigma dW_t,$$

donde $\sigma > 0$ y r_0 es la condición inicial de la ecuación. Una ventaja del modelo de Merton es que proporciona soluciones analíticas para la estructura temporal de los tipos de interés, pero, por otro lado, bajo este modelo el tipo de interés puede tomar valores negativos. Además, se puede demostrar que con este proceso el precio de un bono cupón cero es una función creciente del tipo de interés, por lo que un bono cupón cero con vencimiento infinito tendría un valor infinito, que no es un resultado óptimo.

2.2.2. Modelo de Black-Scholes o Geométrico Browniano

Desarrollado por Black y Scholes (1973), se introdujo como modelo para la valoración de derivados financieros, al que también se le conoce como *Black-Scholes-Merton* tras su introducción en el contexto financiero. El movimiento Browniano geométrico es el proceso asociado al modelo, de modo que caracteriza el tipo de interés mediante la solución de la ecuación diferencial estocástica

$$dr_t = \alpha r_t dt + \sigma r_t dW_t,$$

donde $\sigma > 0$ y r_0 es la condición inicial de la ecuación. El parámetro α se puede interpretar como el tipo de interés constante y σ como la volatilidad asociada a los activos con riesgo. La solución explícita del modelo es

$$r_t = r_0 \exp \left\{ \left(\alpha - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right\}.$$

La solución r_t tiene densidad condicional log-normal con media y varianza (de su transformación logarítmica) dada por

$$\begin{aligned} \mu_{\log r} &= \log(r_0) + \left(\alpha - \frac{1}{2} \sigma^2 \right) t, \\ \sigma_{\log r}^2 &= \sigma^2 t, \end{aligned}$$

por lo que su media y varianza son

$$\begin{aligned} \mathbb{E}[r_t | r_0] &= \exp \left\{ \mu_{\log r} + \frac{1}{2} \sigma_{\log r}^2 \right\} = r_0 e^{\alpha t}, \\ \text{Var}[r_t | r_0] &= \exp \left\{ 2\mu_{\log r} + \sigma_{\log r}^2 \right\} \left(\exp \left\{ \sigma_{\log r}^2 \right\} - 1 \right) = r_0^2 e^{2\alpha t} (e^{t\sigma^2} - 1). \end{aligned}$$

La densidad condicional, o de transición, es

$$p_\theta [t, x | r_0] = \frac{1}{x\sigma\sqrt{2\pi t}} \exp \left\{ -\frac{[\log(x) - (\log(r_0) + (\alpha - \frac{1}{2}\sigma^2)t)]^2}{2\sigma^2 t} \right\}.$$

A diferencia del modelo de Merton, en esta especificación el tipo de interés nunca toma valores negativos. Además, para $\alpha > 0$ el proceso revierte a la media y la volatilidad se caracteriza por ser cambiante en el tiempo en función del nivel del tipo de interés.

2.2.3. Modelo de Vasicek o proceso de Ornstein-Uhlenbeck

Vasicek (1977) propuso un modelo basado en el proceso de *Ornstein-Uhlenbeck* (Ornstein y Uhlenbeck, 1930) para modelizar el tipo de interés. El proceso es la solución única de la ecuación diferencial estocástica

$$dr_t = (\alpha - \beta r_t) dt + \sigma dW_t,$$

donde α , β y σ son constantes positivas y r_0 es la condición inicial de la ecuación. Cuando $\beta > 0$ se dice que el proceso *revierte a la media* y, a diferencia del movimiento Browniano geométrico, el proceso tiene varianza finita para todo $t \geq 0$. Con la parametrización anterior, σ se interpreta como la volatilidad, α/β es el valor de equilibrio a largo plazo y β es la velocidad de la reversión. Aplicando la fórmula de Itô, obtenemos la solución explícita de la ecuación,

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

Para cualquier $t \geq 0$, el proceso tiene función de densidad condicional, o de transición, $p_\theta [t, r_t | r_0]$ Gaussiana con media

$$\mathbb{E}[r_t | r_0] = \frac{\alpha}{\beta} + \left(r_0 - \frac{\alpha}{\beta}\right) e^{-\beta t},$$

varianza

$$\text{Var}[r_t | r_0] = \frac{\sigma^2}{2\beta} \left(1 - e^{-2\beta t}\right),$$

y covarianza

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

Cuando $\beta > 0$, el proceso es ergódico y admite una distribución de probabilidad estacionaria. Notar que

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E}[r_t | r_0] &= \frac{\alpha}{\beta}, \\ \lim_{t \rightarrow \infty} \text{Var}[r_t | r_0] &= \frac{\sigma^2}{2\beta}. \end{aligned}$$

La densidad estacionaria del proceso es Gaussiana con media α/β y $\sigma^2/2\beta$,

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

Mientras que el hecho de que el tipo de interés posea reversión a la media es una característica defendida en la literatura, es habitual argumentar que la volatilidad depende del nivel del tipo de interés, en lugar de ser constante. Otro inconveniente del modelo es que el tipo de interés podría tomar valores negativos. Pero las ventajas del modelo – entre las que destaca que es analíticamente tratable, proporcionando una solución exacta para la estructura temporal de los tipos de interés– han contribuido a extender su uso en la literatura.

2.2.4. Modelo de Dothan

Dothan (1978) introdujo una modificación en el modelo de Vasicek, asumiendo que la volatilidad del tipo de interés es proporcional a la propia tasa, por lo que la volatilidad varía a lo largo del tiempo,

$$dr_t = \sigma r_t dW_t.$$

En este modelo el tipo de interés se rige por un movimiento Browniano sin tendencia, de modo que el tipo de interés se distribuye como una variable aleatoria lognormal, pudiendo, por tanto, tomar valores negativos. El modelo de Dothan también se denomina *camino aleatorio geométrico* o *camino aleatorio elástico*, y fue utilizado con anterioridad por Brennan y Schwartz (1977) para la valoración de bonos con opciones incorporadas. Courtadon (1982a) demostró que el modelo no es adecuado para representar el comportamiento a largo plazo del tipo de interés porque $\lim_{n \rightarrow \infty} r_{t_n} = 0$.

El uso del modelo no está muy extendido, ya que modelizar el comportamiento de los tipos de interés mediante una lognormal eleva la complejidad y presenta problemas de estabilidad. Además, no es posible obtener una solución exacta para el precio de los bonos cupón cero ni para las opciones, aunque Brennan y Schwartz (1977) y Dothan (1978) han utilizado el modelo para la valoración de distintos bonos utilizando técnicas de resolución numéricas.

2.2.5. Modelo de Cox-Ingersoll-Ross

El proceso Cox-Ingersoll-Ross fue introducido por Feller (1951) para el estudio del crecimiento de poblaciones y, posteriormente, Cox, Ingersoll y Ross (1985) popularizaron su uso en la modelización de los tipos de interés a corto plazo. Los autores introducen una modificación en el modelo de Vasicek, considerando en el coeficiente de difusión la raíz cuadrada del tipo de interés. Así, el proceso CIR es la solución a la ecuación diferencial estocástica

$$dr_t = (\alpha - \beta r_t) dt + \sigma \sqrt{r_t} dW_t, \quad \text{con } r_0 > 0, \quad (2.3)$$

donde α , β y σ son constantes positivas. Bajo la condición $2\alpha > \sigma^2$, el proceso es estrictamente positivo (en otro caso, es no negativo, lo que significa que puede alcanzar el estado 0). También es habitual utilizar la parametrización

$$dr_t = \kappa(\mu - r_t) dt + \sigma \sqrt{r_t} dW_t,$$

donde $\kappa = \beta$ es la velocidad de la reversión a la media y $\mu = \alpha/\beta$ es la tasa a la que revierte el tipo de interés. La ecuación diferencial estocástica (2.3) tiene solución explícita, dada por

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

Bajo la hipótesis $2\alpha > \sigma^2$, el proceso es estacionario y la densidad de transición condicional existe de forma explícita. Dada la información hasta el instante s , \mathcal{F}_s , el tipo de interés a corto plazo r_t se distribuye como una ji-cuadrado no central,

$$r_t \mid \mathcal{F}_s \sim \chi^2(2cr_t, 4\alpha/\sigma^2, 2u),$$

con $4\alpha/\sigma^2$ grados de libertad y parámetro no central $2u$, donde

$$c = \frac{2\beta}{\sigma^2(1 - e^{-\beta t})}, \quad u = cr_0 e^{-\beta t}, \quad v = cr, \quad q = \frac{2\alpha}{\sigma^2} - 1.$$

La distribución puede ser escrita de forma explícita como

$$p(t, r | r_0) = ce^{u+v} \left(\frac{u}{v}\right)^{q/2} I_q(2\sqrt{uv})$$

donde

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

es la función de Bessel modificada de primera especie y orden q y $\Gamma(\cdot)$ es la función gamma, $\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, p \in \mathbb{R}_+$. La media y varianza condicional de r_t son

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

y la función de covarianzas es

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

La distribución estacionaria del proceso CIR sigue una Gamma con parámetro de forma $2\alpha/\sigma^2$ y parámetro de escala $\sigma^2/2\beta$, por lo que la distribución estacionaria tiene media α/β , varianza $\alpha\sigma^2/(2\beta^2)$ y covarianza

$$\text{Cov} = \frac{\alpha\sigma^2}{2\beta^2} e^{-\beta(t-s)}.$$

El proceso, al igual que el planteado por Vasicek (1977), supone que el tipo de interés posee reversión a la media y que la volatilidad depende del tipo de interés, hipótesis más realista que la de volatilidad constante. El modelo CIR ha sido ampliamente desarrollado y utilizado en la literatura, ya que proporciona una solución analítica para la estructura temporal de los tipos de interés, así como para distintos derivados de los tipos de interés.

2.2.6. Modelo de elasticidad constante de la varianza

El proceso de elasticidad constante de la varianza (CEV) fue introducido para la valoración de opciones (Beckers, 1980). Es la solución de la ecuación diferencial estocástica

$$dr_t = \beta r_t dt + \sigma r_t^\gamma dW_t$$

con $\gamma \geq 0$. Para $\gamma < 1$ la volatilidad presenta asimetría negativa y para $\gamma > 1$ la asimetría es positiva. Notar que cuando $\gamma = 1$, es un caso particular del movimiento Browniano geométrico. El proceso se asume que es positivo, pero la versión discretizada del mismo puede tomar valores negativos. En concreto, para $\gamma < 0.5$ hay probabilidad positiva de que el proceso sea absorbido en el cero.

2.2.7. Modelo de Chan-Karolyi-Longstaff-Sanders

Chan et al. (1992) propusieron el modelo CKLS, que generaliza el proceso de volatilidad constante con reversión a la media, dado por la ecuación diferencial estocástica

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

con $\sigma > 0$. Es una extensión del proceso CIR y generaliza los modelos vistos hasta ahora, como podemos ver en el Cuadro 2.1. No admite una función de densidad de transición explícita y para valores $\alpha, \beta > 0$ y $\gamma > 0.5$, el proceso toma valores en $(0, +\infty)$.

Modelos	Restricciones				Ecuación diferencial estocástica
	α	β	σ^2	γ	
Merton (ABM)		0		0	$dr_t = \alpha dt + \sigma dW_t$
Vasicek				0	$dr_t = (\alpha + \beta r_t) dt + \sigma dW_t$
CIR SR				1/2	$dr_t = (\alpha + \beta r_t) dt + \sigma r_t^{1/2} dW_t$
Dothan	0	0		1	$dr_t = \sigma r_t dW_t$
GBM	0			1	$dr_t = \beta r_t dt + \sigma r_t dW_t$
Brennan				1	$dr_t = (\alpha + \beta r_t) dt + \sigma r_t dW_t$
CIR VR	0	0		3/2	$dr_t = \sigma r_t^{3/2} dW_t$
CEV	0				$dr_t = \beta r_t dt + \sigma r_t^\gamma dW_t$
CKLS					$dr_t = (\alpha + \beta r_t) dt + \sigma r_t^\gamma dW_t$

Cuadro 2.1: Especificaciones paramétricas de la familia de procesos CKLS.

En estos modelos tenemos que

$$m(r_t; \theta) = \theta_1 + \theta_2 r_t,$$

$$\sigma^2(r_t; \theta) = \theta_3^2 r_t^{2\theta_4},$$

donde θ_3^2 es el parámetro de volatilidad que representa un factor de escala de las variaciones no anticipadas de los tipos de interés y $2\theta_4$ es la *elasticidad de la varianza*, la sensibilidad de la volatilidad ante variaciones de los tipos de interés. Este tipo de modelos recogen la relación entre la volatilidad y el nivel de los tipos de interés, ya que a medida que el nivel de los tipos de interés aumenta se espera que la volatilidad también se incremente. Aunque, en ocasiones, valores muy elevados de θ_4 no dan lugar a este efecto en determinados períodos de tiempo.

2.2.8. Modelo de Aït-Sahalia

Aït-Sahalia (1996a) propone un modelo no lineal con reversión a la media para modelizar el tipo de interés. El modelo satisface la ecuación diferencial estocástica

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

Aunque no es posible obtener la distribución exacta del modelo, es factible aproximar la densidad de transición mediante una expansión de polinomios de Hermite, como veremos más adelante. El autor también propone una generalización del modelo, incluyendo una función de difusión más compleja,

$$dr_t = (\alpha_{-1} r_t^{-1} + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2) dt + (\beta_0 + \beta_1 r_t + \beta_2 r_t^\gamma) dW_t.$$

En este modelo, la propiedad de Markov sólo está garantizada bajo ciertas restricciones. Para obtener especificaciones razonables del modelo es necesario establecer restricciones sobre los coeficientes, éstas pueden consultarse en Aït-Sahalia (1996a). Notar que el modelo CKLS es un caso particular del modelo anterior.

2.2.9. Modelo Inverse-Feller

El modelo de Ahn y Gao, o modelo *Inverse-Feller*, fue propuesto por Ahn y Gao (1999) basándose en los trabajos de Aït-Sahalia (1996a) y Aït-Sahalia (1996b). El modelo es una transformación del CIR, donde el tipo de interés sigue una tendencia no lineal y la volatilidad posee elasticidad constante de la varianza,

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

Su distribución condicional está relacionada con la del modelo CIR,

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

donde p^{CIR} es la densidad condicional del proceso CIR, dada por (2.2.5).

Capítulo 3

Métodos de estimación

En esta sección revisaremos distintos métodos de estimación de procesos de difusión en tiempo continuo con muestras en tiempo discreto. Es importante destacar que sólo las observación están en tiempo discreto, ya que el modelo subyacente es en tiempo continuo. Consideraremos la ecuación diferencial estocástica homogénea, en su forma paramétrica,

$$dr_t = m(r_t, \theta) dt + \sigma(r_t, \theta) dW_t,$$

para el parámetro desconocido $\theta \in \Theta \subset \mathbb{R}^d$ con d un entero positivo. Las funciones drift y difusión, $m : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ y $\sigma : \mathbb{R} \times \Theta \rightarrow (0, \infty)$, respectivamente, se consideran conocidas, de modo que existe la solución de la ecuación diferencial estocástica¹. Cuando la densidad estacionaria $\pi_\theta(\cdot)$ existe, tiene la forma

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

donde

$$s(r, \theta) = \exp \left\{ -2 \int_{r_0}^r \frac{m(y, \theta)}{\sigma^2(y, \theta)} dy \right\}$$

para algún $r_0 \in I$, con I el espacio de estado del proceso, y $M(\theta)$ una constante de normalización. Tal y como vimos en la Sección 2.1, la función $s(\cdot, \theta)$ es la *medida de escala* y $b(r) = \pi_\theta(r)M(\theta)$ es la *medida de rapidez*. La distribución de r , con $r_0 \sim \pi_\theta$, se denota por P_θ y, bajo esta medida, $r_t \sim \pi_\theta$ para todo t .

En los modelos especificados en tiempo continuo podemos calcular, en lugar de estimar, el subconjunto de parámetros de θ correspondiente a la difusión $\sigma(r_t, \theta)$ a partir de la variación cuadrática del proceso, ya que para todo $t \geq 0$,

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

cuando $n \rightarrow \infty$ en probabilidad bajo P_θ . Así, podríamos considerar $\sigma(r, \theta) = \sigma(r)$ conocido. El resto de los parámetros, correspondientes al drift $m(r_t, \theta)$, se pueden estimar por máxima verosimilitud. Dado el coeficiente de difusión, la función de verosimilitud de r viene dada por

$$\mathcal{L}_T(\theta) = \exp \left(\int_0^T \frac{m(r_s, \theta)}{\sigma^2(r_s)} dr_s - \frac{1}{2} \int_0^T \frac{m^2(r_s, \theta)}{\sigma^2(r_s)} ds \right).$$

¹Para garantizar la existencia de la solución de la ecuación diferencial estocástica, es necesario asumir ciertas suposiciones sobre el modelo (ver Iacus, 2008)

Maximizando $\mathcal{L}_T(\theta)$ obtendremos la estimación del verdadero parámetro θ_0 . Pero, aunque los modelos están formulados en tiempo continuo, los datos son registrados en puntos discretos del tiempo. Por ello, para la estimación de los parámetros del modelo en tiempo continuo deberemos considerar una versión discreta del mismo. El esquema de observación que asumiremos en este trabajo será el de *muestras numerosas*, bajo el cual el tiempo Δ entre dos observaciones consecutivas es fijo y el número de observaciones n es creciente, así como la ventana de observación $[0, T = n\Delta]$, que también se incrementa con n . Destacar que, en este enfoque, se requieren suposiciones adicionales sobre el modelo continuo subyacente, tales como estacionariedad y ergodicidad.

Uno de los esquemas de aproximación más utilizado es el método de Euler-Maruyama: dado un proceso de Itô $\{r_t, 0 \leq t \leq T\}$, solución de la ecuación diferencial estocástica

$$dr_t = m(r_t) dt + \sigma(r_t) dW_t, \quad (3.1)$$

con valor inicial $r_{t_0} = r_0$ y la discretización $\Pi_n = \Pi_n([0, T])$ del intervalo $[0, T]$, $0 = t_0 < t_1 < \dots < t_n = T$, la aproximación de Euler-Maruyama de r es un proceso estocástico continuo X que satisface el esquema iterativo

$$X_{t_{i+1}} - X_{t_i} = m(t_i, X_{t_i})(t_{i+1} - t_i) + \sigma(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i}), \quad i = 0, 1, \dots, n-1,$$

con $X_{t_0} = r_0$. Notar que los incrementos del tiempo $\Delta = t_{i+1} - t_i$ se han supuesto constantes. Entre dos puntos del tiempo t_i y t_{i+1} , para definir el proceso podemos considerar una interpolación lineal, de modo que X_t se definiría como

$$X(t) = X_{t_i} + \frac{t - t_i}{t_{i+1} - t_i}(X_{t_{i+1}} - X_{t_i}), \quad t \in [t_i, t_{i+1}).$$

Otros métodos de uso frecuente son los propuestos por Milstein (1978) y Kloeden et al. (2000), conocido como KPS (las iniciales de los autores, Kloeden, Platen y Schurz). El método de Milstein utiliza el lema de Itô para incrementar la precisión de la aproximación añadiendo un término de segundo orden,

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= m(t_i, X_{t_i})(t_{i+1} - t_i) + \sigma(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i}) \\ &\quad + \frac{1}{2}\sigma(t_i, X_{t_i})\sigma_r(t_i, X_{t_i})[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)], \end{aligned}$$

$i = 0, 1, \dots, n-1$, donde $\sigma_r(t, r) = \partial\sigma/\partial r$.

El método KPS añade más términos a la expansión Itô-Taylor, definiendo $\Delta W_t = W_{t_{i+1}} - W_{t_i}$ y $\Delta t = t_{i+1} - t_i$, tenemos

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= m(t_i, X_{t_i})\Delta t + \sigma(t_i, X_{t_i})\Delta W_t \\ &\quad + \frac{1}{2}\sigma(t_i, X_{t_i})\sigma_r(t_i, X_{t_i})[(\Delta W_t)^2 - \Delta t] \\ &\quad + \sigma(t_i, X_{t_i})m_r(t_i, X_{t_i})\Delta V_t + \frac{1}{2}\left[m(t_i, X_{t_i})m_r(t_i, X_{t_i})\right. \\ &\quad \left.+ \frac{1}{2}\sigma^2(t_i, X_{t_i})m_{rr}(t_i, X_{t_i})\right]\Delta t^2 + \left[m(t_i, X_{t_i})\sigma_r(t_i, X_{t_i})\right. \\ &\quad \left.+ \frac{1}{2}\sigma^2(t_i, X_{t_i})\sigma_{rr}(t_i, X_{t_i})\right](\Delta W_t\Delta t - \Delta V_t) \\ &\quad + \frac{1}{2}\sigma(t_i, X_{t_i})(\sigma(t_i, X_{t_i})\sigma_r(t_i, X_{t_i}))_r\left[\frac{1}{3}(\Delta W_t)^2 - \Delta t\right]\Delta W_t, \end{aligned}$$

donde

$$\Delta V_t = \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_s} dW_u ds$$

es una variable aleatoria Gaussiana con media cero y varianza $\frac{1}{3}\Delta t^3$, y donde $\mathbb{E}(\Delta V_t \Delta W_t) = \frac{1}{2}\Delta t^2$. Los pares $(\Delta W_t, \Delta V_t)$ son independientes para todos los t_i .

Los métodos de discretización pueden ser caracterizados por su orden de convergencia. Un esquema de discretización y^δ con paso de discretización de tamaño $\delta = \Delta t$ se dice *fuertemente convergente* a x en el instante T con orden $\gamma > 0$ si existe una constante positiva C que no depende de δ tal que

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

Esta expresión mide cuan próximas están las trayectorias en el instante T . Por otro lado, y^δ se dice *débilmente convergente* a x en el instante T con orden $\beta > 0$ si para cada función g suficientemente regular, existe una constante positiva C que no depende de δ tal que

$$\left| \mathbb{E}[g(x_T)] - \mathbb{E}[g(y^\delta(T))] \right| \geq C\delta^\beta.$$

La convergencia débil de orden β implica la convergencia de los momentos de todos los órdenes (tomando $g(x) = |x|^q$). En la aproximación fuerte, el orden de convergencia lo determina el término donde truncamos la serie. Pero el criterio de convergencia débil concierne aspectos probabilísticos de la trayectoria de la muestra, pero no la trayectoria en sí. De este modo, para un cierto orden de convergencia, el número requerido de términos de la expansión es menor para el caso de la convergencia débil que en el caso de la convergencia fuerte si queremos un determinado orden de convergencia. Bajo condiciones adecuadas de suavidad y crecimiento sobre las funciones drift y difusión, el método de Euler-Maruyama es fuertemente convergente de orden $\gamma = \frac{1}{2}$ y débilmente convergente de orden 1. Notar que este método es la aproximación de Taylor más simple, si se consideraran más términos en el desarrollo de Taylor, el orden fuerte de convergencia se incrementa. Así, el método de Milstein tiene orden de convergencia débil y fuerte de $\gamma = 1$ y el método KPS tiene orden de convergencia fuerte $\gamma = 1.5$. En este trabajo utilizaremos el método de Euler-Maruyama para aproximar los modelos de difusión en tiempo continuo.

3.1. Estimación paramétrica

La ecuación diferencial estocástica (3.1) puede ser representada de forma paramétrica como

$$dr_t = m(r_t, \theta) dt + \sigma(r_t, \theta) dW_t,$$

donde θ es un parámetro d -dimensional desconocido que pertenece a un espacio paramétrico $\Theta \subset \mathbb{R}^d$, con d un entero positivo. La versión discreta, con el método de Euler-Maruyama, es

$$r_{t_{i+1}} - r_{t_i} = m(r_{t_i}, \theta)\Delta + \sigma(r_{t_i}, \theta)\Delta^{1/2}\varepsilon_{t_i}, \quad i = 1, \dots, n-1 \quad (3.2)$$

donde ε_{t_i} son iid $N(0, 1)$. A continuación se introducen los supuestos básicos relativos a los coeficientes del proceso de difusión.

Supuesto 1. (Lipschitz Global)

Existe una constante K independiente de θ tal que,

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

para todo r y s .

Supuesto 2. (Crecimiento lineal)

Existe una constante K independiente de θ tal que,

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

Supuesto 3. (Positividad del coeficiente de difusión)

Existe una constante K independiente de θ tal que,

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

para todo r .

Supuesto 4. (Momentos acotados)

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

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

Supuesto 5. (Suavidad de los coeficientes)

Las funciones m y σ y sus derivadas en θ (hasta orden 3) son suaves en r y de crecimiento de orden polinomial en r uniformemente en θ .

Supuesto 6. (Crecimiento polinómico)

Existen $K > 0$ y $b > 0$, independiente de θ , tal que,

$$|m(r, \theta)| \leq K(1 + |r|^b),$$

con $\theta \in \Theta$.

Este conjunto de supuestos se completa con condiciones que aseguren la adecuada tasa de convergencia y la existencia de la información de fisher asociada al método de estimación.

3.1.1. Máxima verosimilitud exacta

Por ser r_t un proceso de Markov, podemos obtener la función de verosimilitud $\mathcal{L}_n(\theta)$ del proceso discreto utilizando la regla de Bayes,

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, r_{t_i} | r_{t_{i-1}}) p_\theta(r_{t_0}),$$

donde $p_\theta(\Delta, r_{t_i} | r_{t_{i-1}})$ denota la función de densidad de transición, o densidad condicional, asociada al modelo de difusión paramétrico, con parámetro desconocido θ . El logaritmo de la función de verosimilitud es

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta) = \sum_{i=1}^n \ell_i(\theta) + \log(p_\theta(r_{t_0})) = \sum_{i=1}^n \log p_\theta(\Delta, r_{t_i} | r_{t_{i-1}}) + \log(p_\theta(r_{t_0})).$$

Si el número de observaciones es suficientemente grande, podemos asumir que el peso relativo de $p_\theta(r_{t_0})$ en la función de verosimilitud decrece, por lo que $p_\theta(r_{t_0}) = 1$.

Si la forma paramétrica del modelo que genera las observaciones $\{r_{t_i}\}_{i=0}^n$ es conocida, podemos utilizar el método de máxima verosimilitud, de modo que el estimador de máxima verosimilitud (MLE) del verdadero parámetro es

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

Notar que aunque los modelos son descritos en tiempo continuo, la disponibilidad de los datos es en tiempo discreto, por lo que ignorar esta diferencia puede resultar en estimaciones inconsistentes. Pero no siempre es posible determinar explícitamente la verosimilitud para observaciones discretas generadas por (3.2), ya que existen modelos en los que no disponemos de fórmulas cerradas para la densidad de transición o de probabilidad condicional. Por ello, este método de estimación sólo será aplicable en aquellos modelos de difusión en los que se conoce la distribución asociada al modelo. Los modelos de Black and Scholes (1973), Vasicek (1977), Cox, Ingersoll y Ross (1985) y Cox (1975) se basan en expresiones cerradas conocidas.

3.1.2. Máxima verosimilitud aproximada

Para estimar el modelo (3.1), podemos utilizar un esquema de aproximación, como el método de Euler-Maruyama. Con este método no aproximamos directamente la densidad de transición, sino la trayectoria del proceso de modo que podamos utilizar la verosimilitud de la versión discretizada del modelo, dada por (3.2). Así, para estimar el modelo (3.1), procedemos como si las observaciones proviniesen de una distribución Gaussiana, de media el drift y desviación estándar la función de difusión, por lo que la densidad de transición es

$$p_{r,\theta}(\Delta, x | r) = \frac{1}{\sqrt{2\pi\Delta\sigma^2(r,\theta)}} \exp \left\{ -\frac{1}{2} \frac{(x - r - m(r,\theta)\Delta)^2}{\Delta\sigma^2(r,\theta)} \right\}.$$

De este modo, podemos obtener el estimador de máxima verosimilitud. Notar que esta aproximación es eficiente si el paso de discretización Δ es lo suficientemente pequeño, en caso contrario, pueden producirse sesgos significativos en las estimaciones.

En caso de que se cumplan los supuestos 3, 4 y 6 y asumiendo que $\sigma(r,\theta) = \sigma > 0$ es constante y que el resto de parámetros a estimar están en la función drift $m(r,\theta)$, el logaritmo de la verosimilitud del proceso discretizado es

$$\ell_n(\theta) = -\frac{1}{2} \left\{ \sum_{i=1}^n \frac{r_{t_i} - r_{t_{i-1}} - m(r_{t_{i-1}},\theta)\Delta^2}{\sigma^2\Delta} + n \log(2\pi\sigma^2\Delta) \right\}.$$

A la ecuación anterior se le denomina *aproximación Gaussiana local*. Dado que σ^2 es constante, maximizar la log-verosimilitud equivale a maximizar la función

$$\sum_{i=1}^n (r_{t_i} - r_{t_{i-1}})m(r_{t_{i-1}},\theta) - \frac{\Delta}{2} \sum_{i=1}^n m^2(r_{t_{i-1}},\theta).$$

Bajo el supuesto adicional de que $n\Delta_n^3 \rightarrow 0$, el estimador de máxima verosimilitud de la pseudo-verosimilitud es consistente y asintóticamente normal. En Florens-Zmirou (1989) y Yoshida (1992) se demuestra que un estimador consistente de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{n\Delta} \sum_{i=1}^n (r_{t_i} - r_{t_{i-1}})^2.$$

3.1.3. Método de máxima verosimilitud de Aït-Sahalia

La regla de Bayes combinada con la naturaleza Markoviana de (3.1), heredada por los datos discretos, implica que la log-verosimilitud tiene la forma

$$\ell_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^n \ln \{p_{r,\theta}(\Delta, r_{i\Delta} | r_{(i-1)\Delta})\}, \quad (3.3)$$

asumiendo que el proceso es observado en el intervalo $\{t = i\Delta | i = 0, \dots, n\}$, con Δ fijo. Aït-Sahalia (2002) propone un método de máxima verosimilitud para procesos de difusión con muestras discretas, basado en una aproximación de la función de verosimilitud utilizando polinomios de Hermite. El autor construye una sucesión de aproximaciones $p_{r,\theta}^{(j)}$ de la densidad de transición, de modo que (3.3) es una sucesión de aproximaciones $\ell_n^{(j)}$. Para ello, necesitaremos estandarizar el coeficiente de difusión de r , que llevaremos a cabo utilizando la transformación de Lamperti.

Transformación de Lamperti

La ecuación diferencial estocástica (3.1), puede transformarse en otro proceso de difusión, U_t , con coeficiente de difusión unitario aplicando la transformación de Lamperti,

$$U_t = \psi(r_t; \theta) = \int^{r_t} \frac{1}{\sigma(s, \theta)} ds. \quad (3.4)$$

Utilizando la fórmula de Itô en el nuevo proceso U_t , tenemos la difusión unitaria

$$dU_t = m_U(U_t; \theta) dt + dW_t,$$

con función drift m_U dada por

$$m_U(u; \theta) = \frac{m(\psi^{-1}(u; \theta); \theta)}{\sigma(\psi^{-1}(u; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial r} (\psi^{-1}(u; \theta); \theta).$$

Notar que hemos utilizado la fórmula de Itô con

$$f(t, x) = \int^x \frac{1}{\sigma(s)} ds, \quad f_t(t, x) = 0, \quad f_x(t, x) = \frac{1}{\sigma(x)}, \quad f_{xx}(t, x) = -\frac{\sigma_x(x)}{\sigma^2(x)},$$

donde $f_t(t, r) = \partial f / \partial t$, $f_r(t, r) = \partial f / \partial r$ y $f_{rr}(t, r) = \partial^2 f / \partial r^2$. Por tanto,

$$\begin{aligned} df(t, r) &= 0 \cdot dt + f_r(t, r_t) dr_t + \frac{1}{2} f_{rr}(t, r_t) (dr_t)^2 \\ &= \frac{m(t, r_t) dt + \sigma(r_t) dW_t}{\sigma(r_t)} - \frac{1}{2} \frac{\sigma_r(t, r_t)}{\sigma^2(r_t)} (m(t, r_t) dt + \sigma(r_t) dW_t)^2 \\ &= \frac{m(t, r_t)}{\sigma(r_t)} dt + dW_t - \frac{1}{2} \frac{\sigma_r(t, r_t)}{\sigma^2(r_t)} \left[m(t, r_t) (dt)^2 \right. \\ &\quad \left. + 2m(t, r_t) \sigma(r_t) dt \cdot dW_t + \sigma^2(r_t) (dW_t)^2 \right], \end{aligned}$$

donde $(dt)^2$ y $dt \cdot dW_t$ son de orden $\mathcal{O}(dt)$, por lo que en la fórmula anterior, los términos cuya parte diferencial es $(dt)^2$ y $dt \cdot dW_t$ podemos ignorarlos. Además, los términos de

orden $(dW_t)^2$ se comportan como dt por las propiedades del movimiento Browniano, por lo que podemos sustituir $(dW_t)^2$ por dt . Así, podemos escribir,

$$dU_t = \left(\frac{m(\psi^{-1}(u; \theta); \theta)}{\sigma(\psi^{-1}(u; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial r}(\psi^{-1}(u; \theta); \theta) \right) dt + dW_t.$$

Las transformaciones que, en las ecuaciones diferenciales estocásticas, dejan el término de difusión constante son importantes para simulación y estimación. Cabe notar que la aplicación práctica de la transformación de Lamperti está limitada por la existencia de una solución explícita de la inversa de la transformación,

$$r_t = \psi^{-1}(U_t; \theta).$$

Pero en el contexto financiero es habitual utilizar esta transformación porque cuando $\sigma(r_t; \theta) = \sigma r_t^\gamma$, siempre existe solución explícita de ψ^{-1} para cualquier constante γ .

Expansión de la verosimilitud mediante polinomios de Hermite

Los métodos para aproximar la densidad de transición de los procesos de difusión permiten obtener expresiones en forma cerrada para la densidad y, por ello, estas expansiones pueden utilizarse en la estimación paramétrica. Aït-Sahalia (2002b) realiza una expansión de Hermite univariante que, realizando un cambio de variable, utiliza una base polinomial común para todos los modelos, que podremos utilizar para aproximar la verosimilitud.

Este método posee una clara analogía que permitirá comprender su construcción: consideremos una suma de variables aleatorias estandarizadas donde se cumple el Teorema Central del Límite (TCL). Para un tamaño muestral n elevado, por el TCL podemos utilizar como distribución límite la $N(0, 1)$, en otro caso, podemos expandir la distribución límite (con las series de Edgeworth basadas en los polinomios de Hermite, por ejemplo) considerando términos de orden superior para mejorar la precisión de la aproximación. La idea propuesta por Aït-Sahalia es utilizar este enfoque para aproximar la densidad de transición desconocida del proceso de difusión, donde $\Delta \rightarrow 0$ juega el papel de $n \rightarrow \infty$ en el TCL. Así, para un Δ pequeño, la distribución condicional está más próxima a ser normal por la contribución del incremento Browniano. Si estandarizamos los datos, podremos encontrar la distribución límite de las observaciones estandarizadas cuando Δ tiende a cero, por analogía con lo que ocurre en el TCL cuando el tamaño muestral tiende a infinito. En los procesos de difusión, estandarizar los datos implica transformar la difusión original r en otra, U , tal y como hemos visto con la transformación de Lamperti en la sección anterior. Tanto en el TCL como en este caso, la estandarización hace que la función de distribución se aproxime a una $N(0, 1)$. Esta aproximación a la normal estándar se puede refinar añadiendo términos de orden superior basados en los polinomios de Hermite, que son ortogonales con respecto a la función de densidad de probabilidad normal.

Una propiedad de los procesos de difusión es que en pequeños intervalos de tiempo, la primera diferencia dividida por $\sqrt{\Delta}$ es aproximadamente normal. Esta aproximación mejora a medida que Δ se reduce, pero la varianza puede ser dependiente del estado. Por este motivo, previamente deberemos transformar r para que la aproximación límite sea una normal estándar. Para ello, utilizamos la transformación de Lamperti (3.4), que nos permite eliminar la heterocedasticidad convirtiendo r_t en un proceso U_t con coeficiente de difusión unitario. La densidad estacionaria q_U del proceso transformado, generalmente,

no es normal, pero satisface

$$q_U(u) \propto \left[2 \int^U m_U(s) ds \right].$$

La densidad estacionaria tiene colas poco pesadas, por lo que los polinomios tienen momentos de segundo orden finitos una vez aplicada esta transformación. La primera diferencia de U_t tendrá densidad normal estándar una vez dividida por $\sqrt{\Delta}$. Notar que si no realizamos este ajuste, la primera diferencia de U_t convergerá a una distribución degenerada con probabilidad uno en el cero (función delta de Dirac). Sea p_U la función de densidad de transición del proceso U_t , calcularemos los incrementos pseudo-normalizados de U para adecuar p_U a la expansión de Hermite,

$$Z_\Delta \doteq \Delta^{-1/2}(U_\Delta - U_0),$$

condicionando en $U_0 = u_0 = \psi(r_0)$. Sea p_z la densidad condicionada de Z_Δ , donde Δ denota el intervalo de tiempo utilizado para la aproximación. Como Z_Δ es una transformación conocida de r , podemos recuperar la densidad de transición de r a partir de la densidad de Z_Δ utilizando la fórmula del jacobiano

$$\begin{aligned} p_r(\Delta, r | r_0) &= \frac{\partial}{\partial r} \mathbb{P}[r_{t+\Delta} \leq r | r_t = r_0] \\ &= \frac{p_U[\Delta, \psi(r) | \psi(r_0)]}{\sigma(r)} = \frac{p_z[\Delta, \Delta^{-1/2}(\psi(r) - \psi(r_0)) | \psi(r_0)]}{\sigma(r)\Delta^{-1/2}}. \end{aligned}$$

Necesitaremos, por tanto, aproximar la función de densidad p_z . Sean h_j los polinomios de Hermite, polinomios ortogonales respecto a la densidad normal estándar, podemos calcularlos como

$$h_j \doteq \exp\left(\frac{z^2}{2}\right) \frac{d^j}{dz^j} \left[\exp\left(-\frac{z^2}{2}\right) \right], \quad j \geq 0.$$

La expansión de Hermite es

$$p_z(\Delta, z | u_0) = \frac{\exp(-z^2/2)}{\sqrt{2\pi}} \sum_{j=0}^{\infty} \eta_j(\Delta, u_0) h_j(z), \quad (3.5)$$

cuyos coeficientes vienen dados por

$$\begin{aligned} \eta_j(\Delta, u_0) &= \frac{1}{j!} \int_{-\infty}^{+\infty} h_j(z) p_z(\Delta, z | u_0) dz \\ &= \frac{1}{j!} \int_{-\infty}^{+\infty} h_j(\Delta^{-1/2}(u - u_0)) p_U(\Delta, u | u_0) du \\ &= \frac{1}{j!} \mathbb{E}[h_j(\Delta^{-1/2}(U_\Delta - U_0)) | U_0 = u_0] \\ &= \frac{1}{j!} \mathcal{T}_\Delta h_j(\Delta^{-1/2}(U_\Delta - u_0)). \end{aligned}$$

La aproximación de Hermite a p_z utiliza un número de términos finitos en la expansión (3.5), de modo que $p_z^{(J)}(\Delta, z | u_0)$ denota la suma parcial de (3.5) hasta $j = J$. Habitualmente se toma $J = 6$, ya que este método converge bastante rápido. Como los

coeficientes η_j son momentos condicionales específicos del proceso $\{U_t\}$, pueden calcularse utilizando métodos numéricos como la integración Monte Carlo. Aït-Sahalia (2002b) propuso utilizar una expansión exponencial, ya que cuando los coeficientes de difusión son suaves, se pueden utilizar series de potencias para un subconjunto de funciones en el dominio del generador. Sea $g(u)$ un polinomio (notar que aunque depende de u_0 , hemos tomado u_0 como fijo) y \mathcal{A} el generador infinitesimal de la difusión U , definido como el operador $\mathcal{A}(\theta) : g \mapsto m_U(\cdot; \theta) \partial g / \partial u + (1/2) \partial^2 g / \partial u^2$. Dada la suavidad de las funciones drift y difusión, los polinomios y sus iteraciones se obtienen aplicando repetidamente el generador \mathcal{A} , bajo condiciones de regularidad del comportamiento límite del proceso. Esto garantiza que las series de Taylor son una aproximación adecuada de \mathcal{T}_Δ ,

$$\mathcal{T}_{\Delta, g} \approx \sum_{k=0}^K \frac{\Delta^k \mathcal{A}^k g}{k!}.$$

Tomando g como h_j y $K = 3$, obtenemos las expresiones explícitas de la aproximación de Taylor de η_j , con $\eta_0 = 1$ y

$$\begin{aligned} \eta_1 &= -m_U \Delta^{1/2} - (2m_U m_U^{[1]} + m_U^{[2]}) \Delta^{3/2} / 4 \\ &\quad - (4m_U m_U^{[1]2} + 4m_U^2 m_U^{[2]} + 6m_U^{[1]} m_U^{[2]} + 4m_U m_U^{[3]} + m_U^{[4]}) \Delta^{5/2} / 24, \\ \eta_2 &= (m_U^2 + m_U^{[1]}) \Delta / 2 + (6m_U^2 m_U^{[1]} + 4m_U^{[1]2} + 7m_U m_U^{[2]} + 2m_U^{[3]}) \Delta^2 / 12 \\ &\quad + (28m_U^2 m_U^{[1]2} + 28m_U^2 m_U^{[3]} + 16m_U^{[1]3} + 16m_U^3 m_U^{[2]} + 88m_U m_U^{[1]} m_U^{[2]} \\ &\quad + 21m_U^{[2]2} + 32m_U^{[1]} m_U^{[3]} + 16m_U m_U^{[4]} + 3m_U^{[5]}) \Delta^3 / 96, \\ \eta_3 &= - (m_U^3 + 3m_U m_U^{[1]} + m_U^{[2]}) \Delta^{3/2} / 6 - (12m_U^3 m_U^{[1]} + 28m_U m_U^{[1]2} \\ &\quad + 22m_U^2 m_U^{[2]} + 24m_U^{[1]} m_U^{[2]} + 14m_U m_U^{[3]} + 3m_U^{[4]}) \Delta^{5/2} / 48, \\ \eta_4 &= (m_U^4 + 6m_U^2 m_U^{[1]} + 3m_U^{[1]2} + 4m_U m_U^{[2]} + m_U^{[3]}) \Delta^2 / 24 \\ &\quad + (20m_U^4 m_U^{[1]} + 50m_U^3 m_U^{[2]} + 100m_U^2 m_U^{[1]2} + 50m_U^2 m_U^{[3]} + 23m_U m_U^{[4]} \\ &\quad + 180m_U m_U^{[1]} m_U^{[2]} + 40m_U^{[1]3} + 34m_U^{[2]2} + 52m_U^{[1]} m_U^{[3]} + 4m_U^{[5]}) \Delta^3 / 240, \\ \eta_5 &= - (m_U^5 + 10m_U^3 m_U^{[1]} + 15m_U m_U^{[1]2} + 10m_U^2 m_U^{[2]} \\ &\quad + 10m_U^{[1]} m_U^{[2]} + 5m_U m_U^{[3]} + m_U^{[4]}) \Delta^{5/2} / 120, \\ \eta_6 &= (m_U^6 + 15m_U^4 m_U^{[1]} + 15m_U^{[1]3} + 20m_U^3 m_U^{[2]} + 15m_U^{[1]} m_U^{[3]} + 45m_U^2 m_U^{[1]2} \\ &\quad + 10m_U^{[2]2} + 15m_U^2 m_U^{[3]} + 60m_U m_U^{[1]} m_U^{[2]} + 6m_U m_U^{[4]} + m_U^{[5]}) \Delta^3 / 720, \end{aligned}$$

donde se ha empleado la notación $m_U^{[k]m} = (\partial^k m_U(u_0; \theta) / \partial u_0^k)^m$.

Podemos derivar expresiones explícitas de la densidad de transición del proceso U basadas en expansiones de Hermite. La expansión de p_U de orden K viene dada por

$$\tilde{p}_U^{(K)}(\Delta, u | u_0) = \Delta^{-1/2} \phi\left(\frac{u - u_0}{\Delta^{1/2}}\right) \exp\left\{\int_{u_0}^u m_U(w) dw\right\} \sum_{k=0}^K c_k(u | u_0) \frac{\Delta^k}{k!},$$

donde $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$ es la densidad normal $N(0, 1)$, $c_0(u | u_0) = 1$ y, para todo $j \geq 1$,

$$c_j(u | u_0) = j(u - u_0)^{-j} \left(\int_{u_0}^u (w - u_0)^{j-1} \{ \lambda_u(w) c_{j-1}(w | u_0) + [\partial^2 c_{j-1}(w | u_0) / \partial w^2] / 2 \} dw \right),$$

con $\lambda_u(u; \theta) \equiv -(m_{U^2}(u) + \partial m_U(u) / \partial u) / 2$. Así, tenemos que

$$\tilde{p}_r^{(K)}(\Delta, r | r_0) \equiv \sigma(r)^{-1} \tilde{p}_U^{(K)}(\Delta, \psi(r) | \psi(r_0)),$$

la sucesión de funciones explícitas $\tilde{p}_U^{(K)}$ aproxima a p_U . En Aït-Sahalia (1999) se demuestra que

$$\tilde{p}_r^{(K)}(\Delta, r | r_0) \rightarrow p_r(\Delta, r | r_0).$$

Para calcular los estimadores de máxima verosimilitud, maximizamos la aproximación del logaritmo de la verosimilitud

$$\ell_n^{(K)} \equiv \frac{1}{n} \sum_{i=1}^n \ln \left\{ \tilde{p}_{r, \theta}^{(K)}(\Delta, r_{i\Delta} | r_{(i-1)\Delta}) \right\}.$$

De esta manera obtenemos un estimador $\hat{\theta}_n^{(K)}$ que es próximo al exacto $\hat{\theta}_n$ (ver Aït-Sahalia 1999). Cuando el proceso admite una distribución estacionaria, la varianza asintótica del estimador de máxima verosimilitud viene dada por la inversa de la matriz de información de Fisher. Para procesos estacionarios, el estimador satisface

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1}),$$

donde la matriz de información de Fisher es

$$i(\theta_0) \equiv \mathbb{E} \left[\dot{\mathcal{L}}(\theta) \dot{\mathcal{L}}(\theta)' \right] = -\mathbb{E} \left[\ddot{\mathcal{L}}(\theta) \right],$$

con $\mathcal{L}(\theta) = \ln [p_{r, \theta}(\Delta, r_\Delta | r_0)]$, $\dot{\mathcal{L}}(\theta) = \partial \mathcal{L}(\theta) / \partial \theta$ y $\ddot{\mathcal{L}}(\theta) = \partial^2 \mathcal{L}(\theta) / \partial \theta \partial \theta'$.

3.1.4. Método Generalizado de los Momentos

Chan et al. (1992) introducen el modelo CKLS y para la estimación de los parámetros del modelo consideran el método generalizado de los momentos (GMM). Aproximando el proceso en tiempo continuo con una especificación en tiempo discreto tenemos

$$r_{t+1} - r_t = \alpha + \beta r_t + \varepsilon_{t+1}, \tag{3.6}$$

$$\mathbb{E}(\varepsilon_{t+1}) = 0, \quad \mathbb{E}(\varepsilon_{t+1}^2) = \sigma^2 r_t^{2\gamma}. \tag{3.7}$$

El método generalizado de los momentos no requiere que la distribución de los cambios en el tipo de interés sea normal, la justificación asintótica para el procedimiento GMM sólo requiere que la distribución de los cambios en el tipo de interés sea estacionaria y ergódica y que las esperanzas relevantes existan.

Sea $\boldsymbol{\theta}$ el vector de parámetros con los elementos α, β, σ^2 y γ . Dado $\varepsilon_{t+1} = r_{t+1} - r_t - \alpha - \beta r_t$, denotamos por $f_t(\boldsymbol{\theta})$ el vector

$$f_t(\boldsymbol{\theta}) = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} r_t \\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \\ (\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}) r_t \end{bmatrix}.$$

Bajo la hipótesis nula de que las restricciones que implican (3.7) son ciertas, tenemos las *condiciones de ortogonalidad* $\mathbb{E}[f_t(\boldsymbol{\theta})] = 0$. El procedimiento GMM consiste en reemplazar $\mathbb{E}[f_t(\boldsymbol{\theta})]$ con su contrapartida muestral, $g_n(\boldsymbol{\theta})$, utilizando las n observaciones, donde

$$g_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n f_t(\boldsymbol{\theta}),$$

Por la Ley de los Grandes Números, tenemos que

$$g_n(\boldsymbol{\theta}) \xrightarrow{c.s.} \mathbb{E}[f_t(\boldsymbol{\theta})]$$

uniformemente en $\boldsymbol{\theta}$. El método generalizado de los momentos consiste en reemplazar las condiciones de ortogonalidad por sus correspondientes momentos muestrales, escogiendo las estimaciones de los parámetros que minimizan la forma cuadrática

$$J_n(\boldsymbol{\theta}) = g_n'(\boldsymbol{\theta}) W_n(\boldsymbol{\theta}) g_n(\boldsymbol{\theta}),$$

donde $W_n(\boldsymbol{\theta})$ es una matriz de pesos simétrica y definida positiva. Para una matriz de pesos $W_n(\boldsymbol{\theta})$, el estimador GMM es

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} J_n(\boldsymbol{\theta}).$$

Hansen (1982) demostró que escogiendo $W_n(\boldsymbol{\theta}) = S^{-1}(\boldsymbol{\theta})$, donde

$$S(\boldsymbol{\theta}) = \mathbb{E}[f_t(\boldsymbol{\theta}) f_t'(\boldsymbol{\theta})],$$

resulta en el estimador GMM de $\boldsymbol{\theta}$ con la matriz de covarianzas asintótica más pequeña.

Notar que los momentos condicionales del modelo en tiempo continuo y los momentos condicionales de la discretización no coinciden. Por ello, utilizar los momentos condicionales puede conllevar estimaciones inconsistentes. Sin embargo, para $\Delta \rightarrow 0$, los momentos condicionales convergen, al igual que los estimadores.

3.1.5. Filtro de Kalman

Los modelos de espacio de estados, introducidos en Kalman (1960), proporcionan una metodología unificada para el tratamiento de una amplia variedad de problemas en el análisis de series temporales. Este modelo utiliza un vector de autorregresión de orden uno como ecuación de estados, \boldsymbol{x}_t , que determina la regla para la generación del vector de estados \boldsymbol{x}_t respecto del estado anterior \boldsymbol{x}_{t-1} , para $t = 1, \dots, n$. El modelo de espacio de estados asume que no observamos el vector de estados \boldsymbol{x}_t directamente, sino una transformación lineal del mismo con un error asociado, \boldsymbol{y}_t .

Dadas las observaciones $Y_n = \{\mathbf{y}_1 \dots \mathbf{y}_n\}$, la ecuación de estados y observaciones son:

$$\begin{aligned} \text{Ecuación de estados} \quad \mathbf{x}_t &= \Phi \mathbf{x}_{t-1} + \Upsilon \mathbf{u}_t + \mathbf{w}_t & \mathbf{w}_t &\sim \text{iid } N(0, Q), \\ \text{Ecuación de observaciones} \quad \mathbf{y}_t &= A_t \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{v}_t & \mathbf{v}_t &\sim \text{iid } N(0, R), \end{aligned}$$

donde A_t es la matriz de medidas u observaciones y \mathbf{u}_t es un vector de inputs (variables exógenas o inputs constantes que pueden incorporarse en la ecuación de observaciones o de estados). Por simplicidad, asumimos que $\{\mathbf{w}_t\}$ y $\{\mathbf{v}_t\}$ están incorrelados.

Si definimos $\mathbf{x}_t^n = \mathbb{E}(\mathbf{x}_t | Y_n)$ y $P_t^n = \mathbb{E}\{(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)'\}$, las ecuaciones del Filtro de Kalman, con estado inicial $\mathbf{x}_0 \sim N(\mathbf{x}_0^0, \mathbf{P}_0^0)$, son

$$\begin{array}{ccc} & \text{Estado inicial} & \\ & \mathbf{x}_0^0, P_0^0 & \\ & \searrow & \\ \text{Ecuaciones de actualización} & & \text{Ecuaciones de predicción} \\ \begin{aligned} K_t &= P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} \\ \mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + K_t [\mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t] \\ P_t^t &= [I - K_t A_t] P_t^{t-1} \end{aligned} & & \begin{aligned} \mathbf{x}_t^{t-1} &= \Phi \mathbf{x}_{t-1}^{t-1} + \Upsilon \mathbf{u}_t \\ P_t^{t-1} &= \Phi P_{t-1}^{t-1} \Phi' + Q \end{aligned} \\ & \swarrow & \end{array}$$

El Filtro de Kalman produce la proyección lineal de \mathbf{x}_t en Y_{t-1} , como denota $\mathbf{x}_t^{t-1} = \mathbb{E}(\mathbf{x}_t | Y_{t-1})$, junto con el error cuadrático medio, dado por $P_t^{t-1} = \mathbb{E}\{(\mathbf{x}_t - \mathbf{x}_t^{t-1})(\mathbf{x}_t - \mathbf{x}_t^{t-1})'\}$. Los estimadores son óptimos en el sentido de que minimizan el error cuadrático medio dentro de la clase de todos los estimadores lineales del vector de estados.

Para derivar las ecuaciones de predicción tenemos:

$$\mathbf{x}_t^{t-1} = \mathbb{E}(\mathbf{x}_t | Y_{t-1}) = \mathbb{E}(\Phi \mathbf{x}_{t-1} + \Upsilon \mathbf{u}_t + \mathbf{w}_t | Y_{t-1}) = \Phi \mathbf{x}_{t-1}^{t-1} + \Upsilon \mathbf{u}_t,$$

y, por lo tanto,

$$\begin{aligned} P_t^{t-1} &= \mathbb{E}\{(\mathbf{x}_t - \mathbf{x}_t^{t-1})(\mathbf{x}_t - \mathbf{x}_t^{t-1})'\} \\ &= \mathbb{E}\{[\Phi(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^{t-1}) + \mathbf{w}_t][\Phi(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^{t-1}) + \mathbf{w}_t]'\} \\ &= \Phi P_{t-1}^{t-1} \Phi' + Q. \end{aligned}$$

Para derivar las ecuaciones de actualización, escribimos las innovaciones como

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbb{E}\{\mathbf{y}_t | Y_{t-1}\} = \mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t,$$

y las correspondientes matrices de varianzas-covarianzas

$$\Sigma_t = \text{Var}(\boldsymbol{\varepsilon}_t) = \text{Var}[A_t(\mathbf{x}_t - \mathbf{x}_t^{t-1}) + \mathbf{v}_t] = A_t P_t^{t-1} A_t' + R \quad \text{para } t = 1, \dots, n.$$

Notar que $\mathbb{E}(\varepsilon_t \mathbf{y}'_s) = 0$ para $s < t$, y como la secuencia de innovaciones es un proceso Gaussiano, las innovaciones son independientes de las observaciones pasadas. La covarianza condicional entre \mathbf{x}_t y ε_t dado Y_{t-1} es $\text{Cov} = \{\mathbf{x}_t, \varepsilon_t \mid Y_{t-1}\} = P_t^{t-1} A'_t$. Entonces, la distribución condicional conjunta de \mathbf{x}_t y ε_t dado Y_{t-1} es normal,

$$\begin{pmatrix} \mathbf{x}_t \\ \varepsilon_t \end{pmatrix} \Big| Y_{t-1} \sim N \left(\begin{bmatrix} \mathbf{x}_t^{t-1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} P_t^{t-1} & P_t^{t-1} A'_t \\ A_t P_t^{t-1} & \Sigma_t \end{bmatrix} \right).$$

Podemos escribir ²

$$\mathbf{x}_t^t = \mathbb{E} \{ \mathbf{x}_t \mid \mathbf{y}_1, \dots, \mathbf{y}_{t-1}, \mathbf{y}_t \} = \mathbb{E} \{ \mathbf{x}_t \mid Y_{t-1}, \varepsilon_t \} = \mathbf{x}_t^{t-1} + K_t \varepsilon_t,$$

donde

$$K_t = P_t^{t-1} A'_t \Sigma_t^{-1} = P_t^{t-1} A'_t [A_t P_t^{t-1} A'_t + R]^{-1}$$

es la Ganancia de Kalman, que minimiza la matriz de covarianza del error,

$$P_t^t = \text{Cov} \{ \mathbf{x}_t \mid Y_{t-1}, \varepsilon_t \} = P_t^{t-1} - P_t^{t-1} A'_t \Sigma_t^{-1} A_t P_t^{t-1} = [I - K_t A_t] P_t^{t-1}.$$

Estimación por máxima verosimilitud

La estimación de los parámetros que especifican el modelo de espacio de estados se realiza por máxima verosimilitud. Sea $\boldsymbol{\theta} = \{\mathbf{x}_0^0, P_0^0, \Phi, Q, R, \Upsilon, \Gamma\}$ el vector de parámetros que contiene los elementos de media y covarianza inicial \mathbf{x}_0^0 y P_0^0 , la matriz de transición Φ , las matrices de covarianzas de estado y observaciones Q y R y la matriz de coeficientes del vector de inputs Υ y Γ . Podemos utilizar máxima verosimilitud bajo el supuesto de que el estado inicial es Gaussiano y los errores $\mathbf{w}_1, \dots, \mathbf{w}_n$ y $\mathbf{v}_1, \dots, \mathbf{v}_n$ son vectores de variables normales e incorrelados. La verosimilitud se computa utilizando las innovaciones $\varepsilon_1, \dots, \varepsilon_n$, que por ser vectores aleatorios Gaussianos independientes con media cero y matriz de covarianzas $\Sigma_t = A_t P_t^{t-1} A'_t + R$, podemos escribir la verosimilitud, $\mathcal{L}_Y(\boldsymbol{\theta})$, (ignorando la constante) como

$$-\ln \mathcal{L}_Y(\boldsymbol{\theta}) = \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t(\boldsymbol{\theta})| + \frac{1}{2} \sum_{t=1}^n \varepsilon_t(\boldsymbol{\theta})' \Sigma_t(\boldsymbol{\theta})^{-1} \varepsilon_t(\boldsymbol{\theta}).$$

El procedimiento habitual es, dado \mathbf{x}_0^0 , desarrollar un conjunto de recursiones para la función del logaritmo de la verosimilitud y sus dos primeras derivadas. Posteriormente, podemos utilizar el algoritmo de Newton-Raphson sucesivamente para actualizar los valores de los parámetros hasta que se minimice el negativo del logaritmo de la verosimilitud.

Para utilizar el filtro de Kalman para obtener la estimación de los parámetros del proceso de difusión

$$d\mathbf{r}_t = m(\mathbf{r}_t, \boldsymbol{\theta}) dt + \sigma(\mathbf{r}_t, \boldsymbol{\theta}) dW_t,$$

²La distribución normal multivariante tiene la siguiente propiedad:
Sea $\mathbf{y} = (y_1, \dots, y_m)'$, $\mathbf{x} = (x_1, \dots, x_n)'$, y supongamos que el vector $(\mathbf{y}', \mathbf{x}')'$ es normal,

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim N \left(\begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \right).$$

Entonces, $\mathbf{y} \mid \mathbf{x}$ es normal con

$$\begin{aligned} \boldsymbol{\mu}_{y|x} &= \boldsymbol{\mu}_y + \Sigma_{yx} \Sigma_{xx}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x), \\ \Sigma_{y|x} &= \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}, \end{aligned}$$

donde se asume que Σ_{xx} es no singular.

consideramos su versión discreta,

$$r_{t_{i+1}} - r_{t_i} = m(r_t, \theta)\Delta + \sigma(r_t, \theta)(W_{t_{i+1}} - W_{t_i}),$$

y reexpresamos el proceso como

$$\frac{z_{t_i}}{\Delta} = m(r_t, \theta) + \sigma(r_t, \theta)\Delta^{-1/2}\varepsilon_{t_i},$$

con $z_{t_i} = r_{t_{i+1}} - r_{t_i}$ y donde ε_{t_i} son variables aleatorias iid $N(0,1)$.

Distribución asintótica de los estimadores

Bajo condiciones generales, sea $\hat{\theta}_n$ el estimador del verdadero parámetro θ_0 obtenido maximizando la verosimilitud de las innovaciones, $\mathcal{L}_Y(\theta)$. Entonces, cuando $n \rightarrow \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1}),$$

donde

$$i(\theta_0) = \lim_{n \rightarrow \infty} n^{-1} \mathbb{E} [-\partial^2 \ln \mathcal{L}_Y(\theta_0) / \partial \theta_0 \partial \theta_0']$$

es la matriz de información asintótica de Fisher.

3.2. Estimación no paramétrica

Los modelos expuestos en el Capítulo 2.1 se basan en una serie de hipótesis sobre el posible comportamiento de las funciones drift y difusión del proceso estocástico que recoge la dinámica de los tipos de interés. Pero no existe consenso ni evidencia empírica de que un determinado tipo de proceso estocástico sea más adecuado para modelizar el comportamiento de los tipos de interés. Los modelos paramétricos intentan capturar distintas propiedades particulares de la dinámica de los tipos de interés, lo que ha llevado a utilizar también técnicas de estimación no paramétricas, que permiten capturar la dinámica sin necesidad de especificar una relación paramétrica. A diferencia de los métodos paramétricos, se requieren menos hipótesis y restricciones, pero para su adecuado funcionamiento necesitaremos tamaños muestrales elevados.

Las técnicas no-paramétricas aplicadas en el contexto de los modelos de difusión están orientadas, principalmente, a la determinación de la función de densidad estacionaria y las funciones drift y difusión.

Sea r un proceso de difusión ergódico, solución de la ecuación diferencial estocástica

$$dr_t = m(r_t) dt + \sigma(r_t) dW_t,$$

donde $m(\cdot)$ y $\sigma(\cdot)$ satisfacen las condiciones de regularidad y existe una medida invariante $\pi(r)$ para el proceso r .

3.2.1. Estimación de la densidad estacionaria

Dada una muestra aleatoria simple (X_1, X_2, \dots, X_n) de una población con función de distribución F , absolutamente continua, y función de densidad f , el estimador tipo núcleo

propuesto por Parzen (1962) y Rosenblatt (1956) viene dado por

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

donde $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$, K es una *función Kernel* o *función Núcleo* y $h > 0$ es el *parámetro de suavizado*, llamado *ventana* o *ancho de banda*, que regula el tamaño del entorno que se usa para llevar a cabo la estimación. Es habitual exigir que la función núcleo K sea no negativa y su integral sea uno:

$$K(u) \geq 0, \quad \forall u, \quad \int_{-\infty}^{\infty} K(u) du = 1.$$

También es frecuente exigir que K sea una función simétrica, $K(-u) = K(u)$. Por tanto, el estimador tipo núcleo de la densidad estacionaria es

$$\hat{\pi}_h(r) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{r - r_{t_i}}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(r - r_{t_i}).$$

Es usual utilizar el núcleo Gaussiano

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}.$$

Respecto a la distribución asintótica del estimador de Parzen-Rosenblatt, las condiciones mínimas necesarias para que el sesgo y la varianza del estimador tiendan a cero cuando el tamaño muestral tiende a infinito son $h \rightarrow 0$, $nh \rightarrow \infty$. En tales circunstancias se tiene

$$\sqrt{nh} (\hat{\pi}_h(r) - \pi(r)) \xrightarrow{d} N(B, V).$$

Puede probarse que el valor asintóticamente óptimo de h , en el sentido del error cuadrático medio (MSE), es $h = c_0 n^{-1/5}$, con

$$c_0 = \left(\frac{\pi(r) \int K(t)^2 dt}{d_K^2 \pi''(r)^2} \right)^{\frac{1}{5}},$$

$d_K = \int t^2 K(t) dt$. Con este valor de h los valores de media y varianza de la distribución normal límite son

$$B = \frac{1}{2} c_0^{5/2} d_K \pi''(r),$$

$$V = \pi(r) \int K(t)^2 dt.$$

Asumiendo $\lim_{n \rightarrow \infty} nh^{4.5} = 0$ y bajo condiciones de regularidad, el estimador de la densidad estacionaria $\hat{\pi}_h$ se comporta como en el caso de variables aleatorias independientes e idénticamente distribuidas, tenemos que

$$\sqrt{nh} (\hat{\pi}_h(r) - \pi(r)) \xrightarrow{d} N(0, V).$$

Podemos encontrar resultados referidos a la consistencia asintótica en Boente y Fraiman (1988).

3.2.2. Estimación de las funciones drift y difusión

Existen dos enfoques para la estimación de los coeficientes del proceso de difusión. El primero es un enfoque semiparamétrico que considera el conocimiento de la función drift y el segundo es un enfoque no-paramétrico, que supone el desconocimiento de ambas funciones, drift y difusión.

Enfoque semiparamétrico

Como hemos visto en la Sección 2.1, de las ecuaciones de Fokker-Planck (Aït-Sahalia (1996a)) obtenemos

$$\frac{\partial^2}{\partial r^2} [\sigma^2(r)\pi(r)] = 2 \frac{\partial}{\partial r} [m(r)\pi(r)],$$

integrando y reordenando,

$$m(r) = \frac{1}{2\pi(r)} \frac{\partial}{\partial r} [\sigma^2(r)\pi(r)], \quad \sigma^2(r) = \frac{2}{\pi(r)} \int_0^r m(u)\pi(u) du,$$

bajo la condición $\pi(0) = 0$. Estas ecuaciones permiten estimar el drift $m(\cdot)$ dada una especificación de la función de difusión $\sigma^2(\cdot)$ y de la densidad marginal o estimar el término de difusión dada una especificación de la función drift y de la densidad marginal estimada. Así, tomando $\hat{\pi}_h$ como estimador de la densidad estacionaria y dado un drift $m(\cdot)$ conocido (o cuya forma paramétrica sea conocida y dispongamos de estimadores consistentes), podemos estimar $\sigma^2(\cdot)$ por

$$\hat{\sigma}_h^2(r) = \frac{2}{\hat{\pi}_h(r)} \int_0^r \hat{m}(u)\pi_h(u) du = \frac{2}{\hat{\pi}_h(r)} \int_0^r \hat{m}(u, \hat{\theta})\pi_h(u) du,$$

donde $\hat{m}(\cdot, \hat{\theta})$ es el estimador de la función drift y $\hat{\theta}$ es un estimador \sqrt{n} -consistente de θ . Las propiedades asintóticas de este estimador pueden consultarse en Prakasa Rao (1990) y Aït-Sahalia (1996).

Enfoque no-paramétrico

Si carecemos de información acerca de la función drift, la estimación de los coeficientes se lleva a cabo de forma no-paramétrica. El estimador de la función de difusión propuesto por Florens-Zmirou (1993) es

$$\hat{\sigma}_h^2(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right) (r_{t_{i+1}} - r_{t_i})^2}{\Delta \sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right)},$$

que requiere un esquema de altas frecuencias, con la condición $\Delta \rightarrow 0$ en alguna tasa adecuada para obtener la consistencia y la normalidad asintótica. En el artículo original se utilizó el como K el kernel uniforme. Posteriormente, Jiang y Knight (1998) extendieron el resultado de Florens-Zmirou al kernel Gaussiano y Bandi y Phillips (2003) extendieron los resultados a procesos estacionarios generales con un kernel K . Otros autores han utilizado técnicas diferentes para la estimación del coeficiente de difusión: wavelets (Genon-Catalot et al. 1992), estimación adaptativa (Hoffmann 1999), vecino más próximo, etc.

Podemos obtener un estimador del drift tal y como hemos visto con la función de difusión:

$$\hat{m}_h(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right) (r_{t_{i+1}} - r_{t_i})}{\Delta \sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right)}$$

En el enfoque de Stanton (1997), los dos estimadores empleados son los estimadores de tipo núcleo de Nadaraya-Watson de las esperanzas condicionales

$$m(r) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E} \{ (r_t - r) \mid r_0 = r \},$$

$$\sigma^2(r) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E} \{ (r_t - r)^2 \mid r_0 = r \}.$$

En este enfoque, $m(r)$ y $\sigma^2(r)$ equivalen a la media condicional y varianza condicional, respectivamente, del proceso cuando $r_0 = r$. Para un Δ_n fijo, estas ecuaciones se pueden reescribir como

$$m(r) = \frac{1}{\Delta} \mathbb{E} \{ (r_{t_{i+1}} - r_{t_i}) \mid r_{t_i} = r \} + \frac{\mathcal{O}(\Delta_n)}{\Delta},$$

$$\sigma^2(r) = \frac{1}{\Delta} \mathbb{E} \{ (r_{t_{i+1}} - r_{t_i})^2 \mid r_{t_i} = r \} + \frac{\mathcal{O}(\Delta_n)}{\Delta}.$$

Respecto a la elección del parámetro de suavizado h , es habitual utilizar la *regla del pulgar*, como la regla de Scott (1992), de modo que h es proporcional a $s \cdot n^{-1/5}$, donde s es la desviación estándar de la muestra,

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (r_{t_i} - \bar{r})^2}.$$

Capítulo 4

Aplicación a datos simulados y reales

Actualmente, no existe un consenso sobre qué modelo es más apropiado para representar la dinámica de los tipos de interés, ni qué método de estimación es más adecuado. En este capítulo se presenta una comparación empírica de los métodos de estimación, paramétricos y no-paramétricos, descritos en el Capítulo 3.

Para la simulación se ha optado por el proceso CIR, ya que, como la distribución condicional del proceso es conocida, podemos basar el método de simulación en su distribución y no recurrir a un método de discretización. También nos permitirá considerar entre los métodos de estimación paramétrica el de máxima verosimilitud exacta y, además, en la Sección 4.2 podremos comparar la estimación no-paramétrica de la densidad estacionaria con la verdadera función de densidad $\pi(r)$.

En el capítulo también se introducen técnicas de remuestreo para aproximar la distribución de los estimadores. Para ello, emplearemos la metodología bootstrap en el filtro de Kalman. Es importante notar que un aspecto fundamental del bootstrap es su corrección asintótica, es decir, que el orden de aproximación de la distribución límite sea igual o mejor que el de la teoría asintótica convencional. Veremos que el orden de aproximación es igual en ambos casos, por lo que el interés de la aplicación de la metodología bootstrap en el filtro de Kalman radica en la posibilidad de estudiar la distribución de los estimadores.

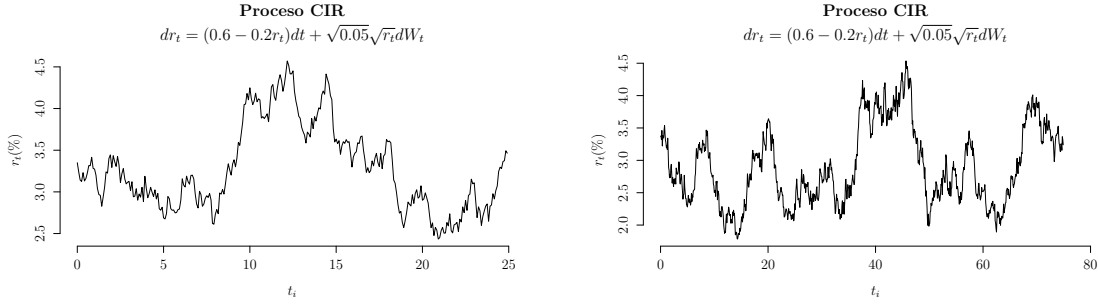
4.1. Estimación paramétrica

En esta sección llevaremos a cabo una comparación empírica de los métodos de estimación paramétricos descritos en la Sección 3.1: máxima verosimilitud exacta, máxima verosimilitud aproximada, Método Generalizado de los Momentos (GMM), máxima verosimilitud de Ait-Sahalia y filtro de Kalman.

Consideremos el proceso Cox-Ingersoll-Ross dado por

$$dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t, \quad (4.1)$$

con $r_0 = 3\%$, $\theta_1 = 0.6$, $\theta_2 = 0.2$, $\theta_3 = \sqrt{0.05} \approx 0.2236$. Se cumple que $2\theta_1 > \theta_3^2$, por lo que el proceso anterior es estacionario. Se ha simulado el proceso para $\Delta = 1/12$ y dos tamaños muestrales, $n = 300$ y $n = 900$, cuyas trayectorias pueden verse en la Figura 4.1.



(a) Trayectoria simulada del proceso con tamaño muestral $n = 300$ y $\Delta = 1/12$, donde $t_i = i\Delta/T$, con $i = 0, 1, \dots, 300$ y $T = \Delta n = 25$.

(b) Trayectoria simulada del proceso con tamaño muestral $n = 900$ y $\Delta = 1/12$, donde $t_i = i\Delta/T$, con $i = 0, 1, \dots, 900$ y $T = \Delta n = 75$.

Figura 4.1: Trayectorias simuladas del proceso CIR $dr_t = (0.6 - 0.2r_t)dt + \sqrt{0.05r_t}dW_t$ para dos tamaños muestrales, $n = 300$ y $n = 900$.

Para la estimación por el Método Generalizado de los Momentos, consideramos el proceso discretizado

$$r_{t_{i+1}} - r_{t_i} = (\theta_1 - \theta_2 r_t)\Delta + \varepsilon_{t_{i+1}} \quad (4.2)$$

y asumimos

$$\mathbb{E}[\varepsilon_{t_{i+1}} | \mathcal{F}_{t_i}] = 0, \quad \mathbb{E}[\varepsilon_{t_{i+1}}^2 | \mathcal{F}_{t_i}] = \Delta\theta_3^2 r_{t_i}, \quad \mathbb{E}[\varepsilon_{t_i}\varepsilon_{t_j} | \mathcal{F}_{t_i}] = 0,$$

por lo que $\mathbb{E}[r_{t_{i+1}} | \mathcal{F}_{t_i}] = r_{t_i} + (\theta_1 - \theta_2 r_t)\Delta$ y $r_{t_{i+1}} - \mathbb{E}[r_{t_{i+1}} | \mathcal{F}_{t_i}] = \varepsilon_{t_{i+1}}$. El vector $f_t(\boldsymbol{\theta})$ será, por tanto,

$$f_t(\boldsymbol{\theta}) = \begin{bmatrix} r_{t_{i+1}} - \mathbb{E}(r_{t_{i+1}} | \mathcal{F}_{t_i}) \\ r_{t_i} [r_{t_{i+1}} - \mathbb{E}(r_{t_{i+1}} | \mathcal{F}_{t_i})] \\ \text{Var}(r_{t_{i+1}} | \mathcal{F}_{t_i}) - [r_{t_{i+1}} - \mathbb{E}(r_{t_{i+1}} | \mathcal{F}_{t_i})]^2 \\ r_{t_i} (\text{Var}(r_{t_{i+1}} | \mathcal{F}_{t_i}) - [r_{t_{i+1}} - \mathbb{E}(r_{t_{i+1}} | \mathcal{F}_{t_i})]^2) \end{bmatrix}$$

donde

$$\mathbb{E}(r_{t_{i+1}} | r_{t_i} = x) = \frac{\theta_1}{\theta_2} + \left(x - \frac{\theta_1}{\theta_2}\right)e^{-\theta_2\Delta},$$

$$\text{Var}(r_{t_{i+1}} | r_{t_i} = x) = x \frac{\theta_3^2(e^{-\theta_2\Delta} - e^{-2\theta_2\Delta})}{\theta_2} + \frac{\theta_1\theta_3^2(1 - e^{-2\theta_2\Delta})}{2\theta_2^2}.$$

Por otro lado, para la estimación con el método de máxima verosimilitud de Ait-Sahalia necesitaremos calcular la transformación de Lamperti del proceso CIR,

$$U_t = \psi(r_t; \boldsymbol{\theta}) = \int^{r_t} \frac{1}{\sigma(s; \boldsymbol{\theta})} ds = \int^{r_t} \frac{1}{\theta_3\sqrt{s}} ds = \frac{2\sqrt{r_t}}{\theta_3},$$

y, por tanto,

$$r_t = \psi^{-1}(U_t; \boldsymbol{\theta}) = \left(\frac{\theta_3 U_t}{2}\right)^2.$$

Aplicando la fórmula de Itô al nuevo proceso U ,

$$\begin{aligned} dU_t &= m_U(U_t; \boldsymbol{\theta}) dt + dW_t \\ &= \left(\frac{m(\psi^{-1}(u; \boldsymbol{\theta}); \boldsymbol{\theta})}{\sigma(\psi^{-1}(u; \boldsymbol{\theta}); \boldsymbol{\theta})} - \frac{1}{2} \frac{\partial \sigma}{\partial r}(\psi^{-1}(u; \boldsymbol{\theta}); \boldsymbol{\theta}) \right) dt + dW_t \\ &= \frac{4\theta_1 - \theta_3^2}{2\theta_3^2 U_t} dt - \frac{\theta_2}{2} U_t dt + dW_t, \end{aligned}$$

obtenemos un proceso con coeficiente de difusión unitario. Tenemos que el drift del proceso U_t es

$$m_U(u_0; \boldsymbol{\theta}) = \frac{4\theta_1 - \theta_3^2}{2\theta_3^2 u_0} - \frac{\theta_2}{2} u_0,$$

y sus derivadas parciales vienen dadas por

$$\begin{aligned} m_U^{[1]} &= \frac{\partial m_U(u_0; \boldsymbol{\theta})}{\partial u_0} = \frac{-4\theta_1 + \theta_3^2}{2\theta_3^2 u_0^2} - \frac{\theta_2}{2}, & m_U^{[4]} &= \frac{\partial^4 m_U(u_0; \boldsymbol{\theta})}{\partial u_0^4} = \frac{48\theta_1 - 12\theta_3^2}{\theta_3^2 u_0^5}, \\ m_U^{[2]} &= \frac{\partial^2 m_U(u_0; \boldsymbol{\theta})}{\partial u_0^2} = \frac{4\theta_1 - \theta_3^2}{\theta_3^2 u_0^3}, & m_U^{[5]} &= \frac{\partial^5 m_U(u_0; \boldsymbol{\theta})}{\partial u_0^5} = \frac{-240\theta_1 + 60\theta_3^2}{\theta_3^2 u_0^6}, \\ m_U^{[3]} &= \frac{\partial^3 m_U(u_0; \boldsymbol{\theta})}{\partial u_0^3} = \frac{-12\theta_1 + 3\theta_3^2}{\theta_3^2 u_0^4}, & m_U^{[6]} &= \frac{\partial^6 m_U(u_0; \boldsymbol{\theta})}{\partial u_0^6} = \frac{1440\theta_1 - 360\theta_3^2}{\theta_3^2 u_0^7}. \end{aligned}$$

En el Cuadro 4.1 se muestran las estimaciones de los parámetros y el error estándar asociado (entre paréntesis) utilizando los distintos métodos de estimación paramétrica vistos en la Sección 3.

$$dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 \sqrt{r_t} dW_t$$

$$\theta_1 = 0.6 \quad \theta_2 = 0.2 \quad \theta_3 = \sqrt{0.05} \approx 0.2236$$

Método	$n = 300$			$n = 900$		
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
MV exacta	1.0114 (0.5186)	0.3074 (0.1601)	0.2150 (0.0089)	0.6297 (0.2170)	0.2178 (0.0763)	0.2265 (0.0054)
MV aproximada	1.0107 (0.5048)	0.3073 (0.1559)	0.2122 (0.0087)	0.6185 (0.2132)	0.2138 (0.0750)	0.2246 (0.0053)
GMM	0.9788 (2.1047)	0.3032 (0.7280)	0.1518 (0.0270)	0.7691 (0.8558)	0.2793 (0.4326)	0.1523 (0.0474)
Aït-Sahalia	1.0105 (0.5184)	0.3072 (0.1601)	0.2150 (0.0089)	0.6274 (0.2170)	0.2168 (0.0764)	0.2265 (0.0054)
Filtro de Kalman	1.0107 (0.5048)	0.3072 (0.1560)	0.2122 (0.0087)	0.6180 (0.2132)	0.2136 (0.0750)	0.2246 (0.0053)

Cuadro 4.1: *Estimación paramétrica de las componentes del proceso CIR, utilizando distintos métodos de estimación.*

Para simular el proceso, se ha generado una muestra de tamaño 45 000 y nos hemos quedado con las últimas n observaciones para eliminar el efecto del valor inicial r_0 y para asegurarnos de que las observaciones del proceso son estacionarias. En el Cuadro 4.1, podemos apreciar como al incrementar el tamaño muestral, las estimaciones de los parámetros se acercan a sus verdaderos valores. La estimación del coeficiente de difusión, θ_3 , se aproxima a su verdadero valor incluso con tamaño muestral $n = 300$. En general, las estimaciones obtenidas con los cinco métodos son muy similares, a excepción del Método Generalizado de los Momentos (notar que el error estándar que obtenemos es bastante elevado). Es importante resaltar que, tal y como vimos en la Sección 3.1.4, los momentos condicionales del modelo en tiempo continuo (4.1) y los momentos condicionales del proceso discretizado (4.2) no coinciden, por lo que las estimaciones obtenidas pueden ser inconsistentes. Sea $\boldsymbol{\theta}$ el verdadero parámetro de (4.1) y $\tilde{\boldsymbol{\theta}}$ el verdadero parámetro de (4.2). Si comprobamos la condición de ortogonalidad, $\mathbb{E}[f_t(\boldsymbol{\theta})] = 0$, para los dos primeros elementos de f_t , $\mathbb{E}(\varepsilon_{t_{i+1}}) = 0$ y $\mathbb{E}(r_t \varepsilon_{t_{i+1}}) = 0$, obtenemos

$$\tilde{\theta}_1 = \frac{\theta_1}{\theta_2 \Delta} (e^{\theta_2 \Delta} - 1), \quad \tilde{\theta}_2 = \frac{e^{\theta_2 \Delta} - 1}{\Delta},$$

es decir, este método proporciona estimadores consistentes de $\tilde{\theta}_1$ y $\tilde{\theta}_2$, pero no de θ_1 y θ_2 . Cuando $\Delta \rightarrow 0$, los parámetros $\tilde{\theta}_1$ y $\tilde{\theta}_2$ son prácticamente iguales a sus homólogos en tiempo continuo, θ_1 y θ_2 . En este caso, tenemos que $\Delta = 1/12$, por lo que parece que no es lo suficientemente pequeño para obtener estimaciones razonables con el Método Generalizado de los Momentos.

4.1.1. Estudio de simulación

En esta sección llevaremos a cabo un estudio de simulación para comparar las estimaciones de los parámetros con los cinco métodos de estimación paramétrica introducidos a lo largo de este trabajo. Para ello, realizaremos 500 simulaciones del proceso CIR expuesto en el apartado anterior,

$$dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 \sqrt{r_t} dW_t, \\ \theta_1 = 0.6 \quad \theta_2 = 0.2 \quad \theta_3 = \sqrt{0.05} \approx 0.2236,$$

con $r_0 = 3\%$. Se han generado 500 muestras de tamaño 45 000 y nos hemos quedado con las últimas $n = 900$ observaciones, para las que hemos estimado los parámetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ con los cinco métodos de estimación.

En el Cuadro 4.2 se muestran distintos estadísticos de las estimaciones. En media, el método generalizado de los momentos (GMM) es el que peor funciona, ya que, como vimos en la sección anterior, las estimaciones obtenidas con este método pueden ser inconsistentes. Respecto a los otros métodos de estimación, presentan un comportamiento bastante similar, siendo el método de máxima verosimilitud exacta el que, en media, más se acerca al verdadero valor de los parámetros.

En la Figura 4.2 se han representado los diagramas de caja de las estimaciones y en la Figura 4.3 se muestran los histogramas de las estimaciones de los tres parámetros para los cinco métodos de estimación, así como la media (línea punteada) y el valor del verdadero parámetro. Notar que, en media, la estimación del coeficiente de difusión, θ_3 , por máxima verosimilitud exacta y por el método de Ait-Sahalia son muy parecidas y prácticamente iguales al parámetro teórico. Para los parámetros θ_1 y θ_2 , los histogramas de los cinco

métodos son muy similares, excepto para el método generalizado de los momentos que presenta un sesgo bastante mayor.

Método	Parámetro	Cuartiles					Máx.	Media
		Mín.	25 %	50 %	75 %			
MV exacta	$\hat{\theta}_1$	0.2186	0.5804	0.7533	0.9330	1.5992	0.7739	
	$\hat{\theta}_2$	0.0663	0.1921	0.2522	0.3123	0.5406	0.2577	
	$\hat{\theta}_3$	0.2104	0.2202	0.2238	0.2274	0.2399	0.2237	
MV aproximada	$\hat{\theta}_1$	0.2136	0.5736	0.7420	0.9202	1.5368	0.7645	
	$\hat{\theta}_2$	0.0649	0.1900	0.2490	0.3052	0.5201	0.2546	
	$\hat{\theta}_3$	0.2086	0.2177	0.2214	0.2248	0.2368	0.2214	
GMM	$\hat{\theta}_1$	0.0083	0.7278	0.9596	1.2934	3.7294	1.0696	
	$\hat{\theta}_2$	0.0328	0.2415	0.3297	0.4232	1.1050	0.3543	
	$\hat{\theta}_3$	0.1396	0.1545	0.1585	0.1628	0.2235	0.1589	
Ait-Sahalia	$\hat{\theta}_1$	0.2186	0.5804	0.7533	0.9330	1.5780	0.7736	
	$\hat{\theta}_2$	0.0663	0.1921	0.2522	0.3123	0.5314	0.2576	
	$\hat{\theta}_3$	0.2104	0.2202	0.2238	0.2274	0.2399	0.2237	
Filtro de Kalman	$\hat{\theta}_1$	0.2127	0.5737	0.7421	0.9194	1.5361	0.7644	
	$\hat{\theta}_2$	0.0646	0.1901	0.2490	0.3053	0.5203	0.2545	
	$\hat{\theta}_3$	0.2086	0.2177	0.2214	0.2248	0.2368	0.2214	

Cuadro 4.2: Estadísticos de las estimaciones en 500 muestras de tamaño $n = 900$ del proceso CIR $dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 \sqrt{r_t} dW_t$ con parámetros $\theta_1 = 0.6, \theta_2 = 0.2, \theta_3 = \sqrt{0.05} \approx 0.2236$, mediante cinco métodos de estimación. Se muestran los valores mínimos y máximos, los cuartiles y la media.

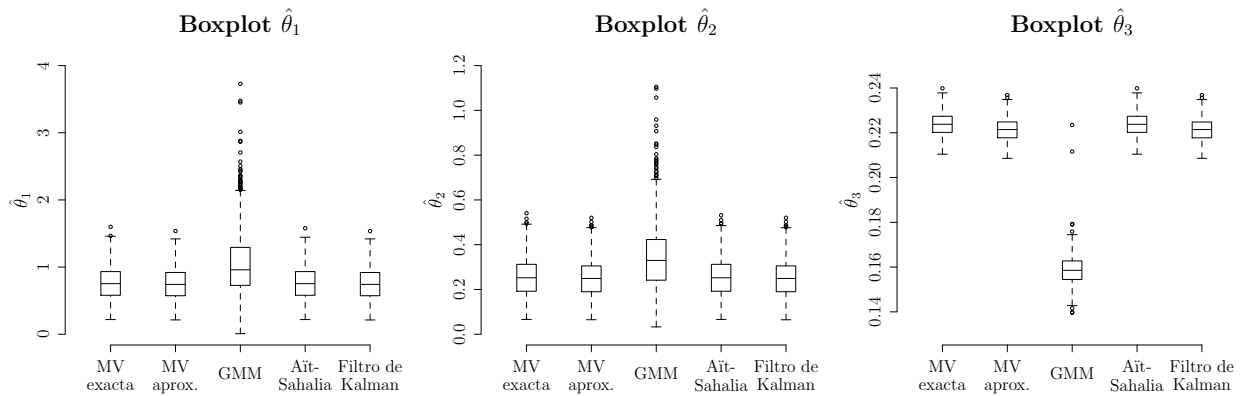


Figura 4.2: Diagramas de caja de las estimaciones de $\theta = (\theta_1, \theta_2, \theta_3)' = (0.6, 0.2, \sqrt{0.05})'$ de 500 réplicas del proceso CIR $dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 \sqrt{r_t} dW_t$.

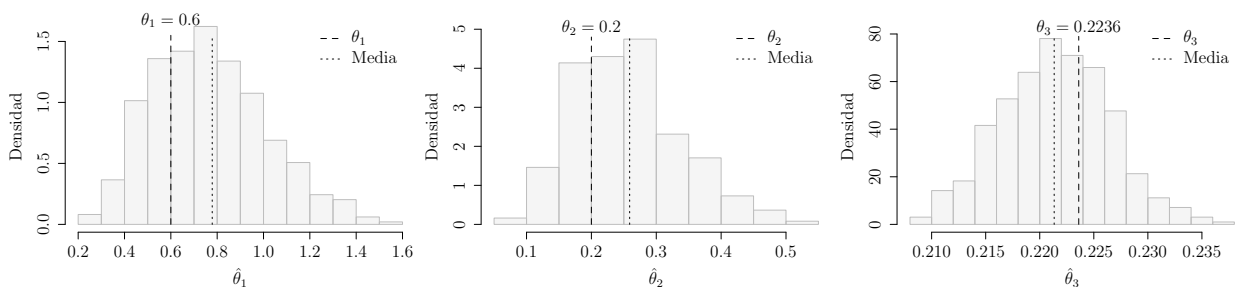
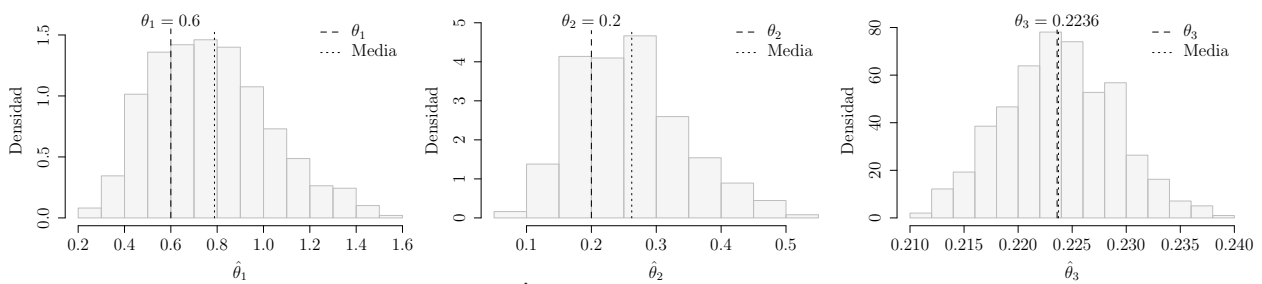
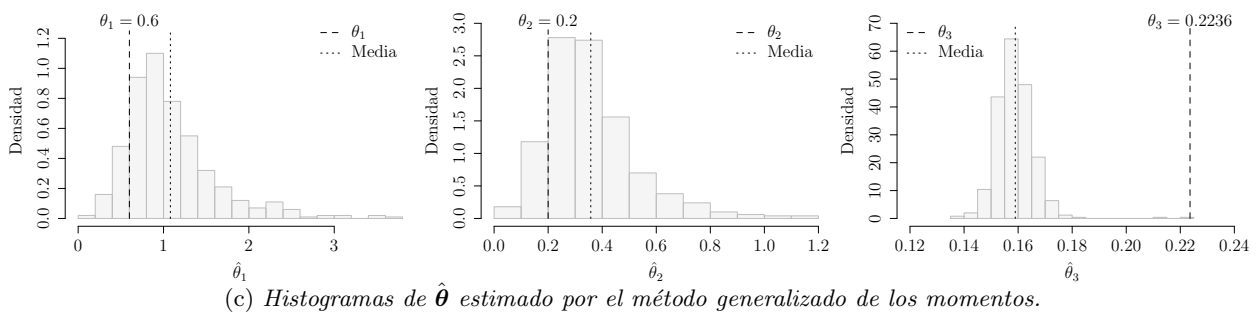
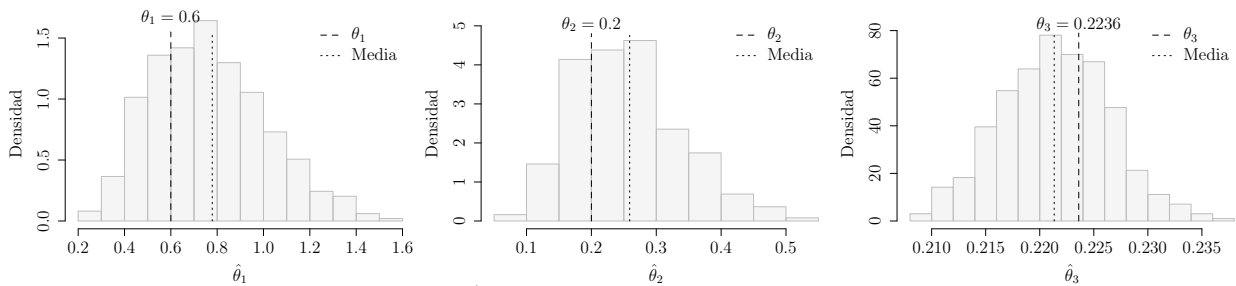
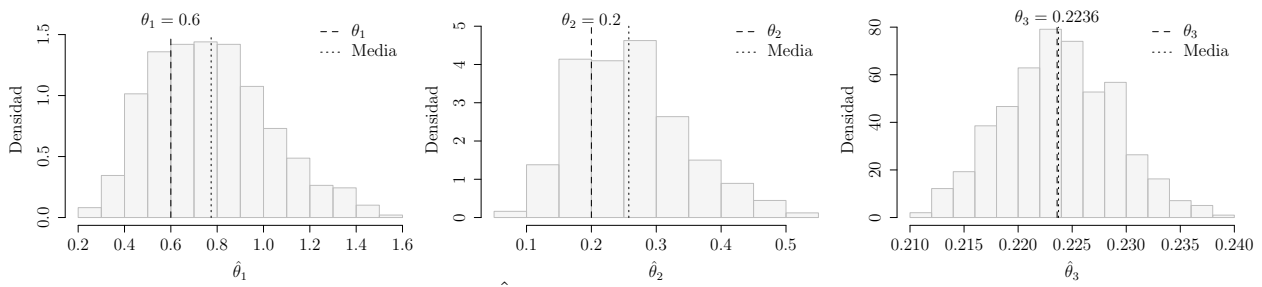


Figura 4.3: Histogramas de $\hat{\theta}$ para los cinco métodos de estimación paramétrica, basados en 500 simulaciones.

4.2. Estimación no-paramétrica

Dado el proceso (4.1),

$$dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t,$$

en la Figura 4.4 se muestra el estimador no paramétrico (línea punteada) de las funciones drift y difusión (línea sólida), para $n = 900$. Los estimadores de la función drift y difusión vienen dados por

$$\hat{m}_h(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right) (r_{t_{i+1}} - r_{t_i})}{\Delta \sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right)},$$

$$\hat{\sigma}_h^2(r) = \frac{\sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right) (r_{t_{i+1}} - r_{t_i})^2}{\Delta \sum_{i=0}^{n-1} K\left(\frac{r-r_{t_i}}{h}\right)},$$

respectivamente.

Se ha utilizado el núcleo Gaussiano y la ventana se ha calculado utilizando la regla de Scott, obteniendo $h = 0.164$. A la vista de la figura, la estimación que obtenemos con este parámetro de ventana es demasiado ruidosa para ambas funciones, por lo que también se ha considerado un ancho de banda mayor, a efectos comparativos. Notar que para estimar la función drift parece conveniente utilizar una ventana mayor que para la función de difusión.

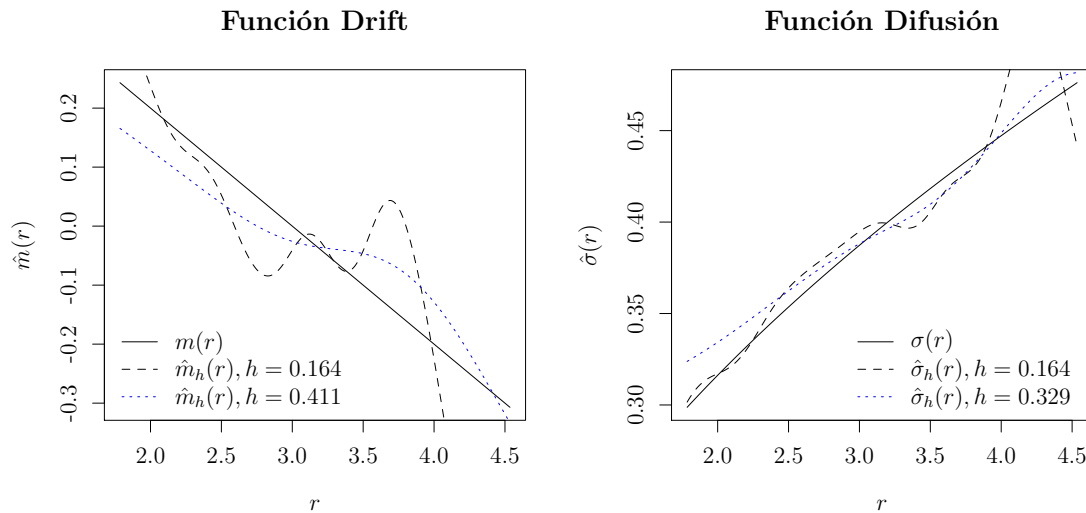


Figura 4.4: Estimación no paramétrica de las funciones drift (izquierda) y difusión (derecha) del proceso $dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t$, para tamaño muestral $n = 900$ y $\Delta = 1/12$. Las líneas sólidas son las funciones teóricas y las líneas punteadas son estimaciones no paramétricas utilizando dos parámetros ventana h diferentes. El ancho de banda de la línea punteada negra, $h = 0.164$, ha sido calculado con la regla de Scott.

Densidad estacionaria CIR

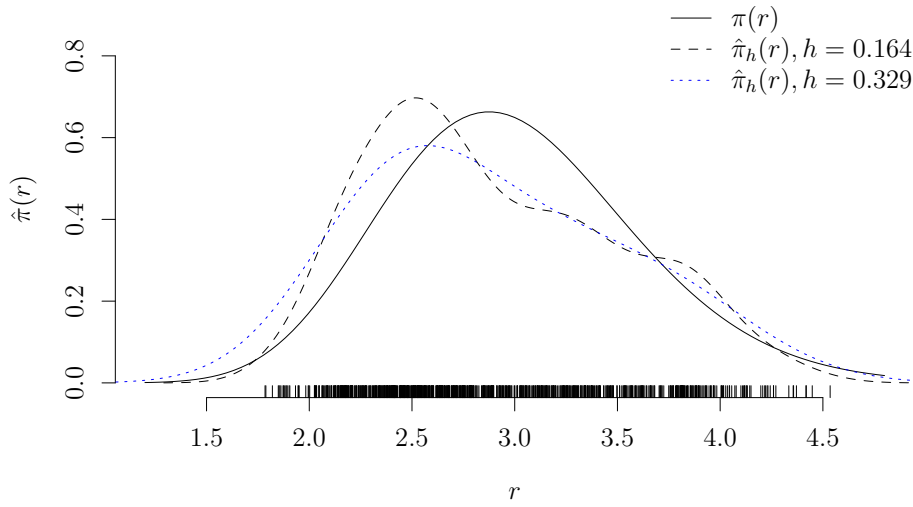


Figura 4.5: *Estimación no paramétrica de la densidad estacionaria del proceso $dr_t = (0.6 - 0.2r_t)dt + \sqrt{0.05r_t}dW_t$, para tamaño muestral $n = 900$ y $\Delta = 1/12$. La línea sólida es la verdadera densidad y las líneas punteadas son estimaciones utilizando dos parámetros ventana h diferentes. El ancho de banda de la línea punteada negra, $h = 0.164$, ha sido calculado con la regla de Scott.*

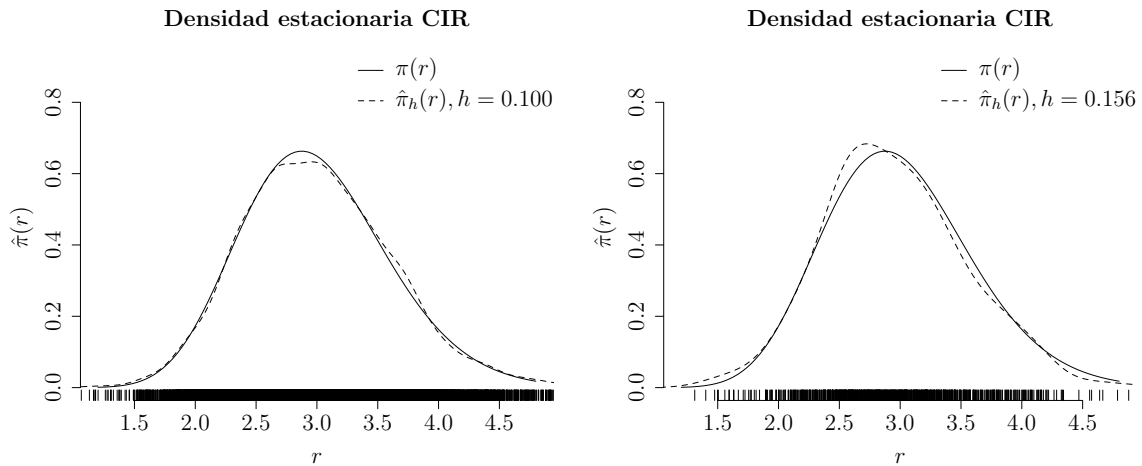
Para la estimación de la densidad estacionaria, Figura 4.5, se ha utilizado el estimador tipo núcleo

$$\hat{\pi}_h(r) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{r - r_{t_i}}{h}\right).$$

Al igual que para las funciones drift y difusión, se ha utilizado el núcleo Gaussiano y la ventana se ha calculado con la regla de Scott. Parece que, como sucedía con las funciones drift y difusión, la ventana es un poco pequeña, ya que la estimación obtenida (línea punteada negra) no es tan suave como la verdadera densidad (línea sólida). A efectos comparativos, se ha incluido una estimación con un ventana mayor (en azul).

Aunque este método no requiere que $\Delta = 1$, un análisis empírico puede mostrar que la estimación tipo núcleo de la densidad $\pi(r)$ cuando $\Delta < 1$ es menos precisa. En la Figura 4.6b, se muestra la estimación tipo núcleo de la densidad estacionaria para el mismo tamaño muestral, $n = 900$, pero con $\Delta = 1$. Se puede apreciar como la estimación es más precisa que con $\Delta = 1/12$.

Para mejorar la estimación cuando $\Delta < 1$, podemos incrementar el tamaño muestral. En la Figura 4.6a se muestra la estimación tipo núcleo para el proceso CIR con $\Delta = 1/12$ y tamaño muestral $n = 12000$. Si la comparamos con la Figura 4.5, vemos que al incrementar el tamaño muestral, manteniendo el mismo valor de Δ , la estimación tipo núcleo es más precisa. De nuevo, el parámetro de suavizado se ha calculado con la regla de Scott.



(a) Estimación no paramétrica de la densidad estacionaria del proceso $dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t$, para tamaño muestral $n = 12000$ y $\Delta = 1/12$. Se puede apreciar que al incrementar el tamaño muestral, la estimación es más precisa que en la Figura 4.5.

(b) Estimación no paramétrica de la densidad estacionaria del proceso $dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t$, para tamaño muestral $n = 900$ y $\Delta = 1$. Manteniendo el mismo tamaño muestral que en la Figura 4.5, al tomar $\Delta = 1$ la estimación es más precisa.

Figura 4.6: Efecto del tamaño muestral n y del paso de discretización Δ en la estimación tipo núcleo de la densidad estacionaria $\pi(r)$.

4.2.1. Estudio de simulación

En esta sección se plantea un estudio de simulación para comparar la estimación paramétrica y no paramétrica de la densidad estacionaria del proceso CIR. Consideraremos el proceso CIR

$$dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t,$$

y su estimación no paramétrica utilizando el estimador tipo núcleo visto en la sección anterior, $\hat{\pi}_h$. Para la estimación paramétrica, tenemos que la densidad estacionaria del modelo CIR sigue una Gamma con parámetro de forma $2\theta_1/\theta_3^2$ y parámetro de escala $\theta_3^2/2\theta_2$, por lo que sustituiremos los parámetros por sus respectivas estimaciones $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$. En la Figura 4.7 se muestra la densidad estacionaria teórica y su estimación paramétrica y no paramétrica para el ejemplo de la sección anterior.

Para la simulación, consideraremos 1000 réplicas, con $n = 900$, y como criterio comparativo entre las estimaciones de la densidad y la verdadera densidad estacionaria utilizaremos la divergencia de Kullback-Leibler (KL), también conocida como divergencia de la información o entropía relativa, que es una medida de la similitud o diferencia entre dos funciones de distribución de probabilidad. Para dos variables aleatoria continuas, la divergencia KL se define como la integral

$$D_{KL}(p, q) = \int p(x) \log \frac{p(x)}{q(x)} dx,$$

donde p y q son funciones de densidad. Notar que D_{KL} es cero si y sólo si $p = q$. Tomamos como $q(x)$ la densidad estacionaria teórica $\pi(x)$ y como $p(x)$ el estimador $\hat{\pi}_h$ para la estimación no paramétrica y $\pi_{\hat{\theta}}$ para la estimación paramétrica. En el Cuadro 4.3 se muestran los estadísticos de la medida D_{KL} para la estimación paramétrica y para la

estimación no paramétrica respecto de la densidad estacionaria teórica. Además, en la Figura 4.8 se han representado los diagramas de caja para la D_{KL} de ambas estimaciones respecto a la densidad estacionaria teórica.

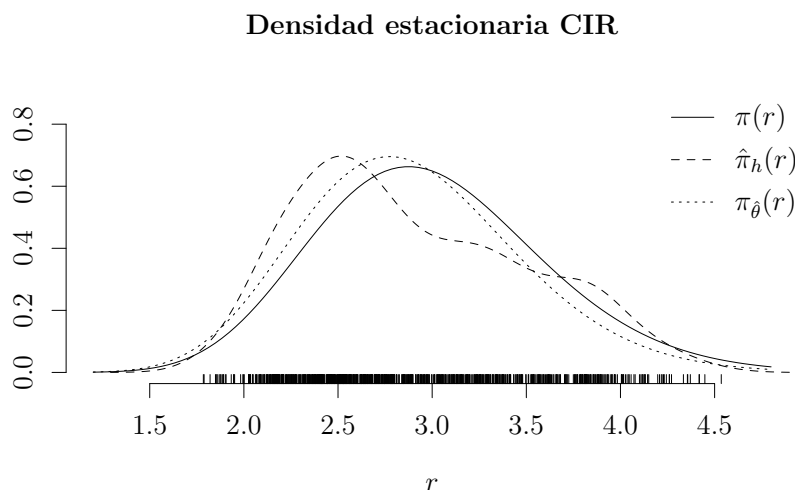


Figura 4.7: Densidad estacionaria del proceso $dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t$, para tamaño muestral $n = 900$ y $\Delta = 1/12$. La línea sólida es la verdadera densidad, la línea rayada es la estimación no paramétrica ($\hat{\pi}_h$), con el parámetro de suavizado calculado con la regla de Scott, y la línea punteada es la estimación paramétrica ($\pi_{\hat{\theta}}$), es decir, la densidad estacionaria con parámetro θ igual a la estimación $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)'$.

Estimación	Cuartiles					Media
	Mín.	25 %	50 %	75 %	Máx.	
Paramétrica	0.0001	0.0257	0.0642	0.1325	0.7298	0.0942
No paramétrica	0.0053	0.0511	0.0881	0.1492	0.6519	0.1141

Cuadro 4.3: Estadísticos para la D_{KL} entre la estimación paramétrica de la densidad estacionaria y la densidad teórica y la estimación no paramétrica de la densidad estacionaria y la verdadera densidad, en 1000 muestras de tamaño $n = 900$ del proceso CIR $dr_t = (0.6 - 0.2r_t) dt + \sqrt{0.05r_t} dW_t$. Se muestran los valores mínimos y máximos, los cuartiles y la media.

En media, la estimación paramétrica presenta menor divergencia con la densidad teórica y llega a alcanzar valores muy cercanos a cero. Cabe notar que el parámetro de suavizado en la estimación no paramétrica se ha calculado con la regla de Scott y, como vimos en la sección anterior (ver Figura 4.5), para el tamaño muestral de $n = 900$ y $\Delta = 1/12$ parece que sería más adecuado un ancho banda mayor. En la Figura 4.9 se muestra el diagrama de caja de los anchos de banda, calculados con la regla de Scott, de las 1000 simulaciones. La media de los h está en torno a 0.15, por lo que consideraremos distintos anchos de banda y realizaremos la simulación de nuevo para ver cómo afecta el parámetro de suavizado en la estimación no paramétrica.

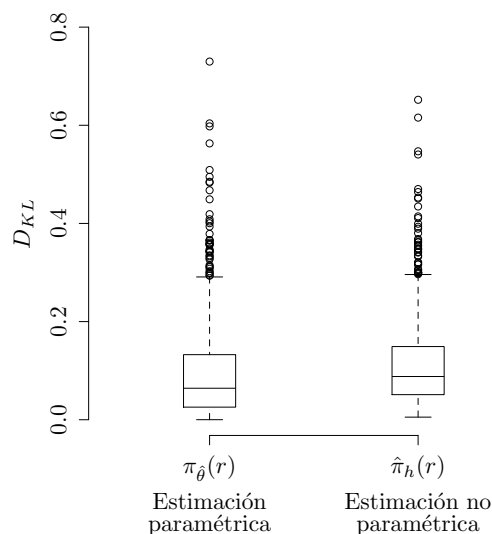


Figura 4.8: Diagramas de caja de D_{KL} entre la estimación paramétrica de la densidad estacionaria y la densidad teórica y la estimación no paramétrica de la densidad estacionaria y la verdadera densidad, con 1000 réplicas del proceso CIR.

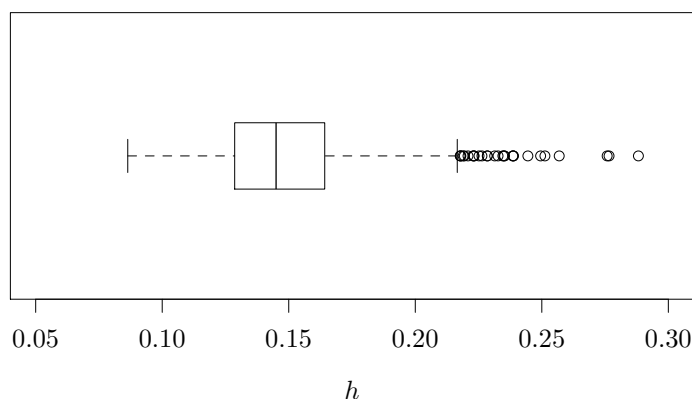


Figura 4.9: Diagrama de caja de los parámetros de suavizado h , calculados con la regla de Scott, de las 1000 simulaciones.

En el Cuadro 4.4 se muestran los estadísticos de la divergencia de Kullback-Leibler de la densidad estacionaria teórica π a la estimación no paramétrica de la densidad con distintos parámetros de suavizado h . Se ha incluido en la simulación el ancho de banda calculado con la regla de Scott (*regla del pulgar*) a efectos comparativos. Notar que el parámetro $h = 0.30$ tiene, en media, una menor divergencia y, si la comparamos con la obtenida por la estimación paramétrica (ver Cuadro 4.3) es incluso menor. El parámetro h calculado con la regla de Scott, en media, está en torno a 0.15, por lo que es necesaria duplicarlo para obtener una D_{KL} menor que con la estimación paramétrica. En la Figura 4.10 se han representado los diagramas de caja correspondientes a las 1000 D_{KL} de los distintos parámetros de suavizado considerados. Notar que en la representación se han suprimido los valores atípicos.

Parámetro ventana	Cuartiles					Máx.	Media
	Mín.	25 %	50 %	75 %			
Regla del pulgar	0.0051	0.0541	0.0851	0.1430	0.7240	0.1111	
$h = 0.15$	0.0055	0.0519	0.0818	0.1338	0.7082	0.1074	
$h = 0.20$	0.0034	0.0419	0.0723	0.1199	0.7088	0.0971	
$h = 0.25$	0.0012	0.0349	0.0654	0.1136	0.7162	0.0906	
$h = 0.30$	0.0010	0.0321	0.0623	0.1133	0.7240	0.0886	
$h = 0.35$	0.0010	0.0337	0.0649	0.1212	0.7298	0.0917	
$h = 0.40$	0.0031	0.0395	0.0742	0.1321	0.7335	0.1001	
$h = 0.45$	0.0045	0.0517	0.0910	0.1482	0.7358	0.1137	
$h = 0.50$	0.0112	0.0699	0.1118	0.1685	0.7376	0.1322	

Cuadro 4.4: Estadísticos de D_{KL} entre la estimación no paramétrica de la densidad estacionaria y la verdadera densidad, en 1000 muestras de tamaño $n = 900$ del proceso CIR $dr_t = (0.6 - 0.2r_t)dt + \sqrt{0.05r_t}dW_t$, para distintos valores del parámetro de suavizado h . Se muestran los valores mínimos y máximos, los cuartiles y la media.

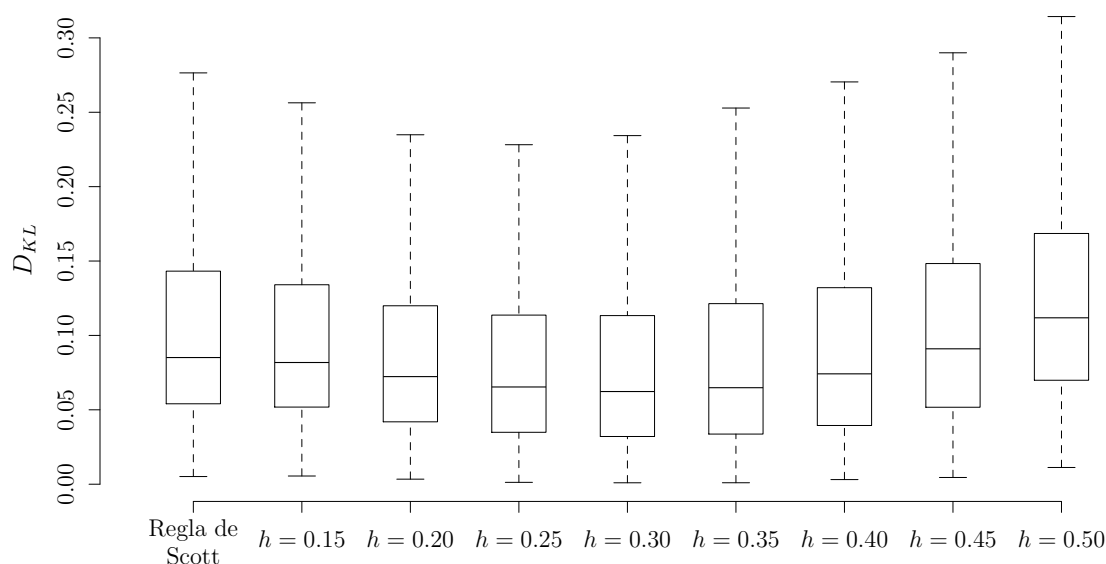


Figura 4.10: Diagramas de caja de D_{KL} entre la estimación no paramétrica de la densidad estacionaria y la verdadera densidad, con 1000 réplicas del proceso CIR, para distintos valores del parámetro de suavizado h .

4.3. Procedimiento bootstrap

En esta sección veremos cómo utilizar técnicas bootstrap en el filtro de Kalman que nos permitirán aproximar la distribución de los estimadores de los parámetros. Para ello, introduciremos un algoritmo para remuestrear en modelos de espacio de estados que podemos encontrar en Stoffer y Wall (1991). Este enfoque utiliza el bootstrap mediante métodos de Monte Carlo no paramétricos y se centra en los estimadores Gaussianos máximo-verosímiles.

La inferencia en los modelos de espacio de estados estimados mediante el filtro de Kalman es factible porque existe una teoría asintótica. Como hemos visto en la Sección 3.1.5, bajo condiciones adecuadas tanto las estimaciones de los parámetros obtenidas por técnicas de máxima verosimilitud como la estimación del estado del filtro de Kalman son consistentes y asintóticamente normales.

Si recordamos la ecuación de las innovaciones (3.1.5), podemos reescribir el modelo para obtener la denominada *representación en forma de innovaciones*,

$$\begin{aligned}\mathbf{x}_{t+1}^t &= \Phi \mathbf{x}_t^{t-1} + \Upsilon \mathbf{u}_{t+1} + K_t \boldsymbol{\varepsilon}_t, \\ \mathbf{y}_t &= A_t \mathbf{x}_t^{t-1} + \Gamma \mathbf{u}_t + \boldsymbol{\varepsilon}_t.\end{aligned}$$

Es importante notar que, aunque las innovaciones $\boldsymbol{\varepsilon}_t$ están incorreladas, inicialmente, Σ_t puede diferir mucho para distintos momentos del tiempo t . Por ello, en un procedimiento de remuestreo podemos optar por ignorar los primeros valores de $\boldsymbol{\varepsilon}_t$ hasta que Σ_t se establezca o trabajar con las innovaciones estandarizadas,

$$\mathbf{e}_t = \Sigma_t^{-1/2} \boldsymbol{\varepsilon}_t,$$

de modo que garantizamos que las innovaciones \mathbf{e}_t tienen, al menos, los primeros dos momentos iguales.

El procedimiento bootstrap para los modelos de espacio de estados se define por un algoritmo de 5 etapas. Sea $\hat{\boldsymbol{\theta}}$ el estimador de máxima verosimilitud de $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \mathcal{L}_Y(\boldsymbol{\theta})$, y $\boldsymbol{\varepsilon}_t(\hat{\boldsymbol{\theta}})$ y $\Sigma_t(\hat{\boldsymbol{\theta}})$ los valores obtenidos tras ejecutar el filtro bajo $\hat{\boldsymbol{\theta}}$. El procedimiento bootstrap consiste en las siguientes etapas:

- (i) Construir las innovaciones estandarizadas $\mathbf{e}_t = \Sigma_t^{-1/2} \boldsymbol{\varepsilon}_t$.
- (ii) Remuestrear, con reemplazamiento, n veces $\{\mathbf{e}_1(\hat{\boldsymbol{\theta}}), \dots, \mathbf{e}_n(\hat{\boldsymbol{\theta}})\}$ para obtener una remuestra bootstrap de las innovaciones estandarizadas, $\{\mathbf{e}_1^*(\hat{\boldsymbol{\theta}}), \dots, \mathbf{e}_n^*(\hat{\boldsymbol{\theta}})\}$.
- (iii) Utilizando la *representación en forma de innovaciones*, construir el conjunto de datos $\{\mathbf{y}_1^*, \dots, \mathbf{y}_n^*\}$. Para ello, se define el vector $\boldsymbol{\zeta}_t = (\mathbf{x}_{t+1}^t, \mathbf{y}_t^t)'$, de modo que

$$\boldsymbol{\zeta}_t = F_t \boldsymbol{\zeta}_{t-1} + G \mathbf{u}_t + H_t \mathbf{e}_t, \quad (4.3)$$

donde

$$F_t = \begin{bmatrix} \Phi & 0 \\ A_t & 0 \end{bmatrix}, \quad G = \begin{bmatrix} \Upsilon \\ \Gamma \end{bmatrix}, \quad H_t = \begin{bmatrix} K_t \Sigma_t^{1/2} \\ \Sigma_t^{1/2} \end{bmatrix}.$$

Para construir el conjunto de datos $\{\mathbf{y}_1^*, \dots, \mathbf{y}_n^*\}$, utilizamos la ecuación (4.3) con $\mathbf{e}_t^*(\hat{\boldsymbol{\theta}})$ en lugar de \mathbf{e}_t . La variable exógena \mathbf{u}_t y los valores iniciales permanecen fijos, así como el vector de parámetros, fijado en $\hat{\boldsymbol{\theta}}$.

- (iv) Utilizando la remuestra bootstrap $\{\mathbf{y}_t^*\}_{t=1}^n$ construir la verosimilitud $\mathcal{L}_{Y^*}(\boldsymbol{\theta})$ y obtener el estimador de máxima verosimilitud de $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^*$.
- (v) Repetir B veces los pasos 2-4 para obtener las réplicas bootstrap $\{\hat{\boldsymbol{\theta}}_b^*\}_{b=1}^B$. La distribución en el muestreo de $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ puede ser aproximada por la distribución empírica de $\{\hat{\boldsymbol{\theta}}_b^* - \hat{\boldsymbol{\theta}}\}_{b=1}^B$.

4.3.1. Justificación asintótica del procedimiento

En esta sección veremos que si el número de réplicas bootstrap $B \rightarrow \infty$ y el tamaño muestral $n \rightarrow \infty$, la distribución asintótica del estimador de máxima verosimilitud y el estimador bootstrap son equivalentes (se puede consultar en Stoffer y Wall, 1991). Para garantizar la estabilidad asintótica del filtro, asumiremos que $\Phi(\boldsymbol{\theta})$ posee todos sus autovalores dentro del círculo unitario y que $A(\boldsymbol{\theta})Q(\boldsymbol{\theta})A(\boldsymbol{\theta})' + R(\boldsymbol{\theta})$ es definida positiva para $\boldsymbol{\theta} \in \mathcal{P}$, donde \mathcal{P} es un espacio compacto, normalmente un subconjunto del espacio euclídeo \mathbb{R}^d .

Los resultados para el procedimiento bootstrap dependen de la existencia de teoría asintótica para los estimadores de máxima verosimilitud del modelo de espacio de estados. Por ello, se presentarán brevemente los principales resultados.

Consistencia y distribución asintótica de los estimadores de máximo verosimilitud

Sea $\boldsymbol{\theta}_0$ el verdadero vector de parámetros desconocido que determina $\Phi(\boldsymbol{\theta}_0)$, $\Upsilon(\boldsymbol{\theta}_0)$, $A(\boldsymbol{\theta}_0)$, $\Gamma(\boldsymbol{\theta}_0)$, $Q(\boldsymbol{\theta}_0)$ y $R(\boldsymbol{\theta}_0)$. Sea $\hat{\boldsymbol{\theta}}_n$ la estimación consistente y asintóticamente normal de $\boldsymbol{\theta}_0$ obtenida al minimizar

$$V_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^n \log |\Sigma_t(\boldsymbol{\theta})| + \mathbf{e}_t'(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\theta}) = n^{-1} \ell_Y(\boldsymbol{\theta}),$$

donde $\ell_Y(\boldsymbol{\theta}) = 2 \ln \mathcal{L}_Y(\boldsymbol{\theta})$. Notar que $\dot{V}_n(\bar{\boldsymbol{\theta}}_n) = \partial V_n(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_n} = 0$. Sea $W_n(\boldsymbol{\theta}) = \mathbb{E}[V_n(\boldsymbol{\theta})]$ y asumiendo que $W_n(\boldsymbol{\theta})$ tiene un mínimo global en $\bar{\boldsymbol{\theta}}_n$, entonces, bajo condiciones adecuadas, cuando $n \rightarrow \infty$, $\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_n \rightarrow 0$ casi seguro y

$$\sqrt{n} B_n^{-1/2} (\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_n) \xrightarrow{d} N(0, I), \quad (4.4)$$

donde

$$B_n = \left[\ddot{W}_n(\bar{\boldsymbol{\theta}}_n) \right]^{-1} U_n(\bar{\boldsymbol{\theta}}_n) \left[\ddot{W}_n(\bar{\boldsymbol{\theta}}_n) \right]^{-1}$$

y

$$U_n(\bar{\boldsymbol{\theta}}_n) = n \mathbb{E} \left[\dot{V}_n(\bar{\boldsymbol{\theta}}_n) \dot{V}_n(\bar{\boldsymbol{\theta}}_n)' \right],$$

con $\ddot{W}_n(\bar{\boldsymbol{\theta}}_n) = \partial^2 W_n(\bar{\boldsymbol{\theta}}_n) / \partial \bar{\boldsymbol{\theta}}_n \partial \bar{\boldsymbol{\theta}}_n'$. Si, además, $W_n(\boldsymbol{\theta}) \rightarrow W(\boldsymbol{\theta})$ uniformemente en $\boldsymbol{\theta}$ cuando $n \rightarrow \infty$ y $W(\boldsymbol{\theta})$ tiene un mínimo global en $\boldsymbol{\theta}_0$, entonces bajo condiciones adecuadas,

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, B),$$

donde

$$B = \left[\ddot{W}(\boldsymbol{\theta}_0) \right]^{-1} U(\boldsymbol{\theta}_0) \left[\ddot{W}(\boldsymbol{\theta}_0) \right]^{-1}$$

y $U(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} U_n(\bar{\boldsymbol{\theta}}_n)$. Por último, bajo las condiciones en las que se cumple (4.4),

$$V_n(\boldsymbol{\theta}) - W_n(\boldsymbol{\theta}) \rightarrow 0 \quad \text{casi seguro} \quad (4.5)$$

y

$$\begin{aligned} \dot{V}_n(\boldsymbol{\theta}) - \dot{W}_n(\boldsymbol{\theta}) &\rightarrow 0, \\ \ddot{V}_n(\boldsymbol{\theta}) - \ddot{W}_n(\boldsymbol{\theta}) &\rightarrow 0, \end{aligned}$$

casi seguro cuando $n \rightarrow \infty$ uniformemente en $\boldsymbol{\theta} \in \mathcal{P}$.

Justificación asintótica del bootstrap

En esta sección asumiremos que se cumplen las condiciones necesarias para el estimador máximo verosímil $\hat{\theta}_n$ expuestas en la sección 4.3.1. Sea $\{e_t^*(\theta)\}_{t=1}^n$ una remuestra bootstrap de innovaciones estandarizadas, con $V_n^*(\theta) = n^{-1} \sum_{t=1}^n \log |\Sigma_t(\theta)| + e_t^{*\prime}(\theta)e_t^*(\theta)$ y sea $W_n^*(\theta) = \mathbb{E}^*[V_n^*(\theta)]$ donde \mathbb{E}^* denota la esperanza respecto de la distribución bootstrap.

Lema 1. $W_n^*(\theta) = V_n^*(\theta)$ para todo $\theta \in \mathcal{P}$, por lo que $\hat{\theta}_n$ minimiza $W_n^*(\theta)$.

Demostración. Notar que,

$$\mathbb{E}^*[e_t^{*\prime}(\theta)e_t^*(\theta)] = n^{-1} \sum_{j=1}^n e_j'(\theta)e_j(\theta),$$

a lo que le sigue que

$$V_n(\theta) - W_n^*(\theta) = n^{-1} \sum_{t=1}^n e_t'(\theta)e_t(\theta) - \mathbb{E}^*[e_t^{*\prime}(\theta)e_t^*(\theta)] = 0.$$

Sea $\hat{\theta}_n^*$ el vector que minimiza $V_n^*(\theta)$. Entonces,

$$\mathbf{0} = [\partial V_n^*(\theta)/\partial\theta_1, \dots, \partial V_n^*(\theta)/\partial\theta_k]' |_{\theta=\hat{\theta}_n^*} = \dot{V}_n^*(\hat{\theta}_n^*).$$

Desarrollamos $\dot{V}_n^*(\hat{\theta}_n^*)$ en $\hat{\theta}_n^*$ para obtener

$$\mathbf{0} = \dot{V}_n^*(\hat{\theta}_n^*) = \dot{V}_n^*(\hat{\theta}_n) + \ddot{V}_n^*(\xi_n)(\hat{\theta}_n^* - \hat{\theta}_n),$$

donde $\ddot{V}_n^*(\xi_n) = \partial^2 V_n^*(\xi_n)/\partial\theta\partial\theta'$ de modo que ξ_n se encuentra en el segmento que conecta los puntos $\hat{\theta}_n^*$ y $\hat{\theta}_n$ en \mathcal{P} . De este modo, siempre que $\ddot{V}_n^*(\xi_n)$ sea invertible, tenemos que $(\hat{\theta}_n^* - \hat{\theta}_n) = [\ddot{V}_n^*(\xi_n)]^{-1} \dot{V}_n^*(\hat{\theta}_n)$. Entonces, como vimos en (4.4), tenemos que

$$\sqrt{n} B_n^{*-1/2} (\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d} N(0, I)$$

cuando $n \rightarrow \infty$, donde

$$\begin{aligned} B_n^* &= [\ddot{W}_n^*(\hat{\theta}_n)]^{-1} U_n^*(\hat{\theta}_n) [\ddot{W}_n^*(\hat{\theta}_n)]^{-1}, \\ U_n^*(\hat{\theta}_n) &= n \mathbb{E}^*[\dot{V}_n^*(\hat{\theta}_n) \dot{V}_n^*(\hat{\theta}_n)'], \\ \ddot{W}_n^*(\hat{\theta}_n) &= \mathbb{E}^*[\ddot{V}_n^*(\hat{\theta}_n)]. \end{aligned}$$

□

Para establecer los principios del bootstrap, por lo visto en la sección 4.3.1 tenemos que $B_n^* - B_n \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$. Por el Lema 1 tenemos que $\ddot{W}_n^*(\hat{\theta}_n) = \ddot{V}_n^*(\hat{\theta}_n)$ y, por tanto, $\ddot{W}_n^*(\hat{\theta}_n) - \ddot{W}_n^*(\bar{\theta}_n) \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$. A continuación expon-dremos que $U_n^*(\hat{\theta}_n)$ y $U_n(\bar{\theta}_n)$ difieren en una cantidad pequeña para un n suficientemente grande.

Lema 2. $U_n^*(\hat{\theta}_n) - U_n(\bar{\theta}_n) \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$.

Demostración. Sea $U_n^*(a, b) = n\mathbb{E}^*[(\partial V^*/\partial\theta_a)(\partial V^*/\partial\theta_b)]|_{\theta=\hat{\theta}_n}$ el (a, b) -ésimo elemento de la matriz $U_n^*(\hat{\theta}_n)$ para $a, b = 1, \dots, k$. Denotando $C_{ta} = \partial \log |\Sigma_t| / \partial\theta_a$ y $Z_{ta}^* = \partial(e_t^* e_t^*) / \partial\theta_a$ de modo que $nU_n^*(a, b) = \mathbb{E}[\sum_{t=1}^n \sum_{s=1}^n (C_{ta} + Z_{ta}^*)(C_{sb} + Z_{sb}^*)] |_{\theta=\hat{\theta}_n}$, escribimos

$$nU_n^*(a, b) = \sum_{t=1}^n \sum_{s=1}^n [(C_{ta} + \mathbb{E}^*(Z_{ta}^*))(C_{sb} + \mathbb{E}^*(Z_{sb}^*)) - \mathbb{E}^*(Z_{ta}^*)\mathbb{E}^*(Z_{sb}^*) + \mathbb{E}^*(Z_{ta}^*Z_{sb}^*)] |_{\theta=\hat{\theta}_n},$$

por lo que

$$\begin{aligned} U_n^*(a, b) &= n\mathbb{E}^*(\partial V^*/\partial\theta_a)\mathbb{E}^*(\partial V^*/\partial\theta_b) |_{\theta=\hat{\theta}_n} \\ &\quad + n^{-1} \sum_{t=1}^n \sum_{s=1}^n [\mathbb{E}^*(Z_{ta}^*Z_{sb}^*) - \mathbb{E}^*(Z_{ta}^*)\mathbb{E}^*(Z_{sb}^*)] |_{\theta=\hat{\theta}_n}. \end{aligned} \quad (4.6)$$

El primer término en (4.6) es el (a, b) -ésimo elemento de $n[\dot{W}_n^*(\hat{\theta}_n)][\dot{W}_n^*(\hat{\theta}_n)]'$ y, por el Lema 1, es cero. Para evaluar el segundo término, notar que

$$\begin{aligned} \mathbb{E}^*[Z_{ta}^*] &= n^{-1} \sum_{j=1}^n Z_{ja}, \\ \mathbb{E}^*[Z_{ta}^*Z_{tb}^*] &= n^{-1} \sum_{j=1}^n Z_{ja}Z_{jb}, \\ \mathbb{E}^*[Z_{ta}^*Z_{sb}^*] &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n Z_{ia}Z_{jb}, \quad s \neq t, \end{aligned}$$

donde $Z_{ta} = \partial e_t^* e_t^* / \partial\theta_a$. Sustituyendo en la ecuación (4.6) tenemos

$$U_n^*(a, b) = \left[\left(n^{-1} \sum_{t=1}^n Z_{ta}Z_{tb} \right) - \left(n^{-1} \sum_{t=1}^n Z_{ta} \right) \left(n^{-1} \sum_{s=1}^n Z_{sb} \right) \right] |_{\theta=\hat{\theta}_n}.$$

A continuación evaluamos $U_n(a, b) = n\mathbb{E}[(\partial V/\partial\theta_a)(\partial V/\partial\theta_b)]|_{\theta=\hat{\theta}_n}$. Como hemos visto en (4.6), se puede mostrar que

$$U_n(a, b) = n^{-1} \sum_{t=1}^n \sum_{s=1}^n [\mathbb{E}(Z_{ta}Z_{sb}) - \mathbb{E}(Z_{ta})\mathbb{E}(Z_{sb})] |_{\theta=\hat{\theta}_n}, \quad (4.7)$$

donde se ha utilizado que $\dot{W}_n = \mathbf{0}$. Como $U_n(\hat{\theta}_n) \rightarrow U(\theta_0)$ cuando $n \rightarrow \infty$, entonces $U_n(a, b)$ es asintóticamente equivalente a

$$n^{-1} \sum_{t=1}^n \sum_{s=1}^n [\mathbb{E}(Z_{ta}Z_{sb}) - \mathbb{E}(Z_{ta})\mathbb{E}(Z_{sb})] |_{\theta=\theta_0}.$$

Utilizando que las verdaderas innovaciones, $e_t(\theta_0)$ y $e_s(\theta_0)$, están incorreladas, $s \neq t$, tenemos que (4.7) es asintóticamente equivalente a

$$n^{-1} \sum_{t=1}^n [\mathbb{E}(Z_{ta}Z_{tb}) - \mathbb{E}(Z_{ta})\mathbb{E}(Z_{tb})] |_{\theta=\theta_0}.$$

Por la estabilidad del filtro, cuando $n \rightarrow \infty$, tenemos que

$$n^{-1} \sum_{t=1}^n \mathbb{E}[Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0)] \rightarrow \mu_4(a, b),$$

$$n^{-1} \sum_{t=1}^n \mathbb{E}[Z_{ta}(\boldsymbol{\theta}_0)]\mathbb{E}[Z_{tb}(\boldsymbol{\theta}_0)] \rightarrow \mu_2(a)\mu_2(b),$$

con

$$\mu_4(a, b) = \mathbb{E} \left[\frac{\partial \mathbf{u}'_t \Sigma^{-1} \mathbf{u}_t}{\partial \theta_a} \frac{\partial \mathbf{u}'_t \Sigma^{-1} \mathbf{u}_t}{\partial \theta_b} \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

$$\mu_2(a) = \mathbb{E}[\partial \mathbf{u}'_t \Sigma^{-1} \mathbf{u}_t / \partial \theta_a] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

donde $\{\mathbf{u}_t\}$ denota la secuencia de innovaciones del estado estacionario y Σ es la matriz de varianzas-covarianzas del error del estado estacionario, $\Sigma = \mathbb{E}[\mathbf{u}_t \mathbf{u}'_t]$.

Para la demostración de que $n^{-1} \sum_{t=1}^n Z_{ta}(\hat{\boldsymbol{\theta}}_n)Z_{tb}(\hat{\boldsymbol{\theta}}_n) \rightarrow \mu_4(a, b)$ y que $n^{-1} \sum_{t=1}^n Z_{ta}(\hat{\boldsymbol{\theta}}_n) \rightarrow \mu_2(a)$ casi seguro cuando $n \rightarrow \infty$, utilizaremos el resultado del Teorema 1, (b) y (d) de Watanabe (1984), por lo que tenemos $n^{-1} \sum_{t=1}^n \mathbf{e}_t(\hat{\boldsymbol{\theta}}_n) = n^{-1} \sum_{t=1}^n \mathbf{e}_t(\boldsymbol{\theta}_0) + o_{c.s.}(1)$ y, por las condiciones de diferenciabilidad y acotación en $Z_{ta}(\boldsymbol{\theta})$,

$$n^{-1} \sum_{t=1}^n Z_{ta}(\hat{\boldsymbol{\theta}}_n) = n^{-1} \sum_{t=1}^n Z_{ta}(\boldsymbol{\theta}_0) + o_{c.s.}(1),$$

$$n^{-1} \sum_{t=1}^n Z_{ta}(\hat{\boldsymbol{\theta}}_n)Z_{tb}(\hat{\boldsymbol{\theta}}_n) = n^{-1} \sum_{t=1}^n Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0) + o_{c.s.}(1).$$

Por (4.5), podemos escribir

$$n^{-1} \sum_{t=1}^n Z_{ta}(\boldsymbol{\theta}_0) = n^{-1} \sum_{t=1}^n \mathbb{E}[Z_{ta}(\boldsymbol{\theta}_0)] + o_{c.s.}(1),$$

$$n^{-1} \sum_{t=1}^n Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0) = n^{-1} \sum_{t=1}^n \mathbb{E}[Z_{ta}(\boldsymbol{\theta}_0)Z_{tb}(\boldsymbol{\theta}_0)] + o_{c.s.}(1),$$

y, por la estacionaridad de la secuencia de innovaciones, $U_n^*(\hat{\boldsymbol{\theta}}_n) - U_n(\bar{\boldsymbol{\theta}}_n) \rightarrow 0$ casi seguro cuando $n \rightarrow \infty$. \square

4.3.2. Aplicación del procedimiento a la serie simulada

En la sección 4.1 utilizamos el filtro de Kalman para obtener la estimación de los parámetros del proceso de difusión

$$dr_t = (\theta_1 - \theta_2 r_t)dt + \theta_3 \sqrt{r_t} dW_t,$$

con $\boldsymbol{\theta} = (0.6, 0.2, \sqrt{0.05})'$. Para ello, considerábamos el proceso discretizado,

$$z_{t_i} = (\theta_1 - \theta_2 r_{t_i})\Delta + \theta_3 \sqrt{r_{t_i}} \Delta^{1/2} \varepsilon_{t_i}, \quad i = 1, \dots, n,$$

donde $z_{t_i} = r_{t_{i+1}} - r_{t_i}$. Utilizando notación de modelo de espacio de estados,

$$\begin{aligned}x_t &= \Phi x_{t-1} + \Upsilon \mathbf{u}_t + w_t, & w_t &\sim \text{iid } N(0, Q), \\y_t &= Ax_t,\end{aligned}\tag{4.8}$$

donde, $y_t = z_t \Delta^{-1}$, $A = 1$, $x_t = y_t$, $\Phi = 0$, $\mathbf{u}_t = (1, r_{t-1})'$, $\Upsilon = (\theta_1, -\theta_2)$, $Q = \theta_3^2 r_{t-1} \Delta^{-1}$ y $\Delta = 1/12$. El vector de parámetros es $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$.

En el Cuadro 4.5 se muestra el resultado de la estimación de los parámetros del proceso, con $n = 900$, a través del procedimiento Newton-Raphson, así como los errores estándar obtenidos de $B = 1000$ remuestras bootstrap. Destacar que el procedimiento bootstrap empleado se ha basado en las innovaciones estandarizadas, pero se han excluido del muestreo los tres primeros valores, $\mathbf{e}_t^*(\hat{\boldsymbol{\theta}}) = \mathbf{e}_t(\hat{\boldsymbol{\theta}})$ para $1 \leq t \leq 3$, para evitar problemas causados por el comportamiento transitorio del filtro en el intervalo $1 \leq t \leq 3$. Esta modificación es recomendable para evitar remuestras bootstrap \mathbf{y}_t^* no representativas, debido a valores inusuales de Σ_t y K_t (ver Stoffer y Wall, 1991).

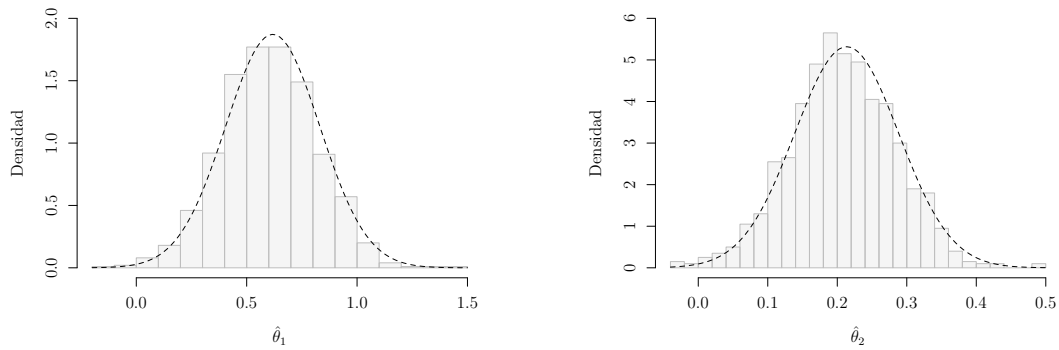
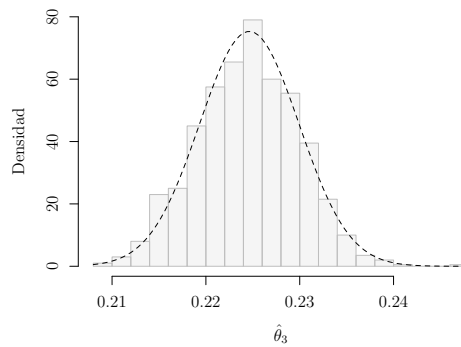
Parámetro	Valor teórico	Estimación	Error estándar asintótico	Error estándar bootstrap
θ_1	0.6000	0.6180	0.2132	0.2144
θ_2	0.2000	0.2136	0.0750	0.0759
θ_3	0.2236	0.2246	0.0053	0.0054

Cuadro 4.5: Estimación de los parámetros, error estándar asintótico asociado y error estándar bootstrap del modelo $dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 \sqrt{r_t} dW_t$.

Los errores estándar bootstrap son la desviación estándar de las estimaciones bootstrap, es decir, la raíz cuadrada de

$$\sum_{b=1}^B \frac{(\theta_{ib}^* - \bar{\theta}_i^*)^2}{(B-1)},$$

donde θ_i es el i -ésimo parámetro, con $i = 1, 2, 3$, y $\bar{\theta}_i^* = \sum_{b=1}^B \theta_{ib}^* / B$, en concreto, $\bar{\boldsymbol{\theta}}^* = (0.6017, 0.2075, 0.2244)'$. Es habitual que el error estándar asintótico sea inferior al obtenido con el bootstrap, aunque en este caso ambos son muy similares, y la teoría asintótica recomienda el uso de la teoría normal para la estimación de los parámetros. Lo que nos aporta adicionalmente el bootstrap es la posibilidad de estudiar la distribución de los estimadores. En la Figura 4.11, podemos analizar la distribución bootstrap de $\hat{\boldsymbol{\theta}}$, ya que se ha representado el histograma bootstrap y la distribución normal asintótica (línea punteada) de $\hat{\theta}_i$, con $i = 1, 2, 3$. Podemos observar cómo la distribución bootstrap y la distribución asintótica son muy similares en los tres parámetros.

(a) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_1$.*(b) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_2$.*(c) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_3$.*Figura 4.11: *Histograma bootstrap de $\hat{\theta}_i$, con $i = 1, 2, 3$, y distribución normal asintótica.*

4.3.3. Aplicación del procedimiento a datos reales

A continuación consideraremos una serie de datos reales. Estimaremos los parámetros de los modelos CIR y CKLS de un tipo de interés del mercado interbancario del euro, el euríbor (acrónimo de Euro Interbank Offered Rate, tipo europeo de oferta interbancaria). Consideraremos las series diarias del euríbor a 3 meses, 6 meses, 9 meses y 12 meses en el período comprendido entre el 15 de octubre del 2001 y el 30 de diciembre de 2005 (tamaño muestral de 1077), para evitar el período de la crisis económica y el intervencionismo del Banco Central Europeo. En la Figura 4.12 podemos ver la evolución diaria de la serie para los plazos mencionados.

En el Cuadro 4.6 se muestran las estimaciones de los parámetros del modelo CIR,

$$dr_t = (\theta_1 - \theta_2 r_t)dt + \theta_3 \sqrt{r_t} dW_t,$$

para la serie del euríbor, junto con el error estándar (entre paréntesis). Además del filtro de Kalman, también se han estimado los parámetros con el método de Aït-Sahalia. Las estimaciones por ambos métodos son prácticamente iguales, apenas difieren entre los dos métodos de estimación. Podemos ver que los parámetros del drift para los plazos de 3, 6 y 9 meses son muy similares, pero la volatilidad aumenta al incrementar el plazo, como es

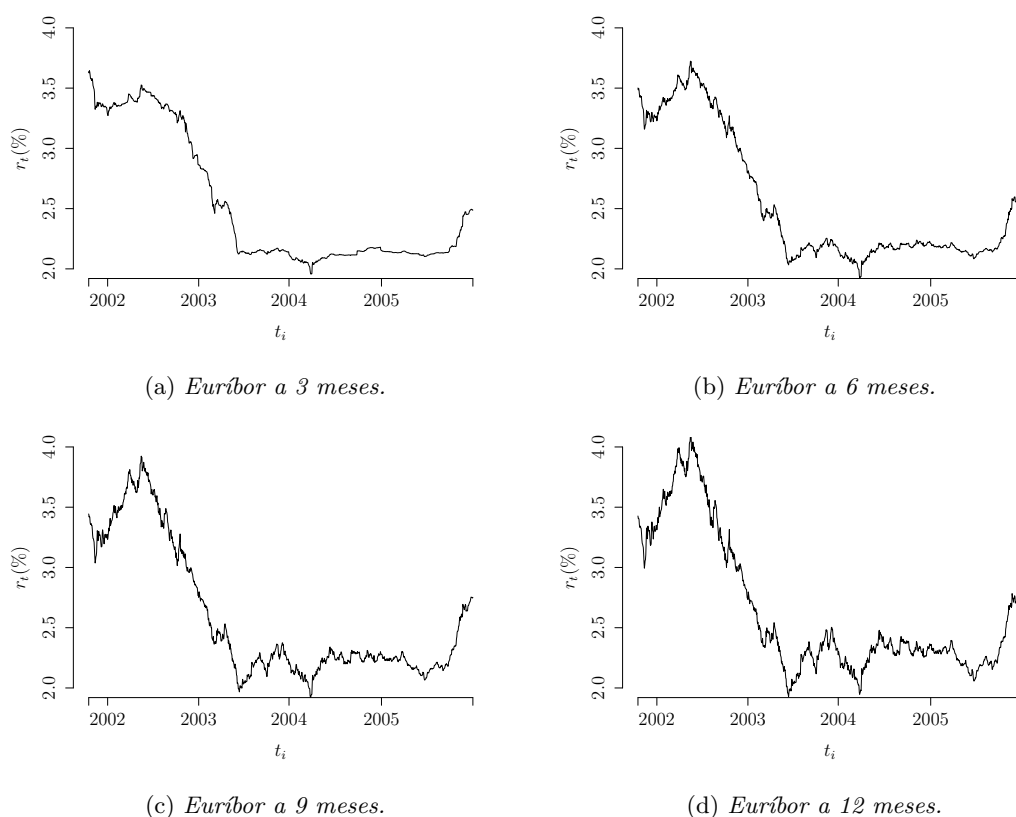


Figura 4.12: Evolución diaria del euríbor en el período comprendido entre el 15 de octubre del 2001 y el 30 de diciembre de 2005. El tamaño muestral para cada serie es de $n = 1077$.

de esperar. El euríbor a 12 meses presenta valores mayores en los parámetros del drift y mayor coeficiente de difusión.

Si analizamos las estimaciones obtenidas para el euríbor a 3 meses, en el Cuadro 4.7 se recogen los errores estándar obtenidos de $B = 1000$ remuestras bootstrap y en la Figura 4.13 tenemos la distribución bootstrap de $\hat{\theta}$ y la distribución normal asintótica (línea punteada) de $\hat{\theta}_i$, con $i = 1, 2, 3$. En el caso de $\hat{\theta}_1$ y $\hat{\theta}_2$, la distribución bootstrap se asemeja bastante a la asintótica, pero para θ_3 difieren. El error estándar bootstrap es mayor que el asintótico, por lo que la distribución asintótica es muy apuntada comparada con la bootstrap, aunque esta última es simétrica con media $\bar{\theta}_3^* = 0.1034$, muy cercana a la estimación del parámetro.

Planteando otro modelo para los datos, en el Cuadro 4.8 se muestran las estimaciones de los parámetros del modelo CKLS,

$$dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 r_t^{\theta_4} dW_t,$$

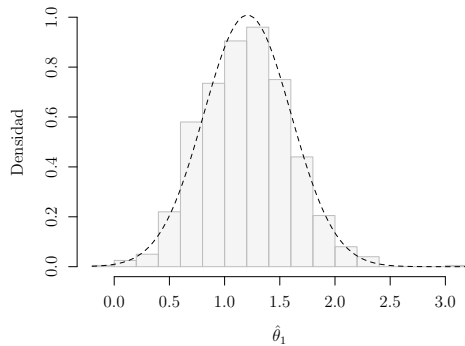
para la serie del euríbor, junto con el error estándar (entre paréntesis). Las estimaciones por ambos métodos son muy similares, aunque difieren ligeramente en el parámetro θ_4 . En Pagan et al. (1995) se muestra que el tamaño del parámetro θ_4 depende del método de estimación utilizado, obteniendo resultados diferentes al utilizar el Método de Máxima Verosimilitud y el Método Generalizado de los Momentos.

Euríbor	Filtro de Kalman			Aït-Sahalia		
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
3 meses	1.2069 (0.3956)	0.5832 (0.1595)	0.1038 (0.0022)	1.2220 (0.3963)	0.5894 (0.1598)	0.1039 (0.0022)
6 meses	1.1559 (0.6087)	0.5331 (0.2435)	0.1556 (0.0034)	1.1810 (0.6097)	0.5430 (0.2440)	0.1557 (0.0034)
9 meses	1.2992 (0.8776)	0.5654 (0.3461)	0.2228 (0.0048)	1.3357 (0.8789)	0.5793 (0.3466)	0.2229 (0.0048)
12 meses	1.4900 (1.0885)	0.6163 (0.4205)	0.2759 (0.0059)	1.5284 (1.0899)	0.6311 (0.4211)	0.2760 (0.0060)

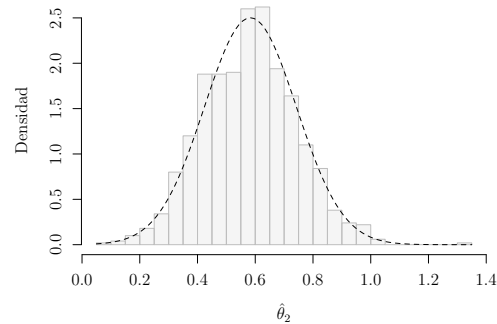
Cuadro 4.6: *Parámetros estimados del modelo CIR para la serie del euríbor utilizando dos métodos de estimación.*

Parámetro	Estimación	Error estándar asintótico	Error estándar bootstrap
θ_1	1.2069	0.3956	0.4042
θ_2	0.5832	0.1595	0.1623
θ_3	0.1038	0.0022	0.0075

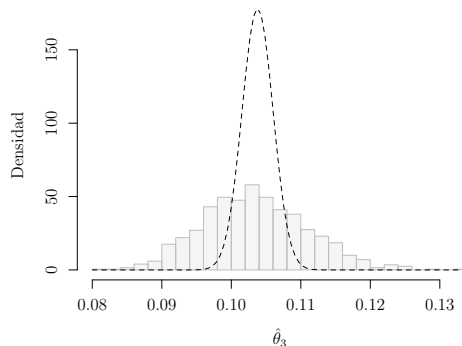
Cuadro 4.7: *Estimación de los parámetros, error estándar asintótico asociado y error estándar bootstrap del modelo CIR para la serie del euríbor a 3 meses.*



(a) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_1$.*



(b) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_2$.*



(c) *Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_3$.*

Figura 4.13: *Histograma bootstrap de $\hat{\theta}_i$, con $i = 1, 2, 3$, y distribución normal asintótica de la serie del euríbor a 3 meses modelizada con un proceso CIR.*

Si comparamos el valor de los parámetros con los obtenidos para el modelo CIR, el parámetro θ_3 difiere en mayor medida que θ_1 y θ_2 , debido a la capacidad de los modelos para captar la dinámica de la serie (notar que en el modelo CIR tenemos que $\theta_4 = 0.5$). En el CKLS, el parámetro θ_4 controla la relación entre el tipo de interés y la volatilidad, de modo que cuando $\theta_4 < 1$ la volatilidad aumenta cuando el tipo de interés disminuye y viceversa, cuando $\theta_4 > 1$ la volatilidad tiende a incrementar a medida que aumentan los tipos de interés. Podemos ver que para la serie del euríbor hay una alta sensibilidad de la volatilidad al nivel de los tipos. A medida que incrementa el plazo, θ_4 se reduce.

Si analizamos las estimaciones obtenidas para el euríbor a 3 meses en el modelo CKLS, en el Cuadro 4.9 se recogen los errores estándar obtenidos de $B = 1\,000$ remuestras bootstrap y en la Figura 4.14 tenemos la distribución bootstrap de $\hat{\theta}$ y la distribución normal asintótica (línea punteada) de $\hat{\theta}_i$, con $i = 1, 2, 3, 4$. En el caso de $\hat{\theta}_1$ y $\hat{\theta}_2$, los parámetros de la función drift, la distribución bootstrap se asemeja bastante a la asintótica, pero para $\hat{\theta}_3$ y $\hat{\theta}_4$ difieren. Esto ya sucedió cuando ajustamos el modelo CIR, donde la distribución asintótica de los parámetros de la función drift coincidía con la distribución bootstrap, pero para el parámetro de la función de difusión, la distribución asintótica y la bootstrap diferían. En el caso del CKLS, el error estándar bootstrap de $\hat{\theta}_3$ y $\hat{\theta}_4$ (los parámetros de la función de difusión) es mayor que el asintótico, por lo que la distribución asintótica es muy apuntada comparada con la bootstrap. Además, la distribución bootstrap de θ_3 presenta una asimetría positiva (ver Figura 4.14c) y la media, $\bar{\theta}_3^* = 0.0343$, se aleja ligeramente de la estimación del parámetro, $\hat{\theta}_3 = 0.0304$. Si analizamos la Figura 4.14d, vemos que la distribución bootstrap de θ_4 tiene una ligera asimetría negativa y la media, $\bar{\theta}_4^* = 1.7392$, difiere ligeramente de la estimación del parámetro, $\hat{\theta}_4 = 1.7884$.

Euríbor	Filtro de Kalman				Ait-Sahalia			
	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
3 meses	1.4998 (0.4494)	0.7070 (0.1965)	0.0304 (0.0032)	1.7884 (0.1129)	1.5371 (0.4551)	0.7235 (0.1998)	0.0280 (0.0029)	1.8768 (0.1119)
6 meses	1.3948 (0.6673)	0.6322 (0.2828)	0.0668 (0.0068)	1.3887 (0.1090)	1.4054 (0.6700)	0.6364 (0.2844)	0.0652 (0.0067)	1.4154 (0.1097)
9 meses	1.5627 (0.9484)	0.6717 (0.3921)	0.1139 (0.0117)	1.1996 (0.1079)	1.5671 (0.9521)	0.6737 (0.3942)	0.1110 (0.0114)	1.2271 (0.1085)
12 meses	1.7616 (1.1622)	0.7238 (0.4667)	0.1613 (0.0169)	1.0515 (0.1078)	1.7782 (1.1673)	0.7303 (0.4697)	0.1563 (0.0164)	1.0838 (0.1085)

Cuadro 4.8: Parámetros estimados del modelo CKLS, $dr_t = (\theta_1 - \theta_2 r_t) dt + \theta_3 r_t^{\theta_4} dW_t$, para la serie del euríbor utilizando dos métodos de estimación.

Parámetro	Estimación	Error estándar asintótico	Error estándar bootstrap
θ_1	1.4998	0.4494	0.4794
θ_2	0.7070	0.1965	0.2105
θ_3	0.0304	0.0032	0.0653
θ_4	1.7884	0.1129	0.4110

Cuadro 4.9: Estimación de los parámetros, error estándar asintótico asociado y error estándar bootstrap del modelo CKLS para la serie del euríbor a 3 meses.

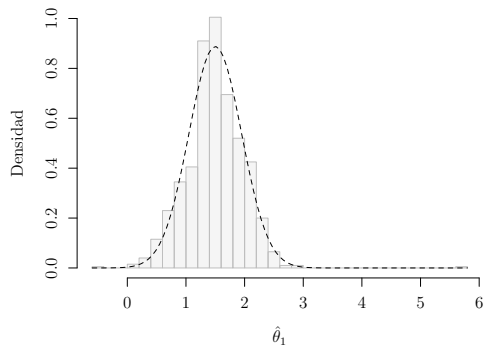
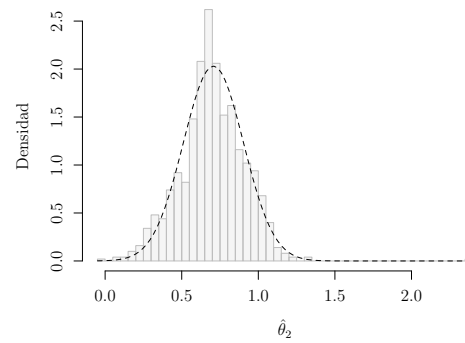
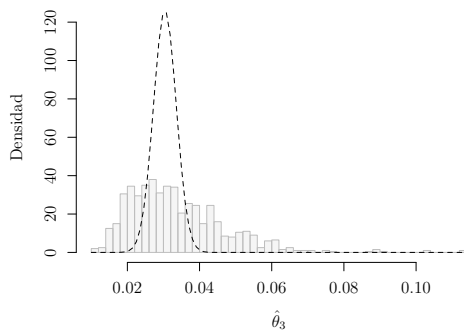
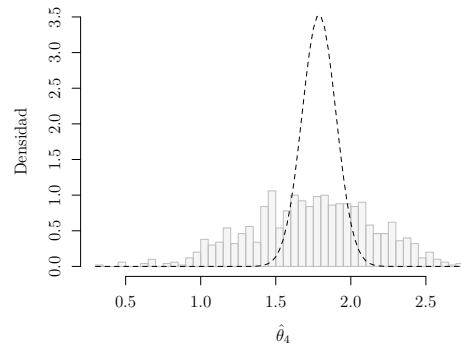
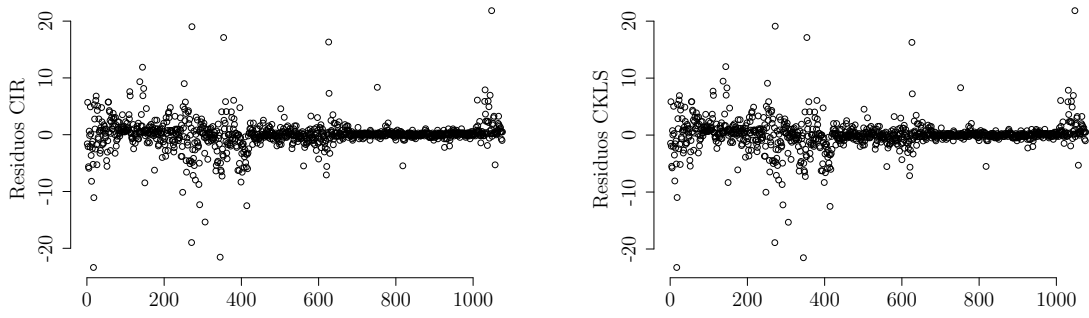
(a) Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_1$.(b) Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_2$.(c) Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_3$.(d) Histograma bootstrap y distribución normal asintótica (línea punteada) de $\hat{\theta}_4$.

Figura 4.14: Histograma bootstrap de $\hat{\theta}_i$, con $i = 1, 2, 3$, y distribución normal asintótica de la serie del euríbor a 3 meses modelizada con un proceso CKLS.

Si nos centramos los histogramas correspondientes a los parámetros de la función de difusión, podemos apreciar que, tanto en la modelización con el proceso CIR ($\hat{\theta}_3$) como con el CKLS ($\hat{\theta}_3$ y $\hat{\theta}_4$), los errores estándar asintóticos difieren considerablemente de los errores estándar bootstrap. Esto puede indicar que los modelos tienen una especificación incorrecta. Por ello, vamos a analizar los residuos de ambos ajustes para detectar presencia de atípicos.

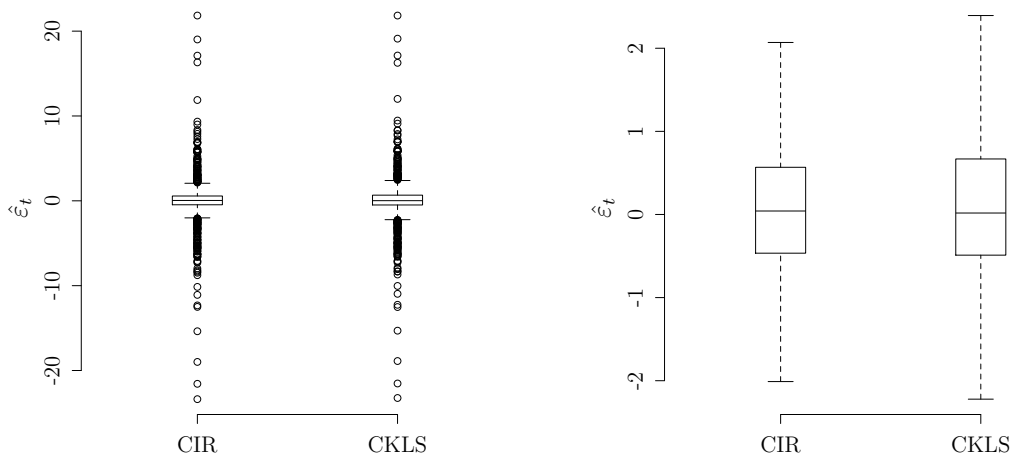
En la Figura 4.15 se han representado los residuos del ajuste del modelo CIR (Figura 4.15a) y del CKLS (Figura 4.15b). En ambos modelos tenemos como hipótesis que los residuos son variables aleatorias iid $N(0, 1)$. A la vista de los gráficos vemos que esta hipótesis se incumple, aparecen muchos outliers (algunos muy extremos) y parece haber una estructura de dependencia. En la Figura 4.16a se muestran los diagramas de caja de los residuos para el ajuste del CKLS y del CIR, donde destaca la presencia de outliers y es importante notar la escala del gráfico. En la Figura 4.16b se muestra el mismo diagrama de caja pero eliminando los outliers.



(a) Residuos del ajuste del modelo CIR.

(b) Residuos del ajuste del modelo CKLS.

Figura 4.15: Residuos de los ajustes del modelo CIR y CKLS para la serie del euríbor a 3 meses.



(a) Diagrama de caja de los residuos de los ajustes del modelo CIR y CKLS para la serie del euríbor a 3 meses.

(b) Diagrama de caja de los residuos de los ajustes del modelo CIR y CKLS para la serie del euríbor a 3 meses, eliminando los outliers.

Figura 4.16: Boxplot de los residuos de los ajustes del modelo CIR y CKLS para la serie del euríbor a 3 meses.

Capítulo 5

Conclusiones

A lo largo de este trabajo se han revisado distintos modelos en tiempo continuo, en el contexto estocástico, destinados a explicar la dinámica del tipo de interés, junto con diferentes métodos de estimación, paramétricos y no paramétricos. Se han aplicado los métodos de estimación a datos reales y simulados y se ha empleado la metodología bootstrap para aproximar la distribución de los estimadores.

En este trabajo nos hemos centrado en los modelos cuya función de difusión es determinista, pero una posible extensión sería considerar modelos de difusión con volatilidad estocástica, es decir,

$$\begin{aligned}dr_t &= m_1(r_t, \theta) dt + \sigma_t \nu_1(r_t, \theta) dW_{1,t}, \\dg(\sigma_t) &= m_2(g(\sigma_t), \vartheta) dt + \nu_2(g(\sigma_t), \vartheta) dW_{2,t},\end{aligned}$$

donde g, m_1, ν_1, m_2 y ν_2 son funciones conocidas y $\Theta = (\theta, \vartheta)' \in \mathbb{R}^d$ es el vector de parámetros desconocido a estimar. El filtro de Kalman, método de estimación paramétrico introducido en este trabajo, permitiría estimar este modelo (González-Manteiga et al., 2017) y, además, es posible adaptar el procedimiento bootstrap presentado para utilizarlo en esta clase de modelos.

El código de **R** desarrollado para implementar los métodos de estimación se muestra en el Apéndice A.

Apéndice A

Código R

En este Apéndice se incluyen las funciones utilizadas en la Sección 4.1 para simular trayectorias del proceso CIR y para la estimación paramétrica de los procesos de difusión.

A.1. Simulación de trayectorias del proceso CIR

El proceso CIR es la solución a la ecuación diferencial estocástica

$$dr_t = (\alpha - \beta r_t) dt + \sigma \sqrt{r_t} dW_t, \quad \text{con } r_0 > 0, \quad (\text{A.1})$$

cuya solución viene dada por

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

Dada la información hasta el instante s , \mathcal{F}_s , el tipo de interés a corto plazo r_t se distribuye como una ji-cuadrado no central,

$$r_t \mid \mathcal{F}_s \sim \chi^2(2cr_t, 4\alpha/\sigma^2, 2u),$$

con $4\alpha/\sigma^2$ grados de libertad y parámetro no central $2u$, donde

$$c = \frac{2\beta}{\sigma^2(1 - e^{-\beta t})}, \quad u = cr_0 e^{-\beta t}, \quad v = cr, \quad q = \frac{2\alpha}{\sigma^2} - 1.$$

En el Código A.1 se muestra la función programada para generar trayectorias del proceso CIR. Los argumentos de la función son los siguientes:

n	Número de observaciones.
t ₀	Momento inicial de la serie.
T	Momento final de la serie, $t_n = T$.
r ₀	Valor inicial del proceso.
alpha	Parámetro α del proceso CIR A.1.
beta	Parámetro β del proceso CIR A.1.

sigma Parámetro σ del proceso CIR A.1.

```

1 sim.CIR <- function(n, t0, T, r0, alpha, beta, sigma){
2   Delta <- (T - t0)/n
3   c <- 2 * beta / ((1 - exp(-beta * Delta)) * sigma^2)
4   df <- 4 * alpha / sigma^2
5   r <- c(r0, rep(0, n - 1))
6   for(i in 2:n){
7     ncp <- 2 * c * r[i-1] * exp(-beta * Delta)
8     r[i] <- rchisq(1, df = df, ncp = ncp)/(2 * c)
9   }
10  r <- ts(r, start = t0, deltat = Delta)
11  return(r)
12 }

```

Código A.1: *Función para generar trayectorias del proceso CIR.*

A.2. Máxima verosimilitud exacta

La función de verosimilitud $\mathcal{L}_n(\theta)$ del proceso discreto es

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n p_{\theta}(\Delta, r_{t_i} | r_{t_{i-1}}) p_{\theta}(r_{t_0}),$$

donde $p_{\theta}(\Delta, r_{t_i} | r_{t_{i-1}})$ denota la función de densidad de transición asociada al modelo de difusión paramétrico, con parámetro desconocido θ . El logaritmo de la función de verosimilitud es

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta) = \sum_{i=1}^n \ell_i(\theta) + \log(p_{\theta}(r_{t_0})) = \sum_{i=1}^n \log p_{\theta}(\Delta, r_{t_i} | r_{t_{i-1}}) + \log(p_{\theta}(r_{t_0})).$$

En el caso del proceso CIR, la forma paramétrica del modelo que genera las observaciones $\{r_{t_i}\}_{i=0}^n$ es conocida, por lo que el estimador de máxima verosimilitud (MLE) del verdadero parámetro es

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

En el Código A.2 se muestra la función de densidad condicional del proceso CIR, cuya distribución es

$$p(t, r | r_0) = c e^{u+v} \left(\frac{u}{v}\right)^{q/2} I_q(2\sqrt{uv}),$$

con

$$c = \frac{2\beta}{\sigma^2(1 - e^{-\beta t})}, \quad u = cr_0 e^{-\beta t}, \quad v = cr, \quad q = \frac{2\alpha}{\sigma^2} - 1.$$

y I_q la función de Bessel modificada de primera especie y orden q . En Iacus (2008) se sugiere utilizar la versión exponencial de la función de Bessel para evitar problemas computacionales. Por ello, se hace uso de la función `bessel.exp`, sugerida por el autor. La función `MVE.CIR` tiene como único argumento el vector de parámetros, $\boldsymbol{\theta} = (\alpha, \beta, \sigma)'$, por lo que para obtener las estimaciones por el método de máxima verosimilitud exacta tendríamos que optimizar esta función (utilizando `optim` del paquete `base stats`, por ejemplo).

```

1  |bessel.exp <- function(x, nu){
2  |  mu <- 4 * nu^2
3  |  A1 <- 1
4  |  A2 <- A1 * (mu - 1) / (1 * (8 * x))
5  |  A3 <- A2 * (mu - 9) / (2 * (8 * x))
6  |  A4 <- A3 * (mu - 25) / (3 * (8 * x))
7  |  A5 <- A4 * (mu - 49) / (4 * (8 * x))
8  |  A6 <- A5 * (mu - 81) / (5 * (8 * x))
9  |  A7 <- A6 * (mu - 121) / (6 * (8 * x))
10 |  1/sqrt(2 * pi * x) * (A1 - A2 + A3 - A4 + A5 - A6 + A7)
11 |}
12 |
13 |den.CIR <- function (r, r0, delta, Theta, log = FALSE) {
14 |  c <- 2 * Theta[2]/((1 - exp(-Theta[2] * delta)) * Theta[3]^2)
15 |  u <- c * r0 * exp(-Theta[2] * delta)
16 |  v <- c * r
17 |  q <- 2 * Theta[1]/Theta[3]^2 - 1
18 |  verosimilitud <- (log(c) - (u + v) + q/2 * log(v/u) +
19 |                  log(bessel.exp(2 * sqrt(u * v), q)) + 2 * sqrt(u * v))
20 |  if (!log) {
21 |    verosimilitud <- exp(verosimilitud)
22 |  }
23 |  verosimilitud
24 |}
25 |
26 |MVE.CIR <- function(Theta) {
27 |  n <- length(r)
28 |  Delta <- deltat(r)
29 |  - sum(den.CIR(r = r[2:n], r0 = r[1:(n-1)], delta = Delta, Theta = Theta, log = TRUE))
30 |}

```

Código A.2: Función de densidad y de verosimilitud del proceso CIR.

A.3. Máxima verosimilitud aproximada

El método de máxima verosimilitud aproximada procede como si las observaciones proviniere de una distribución Gaussiana, de media el drift y desviación estándar la función de difusión, por lo que la densidad de transición es

$$p_{r,\theta}(\Delta, x | r) = \frac{1}{\sqrt{2\pi\Delta\sigma^2(r,\theta)}} \exp\left\{-\frac{1}{2} \frac{(x - r - m(r,\theta)\Delta)^2}{\Delta\sigma^2(r,\theta)}\right\}.$$

En el Código A.3 se muestra la función que devuelve la verosimilitud (en negativo) del proceso CIR, utilizando este método. Los argumentos de la función son:

- r Serie.
- Delta Parámetro Δ .
- Theta Vector de parámetros $\theta = (\alpha, \beta, \sigma)'$.

```

1  |MVA.Euler <- function(r, delta, Theta){
2  |  verosimilitud <- sum(dnorm(r[-1], r[-length(r)] + (Theta[1] - Theta[2] *
3  |                      r[-length(r)]) * delta, sqrt(delta) * Theta[3] *
4  |                      sqrt(r[-length(r)]), TRUE))
5  |  return(-verosimilitud)
6  |}

```

Código A.3: Método de máxima verosimilitud aproximada, con el método de Euler, para el proceso CIR.

A.4. Método de máxima verosimilitud de Ait-Sahalia

En el Código A.4 se muestra la función `AitSahalia`, que devuelve la verosimilitud (en negativo) calculada por el método de máxima verosimilitud de Ait-Sahalia. Los argumentos de la función son:

<code>r</code>	Serie.
<code>theta</code>	Parámetros del proceso.
<code>f.mu*</code>	Función drift del proceso transformado, m_U , donde ‘*’ indica la derivada parcial. Estas funciones tienen dos argumentos: el proceso transformado u_0 y el vector de parámetros θ .
<code>Lamperti</code>	Transformación de Lamperti. Función con dos argumentos: la serie r y el vector de parámetros θ .
<code>f.difusion</code>	Función de difusión del proceso.

```

1 Hermite.AS <- function(Delta, x, dif.delta, mu0, mu1, mu2, mu3, mu4, mu5, mu6){
2   mu02 <- mu0^2
3   mu03 <- mu0^3
4   mu04 <- mu0^4
5   mu05 <- mu0^5
6   mu06 <- mu0^6
7   mu12 <- mu1^2
8   mu13 <- mu1^3
9   mu22 <- mu2^2
10
11  H0 <- function(x) return(1.0)
12  H1 <- function(x) return(-x)
13  H2 <- function(x) return(x^2 - 1)
14  H3 <- function(x) return(-x^3 + 3 * x)
15  H4 <- function(x) return(x^4 - 6 * x^2 + 3)
16  H5 <- function(x) return(-x^5 + 10 * x^3 - 15 * x)
17  H6 <- function(x) return(x^6 - 15 * x^4 + 45 * x^2 - 15)
18
19  eta1 <- -mu0*sqrt(Delta) - (2*mu0*mu1+mu2)*(Delta^1.5)/4 -
20    (4*mu0*mu12 + 4*mu02*mu2 + 6*mu1*mu2 + 4*mu0*mu3 + mu4)*(Delta^2.5)/24
21
22  eta2 <- (mu02+mu1)*Delta/2 + (6*mu02*mu1 + 4*mu12 + 7*mu0*mu2 + 2*mu3) *
23    (Delta^2)/12 + (28*mu02*mu12 + 28*mu02*mu3 + 16*mu13 + 16*mu03*mu2 +
24    88*mu0*mu1*mu2 + 21*mu22 + 32*mu1*mu3 + 16*mu0*mu4 + 3*mu5)*(Delta^3)/96
25
26  eta3 <- -(mu03 + 3*mu0*mu1 + mu2)*(Delta^1.5)/6 -
27    (12*mu03 * mu1 + 28*mu0*mu12 + 22*mu02*mu2 + 24*mu1*mu2 + 14*mu0*mu3 + 3*mu4)*
28    (Delta^2.5)/48
29
30  eta4 <- (mu04 + 6*mu02 * mu1 + 3*mu12 + 4*mu0*mu2 + mu3)*(Delta^2)/24 +
31    (20*mu04*mu1 + 50*mu03*mu2 + 100*mu02*mu12 + 50*mu02*mu3 + 23*mu0*mu4 +
32    180*mu0*mu1*mu2 + 40*mu13 + 34*mu22 + 52*mu1*mu3 + 4*mu5)*(Delta^3.0)/240
33
34  eta5 <- -(mu05 + 10*mu03*mu1 + 15*mu0*mu12 + 10*mu02*mu2 + 10*mu1*mu2 + 5*mu0*mu3 +
35    mu4) * (Delta^2.5)/120
36
37  eta6 <- (mu06 + 15*mu04*mu1 + 15*mu13 + 20*mu03*mu2 + 15*mu0*mu3
38    + 45*mu02*mu12 + 10*mu22 + 15*mu02*mu3 + 60*mu0*mu1*mu2
39    + 6*mu0*mu4 + mu5)*(Delta^3)/720
40
41  valor <- dnorm(x, 0, 1, FALSE)*(1 + eta1*H1(x) + eta2*H2(x) + eta3*H3(x) +
42    eta4*H4(x) + eta5*H5(x) + eta6*H6(x)) / dif.delta
43

```

```

44   return(valor)
45 }
46
47 AitSahalia <- function(r, theta, f.mu0, f.mu1, f.mu2, f.mu3, f.mu4, f.mu5, f.mu6,
48                       Lamperti, f.difusion){
49   Delta <- deltat(r)
50   n <- length(r)
51   y0 <- Lamperti(r[-n], theta)
52   psi <- Lamperti(r[-1], theta)
53   dif.delta <- f.difusion(r[-1], theta) * sqrt(Delta)
54   mu0 <- f.mu0(y0, theta)
55   mu1 <- f.mu1(y0, theta)
56   mu2 <- f.mu2(y0, theta)
57   mu3 <- f.mu3(y0, theta)
58   mu4 <- f.mu4(y0, theta)
59   mu5 <- f.mu5(y0, theta)
60   mu6 <- f.mu6(y0, theta)
61   verosimilitud <- -sum(log(Hermite.AS(Delta, (psi-y0)/sqrt(Delta),
62                                     dif.delta, mu0, mu1, mu2, mu3, mu4, mu5, mu6)))
63   return(verosimilitud)
64 }

```

Código A.4: Método de máxima verosimilitud de Ait-Sahalia.

A.5. Método Generalizado de los Momentos

En el Código A.5 se muestra el Método Generalizado de los Momentos, la función GMM, que hace uso de la función f . Sea θ el vector de parámetros con los elementos α, β, σ^2 y γ , dado $\varepsilon_{t+1} = r_{t+1} - r_t - \alpha - \beta r_t$, denotábamos por $f_t(\theta)$ el vector

$$f_t(\theta) = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} r_t \\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma} \\ (\varepsilon_{t+1}^2 - \sigma^2 r_t^{2\gamma}) r_t \end{bmatrix}.$$

En el Código A.5, la función f es la correspondiente al proceso CIR. Los argumentos de la función del Método Generalizado de los Momentos son los siguientes:

<code>r</code>	Serie.
<code>f</code>	Función $f_t(\theta)$, con cuatro argumentos: r_{t_i} , $r_{t_{i+1}}$, vector de parámetros θ y parámetro Δ .
<code>param.inicial</code>	Vector de parámetros inicial.
<code>max.i</code>	Número máximo de iteraciones del algoritmo. Por defecto 50.

```

1 f <- function(x, y, Theta, delta){
2   c.media <- Theta[1] / Theta[2] + (y - Theta[1] / Theta[2]) * exp(-Theta[2] * delta)
3   c.var <- ((y * Theta[3]^2 * (exp(-Theta[2] * delta) - exp(-2 * Theta[2] * delta)) /
4             Theta[2] + Theta[1] * Theta[3]^2 * (1 - exp(-2 * Theta[2] * delta)) /
5             (2 * Theta[2]^2)))
6   cbind(x - c.media, y * (x - c.media), c.var - (x - c.media)^2,
7         y * (c.var - (x - c.media)^2))
8 }
9

```

```

10 GMM <- function (r, f, param.inicial, max.i = 50) {
11   tol <- .Machine$double.eps
12   Delta <- deltat(r)
13   n <- length(r)
14   g <- function(theta) apply(f(r[2:n], r[1:(n - 1)]), theta, Delta), 2, mean)
15   J <- function(theta) sum(g(theta)^2)
16   S <- function(j, theta) {
17     ((t(f(r[(j + 2):n], r[(j + 1):(n - 1)]), theta, Delta)) **%
18     f(r[2:(n - j)], r[1:(n - j - 1)]), theta, Delta))/n)
19   }
20   w <- 1 - (1:(n - 2))/((n - 2) + 1)
21   thetal <- optim(param.inicial, J, method = "BFGS")$par
22   seguir <- TRUE
23   iter <- 0
24   while (seguir) {
25     iter <- iter + 1
26     S.est <- S(0, thetal)
27     for (i in 1:(n - 2)) S.est = S.est + w[i] * (S(i, thetal) + t(S(i, thetal)))
28     W <- solve(S.est)
29     J1 <- function(theta) g(theta) **% W **% g(theta)
30     fit <- optim(thetal, J1, method = "BFGS", hessian = TRUE)
31     theta2 <- fit$par
32     valor <- fit$value
33     hessian <- fit$hessian
34     if (sum(abs(thetal - theta2)) < tol || valor < tol || iter > max.i) seguir <- FALSE
35     thetal <- theta2
36   }
37   list(theta = thetal, valor = valor, hessian = hessian)
38 }

```

Código A.5: *Método Generalizado de los Momentos.*

A.6. Filtro de Kalman

En el Código A.6 se muestra el algoritmo del Filtro de Kalman. El código se ha programado en base al de la función `Kfilter1` del paquete `astsa`, adaptándola para su uso con los procesos de difusión, con la forma paramétrica general del modelo CKLS,

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

Los argumentos de la función son los siguientes:

- y Serie transformada: $y_{t_i} = r_{t_{i+1}} - r_{t_i}$.
- r*t*i Serie original, r_{t_i} .
- delta Parámetro Δ .
- alpha Parámetro α del proceso.
- beta Parámetro β del proceso.
- sigma Parámetro σ del proceso.
- gamma Parámetro γ del proceso.
- mu0 Parámetro x_0^0 del estado inicial.
- Sigma0 Parámetro P_0^0 del estado inicial.


```

1 FiltroKalman <- function (y, rti, delta, alpha, beta, sigma, gamma, mu0, Sigma0) {
2   A <- 1
3   Phi <- 0
4   R <- 0
5   n <- length(y)
6   Q <- t(sigma) %*% sigma
7   ut <- array(rbind(rep(1, n), rti[-length(rti)]), dim = c(2, 1, n))
8   Upsilon <- matrix(c(alpha, -beta), ncol = 2)
9   y <- as.matrix(y)
10  qdim <- ncol(y)
11  pdim <- 1
12  xp <- array(NA, dim = c(pdim, 1, n))
13  Pp <- array(NA, dim = c(pdim, pdim, n))
14  xf <- array(NA, dim = c(pdim, 1, n))
15  Pf <- array(NA, dim = c(pdim, pdim, n))
16  K <- array(NA, dim = c(pdim, qdim, n))
17  innov <- array(NA, dim = c(qdim, 1, n))
18  sig <- array(NA, dim = c(qdim, qdim, n))
19
20  x00 <- as.matrix(mu0, nrow = pdim, ncol = 1)
21  P00 <- as.matrix(Sigma0, nrow = pdim, ncol = pdim)
22  xp[, , 1] <- Phi %*% x00 + Upsilon %*% ut[, , 1, drop = FALSE]
23  Pp[, , 1] <- Phi %*% P00 %*% t(Phi) + Q * (rti[1]^(2 * gamma)) / delta
24  sigtemp <- A %*% Pp[, , 1] %*% t(A) + R
25  sig[, , 1] <- (t(sigtemp) + sigtemp)/2
26  siginv <- 1 / (sig[, , 1])
27  K[, , 1] <- Pp[, , 1] %*% t(A) %*% siginv
28  Kt <- as.matrix(K[, , 1], nrow = qdim, ncol = pdim)
29  innov[, , 1] <- y[1, ] - A %*% xp[, , 1]
30  xf[, , 1] <- xp[, , 1] + Kt %*% innov[, , 1]
31  Pf[, , 1] <- Pp[, , 1] - Kt %*% A %*% Pp[, , 1]
32  sigmat <- as.matrix(sig[, , 1], nrow = qdim, ncol = qdim)
33  like <- log(det(sigmat)) + t(innov[, , 1]) %*% siginv %*% innov[, , 1]
34  for (i in 2:n) {
35    xp[, , i] <- Phi %*% xf[, , i - 1] + Upsilon %*% ut[, , i, drop = FALSE]
36    Pp[, , i] <- Phi %*% Pf[, , i - 1] %*% t(Phi) + Q * (rti[i]^(2*gamma)) / delta
37    sigtemp <- A %*% Pp[, , i] %*% t(A) + R
38    sig[, , i] <- (t(sigtemp) + sigtemp)/2
39    siginv <- 1 / (sig[, , i])
40    K[, , i] <- Pp[, , i] %*% t(A) %*% siginv
41    Kt <- as.matrix(K[, , i], nrow = qdim, ncol = pdim)
42    innov[, , i] <- y[i, ] - A %*% xp[, , i]
43    xf[, , i] <- xp[, , i] + Kt %*% innov[, , i]
44    Pf[, , i] <- Pp[, , i] - Kt %*% A %*% Pp[, , i]
45    sigmat <- as.matrix(sig[, , i], nrow = qdim, ncol = qdim)
46    like <- like + log(det(sigmat)) + t(innov[, , i]) %*% siginv %*% innov[, , i]
47  }
48  like <- 0.5 * like
49  list(xp = xp, Pp = Pp, xf = xf, Pf = Pf, like = like, innov = innov, sig = sig, Kn = K)
50 }

```

Código A.6: *Filtro de Kalman.*

Bibliografía

- [1] Ahn, D. H. y Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12(4):721-762.
- [2] Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64:527-560.
- [3] Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9(2):385-426.
- [4] Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal Finance*, 54(4):1361-1395.
- [5] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closedform approximation approach. *Econometrica*, 70(1):223-262.
- [6] Bandi, F. y Phillips, P. (2003). Fully nonparametric estimation of scalar diffusion models, *Econometrica*, 71, 241-283.
- [7] Beckers, S. (1980). The Constant Elasticity of Variance Model and Its Implications for Option Pricing. *Journal of Finance*, 35(3):661-73.
- [8] Black, F. y Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637-54.
- [9] Boente, G. y Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis*, 25(1):90-99.
- [10] Brennan, M. y Schwartz, E. (1979). A continuous time approach to the pricing of bonds. *Journal of Banking and Finance*, 3(2):133-155.
- [11] 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.
- [12] Courtadon, G. (1982a). The pricing of options on default-free bonds. *Journal of Financial and Quantitative Analysis* 17(1): 75-100.
- [13] Cox, J. C. (1975). Notes on option pricing I: constant elasticity of variance difussions. Working paper, Stanford University.
- [14] Cox, J., Ingersoll, J. E., y Ross, S. A. (1980). An analysis of variable rate loan contracts. *Journal of Finance*, 35(2):389-403.

- [15] Cox, J., Ingersoll, J. E., y Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385-407.
- [16] Dothan, L. U. (1978). On the term structure of interest rates. *Journal of Financial Economics*, (6):59-69.
- [17] Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54(1):173-182.
- [18] Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied of Probability*, 30:790-804.
- [19] Genon-Catalot, V., Laredo, C. y Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelets methods, *Scand. J. Stat.*, 19, 317-335.
- [20] González-Manteiga, W., Zubelli, J., Monsalve-Cobis, Abelardo E., Febrero-Bande, M. (2017). Goodness of Fit Test for Stochastic Volatility Models. *From Statistics to Mathematical Finance. Festschrift in Honour of Winfried Stute*. Springer.
- [21] Hoffmann, M. (1999). Adaptive estimation in diffusion processes, *Stochastic Process. Appl.*, 79, 135-163.
- [22] Iacus, S. M. (2008). *Simulation and Inference for stochastic Differential Equations*. Springer.
- [23] Jiang, G. J. (1998a). A generalized one-factor term structure model and pricing of interest rate derivative securities. Research report 97a35, University of Gromingen. Research Institute SOM.
- [24] Jiang, G. J. (1998b). Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysis* 33 (4): 465-497.
- [25] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Trans ASME J. Basic Eng.*, 82, 35-45.
- [26] Kloeden, P., Platen, E. y Schurz, H. (2000). *Numerical Solution of SDE through Computer Experiments*. Springer, Berlin.
- [27] Merton, R. C. (1973). The theory of rational option pricing. *Bell Journal of Economics*, 4(1):141-183.
- [28] 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.
- [29] Ornstein, L. S. y Uhlenbeck, G. E. (1930). On the Theory of the Brownian Motion. *Physical Review* 36, no. 5: 823.
- [30] Pagan, A. R., Hall, A. D. y Martin, V. (1995). Modeling the term structure. Technical Report, Australian National University.
- [31] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*. 33. 1065-1076.
- [32] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Annals of Mathematics and Statistics*. 27. 832-837.

- [33] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley, New York.
- [34] Shumway, R. H. y Stoffer, D. S. (2010): *Time Series Analysis and Its Applications, whit R examples*, 3rd edition. Springer.
- [35] 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.
- [36] Steele, J. M. (2001). *Stochastic calculus and financial applications*. Springer.
- [37] Stoffer, D.S. y Wall, K.D. (1991). Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Am. Stat. Assoc.*, 86, 1024-1033.
- [38] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177-188.
- [39] Watanabe, N. (1985). Note on the Kalman Filter with estimated parameters. *Journal of Time Series Analysis*, 6, 269-278.