INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Alejandra María
López Pérez

PhD Thesis

Advances in Statistical Inference for Econometric Diffusion Models

Santiago de Compostela, 2022

Doctoral Programme in Statistics and Operations Research

TESE DE DOUTORAMENTO

ADVANCES IN STATISTICAL INFERENCE FOR ECONOMETRIC DIFFUSION MODELS

Alejandra María López Pérez

ESCOLA DE DOUTORAMENTO INTERNACIONAL
DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO
EN ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

SANTIAGO DE COMPOSTELA
2022

DECLARACIÓN DA AUTORA DA TESE

Dna. **Alejandra María López Pérez**
Título da tese: **Advances in statistical inference for econometric diffusion models**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide coa versión enviada en formato electrónico.

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela, a 22 de xullo de 2022.**

Sinatura electrónica

AUTORIZACIÓN DOS DIRECTORES/TITOR DA TESE

Advances in statistical inference for econometric diffusion models

D. Wenceslao González Manteiga
D. Manuel Febrero Bande

INFORMAN:

Que a presente tese se corresponde co traballo realizado por Dna. Alejandra María López Pérez, baixo a nosa dirección/titorización, e autorizamos a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directores/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declaran tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de Monográfica con reproducción de publicacións, nos que a participación da doutoranda foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

En Santiago de Compostela,
a 22 de xullo de 2022.

Asdo.: Wenceslao González Manteiga

En Santiago de Compostela,
a 22 de xullo de 2022.

Asdo.: Manuel Febrero Bande

Acknowledgments

I would like to thank my supervisors Prof. Wenceslao González Manteiga and Prof. Manuel Febrero Bande for all their support and advice during this PhD. I am also grateful to professors Mario Francisco Fernández, Juan Carlos Pardo Fernández, Beatriz Pateiro López and Rosa María Crujeiras Casais for their feedback in each committee. I would also like to thank Prof. Jorge Passamani Zubelli for inviting me at the *Instituto de Matemática Pura e Aplicada* (IMPA) and his hospitality during my stay. Lastly, my gratitude extends to Dr. Javier Álvarez Liébana, for the discussion and collaboration in Chapter 8, and Prof. María Dolores Ruiz Medina for her suggestions on that chapter.

Alejandra López Pérez
Santiago de Compostela, xullo de 2022.

Funding

This work has been supported by the PhD grant BES-2017-080581, as well as by Spanish Ministerio de Economía y Competitividad, through grant numbers MTM2016-76969-P and PID2020-116587GB-I00, which includes support from the European Regional Development Fund (ERDF). Supercomputing Center of Galicia (CESGA) is acknowledged for providing the computational resources that allowed to run most of the simulations.

Contents

Chapter 1

Introduction

Diffusion models described by stochastic differential equations are a natural extension of deterministic models described by ordinary differential equations. Diffusion processes have been used in different fields to model dynamic phenomena in continuous-time, such as physics (Van Kampen, 1992), neuroscience (Mattia and Del Giudice, 2002), finance (Aït-Sahalia and Jacod, 2014), or population dynamics (Brites and Braumann, 2020), demanding the development of new statistical methods. The use of a continuous-time framework was spread in the financial literature with the seminal work of Merton (1975), and, thenceforth, interest rates models have been treated in continuous-time. Interest rate time series are a core element in finance, as their evolution affects a wide range of economic players and variables, from policy makers to the valuation of financial securities.

The goal of this dissertation is the development of new goodness-of-fit tests for diffusion models, addressing the estimation of the processes, as it can hinder the tests procedures. This introductory chapter is devoted to provide some background on diffusion models and its use in the econometric field. The main objectives of this dissertation and its structure are described in Section 1.2.

Contents

1.1 Continuous-time series

Due to their analytical tractability, continuous-time models have become a centerpiece in asset pricing theory, as they can lead to closed form solutions for prices or portfolio weights. This importance is also present from the empirical perspective, as many price dynamics can be accommodated in this framework. However, the continuous-time nature has proved to be difficult, challenging the identification and estimation of the diffusion models. Diffusion processes are described by stochastic differential equations, which are briefly introduced in the next section, though more rigorous definition is provided in Chapter 2.

1.1.1 Stochastic differential equations

Given a time-homogeneous stochastic differential equation defined in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geqslant 0}, \mathbb{P})$,

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t, \qquad t \in [0, T],$$

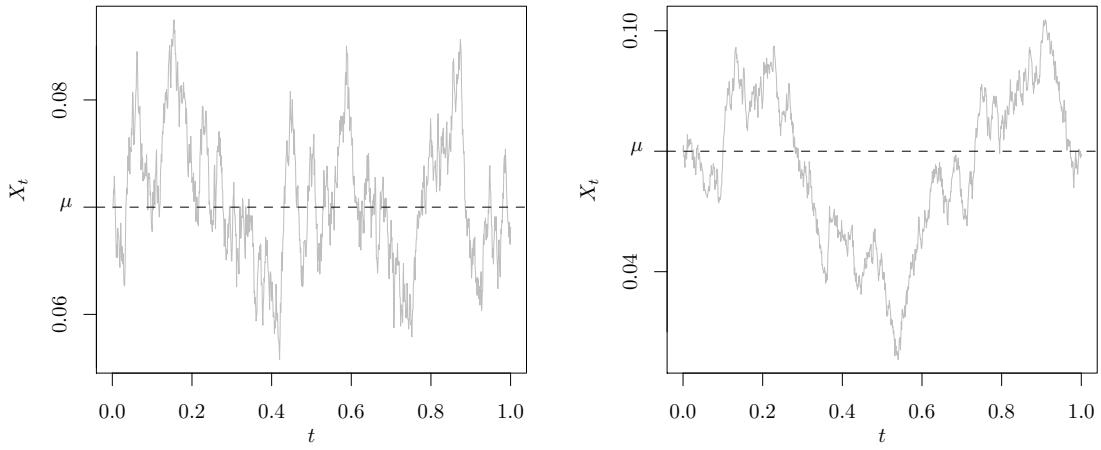
with $X_0 = x_0 \in \mathbb{R}$, W_t a $\{\mathcal{F}\}_{t\geqslant 0}$-adapted Wiener process and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, with $d \in \mathbb{Z}^+$, an unknown parameter vector, X_t is an Itô process. The function $m(\cdot)$ is the deterministic trend, describing the changes in X_t on time t, and the stochastic perturbations are introduced by the function $\sigma(\cdot)$.

The continuous-time paradigm has proved to be an specially useful tool, however, the continuous-time nature of the process does complicate the estimation of the different components of the model because available data are sampled in discrete time. Different discretization schemes have been proposed (Maruyama, 1955; Milshtein, 1979; Kloeden and Platen, 1992a), but discretization implies that the parameter estimates are subject to discretization bias, regardless of the estimation method used. This, together with sample bias, complicates the estimation of diffusion models, deteriorating the calibration of goodness-of-fit test and the bootstrap resampling techniques.

1.1.2 Econometric diffusion models

The finite sample bias is specially acute when the process is highly persistent, as displayed by time series of interest rates, altering the valuation of derivatives since short-term interest rate models are used to price these instruments.

Economic theory finds plausible the mean reversion of interest rates time series, that is, that the time series reverts to a long run mean level over time. Many interest rate models



(a) High mean reverting (low persistence) Ornstein-Uhlenbeck process.

(b) Low mean reverting (high persistence) Ornstein-Uhlenbeck process.

Figure 1.1: Simulated paths of Ornstein-Uhlenbeck processes with different degrees of mean-reversion or persistence.

include mean reversion in their parametric specifications. A common drift function is given by $m(X_t, \boldsymbol{\theta}) = \kappa(\mu - X_t)$, with κ and μ non-negative constants. X_t is the current interest rate level, μ is the long run mean and κ is the speed of mean reversion, the speed at which the interest rate reverts towards μ. Mean reversion implies that when the interest rate level is above the mean, $X_t > \mu$, the drift is negative, driving the rate closer to μ; if the rate level is below the long run mean, $X_t < \mu$, the drift is positive, pushing the rate to the level μ. As an example, Figure 1.1 show two sample paths of the Ornstein-Uhlenbeck process with different levels of mean reversion. Figure 1.1a is a high mean reverting process, that is, shows a low persistence, where the path evolves close to the long run mean μ. On the other hand, Figure 1.1b is a persistent process, with low mean reversion, reverting towards the mean but at a lower speed.

Empirical research regarding interest rates mean reversion is inconclusive, as some long term interest rates can deviate from this assumption and the unit root hypothesis is not always rejected. This can be due to the fact that interest rates exhibit very strong persistence, leaning almost to unit root time series, which complicates even further the estimation of the model.

1.1.3 Methodology

When modeling interest rates dynamics through diffusion processes, their parametric specification allows us to provide an interpretation to the data generating process mechanism, which is possible if the model has a correct specification. It is necessary to validate the adequacy

of the diffusion model specification to the data in order to avoid misspecification errors. Let $\mathcal{M}$ be the unknown feature behind the data generating mechanism and $\mathcal{M}_\theta$ be a given specification, with unknown parameter θ. The goodness-of-fit test or specification test consists on testing the null hypothesis $\mathcal{H}_0 : \mathcal{M} = \mathcal{M}_\theta$. In this dissertation, new goodness-of-fit tests are proposed for the drift and volatility functions of the diffusion models, carrying out simulation studies to verify their finite sample performance.

The computational implementation is developed with the R software (R Core Team, 2014).

1.2 Objectives of the dissertation

The goal of this thesis is the development of new goodness-of-fit test for continuous-time diffusion models, considering diffusion processes with deterministic and stochastic volatility and Itô diffusions as functional time series. The proposal of novel goodness-of-fit tests allows to study the adequacy of the diffusion model specification towards the dynamics of interest rates. Notwithstanding the importance of goodness-of-fit tools for continuous-time models, latent factors and a continuous-time setting with observations occurring at discrete time points challenge the estimation of the models. Therefore, the estimation problem should be addressed, as it hinders the goodness-of-fit procedures. We attempt to discuss the intricacies of different estimation implementations prior to the methodological contribution of the test procedures.

1.2.1 Manuscript distribution

This thesis document is structured in two main blocks: the first covers the estimation of diffusion models (Chapters 3–5) and the second addresses goodness-of-fit tests for diffusion processes (Chapters 6–8). A theoretical background is provided in Chapter 2, while Chapter 9 summarizes the main conclusions of the dissertation. Chapters were designed to be self-contained to the extent possible.

Chapter 2: Diffusion processes and the term structure of interest rates

Chapter 2 introduces the fundamental concepts on which diffusion models are built. A discussion of the models used in the econometric field to explain the dynamics of interest rates is provided, along the main parametric models used in the literature. The formulation of stochastic differential equations relies on stochastic calculus, briefly covered in this chapter

to rigorously define the continuous-time diffusion process used throughout this dissertation. Some notions on economic theory are introduced, though in this dissertation the aim is the modeling of the empirical behavior of interest rates from a statistical perspective.

Chapter 3: Parametric estimation of diffusion models

Chapter 3 provides a review of parametric estimation methods for diffusion models, with deterministic volatility function, that is, one-factor models. Several estimation methods are reviewed and implemented to assess the role of persistence and study the impact of the sampling frequency. A simulation study is conducted under different settings and Euribor time series are analyzed. The implemented methods were collected in the R package `estsde`, included in Appendix B.

This chapter is mainly based on López-Pérez et al. (2021).

Chapter 4: Parametric estimation of diffusion models with stochastic volatility

In Chapter 4, the review of estimation methods of Chapter 3 is extended to two-factor diffusion models, where the volatility function is stochastic. The literature has moved towards more flexible specification in order to capture the dynamics of interest rates, but the latent variable hinders the estimation of the parameter vector. A Monte Carlo simulation study of different estimation procedures is provided, along a discussion on the advantages of the procedures. The Euribor interest rate time series of Chapter 3 is used to estimate a two-factor model, extending the estimated model with deterministic diffusion function of the previous chapter to a stochastic volatility function.

This chapter is mainly based on López-Pérez et al. (2022b).

Chapter 5: Nonparametric estimation of diffusion models

Chapter 5 discusses nonparametric estimation of diffusion models, including the drift, diffusion and stationary density function. The focus is on smoothing techniques and a comparison of several bandwidths is carried out by means of a Monte Carlo study. A bandwidth selection criterion developed for a random setting with correlated errors is considered for the nonparametric estimation of the drift and diffusion function in order to lessen the impact of persistence. As real data application, treasury securities are analyzed.

Chapter 6: Goodness-of-fit test for diffusion models

In Chapter 6 a survey is provided to collect some developments on goodness-of-fit test for diffusion models, implementing the methods to illustrate their finite sample behavior. A new goodness-of-fit test for the parametric specification of diffusion models is proposed, based on the ideas of distance correlation, and its performance is compared to the reviewed procedures, analyzing the effect of the curse of dimensionality. The treasury securities studied in Chapter 5 are considered as real data examples.

This chapter is mainly based on López-Pérez et al. (2022a).

Chapter 7: Goodness-of-fit test for diffusion models with stochastic volatility

A goodness-of-fit test for diffusion models with stochastic volatility is proposed in Chapter 7, to test the parametric specification of the drift and diffusion functions. The test is based on a marked empirical process of the residuals and the test statistics are constructed using a continuous functional (Kolmogorov-Smirnov and Cramér-von Mises) over the empirical process. As discussed in Chapter 4, the use of the Kalman filter algorithm allows the implementation of a bootstrap resampling procedure for the calibration of the tests. The Euribor series were analyzed in Chapters 3 and 4 regarding the estimation of diffusion models with deterministic and stochastic volatility, these time series are analyzed again to test the stochastic volatility specification.

This chapter is mainly based on López-Pérez et al. (2022b).

Chapter 8: Specification test for diffusion models as functional time series

In Chapter 8 a novel specification test for diffusion models as functional time series is proposed. In this chapter, the diffusion model is used in a high-frequency setting, as these data can be collected as a sequence of curves over time (e.g., intra-day prices). In the previous chapters, the whole observation sequence of the diffusion process was regarded as one sample path, ignoring the pattern of the process across days. Working in a functional data setting, we take advantage of the information retained in the shape of the curve, treating the daily curves as a statistical object, instead of a collection of individual observations. The contribution of this chapter is twofold: a goodness-of-fit test for autoregressive Hilbertian (ARH) time series and a specification test for diffusion models characterized as ARH models. Both proposals are illustrated by means of a Monte Carlo simulation, regarding power and size,

and application to intra-day currency exchange rates is analyzed.

This chapter is mainly based on Álvarez-Liébana et al. (2022).

Chapter 9: Conclusions and discussion

Lastly, Chapter 9 collects the conclusions drawn from the dissertation and provides some insights on open research topics.

Appendices

Appendix A and C contain supplementary material for Chapters 3 and 8, respectively. Appendix B collects the documentation of the R package `estsde`, where the main computational implementations of the methods reviewed in the dissertation were included.

Chapter 2

Diffusion processes and the term structure of interest rates

The choice of a model to explain the dynamics of short-term rates is crucial, as it impacts financial data at many levels (e.g., bond pricing, derivatives of interest rates or hedging). The most adequate framework is using continuous-time models and, among them, Itô diffusion processes have been used to explain the evolution of interest rates. Due to the importance of interest rates, several diffusion models have been proposed to describe them. In this chapter, the formulation of diffusion models described by stochastic differential equations is introduced, for which Itô calculus is needed and, therefore, briefly covered.

Contents

2.1 Modeling the term structure of interest rates

Interest rate time series are one of the core elements in Finance. Its evolution affects policy makers, investors, banks and the valuations of financial securities. The term structure of interest rates measures the relationship between interest rates (or bond yields) and a range of

terms to maturity. The understanding of the term structure, which depicts the anticipation of the market of future events, helps to predict how the yield curve will be affected by changes in underlying assets (see Cox et al., 1985). As a consequence, several time series models have been proposed to model, estimate and forecast the term structure of interest rate.

Among the different frameworks, continuous-time modeling has made the most impact when dealing with interest rates since Merton (1975) study of a stochastic growth model. Thenceforth, equilibrium models of term structure have been treated in continuous-time, where the stochastic evolution of the interest rate is captured by an Itô diffusion process. The popularity of the continuous-time Itô process is also supported by its desirable properties. Under quite mild regularity conditions, nonlinear functions of the Itô diffusion are also Itô processes. As the prices of derivative securities are functions of the underlying asset prices (such as bond or treasury bill prices as a function of spot interest rates), through stochastic calculus is possible to derive a partial differential equation to price derivative securities, under a general equilibrium approach or no-arbitrage arguments. Based on this partial differential equation, the derivative prices are routinely solved, with certain boundary and initial conditions. This makes Itô diffusion extremely well suited for asset pricing derivative evaluation problems, making more tractable and convenient the use of a continuous-time framework when modeling asset prices and interest or exchange rates.

In the next section, a brief introduction to stochastic calculus is provided to subsequently define the continuous-time diffusion process, subject of study along this thesis.

2.2 Stochastic calculus

A stochastic process $X = \{X_t \colon 0 \leqslant t < \infty\}$ can be described as a collection of random variables on a measurable space $(\Omega, \mathcal{F})$, the *sample space*, taking values in another measurable space $(S, \mathcal{S})$, the *state space*. In this chapter, the state space considered is the d-dimensional Euclidean space, that is, $S = \mathbb{R}^d$ and $\mathcal{S} = \mathcal{B}(\mathbb{R}^d)$, equipped with the σ-algebra of Borel sets, where $\mathcal{B}(\mathbb{R}^d)$ denotes the smallest σ-algebra containing all open sets of the Euclidean space $\mathbb{R}^d$. Regarding the sample space, we will work in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \{\omega\}$ is a space of elementary events, the filtration $\mathcal{F}$ is a σ-algebra of subsets of Ω and $\mathbb{P}$ is a probability measure defined on the sets from $\mathcal{F}$, where $\mathbb{P}$ is a measurable mapping $\mathbb{P} \colon \mathcal{F} \to [0, 1]$, countably additive and $\mathbb{P}(\Omega) = 1$.

To be able to formulate diffusion models described by stochastic differential equations precisely, we need to use Itô calculus. Therefore, basic concepts from stochastic calculus will be briefly introduced (see Billingsley, 1995; Karatzas and Shreve, 1998; Steele, 2001;

Bass, 2011; Durrett, 2012; among others). As the formulation is through Brownian motion, we will start introducing its definition, continuing with the requirements needed to define a stochastic integral, leading to the definition of the integral equation and its derivative form.

2.2.1 Brownian motion and the Itô integral

The core of continuous time financial modeling uses the Brownian motion, or Wiener process, to model the dynamics of asset prices. The connection between Brownian motion and asset prices has been of fundamental importance since the works of Merton (1973) and Black and Scholes (1973), which led to propose a wide class of models for hedging and pricing. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, this sample space is equipped with a filtration, that is, a nondecreasing family $\{\mathcal{F}_t\}_{t\geqslant 0}$ of sub-σ-algebras of $\mathcal{F}$.

Definition 1 (Filtration)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{\mathcal{F}_t\}_{t\geqslant 0}$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ if

(i) $\mathcal{F}_t$ is a sub σ-algebra of $\mathcal{F}$ for each t.

(ii) $\mathcal{F}_t \subset \mathcal{F}_s$ for all $t \leqslant s$.

The simplest filtration choice for a stochastic process is that generated by the process itself, $\mathcal{F}_t := \sigma(X_s \colon 0 \leqslant s \leqslant t)$. When dealing with stochastic processes, the concept of measurability is usually relegated because the introduction of the filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ unfolds more useful concepts. We say that a measurable function $f \colon [0, \infty) \times \Omega \to \mathbb{R}$ is *adapted to the filtration* $\{\mathcal{F}_t\}_{t\geqslant 0}$ if $\omega \to f(t, \omega)$ is a $\mathcal{F}_t$-measurable function for each t.

Financial assets are modeled using the Brownian motion or Wiener process $\{W_t\}_{0\leqslant t\leqslant T}$, which is a continuous-time stochastic process defined on the time interval $[0, T]$. The Wiener process is obtained as the limit of scaled random walks as $n \to \infty$, inheriting properties from random walks.

Definition 2 (Brownian motion)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a continuous-time stochastic process $\{W_t \colon 0 \leqslant t \leqslant T\}$ is a standard Brownian Motion or Wiener process on $[0, T]$ with respect to the filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ if

(i) W_t is $\mathcal{F}_t$-measurable for each t.

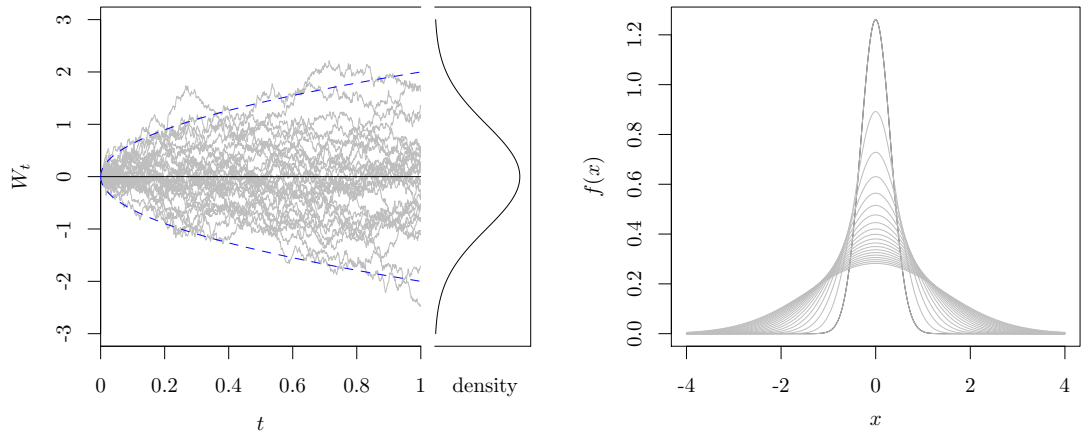
(ii) $W_0 = 0$ almost sure (a.s.).

(iii) For any $0 \leqslant s < t \leqslant T$ the increment $W_t - W_s$ has Gaussian distribution with zero mean and variance $(t - s)$.

(iv) The increments of W_t are independent, that is, $W_t - W_s$ is independent of $\mathcal{F}_s$ for each $s < t$.

(v) For all ω in a set of probability one, $W_t(\omega)$ is a continuous function of t.

Figure 2.1 shows simulated paths of the standard Wiener process in the time interval $[0, 1]$. The mean of the processes is zero and ± 2 standard deviation at each $t \in [0, T]$ is represented in blue.



(a) Simulated paths of a Wiener process and density at $t = 1$.

(b) Evolution of the probability density function for different values of t.

Figure 2.1: Simulated paths of a Wiener process or Brownian motion and evolution of the probability density function.

An important property of the Wiener process is that it is a martingale (with respect to its own filtration).

Definition 3 (Martingale)**.** Let $\{M_t \colon 0 \leqslant t < \infty\}$ be a sequence of random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and adapted to the filtration $\{\mathcal{F}_t\}$. Given that $\{M_t\}$ is $\mathcal{F}_t$-measurable and integrable, $\mathbb{E}\big[|M_t|\big] < \infty$, for all t, we say that the sequence $\{M_t\}$ is a martingale with respect to $\mathcal{F}_t$ if it satisfies the martingale identity,

$$\mathbb{E}\big[M_t \mid \mathcal{F}_s\big] = M_s,$$

a.s. for all $s \leqslant t$, with $s, t \in [0, \infty)$.

The Itô integral

$$I(f)(\omega) = \int_0^t f(u)\, \mathrm{d}W_u,$$

is also a martingale (as long as additional integrability conditions hold). Both the integrand and the variable of integration (the Wiener process) are random, with f adapted to the filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$.

Stochastic calculus provides a theoretical framework to assign meaning to differential equations involving continuous-time stochastic processes, with the Wiener process being a fundamental continuous process. In the next section, the Itô integral is used to build stochastic differential equations driven by the Wiener process, as integration with respect to Wiener process gives rise to a unifying representation for a wide class of diffusion processes (and martingales).

2.2.2 Itô diffusion process

To build the diffusion model, we first express the stochastic differential equation in its integral form, as the integral equation can be rigorously interpreted in the Itô sense.

Definition 4 (Itô process)**.** The continuous-time process $\{X_t\colon 0 \leqslant t \leqslant T\}$ is a real-valued Itô process if there exists a process $m(t)$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ satisfying

$$\int_0^T |m(s)|\,\mathrm{d}s < \infty \quad \text{a.s.},$$

and a predictable process $\sigma(t)$ satisfying

$$\int_0^T |\sigma^2(s)|\,\mathrm{d}s < \infty \quad \text{a.s.},$$

such that

$$X_t = X_0 + \int_0^t m(s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s,$$

with $t \in [0, T]$ and $\{W_t\colon t \geqslant 0\}$ a real-valued Wiener process.

The stochastic differential equation (SDE) of the Itô process,

$$\mathrm{d}X_t = m(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t,$$

with $X_0 \in \mathbb{R}$, is often used as a shortened form of the integral representation. The function $m(\cdot)$ is known as the drift coefficient and $\sigma(\cdot)$ is the diffusion or volatility function. The conditions of existence and uniqueness of the solution of the equation (see, e.g., Karatzas and Shreve, 1998) are given below.

Theorem 2.1 (Existence and uniqueness conditions of solutions to SDEs)**.** *Let* $0 \leqslant T < \infty$ *and* $m\colon \mathbb{R} \to \mathbb{R}$, $\sigma\colon \mathbb{R} \to \mathbb{R}$ *be measurable functions of the stochastic differential equation*

$$\mathrm{d}X_t = m(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t,$$

with $t \in [0, T]$ *and initial condition* $X_0 = x_0 \in \mathbb{R}$*. Sufficient conditions for the existence and uniqueness of squared-integrable solutions,* $\mathbb{E}\left[\int_0^T |X_t|^2\,\mathrm{d}t\right] < \infty$*, to the stochastic differential equation are linear growth,*

$$|m(t,x)| + |\sigma(t,x)| \leqslant C_1(1 + |x|),$$

for all $x \in \mathbb{R}$ *and some finite constant* $C_1 > 0$*; and Lipschitz continuity,*

$$|m(t,x) - m(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant C_2|x - y|,$$

for all $x, y \in \mathbb{R}$ *and where* $0 < C_2 < \infty$ *is the Lipschitz constant.*

The solution obtained in Theorem 2.1 is of strong type, which implies pathwise uniqueness. Under different assumptions, a weak solution can also be obtained. For inference purposes, usually a weak solution is sufficient, because, though two weak solutions $X_t^{(1)}$ and $X_t^{(2)}$ are not necessarily identical pathwise, their distributions are. Next, the Itô diffusion is defined.

Definition 5 (Itô diffusion)**.** An Itô diffusion is a stochastic process $X_t \in \mathbb{R}$ that satisfies a stochastic differential equation

$$\mathrm{d}X_t = m(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{2.1}$$

where the functions $m\colon \mathbb{R} \to \mathbb{R}$ and $\sigma\colon \mathbb{R} \to \mathbb{R}$ satisfy the linear growth and Lipschitz continuity conditions of Theorem 2.1.

As some regularity conditions are needed in order to ensure the existence and uniqueness of a strong stationary solution to (2.1), the following conditions are assumed along the different chapters of this dissertation.

Assumption 1. The time-homogeneous drift $m(\cdot)$ and diffusion $\sigma(\cdot)$ functions are continuous differentiable and $\mathcal{B}$-measurable for $X_{t_0} \in (-\infty, +\infty)$, with $\mathcal{B}$ the σ-field of the Borel sets on $\mathbb{R}$, and

$$\int_0^T |m(X_t)|\,\mathrm{d}t < \infty \quad \text{and} \quad \int_0^T |\sigma^2(X_t)|\,\mathrm{d}t < \infty,$$

a.s.

Assumption 2. The drift $m(\cdot)$ and diffusion $\sigma(\cdot)$ functions satisfy the linear growth condition, that is, for all $x \in \mathbb{R}$, there exists a constant $C_1 > 0$ such that

$$|m(x)| \leqslant C_1(1 + |x|), \qquad 0 \leqslant \sigma(x) \leqslant C_1(1 + |x|).$$

Assumption 3. The drift $m(\cdot)$ and diffusion $\sigma(\cdot)$ functions satisfy the uniform Lipschitz condition in x, that is, for all $x, y \in \mathbb{R}$ there exists a constant $C_2 > 0$ such that

$$|m(x) - m(y)| \leqslant C_2|x - y|, \quad \text{and} \quad |\sigma(x) - \sigma(y)| \leqslant C_2|x - y|.$$

Many diffusion models have been proposed to describe the dynamics of interest rates in a continuous-time setting. Trying to develop a common framework, Chan et al. (1992) proposed a general parametrization,

$$\mathrm{d}X_t = (\alpha + \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

with $\boldsymbol{\theta} = (\alpha, \kappa, \sigma, \gamma)^{\top}$, where many proposal can be nested. These are included in Table 2.1, where in all models we have $\sigma > 0$.

Model	$m(X_t, \boldsymbol{\theta})$	$\sigma(X_t, \boldsymbol{\theta})$	Parameter restrictions
Merton (1973)	α	σ	$\kappa = \gamma = 0$
Vasicek (1977)	$\alpha + \kappa X_t$	σ	$\gamma = 0$
Cox (1975),	κX_t	σX_t^{γ}	$\alpha = 0$
Cox and Ross (1976)			
Dothan (1978)	0	σX_t	$\alpha = \kappa = 0, \gamma = 1$
Cox et al. (1980),	0	$\sigma X_t^{1.5}$	$\alpha = \kappa = 0, \gamma = 1.5$
Constantinides and Ingersoll (1982)			
Rendleman and Bartter (1980)	κX_t	σX_t	$\alpha = 0, \gamma = 1$
Brennan and Schwartz (1980),	$\alpha + \kappa X_t$	σX_t	$\gamma = 1$
Courtadon (1982)			
Cox et al. (1985),	$\alpha + \kappa X_t$	$\sigma\sqrt{X_t}$	$\gamma = 0.5$
Brown and Dybvig (1986),			
Gibbons and Ramaswamy (1993)			
Chan et al. (1992)	$\alpha + \kappa X_t$	σX_t	None

Table 2.1: Different parametrizations of diffusion models for the short-term interest rate proposed in the literature.

2.2.3 Equivalent measures and Girsanov's Theorem

The change of measure based on Girsanov's Theorem is a relevant feature in stochastic calculus, as it allows to solve SDEs driven by Wiener process by changing its underlying probability measure (obtaining a weak solution). Thus, under the new probability measure, the process becomes the solution to the SDE. In financial markets models, we need to change from the real world probability measure $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$. To achieve this, a positive random variable Z can be used to change probability measures on the space Ω. Therefore, to change from $\mathbb{P}$ to $\mathbb{Q}$, probabilities in Ω are reassigned using Z to find out in Ω whether we should revise upwards the probability (where $Z > 1$) or revise it downwards (where $Z < 1$). This process is done set by set and it is described in the following theorem.

Theorem 2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and Z an a.s. nonnegative random variable with $\mathbb{E}[Z] = 1$. For $\mathcal{A} \in \mathcal{F}$, $\mathbb{Q}$ is a probability measure if*

$$\mathbb{Q}(\mathcal{A}) = \int_{\mathcal{A}} Z(\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

Given a nonnegative random variable X, then

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}[XZ],$$

and if Z is strictly positive a.s., we have

$$\mathbb{E}[Y] = \mathbb{E}^{\mathbb{Q}}\left[\frac{Y}{Z}\right],$$

for any nonnegative random variable Y.

The expectation operator $\mathbb{E}^{\mathbb{Q}}$ is the expectation under the probability measure $\mathbb{Q}$, that is,

$$\mathbb{E}^{\mathbb{Q}} = \int_{\Omega} X(\omega)\, \mathrm{d}\mathbb{Q}(\omega).$$

Assuming that $Z > 0$ a.s., under this theorem the measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent.

Definition 6 (Equivalent measure)**.** Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined on $(\Omega, \mathcal{F})$ are equivalent if they agree on which sets in $\mathcal{F}$ have probability zero.

As sets with probability one are complements of sets with probability zero, $\mathbb{P}$ and $\mathbb{Q}$ do also agree on which sets in $\mathcal{F}$ have probability one.

When working with financial models, first the sample space Ω is set up, which can be interpreted as the set of future possible scenarios. This set has a real probability measure $\mathbb{P}$ but, when pricing derivative securities, a risk-neutral probability measure $\mathbb{Q}$ is used. Both the real world $\mathbb{P}$ and the risk-neutral $\mathbb{Q}$ probability measures must agree on the possible or impossible events and may disagree on the probability of those possibilities, that is, they are equivalent measures. The risk-neutral measure is used to compute prices and hedges can be obtained working with probability one under the real world probability measure (and, implicitly, under the risk-neutral measure). The existence of the risk-neutral probability fulfills the *First Fundamental Theorem of Asset Pricing*, that is, the financial model is arbitrage-free. This probability measure arises from the following theorem.

Theorem 2.3 (Radon-Nikodym)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be equivalent probability measures defined in the probability space $(\Omega, \mathcal{F})$. There exists an a.s. positive random variable Z such that $\mathbb{E}[Z] = 1$ and*

$$\mathbb{Q}(\mathcal{A}) = \int_{\mathcal{A}} Z(\omega)\, \mathrm{d}\mathbb{P}(\omega),$$

for every $\mathcal{A} \in \mathcal{F}$.

Definition 7 (Radon-Nikodym derivative)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F})$ equivalent to $\mathbb{P}$, and let Z be an a.s. positive random variable such that

$$\mathbb{Q}(\mathcal{A}) = \int_{\mathcal{A}} Z(\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

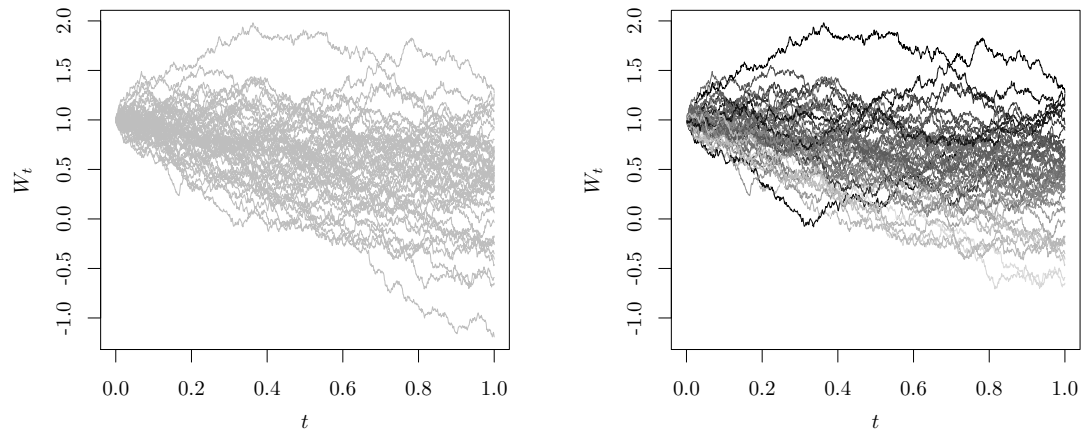
The random variable Z is the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$, that is,

$$Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}.$$

The definition of the Radon-Nikodym derivative process means that the same probability measure can be used for all time horizons, without the need to define, for each filtration or horizon, a new probability measure. An important application of these results to the case of the Wiener process is given by Girsanov's Theorem.

Theorem 2.4 (Girsanov's Theorem)**.** *Let $\{W_t \colon 0 \leqslant t \leqslant T\}$ be a Wiener process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\{\mathcal{F}_t \colon 0 \leqslant t \leqslant T\}$ be a filtration for the Wiener process. Let $\{\mu(t) \colon 0 \leqslant t \leqslant T\}$ be an adapted process, define*

$$Z(t) = \exp\left(-\int_0^t \mu(u)\, \mathrm{d}W_u - \frac{1}{2}\int_0^t |\mu(u)|^2\, \mathrm{d}u\right),$$



(a) Simulated paths of a Wiener process or Brownian motion.

(b) Same paths reweighted according to Radon-Nikodym derivative Z_T.

Figure 2.2: Radon-Nikodym derivative example: simulated paths of a Wiener process with negative drift (left) and same paths where darker color corresponds to higher values of the derivative (right).

and assume that

$$\mathbb{E}\left[\int_0^T \mu^2(u)Z^2(u)\,\mathrm{d}u\right] < \infty.$$

Set $Z = Z(T)$, then $\mathbb{E}[Z] = 1$ and under the probability measure $\mathbb{Q}$, given by

$$\mathbb{Q}(\mathcal{A}) = \int_{\mathcal{A}} Z(\omega)\,\mathrm{d}\mathbb{P}(\omega),$$

the process

$$W_t^{\mathbb{Q}} = W_t + \int_0^t \mu(u)\,\mathrm{d}u,$$

$0 \leqslant t \leqslant T$, is a Wiener process under the measure $\mathbb{Q}$ up to time T.

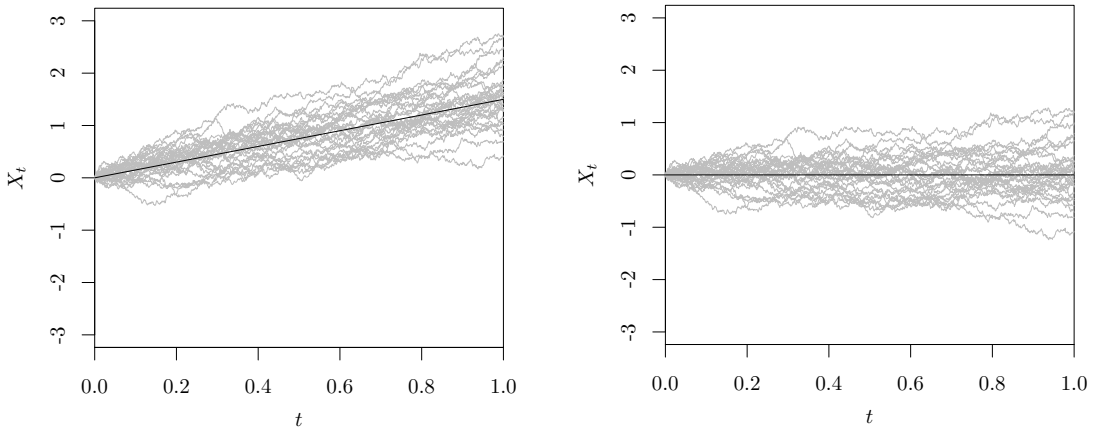
Given that $\mathbb{P}(Z > 0) = 1$, the probability measures $\mathbb{Q}$ and $\mathbb{P}$ of Girsanov's Theorem 2.4 are equivalent,

$$\mathbb{P}(\omega) > 0 \Leftrightarrow \mathbb{Q}(\omega) > 0,$$

for all $\omega \in \Omega$. Both measures agree about the sets that have probability zero and, therefore, about the sets that have probability one.

As an example, consider a standard Wiener process W_t under $\mathbb{P}$ and a constant drift μ, then under the Radon-Nikodym derivative we have the exponential martingale

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z_t = \exp\left(\mu W_t - \frac{1}{2}\mu^2 t\right),$$



(a) Diffusion process $\mathrm{d}X_t = 1.5\,\mathrm{d}t + 0.6\,\mathrm{d}W_t$ *under the measure* $\mathbb{P}$.

(b) Diffusion process $\mathrm{d}X_t = 0.6\,\mathrm{d}W_t^{\mathbb{Q}}$ *under the measure* $\mathbb{Q}$.

Figure 2.3: Change of probability measure for the diffusion $\mathrm{d}X_t = 1.5\,\mathrm{d}t + 0.6\,\mathrm{d}W_t$.

where the process $W_t^{\mathbb{Q}} = W_t - \mu t$ under $\mathbb{Q}$ is a Wiener process. Note that $W_t^{\mathbb{Q}}$ under the original probability measure $\mathbb{P}$ is not a standard Wiener process. Girsanov's Theorem has an important practical relevance as it allows the creation of a new standard Wiener process $W_t^{\mathbb{Q}}$ under the probability measure $\mathbb{Q}$, whatever the drift μ is.

Taking the constant drift example, we define the diffusion process $\mathrm{d}X_t = 1.5\,\mathrm{d}t + 0.6\,\mathrm{d}W_t$ under the real measure $\mathbb{P}$. Using Girsanov's transformation will alter the drift but the diffusion coefficient is left unchanged,

$$\begin{aligned}\mathrm{d}X_t &= 1.5\,\mathrm{d}t + 0.6\,\mathrm{d}W_t \\ &= 0.6\left(\frac{1.5}{0.6}\,\mathrm{d}t + \mathrm{d}W_t\right) \\ &= 0.6\,\mathrm{d}W_t^{\mathbb{Q}}.\end{aligned}$$

Figure 2.3 shows various simulated paths of the diffusion process X_t under the real probability measure $\mathbb{P}$, see Figure 2.3a, where the processes grow accordingly the drift $\mu t = 1.5t$ (black solid line); while those paths under the probability measure $\mathbb{Q}$ are shifted, see Figure 2.3b, and the process has no drift.

In the asset price theory, $\mathbb{P}$ is the actual probability measure, or the real world probability measure, and $\mathbb{Q}$, the measure defined in Girsanov's Theorem 2.4, is the risk-neutral probability measure.

2.2.4 The risk neutral measure

Girsanov's Theorem is an important link between SDEs and asset pricing theory, as it provides a means for the definition of the equivalent martingale measure. In the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ the risky assets are modeled as random variables on Ω. To price assets we can use the risk-neutral measure or an equivalent martingale measure (EMM).

Definition 8 (Equivalent martingale measure)**.** A probability measure $\mathbb{Q}$ defined on the space $(\Omega, \mathcal{F}, \mathbb{P})$ is an equivalent martingale measure if $\mathbb{Q}$ is an equivalent measure to $\mathbb{P}$ and such that, under $\mathbb{Q}$, the discounted prices are martingales. This measure is known as risk-neutral measure.

This is a convenient pricing method because asset prices are an expectation of the payoff. Let X_t be the instantaneous interest rate under $\mathbb{Q}$, defined by an stochastic differential equation and choose the bond price B as

$$B_t = \exp\left(-\int_0^t X_u \, \mathrm{d}u\right).$$

There exists a probability measure $\mathbb{Q}$ under which prices of zero coupon bonds $D(t, T, X_t)$ divided by the bond price B_t are martingales. Therefore, under this martingale representation method, the discount bond price is

$$\frac{D(0,T,X_t)}{B_s} = \mathbb{E}^{\mathbb{Q}}\left[\frac{D(0,T,X_t)}{B_s} \,\middle|\, \mathcal{F}_s\right] = D(s,T,X_t) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_s^T X_t \, \mathrm{d}t\right)\right],$$

with $\mathcal{F}_s$ the filtration of the diffusion process X_t. The existence of the risk neutral measure implies no arbitrage opportunities, while the uniqueness of the risk neutral measure implies that every derivative can be hedged. The implications of the existence and uniqueness of the risk neutral measure are the fundamental theorems of asset pricing theory.

Among others, the Vasicek (1977) model characterized the short-term interest rate X_t under the risk-neutral probability measure $\mathbb{Q}$ as

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma \, \mathrm{d}W_t^{\mathbb{Q}},$$

with $\kappa, \mu, \sigma > 0$. Considering a constant risk premium λ is equivalent to change the probability measure,

$$\sigma \, \mathrm{d}W_t^{\mathbb{Q}} = \sigma\left(\mathrm{d}W_t^{\mathbb{P}} + \lambda \, \mathrm{d}t\right),$$

where under the historical probability measure $\mathbb{P}$ the process satisfy

$$\begin{aligned} \mathrm{d}X_t &= \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{Q}} \\ &= \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\left(\mathrm{d}W_t^{\mathbb{P}} + \lambda\,\mathrm{d}t\right) \\ &= \kappa\left(\mu + \frac{\lambda\sigma}{\kappa} - X_t\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}} \\ &= \kappa(\tilde{\mu} - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^{\mathbb{P}}, \end{aligned}$$

with $\tilde{\mu} = \mu + \lambda\sigma/\kappa$. Thus, applying Girsanov's theory we obtain the dynamics under the real world probability measure $\mathbb{P}$, introducing a risk premium. Unless stated otherwise, through this document the Wiener process is defined under the probability measure $\mathbb{Q}$. The measure change from $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$ modifies the mean rate but not the volatility, which tells us which paths are possible. After this change of measure, the same sets of paths are considered, but the probability has shifted, as the measure change puts more probability on the lower return paths.

The absence of arbitrage-free opportunities does imply the existence of the risk-neutral measure $\mathbb{Q}$, which means that models of the term structure (e.g., those in Table 2.1) are equivalent to risk-neutral pricing. The absence of arbitrage opportunities on the market is an economically justifiable hypothesis, as in a financial market an expected profit without any risk and zero net investment should not be possible. On the other hand, the existence of the equivalent measure $\mathbb{Q}$ is not a priori guaranteed. However, many works in the literature deal with the relations between martingale theory and arbitrage-free arguments, such as Harrison and Kreps (1979) and Harrison and Pliska (1981), where the fundamental theorems of asset pricing were established. The first theorem connects the martingale theory to the absence of arbitrage argument, discussing that no-arbitrage opportunities characterizes the existence of the equivalent martingale measure $\mathbb{Q}$ in a frictionless security market.

Chapter 3

Parametric estimation of diffusion models

This chapter provides a review about parametric estimation methods for stationary stochastic differential equations with discrete time observations. The short-term interest rate dynamics are commonly described by continuous-time diffusion processes, whose parameters are subject to estimation bias, as data are highly persistent, and discretization bias, as data are discretely sampled despite the continuous-time nature of the model. To assess the role of persistence and the impact of sampling frequency on the estimation, we conducted a simulation study under different settings to compare the performance of the procedures and illustrate the finite sample behavior. To complete the survey, an application of the procedures to real data is provided. All the methods were implemented in a new R package, `estsde`.

The contents of this chapter appear on López-Pérez et al. (2021).

Contents

3.1 Introduction

Diffusion processes described by stochastic differential equations are frequently applied in physical, biological and financial fields to model dynamical systems with a disturbance term. Its use in mathematical finance for modeling the evolution of important economic variables, such as the interest rate, has been increasingly important over the last decades. The need to develop the analysis of the term structure of interest rates in a stochastic environment emerges as a consequence of turbulences in the financial market throughout the seventies. New theories of the term structure of interest rate based on pricing models in absence of arbitrage under stochastic environment were emerging: Merton (1973) used the interest rate in option pricing modeling it as a stochastic process. Subsequently, Black and Scholes (1973) had an important impact on arbitrage models of the term structure of interest rates, as shown in Vasicek (1977), Brennan and Schwartz (1979), Cox et al. (1985), Chan et al. (1992) or Ahn and Gao (1999). In these models, the interest rate is the solution of the stochastic differential equation, therefore we can use the framework of Markov processes theory for its analytical treatment. The continuous time paradigm proves to be an especially useful tool, but the continuous time nature of the model does complicate the estimation of the parameters because available data are sampled in discrete time. Thus, parameter estimates are subject to discretization bias together with estimation bias and this issues have been addressed using different estimation methods. The finite sample bias is especially acute when the process is highly persistent, such as time series of interest rates, and alters the valuation of derivatives since short term interest rate models are used to price these instruments (Phillips and Yu, 2005).

We consider the SDE defined in a filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, \mathbb{P}\right)$, where Ω is a nonempty set, $\mathcal{F}$ is a σ-algebra of subsets of Ω and $\mathbb{P}$ is a probability measure, $\mathbb{P}(\Omega) = 1$. We will focus on parametric time-homogeneous stochastic differential equations, where X_t is an Itô process,

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t, \qquad \text{with } X_0 = x_0, \quad 0 \leqslant t \leqslant T, \tag{3.1}$$

with $X_t \in \mathbb{R}$, W_t a $\{\mathcal{F}_t\}_{t\geqslant 0}$-adapted standard Wiener process and $\boldsymbol{\theta}$ an unknown parameter vector such that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ with d a positive integer and Θ a compact set. We assume

the knowledge of the drift and diffusion functions parametric structure, $m(\cdot, \boldsymbol{\theta}) \colon \mathbb{R} \times \Theta \to \mathbb{R}$ and $\sigma(\cdot, \boldsymbol{\theta}) \colon \mathbb{R} \times \Theta \to (0, \infty)$, respectively, where none is time dependent.

A parametric specification of (3.1) that encloses different interest rate models is the Chan-Karolyi-Longstaff-Sanders (CKLS) model proposed in Chan et al. (1992), given by

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

which is a mean-reverting process that allows the conditional mean and variance to depend on the interest rate level X. The drift parameter μ is the long-term mean and κ is the rate of reversion, while the diffusion parameter σ is the volatility and γ is the proportional volatility exponent that measures the sensitivity of the volatility regarding the process in time t. This model generalizes prior interest rate models by imposing restrictions on the parameters. When γ is 0 or 0.5—which yields the Vasicek (1977) model and the Cox et al. (1985), respectively—the process is tractable and admits an analytical solution.

In this chapter, we consider SDEs with deterministic volatility function, though current option pricing literature withdraws the constant volatility assumption and adopts instead a stochastic volatility framework. Nevertheless, the issues addressed here are inherited by stochastic volatility models and the estimation methods can be extended to SDEs with volatility described by a stochastic process.

Several studies of estimation methods for diffusion processes can be found in the literature, see Sørensen (2004) for a theoretical comparison or Hurn et al. (2007) for a more practical approach. In Shoji and Ozaki (1997) a comparative study of different discretization methods and moment-based estimation is carried out, while Durham and Gallant (2002) focused on simulation-based approaches. The aim of this chapter is to complement these comparative studies extending the evaluation of the finite sample performance to different settings—near unit-root time series, various degrees of persistence without changes in the marginal density or different sampling intervals and observation times—whose impact on estimation have been hinted at in the financial literature. The methods here considered are maximum likelihood estimation, local linearization (Ozaki, 1992; Shoji and Ozaki, 1998), Hermite polynomial expansion (Aït-Sahalia, 2002), Kalman filter (Kalman, 1960), Markov Chain Monte Carlo (Elerian et al., 2001; Eraker, 2001) and generalized method of moments (Hansen, 1982; Chan et al., 1992). The procedures are provided in the companion `estsde` R (R Core Team, 2014) package (see Appendix B), implemented in C and C++ for the sake of efficiency.

The sections are organized as follows: Section 3.2 provides an outline of the estimation methods, Section 3.3 designs the Monte Carlo experiment and discusses the finite sample

performance of the procedures. Real data applications to interest rate series are presented in Section 3.4 and conclusions are drawn in Section 3.5. Tabulated simulation results are deferred to Appendix A.

3.2 Estimation methods

The unique strong solution X_t of the SDE in (3.1),

$$X_t = X_{t_0} + \int_0^T m(X_u, \boldsymbol{\theta})\,\mathrm{d}u + \int_0^T \sigma(X_u, \boldsymbol{\theta})\,\mathrm{d}W_u,$$

exists under the assumptions that both drift $m(\cdot)$ and volatility $\sigma(\cdot)$ functions satisfy global Lipschitz continuous and growth conditions (see Assumptions 1–3 in Chapter 2).

Although the model is formulated in continuous time, data are registered in discrete time points. In this regard, for the estimation of the continuous time model parameters we should consider a discrete version of it. The observation scheme assumed is the *fixed-Δ scheme*, in which the time step Δ between two consecutive observations is fixed and the sample size $n \in \mathbb{N}$ increases, as well as the time interval $[0, T = n\Delta]$. One of the most used approximation schemes is the Euler-Maruyama method (Maruyama, 1955): given an Itô process $\{X_t, 0 \leqslant t \leqslant T\}$, solution of the SDE in (3.1) with initial value $X_{t_0} = x_0$ and the discretization of the time interval $[0, T]$, $0 = t_0 < t_1 < \ldots < t_n = T$, the Euler-Maruyama approximation of X is a continuous stochastic process that satisfies the iterative scheme

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

with $i = 0, 1, \ldots, n-1$, $t_i = i\Delta$ and $X_{t_0} = x_0 \in \mathbb{R}$.

In the remainder of this section, we provide an outline of the procedures and their implementation to estimate the unknown parameter vector $\boldsymbol{\theta}$. The methods described fall into two categories: likelihood-based and method of moments. The method of moments provides estimates by matching population and sample moments and minimizing a quadratic form, hence we assume that the moment conditions of X_t are bounded:

Assumption 4 (Bounded moments)**.** For all $k > 0$, all order k moments of the diffusion process exist and are such that

$$\sup_{t \geqslant 0} \mathbb{E}\left[\|X_t\|^k\right] < \infty.$$

The maximum likelihood (ML) estimates are obtained by different methods: exact and discrete (piecewise constant or linear approximations) ML, univariate Hermite expansion

of the transition function, filtering algorithm (linear quadratic estimation) and Bayesian approach.

3.2.1 Exact maximum likelihood

As X_t is a Markov process, we can obtain the likelihood function $\mathcal{L}_n(\boldsymbol{\theta})$ of the discrete process using Bayes' rule,

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

where $p_{\boldsymbol{\theta}}(\Delta, X_{t_i} \mid X_{t_{i-1}})$ denotes the transition density function associated to the parametric diffusion model, with unknown parameter $\boldsymbol{\theta}$. If the parametric form of the model that generates the observations $\{X_{t_i}\}_{i=0}^n$ is known, we can use a maximum likelihood method, so that the maximum likelihood estimator (MLE) of the true parameter is

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

where $\ell_n(\boldsymbol{\theta}) = \log \mathcal{L}_n(\boldsymbol{\theta})$ is the log-likelihood function. This estimation method can seldom be used with diffusion processes, as few models have a closed-form solution (e.g., Black and Scholes, 1973; Vasicek, 1977; Cox et al., 1985). As a consequence, new procedures have been proposed in the ML framework based on different approximations of the transition density function.

3.2.2 Discrete maximum likelihood

Discrete time likelihood (also known as pseudo-likelihood) methods emerge as an alternative approach to approximate the unknown transition density of SDEs, where the diffusion model is discretized with a certain numerical scheme, when exact maximum likelihood is unfeasible. In some cases, analytical expressions for the parameters estimates can be obtained, otherwise a numerical optimization routine is needed to maximize (minimize) the (negative) log-likelihood. Several algorithms have been proposed to approximate SDEs (see, e.g., Kloeden and Platen, 1992b; Elerian, 1998; Shoji and Ozaki, 1998), here we briefly detail two of them.

Euler method

To estimate the model we can use an approximation scheme, such as the Euler-Maruyama (Maruyama, 1955) method. With this method we do not approximate the transition density directly, instead the trajectory of the process is approximated so that we can use the likelihood of the discretized version of the model, given by

$$X_{t_{i+1}} - X_{t_i} = m(X_{t_i}, \boldsymbol{\theta})\Delta + \sigma(X_{t_i}, \boldsymbol{\theta})\Delta^{1/2}\varepsilon_{t_i}, \quad i = 0, 1, \ldots, n-1, \tag{3.2}$$

where ε_{t_i} are independent identically distributed (i.i.d.) $\mathcal{N}(0,1)$ and $\Delta = (t_{i+1} - t_i)$. Therefore, the estimation with the Euler-Maruyama method proceeds as if the observations follow a Gaussian distribution, with mean the drift function and standard deviation the diffusion function. Thus, the transition density is given by

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

The implementation of this method is straightforward, however, Euler-type schemes depend on the sampling interval $\Delta > 0$ and introduce discretization bias in the estimates, although they converge to exact ML estimates as $\Delta \to 0$. Departures from Gaussian distribution can also increase bias since the Euler scheme increments are conditionally Gaussian.

Local linearization

While the Euler-Maruyama approximation method restricts coefficients of the drift and diffusion terms to be piecewise constant, the local linearization instead uses a linear approximation. Considering the SDE

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{3.3}$$

where $\sigma > 0 \in \mathbb{R}$ is assumed constant, the local linearization method developed in Ozaki (1992) and Shoji and Ozaki (1998) is an approximation method by which the drift function $m(\cdot, \boldsymbol{\theta})$ is locally approximated by a linear function of X_t (no expansion is required for the diffusion function, as it is constant). The numerical scheme is based on the local linearization of the SDE's drift coefficient by means of a truncated Itô-Taylor expansion. The process discretized by the local linearization method is

$$X_{(i+1)\Delta} = X_{i\Delta} + \frac{m(X_{i\Delta}, \boldsymbol{\theta})}{L_{i\Delta}}\left(e^{\Delta L_{i\Delta}} - 1\right) + \sigma \int_{i\Delta}^{(i+1)\Delta} e^{K_{i\Delta}[(i+1)\Delta - u]}\,\mathrm{d}W_u,$$

where

$$K_t = \frac{1}{\Delta} \log \left(1 + \frac{m(X_t, \boldsymbol{\theta})}{X_t L_t} \left(e^{\Delta L_t} - 1 \right) \right)$$

and $L_t = \partial m(X_t, \boldsymbol{\theta}) / \partial X$.

The linear function K_t approximates the drift function $m(\cdot)$, with K_t constant in the interval $\left[i\Delta, (i+1)\Delta \right)$. Given that the stochastic integral is a Gaussian random variable, the transition density for $X_{(i+1)\Delta}$ given $X_{i\Delta}$ is indeed Gaussian. Thus, we have that $\left(X_{(i+1)\Delta} \mid X_{i\Delta} \right)$ follows a normal distribution with mean and variance given by

$$\begin{aligned} \mathbb{E}\left[X_{(i+1)\Delta} \mid X_{i\Delta} \right] &= X_{i\Delta} + \frac{m(X_{i\Delta}, \boldsymbol{\theta})}{L_{i\Delta}} \left(e^{\Delta L_{i\Delta}} - 1 \right), \\ \mathbb{V}\mathrm{ar}\left[X_{(i+1)\Delta} \mid X_{i\Delta} \right] &= \sigma^2 \left(\frac{e^{2\Delta K_{i\Delta}} - 1}{2K_{i\Delta}} \right), \end{aligned}$$

respectively, and therefore, maximum likelihood can be used to obtain the parameter estimates. As the SDE in (3.3) has a constant diffusion function, if the parametric specification we want to estimate is more intricate, a transformation is needed (e.g., standardizing the diffusion term with the Lamperti transform, see Section 3.2.3). The local linearization approximation can provide more accurate estimates than the Euler scheme (specially for nonlinear drifts), though the implementation is more troublesome as a prior transformation of the process is needed, as well as the derivative of the transformed process drift function (which can be computed numerically or analytically, for higher efficiency).

3.2.3 Hermite polynomial expansion

Bayes' rules combined with the Markovian nature of the diffusion process, inherited by discrete data, implies that the log-likelihood is of the form

$$\ell_n(\boldsymbol{\theta}) \equiv \frac{1}{n} \sum_{i=1}^{n} \ln \left(p_{X,\boldsymbol{\theta}}(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta}) \right), \tag{3.4}$$

assuming that the process is observed in the time interval $\{i\Delta \mid i = 0, \ldots, n\}$, with fixed Δ.

Aït-Sahalia (2002) proposed a maximum likelihood method for diffusion processes with discrete samples, based on an approximation of the likelihood function using Hermite polynomials. The author constructed a succession of approximations $\{p^{(K)}_{X,\boldsymbol{\theta}} \mid K \geqslant 0\}$ of the transition density, such that (3.4) is a succession of approximations of $\ell^{(K)}_n$.

We will need to standardize the diffusion coefficient of X, which is achieved using the

Lamperti transform,

$$U_t = \psi(X_t, \boldsymbol{\theta}) = \int^{X_t} \frac{1}{\sigma(s, \boldsymbol{\theta})}\, \mathrm{d}s,$$

and using Itô's formula in the new process U_t, we have the unitary diffusion

$$\mathrm{d}U_t = \left(\frac{m(\psi^{-1}(u, \boldsymbol{\theta}); \boldsymbol{\theta})}{\sigma(\psi^{-1}(u, \boldsymbol{\theta}); \boldsymbol{\theta})} - \frac{1}{2} \frac{\partial \sigma}{\partial x} \left(\psi^{-1}(u, \boldsymbol{\theta}); \boldsymbol{\theta}\right) \right) \mathrm{d}t + \mathrm{d}W_t,$$

provided that $\psi^{-1}(X_t, \boldsymbol{\theta})$ exists. This transformation allows the computation of the transition density $p_{X,\boldsymbol{\theta}}$ from $p_{U,\boldsymbol{\theta}}$ through the Jacobian formula

$$\tilde{p}_X^{(J)}(\Delta, x \mid x_0; \boldsymbol{\theta}) \equiv \sigma(\psi(x; \boldsymbol{\theta}); \boldsymbol{\theta})^{-1}\, \tilde{p}_U^{(J)}(\Delta, \psi(x; \boldsymbol{\theta}) \mid \psi(x_0; \boldsymbol{\theta}); \boldsymbol{\theta}),$$

where the succession of explicit functions $\tilde{p}_U^{(J)}$, based on Hermite expansions of the density p_U around a Gaussian density function up to order J, approximates p_U. In Aït-Sahalia (2002, Theorem 1) it is proved that

$$p_X^{(J)}(\Delta, x \mid x_0; \boldsymbol{\theta}) \underset{J \to \infty}{\longrightarrow} p_X(\Delta, x \mid x_0; \boldsymbol{\theta}).$$

The coefficients of the density expansion terms can be calculated with a Taylor series expansion in Δ, denoting $p_U^{(J,K)}$ as the order K Taylor series in Δ of $p_X^{(J)}$. Usually, $J = 6$ is taken, so the first seven Hermite coefficients $(j = 0, \ldots, 6)$ are used, along with Taylor series up to order $K = 3$.

To obtain the MLE, the approximation of the log-likelihood function,

$$\ell_n^{(J)} \equiv \frac{1}{n} \sum_{i=1}^{n} \ln \left(\tilde{p}_{X,\theta}^{(J)}(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta}) \right)$$

is maximized. Thus, we obtain an estimator $\hat{\boldsymbol{\theta}}_n^{(J)}$ close to the exact $\hat{\boldsymbol{\theta}}_n$ (Aït-Sahalia, 2002, Theorem 2). For stationary processes, the estimator satisfies

$$\sqrt{n}\, (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\big(0,\, i(\boldsymbol{\theta}_0)^{-1}\big),$$

where $i(\boldsymbol{\theta}_0)$ is the Fisher information matrix.

The practical implementation of this procedure is limited to the existence of an explicit inverse $\psi^{-1}(X_t, \boldsymbol{\theta})$ and its complexity emerges from the analytically approximation of the Hermite expansion coefficients, though in return the accuracy of the estimates is high.

3.2.4 Kalman filter

The state-space model or dynamic linear model, introduced in Kalman (1960), employs an order one vector autoregression as a state equation and assumes that we do not observe the state vector $\boldsymbol{x}_t$ directly, but a linear transformation of it with noise added, $\boldsymbol{y}_t$. The *state-space* representation of the dynamics of $\boldsymbol{y}_t$ is given by the system of equations:

$$\boldsymbol{x}_t = \Phi \boldsymbol{x}_{t-1} + \Upsilon \boldsymbol{u}_t + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \overset{\text{iid}}{\sim} N(0, \mathbf{Q}), \tag{3.5}$$

$$\boldsymbol{y}_t = \mathbf{A}_t \boldsymbol{x}_t + \Gamma \boldsymbol{u}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \overset{\text{iid}}{\sim} N(0, \mathbf{R}), \tag{3.6}$$

where the state vector $\boldsymbol{x}_t$ is $p \times 1$, the observed data vector $\boldsymbol{y}_t$ is $q \times 1$, the observation matrix $\mathbf{A}_t$ is $q \times p$, $\boldsymbol{u}_t$ is a $r \times 1$ vector of inputs, Υ is $p \times r$, Γ is $q \times r$ and, for simplicity, we assume that $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_t\}$ are uncorrelated. Equation (3.5) is known as the *state equation* and equation (3.6) as *observation equation*.

Let $\boldsymbol{x}_{t|t-1} = \mathbb{E}\left[\boldsymbol{x}_t \mid \boldsymbol{y}_{t-1}\right]$ and $\mathbf{P}_{t|t-1} = \mathbb{E}\left[(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})^\top\right]$, the Kalman filter equations, with initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{x}_{0|0}, \mathbf{P}_{0|0})$, are given by

$$\boldsymbol{x}_{t|t-1} = \Phi \boldsymbol{x}_{t-1|t-1} + \Upsilon \boldsymbol{u}_t, \tag{3.7}$$

$$\mathbf{P}_{t|t-1} = \Phi \mathbf{P}_{t-1|t-1} \Phi^\top + \mathbf{Q}, \tag{3.8}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{A}_t^\top \left(\mathbf{A}_t \mathbf{P}_{t|t-1} \mathbf{A}_t^\top + \mathbf{R}\right)^{-1}, \tag{3.9}$$

$$\boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + \mathbf{K}_t \left(\boldsymbol{y}_t - \mathbf{A}_t \boldsymbol{x}_{t|t-1} - \Gamma \boldsymbol{u}_t\right),$$

$$\mathbf{P}_{t|t} = \left(\mathbf{I} - \mathbf{K}_t \mathbf{A}_t\right) \mathbf{P}_{t|t-1}. \tag{3.10}$$

The Kalman filter is a recursive algorithm, equations (3.7)-(3.8) are the *time update* equations, where the state in time t is estimated with the information until time $t-1$, and equations (3.9)-(3.10) are the *measurement update* equations, where the new information of the estimation is incorporated and the mean squared error is minimized.

Let $\boldsymbol{\theta} = (\Phi, \mathbf{Q}, \mathbf{R}, \Upsilon, \Gamma)^\top$ be the vector of parameters, we can use maximum likelihood under the assumption that the initial state is Gaussian and the errors $\left\{\boldsymbol{w}_i\right\}_{i=1}^n$ and $\left\{\boldsymbol{v}_i\right\}_{i=1}^n$ are uncorrelated Gaussian vectors. The Kalman filter can be set up to evaluate the likelihood function, which can be computed with the innovations $\left\{\boldsymbol{\varepsilon}_i\right\}_{i=1}^n$, where $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \mathbf{A}_t \boldsymbol{x}_{t|t-1} - \Gamma \boldsymbol{u}_t$, and because they are independent Gaussian random vectors with zero mean and covariance matrix $\Sigma_t = \mathbf{A}_t \mathbf{P}_{t|t-1} \mathbf{A}_t^\top + \mathbf{R}$, we can write the likelihood, $\mathcal{L}_Y(\boldsymbol{\theta})$, as

$$\ln \mathcal{L}_Y(\boldsymbol{\theta}) = \frac{1}{2} \sum_{t=1}^{n} \ln |\Sigma_t(\boldsymbol{\theta})| + \frac{1}{2} \sum_{t=1}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta})^\top \Sigma_t(\boldsymbol{\theta})^{-1} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}). \tag{3.11}$$

The log-likelihood in (3.11) can be maximized by numerical search procedures to obtain the estimation of $\boldsymbol{\theta}$, where the derivatives of (3.11) can be calculated numerically or analytically. The analytical derivatives can be obtained recursively by differentiating the Kalman filter recursion (see Caines, 1988). Algorithms like the Expectation-Maximization (Dempster et al., 1977) or the Newton-Raphson can be used to maximize the log-likelihood, as in Shumway and Stoffer (1982) and Jones (1980), for an example of both approaches.

Under general conditions, let $\hat{\boldsymbol{\theta}}_n$ be the estimator of the true parameters $\boldsymbol{\theta}_0$ obtained maximizing the innovation log-likelihood in (3.11). Subject to certain regularity conditions, when $n \to \infty$

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\Big(0,\, i(\boldsymbol{\theta}_0)^{-1}\Big),$$

where $i(\boldsymbol{\theta}_0)$ is the asymptotic Fisher information matrix. The Kalman filter will generate strongly consistent estimates of $\boldsymbol{\theta}_0$ when the parameters lie in a compact set. More details on regularity conditions, convergence and asymptotic properties can be found in Pagan (1980), Ljung and Caines (1980), Caines (1988), Hannan and Deistler (1988) and Harvey (1990).

The Kalman filter can be used to estimate the parameters of the diffusion process (3.1) writing in state-space form the discrete version given in (3.2),

$$Y_{t_i} = m(X_{t_i}, \boldsymbol{\theta}) + \frac{\sigma(X_{t_i}, \boldsymbol{\theta})}{\sqrt{\Delta}}\, \varepsilon_{t_i}, \tag{3.12}$$

with $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ and ε_{t_i} are i.i.d. $\mathcal{N}(0,1)$ random variables. The discretized model (3.12) is obtained by means of a Euler-Maruyama scheme, which makes the likelihood function in (3.11) equivalent to the one obtained using the Euler method (see Section 3.2.2), hence the parameter estimates for both methods will be close. The Kalman filter algorithm provides a computationally efficient method for evaluating the log-likelihood. One of the advantages of the state-space representation of the system (3.5)–(3.6) is that it allows latent variables, which simplifies the extension to SDEs with stochastic volatility. Furthermore, this framework does also admit the estimation of multidimensional models. Some extensions of the filter to nonlinear systems have been proposed in the literature, such as the extended Kalman filter (Jazwinski, 1970), which is close related to the local linearization method introduced in Section 3.2.2 (see Singer, 2002).

3.2.5 Markov Chain Monte Carlo

The estimation of the parameters for the continuous time model by means of the Markov Chain Monte Carlo (MCMC) procedure, given the discontinuous data, implies finding the

discrete version of the model. Discretizing the model with the Euler-Maruyama approach, as in (3.2), we have

$$X_{t_i} - X_{t_{i-1}} = m(X_t, \boldsymbol{\theta})\Delta + \sigma(X_t, \boldsymbol{\theta})(W_{t_i} - W_{t_{i-1}}),$$

where $(W_{t_i} - W_{t_{i-1}})$ are i.i.d. $\mathcal{N}(0, \Delta)$. This discrete time approximation of the SDE can be too coarse to approximate the true transition density accurately (see Kloeden and Platen, 1992a, for the strong convergence criterion for SDE). Elerian et al. (2001) and Eraker (2001) proposed MCMC approaches involving data augmentation, where missing data between two neighbor observations is treated as unknown parameters. Dividing the interval $[0, T]$ into $n = mT$ equidistant points $0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = T$ implies that $T(m-1)$ data points are missing, such that $X_{t_i} = x^*_{t_i,0}, x^*_{t_i,1}, \ldots, x^*_{t_i,m-1}, x^*_{t_i,m} = X_{t_{i+1}}$, as seen in Figure 3.1. The values of the unobserved data, $x^*_{t_i,j}$, are updated using the Metropolis-Hastings algorithm. The unobserved data between two observations, X_{t_i} and $X_{t_{i+1}}$, is updated in random sized blocks, where the block size M follows a Poisson distribution with mean λ, which leads to an average block size of $\lambda + 1$. Blocks of latent points, $x^*_{t_i,k}, \ldots, x^*_{t_i,k+M-1}$, preceded by the observation $x^*_{t_i,k-1}$ and followed by $x^*_{t_i,k+M}$, have density conditioned on $(x^*_{t_i,k-1}, x^*_{t_i,k+M})$ given by

$$f(x^*_{t_i,k}, \ldots, x^*_{t_i,k+M-1} \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta}) \propto \prod_{j=k-1}^{k+M-1} \mathcal{N}\left(x^*_{t_i,j} + m(x^*_{t_i,j}; \boldsymbol{\theta})\Delta, \sigma^2(x^*_{t_i,j}; \boldsymbol{\theta})\Delta\right). \tag{3.13}$$

Each block is sampled in sequence by the Metropolis-Hastings algorithm. Therefore, new values for the unobserved block are drawn from the multivariate Gaussian distribution in equation (3.13), as suggested by Elerian et al. (2001). The probability of move used to determine if the proposal value should be taken as the next item of the chain is given by

$$\alpha = \min\left\{1, \frac{f\left(w \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta}\right) q\left(x^{*(p)}_{t_i,k}, \ldots, x^{*(p)}_{t_i,k+M-1} \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta}\right)}{f\left(x^{*(p)}_{t_i,k}, \ldots, x^{*(p)}_{t_i,k+M-1} \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta}\right) q\left(w \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta}\right)}\right\},$$

where $w \sim q(x^*_{t_i,k}, \ldots, x^*_{t_i,k+M-1} \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta})$ and $x^{*(p)}_{t_i,k}$ is the current value of $x^*_{t_i,k}$ at the end of the pth iteration. We then set $w = \left\{x^{*(p+1)}_{t_i,j}\right\}_{j=k}^{k+M-1}$ with probability α and $\left\{x^{*(p+1)}_{t_i,j}\right\}_{j=k}^{k+M-1} = \left\{x^{*(p)}_{t_i,j}\right\}_{j=k}^{k+M-1}$ with probability $(1-\alpha)$. The mean and covariance matrix of the multivariate Gaussian distribution are obtained by a Newton-Raphson iterative procedure, where the mean is given by the mode of $\ln f(\cdot \mid x^*_{t_i,k-1}, x^*_{t_i,k+M}; \boldsymbol{\theta})$ and the covariance matrix is the negative of the inverse Hessian evaluated at the mode (see Elerian et al., 2001,

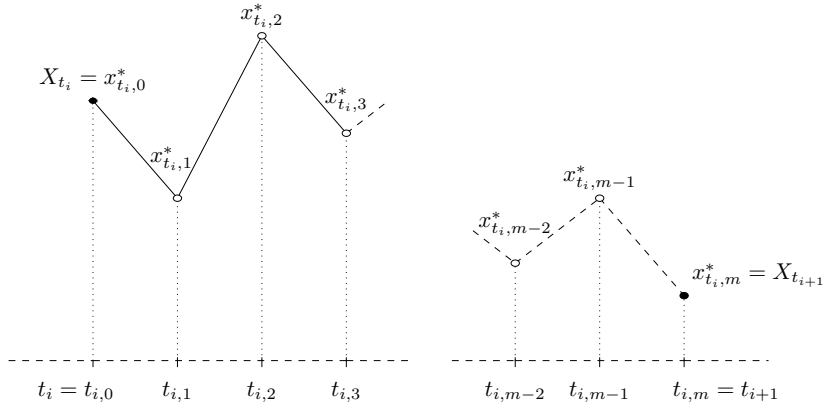


Figure 3.1: Augmented data in the discretization scheme: the observed data, X_{t_i} and $X_{t_{i+1}}$, is augmented by introducing $(m-1)$ unobserved data points.

for details regarding the gradient and Hessian matrix of the target density).

To complete one cycle of the MCMC sampler, we need to sample $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta} \mid X, X^*)$ conditioned on the augmented sample, both the observed states X and the simulated auxiliary states X^*. Assuming a non informative prior, the likelihood of the augmented sample, under the Euler-Maruyama discretization scheme, is

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=0}^{n-1} \left[\prod_{j=0}^{M} \frac{1}{\sqrt{2\pi\, m^2(x^*_{t_i,j}, \boldsymbol{\theta})\Delta}} \exp\left(- \frac{\left[x^*_{t_i,j+1} - x^*_{t_i,j} - m(x^*_{t_i,j}, \boldsymbol{\theta})\Delta\right]^2}{2\sigma^2(x^*_{t_i,j}, \boldsymbol{\theta})\Delta} \right) \right]. \quad (3.14)$$

This method shows accuracy and can be extended to multi-dimensional models—at the cost of further computational demand—and to partially observed processes, such as stochastic volatility models. The main drawbacks are related to its model-specific nature and its more troublesome implementation. Moreover, determining the convergence of the algorithm is not straightforward, nor is the number of parameter draws and the initial iterations to be discarded.

3.2.6 Generalized method of moments

The generalized method of moments (GMM), introduced by Hansen (1982), is a special case of minimum distance estimation based in moment conditions. Let $\boldsymbol{\theta}$ be the vector of parameters, we denote $f_t(\boldsymbol{\theta})$ as

$$f_t(\boldsymbol{\theta}) = \boldsymbol{u}_t \otimes \boldsymbol{z}_t, \quad (3.15)$$

where $\boldsymbol{u}_t$ is an unobservable vector of disturbance terms, $\boldsymbol{z}_t$ is a vector of instrumental variables and "$\otimes$" denotes the Kronecker product. We obtain the moment conditions assuming

that the error term is uncorrelated with the instrumental variables, thus we have the *orthogonality conditions* $\mathbb{E}\left[f_t(\boldsymbol{\theta})\right] = 0$. Replacing the theoretical moment condition $\mathbb{E}\left[f_t(\boldsymbol{\theta})\right]$ with its sample counterpart,

$$g_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} f_{t_i}(\boldsymbol{\theta}),$$

the GMM estimator is one that minimizes a squared Euclidean distance of sample moments from their population counterpart of zero, given by the quadratic form

$$Q_n(\boldsymbol{\theta}) = g_n(\boldsymbol{\theta})^\top W_n(\boldsymbol{\theta}) g_n(\boldsymbol{\theta}), \tag{3.16}$$

where $W_n(\boldsymbol{\theta})$ is a positive semi-definite weighting matrix. Choosing $W_n(\boldsymbol{\theta}) = S^{-1}(\boldsymbol{\theta})$, where $S(\boldsymbol{\theta}) = \mathbb{E}\left[f_t(\boldsymbol{\theta}) f_t^\top(\boldsymbol{\theta})\right]$, gives the GMM estimator of $\boldsymbol{\theta}$ with the smallest asymptotic covariance matrix (see Hansen, 1982). For a weighting matrix $W_n(\boldsymbol{\theta})$, the GMM estimator is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}).$$

We assume that there are, at least, as many moment functions as parameters and that on the compact parameter space Θ,

$$\mathbb{E}\left[f_t(\boldsymbol{\theta})\right] = 0 \quad \text{ if, and only if } \ \boldsymbol{\theta} = \boldsymbol{\theta}_0,$$

to achieve the identification condition for consistency of the GMM estimator. Consistency results and conditions for asymptotic normality are given in Newey and McFadden (1994, Theorems 2.6 and 3.1). When $W(\boldsymbol{\theta}) = S^{-1}(\boldsymbol{\theta})$, we have that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\big(0, V\big),$$

where $V = \left(G^\top W(\boldsymbol{\theta}) G\right)^{-1}$, with $G = \partial \mathbb{E}\left[f_t(\boldsymbol{\theta}_0)\right]/\partial\boldsymbol{\theta}$.

To implement the GMM for the diffusion model in (3.1), we can consider the discretized process

$$X_{t_{i+1}} = X_{t_i} + m(X_{t_i}, \boldsymbol{\theta})\Delta + \varepsilon_{t_{i+1}}.$$

The error term can be defined as $\varepsilon_{t_{i+1}} = X_{t_{i+1}} - \mathbb{E}\left[X_{t_{i+1}} \mid \mathcal{F}_{t_i}\right] = X_{t_{i+1}} - X_{t_i} - m(X_{t_i}, \boldsymbol{\theta})\Delta$, and the first and second moments under the time period $\Delta = t_{i+1} - t_i$ are

$$\mathbb{E}\left[\varepsilon_{t_{i+1}} \mid \mathcal{F}_{t_i}\right] = 0, \qquad \mathbb{E}\left[\varepsilon_{t_{i+1}}^2 \mid \mathcal{F}_{t_i}\right] = \sigma(X_{t_i}, \boldsymbol{\theta})^2\Delta,$$

respectively. Due to the independence of increments property of the Wiener process, we can

define (3.15) as the moments vector

$$f_t(\boldsymbol{\theta}) = \begin{bmatrix} \varepsilon_{t_{i+1}} \\ \varepsilon_{t_{i+1}}^2 - \mathbb{E}\left[\varepsilon_{t_{i+1}}^2 \mid \mathcal{F}_{t_i}\right] \end{bmatrix} \otimes \begin{bmatrix} 1 \\ X_{t_i} \end{bmatrix},$$

therefore, we have the orthogonality condition $\mathbb{E}\left[f_t(\boldsymbol{\theta})\right] = 0$ to construct the GMM estimator of $\boldsymbol{\theta}$. In the rare cases that true moments are known, they should be used instead of their discretized counterpart to avoid discretization bias.

The GMM has a more flexible framework than maximum likelihood methods, as no prior knowledge of the transition density of the SDE is assumed. The simple empirical implementation of method-of-moments type estimators, along a rather low computational cost, have motivated the development of related methods, for instance the GMM. However, these procedures have poor finite sample properties and if moment conditions provide weak parameter identification, which can happen with highly persistent time series, estimates are subject to large finite sample bias. Besides, the occurrence of local minima in the quadratic form (3.16) can likely lead to optimization problems.

3.3 Simulation study

In this section, a simulation study is conducted to compare the performance of the estimation methods described in this chapter. The different settings were designed to (i) assess the role of persistence in estimation bias, (ii) compare the accuracy of the procedures, (iii) confirm that $T \to \infty$ plays the role of the discrete-time standard asymptotic of $n \to \infty$, (iv) examine discretization bias and the impact of different sampling frequencies Δ, and (v) study the effect of volatility on the estimators performance.

3.3.1 Experimental design

We consider two models, one proposed by Vasicek (1977) based in the Uhlenbeck and Ornstein (1930) process and the CKLS proposed by Chan et al. (1992). The latter lacks a tractable likelihood function, while the former admits a closed-form expression for transition and marginal density, which allows us to perform exact maximum likelihood estimation and avoid simulation errors sampling directly from the continuous time model. The Vasicek model is given by

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

Vasicek	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**
(μ, κ, σ^2)	$(0.09, 0.2, 4{\times}10^{-5})$	$(0.09, 0.2, 4{\times}10^{-4})$	$(0.09, 0.9, 1.8{\times}10^{-4})$	$(0.09, 0.9, 1.8{\times}10^{-3})$
$\mathbb{E}[X_t]$	0.090	0.090	0.090	0.090
$\mathbb{V}\mathrm{ar}[X_t]$	10^{-4}	10^{-3}	10^{-4}	10^{-3}
$\mathbb{E}\left[X_{(i+1)\Delta} \mid X_{i\Delta}\right]$	$3.4{\times}10^{-4} + 0.996X_t$	$3.4{\times}10^{-4} + 0.996X_t$	$0.002 + 0.98X_t$	$0.002 + 0.98X_t$
$\mathbb{V}\mathrm{ar}\left[X_{(i+1)\Delta} \mid X_{i\Delta}\right]$	$7.6{\times}10^{-7}$	$7.6{\times}10^{-6}$	$3.4{\times}10^{-6}$	$3.4{\times}10^{-5}$
Vasicek	**Scenario 5**	**Scenario 6**	**Scenario 7**	**Scenario 8**
(μ, κ, σ^2)	$(0.09, 5, 10^{-3})$	$(0.09, 5, 10^{-2})$	$(0.09, 9.8, 1.96{\times}10^{-3})$	$(0.09, 9.8, 1.96{\times}10^{-2})$
$\mathbb{E}[X_t]$	0.090	0.090	0.090	0.090
$\mathbb{V}\mathrm{ar}[X_t]$	10^{-4}	10^{-3}	10^{-4}	10^{-3}
$\mathbb{E}\left[X_{(i+1)\Delta} \mid X_{i\Delta}\right]$	$8.3{\times}10^{-3} + 0.908X_t$	$8.3{\times}10^{-3} + 0.908X_t$	$0.015 + 0.83X_t$	$0.015 + 0.83X_t$
$\mathbb{V}\mathrm{ar}\left[X_{(i+1)\Delta} \mid X_{i\Delta}\right]$	$1.7{\times}10^{-5}$	$1.7{\times}10^{-4}$	$3.1{\times}10^{-5}$	$3.1{\times}10^{-4}$
CKLS	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**
$(\mu, \kappa, \sigma^2, \gamma)$	$(0.09, 0.2, 0.25, 1.5)$	$(0.09, 0.2, 1, 1.5)$	$(0.09, 0.9, 0.5, 1.5)$	$(0.09, 0.9, 2, 1.5)$

Table 3.1: Scenarios for the Monte Carlo study for the Vasicek, $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$*, and CKLS,* $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t$*, models with* $\Delta = 1/52$*.*

and the CKLS is

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t, \tag{3.17}$$

where μ, κ and σ are positive constants and $X_{t_0} = x_0$ is the initial condition. The parameter μ represents the long time mean, κ is the speed of mean reversion, σ is the standard deviation of volatility and γ is the elasticity of variance.

The Monte Carlo setup consist in the Vasicek and CKLS models, for a low mean reversion scenario (*high persistent dependence*) and a high mean reversion scenario (*low persistent dependence*). As it is common to record X_t annualized, we use weekly ($\Delta = 1/52$) and monthly ($\Delta = 1/12$) frequency. For each case, we also consider two different volatility scenarios, with increasing unconditional variance.

A thousand realizations of random sample paths $\{X_{i\Delta}\}_{i=1}^{n}$ are generated for sample sizes $n \in \{520, 2\,600\}$, with $\Delta = 1/52$, which corresponds to weekly data on an observation window of $T = 10$ and 50 years, respectively, along with sample paths with $n = 520$ and $\Delta = 1/12$, which corresponds to monthly data for, approximately, 43 years. With this design we can evaluate the performance of the estimation methods with a larger sample size n and when the total observation time T is increased, keeping the sample size constant. In addition, the different values of Δ could give rise to discretization bias, which can appear jointly with estimation bias in those methods that rely on discretization schemes.

Table 3.1 shows the eight simulated scenarios for the Vasicek model, along with uncondi-

tional and conditional mean and variance, as the analytical density is available. The first two scenarios correspond to the low mean reversion, with increasing unconditional volatility, and the third and fourth are the high mean reversion cases. For the high persistent dependence scenarios (1 and 2) the parameter values are $\boldsymbol{\theta}_1 = (\mu, \kappa, \sigma^2)^\top = (0.09, 0.2, 4 \times 10^{-5})^\top$ and $\boldsymbol{\theta}_2 = (0.09, 0.2, 4 \times 10^{-4})^\top$, and for the low persistence (scenarios 3 and 4) the parameter are $\boldsymbol{\theta}_3 = (0.09, 0.9, 1.8 \times 10^{-4})^\top$ and $\boldsymbol{\theta}_4 = (0.09, 0.9, 1.8 \times 10^{-3})^\top$. The marginal density was kept unchanged while varying the speed of mean reversion, therefore scenarios 1 and 3 and scenarios 2 and 4 have the same marginal density, as illustrated in Figure A.1a. The settings of scenarios 1 and 2 allow us to check the performance in a near unit-root case, as the regressive coefficient of $\mathbb{E}\left[X_{t_{i+1}} \mid X_{t_i}\right]$ is 0.996. Scenarios 5–8, although unrelated to real interest rate processes, are limiting cases to quantify the impact of persistence separately from changes in the marginal density. In scenarios 5 and 6 (7 and 8), the mean-reverting force is increased by a factor of 25 (49) from scenarios 1 and 2 and σ is changed in the same proportion to hold the marginal density fixed.

In regards to the scenarios for the CKLS model, a similar scheme was set, as shown in Table 3.1, keeping for all scenarios the same parameters for the drift function as the Vasicek. First two scenarios correspond to high persistent dependence (or low mean reversion) with parameter values $\boldsymbol{\theta}_1 = (\mu, \kappa, \sigma^2, \gamma)^\top = (0.09, 0.2, 0.25, 1.5)^\top$ and $\boldsymbol{\theta}_2 = (0.09, 0.2, 1, 1.5)^\top$, and for the low persistence (scenarios 3 and 4) the parameter are $\boldsymbol{\theta}_3 = (0.09, 0.9, 0.5, 1.5)^\top$ and $\boldsymbol{\theta}_4 = (0.09, 0.9, 2, 1.5)^\top$. Figure A.1b illustrates the stationary density for scenarios 1 and 2, where the parameter σ^2 was increased by a factor of 4, just like from scenario 3 to 4.

Tables 3.2–3.3 show the simulation study for the Vasicek model (along with Tables A.1–A.4, deferred to Appendix A), while Tables 3.4–3.5 (A.5–A.6) refer to the CKLS model. The mean of the parameter estimates for the $1\,000$ replications of the experiment is included, as well as standard deviation (SD) and root mean squared error (RMSE) for the estimation methods in Section 3.2:

(i) Exact maximum likelihood (EML)

(ii) Euler method (DML)

(iii) Local linearization (LL)

(iv) Hermite polynomial expansion (HP)

(v) Generalized Method of Moments (GMM)

(vi) Kalman Filter (KF)

(vii) Markov Chain Monte Carlo (MCMC)

Scenario 1	θ	EML	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	0.0915	0.0915	0.0915	0.0915	0.0915	**0.0913**	0.0917
	$\hat{\kappa}$	0.6762	0.6703	0.6696	0.6720	**0.6696**	0.6789	0.6701
	$\hat{\sigma}$	0.0063	0.0063	0.0063	0.0063	**0.0063**	0.0063	0.0063
SD	$\hat{\mu}$	0.0177	0.0180	0.0179	0.0177	0.0179	**0.0148**	0.0178
	$\hat{\kappa}$	0.4579	**0.4511**	0.4518	0.4548	0.4518	0.4609	0.4518
	$\hat{\sigma}$	1.98×10^{-4}	1.97×10^{-4}	1.97×10^{-4}	1.98×10^{-4}	$\mathbf{1.97\times10^{-4}}$	1.99×10^{-4}	2.04×10^{-4}
RMSE	$\hat{\mu}$	0.0178	0.0180	0.0180	0.0178	0.0180	**0.0148**	0.0178
	$\hat{\kappa}$	0.6606	0.6517	0.6517	0.6554	**0.6517**	0.6647	0.6520
	$\hat{\sigma}$	1.99×10^{-4}	1.99×10^{-4}	1.98×10^{-4}	1.99×10^{-4}	$\mathbf{1.98\times10^{-4}}$	2.01×10^{-4}	2.07×10^{-4}
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	**0.0900**	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900
	$\hat{\kappa}$	0.2865	0.2837	0.2836	0.2864	**0.2836**	0.2847	0.2837
	$\hat{\sigma}$	**0.0063**	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063
SD	$\hat{\mu}$	**0.0046**	0.0046	0.0046	0.0046	0.0046	0.0046	0.0046
	$\hat{\kappa}$	0.1274	0.1262	0.1261	0.1271	**0.1261**	0.1270	0.1261
	$\hat{\sigma}$	8.30×10^{-5}	8.29×10^{-5}	$\mathbf{8.29\times10^{-5}}$	8.31×10^{-5}	$\mathbf{8.29\times10^{-5}}$	8.33×10^{-5}	8.31×10^{-5}
RMSE	$\hat{\mu}$	**0.0046**	0.0046	0.0046	0.0046	0.0046	0.0046	0.0046
	$\hat{\kappa}$	0.1540	0.1514	0.1513	0.1537	**0.1513**	0.1527	0.1513
	$\hat{\sigma}$	$\mathbf{8.33\times10^{-5}}$	8.37×10^{-5}	8.34×10^{-5}	8.33×10^{-5}	8.34×10^{-5}	8.36×10^{-5}	8.41×10^{-5}
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	**0.0902**	0.0902	0.0902	0.0902	0.0902	0.0902	0.0902
	$\hat{\kappa}$	0.2973	**0.2923**	0.2924	0.2939	0.2924	0.2978	0.2932
	$\hat{\sigma}$	**0.0063**	0.0063	0.0063	0.0063	0.0063	0.0063	0.0063
SD	$\hat{\mu}$	**0.0050**	0.0050	0.0050	0.0050	0.0050	0.0050	0.0051
	$\hat{\kappa}$	0.1377	**0.1345**	0.1346	0.1349	0.1346	0.1389	0.1347
	$\hat{\sigma}$	1.99×10^{-4}	1.95×10^{-4}	$\mathbf{1.95\times10^{-4}}$	1.99×10^{-4}	1.95×10^{-4}	1.98×10^{-4}	2.05×10^{-4}
RMSE	$\hat{\mu}$	**0.0050**	0.0050	0.0050	0.0050	0.0050	0.0050	0.0051
	$\hat{\kappa}$	0.1686	**0.1632**	0.1633	0.1644	0.1633	0.1698	0.1638
	$\hat{\sigma}$	$\mathbf{1.99\times10^{-4}}$	2.06×10^{-4}	2.03×10^{-4}	1.99×10^{-4}	2.03×10^{-4}	1.99×10^{-4}	2.15×10^{-4}

Table 3.2: Monte Carlo simulation for Vasicek model, $(\mu, \kappa, \sigma)^\top = (0.09, 0.2, 0.00632)^\top$, scenario 1: low mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

3.3.2 Implementation details

The simulated sample paths for the Vasicek model were constructed from the closed-form transition density and for the CKLS model the Milstein scheme was used. To reduce discretization bias, paths were generated with daily frequency ($\Delta = 1/364$) and subsamples were taken on a weekly ($\Delta = 1/52$) or monthly ($\Delta = 1/12$) basis. The initial condition X_{t_0} was set to be the average and the first $1\,000$ data were discarded, as a burn-in period to remove the dependence on the initial value.

As in the CKLS model a closed form expression for the transition density is not available,

Scenario 3	θ	EML	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	$\mathbf{0.0902}$	0.0902	0.0902	0.0902	0.0902	0.0902	0.0902
	$\hat{\kappa}$	1.3193	$\mathbf{1.2982}$	1.2984	1.3063	1.2984	1.3224	1.3127
	$\hat{\sigma}$	0.0134	0.0133	0.0133	$\mathbf{0.0134}$	0.0133	0.0135	0.0133
SD	$\hat{\mu}$	$\mathbf{0.0049}$	0.0049	0.0049	0.0049	0.0049	0.0049	0.0051
	$\hat{\kappa}$	0.6031	$\mathbf{0.5884}$	0.5888	0.5893	0.5888	0.6077	0.5937
	$\hat{\sigma}$	4.21×10^{-4}	4.14×10^{-4}	$\mathbf{4.14\times10^{-4}}$	4.21×10^{-4}	4.14×10^{-4}	4.21×10^{-4}	4.47×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0049}$	0.0049	0.0049	0.0049	0.0049	0.0049	0.0051
	$\hat{\kappa}$	0.7346	$\mathbf{0.7105}$	0.7110	0.7158	0.7110	0.7401	0.7231
	$\hat{\sigma}$	4.22×10^{-4}	4.38×10^{-4}	4.31×10^{-4}	$\mathbf{4.22\times10^{-4}}$	4.31×10^{-4}	4.22×10^{-4}	4.60×10^{-4}
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	$\mathbf{0.0900}$	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900
	$\hat{\kappa}$	0.9782	0.9637	$\mathbf{0.9635}$	0.9737	0.9635	0.9724	0.9646
	$\hat{\sigma}$	$\mathbf{0.0134}$	0.0133	0.0133	$\mathbf{0.0134}$	0.0133	0.0134	0.0133
SD	$\hat{\mu}$	$\mathbf{0.0022}$	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022
	$\hat{\kappa}$	0.2116	0.2088	$\mathbf{0.2085}$	0.2088	0.2085	0.2125	0.2098
	$\hat{\sigma}$	1.77×10^{-4}	1.75×10^{-4}	$\mathbf{1.75\times10^{-4}}$	1.77×10^{-4}	1.75×10^{-4}	1.77×10^{-4}	1.78×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0022}$	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022
	$\hat{\kappa}$	0.2256	0.2183	$\mathbf{0.2180}$	0.2215	0.2180	0.2245	0.2195
	$\hat{\sigma}$	1.78×10^{-4}	2.08×10^{-4}	2.05×10^{-4}	$\mathbf{1.77\times10^{-4}}$	2.05×10^{-4}	1.77×10^{-4}	1.95×10^{-4}
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	$\mathbf{0.0901}$	0.0901	0.0901	0.0901	0.0901	0.0901	0.0901
	$\hat{\kappa}$	0.9851	$\mathbf{0.9462}$	0.9465	0.9664	0.9465	0.9868	0.9704
	$\hat{\sigma}$	0.0134	0.0129	0.0129	$\mathbf{0.0134}$	0.0129	0.0134	0.0132
SD	$\hat{\mu}$	$\mathbf{0.0023}$	0.0023	0.0023	0.0023	0.0023	0.0023	0.0025
	$\hat{\kappa}$	0.2362	0.2182	$\mathbf{0.2178}$	0.237	0.2178	0.2358	0.2365
	$\hat{\sigma}$	4.34×10^{-4}	$\mathbf{4.02\times10^{-4}}$	4.03×10^{-4}	4.32×10^{-4}	4.03×10^{-4}	4.26×10^{-4}	4.95×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0023}$	0.0023	0.0023	0.0023	0.0023	0.0023	0.0025
	$\hat{\kappa}$	0.2511	0.2230	$\mathbf{0.2227}$	0.2461	0.2227	0.2513	0.2467
	$\hat{\sigma}$	4.35×10^{-4}	6.52×10^{-4}	6.33×10^{-4}	$\mathbf{4.32\times10^{-4}}$	6.33×10^{-4}	4.28×10^{-4}	5.53×10^{-4}

Table 3.3: Monte Carlo simulation for Vasicek model, $(\mu, \kappa, \sigma)^\top = (0.09, 0.9, 0.0134)^\top$, scenario 3: high mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

the exact maximum likelihood method is not included, as it is unfeasible. The GMM was implemented with four moment conditions, where the first two identify the marginal distribution and the higher order are nonlinear functions of the first two moments. As the Vasicek model does have a closed form for the transition density, the true moments were used. Regarding the MCMC setup, for the Vasicek model the algorithm is iterated $2\,500$ times and the first 500 iterations were discarded. The iterations were increased for the CKLS model as there is an additional parameter to sample, thus $5\,000$ iterations were executed and the first $1\,000$ were discarded. For both models, $m = 5$ was fixed in the data augmentation step. For all

Scenario 1	θ	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$							
Mean	$\hat{\mu}$	**0.0950**	0.0953	0.0967	0.0956	0.0963	0.1001
	$\hat{\kappa}$	0.6596	0.6683	**0.6438**	0.6590	0.6688	0.7208
	$\hat{\sigma}$	0.5776	0.5985	0.6331	0.5774	0.6860	**0.5534**
	$\hat{\gamma}$	1.4904	1.5052	1.5130	1.4905	**1.4961**	1.4605
SD	$\hat{\mu}$	**0.0420**	0.0483	0.0547	0.0518	0.0730	0.2047
	$\hat{\kappa}$	0.4630	0.4766	0.4649	**0.4615**	0.4729	0.4675
	$\hat{\sigma}$	0.3783	0.3938	0.4799	**0.3778**	0.5906	0.3965
	$\hat{\gamma}$	0.2453	0.2459	0.2588	**0.2446**	0.2913	0.2644
RMSE	$\hat{\mu}$	**0.0423**	0.0486	0.0551	0.0522	0.0733	0.2050
	$\hat{\kappa}$	0.6524	0.6681	**0.6427**	0.6509	0.6659	0.6998
	$\hat{\sigma}$	0.3862	0.4060	0.4980	**0.3857**	0.6192	0.4001
	$\hat{\gamma}$	0.2455	0.2460	0.2591	**0.2447**	0.2913	0.2673
Weekly sampling ($\Delta = 1/52$) and $n = 2600$							
Mean	$\hat{\mu}$	0.0912	0.0911	0.0912	0.0912	0.0912	**0.0904**
	$\hat{\kappa}$	**0.2816**	0.2829	0.2828	0.2820	0.2828	0.3216
	$\hat{\sigma}$	0.5015	0.5095	0.5078	**0.5012**	0.5169	0.4972
	$\hat{\gamma}$	1.4956	1.5025	**1.4997**	1.4954	1.5004	1.4868
SD	$\hat{\mu}$	0.0131	0.0131	0.0131	0.0131	0.0131	**0.0114**
	$\hat{\kappa}$	0.1356	0.1368	0.1378	0.1357	0.1365	**0.1353**
	$\hat{\sigma}$	0.0865	0.0875	0.0885	**0.0863**	0.1078	0.1166
	$\hat{\gamma}$	0.0701	**0.0694**	0.0711	0.0700	0.0832	0.0954
RMSE	$\hat{\mu}$	0.0131	0.0131	0.0131	0.0131	0.0131	**0.0114**
	$\hat{\kappa}$	**0.1583**	0.1599	0.1608	0.1586	0.1597	0.1819
	$\hat{\sigma}$	0.0865	0.0880	0.0888	**0.0863**	0.1091	0.1167
	$\hat{\gamma}$	0.0702	**0.0695**	0.0711	0.0702	0.0832	0.0963
Monthly sampling ($\Delta = 1/12$) and $n = 520$							
Mean	$\hat{\mu}$	0.0930	0.0929	0.0936	0.0930	0.0930	**0.0919**
	$\hat{\kappa}$	**0.2844**	0.2892	0.2891	0.2845	0.2888	0.3248
	$\hat{\sigma}$	0.4722	0.5055	**0.5043**	0.4722	0.5395	0.4498
	$\hat{\gamma}$	1.4543	**1.4846**	1.4783	1.4543	1.4739	1.4196
SD	$\hat{\mu}$	0.0145	**0.0143**	0.0236	0.0145	0.0144	0.0153
	$\hat{\kappa}$	0.1446	0.1497	0.1516	0.1448	0.1492	**0.1437**
	$\hat{\sigma}$	0.1690	0.1785	0.1814	**0.1688**	0.2521	0.2068
	$\hat{\gamma}$	0.1466	**0.1434**	0.1466	0.1464	0.1702	0.1803
RMSE	$\hat{\mu}$	0.0148	**0.0146**	0.0239	0.0148	0.0147	0.0154
	$\hat{\kappa}$	**0.1674**	0.1743	0.1758	0.1677	0.1736	0.1904
	$\hat{\sigma}$	0.1713	0.1786	0.1815	**0.1711**	0.2552	0.2128
	$\hat{\gamma}$	0.1535	**0.1442**	0.1482	0.1534	0.1721	0.1974

Table 3.4: Monte Carlo simulation for CKLS model, $(\mu, \kappa, \sigma, \gamma)^\top = (0.09, 0.2, 0.5, 1.5)^\top$, scenario 1: low mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

Scenario 3	θ	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$							
Mean	$\hat{\mu}$	0.0908	0.0908	0.0909	0.0908	0.0909	**0.0905**
	$\hat{\kappa}$	1.2893	1.3112	1.2927	**1.2882**	1.3121	1.3760
	$\hat{\sigma}$	0.7701	0.8260	0.9005	0.7711	0.9479	**0.7452**
	$\hat{\gamma}$	1.4769	**1.5073**	1.5281	1.4773	1.4900	1.4486
SD	$\hat{\mu}$	0.0086	0.0087	0.0089	0.0086	0.0089	**0.0075**
	$\hat{\kappa}$	0.6072	0.6277	0.6182	**0.6059**	0.6274	0.6121
	$\hat{\sigma}$	**0.4582**	0.4938	0.6338	0.4605	0.8601	0.5017
	$\hat{\gamma}$	0.2277	**0.2268**	0.2332	0.2278	0.2805	0.2514
RMSE	$\hat{\mu}$	0.0087	0.0087	0.0089	0.0086	0.0090	**0.0075**
	$\hat{\kappa}$	0.7213	0.7504	0.7324	**0.7197**	0.7506	0.7755
	$\hat{\sigma}$	**0.4625**	0.5080	0.6627	0.4650	0.8932	0.5031
	$\hat{\gamma}$	0.2289	**0.2270**	0.2349	0.2289	0.2807	0.2566
Weekly sampling ($\Delta = 1/52$) and $n = 2600$							
Mean	$\hat{\mu}$	0.0901	**0.0900**	0.0902	0.0901	0.0901	0.0900
	$\hat{\kappa}$	**0.9623**	0.9777	0.9688	0.9629	0.9717	1.0023
	$\hat{\sigma}$	**0.6958**	0.7421	0.7396	0.6948	0.7402	0.6837
	$\hat{\gamma}$	1.4863	1.5142	1.5050	1.4857	**1.4985**	1.4732
SD	$\hat{\mu}$	0.0032	**0.0032**	0.0043	0.0032	0.0032	0.0032
	$\hat{\kappa}$	**0.2200**	0.2244	0.2268	0.2209	0.2249	0.2400
	$\hat{\sigma}$	0.1523	0.1564	0.2030	**0.1516**	0.2014	0.1913
	$\hat{\gamma}$	0.0896	**0.0868**	0.1007	0.0895	0.1099	0.1128
RMSE	$\hat{\mu}$	0.0032	**0.0032**	0.0043	0.0032	0.0032	0.0032
	$\hat{\kappa}$	**0.2287**	0.2375	0.2370	0.2297	0.2361	0.2608
	$\hat{\sigma}$	0.1527	0.1603	0.2056	**0.1521**	0.2041	0.1928
	$\hat{\gamma}$	0.0907	**0.0880**	0.1008	0.0906	0.1099	0.1160
Monthly sampling ($\Delta = 1/12$) and $n = 520$							
Mean	$\hat{\mu}$	0.0902	**0.0901**	0.0904	0.0902	0.0903	0.0902
	$\hat{\kappa}$	**0.9481**	0.9917	0.9738	0.9481	0.9852	0.9759
	$\hat{\sigma}$	**0.6411**	0.7908	0.8134	0.6407	0.8705	0.5984
	$\hat{\gamma}$	1.4251	1.5156	**1.5052**	1.4249	1.4894	1.3820
SD	$\hat{\mu}$	0.0034	0.0035	0.0043	**0.0034**	0.0035	0.0034
	$\hat{\kappa}$	**0.2304**	0.2514	0.2565	0.2304	0.2498	0.2450
	$\hat{\sigma}$	0.3094	0.3763	0.4269	**0.3089**	0.5436	0.3535
	$\hat{\gamma}$	0.2005	**0.1999**	0.2055	0.2001	0.2406	0.2253
RMSE	$\hat{\mu}$	0.0034	0.0035	0.0043	**0.0034**	0.0035	0.0034
	$\hat{\kappa}$	**0.2353**	0.2676	0.2669	0.2354	0.2640	0.2565
	$\hat{\sigma}$	0.3163	0.3855	0.4400	**0.3159**	0.5676	0.3698
	$\hat{\gamma}$	0.2140	**0.2005**	0.2055	0.2137	0.2409	0.2544

Table 3.5: Monte Carlo simulation for CKLS model, $(\mu, \kappa, \sigma, \gamma)^\top = (0.09, 0.9, 0.707, 1.5)^\top$, scenario 3: high mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

methods, $\boldsymbol{\theta}$ was jointly estimated. The drift parameter μ is calculated indirectly, as the drift specification was rewritten to estimate the intercept $\kappa\mu$.

Regarding optimization, we chose to minimize the negative log-likelihood function and use the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm (R Core Team, 2014), where the gradient and Hessian matrix are calculated numerically in the optimization routine. As for the choice of initial values, the approach is the following: we fit a linear regression, writing the discrete version of the SDE as in equation (3.12), to obtain a rough estimation of the parameters and set the starting values. For the CKLS diffusion parameter γ, we specified a power variance function for the heteroscedasticity structure. Implementation details regarding the Hermite polynomial expansion and MCMC are provided in Appendix A.2.

Since some of the procedures are computationally expensive, we chose to integrate C and C++ code (Eddelbuettel and François, 2011) in the R routines. Table 3.6 shows run-times for each Monte Carlo iteration, with sample size $n = 2600$, and evidences the computational cost of simulation-based techniques, such as the MCMC. The Kalman filter benefits from using a lower-level programming language like C and has low run-times, though the remainder of the methods have a similar performance in base R, being the GMM the slowest.

Time (seconds)	**EML**	**DML**	**LL**	**HP**	**KF**	**MCMC**	**GMM**
Vasicek	0.0534	0.0469	0.0424	0.4921	0.1097	50.8442	1.0108
CKLS	-	0.0818	0.2011	1.9543	0.2017	140.5385	2.0237
Implementation	R	R	R	R	C	C++	R/C++

Table 3.6: CPU time (in seconds) of the estimation methods per iteration, with $n = 2600$.

3.3.3 Discussion

Tables 3.2 and 3.4 report the estimates for scenario 1, for both Vasicek and CKLS models, which features low mean reversion and the unconditional volatility is very small, a quiet process whose estimation can be challenging. The process is nearly unit-root, which can increase the estimation bias in the drift parameter κ, as reported in Yu and Phillips (2001). This large bias in the estimation of the drift parameter κ, which controls the speed of mean reversion, is encountered in both tables with small sample size n through all estimation procedures, with more than a 200% relative bias (see Figure 3.3). The RMSE in both models for κ is similar, the other parameters estimates have small RMSE but the parameters in the diffusion function incur in more bias in the CKLS model, as this model is relatively hard to identify as it may yield similar volatility functions for different values of σ and γ, while the Vasicek model has

a simpler diffusion function. Both bias and RMSE decrease as n and the total observation time T are increased, especially for κ. As opposed to discrete time series, where the sample size n controls the estimation error, is T who defines the bias and variance of the estimations: the last two columns (weekly and monthly frequency) have similar observation time T but different sample size n, and RMSEs are close. Exact ML is available for the Vasicek model, thus avoiding discretization bias, but nevertheless biases are homogeneous for all methods, with GMM presenting a slightly higher RMSE.

When volatility is increased while keeping the same drift function (Tables A.1 and A.5), bias is moderately increased for κ and variance is higher than in scenario 1 for the Vasicek model, while reducing for the diffusion parameters in the CKLS model, specially γ. All biases shrink as the sample size and observation time increases. The GMM method shows less efficiency, with a larger RMSE in the CKLS model.

The estimation bias of the drift parameter κ increases when the diffusion process has an absence of dynamics, which happens when κ is small. Scenarios 1 and 2 have a mean reversion parameter κ closer to zero, while scenarios 3 (Tables 3.3 and 3.5) and 4 (Tables A.2 and A.6) show high mean reversion. For the low volatility case, scenario 3, the finite sample bias of κ is significantly smaller (close to 40%, see Figure 3.3) than the high persistence dependence scenarios for both models and estimation bias and variance of μ is also reduced. For the CKLS model, the estimation bias of σ is reduced and the estimation error of γ is similar to scenario 1. RMSE and bias diminish with increasing n and T, with less RMSE in the weekly scenario for the CKLS model compared with the monthly simulation, which may be due to the discretization error. The higher volatility scenario shows a similar behavior, with smaller bias in the drift parameters than low mean reversion scenarios, but moderately higher in μ with respect to scenario 3, and the estimation errors for the diffusion parameters of the CKLS model are inferior compared to scenario 3.

Scenario 6[1] (Table A.3) has a mean-reverting force five times higher than scenario 2 while keeping the same marginal distribution. In this setting, the estimation bias shrinks substantially in all schemes, as we are departing from the unit-root case. However, the discretization bias starts arising, as shown in the estimations of DML and KF methods with monthly frequency. This is magnified in the limiting case of scenario 8 (Table A.4), where the estimation bias is very small in all schemes, but the discretization bias in DML, KF and, less so, in LL for κ and σ is large, mostly in the coarser discretization (monthly), noticeably underestimating both parameters[2]. Figure 3.2 illustrates true (black) and discretized (gray) log-likelihood for scenarios 1, 3, 6 and 8, where departures from the true log-likelihood are larger for lower

[1]Scenario 5 was not included, as the results were very close and so too were the conclusions drawn.

[2]The performance of the methods are analogous in scenario 7, therefore results were not reported.

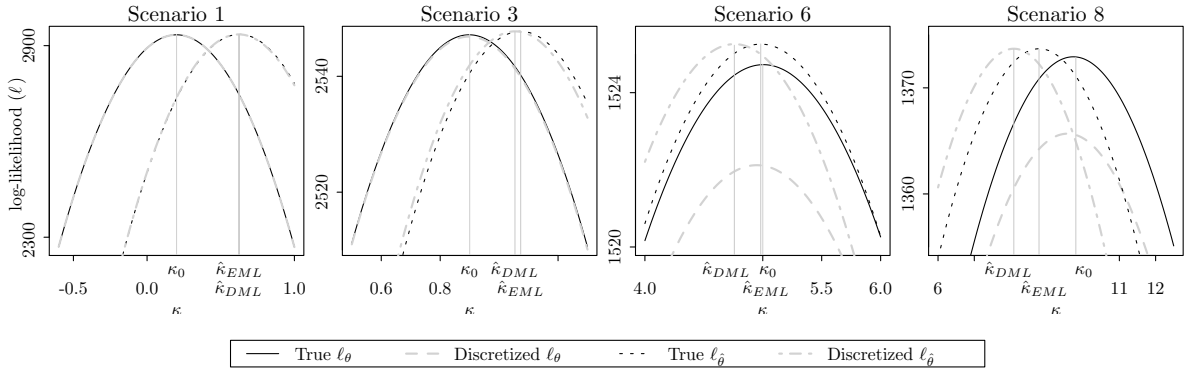


Figure 3.2: Vasicek model log-likelihood (ℓ) for the drift parameter κ with known ($\ell_{\boldsymbol{\theta}}$) and estimated ($\ell_{\hat{\boldsymbol{\theta}}}$) parameters, where κ_0 is the true value. The true density and its discretized version ($\Delta = 1/52$) is illustrated, along with the estimate obtained ($\hat{\kappa}_{EML}$ and $\hat{\kappa}_{DML}$, respectively).

mean reversion scenarios, while scenarios 6 and 8 display discretization bias.

The analysis of the Monte Carlo evidence reveals the following insights:

(i) The dynamic of the process is governed by the drift parameter κ, which determines the persistence of the process by controlling the reversion towards the unconditional mean. As $\kappa \to 0$, mean reversion goes to zero and correlation between observations approaches one. This increases persistence, which introduces sample bias in parametric estimation (Ball and Torous, 1996). Simulations show that increasing κ lowered persistence and, therefore, estimation bias was also diminished (see Figure 3.3). High persistence scenarios, near unit-root cases, revealed significant estimation bias in the drift parameter κ, but almost negligible in the diffusion parameter σ.

(ii) Increasing the volatility parameters had minor effect on the estimators performance. In the Vasicek model, higher values in the volatility parameter slightly increased RMSE in the estimation of σ. On the other hand, the estimation of the parameters in the CKLS diffusion function benefit from richer volatility dynamics, reducing RMSE.

(iii) In discrete time series, the bias and variance of estimators is controlled by the sample size n, so that they reduce as $n \to \infty$. In continuous-time models sampled at discrete time points, bias and variance in the estimation of the drift parameter κ is dominated by the total observation time $T = n\Delta$. Under quite general conditions, the estimators of the drift parameters are of order $\mathcal{O}(T^{-1})$, while the diffusion parameter σ is of order n^{-1} (Tang and Chen, 2009). The simulated scenarios corroborate this, as estimation bias with $T = 50$ and 43 years were close despite the different frequency (weekly and monthly, respectively) and sample size n (2600 and 520, respectively).

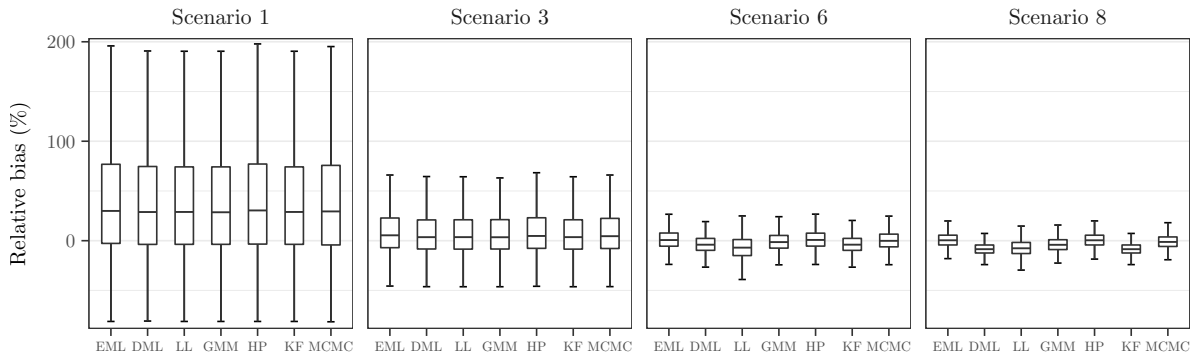


Figure 3.3: Relative bias (in percentage) for the drift parameter κ with weekly data ($\Delta = 1/52$) and $n = 2\,600$.

(iv) Discretization bias arises in DML and KF methods in scenarios with low sampling frequency and low persistence (see Figure 3.2), and correcting the DML estimates with local linearization does not always correct the bias and both κ and σ are underestimated.

(v) There appears to be similar estimation bias in the drift parameters for both Vasicek and CKLS models. However, the more flexible parametric form of the CKLS volatility function makes estimation more challenging, and bias and RMSE for those parameters are higher than for the Vasicek model.

(vi) Regarding efficiency, as exact ML is available for the Vasicek model, it can be regarded as a benchmark. Overall, the estimations of the parameters are close to the EML performance, being the GMM the less efficient. The estimations differ when discretization bias arises.

Table 3.7 provides a summary of the properties and finite sample performance of the methods. Regarding accuracy, differences among the methods are not always clear and the context of application (e.g., the sampling interval) should be considered when choosing the estimation procedure. The HP shows the best trade-off between efficiency and speed—followed by the LL method—and MCMC exhibits good accuracy through all scenarios, however, the inherent time consuming implementation (methodologically and computationally) are major disadvantages. The performance of simpler discretization schemes, like the ones used in DML and KF, is conditioned on the sampling interval and should be avoided for large Δ, otherwise, their efficiency is close to the other alternatives (see Figure 3.3). The generalized method of moments was outperformed in the majority of scenarios, notably with higher dimensions of the parameter vector $\boldsymbol{\theta}$.

Method	Authors	Asymptotic Properties	Finite Sample Performance
DML	Florens-Zmirou (1989)	Asymptotically normal and consistent	Biased when Δ is large
LL	Ozaki (1992) Shoji and Ozaki (1998)	As ML estimators	Outperforms discrete ML and KF
HP	Aït-Sahalia (1999, 2002)	Asymptotically normal and consistent (Aït-Sahalia, 2002)	Outperforms LL
KF	Kalman (1960,1963)	Asymptotically normal and consistent (Caines, 1988; Pagan, 1980)	Similar to discrete ML
MCMC	Elerian et al. (2001) Eraker (2001)	Simulation based	Efficient but the most computationally intensive
GMM	Hansen (1982) Chan et al. (1992)	Asymptotically normal and consistent (Hansen, 1982; Newey and McFadden, 1994)	The least efficient and efficiency depends on moments

Table 3.7: Summary of estimation procedures for SDE parameters.

3.4 Application to Euribor series

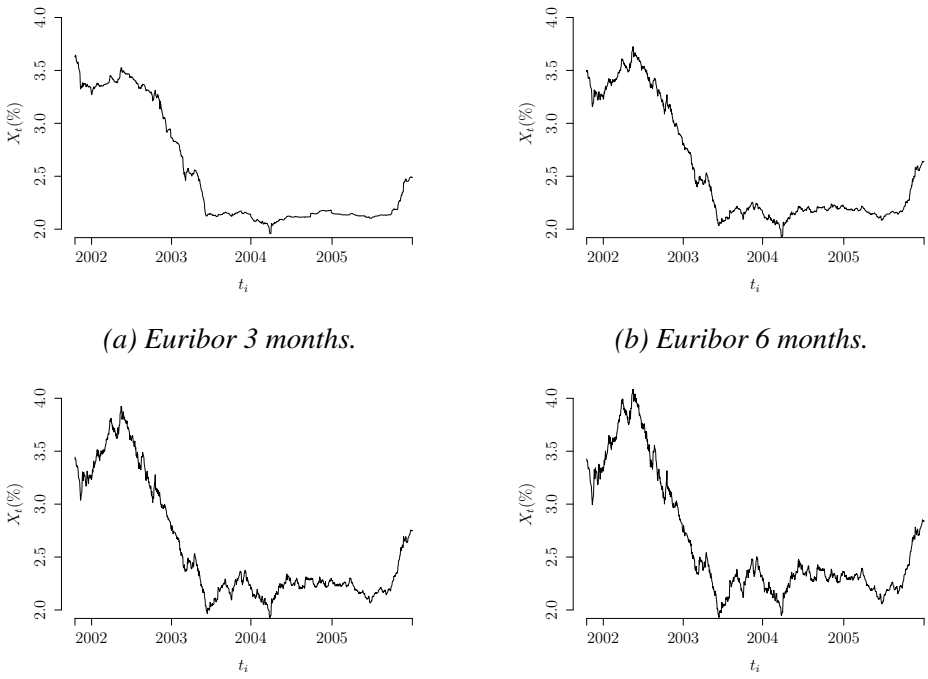
In this section, we consider four data sets corresponding to four maturities (three, six, nine and twelve months) of the Euribor (Euro Interbank Offered Rate) interest rate series. This daily series expand from October 15$^{\text{th}}$ 2001 to December 30$^{\text{th}}$ 2005 (sample size of $n = 1\,077$), see Figure 3.4. Numerous models have been proposed to capture the dynamics of short-term interest rate, including those by Merton (1973), Brennan and Schwartz (1979), Vasicek (1977), Cox et al. (1985) or Brennan and Schwartz (1980). As these models can be nested within the unrestricted model

$$\mathrm{d}X_t = (\theta_1 - \theta_2 X_t)\,\mathrm{d}t + \theta_3 X_t^{\theta_4}\,\mathrm{d}W_t,$$

proposed by Chan et al. (1992), we will estimate the parameters[3] with the methods presented in the previous section.

The results of the parameter estimation for the CKLS model are shown in Table 3.8, with the associated standard error in parentheses. The estimates obtained by the different approaches are considerably close, where the generalized method of moments seems to present the most noticeable disparity, mainly in the level effect parameter θ_4. This was already no-

[3]We chose a different drift parametrization of the CKLS model in (3.17) to provide standard errors for all estimates, so none of them are estimated indirectly.



(a) Euribor 3 months. *(b) Euribor 6 months.*

(c) Euribor 9 months. *(d) Euribor 12 months.*

Figure 3.4: Euribor series. Daily evolution for the time period between October 15th 2001 and December 30th 2005. Sample size for each dataset is $n = 1\,077$.

ticed in Pagan et al. (1996), where different values of θ_4 where obtained using MLE and a GMM estimator. It is important to note that the standard error associated to the GMM estimation is larger than in the other approaches, which was already manifested in the simulations (see Tables 3.4–3.5 and A.5–A.6). Regarding the estimated values for the different maturities, for all time periods the series are persistent and the main variation comes from the parameters of the volatility function: θ_3 increases with maturity and θ_4 decreases. The parameter θ_4 controls the relationship between the interest rate and the volatility, for all series we have $\theta_4 > 1$ which indicates that volatility tends to increase as the rate rises. The estimation in the Euribor 12 months for θ_4 is close to 1, which would correspond to the diffusion process proposed by Brennan and Schwartz (1980).

As goodness-of-fit test for diffusion processes are available in the literature, we will test the parametric form of the drift and diffusion functions estimated for the Euribor series. Table 3.9 shows the p-values for the goodness-of-fit test suggested by Monsalve-Cobis et al. (2011). The empirical p-value is significant for the drift function in every maturity. Conversely, the p-value for the volatility function leads to a strong rejection of the null hypothesis, implying that the model is inadequate to explain the volatility of the series, for every

Method	$\hat{\theta}$	3 months		6 months		9 months		12 months	
DML	$\hat{\theta}_1$	1.5063	(0.4510)	1.3956	(0.6671)	1.5620	(0.9484)	1.7607	(1.1623)
	$\hat{\theta}_2$	0.7097	(0.1974)	0.6323	(0.2828)	0.6713	(0.3920)	0.7232	(0.4668)
	$\hat{\theta}_3$	0.0297	(0.0031)	0.0667	(0.0068)	0.1141	(0.0117)	0.1613	(0.0169)
	$\hat{\theta}_4$	1.8143	(0.1118)	1.3900	(0.1084)	1.1987	(0.1081)	1.0521	(0.1078)
LL	$\hat{\theta}_1$	1.5083	(0.4537)	1.4027	(0.6696)	1.5723	(0.9517)	1.7803	(1.1667)
	$\hat{\theta}_2$	0.7109	(0.1989)	0.6352	(0.2842)	0.6763	(0.3940)	0.7311	(0.4695)
	$\hat{\theta}_3$	0.0286	(0.0030)	0.0653	(0.0067)	0.1109	(0.0114)	0.1560	(0.0164)
	$\hat{\theta}_4$	1.8520	(0.1140)	1.4114	(0.1098)	1.2264	(0.1086)	1.0843	(0.1084)
HP	$\hat{\theta}_1$	1.5255	(0.4553)	1.4092	(0.6701)	1.5650	(0.9521)	1.7844	(1.1673)
	$\hat{\theta}_2$	0.7182	(0.1999)	0.6380	(0.2844)	0.6732	(0.3942)	0.7332	(0.4698)
	$\hat{\theta}_3$	0.0280	(0.0029)	0.0652	(0.0067)	0.1110	(0.0115)	0.1563	(0.0164)
	$\hat{\theta}_4$	1.8764	(0.1129)	1.4156	(0.1097)	1.2267	(0.1086)	1.0841	(0.1084)
KF	$\hat{\theta}_1$	1.4964	(0.4492)	1.3954	(0.6669)	1.5603	(0.9483)	1.9177	(1.1629)
	$\hat{\theta}_2$	0.7049	(0.1964)	0.6321	(0.2826)	0.6711	(0.3920)	0.7827	(0.4672)
	$\hat{\theta}_3$	0.0303	(0.0031)	0.0670	(0.0068)	0.1141	(0.0117)	0.1609	(0.0168)
	$\hat{\theta}_4$	1.7899	(0.1121)	1.3861	(0.1088)	1.1985	(0.1079)	1.0547	(0.1078)
MCMC	$\hat{\theta}_1$	1.5256	(0.4536)	1.4058	(0.6657)	1.5398	(0.9508)	1.7899	(1.1700)
	$\hat{\theta}_2$	0.7183	(0.2016)	0.6366	(0.2821)	0.6657	(0.3949)	0.7325	(0.4683)
	$\hat{\theta}_3$	0.0290	(0.0032)	0.0673	(0.0062)	0.1173	(0.0108)	0.1548	(0.0159)
	$\hat{\theta}_4$	1.8444	(0.1059)	1.3869	(0.1029)	1.1756	(0.1012)	1.0993	(0.1095)
GMM	$\hat{\theta}_1$	1.7081	(0.6782)	1.1439	(0.6100)	1.2429	(0.6646)	1.3074	(0.6633)
	$\hat{\theta}_2$	0.7431	(0.2389)	0.5216	(0.2214)	0.5400	(0.2503)	0.5465	(0.2590)
	$\hat{\theta}_3$	0.0242	(0.0167)	0.0676	(0.0213)	0.1126	(0.0306)	0.1661	(0.0438)
	$\hat{\theta}_4$	1.9562	(0.6262)	1.3704	(0.2735)	1.2068	(0.2230)	1.0227	(0.2118)

Table 3.8: Estimated parameters and standard errors (in parentheses) for the CKLS model, $\mathrm{d}X_t = (\theta_1 - \theta_2 X_t)\,\mathrm{d}t + \theta_3 X_t^{\theta_4}\,\mathrm{d}W_t$, fitted to Euribor series using six estimation methods.

maturity.

Maturity	3 months	6 months	9 months	12 months
p-value drift function	0.267	0.723	0.911	0.950
p-value volatility function	< 0.001	< 0.001	0.001	0.004

Table 3.9: p-values for the Monsalve-Cobis et al. (2011) goodness-of-fit test for the CKLS parametric form of the drift and diffusion functions.

To further analyze the fitted CKLS model, we will use a resampling procedure in the context of state space models (see Section 3.2.4) as it can provide insight into the validity of the

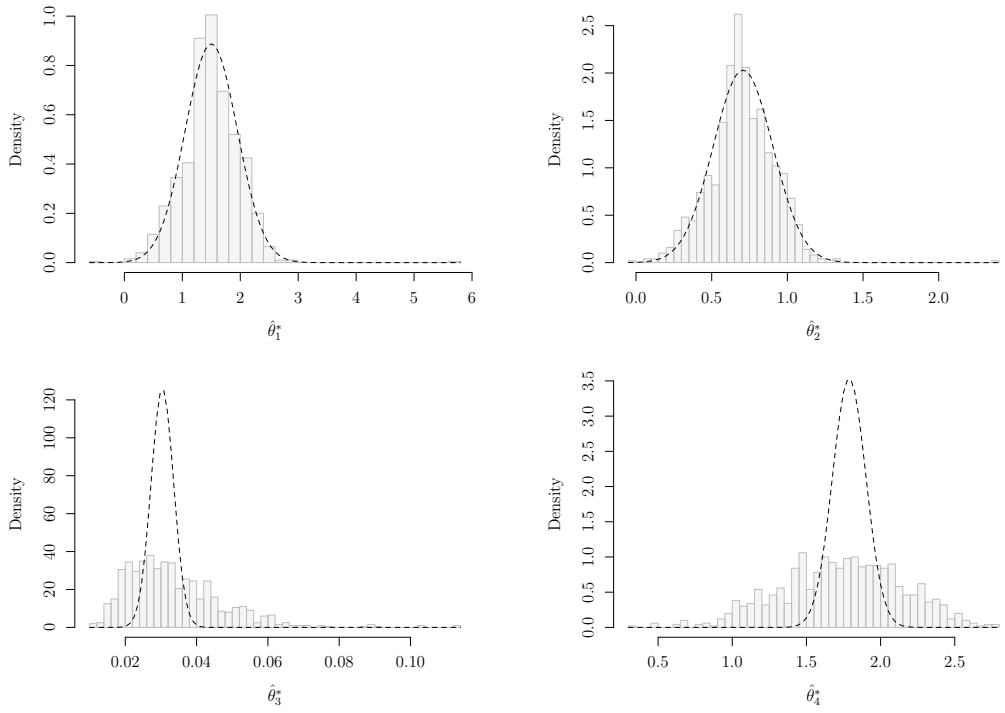


Figure 3.5: Bootstrap histogram of $\hat{\theta}_i$, with $i \in \{1, 2, 3, 4\}$, and asymptotic Gaussian distribution for the CKLS model fitted to Euribor 3 months series.

model. The bootstrap technique developed in Stoffer and Wall (1991) is easily implemented with the innovations form of the Kalman filter and allows us to approximate the sampling distributions of the parameter estimates. Inference in state space models estimated using the Kalman filter is feasible because there exists an asymptotic theory, as seen in Section 3.2.4, under general conditions, the parameter estimates of a state space model are consistent and asymptotically normal. Focusing on the estimates for the Euribor 3 months, Table 3.10 shows the standard errors obtained from $B = 1\,000$ bootstrap resamples, along with the asymptotic standard errors and parameter estimates, and Figure 3.5 illustrates the bootstrap distribution (histogram) and the asymptotic Gaussian distribution (dashed line) of $\hat{\theta}_i$, with $i \in \{1, 2, 3, 4\}$. Regarding the drift parameters ($\hat{\theta}_1$ and $\hat{\theta}_2$), the bootstrap and asymptotic distributions are close, however, for the diffusion parameters ($\hat{\theta}_3$ and $\hat{\theta}_4$), both distributions differ. This is consistent with the conclusions drawn from the goodness-of-fit test (see Table 3.9), where the parametric form of the drift was not rejected, but the diffusion function lead to a strong rejection. The bootstrap standard error of $\hat{\theta}_3$ and $\hat{\theta}_4$ is notable larger and the histograms show a slightly skewed distribution. This implies that the CKLS diffusion function specification is not able to explain the dynamics of the Euribor series, although the linear drift with mean

reversion seems to be adequate. A deterministic parametric form of the diffusion function might be incapable of capturing the volatility behavior (Ferreira and Gil-Bazo, 2004) and a more intricate form, such as stochastic volatility, may be more suitable, which was already pointed out in the econometric literature, that find data are more accurately represented by a stochastic volatility model.

Parameters:	θ_1	θ_2	θ_3	θ_4
Estimation	1.4964	0.7049	0.0303	1.7899
Asymptotic standard error	0.4492	0.1964	0.0031	0.1121
Bootstrap standard error	0.4794	0.2105	0.0653	0.4110

Table 3.10: Kalman filter estimates, asymptotic standard error and bootstrap standard error of the CKLS model fitted to Euribor 3 months serie.

3.5 Conclusions

We reviewed different parametric estimation methods for time-homogeneous SDE. Interest rate time series are highly persistent and this strong correlation through time challenges estimation, as its discretized counterpart corresponds to a near unit-root model. To address the problem of estimation and discretization bias, a comparative study of estimation methods was discussed under different settings. Based on the analysis of the simulation results, the following conclusions can be reached. First, estimation bias is large for the drift parameter κ, which controls the speed of mean-reversion, though smaller for the diffusion parameters. Lack of dynamics, that emerges when κ is small, increases this bias. Discretization bias was discerned in very low persistence scenarios with coarse discretization frequency. Second, the parameter of the diffusion function was accurately estimated in the Vasicek model, but the performance in the CKLS model was worse. Increasing the observation time T did reduce the bias, as expected, and increasing conditional volatility resulted in more accurate estimates. Third, estimation bias and variance is reduced as $T = n\Delta \to \infty$, rather than when only the sample size n is increased. This was illustrated in the simulations, where scenarios with different sample size n but similar T had comparable performances. Finally, regarding the parameter estimators, the GMM is the least efficient, while discrete maximum likelihood methods performed similarly, with LL and HP improving discretization bias over DML and KF. The MCMC method also provides efficient estimates; however, the more *ad hoc* implementation, highly model-dependent, makes its use more intricate than the other methods. The procedures reviewed in this chapter are provided in the `estsde` R package (Appendix B).

Chapter 4

Parametric estimation of diffusion models with stochastic volatility

One-factor diffusion models can be extended adding a more flexible specification of the diffusion function. Considering a stochastic volatility function, we can construct two-factor models, widely used in the financial literature, extending to a continuous-time setting the known discrete time series stochastic volatility specification. However, this flexibility comes with a cost, and the ensuing arduous estimation of the hidden variable has motivated the proposal of several estimation methods addressing latent variables. As the estimation of diffusion models with stochastic volatility based on discretely sampled data has proven difficult, we address this issue by reviewing some parametric estimation methods and carrying out a Monte Carlo study. An application of the procedures to real data is provided.

The contents of this chapter are collected in López-Pérez et al. (2022b).

Contents

4.1 Introduction

The empirical evidence obtained from the goodness-of-fit tests for the one-factor model proved unsatisfactory empirical fits and suggested that more flexible specifications for the volatility function were needed to capture the dynamics of returns of interest rates. Therefore, the literature has moved towards two-factor formulations, allowing the volatility function to incorporate a source of random variation, leading to a continuous-time stochastic volatility (SV) model, such as

$$\mathrm{d}X_t = m_1(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma_t \nu_1(X_t, \boldsymbol{\theta})\,\mathrm{d}W_{1,t}, \tag{4.1}$$

$$\mathrm{d}g(\sigma_t) = m_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}t + \nu_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}W_{2,t}, \tag{4.2}$$

where the functions g, m_1, ν_1, m_2 and ν_2 are sufficiently smooth and satisfy growth conditions to obtain existence and uniqueness for the stochastic differential equation solution, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ is an unknown parameter vector. Several parametrizations of (4.1)–(4.2) have been examined, see Hull and White (1987), Heston (1993), Andersen and Lund (1997a), Gallant and Tauchen (1998), Eraker (2001), or Christoffersen et al. (2009), among others.

The latent factor in the stochastic volatility model challenges its implementation. In addition to the unobserved volatility, the process is specified in continuous-time but the observations occur at discrete time points. We attempt to discuss the intricacies of different implementations, though a comprehensive survey of estimation methods for continuous-time unobserved state-variable models is beyond the scope of this chapter. Several methods have been proposed for the estimation of stochastic volatility models (see Chen, 2003, for a review). One of the earliest proposals was quasi-maximum likelihood, introduced in Harvey et al. (1994) (see, e.g., Ruiz, 1994; Hurn et al., 2013). Another analytical methods include the Kalman (1960) filter (see Broto and Ruiz, 2004 for a survey on likelihood-based methods); the generalized method of moments (see, e.g., Melino and Turnbull, 1990; Andersen and Sørensen, 1996; Sapp, 2009); approximate likelihood methods based on the characteristic function of the transition density (Bates, 2006); closed-form moment-based procedure (Dufour and Valéry, 2009); Fourier transform methods (Cuchiero and Teichmann, 2015; Merkle

et al., 2020); maximum likelihood using closed-form approximations and latent factor filtering (Aït-Sahalia et al., 2020). Simulation-based methods, though more computationally demanding, are increasingly used in the financial context. These methods include Markov Chain Monte Carlo techniques (see, e.g., Jacquier et al., 1994; Shephard and Pitt, 1997; Kim et al., 1998; Eraker, 2001; Chib et al., 2002; Johannes and Polson, 2010; Kastner and Frühwirth-Schnatter, 2014); particle filters (see, e.g., Kotecha and Djuric, 2003; Carvalho et al., 2010; Lopes and Tsay, 2011; Kantas et al., 2015); Expectation-Maximization algorithms (see, e.g., Dempster et al., 1977; Little and Rubin, 2019); simulated maximum likelihood (see, e.g., Danielsson and Richard, 1993; Sandmann and Koopman, 1998; Durham, 2006); integrated nested Laplace approximations (INLA) methods (Rue et al., 2009).

The rest of the chapter is structured as follows. Section 4.2 provides an outline of estimation methods and in Section 4.3 Monte Carlo experiments are designed in order to discuss the finite sample performance of the procedures. Section 4.4 includes the estimation of the two-factor model for the Euribor series and conclusions are drawn in Section 4.5.

4.2 Estimation of diffusion models with latent variables

Given a two-factor continuous-time diffusion model as in (4.1)-(4.2), where the state X_t is observable but the volatility σ_t is a latent factor, although the model is formulated in continuous-time, data are sampled in discrete time points. Therefore, a discretized version of the diffusion equations should be consider to estimate the model parameters. Taking a popular specification, such as

$$\mathrm{d}X_t = (\beta - \kappa X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_{1,t}, \tag{4.3}$$

$$\mathrm{d}\log\sigma_t^2 = (\theta_0 - \theta_1 \log\sigma_t^2)\,\mathrm{d}t + \xi\,\mathrm{d}W_{2,t}, \tag{4.4}$$

where volatility follows an Ornstein-Uhlenbeck (OU) process (Stein and Stein, 1991), we assume that the process $\{X_{t_i}\}_{i=0}^n$ is observed at equispaced discrete time points $0 = t_0 < t_1 < \ldots < t_n = T$ in the interval $[0, T = n\Delta]$, where the time step Δ between consecutive observations is fixed. The Euler-Maruyama (Maruyama, 1955) method is commonly used as an approximation scheme, thereby, considering the SDE in (4.3)–(4.4), its discretized counterpart is given by

$$X_{t_i} - X_{t_{i-1}} = \beta\Delta - \kappa\Delta X_{t_{i-1}} + \sigma_{t_i}\sqrt{\Delta}\,\varepsilon_{1,t_i}, \qquad \varepsilon_{1,t_i} \sim \mathcal{N}(0,1), \tag{4.5}$$

$$\log\sigma_{t_i}^2 - \log\sigma_{t_{i-1}}^2 = \theta_0\Delta - \theta_1\Delta\log\sigma_{t_{i-1}}^2 + \xi\sqrt{\Delta}\,\varepsilon_{2,t_i}, \qquad \varepsilon_{2,t_i} \sim \mathcal{N}(0,1), \tag{4.6}$$

with $i = 0, 1, \ldots, n-1$, $t_i = i\Delta$, $X_{t_0} = x_0 \in \mathbb{R}$, and where ε_{1,t_i} and ε_{2,t_i} are i.i.d. standard Gaussian variables, given the independent and Gaussian increments property of the Wiener process.

In the remainder of this section, we give an outline of estimation methods procedures to estimate the unknown parameter vector. The state-space models allow an easily interpretable and flexible framework for stochastic volatility models, therefore, we focus on estimation procedures that are maximum likelihood-based on a state space model representation. We begin introducing a filtering algorithm and, as several Monte Carlo-based approximations for state-space models are available, we also consider this Bayesian approach, as well as sequential Monte Carlo methods or particle filters.

4.2.1 Kalman filter

As discussed in Chapter 3, the dynamic linear model (Kalman, 1960) considers that the observation vector $\boldsymbol{y}_t$ is generated by the state-space model

$$\boldsymbol{x}_t = \boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{Q}), \tag{4.7}$$

$$\boldsymbol{y}_t = \mathbf{A}_t\boldsymbol{x}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{R}), \tag{4.8}$$

where $\boldsymbol{x}_t \in \mathbb{R}^p$ is the unknown state vector, $\boldsymbol{y}_t \in \mathbb{R}^q$ is the observed data vector, $\mathbf{A}_t \in \mathbb{R}^{q\times p}$ is the observation matrix, $\boldsymbol{\Phi} \in \mathbb{R}^{p\times p}$ is the transition matrix and we assume that $\{\boldsymbol{w}_t\}$ and $\{\boldsymbol{v}_t\}$ are uncorrelated. The state vector $\boldsymbol{x}_t$ is latent, which provides an adequate framework for the stochastic volatility model. Let $\boldsymbol{x}_{t|t-1} = \mathbb{E}\left[\boldsymbol{x}_t \mid Y_{t-1}\right]$ and $\mathbf{P}_{t|t-1} = \mathbb{E}\left[(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})^\top \mid Y_{t-1}\right]$, with initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{x}_{0|0}, \mathbf{P}_{0|0})$, given the data $Y_s = (y_1, \ldots, y_s)$ we have

$$\boldsymbol{x}_t \mid Y_{t-1} \sim \mathcal{N}(\boldsymbol{x}_{t|t-1}, \mathbf{P}_{t|t-1}), \tag{4.9}$$

$$\boldsymbol{y}_t \mid Y_{t-1} \sim \mathcal{N}(\boldsymbol{y}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}),$$

$$\boldsymbol{x}_t \mid Y_t \sim \mathcal{N}(\boldsymbol{x}_{t|t}, \mathbf{P}_{t|t}), \tag{4.10}$$

which are the *propagation*, *predictive* and *filtering density*, respectively, with

$$\boldsymbol{x}_{t|t-1} = \boldsymbol{\Phi}\boldsymbol{x}_{t-1|t-1}, \qquad \mathbf{P}_{t|t-1} = \boldsymbol{\Phi}\mathbf{P}_{t-1|t-1}\boldsymbol{\Phi}^\top + \mathbf{Q},$$

$$\boldsymbol{y}_{t|t-1} = \mathbf{A}_t\boldsymbol{x}_{t|t-1}, \qquad \boldsymbol{\Sigma}_t = \mathbf{A}_t\mathbf{P}_{t|t-1}\mathbf{A}_t^\top + \mathbf{R},$$

$$\boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + \mathbf{K}_t\boldsymbol{\varepsilon}_t, \qquad \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\boldsymbol{\Sigma}_t\mathbf{K}_t^\top,$$

where $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \mathbf{A}_t \boldsymbol{x}_{t|t-1}$ are the prediction errors and $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{A}_t^\top \boldsymbol{\Sigma}_t^{-1}$ is the *Kalman gain*. In state space models the aim is usually the estimation of the latent state vector $\boldsymbol{x}_t$ through filtering, for which we need to estimate the marginal distribution of the state vector given the observations, $p(\boldsymbol{x}_t \mid Y_t)$. In Equations (4.9)–(4.10) linearity and the gaussianity of errors are assumed, but more general state space models can be considered. For this linear model, the Kalman filter (Kalman, 1960) can be used to estimate the distribution $p(\boldsymbol{x}_t \mid Y_t)$ and computing the likelihood using the innovations. The discretized OU model in (4.5)–(4.6) is not linear so, in order to use the Kalman filter with the state-space model (4.7)–(4.8), we first need to linearize it. Taking the residuals from the linear regression, $e_{t_i} = [X_{t_i} - \beta\Delta + (1-\kappa\Delta)X_{t_{i-1}}]/\sqrt{\Delta} = \sigma_{t_i}\varepsilon_{1,t_i}$, we define the logarithm of the squared residuals

$$y_{t_i} = \log e_{t_i}^2$$
$$h_{t_i} = \log \sigma_{t_i}^2,$$

therefore, the model can be linearized as follows, parameterizing $\phi_0 = \Delta\theta_0$, $\phi_1 = (1-\theta_1\Delta)$,

$$y_{t_i} = h_{t_i} + v_{t_i}, \qquad v_{t_i} \sim \log \chi_1^2, \tag{4.11}$$
$$h_{t_i} = \phi_0 + \phi_1 h_{t_{i-1}} + w_{t_i}, \qquad w_{t_i} \sim \mathcal{N}(0, \sigma_w^2), \tag{4.12}$$

with $w_{t_i} = \xi\sqrt{\Delta}\,\varepsilon_{2,t_i}$ a Gaussian distributed variable, as ε_{2,t_i} is a standard Gaussian variable, and where $\sigma_w^2 = \Delta\xi^2$. The error term in the space equation (4.11) follows a log chi-squared with one degree of freedom, as $v_{t_i} = \log \varepsilon_{1,t_i}^2$ and $\varepsilon_{1,t_i} \sim \mathcal{N}(0,1)$. The density of the $\log \chi_1^2$ is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\exp(x) - x\right)\right], \qquad x \in \mathbb{R},$$

with mean $\mathbb{E}[v_{t_i}] = \psi(1) - \log 2 \approx -1.2704$, where $\psi(\cdot)$ is a digamma function, and variance $\mathbb{V}\mathrm{ar}[v_{t_i}] = \pi^2/2 \approx 4.9348$. As this density is skewed (see Figure 4.3), it departs from the Gaussian assumption, thus different approaches have been proposed in the literature. Shumway and Stoffer (2000) proposed modeling the $\log \chi_1^2$ with a mixture of two Gaussian variables, one centered at zero, such as $\eta_{t_i} = I_{t_i} z_{0,t_i} + (1 - I_{t_i}) z_{1,t_i}$, where $z_{0,t_i} \sim \mathcal{N}(0, \sigma_0^2)$ and $z_{1,t_i} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and I_{t_i} is an i.i.d. Bernoulli process, $\mathbb{P}(I_{t_i} = 0) = \pi_0$ and $\mathbb{P}(I_{t_i} = 1) = \pi_1$, with $\pi_0 + \pi_1 = 1$. Substituting the space equation (4.11) with

$$y_{t_i} = h_{t_i} + \eta_{t_i},$$

we have the filtering equations for this model:

$$\begin{aligned} h_{t_{i+1}|t_i} &= \phi_0 + \phi_1 h_{t_i|t_{i-1}} + \sum_{j=0}^{1} \pi_{j,t_i} K_{j,t_i} \varepsilon_{j,t_i} \\ P_{t_{i+1}|t_i} &= \phi_1^2 P_{t_i|t_{i-1}} + \sigma_w^2 - \sum_{j=0}^{1} \pi_{j,t_i} K_{j,t_i}^2 \Sigma_{j,t_i} \end{aligned} \tag{4.13}$$

$$\begin{aligned} \epsilon_{0,t_i} &= y_{t_i} - h_{t_i|t_{i-1}} & \epsilon_{1,t_i} &= y_{t_i} - h_{t_i|t_{i-1}} - \mu_1 \\ \Sigma_{0,t_i} &= P_{t_i|t_{i-1}} + \sigma_0^2 & \Sigma_{1,t_i} &= P_{t_i|t_{i-1}} + \sigma_1^2 \\ K_{0,t_i} &= \phi_1 P_{t_i|t_{i-1}} \, \Sigma_{0,t_i}^{-1} & K_{1,t_i} &= \phi_1 P_{t_i|t_{i-1}} \, \Sigma_{1,t_i}^{-1} \end{aligned} \tag{4.14}$$

where, given the density of y_{t_i} conditional to its past values Y_{t_1}, $f(t_i \mid t_{i-1})$,

$$\pi_{j,t_i} = \frac{\pi_{1,t_i} f_1(t_i \mid t_{i-1})}{\pi_{0,t_i} f_0(t_i \mid t_{i-1}) + \pi_{1,t_i} f_1(t_i \mid t_{i-1})}$$

and $f_j(t_i \mid t_{i-1}) = \mathcal{N}(x_{t_i|t_{i-1}} + \mu_j, \Sigma_{j,t_i})$, for $j = 0, 1$ and $\mu_0 = 0$. The distribution π_{j,t_i}, for $j = 0, 1$, is specified *a priori*, usually uniform priors are chosen, $\pi_{0,t_i} = \pi_{1,t_i} = 1/2$. Let $\boldsymbol{\theta} = (\phi_0, \phi_1, \sigma_w^2, \mu_1, \sigma_0^2, \sigma_1^2)^\top$ be the vector of unknown parameters, maximum likelihood can be used for estimation, maximizing the log-likelihood $\ln \mathcal{L}_Y(\boldsymbol{\theta})$ given by

$$\ln \mathcal{L}_Y(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \left(\sum_{j=0}^{1} \pi_{j,t_i} \mathcal{N}\Big(x_{t_i|t_{i-1}} + \mu_j, \sigma_j^2\Big) \right). \tag{4.15}$$

The Newton-Raphson algorithm can be considered to maximize the log-likelihood (4.15). Another approach is to implement the EM algorithm, where the steps of the procedure can be regarded as the recursions of the Kalman filter and smoother and a multivariate MLE.

4.2.2 Markov Chain Monte Carlo

Given the discrete version (4.11)–(4.12) of the SDE in (4.3)–(4.4), we have the popular parametrization of the discretized stochastic volatility model,

$$e_{t_i} = \exp\left(h_{t_i}/2\right) \varepsilon_{1,t_i}, \qquad \varepsilon_{1,t_i} \sim \mathcal{N}(0,1), \tag{4.16}$$

$$h_{t_i} = \phi_0 + \phi_1 h_{t_{i-1}} + w_{t_i}, \qquad w_{t_i} \sim \mathcal{N}(0, \sigma_w^2), \tag{4.17}$$

where e_{t_i} are the residuals from the linear regression $X_{t_i} = \beta\Delta + (1 - \kappa\Delta) X_{t_{i-1}} + \sqrt{\Delta} e_{t_i}$, with $h_{t_i} = \log \sigma_{t_i}^2$, $\phi_0 = \Delta\theta_0$, $\phi_1 = (1 - \theta_1\Delta)$ and $\sigma_w^2 = \Delta\xi^2$. The model in (4.16) can be linearized

by taking the logarithm of the squared observations, $y_{t_i} = \ln e_{t_i}^2$, as in (4.11). In the Bayesian approach of the estimation problem, the unknown parameter vector $\boldsymbol{\theta} = (\phi_0, \phi_1, \sigma_w^2)^\top$, with $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^3$, we consider a prior distribution of $\boldsymbol{\theta}$ over the parameter space Θ. Given the prior distribution $\pi(\boldsymbol{\theta})$, we use Bayes' Theorem to obtain the posterior distribution of the parameter vector,

$$p(\boldsymbol{\theta} \mid \boldsymbol{e}) = \frac{p(\boldsymbol{e} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{e} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}} \propto p(\boldsymbol{e} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where $\boldsymbol{e} = \{e_{t_i}\}_{i=0}^n$ is the vector of residuals. As a closed form solution might not exist, or its calculation is too difficult, sampling methods can be used to overcome this problem, such as Markov Chain Monte Carlo (MCMC) procedures, for example. Different MCMC procedures have been proposed to estimate the SV model (see Shephard, 1993; Jacquier et al., 1994, for initial proposals), where the focus is targeted to the posterior density $p(\boldsymbol{\theta}, \boldsymbol{h} \mid \boldsymbol{e})$, with $\boldsymbol{h} = \{h_{t_i}\}_{i=0}^n$, as the direct analysis of $p(\boldsymbol{\theta} \mid \boldsymbol{e})$ is not possible and the likelihood function $\mathcal{L}(\boldsymbol{e} \mid \boldsymbol{\theta})$ is intractable. Through Bayes' Theorem we have $p(\boldsymbol{h}, \boldsymbol{\theta} \mid \boldsymbol{e}) \propto p(\boldsymbol{e} \mid \boldsymbol{h}) p(\boldsymbol{h} \mid \boldsymbol{\theta}) p(\theta)$, and via MCMC methods we can sample from $p(\boldsymbol{\theta}, \boldsymbol{h} \mid \boldsymbol{e})$. Using the Gibbs sampler we can produce samples from the posterior $p\big(h_{t_i} \mid \boldsymbol{h}_{-t_i}, e_{t_i}, \boldsymbol{\theta}\big)$, for $i = 1, \ldots, n$ where $\boldsymbol{h}_{-t_i} = (h_{t_0}, \ldots, h_{t_{i-1}}, h_{t_{i+1}}, \ldots, h_{t_n})^\top$, and $p\big(\boldsymbol{\theta} \mid \boldsymbol{e}, \boldsymbol{h}\big)$, so that these samples will converge to those generated from $p\big(\boldsymbol{\theta}, \boldsymbol{h} \mid \boldsymbol{e}\big)$. Let $\boldsymbol{h}_{t_{a:b}} = (h_{t_a}, \ldots, h_{t_b})^\top$, the Gibbs sampler algorithm for the discrete model (4.16)–(4.17) is given in Algorithm 1.

Algorithm 1 (Gibbs sampling algorithm)**.** For $j = 0, \ldots, n$, the Gibbs sampler proceeds as follows, setting $j = 0$:

1. Initialize $\boldsymbol{h}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$.
2. Sample $\boldsymbol{h}^{(j+1)}$ from $p\big(\boldsymbol{h} \mid \boldsymbol{e}, \boldsymbol{\theta}^{(j)}\big)$.
3. Sample $\boldsymbol{\theta}^{(j+1)}$ from $p\big(\boldsymbol{\theta} \mid \boldsymbol{e}, \boldsymbol{h}^{(j+1)}\big)$.
   (a) Sample $\sigma_w^2 \mid \boldsymbol{e}, \boldsymbol{h}, \phi_0, \phi_1$.
   (b) Sample $\phi_1 \mid \boldsymbol{h}, \phi_0, \sigma_w^2$.
   (c) Sample $\phi_0 \mid \boldsymbol{h}, \phi_1, \sigma_w^2$.
4. Set $j = j + 1$ and go to (2).

Step (2) in Algorithm 1 can be implemented by using a Metropolis-Hastings algorithm, where $p\big(\boldsymbol{h} \mid \boldsymbol{e}, \boldsymbol{\theta}^{(j)}\big)$ is decomposed into conditionals, $p\big(h_{t_i} \mid \boldsymbol{h}_{-t_i}^{(j)}, \boldsymbol{e}, \boldsymbol{\theta}^{(j)}\big)$, for $i = 1, \ldots, n$,

to sample $\boldsymbol{h}^{(j+1)}$. To sample $(h_{t_0} \mid h_{t_1}, \boldsymbol{\theta}, \sigma_w^2)$, given $h_{t_0} \sim \mathcal{N}(m_0, c_0)$ and $h_{t_1} \mid h_{t_0} \sim \mathcal{N}(\phi_0 + \phi_1 h_{t_0}, \sigma_w^2)$, we can use Bayes' theorem leading to

$$h_{t_0} \mid h_{t_1} \sim \mathcal{N}(m_1, c_1),$$

where $m_1 = c_1 \left(c_0^{-1} m_0 + \phi_1 \sigma_w^{-2}(h_{t_1} - \phi_0)\right)$ and $c_1 = (c_0^{-1} + \phi_1^2 \sigma_w^{-2})^{-1}$. Regarding the conditional prior distribution of h_{t_i}, for $i = 1, \ldots, n-1$ we have

$$\begin{pmatrix} h_{t_i} \\ h_{t_{i+1}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \phi_0 + \phi_1 h_{t_{i-1}} \\ (1+\phi_1)\phi_0 + \phi_1^2 h_{t_{i-1}} \end{pmatrix}, \sigma_w^2 \begin{pmatrix} 1 & \phi_1 \\ \phi_1 & 1+\phi_1^2 \end{pmatrix} \right),$$

therefore

$$\begin{aligned} (h_{t_i} \mid h_{t_{i-1}}, h_{t_{i+1}}, \boldsymbol{\theta}, \sigma_w^2) &\sim \mathcal{N}(\mu_{t_i}, \nu^2), \\ (h_{t_n} \mid h_{t_{n-1}}, \boldsymbol{\theta}, \sigma_w^2) &\sim \mathcal{N}(\mu_{t_n}, \sigma_w^2), \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}\left[h_{t_i} \mid h_{t_{i-1}}, h_{t_{i+1}}, \boldsymbol{\theta}, \sigma_w^2\right] &= \mu_{t_i} = \left(\frac{1-\phi_1}{1+\phi_1^2}\right)\phi_0 + \left(\frac{\phi_1}{1+\phi_1^2}\right)(h_{t_{i-1}} + h_{t_{i+1}}), \\ \mathbb{V}\text{ar}\left[h_{t_i} \mid h_{t_{i-1}}, h_{t_{i+1}}, \boldsymbol{\theta}, \sigma_w^2\right] &= \nu^2 = \sigma_w^2(1+\phi_1^2)^{-1}, \end{aligned}$$

and $\mathbb{E}\left[h_{t_n} \mid h_{t_{n-1}}, \boldsymbol{\theta}, \sigma_w^2\right] = \mu_{t_n} = \phi_0 + \phi_1 h_{t_{n-1}}$. We can sample h_{t_i} via independent Metropolis-Hastings, let $f_{\mathcal{N}}(t_i \mid a, b)$ denote the Gaussian density distribution with mean a and variance b, the full conditional distribution of h_{t_i} is given by

$$\begin{aligned} p(h_{t_i} \mid \boldsymbol{h}_{-t_i}, \boldsymbol{e}, \boldsymbol{\theta}, \sigma_w^2) &= p(h_{t_i} \mid h_{t_{i-1}}, h_{t_{i+1}}, \boldsymbol{\theta}, \sigma_w^2)\; p(\boldsymbol{e} \mid h_{t_i}) \\ &= f_{\mathcal{N}}(h_{t_i}; \mu_t, \nu^2)\; f_{\mathcal{N}}\left(e_{t_i}; 0, \exp(h_{t_i})\right). \end{aligned}$$

Given that

$$\log p(e_{t_i} \mid h_{t_i}) = \text{const} - \frac{1}{2} h_{t_i} - \frac{e_{t_i}^2}{2} \exp(-h_{t_i})$$

and that a Taylor expansion of $\exp(-h_{t_i})$ around μ_{t_i} leads to

$$\begin{aligned} \log p(e_{t_i} \mid h_{t_i}) &\approx \text{const} - \frac{1}{2} h_{t_i} - \frac{e_{t_i}^2}{2}\big(\exp(-\mu_{t_i}) - (h_{t_i} - \mu_{t_i})\exp(-\mu_{t_i})\big), \\ g(h_{t_i}) &= \exp\left(-\frac{1}{2} h_{t_i}\left(1 - e_{t_i}^2 \exp(-\mu_t)\right)\right), \end{aligned}$$

by combining $f_{\mathcal{N}}(h_{t_i}; \mu_{t_i}, \nu^2)$ and $g(h_{t_i})$ we have the proposal distribution

$$p(h_{t_i} \mid h_{-t_i}, \boldsymbol{e}, \boldsymbol{\theta}, \sigma_w^2) \equiv \mathcal{N}(h_{t_i}; \tilde{\mu}_{t_i}, \nu^2),$$

where $\tilde{\mu}_{t_i} = \mu_{t_i} + \frac{1}{2}\nu^2\big(e_{t_i}^2 \exp(-\mu_{t_i}) - 1\big)$. Algorithm 2 provides the independent Metropolis-Hastings algorithm, where the acceptance probability is given in step (3).

Algorithm 2 (Metropolis-Hastings algorithm)**.** For $i = 1, \ldots, n$ and $j = 0, \ldots, l$, the independent Metropolis-Hastings algorithm proceeds as follows:

1. Current state $h_{t_i}^{(j)}$.

2. Sample $h_{t_i}^*$ from $\mathcal{N}(\tilde{\mu}_{t_i}, \nu^2)$.

3. Compute the acceptance probability

$$\alpha = \min\left\{1, \frac{f_{\mathcal{N}}\big(h_{t_i}^*; \mu_{t_i}, \nu^2\big)\ f_{\mathcal{N}}\big(e_{t_i}; 0, \exp(h_{t_i}^*)\big)}{f_{\mathcal{N}}\big(h_{t_i}^{(j)}; \mu_{t_i}, \nu^2\big)\ f_{\mathcal{N}}\big(e_{t_i}; 0, \exp(h_{t_i}^{(j)})\big)}\ \frac{f_{\mathcal{N}}\big(h_{t_i}^{(j)}; \tilde{\mu}_{t_i}, \nu^2\big)}{f_{\mathcal{N}}\big(h_{t_i}^*; \tilde{\mu}_{t_i}, \nu^2\big)}\right\}.$$

4. New state:

$$h_{t_i}^{(j+1)} = \begin{cases} h_{t_i}^* & \text{w.p.} \quad \alpha, \\ h_{t_i}^{(j)} & \text{w.p.} \quad 1 - \alpha. \end{cases}$$

Regarding the sampling of the hyperparameters $\boldsymbol{\theta} = (\phi_0, \phi_1, \sigma_w^2)^\top$, step (3) in Algorithm 1, setting the initial log volatility $h_{t_0} \sim \mathcal{N}(m_0, c_0)$ and $\boldsymbol{\phi} = (\phi_0, \phi_1)^\top$, the prior distributions of $\boldsymbol{\phi}$ and σ_w^2 are

$$\begin{aligned} \boldsymbol{\phi} \mid \sigma_w^2 &\sim \mathcal{N}(\boldsymbol{\theta}^{(0)}, \sigma_w^2 \mathbf{V}_0), \\ \sigma_w^2 &\sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0 s_0^2}{2}\right), \end{aligned}$$

respectively. Conditional on $\boldsymbol{h}_{t_{0:n}}$, the posterior distribution of $\boldsymbol{\phi}$ and σ_w^2 is

$$\begin{aligned} (\boldsymbol{\phi} \mid \sigma_w^2, \boldsymbol{e}, h_{t_{0:n}}) &\sim \mathcal{N}(\boldsymbol{\phi}^{(1)}, \sigma_w^2 \mathbf{V}_1), \\ (\sigma_w^2 \mid \boldsymbol{e}, \boldsymbol{h}_{t_{0:n}}) &\sim IG\left(\frac{\nu_1}{2}, \frac{\nu_1 s_1^2}{2}\right), \end{aligned}$$

given that $\nu_1 = \nu_0 + n$,

$$\mathbf{X} = \begin{pmatrix} 1 & h_{t_0} \\ \vdots & \vdots \\ 1 & h_{t_{n-1}} \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{V}_1 &= (\mathbf{V}_0^{-1} + \mathbf{X}^\top \mathbf{X})^{-1}, \\ \boldsymbol{\phi}^{(1)} &= \mathbf{V}_1(\mathbf{V}_0^{-1}\boldsymbol{\phi}^{(0)} + \mathbf{X}^\top \boldsymbol{h}_{t_{1:n}}), \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (\boldsymbol{e} - \mathbf{X}\boldsymbol{\phi}^{(1)})^\top (\boldsymbol{e} - \mathbf{X}\boldsymbol{\phi}^{(1)}) + (\boldsymbol{\phi}^{(1)} - \boldsymbol{\phi}^{(0)})^\top \mathbf{V}_0^{-1} (\boldsymbol{\phi}^{(1)} - \boldsymbol{\phi}^{(0)}). \end{aligned}$$

4.2.3 Particle filter

Particle filters incorporate the sequential estimation approach of the Kalman filter algorithms and the flexibility for modeling of MCMC sampling algorithms. Replacing the Kalman filter recursions in (4.9) and (4.10) by

$$p\big(\boldsymbol{x}_{t_i} \mid Y_{t_{i-1}}\big) = \int p\big(\boldsymbol{x}_{t_i} \mid \boldsymbol{x}_{t_{i-1}}\big) p\big(\boldsymbol{x}_{t_{i-1}} \mid Y_{t_{i-1}}\big)\, \mathrm{d}\boldsymbol{x}_{t_{i-1}}, \tag{4.18}$$

$$p\big(\boldsymbol{x}_{t_i} \mid Y_{t_i}\big) = \frac{p\big(Y_{t_i} \mid \boldsymbol{x}_{t_i}\big) p\big(\boldsymbol{x}_{t_i} \mid Y_{t_{i-1}}\big)}{p\big(Y_{t_i} \mid Y_{t_{i-1}}\big)}, \tag{4.19}$$

respectively, leads to a more general dynamic model, where assumptions like normality and linearity can be relaxed. However, both distributions in (4.18) and (4.19) are intractable and computationally costly. Particle filters algorithms approximate $p\big(\boldsymbol{x}_{t_i} \mid Y_{t_i}\big)$ by drawing a set of l i.i.d. particles $\{\boldsymbol{x}_{t_i}^{(j)}\}_{j=1}^{l}$, starting with a set of i.i.d. particles $\{\boldsymbol{x}_{t_{i-1}}^{(j)}\}_{j=1}^{l}$ approximating $p\big(\boldsymbol{x}_{t_{i-1}} \mid Y_{t_{i-1}}\big)$. Since the early sequential Monte Carlo algorithm proposed by West (1992), several filters have been proposed in the literature, like the Bootstrap filter or sequential importance sampling with resampling (SISR) by Gordon et al. (1993) and the auxiliary particle filter or auxiliary SIR (ASIR) proposed by Pitt and Shephard (1999), among others. Liu and West (2001) proposed a filter for sequential learning, a variant of the auxiliary particle filtering (APF) algorithm, that combines the APF together with a kernel approximation to $p(\boldsymbol{\theta} \mid Y_{t_{i-1}})$ using a mixture of multivariate Gaussian distributions and shrinkage parameter to provide artificial evolution for the parameter vector $\boldsymbol{\theta}$. Therefore, the posterior distribution for $\boldsymbol{\theta}$ is approximated by the normal mixture

$$p\big(\boldsymbol{\theta} \mid Y_{t_i}\big) = \sum_{j=1}^{l} \mathcal{N}\big(m^{(j)}, h^2 V_{t_i}\big),$$

where $m^{(j)} = a\boldsymbol{\theta}_{t_i}^{(j)} + (1-a)\tilde{\boldsymbol{\theta}}_{t_i}$, $a = \sqrt{1-h^2}$, $\tilde{\boldsymbol{\theta}}_{t_i} = \sum_{j=1}^{l} \boldsymbol{\theta}_{t_i}^{(j)}/l$ and $V_{t_i} = \sum_{j=1}^{l}(\boldsymbol{\theta}_{t_i}^{(j)} - \tilde{\boldsymbol{\theta}}_{t_i})(\boldsymbol{\theta}_{t_i}^{(j)} - \tilde{\boldsymbol{\theta}}_{t_i})^{\top}/l$. The constant a measures the extent of the shrinkage and h controls the degree of overdispersion of the mixture (the choice of both parameters is discussed in Liu and West, 2001). The general algorithm is displayed in Algorithm 3.

Algorithm 3 (Liu and West filter)**.** For $i = 1, \ldots, n$, a general Liu and West (2001) filter algorithm runs as follows:

1. Set the prior point estimates $\{(\hat{\mu}_{t_{i+1}}, m_{t_i})^{(j)}\}_{j=1}^{l}$ of $(x_{t_i}, \boldsymbol{\theta})$ where $\hat{\mu}_{t_i}^{(j)} = \mathbb{E}\big[x_{t_{i+1}} \mid x_{t_i}^{(j)}, \boldsymbol{\theta}^{(j)}\big]$.
2. Resample $\big\{(\tilde{x}_{t_i}, \tilde{\boldsymbol{\theta}}_{t_i})^{(j)}\big\}_{j=1}^{l}$ from $\big\{(x_{t_i}, \boldsymbol{\theta}_{t_i})^{(j)}\big\}_{j=1}^{l}$ with weights
$$w_{t_{i+1}}^{(j)} \propto p\big(y_{t_{i+1}} \mid \hat{\mu}_{t_{i+1}}^{(j)}, m^{(j)}\big).$$
3. Propagate
   (a) $\Big\{\tilde{\boldsymbol{\theta}}_{t_i}^{(j)}\Big\}_{j=1}^{l}$ via $\mathcal{N}\big(\tilde{m}^{(j)}, h^2 V_{t_i}\big)$,
   (b) $\Big\{\tilde{x}_{t_i}^{(j)}\Big\}_{j=1}^{l}$ via $p\big(x_{t_{i+1}} \mid \tilde{x}_{t_i}^{(j)}, \tilde{\boldsymbol{\theta}}_{t_{i+1}}^{(j)}\big)$.
4. Resample $\big\{(x_{t_{i+1}}, \boldsymbol{\theta}_{t_{i+1}})^{(j)}\big\}_{j=1}^{l}$ from $\big\{(\tilde{x}_{t_{i+1}}, \tilde{\boldsymbol{\theta}}_{t_{i+1}}^{(j)})^{(j)}\big\}_{j=1}^{l}$ with weights
$$w_{t_{i+1}}^{(j)} \propto \frac{p\big(y_{t_{i+1}} \mid \tilde{x}_{t_{i+1}}^{(j)}, \tilde{\boldsymbol{\theta}}_{t_{i+1}}^{(j)}\big)}{p\big(y_{t_{i+1}} \mid \hat{\mu}_{t_{i+1}}^{(j)}, \tilde{m}^{(j)}\big)}.$$

4.2.4 Fourier transform methods

Another Bayesian approach is using Fourier transform methods. Merkle et al. (2020) proposed a procedure using Cuchiero and Teichmann (2015) volatility estimator, which provides an estimate of the latent volatility process to subsequently approximate the likelihood for the parameters. Considering a general two-factor model,

$$\begin{aligned} \mathrm{d}X_t &= \mu\,\mathrm{d}t + f(\sigma_t)\,\mathrm{d}W_{1,t}, \\ \mathrm{d}\sigma_t &= (\theta_0 - \theta_1\sigma_t)\,\mathrm{d}t + \xi g(\sigma_t)\,\mathrm{d}W_{2,t}, \\ \mathrm{d}W_{1,t}\,\mathrm{d}W_{2,t} &= \rho\,\mathrm{d}t. \end{aligned} \tag{4.20}$$

with $\rho \in [-1, 1]$ the correlation between the Wiener processes. In the two-stages procedure of Merkle et al. (2020), first an estimation of the volatility is obtained using the Cuchiero

and Teichmann (2015) estimator of the instantaneous volatility, that is, given a continuous function h and let $s_m^n = m/n$, assuming the process is observed at times $\{s_m^n\}_{m=0}^{\lfloor nT \rfloor}$,

$$\widehat{\sigma}_t^{n,N} = f^{-1}\left(\sqrt{\rho_h^{-1}\left(\widehat{\rho_h(\sigma)}_t^{n,N}\right)}\right),$$

where

$$\widehat{\rho_h(\sigma)}_t^{n,N} = \frac{1}{T}\sum_{k=-N}^{N}\left(1-\frac{|k|}{N}\right)\exp\left(i\frac{2\pi}{T}kt\right)G(X,h,k)_T^n,$$

$$G(X,h,k)_T^n = \frac{1}{n}\sum_{m=1}^{\lfloor nT \rfloor}\exp\left(-i\frac{2\pi}{T}ks_{m-1}^n\right)h\left(\sqrt{n}(X_{s_m^n}-X_{s_{m-1}^n})\right),$$

with $\rho_h(\Sigma) = \mathbb{E}\left[h(Z)\right]$, $Z \sim N(0,\Sigma)$, for any covariance matrix Σ, and usually $N = \lfloor\sqrt{n}\rfloor$. Regarding the choice of the function h, taking $h(x) = \cos(x)$ gives $\rho_h(x) = e^{-\frac{1}{2}x}$ and $\rho_h^{-1}(x) = -2\log x$. Conditioned on the true volatility process path σ_t, the increments of X_t are independent and Gaussian (Merkle et al., 2020, Theorem 2.2). Thus, an expression of the likelihood function conditional on the volatility can be obtained and, by replacing the true path by the estimate $\widehat{\sigma}_t^{n,N}$, the parameter vector can be estimated.

Taking the parametrization in (4.20), by setting $f(x) = g(x) = \sqrt{x}$ we obtain the Heston model, and $f(x) = e^x$, $g(x) = 1$, we have the Ornstein-Uhlenbeck as in (4.3)–(4.4). Figures 4.1–4.2 show the estimated path of the volatility from the Heston model (with $n = 10\,000$) and the Ornstein-Uhlenbeck (with $n = 20\,000$), respectively. As noted in Merkle et al. (2020), in both models the estimated paths are close to the true volatility, but the Heston model is estimated more accurately.

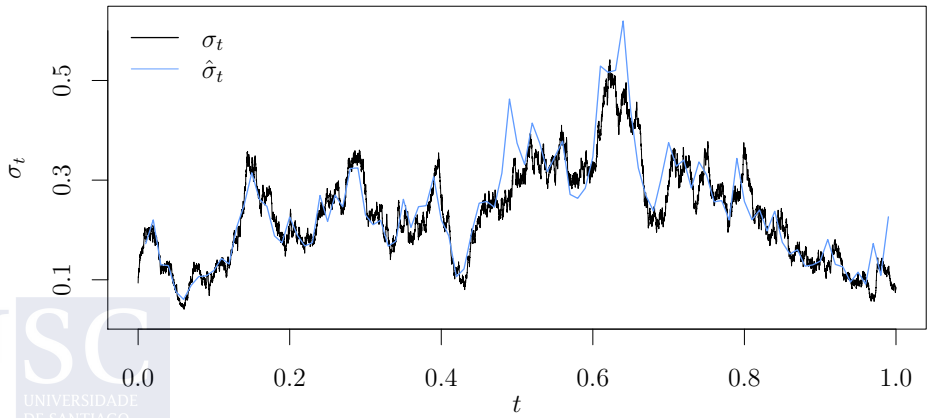


Figure 4.1: Volatility estimated path of the Heston model. True unobserved path in black.

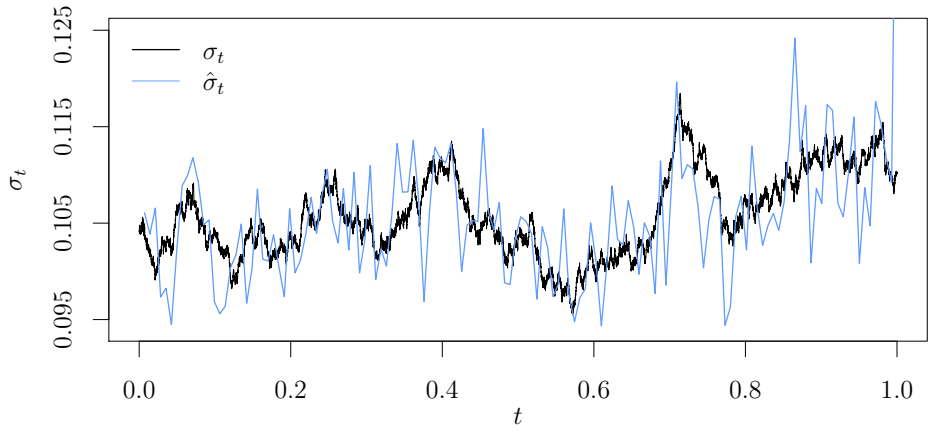


Figure 4.2: Volatility estimated path of the OU model. True unobserved path in black.

4.2.5 Generalized method of moments

The Generalized Method of Moments (GMM) introduced in Chapter 3 for the estimation of diffusion models with deterministic volatility function can be extended to the case of an stochastic volatility function by considering a different choice of moment conditions. Taking the parametrization in (4.16)–(4.17), it is possible to derive 24 moment conditions from the statistical properties of the model. Set $m = \phi_0/(1-\phi_1)$, $s^2 = \sigma_w^2/(1-\phi_1^2)$ and given that

$$\mathbb{E}\left[\sigma_{t_i}^p\right] = \exp\left(\frac{pm}{2} + \frac{p^2 s^2}{8}\right),$$
$$\mathbb{E}\left[\sigma_{t_i}^p \sigma_{t_{i-j}}^q\right] = \mathbb{E}\left[\sigma_{t_i}^p\right]\mathbb{E}\left[\sigma_{t_{i-j}}^q\right] \exp\left(\frac{pq\phi_1^j s^2}{4}\right),$$

the moment conditions expressions are given by

$$\begin{aligned}
\mathbb{E}\left[|e_{t_i}|\right] &= \sqrt{2/\pi}\mathbb{E}\left[\sigma_{t_i}\right], \\
\mathbb{E}\left[e_{t_i}^2\right] &= \mathbb{E}\left[\sigma_{t_i}^2\right], \\
\mathbb{E}\left[|e_{t_i}|\right] &= 2\sqrt{2/\pi}\mathbb{E}\left[\sigma_{t_i}^2\right], \\
\mathbb{E}\left[e_{t_i}^4\right] &= 3\mathbb{E}\left[\sigma_{t_i}^4\right], \\
\mathbb{E}\left[|e_{t_i} e_{t_j}|\right] &= (2/\pi)\mathbb{E}\left[\sigma_{t_i}\sigma_{t_j}\right], \quad j \in \{1,\ldots,10\}, \\
\mathbb{E}\left[e_{t_i}^2 e_{t_j}^2\right] &= \mathbb{E}\left[\sigma_{t_i}^2\sigma_{t_j}^2\right], \qquad j \in \{1,\ldots,10\},
\end{aligned}$$

for positive integers p and q. With the analytic expressions, more intricate models such as the two-factor CKLS model, that is, a diffusion function $\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$, can be estimated within the GMM framework. The following vector of 14 moment conditions $f_t(\boldsymbol{\theta})$ can be

used for the CKLS model:

$$\mathbb{E}\left[f_t(\boldsymbol{\theta})\right] = \mathbb{E}\left[\begin{array}{c}
\varepsilon_{1,t_i} \\
\varepsilon_{1,t_i} X_{t_i} \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_i}^{\gamma}}\right)^2 - \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{2(1-\phi_1^2)}\right) \\
\left[\left(\frac{\varepsilon_{1,t_i}}{X_{t_i}^{\gamma}}\right)^2 - \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{2(1-\phi_1^2)}\right)\right] X_{t_{i-1}} \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right| - \sqrt{\frac{2}{\pi}} \exp\left(\frac{\phi_0}{2(1-\phi_1)} + \frac{\sigma_w^2}{8(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right|^3 - 2\sqrt{\frac{2}{\pi}} \exp\left(\frac{3\phi_0}{2(1-\phi_1)} + \frac{9\sigma_w^2}{8(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^4 - 3\exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{2\sigma_w^2}{(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}} \frac{\varepsilon_{1,t_{i-1}}}{X_{t_{i-2}}^{\gamma}}\right| - \frac{2}{\pi} \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{4(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1 \sigma_w^2}{4(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}} \frac{\varepsilon_{1,t_{i-3}}}{X_{t_{i-4}}^{\gamma}}\right| - \frac{2}{\pi} \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{4(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^3 \sigma_w^2}{4(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}} \frac{\varepsilon_{1,t_{i-5}}}{X_{t_{i-6}}^{\gamma}}\right| - \frac{2}{\pi} \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{4(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^5 \sigma_w^2}{4(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}} \frac{\varepsilon_{1,t_{i-7}}}{X_{t_{i-8}}^{\gamma}}\right| - \frac{2}{\pi} \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{4(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^7 \sigma_w^2}{4(1-\phi_1^2)}\right) \\
\left|\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}} \frac{\varepsilon_{1,t_{i-9}}}{X_{t_{i-10}}^{\gamma}}\right| - \frac{2}{\pi} \exp\left(\frac{\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{4(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^9 \sigma_w^2}{4(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^2 \left(\frac{\varepsilon_{1,t_{i-2}}}{X_{t_{i-3}}^{\gamma}}\right)^2 - \exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^2 \sigma_w^2}{(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^2 \left(\frac{\varepsilon_{1,t_{i-4}}}{X_{t_{i-5}}^{\gamma}}\right)^2 - \exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^4 \sigma_w^2}{(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^2 \left(\frac{\varepsilon_{1,t_{i-6}}}{X_{t_{i-7}}^{\gamma}}\right)^2 - \exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^6 \sigma_w^2}{(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^2 \left(\frac{\varepsilon_{1,t_{i-8}}}{X_{t_{i-9}}^{\gamma}}\right)^2 - \exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^8 \sigma_w^2}{(1-\phi_1^2)}\right) \\
\left(\frac{\varepsilon_{1,t_i}}{X_{t_{i-1}}^{\gamma}}\right)^2 \left(\frac{\varepsilon_{1,t_{i-10}}}{X_{t_{i-11}}^{\gamma}}\right)^2 - \exp\left(\frac{2\phi_0}{(1-\phi_1)} + \frac{\sigma_w^2}{(1-\phi_1^2)}\right) \exp\left(\frac{\phi_1^{10} \sigma_w^2}{(1-\phi_1^2)}\right)
\end{array}\right] = 0$$

The choice of the number of moment conditions has been discussed in the literature when estimating models with latent variables, as an excessive number can hinder the estimation, but a low number can lead to misspecification. Sapp (2009) suggested that a judicious choice is 14 moment conditions, as Andersen and Lund (1997a) stated that at least nine different moment conditions were needed, because an inferior number could made the estimation procedure unstable. However, using a very large number of moment conditions could lead to

inefficiency due to the size of the covariance matrix estimated from the observations. This is evidenced in Melino and Turnbull (1990) and Vetzal (1997), where despite of using 35 moment conditions, the GMM is unable to estimate the level parameter.

4.3 Comparative study

This section compares the results of the different estimation procedures applied to three parametrizations of the continuous-time two-factor model with stochastic volatility: we consider a simpler model, such as (4.3)–(4.4), which does not include a level parameter; a model with a more intricate volatility function, with level parameter, with and without correlated errors. We compare the parameter estimates obtained with the different procedures under Monte Carlo settings to examine their finite sample performance. Through all the models and procedures considered, we first estimate the parameters of the drift function $m_1(\cdot)$ in (4.1) and, subsequently, the residuals obtained are used in the procedures to estimate the parameter vector $\boldsymbol{\theta}$.

4.3.1 Ornstein-Uhlenbeck with stochastic volatility

Considering the Ornstein-Uhlenbeck model with stochastic volatility introduced in (4.3)–(4.4) and its discretized counterpart, (4.5)–(4.6), we have the discrete two-factor model

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= (\beta - \kappa X_{t_i})\Delta + \sigma_{t_i}\sqrt{\Delta}\varepsilon_{1,t_i}, & \varepsilon_{1,t_i} &\sim \mathcal{N}(0,1), \\ \log \sigma^2_{t_{i+1}} &= \phi_0 - \phi_1 \log \sigma^2_{t_i} + w_{t_i}, & w_{t_i} &\sim \mathcal{N}(0, \sigma^2_w), \end{aligned}$$

where $\phi_0 = \theta_0\Delta$, $\phi_1 = 1 - \theta_1\Delta$ and $\sigma^2_w = \Delta\xi^2$, with $i = 0, 1, \ldots, n-1$ and initial condition $X_{t_0} = x_0 \in \mathbb{R}$. To estimate the vector of parameters $\boldsymbol{\theta} = (\beta, \kappa, \phi_0, \phi_1, \sigma^2_w)^\top$ we first obtain the residuals from the linear regression $u_{t_i} = \beta - \kappa X_{t_i} + e_{t_i}$, where $u_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$. Therefore, in this first step we obtain the estimates of β and κ (see Section 3.3.2) and thereafter we use the procedures to estimate the rest of the parameters using the residuals of the linear regression, $e_{t_i} = \sigma_{t_i}\Delta^{-1/2}\varepsilon_{1,t_i}$. The vector of parameter values considered for data simulation is $\boldsymbol{\theta} = (\beta, \kappa, \phi_0, \phi_1, \sigma^2_w)^\top = (0.01, 0.3, -0.006, 0.99, 0.0225)^\top$, with weekly frequency ($\Delta = 1/52$) for sample size $n \in \{520, 1040, 2080\}$, which corresponds to $T = 10, 20$ and 40 years, respectively. A thousand realizations of random sample paths $\{X_{i\Delta}\}_{i=1}^n$ were generated, where the first 1000 observations were discarded to remove the dependence on the initial value.

Parameter	True	Mean	Var	MSE
β	0.01	0.0195	0.6764	0.6765
κ	0.3	0.8888	0.3903	0.7370
		MCMC		
ϕ_0	-0.006	-0.0139	0.0019	0.0020
ϕ_1	0.99	0.9692	0.0065	0.0069
σ_w^2	0.0225	0.0231	4.457×10^{-5}	4.490×10^{-5}
		Particle filter		
ϕ_0	-0.006	-0.0224	0.0011	0.0013
ϕ_1	0.99	0.9370	0.0004	0.0032
σ_w^2	0.0225	0.0376	1.394×10^{-4}	3.663×10^{-4}
		Kalman filter		
ϕ_0	-0.006	-0.0420	0.0506	0.0519
ϕ_1	0.99	0.9457	0.0467	0.0487
σ_w^2	0.0225	0.0410	9.213×10^{-4}	1.265×10^{-3}

Table 4.1: Parameter estimates for the Ornstein-Uhlenbeck process with stochastic volatility, with sample size $n = 520$ which corresponds to ten years ($T = 10$) of weekly data ($\Delta = 1/52$).

Parameter	True	Mean	Var	MSE
β	0.01	0.0089	0.1617	0.1617
κ	0.3	0.5676	0.1240	0.1956
		MCMC		
ϕ_0	-0.006	-0.0083	7.270×10^{-5}	7.781×10^{-5}
ϕ_1	0.99	0.9851	6.075×10^{-5}	8.525×10^{-5}
σ_w^2	0.0225	0.0227	3.341×10^{-5}	3.345×10^{-5}
		Particle filter		
ϕ_0	-0.006	-0.0211	6.144×10^{-4}	8.422×10^{-4}
ϕ_1	0.99	0.9501	2.597×10^{-4}	1.854×10^{-3}
σ_w^2	0.0225	0.0369	1.462×10^{-4}	3.539×10^{-4}
		Kalman filter		
ϕ_0	-0.006	-0.0168	3.885×10^{-3}	5.178×10^{-3}
ϕ_1	0.99	0.9585	3.985×10^{-2}	4.181×10^{-2}
σ_w^2	0.0225	0.0350	1.902×10^{-4}	5.336×10^{-4}

Table 4.2: Parameter estimates for the Ornstein-Uhlenbeck process with stochastic volatility, with sample size $n = 1040$ which corresponds to twenty years ($T = 20$) of weekly data ($\Delta = 1/52$).

Parameter	True	Mean	Var	MSE
β	0.01	0.0192	0.0508	0.0509
κ	0.3	0.4378	0.0433	0.0623
		MCMC		
ϕ_0	-0.006	-0.0071	2.699×10^{-5}	2.825×10^{-5}
ϕ_1	0.99	0.9878	2.121×10^{-5}	2.585×10^{-5}
σ_w^2	0.0225	0.0226	2.323×10^{-5}	2.325×10^{-5}
		Particle filter		
ϕ_0	-0.006	-0.0176	3.167×10^{-4}	4.508×10^{-4}
ϕ_1	0.99	0.9629	1.610×10^{-4}	8.958×10^{-4}
σ_w^2	0.0225	0.0344	1.420×10^{-4}	2.827×10^{-4}
		Kalman filter		
ϕ_0	-0.006	-0.0087	3.638×10^{-5}	1.329×10^{-3}
ϕ_1	0.99	0.9874	2.627×10^{-5}	1.992×10^{-3}
σ_w^2	0.0225	0.0340	7.659×10^{-5}	4.200×10^{-4}

Table 4.3: Parameter estimates for the Ornstein-Uhlenbeck process with stochastic volatility, with sample size $n = 2080$ which corresponds to forty years ($T = 40$) of weekly data ($\Delta = 1/52$).

Tables 4.1–4.3 report the estimates for the three procedures considered: Markov Chain Monte Carlo (MCMC) method, using a Metropolis-Hastings algorithm within the Gibbs sampling; the Liu and West (2001) filter (Particle filter); and the Kalman (1960) filter. Tables include the mean and variance (Var) for one thousand simulations, along with the mean squared error (MSE). The estimates of the drift parameters β and κ were not obtained with the procedures, as mentioned, but rather fitting a linear regression. The simulation-based techniques show low MSE with the different sample sizes considered, while the Kalman filter, thought for the smallest sample size n has higher bias and variance, it particularly decreases when the observation window T is extended, achieving a MSE closer to the other methods. Whilst the MCMC and the particle filter procedures do deliver accurate estimations, their computationally demanding nature and the practical implementation, highly model-dependent, are major disadvantages. On the other hand, the flexibility of the Kalman filter can provide a good trade-off between speed and efficiency.

4.3.2 CKLS with stochastic volatility

In this section, a more intricate model is considered, based on the CKLS model proposed in Chan et al. (1992), where stochastic volatility is incorporated to the diffusion function. The

CKLS model with stochastic volatility described by the Ornstein-Uhlenbeck process is given by

$$\begin{aligned} \mathrm{d}X_t &= (\beta - \kappa X_t)\,\mathrm{d}t + \sigma_t X_t^{\gamma}\,\mathrm{d}W_{1,t}, \\ \mathrm{d}\log\sigma_t^2 &= (\theta_0 - \theta_1 \log\sigma_t^2) + \xi\,\mathrm{d}W_{2,t}, \end{aligned} \tag{4.21}$$

and its discretized counterpart,

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= (\beta - \kappa X_t)\Delta + \sigma_t X_{t_i}^{\gamma}\,(W_{1,t_{i+1}} - W_{1,t_i}), \\ \log\sigma_{t_{i+1}}^2 &= \phi_0 - \phi_1 \log\sigma_{t_i}^2 + \xi\,(W_{2,t_{i+1}} - W_{2,t_i}), \end{aligned} \tag{4.22}$$

where $\phi_0 = \theta_0\Delta$ and $\phi_1 = 1 - \theta_1\Delta$. This process has been proposed in the literature to model the short term interest rate (see Andersen and Lund, 1997a; 1997b; among others), as it represents an extension of the classical stochastic volatility model to a continuous-time setting incorporating *level effect*, implying that volatility depends on the level of the interest rate and inducing conditional heteroskedasticity.

The Kalman filter can be easily extended to allow modifications of the two-factor model. As indicated in Section 4.2.1, the error term in the space equation (4.11) follows a log chi-squared with one degree of freedom, and this has motivated different approaches in the literature. Shumway and Stoffer (2000) modeled the $\log\chi_1^2$ with a mixture of two Gaussian variables (see the filtering equations (4.13)–(4.14)), while Kim et al. (1998) proposed a seven-component Gaussian mixture (see Chib et al., 2002, and Artigas and Tsay, 2004), with weights π_i and mean and variance ($\mathcal{N}(\mu_i, \sigma_i^2)$ for $i \in \{1, \ldots, 7\}$) given in Table 4.4. A comparison of the true $\log\chi_1^2$ distribution with a Gaussian distribution and the two and seven Gaussian mixture is illustrated in Figure 4.3.

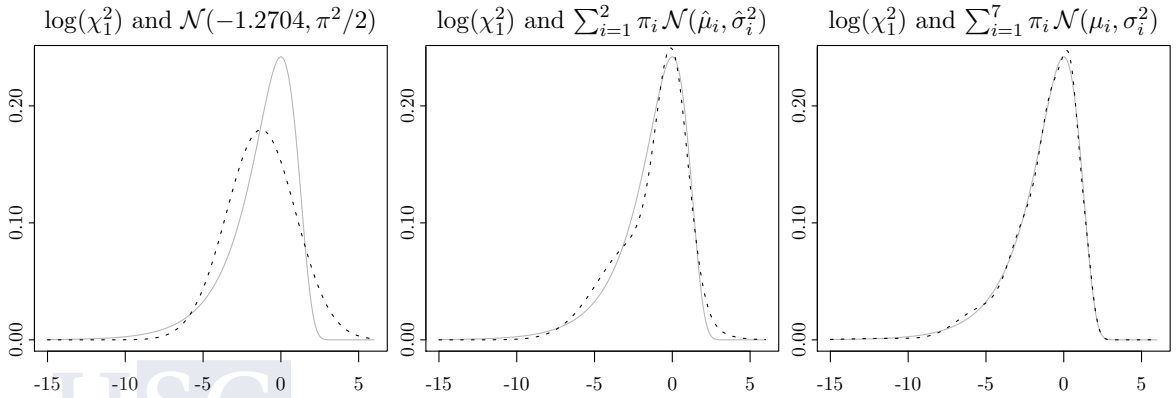


Figure 4.3: Comparison of the $\log\chi_1^2$ *density function (solid line) and Gaussian density (left), mixture of two normal distributions (middle) and mixture of seven normal distributions (right).*

Component i	1	2	3	4	5	6	7
π_i	0.00730	0.10556	0.00002	0.04395	0.34001	0.24566	0.25750
μ_i	-11.400	-5.2432	-9.8373	1.5075	-0.6510	0.5248	-2.3586
σ_i^2	5.7960	2.6137	5.1795	0.1674	0.6401	0.3402	1.2626

Table 4.4: Components of a mixture of seven Gaussian distributions, $\mathcal{N}(\mu_i, \sigma_i^2)$, with weights π_i.

A simulation study was conducted to compare both approaches. A thousand realizations of random sample paths for the CKLS-OU model in (4.21) were generated with weekly frequency, with $\boldsymbol{\theta} = (\beta, \kappa, \gamma, \phi_0, \phi_1, \xi)^\top = (0.04, 0.6, 1.5, -0.013, 0.998, 0.4)^\top$. Note that in the two mixture approach we estimate the parameter vector $\boldsymbol{\theta}$ and the components of the mixture—namely, $(\mu_1, \sigma_0^2, \sigma_1^2)^\top$, as $\mu_0 = 0$—while in the seven mixture approach the means and variances of the Gaussian distributions remain fixed according to the values in Table 4.4.

The parameter estimates are summarized in Tables 4.5–4.7, for the two (Kalman filter 2) and seven (Kalman filter 7) mixture and three sample sizes $n \in \{520, 1040, 2080\}$. Though both methods provide a close performance, for small sample size (Table 4.5) the seven mixture filter provides lower mean squared error. However, larger sample sizes (Table 4.7) show an improvement for the two mixture filter, where the estimation of the parameter ξ, known as the *volatility of volatility* and hard to estimate accurately, outperforms the seven mixture filter. As an example, Figure 4.4 shows the estimated path (dotted) of the log volatility ($\log \sigma_t^2$) using the two and seven mixture approach, and a 95% confidence interval for the estimated paths (shaded).

		Kalman filter (2)			**Kalman filter (7)**		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0731	0.0017	0.0028	0.0731	0.0017	0.0028
κ	0.6	1.0964	0.3760	0.6224	1.0964	0.3760	0.6224
γ	1.5	1.5305	0.0243	0.0253	1.5009	0.0133	0.0133
ϕ_0	-0.013	-0.4977	1.5604	1.7948	-0.2974	0.8113	0.8919
ϕ_1	0.998	0.9249	0.0333	0.0386	0.9555	0.0190	0.0208
ξ	0.4	0.7284	1.1962	1.3040	0.6963	0.2206	0.3083

Table 4.5: Parameter estimates for the discretized version of (4.21)*, as in* (4.22)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions. A thousand simulations were carried out with weekly data and $n = 520$ ($T = 10$ years).*

		Kalman filter (2)			**Kalman filter (7)**		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0567	0.0006	0.0008	0.0567	0.0006	0.0008
κ	0.6	0.8508	0.1258	0.1887	0.8508	0.1258	0.1887
γ	1.5	1.5248	0.0072	0.0078	1.4983	0.0069	0.0069
ϕ_0	-0.013	-0.1228	0.0996	0.1116	-0.0753	0.0299	0.0337
ϕ_1	0.998	0.9825	0.0018	0.0021	0.9892	0.0006	0.0006
ξ	0.4	0.5865	0.0940	0.1288	0.6672	0.0507	0.1221

Table 4.6: Parameter estimates for the discretized version of (4.21)*, as in* (4.22)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions. A thousand simulations were carried out with weekly data and $n = 1040$ ($T = 20$ years).*

		Kalman filter (2)			**Kalman filter (7)**		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0485	0.0002	0.0003	0.0485	0.0002	0.0003
κ	0.6	0.7281	0.0509	0.0673	0.7281	0.0509	0.0673
γ	1.5	1.5184	0.0045	0.0049	1.4939	0.0057	0.0058
ϕ_0	-0.013	-0.0445	0.0013	0.0022	-0.0360	0.0008	0.0013
ϕ_1	0.998	0.9937	2.5×10^{-5}	4.4×10^{-5}	0.9949	1.7×10^{-5}	2.7×10^{-5}
ξ	0.4	0.5169	0.0208	0.0345	0.6826	0.0245	0.1043

Table 4.7: Parameter estimates for the discretized version of (4.21)*, as in* (4.22)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions. A thousand simulations were carried out with weekly data and $n = 2080$ ($T = 40$ years).*

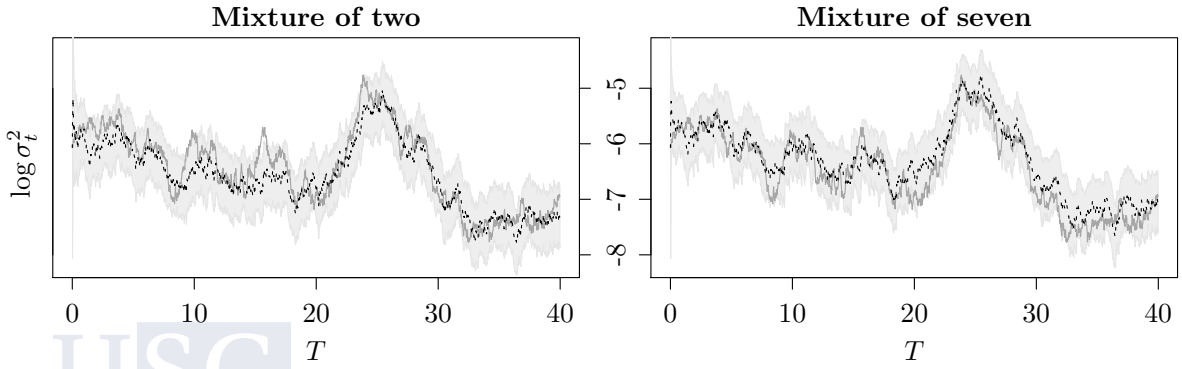


Figure 4.4: Comparison of the true $\log \sigma_t^2$ path (solid line) and estimated (dotted line) with a mixture of two normal distributions (left) and mixture of seven normal distributions (right). Shaded in gray is a 95% confidence interval for both estimations.

4.3.3 CKLS with stochastic volatility and correlated errors

We incorporate leverage effect to the CKLS-OU model in (4.21) by considering correlated Wiener processes, given by

$$\begin{aligned} \mathrm{d}X_t &= (\beta - \kappa X_t)\,\mathrm{d}t + \sigma_t X_t^{\gamma}\,\mathrm{d}W_{1,t}, \\ \mathrm{d}\log\sigma_t^2 &= (\theta_0 - \theta_1 \log\sigma_t^2)\,\mathrm{d}t + \xi\,\mathrm{d}W_{2,t}, \\ \mathrm{d}W_{1,t}\,\mathrm{d}W_{2,t} &= \rho\,\mathrm{d}t. \end{aligned} \tag{4.23}$$

with $\rho \in [-1, 1]$.

Let e_{t_i} be the residuals from the linear regression,

$$e_{t_i} = \left[X_{t_i} - \beta\Delta + (1 - \kappa\Delta)X_{t_{i-1}}\right]/\sqrt{\Delta} = \sigma_{t_{i_1}} X_{t_{i_1}}^{\gamma}\,\varepsilon_{1,t_i},$$

with ε_{1,t_i} standard Gaussian distributed, the discretized version of (4.23), setting $y_{t_i} = \log e_{t_i}^2$ and $h_{t_i} = \log\sigma_{t_i}^2$, is

$$\begin{aligned} y_{t_i} &= h_{t_{i-1}} + 2\gamma\log X_{t_{i-1}} + v_{t_i}, & v_{t_i} &\sim \log\chi^2, \\ h_{t_i} &= \phi_0 + \phi_1 h_{t_{i-1}} + w_{t_i}, & w_{t_i} &\sim \mathcal{N}(0, \sigma_w^2), \end{aligned} \tag{4.24}$$

where $\phi_0 = \theta_0\Delta$, $\phi_1 = 1 - \theta_1\Delta$, $w_{t_i} = \xi\sqrt{\Delta}\varepsilon_{2,t_i}$, $\varepsilon_{2,t_i} \sim \mathcal{N}(0,1)$, and $\sigma_w^2 = \Delta\xi^2$, with $i = 0, 1, \ldots, n-1$. But because of the logarithmic and square transformation ($v_{t_i} = \log\varepsilon_{1,t_i}^2$), we can not retain the correlation between ε_{1,t_i} and w_{t_i}, that is, $\mathrm{Cor}\left[\varepsilon_{1,t_i}, \varepsilon_{2,t_i}\right] = \rho$. To overcome this problem, Artigas and Tsay (2004) proposed maintaining the leverage effect by defining $\eta_{t_i} = \rho w_{t_i} + \tilde{\eta}_{t_i}$, where $\tilde{\eta}_{t_i}$ is a normal random variable independent of ε_{1,t_i} and $\mathbb{V}\mathrm{ar}\left[\tilde{\eta}_{t_i}\right] = \sigma_w^2(1-\rho)^2$. Note that

$$\begin{pmatrix} \varepsilon_{1,t_i} \\ \eta_{t_i} \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma),$$

where

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & \rho\sigma_w \\ \rho\sigma_w & \sigma_w^2 \end{pmatrix},$$

therefore, $\Sigma_{\eta_{t_i}|\varepsilon_{1,t_i}} = \sigma_w^2 - \rho^2\sigma_w^2$. The state-space equations for the Kalman filter algorithm can be written as

$$\begin{aligned} y_{t_i} &= h_{t_{i-1}} + 2\gamma\log X_{t_{i-1}} + v_{t_i}, \\ h_{t_i} &= \phi_0 + \phi_1 h_{t_{i-1}} + \eta_{t_i}, \end{aligned} \tag{4.25}$$

with $\eta_{t_i} = \rho\sigma_w^2\varepsilon_{1,t_i} + \tilde{\eta}_{t_i}$ and $\tilde{\eta}_{t_i} \sim \mathcal{N}\big(0, \sigma_w^2(1-\rho)^2\big)$. Substituting $\varepsilon_{1,t_i} = e_{t_i}\exp\big(-h_{t_{i-1}}/2\big)X_{t_{i-1}}^{-\gamma}$ in the state equation in (4.25), we have the modified state equation

$$h_{t_i} = \phi_0 + \phi_1 h_{t_{i-1}} + \rho\,\sigma_w e_{t_i}\exp\left(-\frac{h_{t_{i-1}}}{2}\right)X_{t_{i-1}}^{-\gamma} + \tilde{\eta}_{t_i} = G(h_{t_{i-1}}) + \tilde{\eta}_{t_i}.$$

This equation is nonlinear, meaning that the filtering equations in (4.13)–(4.14) are no longer applicable. However, Artigas and Tsay (2004) proposed approximating the system by using a time-varying linear Kalman filter, thus, we modify

$$P_{t_{i+1}|t_i} = g\big(h_{t_i|t_i}\big)^2 P_{t_i|t_i} + \sigma_w^2(1-\rho^2),$$

where $g\big(h_{t_i|t_i}\big) = \partial G(x)/\partial x\,|_{x=h_{t_i|t_i}}$ is the derivative of $G(h_{t_i})$ evaluated at $h_{t_i|t_i}$,

$$g\big(h_{t_i|t_i}\big) = \phi_1 - \frac{1}{2}\rho\,\sigma_w e_{t_{i+1}}\exp\left(-\frac{h_{t_i}}{2}\right)X_{t_i}^{-\gamma}.$$

We used this modification of the filtering equations to do a simulation study for the CKLS-OU model with leverage effect, introduced in (4.23). We generated a thousand random sample paths $\{X_{i\Delta}\}_{i=0}^{n}$, discarding the first 1000 observations as a burn-in period, with weekly frequency for sample sizes $n \in \{520, 1040, 2080\}$, that is $T \in \{10, 20, 40\}$ years, and parameters $\boldsymbol{\theta} = (\beta, \kappa, \gamma, \phi_0, \phi_1, \xi, \rho)^\top = (0.04, 0.6, 1.5, -0.010, 0.998, 0.4, -0.5)^\top$.

Tables 4.8–4.10 contain the mean, variance and mean squared error for the estimations obtained using the Kalman filter with a mixture of two (left) and seven (right) Gaussian distributions, for $T = 10, 20$ and 40 years, respectively. Both methods perform similarly, achieving the seven mixture method slightly more accurate estimations, as the MSE is lower. However, they perform poorly when estimating the correlation ρ, even with 40 years of weekly data.

		Kalman filter (2)			**Kalman filter (7)**		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0542	0.0007	0.0009	0.0542	0.0007	0.0009
κ	0.6	0.8127	0.1600	0.2053	0.8127	0.1600	0.2053
γ	1.5	1.5812	0.0560	0.0626	1.5350	0.0443	0.0455
ϕ_0	-0.010	-0.3974	2.2753	2.4257	-0.2071	2.0415	2.0805
ϕ_1	0.998	0.9330	0.1030	0.1072	0.9839	0.0584	0.0586
ξ	0.4	0.6656	0.8315	0.9021	0.5251	0.2638	0.2795
ρ	-0.5	-0.1697	0.1295	0.2386	-0.1855	0.0854	0.1843

Table 4.8: Parameter estimates for the discretized version of (4.23)*, as in* (4.24)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions, with weekly data and* $n = 520$*, which corresponds to* $T = 10$ *years.*

		Kalman filter (2)			Kalman filter (7)		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0485	0.0004	0.0004	0.0485	0.0004	0.0004
κ	0.6	0.7282	0.0836	0.1000	0.7282	0.0836	0.1000
γ	1.5	1.6027	0.0827	0.0933	1.5392	0.0327	0.0343
ϕ_0	-0.010	-0.2130	1.8248	1.8662	-0.0335	0.5314	0.5320
ϕ_1	0.998	0.9842	0.0355	0.0357	1.0155	0.0171	0.0174
ξ	0.4	0.5286	0.2646	0.2811	0.5591	0.1389	0.1642
ρ	-0.5	-0.1418	0.0694	0.4813	-0.1742	0.0238	0.4784

Table 4.9: Parameter estimates for the discretized version of (4.23)*, as in* (4.24)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions, with weekly data and* $n = 1040$*, which corresponds to* $T = 20$ *years.*

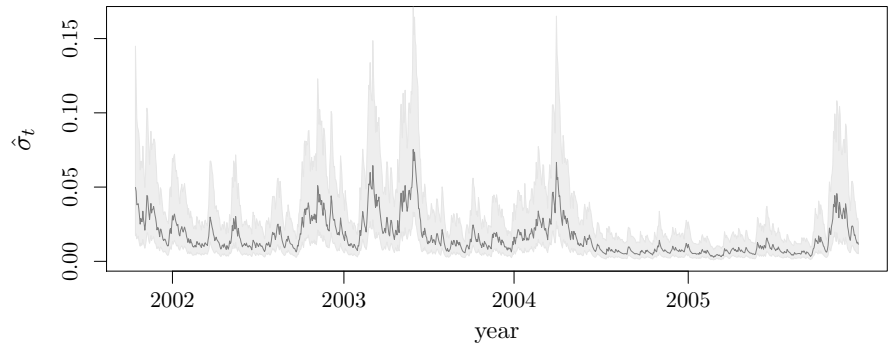
		Kalman filter (2)			Kalman filter (7)		
	True	Mean	Var	MSE	Mean	Var	MSE
β	0.04	0.0452	0.0002	0.0003	0.0452	0.0002	0.0003
κ	0.6	0.6781	0.0502	0.0563	0.6781	0.0502	0.0563
γ	1.5	1.5675	0.1212	0.1257	1.5232	0.0336	0.0342
ϕ_0	-0.010	-0.2519	2.3199	2.3785	-0.0225	0.7764	0.7766
ϕ_1	0.998	0.9776	0.0254	0.0258	1.0154	0.0120	0.0123
ξ	0.4	0.5446	0.1866	0.2075	0.5809	0.0767	0.1094
ρ	-0.5	-0.2564	0.0925	0.6647	-0.2469	0.0588	0.6167

Table 4.10: Parameter estimates for the discretized version of (4.23)*, as in* (4.24)*, for the Kalman filter algorithm with a mixture of two (left) and seven (right) Gaussian distributions, with weekly data and* $n = 2080$*, which corresponds to* $T = 40$ *years.*

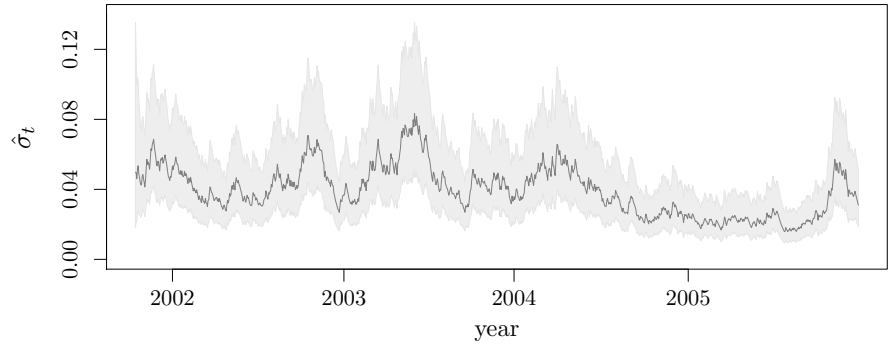
4.4 Real data applications

In this section, we consider the Euribor (Euro Interbank Offered Rate) interest rate series of Chapter 3, corresponding to four maturities (three, six, nine and twelve months), see Figure 3.4. The four datasets expand from October 15th 2001 to December 30th 2005 (sample size of n = 1 077). In Chapter 3 a one-factor CKLS model was fitted, here we fit the CKLS with stochastic volatility in (4.21), as different models can be nested within this unrestricted model.

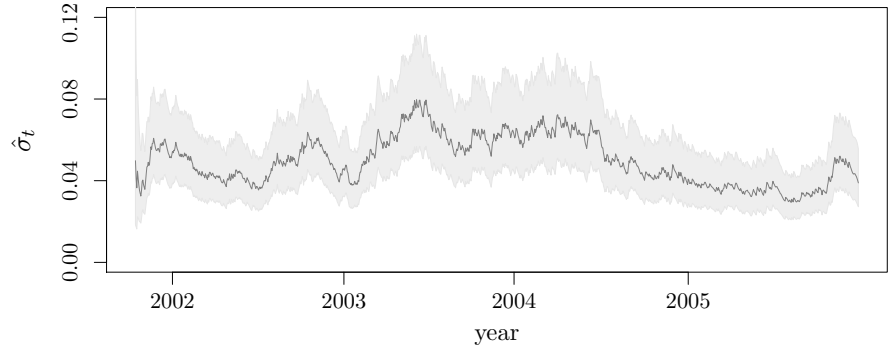
Figure 4.5 illustrates the estimated volatility path $\hat{\sigma}_t$, while Table 4.11 shows the parameter estimations for the CKLS-OU model, using the Kalman filter algorithm, with the



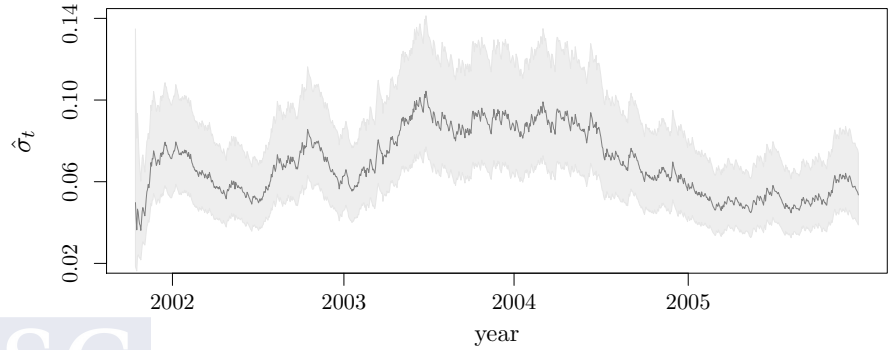
(a) Euribor 3 months.



(b) Euribor 6 months.



(c) Euribor 9 months.



(d) Euribor 12 months.

Figure 4.5: Estimated volatility path of the Euribor series for the CKLS model with stochastic volatility in (4.21)*, with 95% confidence bands.*

associated standard error in parentheses. The values verify the trait usually associated with interest rate time series, that is, persistence, both for the interest rate X_t and for the volatility σ_t equations. Regarding the parameter γ, which controls the relationship between the interest rate level and the volatility, for all maturities we have $\gamma > 1$. This indicates that the volatility tends to increase as the interest rate X_t raises.

Parameters:	β	κ	γ	ϕ_0	ϕ_1	ξ
3 months	1.6546	0.7733	1.6483	-0.2821	0.9674	7.2056
	(0.5980)	(0.2631)	(0.6711)	(0.1696)	(0.0177)	(1.7736)
6 months	1.5139	0.6813	1.4219	-0.0812	0.9878	2.4606
	(0.7567)	(0.3214)	(0.5848)	(0.0728)	(0.0103)	(0.8608)
9 months	1.6850	0.7218	1.7627	-0.0494	0.9919	1.3615
	(1.0196)	(0.4218)	(0.4122)	(0.0479)	(0.0079)	(0.5332)
12 months	1.9218	0.7868	1.6729	-0.0344	0.9937	1.1163
	(1.2384)	(0.4977)	(0.3914)	(0.0358)	(0.0066)	(0.4468)

Table 4.11: Parameter estimates and standard errors (in parentheses) for the CKLS process with stochastic volatility, fitted to Euribor series.

This Euribor time series will be further analyzed in Chapter 7, where a new goodness-of-fit test for two-factor diffusion models is proposed.

4.5 Conclusions

In this chapter we reviewed parametric estimation methods for two-factor continuous-time stochastic volatility models. The continuous time nature of the process does complicate the parameter estimation, as available data are registered in discrete time points. As a consequence, parameters are subject to discretization bias and this, combined with the presence of a latent factor, challenges estimation. We discussed a comparative study of three estimation methods—namely, MCMC, Kalman filter and particle filter—under different settings. The close performance of the procedures, together with the computationally demanding and model-dependent implementation of simulation methods, makes the Kalman filter a computationally efficient estimation method to use in goodness-of-fit testing procedures. Furthermore, the space-state model structure allows to easily implement a bootstrap procedure.

Chapter 5

Nonparametric estimation of diffusion models

With inconclusive empirical studies regarding a correct specification of the different components of diffusion models, the nonparametric framework provides adequate tools to analyze diffusion processes without any assumption on its functional form. This chapter provides a selective review of nonparametric methods for stochastic differential equations, focusing on smoothing techniques. The stationary density, drift and diffusion functions estimation are considered, providing simulation studies of the nonparametric methods and a discussion on bandwidth selection. As real data application, treasury securities with different maturities are analyzed.

Contents

5.1 Introduction

The derivative security pricing theory relies fundamentally on continuous-time arguments, where modeling the dynamics of stock prices, exchange or interest rates is an essential topic in asset pricing. The instantaneous risk-free interest rate, known as the short rate, is of fundamental importance as it determines the yield curve evolution and affects the pricing of fixed-income securities. The underlying process or short rate has been modeled as a time-homogeneous diffusion process, that is, the process $\{X_t\}_{t\geqslant 0}$ is a diffusion solution of the Itô stochastic differential equation defined in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, \mathbb{P})$, with Ω a nonempty set, $\mathcal{F}$ a σ-algebra of subsets of Ω and $\mathbb{P}$ a probability measure such that $\mathbb{P}(\Omega) = 1$,

$$\mathrm{d}X_t = m(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{5.1}$$

with initial condition $X_{t_0} = X_0 \in \mathbb{R}$ and $\{W_t\}_{t\geqslant 0}$ a $\{\mathcal{F}_t\}_{t\geqslant 0}$-adapted standard Wiener process, $m(\cdot)$ the instantaneous mean (drift function) and $\sigma^2(\cdot)$ the instantaneous variance (volatility or diffusion function). Despite the continuous-time nature of the process, continuous sampling in practice is unavailable. In practical situations, the components of the process are estimated from discrete-time observations. Given the Itô process $\{X_t\}_{0\leqslant t\leqslant T}$, solution of the stochastic differential equation in (5.1), the discretized process $\{X_{t_i}\}_{i=0}^n$ is observed at $\{0 = t_0 < t_1 < \ldots < t_n = T\}$, with $\Delta = T/n$ the time interval between consecutive observations $t_i = i\Delta$, with $i = 0, 1, \ldots, n$.

The form of drift and diffusion functions of interest rate are crucial to price securities, but capturing a correct specification by using a functional form of the functions is a difficult task, with inconclusive empirical studies, and there is no a priori information for the unobservable instantaneous volatility. In this context, the nonparametric estimation of the diffusion terms, avoiding any functional specification, constitutes a suitable setup. An overview on the subject is provided in Fan (2005) while Jimenez et al. (2005) focused on the effect of estimating with complete, partial or noisy observations. The first proposal of nonparametric estimation from discrete-time sampling was proposed by Florens-Zmirou (1993), where the diffusion function was estimated nonparametrically using a naive method, leaving the

drift term unspecified. Jiang and Knight (1997) extended the estimator constructed from the indicator function to a Gaussian kernel, while Bandi and Phillips (2003) completed the kernel method approach considering general kernel functions. Aït-Sahalia (1996b) developed a semi-parametric approach, specifying the drift term as a mean-reverting linear function and characterizing the diffusion function through the marginal density and the specified drift. In Stanton (1997), the Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) kernel regression estimator was used to estimate the moments of the process. Other approaches, such as simulated annealing, nearest neighbor, adaptive estimation or wavelet methods, have been proposed (see, e.g., Brugière,1991, 1993; Genon-Catalot et al., 1992; Genon-Catalot and Jacod, 1993, 1994; Hoffmann, 1999; Jacobsen, 2002).

The purpose of this chapter is to provide a selective review of nonparametric methods for diffusion processes, focusing on smoothing techniques. The chapter is organized as follows. Section 5.2 provides some background on diffusion models and the assumptions needed for the nonparametric framework. Section 5.3 introduces the nonparametric estimator of the density function while Section 5.4 focus on the drift and diffusion functions estimation, developing simulation studies and discussing the choice of the bandwidth parameter. Section 5.5 provides an application to the short-term interest rate and conclusions are drawn in Section 5.6.

5.2 Model identification

The stochastic process defined in (5.1) is a strong Markov process with continuous paths where the drift $m(\cdot)$ and diffusion $\sigma^2(\cdot)$ terms are given by

$$m(X_t) = \lim_{\Delta \to 0} \mathbb{E}\left[\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} \;\middle|\; X_t(\omega) = X_t \right], \tag{5.2}$$

$$\sigma^2(X_t) = \lim_{\Delta \to 0} \mathbb{E}\left[\frac{\left(X_{(i+1)\Delta} - X_{i\Delta}\right)^2}{\Delta} \;\middle|\; X_t(\omega) = X_t \right], \tag{5.3}$$

respectively, that is, the mean and variance of X_t for infinitesimal changes in time, with time interval $\{i\Delta \mid i = 0, \ldots, n\}$ and where ω denotes a sample path of X_t. The dynamics of the process due to time changes are described by the drift while the magnitude of the random fluctuation around the drift term are explained by the diffusion function. The limits of the infinitesimal conditional distribution first two moments are given by $m(X_t)\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\sigma(X_t)\colon [0,T] \times \mathbb{R} \to \mathbb{R}^+$, respectively.

Some regularity conditions are needed in order to ensure the existence and uniqueness of

a strong stationary solution to (5.1).

Assumption 5. The initial random variable X_0 defined on the same probability space as X_t is of second order, $\mathbb{E}\left[X_0^2\right] < \infty$.

Assumption 6. The transitional probability $p(X_{(i+1)\Delta} \mid X_{i\Delta})$ is a differentiable, $\mathcal{B}$-measurable function of $X_{i\Delta}$ for fixed $X_{(i+1)\Delta}$ and $\mathcal{B}$-measurable function of $X_{(i+1)\Delta}$ for fixed $X_{i\Delta}$ and satisfies the Chapman-Kolmogorov equation

$$p(X_{(i+1)\Delta} \mid X_{i\Delta}) = \int_{X_\tau} p(X_{(i+1)\Delta} \mid X_\tau) p(\mathrm{d}X_\tau \mid X_{i\Delta}), \qquad i\Delta < \tau < (i+1)\Delta$$

and

$$\lim_{\Delta \to 0} \frac{1}{\Delta} p\left(|X_{(i+1)\Delta} - X_{i\Delta}| \geqslant \epsilon \mid X_{i\Delta}(\omega) = X_{i\Delta}\right) = 0$$

a.s. for every $\epsilon > 0$, with ω denoting a sample path of X_t.

Assumptions 1–3 (see Chapter 2) guarantees that the stochastic differential equation defined in (5.1) has a unique solution $X_t(\omega)$ satisfying

$$X_t = X_0 + \int_0^t m(X_t)\,\mathrm{d}t + \int_0^t \sigma(X_t)\,\mathrm{d}W_t,$$

a.s., with t-continuous components over $[0, T]$ and $\int_0^t \mathbb{E}\left[X_t^2(\omega)\right]\mathrm{d}t < \infty$. Assumptions 1–3 and 5–6 ensure regularity of the underlying Markov process of (5.1), such that both the Kolmogorov backward equation,

$$\begin{aligned} -\frac{\partial p}{\partial t_0}\left(X_t = x \mid X_{t_0} = x_0\right) = m(x_0)\,\frac{\partial}{\partial x_0}\,p\left(X_t = x \mid X_{t_0} = x_0\right) \\ + \frac{1}{2}\,\sigma^2(x_0)\,\frac{\partial^2}{\partial x_0^2}\,p\left(X_t = x \mid X_{t_0} = x_0\right), \end{aligned}$$

and the Kolmogorov forward (Fokker-Planck) equation

$$\begin{aligned} \frac{\partial p}{\partial t}\left(X_t = x \mid X_{t_0} = x_0\right) = -\frac{\partial}{\partial x}\left(m(x)p\left(X_t = x \mid X_{t_0} = x_0\right)\right) \\ + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\left(\sigma^2(x)p\left(X_t = x \mid X_{t_0} = x_0\right)\right), \end{aligned} \tag{5.4}$$

hold true a.s. The transitional probability density function $p(\cdot \mid \cdot)$ of the Markov process is fully characterized by the coefficients under regularity conditions.

5.3 Stationary density estimation

Given the stationarity of the stochastic process X_t, there exists an initial stationary density $\pi(X_{t_0})$ such that

$$\pi\left(X_t = x\right) = \int p\left(X_t = x \mid X_{t_0} = u\right) \pi\left(X_{t_0} = u\right) \mathrm{d}u = \pi\left(X_{t_0} = x\right),$$

for any x. Under the conditions of the previous section, the time-stationary transition density of the Markov process X_t, that is, $p\left(X_{(i+1)\Delta} \mid X_{i\Delta}\right)$, invariant of time t, the Kolmogorov forward equation (5.4) left-hand side becomes zero. Multiplying by the marginal density $\pi(X_{t_0})$ both terms in the right side and integrating with respect to X_{t_0}, we have

$$\pi(X_t) = \frac{\xi}{\sigma^2(X_t)} \exp\left(2 \int_{X^{(0)}}^{X_t} \frac{m(u)}{\sigma^2(u)} \mathrm{d}u\right), \tag{5.5}$$

where $0 < X^{(0)} < +\infty$ is an arbitrary interior point in X_t state space and ξ a standardization factor to ensure $\pi(\cdot)$ integrates to one, with boundary conditions $\pi(+\infty) = \pi'(+\infty) = 0$ or $\sigma^2(+\infty) = \sigma^{2\prime}(+\infty) = 0$. Due to the Gaussian increments of the Winer process, the distributions of the stationary diffusion process X_t, namely the transitional and marginal densities, under regularity conditions are entirely characterized by its first two moments, which are the drift and diffusion functions.

For estimation purposes, more conditions are imposed on the stochastic differential equation (5.1).

Assumption 7. The drift $m(\cdot)$ is a bounded function, twice continuously differentiable and with bounded derivatives.

Assumption 8. The diffusion $\sigma(\cdot)$ is a bounded function, three times continuously differentiable and with bounded derivatives. There exist constants c and C with $0 < c \leqslant \sigma(x) \leqslant C$.

Assumption 9. Let the kernel $K \in L^2(\mathbb{R})$ be a symmetric and bounded probability density function, $\int_{-\infty}^{+\infty} K(x)\,\mathrm{d}x = 1$, twice continuously differentiable and square integrable.

Assumption 8 excludes the case of a singular diffusion process with $\sigma(X_t) = 0$ at some set. Under conditions 1–3 and 5–8, the diffusion process X_t is nonsingular with time-stationary transitional probability function $p(\cdot \mid \cdot)$, and Assumption 9 imposes regularity conditions on the kernel function $K(\cdot)$.

The kernel density estimator of the stationary density $\pi(x)$ is

$$\pi_{n,h}(x) = \frac{1}{nh}\sum_{i=0}^{n} K\left(\frac{x - X_{t_i}}{h}\right) = \frac{1}{n}\sum_{i=0}^{n} K_h(x - X_{t_i}), \tag{5.6}$$

where $K_h(u) = h^{-1}K(u/h)$ is the scaled kernel. The bias and variance of the kernel density estimator at x are given by

$$\mathrm{Bias}\big[\pi_{n,h}(x)\big] = \frac{1}{2}\mu_2(K)\pi''(x)h^2 + \mathcal{O}\left(h^2\right),$$

$$\mathbb{V}\mathrm{ar}\big[\pi_{n,h}(x)\big] = \frac{R(K)}{nh}\pi(x) + \mathcal{O}\left((nh)^{-1}\right),$$

where $\mu_2(K) = \int u^2 K(u)\,\mathrm{d}u$ is the second-order moment of K and $R(K) = \int K(u)^2\,\mathrm{d}u$. Regarding the asymptotic normality of the kernel estimator, assuming that $\int K^{2+\delta}(u)\,\mathrm{d}u < \infty$ for $\delta > 0$, we have

$$\sqrt{nh}\big(\pi_{n,h}(x) - \mathbb{E}\left[\pi_{n,h}(x)\right]\big) \xrightarrow{d} \mathcal{N}\big(0, R(K)\pi(x)\big). \tag{5.7}$$

Regarding the choice of the kernel function and the bandwidth, the Gaussian kernel is commonly used,

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right),$$

and the smoothing parameter is often chosen using Scott's rule (Scott, 1992), where h is proportional to $\mathrm{std}(x)n^{-1/5}$, with $\mathrm{std}(\cdot)$ the standard deviation of the diffusion process. Though Scott's rule is the standard choice in the econometric field, another common choice is to use a cross-validation bandwidth. This bandwidth attempts to minimize the *Mean Integrated Squared Error* (MISE),

$$\begin{aligned}\mathrm{MISE}\left[\pi_{n,h}(\cdot)\right] &= \mathbb{E}\left[\int \left(\pi_{n,h}(x) - \pi(x)\right)^2 \mathrm{d}x\right] \\ &= \mathbb{E}\left[\int \pi_{n,h}(x)^2\,\mathrm{d}x\right] - 2\mathbb{E}\left[\int \pi_{n,h}(x)\pi(x)\,\mathrm{d}x\right] + \int \pi(x)^2\,\mathrm{d}x,\end{aligned}$$

where $\int \pi(x)^2\,\mathrm{d}x$ does not depend on the parameter h, therefore, minimizing $\mathrm{MISE}\left[\pi_{n,h}(\cdot)\right]$ is equivalent to

$$\min \mathbb{E}\left[\int \pi_{n,h}(x)^2\,\mathrm{d}x\right] - 2\mathbb{E}\left[\int \pi_{n,h}(x)\pi(x)\,\mathrm{d}x\right].$$

Due to the unavailability of this quantity, it is estimated by

$$\mathrm{LSCV}(h) \coloneqq \int \pi_{n,h}(x)^2 \,\mathrm{d}x - 2n^{-1} \sum_{i=0}^{n} \pi_{n,h}^{(-i)}(X_{t_i}),$$

with $\pi_{n,h}^{(-i)}(\cdot)$ the *leave-one-out* kernel density estimator, based on the sample without the observation X_{t_i}, that is,

$$\pi_{n,h}^{(-i)}(x) = \frac{1}{n-1} \sum_{\substack{j=0 \\ j \neq i}}^{n} K_h\left(x - X_{t_j}\right).$$

This can also be extended to a *leave-*$(2l+1)$*-out* version, where the observations $\{X_{t_i}\}_{i=-l}^{l}$ are removed from the sample. The Least Squares Cross-Validation (LSCV) selector is defined as

$$\hat{h}_{\mathrm{LSCV}} \coloneqq \arg\min_{h>0} \mathrm{LSCV}(h).$$

The estimation of diffusion models has proven difficult in the financial context, as the intrinsic persistence of interest rate time series can hinder the performance of the estimators. As an example, Figure 5.1 shows the true density ($\pi(x)$) and the nonparametric estimation ($\pi_{n,h}(x)$) for a simulated CIR (Cox et al., 1985) process,

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t,$$

with $(\mu, \kappa, \sigma^2)^\top = (3, 0.2, 0.05)^\top$, with sample size $n = 900$ and monthly ($\Delta = 1/12$) frequency. For the kernel density estimation a Gaussian kernel was used and the smoothing parameter h was estimated using Scott's rule. Persistence, small sample sizes or sampling in a high frequency could provide less accurate estimates. This is illustrated in Figure 5.2, where the effect of sample size n and discretization step Δ on the kernel estimation is displayed by simulating the same CIR process but increasing the sample size to $n = 12\,000$ while keeping Δ fixed (see Figure 5.2a) or extending the observation window $[0, T]$ by sampling less frequency while keeping the same number of observations (see Figure 5.2b). The estimated marginal density in Figure 5.2 is closer to the true density, as increasing the time horizon $T = n\Delta$ provides more accurate estimates (see the discussion for the parametric estimation in Section 3.3.3).

In the next sections, we developed simulation studies to compare the kernel density estimation with its parametric counterpart and to study the performance of different bandwidths, taking into account the persistence of financial diffusion models.

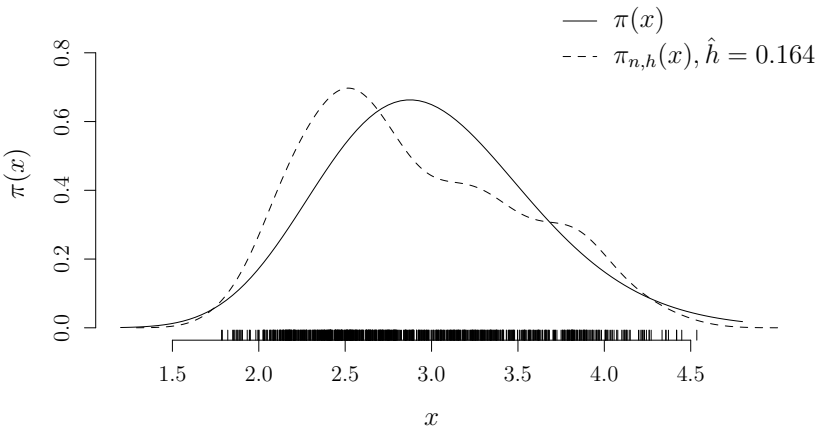
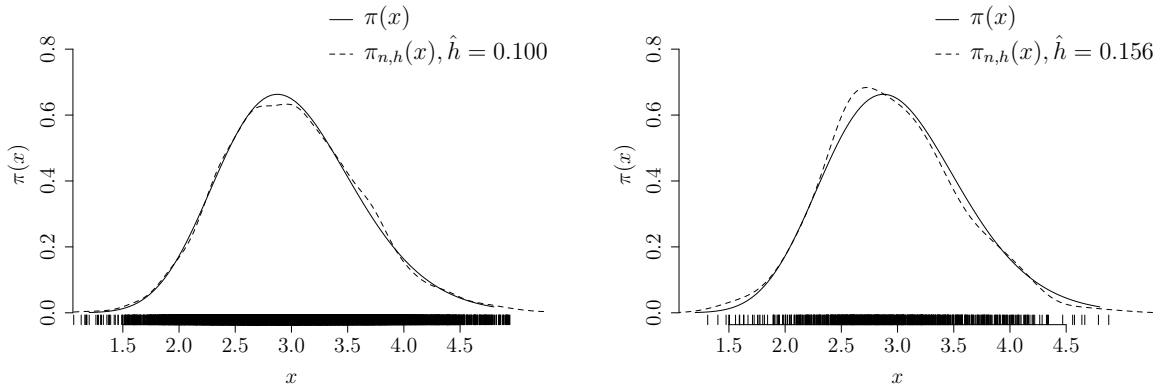


Figure 5.1: Nonparametric estimation of the stationary density for the CIR process, with sample size $n = 900$ and $\Delta = 1/12$.



(a) Sample size $n = 12\,000$ and $\Delta = 1/12$. *(b) Sample size $n = 900$ and $\Delta = 1$.*

Figure 5.2: Effect of the sample size n and discretization step Δ on the stationary density kernel estimation $\pi(r)$ for the CIR process.

5.3.1 Simulation study

A simulation was carried out to compare the parametric and nonparametric estimation of the stationary density. We consider $1\,000$ repetitions, with $n = 900$, and as a criteria for comparing both density estimations we chose the Kullback-Leibler divergence ($\mathrm{D_{KL}}$), which, given two probability density functions $p(\cdot)$ and $q(\cdot)$, is defined as the integral

$$\mathrm{D_{KL}}\big(p(x) \parallel q(x)\big) := \int p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x.$$

Note that the Kullback-Leibler divergence $\mathrm{D_{KL}}$ is zero if and only if $p = q$.

We simulated a thousand paths of the CIR process $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t$, with $(\mu, \kappa, \sigma^2)^\top = (3, 0.2, 0.05)^\top$, sample size $n = 900$ and monthly sampling frequency ($\Delta = 1/12$). The kernel density estimator used was introduced in (5.6), while the parametric estimation of the stationary density (5.5) is

$$\pi_{\hat{\theta}}(X_t) = \frac{\xi(\hat{\boldsymbol{\theta}})}{\sigma^2(X_t, \hat{\boldsymbol{\theta}})} \exp\left(2\int_{X^{(0)}}^{X_t} \frac{m(u, \hat{\boldsymbol{\theta}})}{\sigma^2(u, \hat{\boldsymbol{\theta}})}\,\mathrm{d}u\right), \tag{5.8}$$

with $\hat{\boldsymbol{\theta}}$ a root-n consistent estimator of the true parameter $\boldsymbol{\theta}$ (see Chapter 3). For the CIR process we obtain a Gamma law with $2\mu\kappa/\sigma^2$ as shape parameter and $\sigma^2/(2\kappa)$ as scale parameter. Figure 5.3 shows an example of the true density of the process ($\pi(x)$), the nonparametric estimation ($\pi_{n,h}(x)$), using a Gaussian kernel and Scott's rule to estimate h, and the parametric estimation ($\pi_{\hat{\theta}}(x)$).

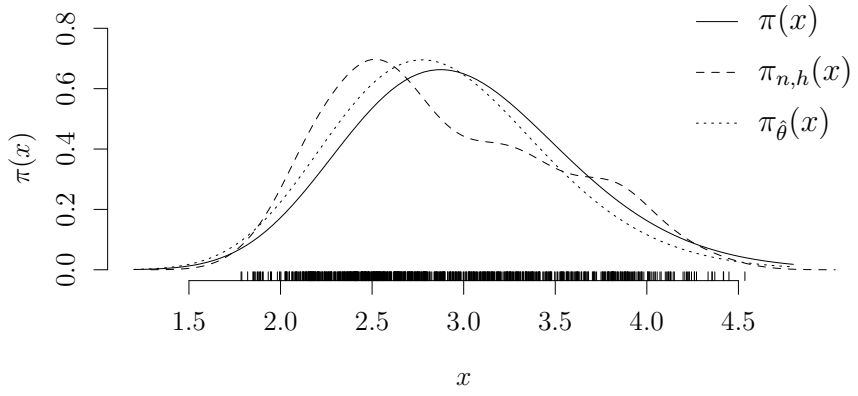


Figure 5.3: Stationary density for the CIR process. Solid line represents the true density $\pi(x)$, dashed line is the nonparametric estimation $\pi_{n,h}(\cdot)$ and dotted line is the parametric estimation $\pi_{\hat{\theta}}(\cdot)$.

We calculated the relative entropy from the parametric estimation $\pi_{\hat{\theta}}(\cdot)$ to the true density $\pi(\cdot)$, that is, $\mathrm{D}_{\mathrm{KL}}\big(\pi(x) \parallel \pi_{\hat{\theta}}(x)\big)$, and the relative entropy from the nonparametric estimation $\pi_{n,h}(\cdot)$ to the true stationary density, that is, $\mathrm{D}_{\mathrm{KL}}\big(\pi(x) \parallel \pi_{n,h}(x)\big)$ for the simulated paths, to measure how close the estimated density is to the true distribution. Results are shown in Table 5.1, including the quartiles and mean of D_{KL} for the thousand simulations, where the first row corresponds to the parametric estimation and the nonparametric estimation was conducted for different bandwidths h. Figure 5.4 depicts the boxplots for the Kullback-Leibler divergence between the true and the estimated—parametric and nonparametric—stationary density.

Bandwidth	Min.	Quartile 25%	Quartile 50%	Quartile 75%	Max.	Mean
Parametric	0.0001	0.0257	0.0642	0.1325	0.7298	0.0942
Scott's rule	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

Table 5.1: Kullback-Leibler divergence for different bandwidths h.

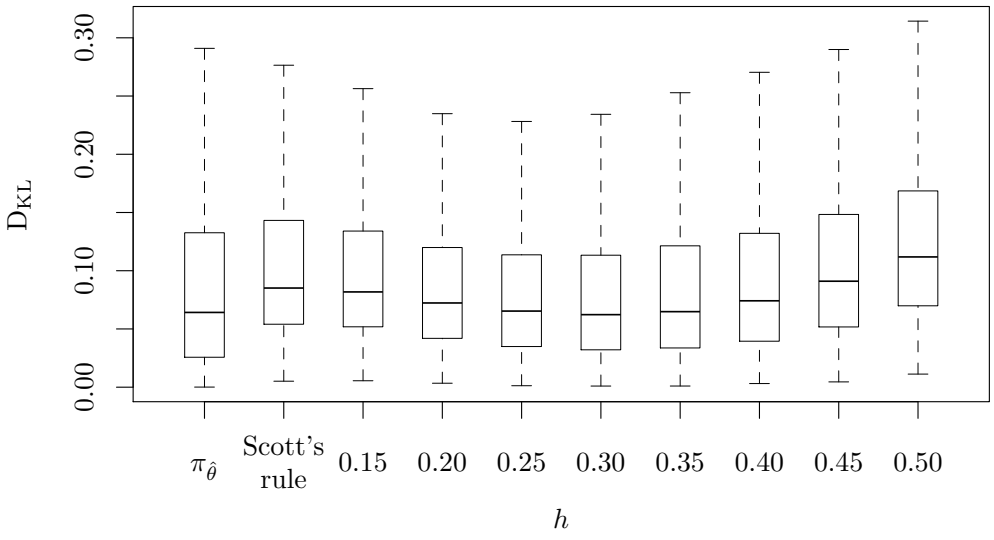


Figure 5.4: Boxplot for the Kullback-Leibler divergence between the true and the estimated stationary density, for the parametric estimation ($\pi_{\hat{\theta}}$) and the kernel estimation with different values of h.

The parametric estimation of the marginal density plays the role of a benchmark, as we are assuming the knowledge of the true parametric specification of the process, unknown in practice, and only the parameter vector $(\mu, \kappa, \sigma)^\top$ needs to be estimated. On the other hand, the nonparametric estimation does not use any prior information of the process X_t and, nonetheless, it achieves a very close performance to the parametric estimation.

5.3.2 Bandwidth selection and persistence

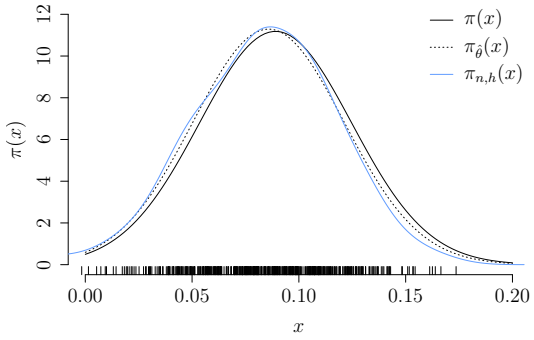
As stated in the previous sections, the intrinsic persistence of financial time series can hinder the estimation of the components of the process. As an illustration of the effect of persistence on the performance of the estimation, Figure 5.5 shows the true stationary density (solid line), the kernel estimation (blue line) and the parametric estimation (dotted line) for three levels of persistence of two diffusion models: the CIR ($\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{x_t}\,\mathrm{d}W_t$) and the Vasicek ($\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$) models, with different parameters (μ, κ, σ). The marginal density was kept unchanged while varying the speed of mean reversion (κ), therefore we have the same stationary density $\pi(\cdot)$ for the Vasicek process (see Figures 5.5a–5.5c) and the CIR model (see Figures 5.5d–5.5f), with increasing (decreasing) levels of persistence (mean reversion). In the less persistent scenarios of both models (Figures 5.5a and 5.5d), the performance of the estimation is very close to the true density, but when persistence increases (Figures 5.5c and 5.5f), mean reversion decreases and both estimates worsen.

To analyze this effect, we conducted a simulation study for both Vasicek and CIR models, simulating three persistence scenarios for either model. Regarding the Vasicek process, we simulate a low $(\mu, \kappa, \sigma^2)^\top = (0.089102, 1.71674, 0.0043708)^\top$, medium $(\mu, \kappa, \sigma^2)^\top = (0.089102, 0.85837, 0.0021854)^\top$ and high $(\mu, \kappa, \sigma^2)^\top = (0.089102, 0.429185, 0.00109274)^\top$ persistence process, while keeping the same marginal density. For the CIR model, we consider low persistence, or high mean reversion, $(\mu, \kappa, \sigma^2)^\top = (0.09045, 1.78436, 0.065484)^\top$, medium $(\mu, \kappa, \sigma^2)^\top = (0.09045, 0.89218, 0.032742)^\top$ and high persistence $(\mu, \kappa, \sigma^2)^\top = (0.09045, 0.44609, 0.016371)^\top$ levels. Again, the three scenarios have the same stationary density.

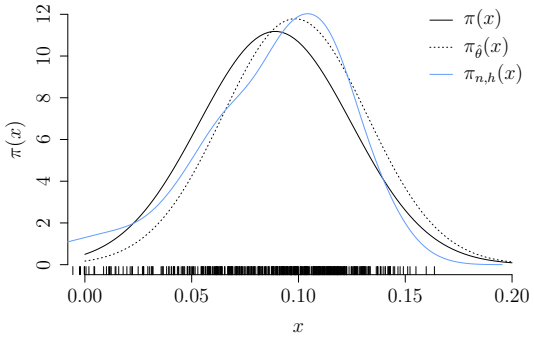
For the simulation study, we consider four bandwidths: Scott's rule and the leave-$(2l+1)$ cross-validation (CV), with $l \in \{0, 1, 2\}$, with Gaussian kernel. As in the previous section, we use the Kullback-Leibler divergence between the true density and the nonparametric estimation, $\mathrm{D_{KL}}\big(\pi(x) \parallel \pi_{n,h}(x)\big)$. We also calculate the *Integrated Squared Error* (ISE),

$$\mathrm{ISE}\left[\pi_{n,h}(\cdot)\right] \coloneqq \int \big(\pi_{n,h}(x) - \pi(x)\big)^2\,\mathrm{d}x$$

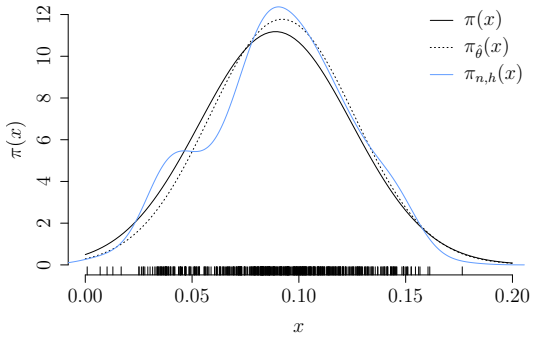
for the thousand path simulated. Table 5.2 shows the results for the Vasicek model, while Table 5.3 for the CIR process, for the different persistence levels, depicting the quartiles and mean of the Kullback-Leibler divergence ($\mathrm{D_{KL}}$) and the MISE and standard deviation (SD) of the ISE.



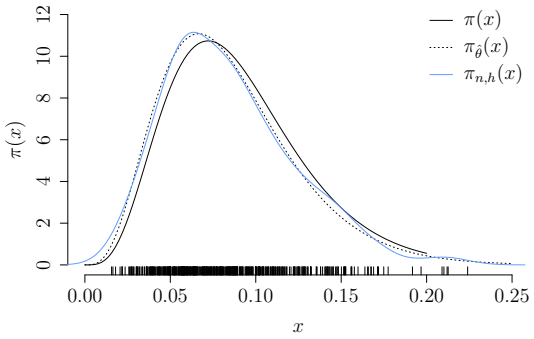
(a) Comparison for the Vasicek model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.089102, 1.71674, 0.0043708)^\top$.



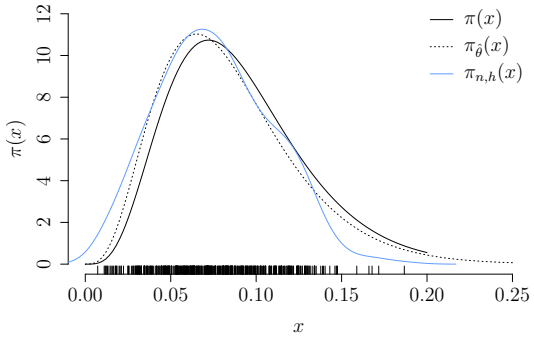
(b) Comparison for the Vasicek model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.089102, 0.85837, 0.0021854)^\top$.



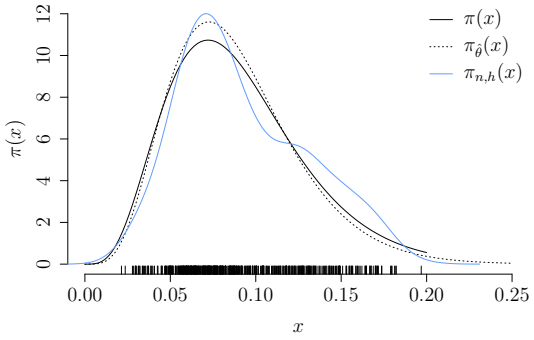
(c) Comparison for the Vasicek model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.089102, 0.42918, 0.0010927)^\top$.



(d) Comparison for the CIR model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.09045, 1.78436, 0.065484)^\top$.



(e) Comparison for the CIR model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.09045, 0.89218, 0.032742)^\top$.



(f) Comparison for the CIR model with parameter values given by $(\mu, \kappa, \sigma^2)^\top = (0.09045, 0.44609, 0.016371)^\top$.

Figure 5.5: Comparison among true stationary density $\pi(x)$*, solid line, parametric estimation* $\pi_{\hat{\theta}}(x)$*, dotted line, and nonparametric estimation* $\pi_{n,h}(x)$*, blue line. Figures (a)–(c) correspond to the Vasicek model, and (d)–(f) to the CIR model, from less* (a *and* d) *to more* (c *and* f) *persistent.*

Regarding the Vasicek model simulation (Table 5.2), there is a clear improvement on the estimates when the mean reversion of the process increases, for all of the bandwidths considered, with the high persistence scenario providing the worst estimation of the marginal density, kept unchanged in all the scenarios. In respect of the bandwidth choice, the leave-$(2l+1)$ cross-validation with $l = 2$ provides a very close performance to Scott's rule, both outperforming the leave-$(2l+1)$ cross-validation with $l = 0$ and $l = 1$.

Pers.	Bandwidth	$\mathrm{D_{KL}}$						ISE	
		Min.	25%	50%	75%	Max.	Mean	MISE	SD
High									
	Scott's rule	0.0059	0.0429	0.0713	0.1248	0.6783	0.0954	0.7247	0.6361
	CV ($l = 0$)	0.0058	0.0505	0.0845	0.1325	0.6759	0.1045	0.8450	0.6884
	CV ($l = 1$)	0.0053	0.0431	0.0767	0.1249	0.6707	0.0972	0.7770	0.6671
	CV ($l = 2$)	0.0053	0.0409	0.0727	0.1227	0.6707	0.0948	0.7533	0.6576
Mid									
	Scott's rule	0.0023	0.0219	0.0389	0.0645	0.3462	0.0508	0.3835	0.3415
	CV ($l = 0$)	0.0024	0.0239	0.0431	0.0704	0.3596	0.0541	0.4217	0.3572
	CV ($l = 1$)	0.0024	0.0218	0.0398	0.0683	0.3575	0.0516	0.4029	0.3522
	CV ($l = 2$)	0.0024	0.0212	0.0393	0.0664	0.3565	0.0508	0.3933	0.3451
Low									
	Scott's rule	0.0011	0.0124	0.0214	0.0352	0.1550	0.0274	0.2004	0.1778
	CV ($l = 0$)	0.0015	0.0132	0.0231	0.0378	0.1526	0.0291	0.2232	0.1874
	CV ($l = 1$)	0.0008	0.0128	0.0222	0.0370	0.1515	0.0285	0.2182	0.1873
	CV ($l = 2$)	0.0008	0.0124	0.0220	0.0369	0.1515	0.0282	0.2170	0.1864

Table 5.2: Kullback-Leibler divergence $\mathrm{D_{KL}}$, with different bandwidths, for the Vasicek model under different persistence scenarios, keeping the marginal density unchanged.

Similar conclusions can be reached concerning the simulation of the CIR process (Table 5.3). Higher persistence levels provide worse estimations and, though the performance of the different bandwidths are close, Scott's rule and cross-validation with $l = 2$ outperform the other bandwidth choices. Figure 5.6 represents the boxplots of the Kullback-Leibler divergence $\mathrm{D_{KL}}$ for the Vasicek (left) and CIR (right) models simulation, for the three persistence scenarios and the four bandwidths. Both figures show how the Kullback-Leibler divergence decreases with persistence.

Pers.	Bandwidth	D_{KL}						ISE	
		Min.	25%	50%	75%	Max.	Mean	MISE	SD
High									
	Scott's rule	0.0089	0.0470	0.0839	0.1357	0.5679	0.1034	0.7887	0.6893
	CV ($l = 0$)	0.0108	0.0530	0.0881	0.1375	0.5735	0.1070	0.8892	0.7282
	CV ($l = 1$)	0.0081	0.0501	0.0864	0.1364	0.5687	0.1051	0.8152	0.7039
	CV ($l = 2$)	0.0081	0.0501	0.0864	0.1367	0.5671	0.1050	0.7885	0.6928
Mid									
	Scott's rule	0.0048	0.0323	0.0520	0.0804	0.3581	0.0632	0.4043	0.3574
	CV ($l = 0$)	0.0048	0.0327	0.0500	0.0802	0.3542	0.0622	0.4378	0.3702
	CV ($l = 1$)	0.0054	0.0336	0.0526	0.0807	0.3543	0.0643	0.4144	0.3694
	CV ($l = 2$)	0.0060	0.0338	0.0540	0.0818	0.3543	0.0653	0.4050	0.3651
Low									
	Scott's rule	0.0043	0.0242	0.0345	0.0527	0.1904	0.0413	0.2137	0.1834
	CV ($l = 0$)	0.0045	0.0229	0.0325	0.0494	0.1914	0.0392	0.2323	0.1945
	CV ($l = 1$)	0.0055	0.0243	0.0349	0.0525	0.1952	0.0416	0.2232	0.1897
	CV ($l = 2$)	0.0053	0.0250	0.0360	0.0538	0.1952	0.0427	0.2202	0.1897

Table 5.3: Kullback-Leibler divergence D_{KL}, with different bandwidths, for the CIR model under different persistence scenarios, keeping the marginal density unchanged.

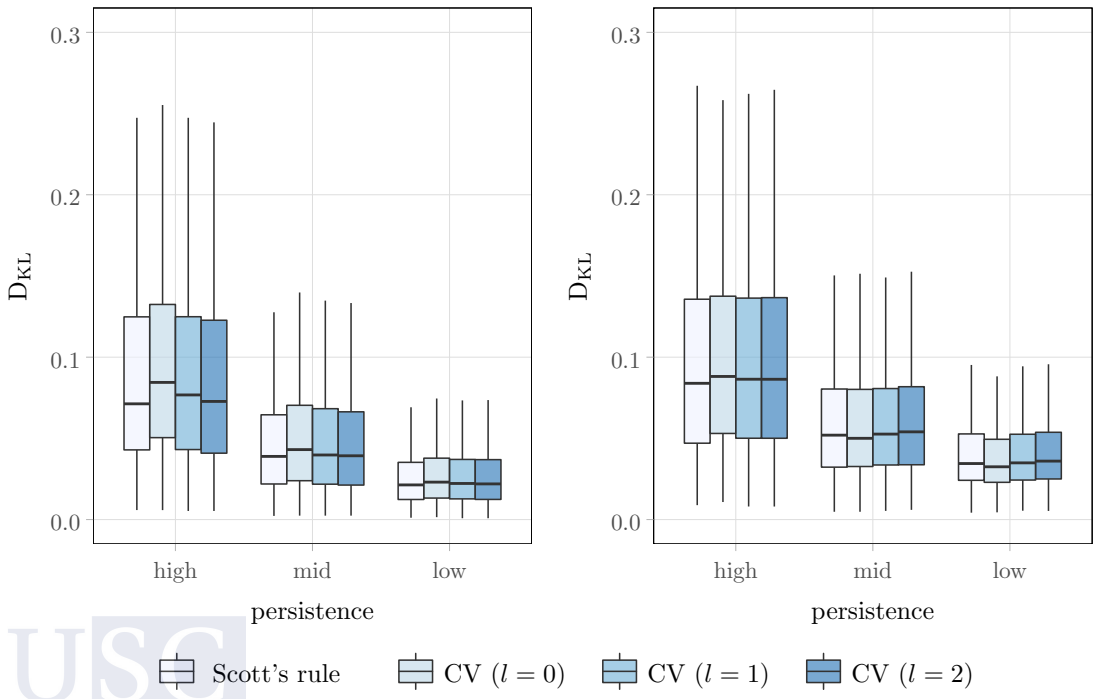


Figure 5.6: Boxplots for Kullback-Leibler divergence D_{KL} for different bandwidths for Vasicek (left) and CIR (right) model with different persistence levels.

5.4 Regression function estimation

Taking the Kolmogorov forward equation in (5.4) and letting $t \to \infty$ we obtain the ordinary differential equation

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\sigma^2(x)\pi(x)\right) = 2\frac{\mathrm{d}}{\mathrm{d}x}\left(m(x)\pi(x)\right), \tag{5.9}$$

which establishes a relationship among the drift function $m(\cdot)$, the diffusion term $\sigma(\cdot)$ and the stationary density $\pi(\cdot)$. With the boundary condition $\pi(+\infty) = \pi'(+\infty) = 0$, by integrating (5.9) we obtain

$$m(x) = \frac{1}{2\pi(x)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\sigma^2(x)\pi(x)\right),$$

and integrating (5.9) twice with the boundary condition $\pi(0) = 0$ yields

$$\sigma^2(x) = \frac{2}{\pi(x)}\int_0^x m(u)\pi(u)\,\mathrm{d}u.$$

Therefore, given a functional specification for either the drift function or the volatility, the other function can be estimated given the stationary density $\pi(\cdot)$ of the process. This semi-parametric approach was studied in Aït-Sahalia (1996b). We will focus on a fully nonparametric procedure, avoiding imposing parametric restrictions on the drift or diffusion functions. Given the infinitesimal generator $\mathscr{L}$ of the process $\{X_t\}$,

$$\begin{aligned}\mathscr{L}f(x,t) &= \lim_{\tau \downarrow t}\frac{\mathbb{E}\left[f(X_\tau,\tau) \mid X_t = x\right] - f(x,t)}{\tau - t} \\ &= \frac{\partial f(x,t)}{\partial t} + \frac{\partial f(x,t)}{\partial x}m(x) + \frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2}\sigma^2(x),\end{aligned}$$

with $f(\cdot)$ an appropriate function (see, e.g., Hille and Phillips, 1957), the conditional expectation $\mathbb{E}\left[f(X_{t_{i+1}}) \mid \mathcal{F}_{t_i}\right]$ can be written in the form of an Itô-Taylor series expansion

$$\begin{aligned}\mathbb{E}\left[f(X_{t_{i+1}}) \mid \mathcal{F}_{t_i}\right] = f(X_{t_i}) + \mathscr{L}f(X_{t_i})\Delta + \frac{1}{2}\mathscr{L}^2 f(X_{t_i})\Delta^2 + \cdots \\ + \frac{1}{n!}\mathscr{L}^n f(X_{t_i})\Delta + \mathcal{O}(\Delta^{n+1}).\end{aligned}$$

As developed in Stanton (1997), given a suitable choice of the function $f(\cdot)$, the latter equation can be used to construct approximations to the drift $m(\cdot)$ and diffusion $\sigma(\cdot)$ functions by rewriting the equation as

$$\mathscr{L}f(X_{t_i}) = \frac{1}{\Delta}\mathbb{E}\left[f(X_{t_{i+1}}) - f(X_{t_i}) \mid \mathcal{F}_{t_i}\right] - \frac{1}{2}\mathscr{L}^2 f(X_{t_i})\Delta - \frac{1}{6}\mathscr{L}^3 f(X_{t_i})\Delta^2 - \cdots.$$

Thus, the first order approximation of $\mathscr{L}f(\cdot)$ is given by

$$\mathscr{L}f(X_{t_i}) = \frac{1}{\Delta}\,\mathbb{E}\big[f(X_{t_{i+1}}) - f(X_{t_i}) \;\big|\; \mathcal{F}_{t_i}\big] + \mathcal{O}(\Delta), \tag{5.10}$$

see Stanton (1997) for higher order approximations. Therefore, to approximate a particular function $g(x,t)$, we need to find a function $f(\cdot)$ satisfying $\mathscr{L}f(x,t) = g(x,t)$. In the next sections, the function $f(\cdot)$ is chosen in order to obtain estimators for the drift and diffusion functions.

5.4.1 Nonparametric estimation of the drift function

An approximation to the drift function $m(\cdot)$ can be derived by considering the function $f_1(x,t) = x$, where from the definition of the infinitesimal operator $\mathscr{L}$ we have

$$\mathscr{L}f_1(x,t) = m(x),$$

and so $\mathscr{L}f_1(X_t) = m(X_t)$. Substituting into equation (5.10) leads to the first order approximation for $m(\cdot)$

$$m(x) = \frac{1}{\Delta}\mathbb{E}\big[X_{t_{i+1}} - X_{t_i} \;\big|\; X_{t_i} = x\big] + \mathcal{O}(\Delta).$$

The conditional expectation can be estimated nonparametrically as

$$\hat{\mathbb{E}}\big[X_{t_{i+1}} - X_{t_i} \;\big|\; X_{t_i} = x\big] = \frac{\sum_{i=0}^{n-1} K\left(\frac{x-X_{t_i}}{h}\right)\left(X_{t_{i+1}} - X_{t_i}\right)}{\sum_{i=0}^{n-1} K\left(\frac{x-X_{t_i}}{h}\right)},$$

obtaining the Nadaraya-Watson (Nadaraya, 1964; Watson, 1964) kernel regression estimator of the instantaneous conditional mean in (5.2), that is, the drift function estimator

$$m_{n,h}(x;0) = \sum_{i=0}^{n-1} W_i^{(0)}(x)\left(X_{t_{i+1}} - X_{t_i}\right), \tag{5.11}$$

with weights

$$W_i^{(0)}(x) = \frac{K_h\left(x - X_{t_i}\right)}{\Delta\sum_{j=1}^{n-1} K_h\left(x - X_{t_j}\right)}. \tag{5.12}$$

The Nadaraya-Watson kernel estimator is a particular case of a wider class of estimators called *local polynomial estimators*, which performs a local constant fit, $m_{n,h}(x;p)$ with $p = 0$.

We can also consider the local linear estimator ($p = 1$), given by

$$m_{n,h}(x;1) = \sum_{i=0}^{n-1} W_i^{(1)}(x) \left(X_{t_{i+1}} - X_{t_i}\right)$$

with weights

$$W_i^{(1)}(x) = \frac{1}{T} \frac{\hat{s}_2(x,h) - \hat{s}_1(x,h)\,(x - X_{t_i})}{\hat{s}_2(x,h)\hat{s}_0(x,h) - \hat{s}_1(x,h)^2} K_h\left(x - X_{t_i}\right), \tag{5.13}$$

where $\hat{s}_r(x,h) := n^{-1} \sum_{i=0}^{n-1} (x - X_{t_i})^r K_h\left(x - X_{t_i}\right)$.

The conditional bias and variance of the local polynomial estimator $m_{n,h}(x;p)$, with $p = 0$ the local constant and $p = 1$ the local linear estimator, are

$$\begin{aligned}
\mathrm{Bias}\big[m_{n,h}(x;p) \mid X_{t_0}, \ldots, X_{t_n}\big] &= B_p(x)h^2 + \mathcal{O}_{\mathbb{P}}\left(h^2\right), \\
\mathbb{V}\mathrm{ar}\big[m_{n,h}(x;p) \mid X_{t_0}, \ldots, X_{t_n}\big] &= \frac{R(K)}{nh\pi(x)}\sigma^2(x) + \mathcal{O}_{\mathbb{P}}\left((nh)^{-1}\right),
\end{aligned}$$

where

$$B_p(x) := \begin{cases} \frac{\mu_2(K)}{2}\left(m''(x) + 2\frac{m'(x)\pi'(x)}{\pi(x)}\right), & \text{if } p = 0, \\ \frac{\mu_2(K)}{2} m''(x), & \text{if } p = 1, \end{cases}$$

Regarding the asymptotic normality of the estimator, assuming that $\mathbb{E}\big[\,(Y_{t_i} - m(x))^{2+\delta} \mid X_{t_i} = x\big] < \infty$ for $\delta > 0$, where $Y_{t_i} = \left(X_{t_i} - X_{t_{i-1}}\right)/\Delta$, we have

$$\sqrt{nh}\Big(m_{n,h}(x;p) - \mathbb{E}\left[m_{n,h}(x;p) \mid X_{t_0}, \ldots, X_{t_n}\right]\Big) \xrightarrow{d} \mathcal{N}\left(0, \frac{R(K)\sigma^2(x)}{\pi(x)}\right). \tag{5.14}$$

5.4.2 Nonparametric estimation of the diffusion function

Analogously to the drift function, an approximation of the diffusion function $\sigma^2(\cdot)$ can be derived by considering the function $f_2(x,t) = (x - X_t)^2$, where from the definition of the infinitesimal operator $\mathscr{L}$ we have

$$\mathscr{L} f_2(x,t) = 2\,(x - X_t)\, m(x) + \sigma^2(x),$$

therefore,

$$\mathscr{L} f_2(X_t) = \sigma^2(X_t).$$

Substituting into equation (5.10) leads to the first order approximation for $\sigma^2(\cdot)$,

$$\sigma^2(x) = \frac{1}{\Delta}\mathbb{E}\big[\left(X_{t_{i+1}} - X_{t_i}\right)^2 \mid X_{t_i} = x\big] + \mathcal{O}(\Delta),$$

and taking the squared root, using a binomial approximation, leads to an approximation of the same order for $\sigma(\cdot)$, that is,

$$\sigma(x) = \sqrt{\Delta^{-1}\mathbb{E}\big[\left(X_{t_{i+1}} - X_{t_i}\right)^2 \mid X_{t_i} = x\big]} + \mathcal{O}(\Delta).$$

Alternatively, replacing $\mathbb{E}\big[\left(X_{t_{i+1}} - X_{t_i}\right)^2 \mid X_{t_i} = x\big]$ by the conditional variance of $X_{t_{i+j}}$, leading to the approximation

$$\sigma(x) = \sqrt{\Delta^{-1}\mathbb{V}\mathrm{ar}\big[X_{t_{i+1}} \mid X_{t_i} = x\big]} + \mathcal{O}(\Delta),$$

which does not change the order of the convergence. The conditional variance can be estimated nonparametrically as

$$\hat{\mathbb{E}}\big[\left(X_{t_{i+1}} - X_{t_i}\right)^2 \mid X_{t_i} = x\big] = \frac{\sum_{i=0}^{n-1} K\left(\frac{x-X_{t_i}}{h}\right)\left(X_{t_{i+1}} - X_{t_i}\right)^2}{\sum_{i=0}^{n-1} K\left(\frac{x-X_{t_i}}{h}\right)},$$

obtaining the Nadaraya-Watson kernel regression estimator of the instantaneous conditional variance in (5.3). Thus, the local constant ($p = 0$) and the local linear ($p = 1$) estimator of the diffusion function is

$$\sigma^2_{n,h}(x;p) = \sum_{i=0}^{n-1} W_i^{(p)}(x)\left(X_{t_{i+1}} - X_{t_i}\right)^2, \tag{5.15}$$

with weights given in (5.12)–(5.13) for $p \in \{0,1\}$. Regarding the asymptotic normality of the estimator we have

$$\sqrt{nh}\Big(\sigma^2_{n,h}(x) - \mathbb{E}\left[\sigma^2_{n,h}(x) \mid X_{t_0},\ldots,X_{t_n}\right]\Big) \xrightarrow{d} \mathcal{N}\left(0, \frac{R(K)V(x)}{\pi(x)}\right), \tag{5.16}$$

where

$$V(x) = \mathbb{E}\left[u^2_{t_i} \mid X_{t_i} = x\right] - \left(\sigma^2(x)\right)^2,$$

with $u_{t_i} = (X_{t_{i+1}} - X_{t_i})^2/\Delta$.

5.4.3 Simulation study and bandwidth selection

As both Nadaraya-Watson ($p = 0$) and local linear ($p = 1$) estimators heavily depend on the smoothing parameter h, we conducted a simulation study to compare the performance of different bandwidths. As in the previous simulation study for the marginal density, we use the rule-of-thumb bandwidth and cross-validation. Let $Y_{t_i} = \left(X_{t_{i+1}} - X_{t_i}\right)/\Delta$, the least squares cross-validation error is given by

$$\mathrm{CV}(h) := \frac{1}{n}\sum_{i=0}^{n-1}\left(Y_{t_i} - m_{n,h}^{(-i)}(X_{t_i}; p)\right)^2,$$

with $m_{n,h}^{(-i)}(X_{t_i}; p)$ the leave-one-out estimate of the drift $m(\cdot)$ computed removing the i-th observation (Y_{t_i}, X_{t_i}), and the cross-validation bandwidth is estimated as

$$\hat{h}_{\mathrm{CV}} := \underset{h>0}{\arg\min}\, \mathrm{CV}(h),$$

and analogously for the diffusion function estimator $\sigma_{n,h}(x; p)$. The leave-one-out cross-validation can be extended to the leave-$(2l+1)$-out version, removing the $\{(Y_{t_i}, X_{t_i})\}_{i=-l}^{l}$ observations from the sample. As we discussed in the previous section, the presence of persistence affects estimation. In order to take this into account when estimating the bandwidth, we use a data-driven bandwidth selection from De Brabanter et al. (2018), where the authors developed a bandwidth selection criterion in a random design setting with correlated errors and unknown correlation structure. The bandwidths selected with the method proposed in De Brabanter et al. (2018) use the three kernel functions in Figure 5.7. In this method, first a bimodal kernel satisfying $K(0) = 0$ is used to obtain a cross-validation pilot bandwidth, $\hat{h}_{\mathrm{bimd}}$. Then, the optimal bandwidth is calculated by

$$\hat{h}_u = \frac{C_p(K)}{C_p(K_u)}\hat{h}_{\mathrm{bimd}},$$

with

$$C_p(K) = \left[\frac{(p+1)!^2 \int K_p^{*2}(t)\,\mathrm{d}t}{2(p+1)\left(\int t^{p+1} K_p^{*}(t)\,\mathrm{d}t\right)^2}\right]^{1/(2p+3)},$$

$u \in \{1, 2, 3\}$ corresponding to the bimodal kernel of choice, see Figure 5.7, with $p = 0$ for the local constant and $p = 1$ the local linear estimators and

$$K_p^*(t) = (1, 0, \ldots, 0)^\top \hat{S}^{-1}(1, t, \ldots, t^p)^\top K(t),$$

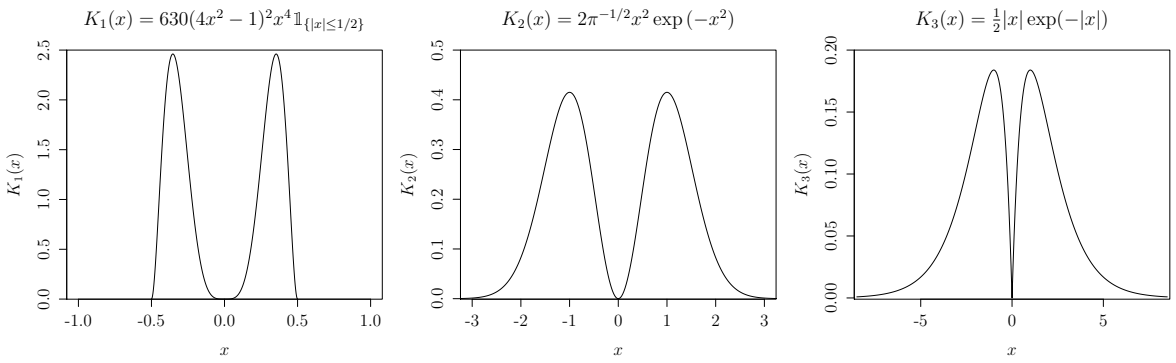


Figure 5.7: Kernel functions K_u satisfying $K(0) = 0$.

where $S = (\mu_{i+j})_{0\leqslant i,\, j\leqslant p}$ and $\mu_j = \int y^j K(y)\,\mathrm{d}y$.

We simulate three models, the CIR process, $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t$, the Vasicek diffusion, $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, for a high and low persistence scenarios (see Section 5.3.2), with weekly sampling ($\Delta = 1/52$) for $n \in \{520, 2080\}$, which correspond to $T = 10$ and 40 years of data, and a diffusion process with nonlinear drift function, with $m(x) = -x + \sin(8x)\,\exp(-x^2)$ and $\sigma(x) = 0.4$. For the drift estimator in (5.11) and the diffusion estimator (5.15) with $p = 0$, a Gaussian kernel is used and five different bandwidths will be calculated:

(i) Rule-of-thumb, $h_{\mathrm{RT}} = \mathrm{std}(x)n^{-1/5}$.

(ii) Leave-$(2l{+}1)$-out cross-validation bandwidth, $h^l_{\mathrm{CV}} = \arg\min \mathrm{CV}(h)$, with $l \in \{0, 1, 2\}$.

(iii) De Brabanter et al. (2018) bandwidth (henceforth denoted as h_1) using the kernel bimodal function $K_1(x) = 630(4x^2 - 1)^2x^4\mathbb{1}_{\{|x|\leqslant 1/2\}}$, see Figure 5.7.

(iv) De Brabanter et al. (2018) bandwidth (henceforth denoted as h_2) using the bimodal kernel function $K_2(x) = 2\pi^{-1/2}x^2e^{-x^2}$, see Figure 5.7.

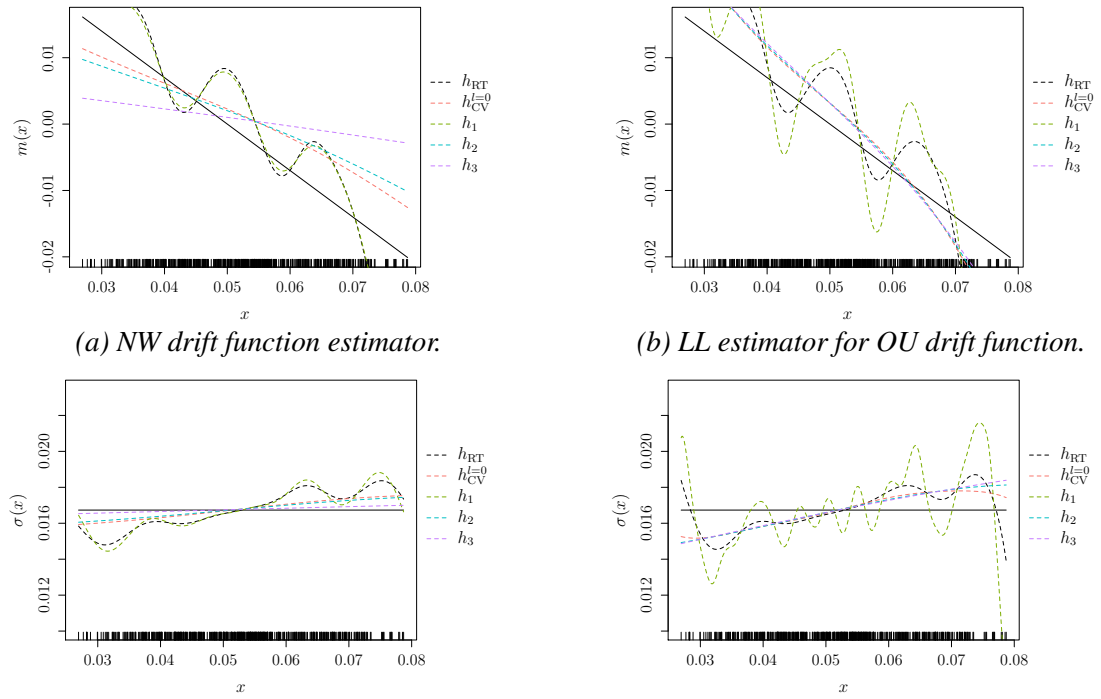
(v) De Brabanter et al. (2018) bandwidth (henceforth denoted as h_3) using the bimodal kernel $K_3(x) = \frac{1}{2}|x|\exp(-|x|)$, see Figure 5.7.

Tables 5.4–5.6 show the MISE and standard deviation (SD) of ISE for the Vasicek, CIR and nonlinear drift models, respectively, for a thousand simulations. Regarding the Vasicek model estimation in Table 5.4, analyzing the drift estimation, the rule-of-thumb bandwidth h_{RT} together with the bimodal kernel K_1 bandwidth h_1 show the highest MISE for all scenarios, while the best performance is achieved by the bimodal kernel bandwidths h_3 and h_2 and the cross-validation bandwidths, with similar results for $l = 0, 1, 2$. For the high persistence scenario, the bimodal kernel for h_2 and h_3 outperforms the cross-validation bandwidth,

Function	**ISE**	h_{RT}	$h_{\mathrm{CV}}^{l=0}$	$h_{\mathrm{CV}}^{l=1}$	$h_{\mathrm{CV}}^{l=2}$	h_1	h_2	h_3
				High persistence				
$n = 520$								
$m_{n,h}(\cdot;0)$	MISE	0.5579	0.0816	0.0814	0.0808	0.5704	0.0834	0.0266
	SD	0.7350	0.1836	0.1812	0.1760	0.7352	0.2495	0.1029
$\sigma_{n,h}(\cdot;0)$	MISE	0.0041	0.0018	0.0017	0.0018	0.0048	0.0014	0.0011
	SD	0.0069	0.0040	0.0036	0.0044	0.0074	0.0039	0.0031
$n = 2080$								
$m_{n,h}(\cdot;0)$	MISE	0.4938	0.0090	0.0090	0.0090	0.4028	0.0116	0.0312
	SD	0.6973	0.0094	0.0094	0.0094	0.5879	0.0107	0.0068
$\sigma_{n,h}(\cdot;0)$	MISE	0.0037	0.0013	0.0015	0.0014	0.0030	0.0008	0.0010
	SD	0.0066	0.0032	0.0035	0.0033	0.0054	0.0024	0.0034
				Low persistence				
$n = 520$								
$m_{n,h}(\cdot;0)$	MISE	1.4920	0.1670	0.1670	0.1670	1.4840	0.2171	0.5053
	SD	2.3030	0.1936	0.1936	0.1936	2.4576	0.2775	0.1133
$\sigma_{n,h}(\cdot;0)$	MISE	0.0160	0.0088	0.0071	0.0061	0.0173	0.0062	0.0071
	SD	0.0253	0.0204	0.0181	0.0154	0.0283	0.0179	0.0201
$n = 2080$								
$m_{n,h}(\cdot;0)$	MISE	1.3654	0.2832	0.2832	0.2832	0.7925	0.3280	0.5618
	SD	2.1480	0.0939	0.0939	0.0939	1.3825	0.0880	0.0483
$\sigma_{n,h}(\cdot;0)$	MISE	0.0151	0.0056	0.0045	0.0038	0.0126	0.0035	0.0033
	SD	0.0254	0.0138	0.0117	0.0093	0.0214	0.0128	0.0108

Table 5.4: MISE ($\times 100$) and standard deviation (SD) of ISE ($\times 100$) for Vasicek models, for high and low persistence and different bandwidths, for weekly data and $n \in \{520, 2080\}$, which corresponds to $T = 10$ and 40 years, respectively. Note that the close performance using some bandwidths may provide the same MISE or SD due to rounding.

however, for a low persistence, the cross-validation provides slightly smaller MISE. Similar conclusions can be reached for the diffusion function estimation, with h_{SR} and h_1 showing similar results and h_{CV}, h_2, h_3 performing closely, and the bimodal kernel h_2 and h_3 outperforming the cross-validation bandwidth. Figure 5.8 shows an example of the drift and diffusion function estimation for the Vasicek model, considering five bandwidths.



(a) NW drift function estimator. *(b) LL estimator for OU drift function.*

(c) NW diffusion function estimator. *(d) LL diffusion function estimator.*

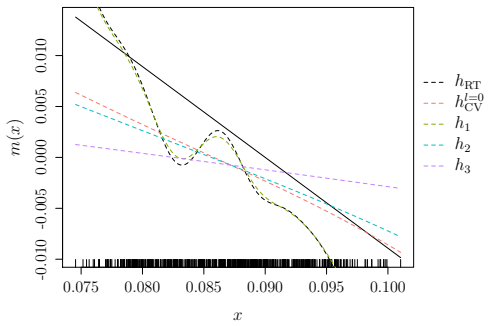
Figure 5.8: Vasicek drift and volatility functions estimates using Nadaraya-Watson (NW) and local linear (LL) estimators with different bandwidths (true functions in black solid line). Weekly sampling ($\Delta = 1/52$ and $n = 520$).

The CIR model simulation in Table 5.5 yields similar conclusions to the Vasicek model simulation: there is a poor performance of h_{SR} and h_1, the cross-validation bandwidths $h_{\mathrm{CV}}^{(l)}$ obtain almost exact results, and h_2 gives the smallest MISE, along with $h_{\mathrm{CV}}^{(l)}$. Figure 5.9 shows an example of the drift and diffusion function estimation for the CIR model, considering five bandwidths.

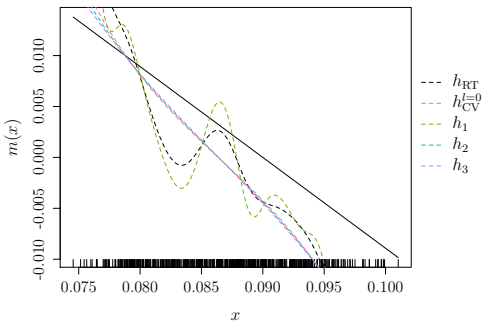
Lastly, Table 5.6 collects the results for the estimation of the nonlinear drift model. When estimating a nonlinear drift, the performance of the cross-validation bandwidth worsens, yielding the highest MISE, unlike the linear drift Vasicek and CIR models performance. However, the bandwidth h_2, which provided good results for linear drift, keeps outperforming when estimating a nonlinear drift. Figure 5.10 shows an example of the drift function estimation for the nonlinear drift model, considering five bandwidths.

Function	**ISE**	h_{RT}	$h_{\mathrm{CV}}^{l=0}$	$h_{\mathrm{CV}}^{l=1}$	$h_{\mathrm{CV}}^{l=2}$	h_1	h_2	h_3
$n = 520$								
$m_{n,h}(\cdot;0)$	MISE	4.7897	0.3300	0.3300	0.3300	4.6203	0.2642	0.4632
	SD	7.1512	0.7284	0.7284	0.7284	6.8331	0.6531	0.1647
$\sigma_{n,h}(\cdot;0)$	MISE	0.0480	0.0215	0.0214	0.0214	0.0491	0.0231	0.0256
	SD	0.0524	0.0158	0.0153	0.0153	0.0542	0.0238	0.0129
$n = 2080$								
$m_{n,h}(\cdot;0)$	MISE	4.8564	0.1530	0.1530	0.1530	3.2662	0.1925	0.5526
	SD	6.7786	0.1149	0.1149	0.1149	4.8881	0.1150	0.0779
$\sigma_{n,h}(\cdot;0)$	MISE	0.0427	0.0158	0.0158	0.0158	0.0310	0.0137	0.0190
	SD	0.0528	0.0062	0.0062	0.0062	0.0394	0.0077	0.0062

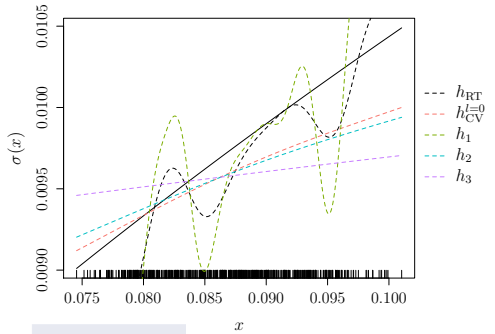
Table 5.5: MISE ($\times 100$) and standard deviation (SD) of ISE ($\times 100$) for CIR model with different bandwidths, for weekly data and $n \in \{520, 2080\}$, which corresponds to $T = 10$ and 40 years, respectively.



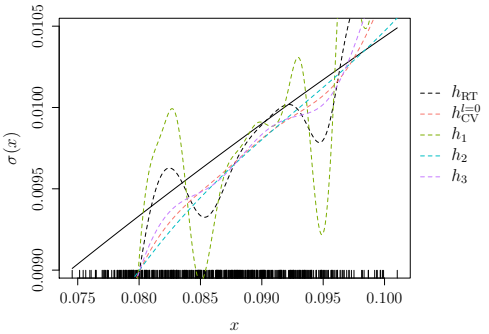
(a) NW drift function estimator.



(b) LL drift function estimator.

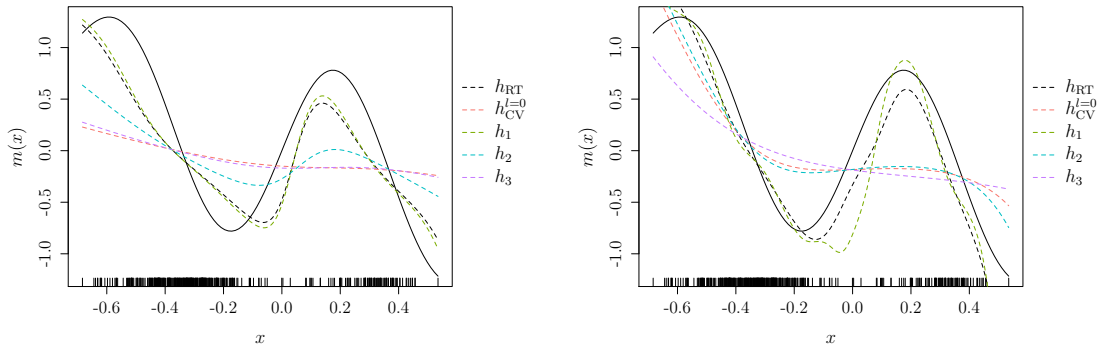


(c) NW diffusion function estimator.



(d) LL diffusion function estimator.

Figure 5.9: CIR drift and volatility functions estimates using Nadaraya-Watson (NW) and local linear (LL) estimators with different bandwidths (true functions in black solid line). Weekly sampling ($\Delta = 1/52$ and $n = 520$).



(a) NW drift function estimator. *(b) LL drift function estimator.*

Figure 5.10: Estimates of a nonlinear drift function diffusion model, with $m(x) = -x + \sin(8x)\,\exp(-x^2)$ *and* $\sigma(x) = 0.4$*, using Nadaraya-Watson (NW) and local linear (LL) estimators with different bandwidths.*

Function	**ISE**	h_{RT}	$h_{\mathrm{CV}}^{l=0}$	$h_{\mathrm{CV}}^{l=1}$	$h_{\mathrm{CV}}^{l=2}$	h_1	h_2	h_3
$m_{n,h}(\cdot;0)$	MISE	0.6693	0.9190	0.9219	0.9233	0.6328	0.6285	0.8293
	SD	0.4193	0.0768	0.0755	0.0746	0.2949	0.1432	0.1497
$\sigma_{n,h}(\cdot;0)$	MISE	0.0181	0.0087	0.0084	0.0083	0.0209	0.0149	0.0098
	SD	0.0219	0.0093	0.0089	0.0088	0.0227	0.0169	0.0108

Table 5.6: MISE ($\times 100$*) and standard deviation (SD) of ISE (*$\times 100$*) for a diffusion model with nonlinear drift,* $m(x) = -x + \sin(8x)\exp(-x^2)$ *and* $\sigma(x) = 0.4$*.*

5.5 Nonparametric estimation of the short-term interest rate

To illustrate the nonparametric methods reviewed in this chapter, we consider short-term interest rate data. We use daily ($\Delta = 1/252$) market yields on U.S. Treasury securities with maturities ranging from 1 month to 30 years (see Figure 5.11), from January 2, 2016 to December 31, 2019, which consist in $n = 999$ observations. We estimate the marginal density, drift and diffusion functions for the eleven maturities.

5.5.1 Marginal density estimation

Using the kernel density estimator $\pi_{n,h}(\cdot)$ introduced in (5.6), we estimate the marginal density for the treasury data. Figure 5.12 shows the estimated density for all maturities along a

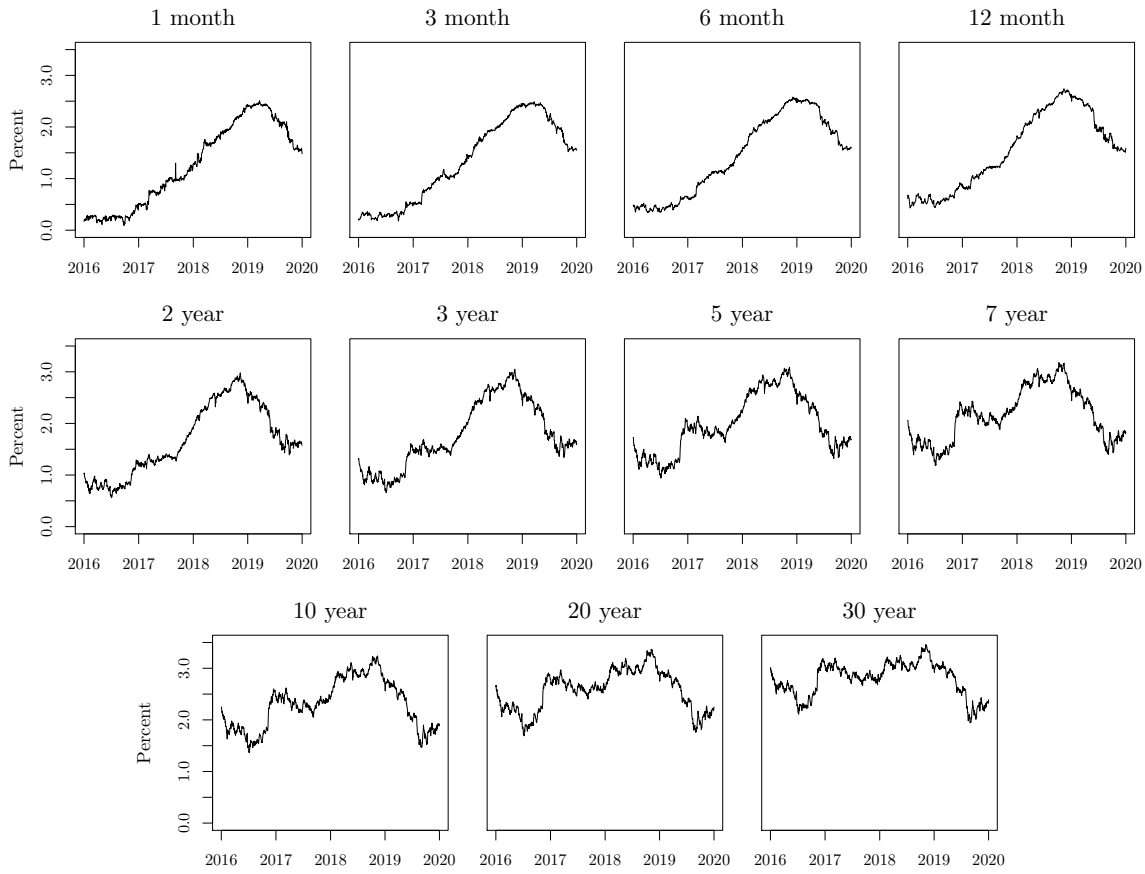


Figure 5.11: U.S. daily Treasury securities rate time series.

pointwise confidence interval. Confidence bands for the stationary density are build by plugging $\pi_{n,h}(x)$ as an estimate of $\pi(x)$ in the variance in (5.7), so that an asymptotic 95% pointwise confidence interval for $\mathbb{E}\left[\pi_{n,h}(x)\right]$ is given by $\left(\pi_{n,h}(x) \pm 1.96\sqrt{R(K)\pi_{n,h}(x)}/\sqrt{nh}\,\right)$. The dashed line of Figure 5.12 correspond to the parametric estimation $\pi_{\hat{\theta}}(\cdot)$ of the density introduced in (5.8), considering the parametric diffusion model,

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

which corresponds to the specification introduced in Chan et al. (1992), known as the CKLS model, commonly use in the econometric literature. The Gaussian kernel and Scott's rule bandwidth were used for the kernel density estimation. For shorter maturities, the parametric and nonparametric estimates differ, with the kernel estimate producing a bimodal distribution. However, larger maturities show close density function estimates for the parametric CKLS model and nonparametric estimation.

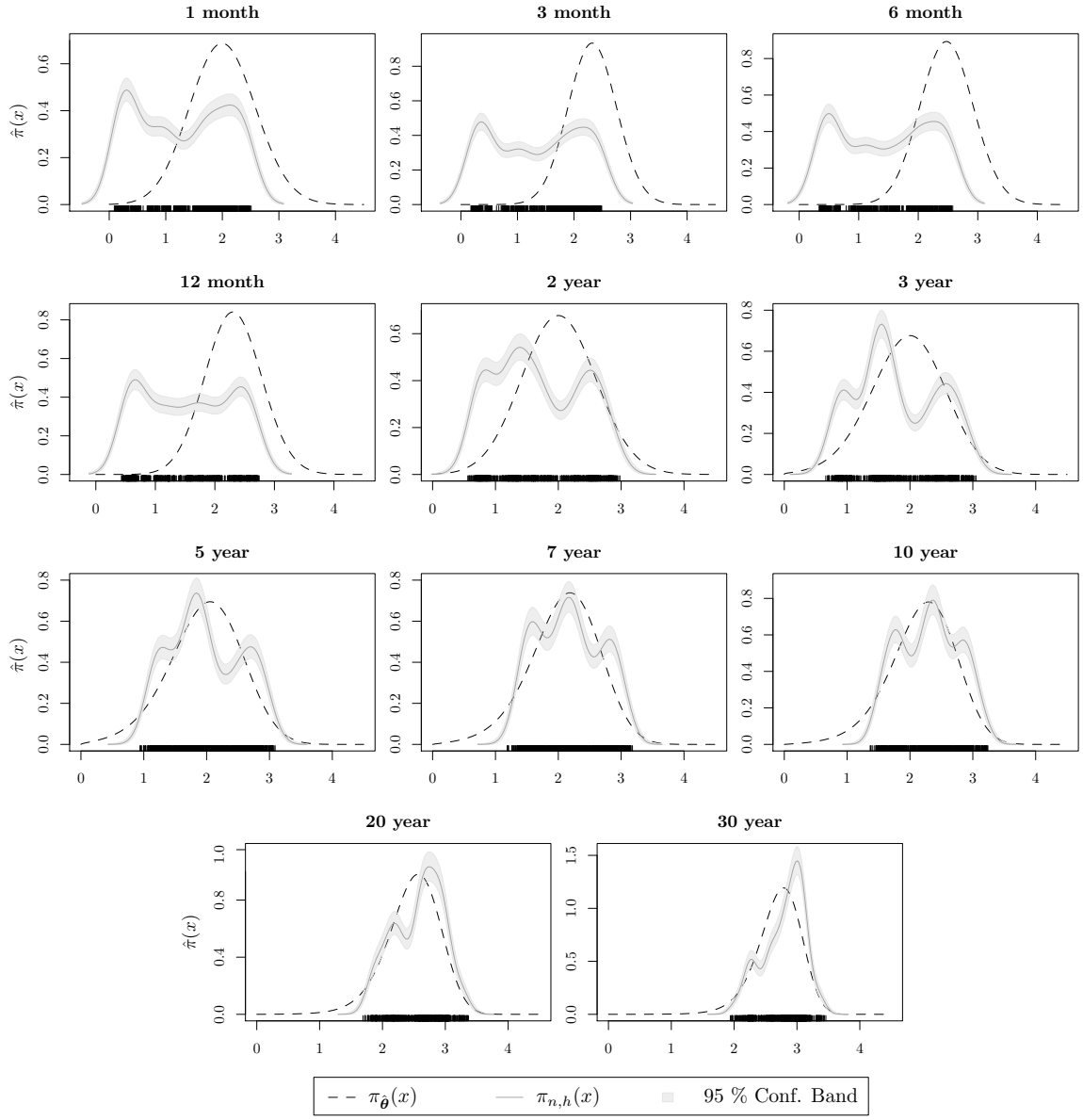


Figure 5.12: Parametric (dashed line) and non parametric (gray solid line) estimation of the marginal density function for the U.S. daily Treasury securities. Shaded area indicates a 95% pointwise confidence interval for the kernel estimation.

5.5.2 Drift function estimation

Regarding the drift function estimation, we use the Nadaraya-Watson estimator introduced in (5.11), using a Gaussian kernel and a leave-one-out cross-validation bandwidth. Figure 5.13 shows the nonparametric estimates for all maturities and the parametric estimation of the drift function for the CKLS model, that is, $m_{\hat{\boldsymbol{\theta}}}(x) = \hat{\kappa}(\hat{\mu} - x)$. The confidence band for the nonparametric estimation was computed using the asymptotic distribution in (5.14), with $p = 0$. The nonparametric estimates of the drift are very close to the parametric estimation of the CKLS drift function, yielding almost linear estimations for the drift.

5.5.3 Diffusion function estimation

Lastly, we provide the nonparametric estimation $\sigma_{n,h}(\cdot)$ of the diffusion function, introduced in (5.15), with $p = 0$, using a Gaussian kernel and a leave-one-out cross-validation bandwidth. Confidence bands for the diffusion function are constructed using the asymptotic distribution in (5.16), plugging the kernel density estimator $\pi_{n,h}(x)$ and estimating $V(x)$ by

$$\hat{V}(x) = \sum_{j=1}^{n-1} W_{nj}(x)\Delta^{-1}\big(X_{t_{i+1}} - X_{t_i}\big)^4 - \big(\sigma^2_{n,h}(x)\big)^2.$$

Figure 5.14 shows the nonparametric estimation of the diffusion function (solid gray line) for the eleven maturities, along a 95% pointwise confidence band (shaded area) and the parametric estimation of the CKLS diffusion function, that is, $\sigma_{\hat{\boldsymbol{\theta}}}(x) = \hat{\sigma}x^{\hat{\gamma}}$. The kernel estimates for short maturities are nonlinear, however, longer maturities show a close behavior of the parametric and nonparametric estimates.

5.6 Conclusions

In this chapter we reviewed nonparametric estimation techniques for diffusion models, and a comparison of different bandwidths was carried out for the marginal density, drift and diffusion functions estimation. The common used bandwidth in the econometric field is obtained using the rule-of-thumb, which showed a similar performance to the cross-validation bandwidth when estimating the marginal density function. However, this bandwidth was outperformed in the drift and diffusion functions estimation. Using cross-validation bandwidths or using a pilot bandwidth with a bimodal kernel function satisfying $K(0) = 0$ yielded lower MISE. The level of persistence hinders the estimation of the components, which may indicate that taking the mean reversion speed into account when computing the bandwidth could

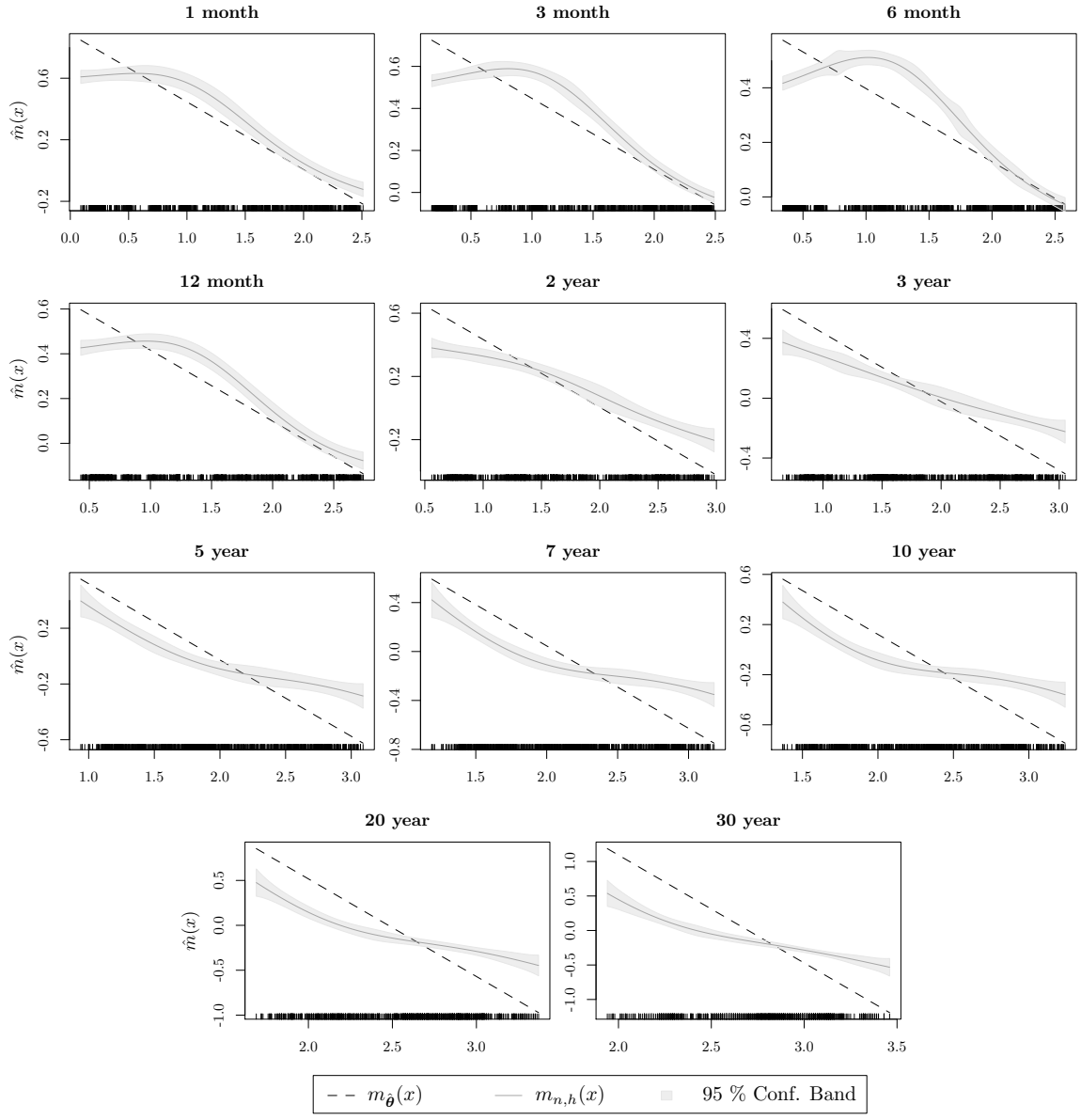


Figure 5.13: Parametric (dashed line) and non parametric (gray solid line) estimation of the drift function for the U.S. daily Treasury securities. Shaded area indicates a 95% pointwise confidence interval for the kernel estimation.

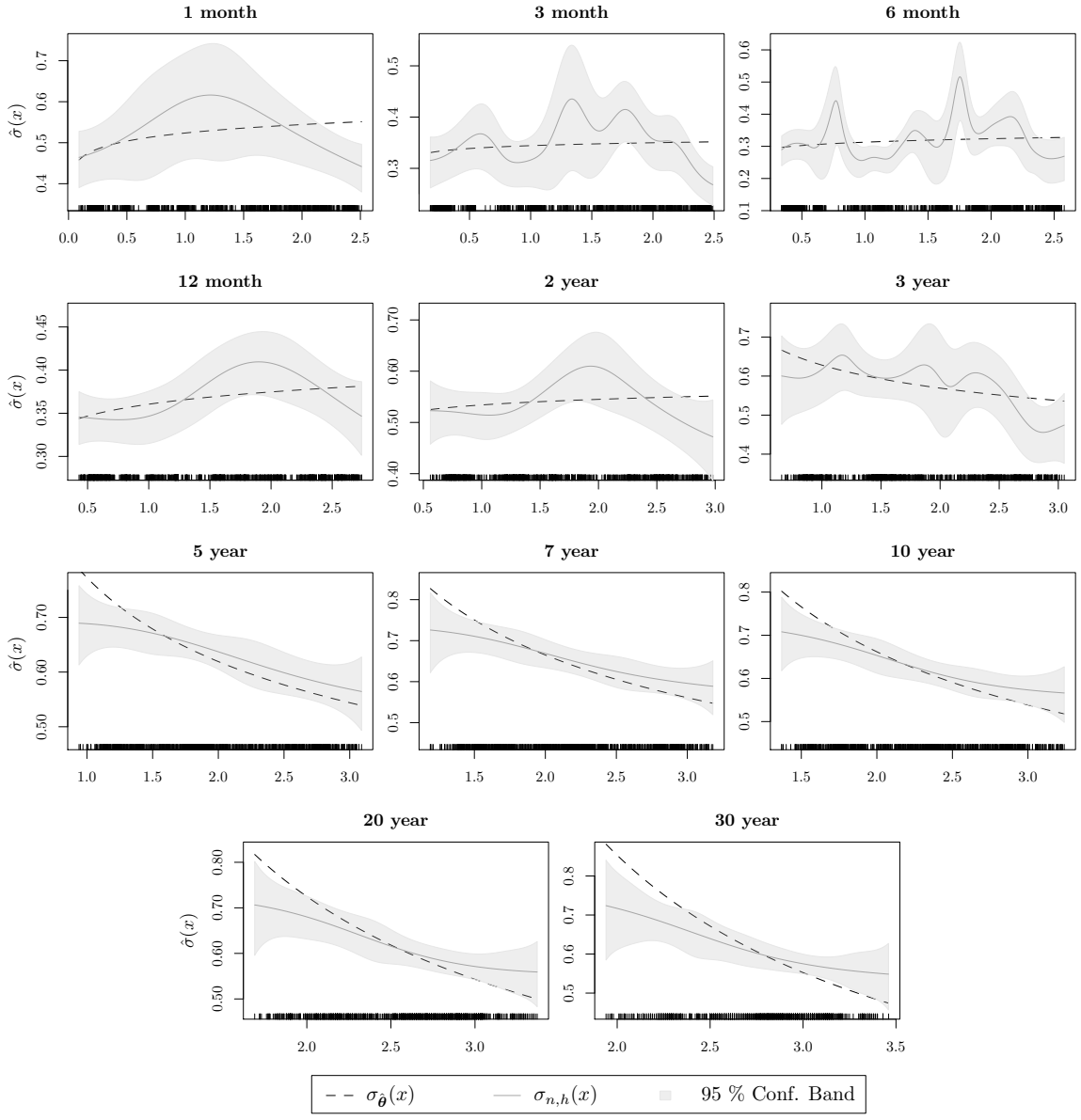


Figure 5.14: Parametric (dashed line) and non parametric (gray solid line) estimation of the diffusion function for the U.S. daily Treasury securities. Shaded area indicates a 95% pointwise confidence interval for the kernel estimation.

provide a bandwidth selection criteria adapted for diffusion models. This is an open research line and, though this problem has been addressed in the literature, not many proposals have been studied. The nonparametric estimation of the short-term interest rate for U.S. Treasury securities showed a close performance to the parametric estimation of the CKLS model for large maturities. The nonparametric estimation of the drift function was almost linear in all maturities and, though empirical evidence has commonly discussed the nonlinearity of the drift, the short period of time considered (four years) might explain the linear behavior.

Chapter 6

Goodness-of-fit test for diffusion models

Diffusion models play an essential role in modeling continuous-time stochastic processes in the financial field. Therefore, several proposals have been developed in the last decades to test the specification of stochastic differential equations. We provide a survey to collect some developments on goodness-of-fit tests for diffusion models and implement these methods to illustrate their finite sample behavior, regarding size and power, by means of a simulation study. We also apply the ideas of distance covariance for testing independence to propose a test for the parametric specification of diffusion models, comparing its performance with the other methods and analyzing the effect of the curse of dimensionality. Short-term interest rate time series are used as real data examples, in particular the treasury securities with different maturities analyzed in Chapter 5 are considered.

The contents of this chapter are collected in López-Pérez et al. (2022a).

Contents

6.1 Introduction

Financial theory has been developing on a continuous-time basis, where the evolution of assets are usually represented in terms of continuous-time stochastic processes, rooted in the foundational works of Black and Scholes (1973) and Merton (1975). With market prices evolving at very short-time scales, continuous-time models have proved to be more appropriate to capture the quick changes observed on real markets than discrete-time models. In this continuous-time paradigm, modeling interest rates represents one of the main topics in the recent financial economics literature, as it lies at the core of the most basic valuations problems in finance. The uncertainty attached to interest rates evolution is a relevant part of the theory of financial decision making, so that understanding the factors that drive interest rates is important for modeling the yield curve, analyzing financial instruments (futures, options on zero coupon bonds, swaps contracts, etc.) or developing hedging interest rate risk strategies. As discussed in Chapters 2 and 3, one of the most common representation is the time-homogeneous diffusion given by a stochastic differential equation driven by a Wiener process $\{W_t\}_{t\geqslant 0}$,

$$\mathrm{d}X_t = m(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{6.1}$$

defined on a complete probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}\big)$, the process $X_t \in \mathbb{R}$ evolves over the interval $[0, T]$ in continuous time, according to the drift $m(\cdot)$ and diffusion or volatility $\sigma(\cdot)$ functions.

The correct specification of the dynamics of short rates is fundamental to much of empirical and theoretical finance, as mis-specification could lead to pricing errors, mishedged or unhedged risks, or incorrect inferences. Over the last four decades, several developments of short-term interest rate models have emerged. Chan et al. (1992) developed a general framework to nest different single-factor models of the short-term interest rate, including those by Merton (1973), Cox (1975), Cox and Ross (1976), Vasicek (1977), Dothan (1978), Brennan and Schwartz (1980), Rendleman and Bartter (1980), Cox et al. (1980, 1985), or Constantinides and Ingersoll (1982). Given the profusion of diffusion-based models, the problem of choosing a family of models to explain the dynamics of interest rates arises naturally. Yet

there is no consensus on the choice of the parametric specification of the short rate process. With model diagnosis playing an essential role to determine which model best captures the short rates dynamics, there exist several proposals for continuous-time model specification with works dealing with goodness-of-fit test in different directions. Aït-Sahalia (1996c) proposed one of the first nonparametric test for univariate diffusion processes by comparing the model-implied parametric stationary density with a kernel density estimator based on discretely sampled observations, as the drift and diffusion functions completely characterize the stationary and marginal density. Years later, Gao and King (2004) developed a procedure to improve its finite sample performance. The comparison between parametric and nonparametric estimates of defining attributes of diffusion models has motivated different proposals, focusing on the marginal density (Gao and King, 2004) and transitional density function (Hong and Li, 2005; Chen et al., 2008), while Corradi and Swanson (2005) compared cumulative distribution functions.

A body of literature is focused in testing the drift or volatility functions of the diffusion processes, as an alternative to test the characterization of the whole dynamics of the process. With respect to the drift function, Lee (2006) adapted the Bickel and Rosenblatt (1973) test for diffusion models, Arapis and Gao (2006) proposal used nonparametric estimation procedures, Negri and Nishiyama (2009, 2010) test is based on the score marked empirical process, Kutoyants (2010) studied direct analogues of the classical Anderson-Darling and Kolmogorov-Smirnov tests and Masuda et al. (2011) used an innovation martingale approach. Regarding the diffusion function, Corradi and White (1999) proposed a homoscedasticity test; Dette and und Wilkau (2003), Dette and Podolskij (2008) and Podolskij and Ziggel (2008) tests are based on stochastic processes of the integrated volatility; Li (2007) and Dette et al. (2006) rely on the L^2-distance between the diffusion function under the null and the alternative hypotheses; Zheng (2009) and Chen et al. (2019) developed nonparametric tests.

Lastly, the specification of both drift and volatility functions is analyzed in Fan and Zhang (2003), Fan et al. (2003) and Aït-Sahalia et al. (2009), based on generalized likelihood ratio test ideas (Fan et al., 2001); Gao and Casas (2008) considered smoothing techniques for semiparametric diffusion models; Chen and Hong (2010) used conditional characteristic function characterization to construct a specification test, while Kristensen (2011) relied on the transition density; Song (2011) test is based on the infinitesimal operator; Chen and Gao (2011) used an empirical likelihood statistic; and Monsalve-Cobis et al. (2011) and Chen et al. (2015) tests are based on empirical regression processes.

In the present chapter, we provide a survey to collect these developments in goodness-of-fit test for diffusion models. Several of these proposals are extensions of earlier works in

the context of continuous-time series, and consequently their drawbacks are also inherited. Due to the curse of dimensionality most of the existing tests cannot be easily extended to multivariate models. For tests based on nonparametric methods, bandwidth selection is a concern and power vitally depends on the dimension of the conditioning variables, decreasing with dimension. Though different adaptations have been addressed in order to avoid the curse of dimensionality, few have been studied for diffusion processes. As a sideways contribution, we apply distance correlation (Székely et al., 2007) ideas to test the specification of diffusion models, which circumvents the curse of dimensionality.

In a seminal series of works, Székely et al. (2007) and Székely and Rizzo (2009, 2012, 2013, 2013) introduced the concepts of distance covariance and distance correlation as measures of the degree of dependence among vectors of random variables and used it to test independence. Since then, several works considered distance correlation to construct test statistics. Zhou (2012) extended the concept to auto-distance correlation in time series to explore nonlinear dependence. Davis et al. (2018) applied distance correlation ideas to stationary univariate and multivariate time series in order to measure lagged auto- and cross-dependencies. Matsui et al. (2017) extended distance covariance notions to continuous-time stochastic processes defined on the same interval, while Dehling et al. (2020) built a measure of independence for the components of i.i.d. pairs of stochastic processes on the unit interval and established consistency of the bootstrap procedure for the sample distance covariance/correlation of two processes. Kroll (2021) derived the asymptotic distribution of sample distance covariance under mixing conditions.

The chapter is organized as follows. Section 6.2 provides a survey of goodness-of-fit test proposals for diffusion models and Section 6.3 concerns the extension of distance covariance ideas to test the specification of diffusion processes. The implementation of some of the methods and the study of the level and power of the tests for finite samples are discussed in Section 6.4 by means of Monte Carlo simulations and real data are analyzed in Section 6.5. Concluding remarks are provided in Section 6.6.

6.2 Goodness-of-fit tests for diffusion models

In this section, we discuss the goodness-of-fit testing problem of diffusion models and introduce different methodologies developed in order to test the specification of the model, many of which are extensions of classical tests ideas from earlier works. We begin by introducing

the parametric version (see Chapter 3) of the SDE in (6.1),

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t. \tag{6.2}$$

Under a parametric framework, $\boldsymbol{\theta}$ is an unknown parameter vector such that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ with d a positive integer and Θ a compact parametric space, and $m(\cdot, \boldsymbol{\theta})\colon \mathbb{R} \times \Theta \to \mathbb{R}$ and $\sigma(\cdot, \boldsymbol{\theta})\colon \mathbb{R} \times \Theta \to (0, \infty)$. A unique strong solution $\{X_t\}_{t\in[0,T]}$ of the stochastic differential equation, with initial condition $X_0 = x_0 \in \mathbb{R}$ adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$,

$$X_t = X_0 + \int_0^T m(X_u, \boldsymbol{\theta})\,\mathrm{d}u + \int_0^T \sigma(X_u, \boldsymbol{\theta})\,\mathrm{d}W_u,$$

exists under Assumptions 1–3 (see Chapter 2). A discrete version of (6.2) can be obtained using the Euler-Maruyama method (see Chapter 3),

$$X_{t_{i+1}} - X_{t_i} \approx m(X_{t_i}, \boldsymbol{\theta})(t_{i+1} - t_i) + \sigma(X_{t_i}, \boldsymbol{\theta})(W_{t_{i+1}} - W_{t_i}), \tag{6.3}$$

with $i = 0, 1, \ldots, n-1$, $X_{t_0} = x_0 \in \mathbb{R}$ and $t_i = i\Delta$, where Δ is the length of the sampling interval, that is $\Delta = T/n = (t_{i+1} - t_i)$.

6.2.1 Test based on the estimation of the distribution and density functions

Many classical goodness-of-fit (GoF) tests proposals are based on the comparison of functions that characterize the probability law given in the null hypothesis and an estimator of this function, where the characterization can take place with the distribution (F), density (f), quantile (F^{-1}) or characteristic function (φ). The GoF methodology is rooted in the comparison of a nonparametric pilot estimator of the distribution or density function of the i.i.d. observations of a random variable X with a consistent estimator of the target function. That is, to test the null hypothesis for the distribution function,

$$\begin{cases} \mathcal{H}_0\colon F \in \mathcal{F}_{\text{dist}} = \{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}, \\ \mathcal{H}_{\text{a}}\colon F \notin \mathcal{F}_{\text{dist}}, \end{cases} \tag{6.4}$$

or the density function

$$\begin{cases} \mathcal{H}_0\colon f \in \mathcal{F}_{\text{den}} = \{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}, \\ \mathcal{H}_{\text{a}}\colon f \notin \mathcal{F}_{\text{den}}, \end{cases}$$

where the pilot estimator is given by the empirical cumulative distribution function (F_n) and the kernel density estimator ($f_{n,h}$), respectively. The construction of the test statistic T_n is based on some distance between the pilot estimator (F_n, $f_{n,h}$) and the estimator under the null hypothesis ($F_{\hat{\boldsymbol{\theta}}}$, $f_{\hat{\boldsymbol{\theta}}}$) of the true function ($F_0$, f_0). For the distribution function, we have the test statistic $T_n = D(F_n, F_{\hat{\boldsymbol{\theta}}})$ and the diversity of tests comes from the diversity of distances D. If we take

$$T_n = n \int_{-\infty}^{\infty} H(x) \left(F_n(x) - F_{\hat{\boldsymbol{\theta}}}(x)\right)^2 \mathrm{d}F_{\hat{\boldsymbol{\theta}}}(x),$$

we have the Cramér-von Mises family of statistics. Taking $H(x) \equiv 1$ we obtain the Cramér-von Mises test, while the Anderson-Darling test is obtained with the weight function $H(x) = \left(F_{\hat{\boldsymbol{\theta}}}(x)\left(1 - F_{\hat{\boldsymbol{\theta}}}\right)\right)^{-1}$. Considering an uniform metric,

$$T_n = \sup_x \sqrt{n} \left|F_n(x) - F_{\hat{\boldsymbol{\theta}}}(x)\right|,$$

we have the Kolmogorov-Smirnov test. Direct analogues of these test can be found for diffusion model, as the drift $m(\cdot)$ and diffusion $\sigma^2(\cdot)$ terms completely characterize the stationary (conditional) or marginal (unconditional) density of the process. To test the specification of the drift and/or volatility function, given a specific parametrization, we have a join parametric family

$$\mathcal{P}_{\boldsymbol{\theta}} \equiv \left\{ \left(m(\cdot, \boldsymbol{\theta}), \sigma^2(\cdot, \boldsymbol{\theta})\right) / \boldsymbol{\theta} \in \Theta \right\},$$

with Θ a compact subset of $\mathbb{R}^d$. The test addresses if the values of the parameters in Θ are a representation of the true process, that is, whether the true functions $\mu_0(\cdot)$ and $\sigma_0^2(\cdot)$ belong to the parametric space $\mathcal{P}_{\boldsymbol{\theta}}$. Regarding the unique characterization of the marginal and transitional densities, any parametrization of $\mathcal{P}_{\boldsymbol{\theta}}$ of $m(\cdot)$ and $\sigma^2(\cdot)$ corresponds to a parametrization of the density,

$$\Pi_{\boldsymbol{\theta}} \equiv \left\{ \left(\pi(\cdot, \boldsymbol{\theta}), p(\cdot, \cdot, \mid \cdot, \cdot; \boldsymbol{\theta})\right) / \left(m(\cdot, \boldsymbol{\theta}), \sigma^2(\cdot, \boldsymbol{\theta})\right) \in \mathcal{P}_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta \right\},$$

where $\pi(x, \boldsymbol{\theta})$ is the marginal density at x and $p(s, y \mid t, x; \boldsymbol{\theta})$ is the transitional density from x at time t to y at time s. Therefore, to test the model specification, the mapping between the drift and volatility functions and the transitional and marginal densities can be used. The marginal density corresponding to the drift and diffusion functions pair $\left(m(\cdot, \boldsymbol{\theta}), \sigma^2(\cdot, \boldsymbol{\theta})\right)$ has a close form expression via Kolmogorov forward equation (see (5.5) in Chapter 5) and is given by

$$\pi(x, \boldsymbol{\theta}) = \frac{\xi(\boldsymbol{\theta})}{\sigma^2(x, \boldsymbol{\theta})} \exp\left(\int_{x_0}^{x} \frac{2m(u, \boldsymbol{\theta})}{\sigma^2(u, \boldsymbol{\theta})} \mathrm{d}u\right),$$

where $\xi(\boldsymbol{\theta})$ is a normalization constant, determined to insure that the density function integrates to one. However, the transition density as defined by the Kolmogorov backward equation may not admit a closed form expression (see Aït-Sahalia, 1999, 2002, for Edgeworth type approximations to overcome this problem). The kernel estimator for the marginal density (see Section 5.3 in Chapter 5) is

$$\pi_{n,h}(x) = n^{-1} \sum_{i=0}^{n} \frac{1}{h} K\left(\frac{x - X_{t_i}}{h}\right),$$

with $K(\cdot)$ the kernel density function and h the bandwidth or smoothing parameter (see Rosenblatt, 1956 and Parzen, 1962). Therefore, the hypotheses in (6.4) can be written as the null and alternative hypotheses

$$\begin{cases} \mathcal{H}_0 \colon \exists \boldsymbol{\theta}_0 \in \Theta \,/\, \pi(\cdot, \boldsymbol{\theta}_0) = \pi_0(\cdot), \\ \mathcal{H}_{\mathrm{a}} \colon \pi_0(\cdot) \notin \Pi_{\boldsymbol{\theta}}, \end{cases}$$

with $\pi_0(\cdot)$ the true marginal density of the process.

When analyzing continuous time models, we can choose to use all the continuous sample path (Kutoyants, 2010) or use a sampling scheme, as in (6.3). Regarding the later, Aït-Sahalia (1996c) and Gao and King (2004) proposed a test to check the specification of both drift and diffusion functions by comparing the parametric stationary density with the same density estimated nonparametrically, while the Bickel and Rosenblatt (1973) test—based on the integrated squared error of the true and kernel density estimate—was extended by Lee (2006) to test the specification of the drift function. Under certain regularity conditions, the test statistics are asymptotically normal. However, in small samples, critical values computed based on the asymptotic distribution could lead to over reject the null hypothesis (see the discussion in Section 6.2.5). Furthermore, the stationary marginal density does not characterize the full process, so tests based on it would not detect alternatives with the same marginal density as the null model. Thus, other authors, such as Hong and Li (2005) and Aït-Sahalia et al. (2009), used the transition density, while the cumulative distribution function was considered in Corradi and Swanson (2005). The former authors established the asymptotic null distributions of the proposed test statistics and the later used a block bootstrap procedure to calibrate the test distribution. The conditional characteristic function can also capture the full dynamic of the process, as the Fourier transform of the transitional density. This is addressed in Chen and Hong (2010), where the authors constructed a nonparametric specification test for continuous-time Markov models using the conditional characteristic function. The test statis-

tics proposed are asymptotically normal but the authors noted that using parametric bootstrap improved the finite-sample performance, as also stated in Aït-Sahalia et al. (2009).

6.2.2 Test based on the estimation of the regression function

The GoF ideas from the later section can be extended to regression models. We can write the diffusion model in (6.3) as

$$Y_{t_i} = m(X_{t_i}) + \sigma(X_{t_i})\Delta^{-1/2}\varepsilon_{t_i}, \tag{6.5}$$

where $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ and $\sqrt{\Delta}\varepsilon_{t_i} = (W_{t_{i+1}} - W_{t_i})$, therefore $\varepsilon_{t_i} \sim \mathcal{N}(0,1)$ for $i = 0, \ldots, n-1$, due to the independent and Gaussian increments of the Wiener process. The goal is to test the specification of the drift

$$\mathcal{H}_0 \colon m \in \{m(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\},$$

and the diffusion function,

$$\mathcal{H}_0 \colon \sigma \in \{\sigma(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}.$$

In the classical regression context, the target functions are usually the regression function of Y_{t_i} over X_{t_i}, that is $m(x) = \mathbb{E}\left[Y_{t_i} \mid X_{t_i} = x\right]$, and the conditional variance $\sigma^2(x) = \mathbb{V}\mathrm{ar}\left[Y_{t_i} \mid X_{t_i} = x\right]$.

Smoothing-based test

The drift and diffusion terms can be estimated nonparametrically. Though different smoothing methods have been proposed, a large body of literature is focused in kernel type estimators (see Nadaraya, 1964 and Watson, 1964). As introduced in Chapter 5, a kernel estimator for the drift and diffusion function (see Florens-Zmirou, 1993) is given by

$$m_{n,h}(x) = \sum_{i=0}^{n-1} W_{ni}(x)(X_{t_{i+1}} - X_{t_i}),$$
$$\sigma^2_{n,h}(x) = \sum_{i=0}^{n-1} W_{ni}(x)(X_{t_{i+1}} - X_{t_i})^2,$$

with

$$W_{ni}(x) = \frac{K\left(\frac{x - X_{t_i}}{h}\right)}{\Delta \sum_{j=0}^{n-1} K\left(\frac{x - X_{t_j}}{h}\right)},$$

where $\Delta \to 0$ is required to obtain consistency and asymptotic normality. Florens-Zmirou (1993) considered a uniform kernel K, later Jiang and Knight (1997) extended the results to the Gaussian kernel and Bandi and Phillips (2003) to general stationary process and kernels. The ideas of the previous section can be extended to the regression context, where an empirical process for the regression problem can be expressed as

$$\begin{aligned} R_{1n}(x) &= \sqrt{nh}\big(m_{n,h}(x) - \mathbb{E}_{\hat{\boldsymbol{\theta}}}\left[m_{n,h}(x)\right]\big) \\ &= \sqrt{nh}\sum_{i=0}^{n-1} W_{ni}(x)\big(Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i})\big), \end{aligned} \tag{6.6}$$

for the drift function (that be interpreted as a smoothing over the residuals $\hat{\eta}_{t_i} = Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i})$), which can easily be extended to the diffusion function. To implement the test, a root-n consistent estimator $\hat{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta}_0$ is needed, that is $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$, providing $\mathbb{E}_{\hat{\boldsymbol{\theta}}}$ an estimate of $\mathbb{E}_{\boldsymbol{\theta}_0}$ (see López-Pérez et al., 2021, for a review about parametric estimation methods for diffusion models). Therefore, a general test based on the empirical process R_{1n} can be constructed by considering a continuous functional,

$$T_n = \int R_{1n}^2(x)\omega(x)\,\mathrm{d}x.$$

For example, denoting $\eta_{t_i 0} = Y_{t_i} - m_{\boldsymbol{\theta}_0}(X_{t_i})$, under the null hypothesis, since $\mathbb{E}\left[\eta_{t_i 0} \mid X_{t_i}\right] = 0$, we have

$$\mathbb{E}\Big[\eta_{t_i 0}\,\mathbb{E}\left[\eta_{t_i 0} \mid X_{t_i}\right]\pi(X_{t_i})\Big] = 0, \tag{6.7}$$

and using a kernel-based sample analogue of (6.7),

$$T_{1n} = \frac{1}{n(n-1)}\sum_{i=0}^{n-1}\sum_{\substack{j=0 \\ j\neq i}}^{n-1} h^{-1}K\left(\frac{X_{t_i} - X_{t_j}}{h}\right)\hat{\eta}_{t_i}\hat{\eta}_{t_j},$$

with $\hat{\eta}_{t_i} = (Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i}))$, we obtain the test statistic introduced in Zheng (1996). Note that in order to apply U-statistic theory, the negligible terms with $i = j$ are dropped. This idea was extended to diffusion models by Gao and Casas (2008), where a nonparametric test for the specification of both drift and diffusion functions was developed for a discretized semiparametric diffusion model. Arapis and Gao (2006) proposed a nonparametric test for the drift coefficient, while Li (2007) and Chen et al. (2019) proposals focused on the diffusion function, and Zheng (2009) developed a nonparametric test for heteroscedasticity build on the work of Zheng (1996). Regarding the test statistic distribution, asymptotic normality is the limiting distribution of the tests under $\mathcal{H}_0$. However, the critical values are usually obtained

using simulation or resampling methods, as with finite samples the convergence is slow, a problem that arises in virtually all smoothing-based tests.

Empirical regression process-based test

Another regression-based approach is inspired by GoF test for distribution, were the selection of a smoothness parameter is avoided. This alternative methodology is based on the empirical estimator of the integrated regression function (Stute, 1997), $I(x) = \mathbb{E}\left[Y \mathbb{1}_{\{X \leqslant x\}}\right] = \int_{-\infty}^{x} m(y)\, \mathrm{d}F_X(y)$, where F_X is the marginal distribution function of X. An empirical estimator of the integrated regression is given by $I_n(x) = n^{-1} \sum_{i=1}^{n} Y_{t_i} \mathbb{1}_{\{X_{t_i} \leqslant x\}}$ and comparing the estimator $I_n(x)$ with its estimated values under the null hypothesis, a new empirical process can be constructed,

$$R_{2n}(x) = \sqrt{n}\big(I_n(x) - \mathbb{E}_{\hat{\boldsymbol{\theta}}}\left[I_n(x)\right]\big) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \big(Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i})\big) \mathbb{1}_{\{X_{t_i} \leqslant x\}}. \tag{6.8}$$

The process R_{2n} is a marked empirical process, where the marks are given by the residuals $(Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i}))$. The test statistics $T_n = \Psi(R_{2n})$ can be constructed based on a distance of the resulting residual marked empirical process R_{2n} in (6.8) from the expected zero mean, measured by Cramér-von Mises and Kolmogorov-Smirnov functionals $\Psi(\cdot)$. Monsalve-Cobis et al. (2011) proposed a test for the drift and diffusion functions based on a marked empirical process of the residuals, constructed using estimators of the integrated regression function and the integrated conditional variance function, respectively. The distribution of the processes are approximated using bootstrap techniques, as tests based on an appropriately scaled discrepancy of the empirical process R_{2n} in (6.8) are not generally asymptotically distribution free. In Koul and Stute (1999) a transformation of R_{2n} is considered in order to construct a process whose limiting distribution is known, so that tests based on it are asymptotically distribution free. To achieve this, a martingale transformation of the underlying process (Khmaladze, 1981) is considered, and this idea was extended to diffusion models by several authors. An approach based on the innovation martingale to test the drift function was introduced in Negri and Nishiyama (2009, 2010) on the basis of the continuous observation and, subsequently, in Masuda et al. (2011) the method was extended to the case of discrete-time observations, obtaining an asymptotically distribution free test, as they proved that the limit distribution is the supremum of the standard Brownian motion. By means of the Khmaladze (1981) martingale transformation of the marked empirical process, Chen et al. (2015) extends the proposal of Monsalve-Cobis et al. (2011) to obtain a limiting distribution-free test for the parametric diffusion function.

Another methodology that avoids the use of smoothing techniques, also based on cumulative sum processes, relies on the integrated variance function to test the parametric form of the variance function $\sigma^2(\cdot)$. Dette and und Wilkau (2003) used the integrated variance function for this purpose, also obtaining a homoscedasticity test, and proved that under the null hypothesis the test statistic had an asymptotic normal distribution. Subsequently, Dette et al. (2006) extended the latter proposal to volatility functions that depend on the variable X_t. Dette and Podolskij (2008) developed a test based on stochastic processes of the integrated volatility, proving its weak convergence to centered processes with Gaussian conditional distributions, and Podolskij and Ziggel (2008) constructed a test statistic with a range-based estimation of the integrated volatility and the integrated quarticity, which converges weakly to a mixed Gaussian distribution.

6.2.3 Generalized likelihood ratio test

The generalized likelihood statistics were introduced in Fan et al. (2001) to overcome the drawbacks of nonparametric maximum likelihood ratio statistics. Considering the regression model in (6.5) and the conditional log-likelihood function $\ell(\cdot)$, the Generalized likelihood ratio test (GLRT) can be build as

$$\Lambda_n = \ell_n(\mathcal{H}_\text{a}) - \ell_n(\mathcal{H}_0) = \ell_n(m_{n,h}) - \ell_n(m_{\hat{\boldsymbol{\theta}}}),$$

under the assumption that the error process is normally distributed. The log-likelihood under the null hypothesis is given by a maximum likelihood estimator and $\ell_n(\mathcal{H}_\text{a})$ is replaced by a generalized maximum likelihood estimator in the nonparametric context. Define the residual sum of squares under the null, $\text{RSS}_0 = n^{-1}\sum_{i=0}^{n-1}\left(Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i})\right)^2$, and the alternative, $\text{RSS}_1 = n^{-1}\sum_{i=0}^{n-1}\left(Y_{t_i} - m_{n,h}(X_{t_i})\right)^2$, hypotheses, the statistic for the GLRT is given by

$$\Lambda_n = \frac{n}{2}\log\frac{\text{RSS}_0}{\text{RSS}_1} \approx \frac{n}{2}\frac{\text{RSS}_0 - \text{RSS}_1}{\text{RSS}_1},$$

resembling the F-test construction for regression models. The test will reject the null hypothesis for large values of Λ_n, as under $\mathcal{H}_0$ the values of RSS_0 and RSS_1 will be close, whereas under the alternative hypothesis RSS_0 should become systematically larger than RSS_1.

The GLRT ideas developed in Fan et al. (2001) were extended to time-homogeneous diffusion models in Fan and Zhang (2003), while Fan et al. (2003) proposal was based on a generalized pseudo-likelihood ratio test and Aït-Sahalia et al. (2009) developed a specification test for continuous-time jump-diffusion process and derived the asymptotic null distribution. The asymptotic distribution of the GLRT exhibits the Wilks phenomenon, that is, the limit

distribution of Λ_n is asymptotically χ^2, and a similar Wilks type of result holds under $\mathcal{H}_0$ in Fan and Zhang (2003) test. Fan et al. (2003) used bootstrap methods to approximate the null distribution of the test statistic, as well as Aït-Sahalia et al. (2009), though they also established its asymptotic null distribution.

6.2.4 Empirical likelihood

The empirical likelihood (EL) technique (Owen, 1988) allows the construction of nonparametric likelihood for a parameter of interest. The idea of the EL is to maximize an objective function defined by a product of probability weights allocated to observations that, under certain constraints, characterize the functional curve to be tested. There exist a diversity of test statistic formulations via the EL, among them the local EL ratio test is a wide framework. The local version of the EL is constructed as

$$\Lambda_n^L = -2 \int \log \left(\mathcal{L}_n\big(\tilde{m}(x, \hat{\boldsymbol{\theta}}) n^n\big) \right) \omega(x) \, \mathrm{d}x,$$

where $\mathcal{L}_n\big(\tilde{m}(x, \hat{\boldsymbol{\theta}})\big) = \max \prod_{i=0}^{n-1} q_i(x)$, subject to $\sum_{i=0}^{n-1} q_i(x) = 1$ and

$$\sum_{i=0}^{n-1} q_i(x) K_h(x - X_{t_i}) \big(Y_{t_i} - \tilde{m}(x, \hat{\boldsymbol{\theta}})\big) = 0,$$

with $\tilde{m}(x, \hat{\boldsymbol{\theta}}) = \mathbb{E}_{\hat{\boldsymbol{\theta}}} \left[m_{n,h}(x) \right]$ the empirical local likelihood under the null hypothesis. This methodology has been extended to diffusion models, namely Chen et al. (2008) developed a EL test to parametrically specify the transitional density of the process based on kernel estimation. They used the EL to formulate a statistic for each bandwidth that is a Studentized L^2-distance between the nonparametric and parametric estimator of the transitional density implied by the parametric process, where a parametric bootstrap procedure was used to obtain the critical values, due to the slow convergence of the asymptotic distribution. Chen and Gao (2011) proposed a nonparametric test for the parametric specification of both the conditional mean and variance functions based on an EL statistic, calibrated using bootstrap techniques.

6.2.5 Test distribution and calibration

The null hypothesis $\mathcal{H}_0$ is rejected at the level α if the test statistic T_n exceeds the α-level critical value $c_{1-\alpha}$, that is $T_n \geqslant c_{1-\alpha}$, where $c_{1-\alpha}$ satisfies

$$\mathbb{P}\left(T_n \geqslant c_{1-\alpha}\right) = \alpha.$$

If the distribution of the test statistic is known, the value $c_{1-\alpha}$ can be determined. Alternatively, the distribution can be approximated by Monte Carlo procedures or resampling methods, such as bootstrap techniques. Though the limiting distribution of the test statistics is available for some procedures, asymptotic critical values may not be accurate in practice, in particular when the sample size is not sufficiently large and the convergence rate to the limit distribution is slow. The intricacies of empirical interest rate data, highly persistent, can further emphasize this, as the near-nonstationarity of these series can lead to over rejection of the null hypothesis when computing critical values based on the asymptotic distribution derived under stationarity. This was noted in Pritsker (1998), who studied the small sample behavior of the asymptotically normally distributed test of Aït-Sahalia (1996c). Short term rates exhibit very low speeds of mean reversion and, despite the process is still formally stationary, the asymptotics based on assuming stationarity are slow to converge, to such an extent that Aït-Sahalia and Park (2012) investigated that it may well be that the asymptotic behavior derived under nonstationarity is a better approximation to the finite sample behavior of the test. Therefore, when the data generating process is stationary but exhibits high persistence, the ability of the limiting distribution to approximate the finite sample behavior of the test deteriorates. As an alternative, the martingale transformation can be used to obtain asymptotically distribution free tests (see, Khmaladze, 1979; Khmaladze, 1981; Stute et al., 1998b; Khmaladze and Koul, 2004). These transformation has been used in the context of diffusion models, but mainly focused on the market empirical process of the residuals (see Section 6.2.2). On the other hand, bootstrap methods can be used to calibrate the distribution of the test statistic T_n under the null hypothesis, where the critical value $c_{1-\alpha}$ is approximated by $c^*_{1-\alpha}$ such that

$$\mathbb{P}^* \left(T^*_n \geqslant c^*_{1-\alpha} \right) = \alpha,$$

with $\mathbb{P}^*$ the probability measure generated by the bootstrap sample and $c^*_{1-\alpha}$ can be obtained by means of Monte Carlo techniques with a statistic of order $\lceil B(1-\alpha) \rceil$ from B bootstrap replicates $\{T^{*j}_n\}_{j=1}^B$, that is, $c^*_{1-\alpha} = T^{*\lceil B(1-\alpha) \rceil}_n$. Parametric bootstrap procedures were used in different works (i.e., Gao and Casas, 2008; Aït-Sahalia et al., 2009; Chen and Hong, 2010; Monsalve-Cobis et al., 2011), which can be implemented using the following simulation procedure:

1. Generate the bootstrap sample $\left\{ \left(X^*_{t_i}, Y^*_{t_i} \right) \right\}_{i=0}^{n-1}$, with $Y^*_{t_i} = m_{\hat{\boldsymbol{\theta}}}(X_{t_i}) + \sigma_{\hat{\boldsymbol{\theta}}}(X_{t_i}) \Delta^{-\frac{1}{2}} \varepsilon^*_{t_i}$, considering a fixed design ($X^*_{t_i} := X_{t_i}$) and where $\{\varepsilon^*_{t_i}\}$ are i.i.d. random variables following a standard Gaussian distribution, independent of $\{X_{t_i}\}$.
2. Estimate $\hat{\boldsymbol{\theta}}^*$ from the bootstrap sample $\left\{ \left(X_{t_i}, Y^*_{t_i} \right) \right\}_{i=0}^{n-1}$ obtained in the previous step.

3. Define the bootstrap statistic T_n^* to be the version of T_n with $\{(X_{t_i}, Y_{t_i})\}_{i=0}^{n-1}$ and $\hat{\boldsymbol{\theta}}$ being replaced by $\left\{\left(X_{t_i}, Y_{t_i}^*\right)\right\}_{i=0}^{n-1}$ and $\hat{\boldsymbol{\theta}}^*$.

4. Repeat B times the above steps, obtaining the bootstrap replicates $\{T_n^{*j}\}_{j=1}^{B}$.

5. Approximate the critical value, $c_{1-\alpha}^* = T_n^{*\lceil B(1-\alpha)\rceil}$, or the p-value, $\#\{T_n^{*j} > T_n\}/B$.

6.2.6 Overcoming the curse of dimensionality

Some of the tests introduced in this section are subject of the curse of dimensionality, which refers to poor performances of local smoothing methods for multivariate data (Lavergne and Patilea, 2008), as the behavior of nonparametric estimators quickly deteriorates as dimension increases, due to the sparsity of data in multidimensional spaces (Stone, 1980). Test statistics built by the comparison of a nonparametric estimator of the model and an estimator under the null hypothesis (based on the R_{1n} process in (6.6)) or by the comparison of the empirical estimator of the integrated regression function with its estimated values under $\mathcal{H}_0$ (based on the R_{2n} process in (6.8)) are subject of the curse of dimensionality as the dimension increases. The effect of the increasing dimension deteriorates the power of the tests. This problem has been addressed by several authors, regarding tests based on smoothing techniques (that is, on the process R_{1n}), see Lavergne and Patilea (2008) and Xia (2009), whose proposals were inspired on the projection pursuit ideas. For the test based on the empirical regression process, in Stute et al. (2008) the process R_{2n} was replaced by a residual empirical process marked by proper functions of the regressors. Stute et al. (2006) addressed this problem in higher order time series and Escanciano (2006) overcame the curse of dimensionality by using a residual marked empirical process based on projections.

Though different adaptations have been proposed in order to avoid the curse of dimensionality, few have been studied for diffusion processes. In the next section, we use the ideas of correlation distance in Székely et al. (2007) to test the parametric specification of diffusion processes, which circumvents the curse of dimensionality problem.

6.3 Specification test using distance covariance

Given the discretized diffusion model in (6.5), that is, $Y_{t_i} = m(X_{t_i}) + \sigma(X_{t_i})\Delta^{-1/2}\varepsilon_{t_i}$, where $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ and $\varepsilon_{t_i} \sim \mathcal{N}(0,1)$ for $i = 0, \ldots, n-1$, we work with a fixed time span with $T = 1$, that is, the unit interval $[0,1]$ and sampling interval $\Delta = 1/n$. Thus, as the sample size n increases, the time span remains fixed and $\Delta \to 0$ as $n \to \infty$. Under

stationarity and ergodicity, we have that $\{\varepsilon_{t_i}\}_{i=0}^{n-1}$,

$$\varepsilon_{t_i} = \frac{Y_{t_i} - m(X_{t_i})}{\sigma(X_{t_i})/\sqrt{\Delta}},$$

are an i.i.d. sequence independent of the stochastic process $\{X_{t_i}\}_{i=0}^{n-1}$ with $\mathbb{E}\left[\varepsilon_{t_i} \,\middle|\, \mathcal{F}_t\right] = 0$. Therefore, the parametric specification of the drift and diffusion terms in (6.5) can be checked by testing the independence of the variables $\{\varepsilon_{t_i}, X_{t_i}\}_{i=0}^{n-1}$.

Based on the distribution of the residuals, we construct a goodness-of-fit test for the parametric specification of the diffusion model motivated by Székely et al. (2007). Let $\varphi_{\varepsilon,X}(t,s) = \mathbb{E}\left[e^{i(\langle t,\varepsilon\rangle + \langle s,X\rangle)}\right]$ denote the joint characteristic function of $X_{t_i} \in \mathbb{R}$ and $\varepsilon_{t_i} \in \mathbb{R}$ and $\varphi_\varepsilon(t) = \varphi_{\varepsilon,X}(t,0)$ and $\varphi_X(s) = \varphi_{\varepsilon,X}(0,s)$ the characteristic functions of ε_{t_i} and X_{t_i}, respectively. In terms of characteristic functions, X_{t_i} and ε_{t_i} are independent if and only if $\varphi_{\varepsilon,X} = \varphi_\varepsilon\varphi_X$. Therefore, to test the hypothesis of independence

$$\mathcal{H}_0 \colon \varphi_{\varepsilon,X} = \varphi_\varepsilon\varphi_X \qquad \text{vs.} \qquad \mathcal{H}_\text{a} \colon \varphi_{\varepsilon,X} \neq \varphi_\varepsilon\varphi_X, \tag{6.9}$$

we can use a distance dependence measure such as the distance covariance (Székely et al., 2007) defined as

$$\begin{aligned}\mathcal{V}^2(\varepsilon_t, X_t; \mu) &:= \left\|\varphi_{\varepsilon,X}(t,s) - \varphi_\varepsilon(t)\varphi_X(s)\right\|_\mu^2 \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \left|\varphi_{\varepsilon,X}(t,s) - \varphi_\varepsilon(t)\varphi_X(s)\right|^2 \mu(t,s)\,\mathrm{d}t\,\mathrm{d}s,\end{aligned} \tag{6.10}$$

where $\mu(t,s) = (c^2|t|^{\alpha+1}|s|^{\alpha+1})^{-1}$ is a non-negative weight function used in the $\|\cdot\|_\mu$-norm, for a constant $c > 0$ and $\alpha \in (0,2)$. Following Székely et al. (2007), we fix $\alpha = 1$ and thus assume $\mu(t,s) = (c^2t^2s^2)^{-1}$, with $c = \pi$, as with this choice of μ the distance correlation, $\mathcal{R}_\mu = \mathcal{V}(\varepsilon_t, X_t; \mu)/(\mathcal{V}(\varepsilon_t, \varepsilon_t; \mu)\mathcal{V}(X_t, X_t; \mu))^{1/2}$, is invariant relative to scale and orthogonal transformations.

As $\mathcal{V}^2(\varepsilon_t, X_t; \mu) = 0$ if and only if ε_t and X_t are independent, that is, $\varphi_{\varepsilon,X} = \varphi_\varepsilon\varphi_X$, a test statistic can be constructed based in the distance covariance. Under the null hypothesis we have $\mathcal{V}_0^2(\varepsilon_{t_i0}, X_{t_i}; \mu) = 0$, with $\varepsilon_{t_i0} = (Y_{t_i} - m_0(X_{t_i}))/(\sigma_0(X_{t_i})/\sqrt{\Delta})$, since $\mathbb{E}\left[\varepsilon_{t_i0} \,\middle|\, \mathcal{F}_t\right] = 0$. To implement the test we need to consider a root-n consistent estimator $\hat{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta}_0$, obtaining the residuals $\hat{\varepsilon}_{t_i} = \big(Y_{t_i} - m_{\hat{\boldsymbol{\theta}}}(X_{t_i})\big)/\big(\sigma_{\hat{\boldsymbol{\theta}}}(X_{t_i})/\sqrt{\Delta}\big)$. An empirical estimator of the distance covariance (6.10) can be obtained by replacing the characteristic

functions φ by their sample analogs

$$\begin{aligned}\mathcal{V}_n^2(\hat{\varepsilon}_t, X_t; \mu) &= \left\| \varphi^{(n)}_{\hat{\varepsilon},X}(t,s) - \varphi^{(n)}_{\hat{\varepsilon}}(t)\varphi^{(n)}_X(s) \right\|^2_\mu \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \left| \varphi^{(n)}_{\hat{\varepsilon},X}(t,s) - \varphi^{(n)}_{\hat{\varepsilon}}(t)\varphi^{(n)}_X(s) \right|^2 \mu(t,s)\,\mathrm{d}t\,\mathrm{d}s,\end{aligned}$$

where the empirical characteristic functions are given by $\varphi^{(n)}_{\hat{\varepsilon},X}(t,s) = \frac{1}{n}\sum_{j=1}^n e^{i\langle t,\hat{\varepsilon}_{t_j}\rangle + i\langle s, X_{t_j}\rangle}$, $\varphi^{(n)}_{\hat{\varepsilon}}(t) = \varphi^{(n)}_{\hat{\varepsilon},X}(t,0)$ and $\varphi^{(n)}_X(s) = \varphi^{(n)}_{\hat{\varepsilon},X}(0,s)$. However, due to the computational cost of this definition, a simpler formula with practical advantages is preferable. An equivalent (see Székely et al., 2007, Theorem 1) empirical distance covariance is defined by

$$\mathcal{V}_n^2(\hat{\varepsilon}_{t_i}, X_{t_i}) = \frac{1}{n^2}\sum_{k,l=0}^{n-1} A_{kl}B_{kl},$$

where A_{kl} and B_{kl} are linear functions of the pairwise distances between sample elements,

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}, \qquad B_{kl} = b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot},$$

with $a_{kl} = |\hat{\varepsilon}_{t_k} - \hat{\varepsilon}_{t_l}|$ and $b_{kl} = |X_{t_k} - X_{t_l}|$ the Euclidean distance matrices and

$$\bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=0}^{n-1} a_{kl}, \qquad \bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=0}^{n-1} a_{kl}, \qquad \bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=0}^{n-1} a_{kl},$$

for $k, l = 0, \ldots, n-1$ and, similarly, define $\bar{b}_{k\cdot}, \bar{b}_{\cdot l}$ and $\bar{b}_{\cdot\cdot}$. Given the empirical distance covariance $\mathcal{V}_n^2$, the test of independence for the null hypothesis in (6.9) is based on the statistic

$$T_n = n\frac{\mathcal{V}_n^2(\hat{\varepsilon}_{t_i}, X_{t_i})}{S_2},$$

where $S_2 = \bar{a}_{\cdot\cdot}\bar{b}_{\cdot\cdot}$. The test rejects independence for large values of T_n and a resampling procedure can be used to determine the quantiles of the test statistic, such as the parametric bootstrap procedure in Section 6.2.5. In Székely et al. (2007), replicates of T_n were computed under random permutations of one of the sample indices.

6.4 Simulation study

In this Section, we carry out a simulation study to compare the performance of the methods described, namely, the empirical regression (henceforth abbreviated as ER) process-based test

of Monsalve-Cobis et al. (2011) and its asymptotically distribution-free (ADF) transformation (Chen et al., 2015), for both Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) test statistics; the nonparametric (NP) proposal of Gao and Casas (2008); and the test based on distance covariance (DCOV) introduced in Section 6.3.

The finite sample behavior of tests for higher order time series is also addressed by studying the effect of the curse of dimensionality in the test based on the process R_{2n} in (6.8), as Monsalve-Cobis et al. (2011) proposal, and the distance covariance test of Section 6.3, designed for avoiding the curse of dimensionality.

6.4.1 Comparative study

We study the finite sample properties of the tests using Monte Carlo simulations, given the diffusion model in (6.2),

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t.$$

The sample paths for the models are simulated using Milsteins scheme (Milshtein, 1979) and we consider the time interval $[0, 1]$, with $\Delta = 1/n$, generating $1\,000$ random samples with sample sizes $n \in \{100, 250, 500, 1\,000\}$. The first thousand observations are discarded to lessen the impact of the initial value. Regarding the methods that require bootstrap resamples to calibrate the test, we chose $B = 1\,000$ replicates. The unknown parameters are estimated using the Kalman (1960) filter, discussed in Chapter 3.

As null hypothesis we test the parametric specification of the diffusion function,

$$\mathcal{H}_0 \colon \sigma^2(x, \boldsymbol{\theta}) = \sigma^2 x^2,$$

where σ is an unknown parameter, and simulate under different scenarios (similar design as in Dette and Podolskij, 2008):

- Size simulation: we consider five scenarios with diffusion function $\sigma^2(X_t) = X_t^2$, therefore $\sigma = 1$, and drift functions $m_1(X_t) = 0$, $m_2(X_t) = 2$, $m_3(X_t) = X_t$, $m_4(X_t) = 2 - X_t$ and $m_5(X_t, t) = tX_t$.
- Power simulation: we consider five scenarios with drift function $m(X_t) = 2 - X_t$ and diffusion functions $\sigma_1^2(X_t) = 1 + X_t^2$, $\sigma_2^2(X_t) = 1$, $\sigma_3^2(X_t) = 5|X_t|^{1.5}$, $\sigma_4^2(X_t) = 5|X_t|$ and $\sigma_5^2(X_t) = (1 + X_t)^2$.

Table 6.1 and 6.2 show the empirical rejection rates for $\alpha = 0.05$. Regarding the sizes in Table 6.1, all methods approximate the nominal level α, though the ADF test shows some

		ER		ADF			
$m(\cdot)$	n	KS	CvM	KS	CvM	NP	DCOV
0	100	0.038	0.044	0.043	**0.069**	0.060	**0.065**
	250	0.048	0.055	0.049	0.057	0.059	0.046
	500	0.052	0.044	0.052	0.061	0.046	0.047
	1000	0.039	0.044	0.041	0.055	0.047	0.051
2	100	0.057	**0.072**	**0.028**	0.044	0.052	**0.076**
	250	0.037	0.040	0.033	0.038	0.055	0.062
	500	0.044	0.041	**0.033**	**0.036**	0.044	0.064
	1000	0.038	0.037	0.033	0.036	0.038	0.049
X_t	100	0.044	0.058	0.031	0.050	0.047	0.060
	250	0.060	0.063	0.042	0.054	0.056	0.056
	500	0.041	0.048	0.034	0.036	0.048	0.053
	1000	0.059	0.061	0.033	0.044	0.044	0.062
$2 - X_t$	100	0.043	0.055	**0.032**	0.048	**0.067**	**0.080**
	250	0.046	0.053	**0.034**	**0.035**	0.061	0.059
	500	0.050	0.053	0.037	0.037	0.053	0.057
	1000	0.047	0.062	0.037	0.046	0.052	0.053
tX_t	100	0.058	**0.071**	**0.030**	**0.033**	**0.066**	**0.065**
	250	0.051	0.056	0.045	0.054	0.060	**0.068**
	500	0.058	**0.067**	0.040	0.042	0.037	0.057
	1000	0.055	0.055	0.037	0.039	0.050	0.048

Table 6.1: Empirical sizes for $\alpha = 0.05$. Boldface indicates the rates of rejection outside a 95% confidence interval for the nominal level α.

under-rejection through the different scenarios, while the distance covariance test slightly over-rejects. Nevertheless, none of the methods show repeated mis-calibration. With respect to power in Table 6.2, all the methods perform closely, with the distance covariance proposal showing less power for the smallest sample size in certain scenarios, as $\sigma_3^2(X_t) = 5|X_t|^{1.5}$ for $n = 100$.

6.4.2 Higher order time series and the curse of dimensionality

To illustrate the effect of the curse of dimensionality, we conduct a simulation study with higher order time series. A continuous-time analogue of the $p \in \mathbb{N}$ order AR(p) process is

		ER		ADF			
$\sigma^2(\cdot)$	n	KS	CvM	KS	CvM	NP	DCOV
$1 + X_t^2$	100	0.605	0.599	0.496	0.637	0.580	0.493
	250	0.867	0.860	0.803	0.854	0.800	0.772
	500	0.976	0.967	0.938	0.962	0.955	0.955
	1000	0.997	0.996	0.970	0.985	0.990	0.993
1	100	0.579	0.582	0.587	0.742	0.678	0.592
	250	0.943	0.953	0.934	0.965	0.964	0.942
	500	1.000	0.999	0.965	0.979	0.997	0.995
	1000	1.000	1.000	0.983	0.989	0.999	0.996
$5\,\lvert X_t\rvert^{1.5}$	100	0.514	0.551	0.529	0.735	0.478	0.235
	250	0.869	0.873	0.872	0.940	0.764	0.564
	500	0.974	0.973	0.990	0.995	0.936	0.893
	1000	1.000	1.000	0.999	1.000	0.990	0.994
$5\,\lvert X_t\rvert$	100	0.830	0.810	0.818	0.909	0.811	0.809
	250	0.961	0.960	0.972	0.993	0.961	0.984
	500	0.989	0.989	0.992	0.998	0.988	1.000
	1000	0.993	0.993	0.999	1.000	0.995	1.000
$\left(1 + X_t\right)^2$	100	0.806	0.790	0.694	0.799	0.778	0.654
	250	0.954	0.950	0.895	0.934	0.940	0.893
	500	0.996	0.992	0.945	0.968	0.990	0.980
	1000	1.000	0.999	0.968	0.980	0.999	0.991

Table 6.2: Empirical power for $\alpha = 0.05$.

the continuous-time Gaussian autoregressive process CAR(p) symbolically defined to be a stationary solution of the SDE

$$a(D)Y(t) = \sigma DW(t),$$

with $a(D) = D^p + \alpha_1 D^{p-1} + \cdots + \alpha_p$, where the operator D denotes differentiation with respect to t and $\{W(t)\}_{t\geqslant 0}$ is a standard Wiener process. Since $DW(t)$ does not exist in the usual sense (it does not exist as a random function), a meaningful interpretation can be given by writing it in the state-space form,

$$Y(t) = \mathbf{b}^\top \mathbf{X}(t),$$

$\mathbf{b}^\top = (1, 0, \ldots, 0)$, where the state vector $\mathbf{X}(t) = \left(X^{(0)}(t), \ldots, X^{(p-1)}(t)\right)^\top$, with $X^{(j)}(t)$ the j^{th} mean-square and pathwise derivative $D^j X^{(0)}(t)$, satisfies the Itô equation

$$\mathrm{d}\mathbf{X}(t) = \mathbf{A}\mathbf{X}(t)\,\mathrm{d}t + \sigma\mathbf{e}\,\mathrm{d}W(t),$$

with $\sigma \in (0, \infty)$,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\alpha_p & -\alpha_{p-1} & -\alpha_{p-2} & \cdots & -\alpha_1 \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

under the initial condition that $\mathbf{X}(0)$ is normally distributed and independent of $W(t)$ for all $t \geqslant 0$. The CAR(p) process $\{Y(t)\}_{t \geqslant 0}$ exists and satisfies

$$Y(t) = \mathbf{b}^\top e^{\mathbf{A}t}\mathbf{X}(0) + \sigma\mathbf{b}^\top \int_0^t e^{\mathbf{A}(t-u)}\mathbf{e}\,\mathrm{d}W(u),$$

for $t \geqslant 0$. Note that the CAR(1) is the Ornstein-Uhlenbeck process, usually parametrized as

$$\mathrm{d}Y_t = \kappa(\mu - Y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad t \in [0, T], \tag{6.11}$$

where μ is the long term mean, κ the rate of mean reversion and $\sigma > 0$ the volatility around the mean. To estimate the autoregressive coefficients $\alpha_1, \alpha_2, \ldots, \alpha_p$, we proceed to estimate the analogue discrete-time ARMA process $\{Y_n^{(\Delta)} := Y(n\Delta), n = 0, 1, 2, \ldots\}$ by maximum likelihood and from that estimates we obtain $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$.

Regarding non-linear time series, we consider the continuous-time analogue of the threshold models of Tong (1983), the continuous-time threshold autoregressive process CTAR(1), given by

$$\begin{aligned} DY(t) + \alpha_1 Y(t) &= DW(t) + \alpha_1\mu_1, \quad \text{if } Y(t) \leqslant r, \\ DY(t) + \alpha_2 Y(t) &= DW(t) + \alpha_2\mu_2, \quad \text{if } Y(t) > r, \end{aligned} \tag{6.12}$$

with threshold r.

To study the behavior of the tests as the dimension p increases we use Monte Carlo simulations. As null hypothesis we test the CAR(p) process, for different order p, to illustrate the finite sample behavior regarding power and size:

- Size simulation: we generate samples from the CAR(1) process, namely the Ornstein-

Uhlenbeck process in (6.11), noting that the CAR(1) may be embedded in CAR(p), and test for CAR(p) with $p \in \{1, 2, 3, 4, 5\}$. The parameter values are $(\mu, \kappa, \sigma^2)^\top = (0.08, 0.5, 0.25)^\top$.

- Power simulation: we consider the CTAR(1) process in (6.12) with parameters given by $(\mu_1, \alpha_1, \mu_2, \alpha_2)^\top = (4, 0.5, -1.25, -0.4)^\top$ and $r = 1$, and test for CAR(p) with $p \in \{1, 2, 3, 4, 5\}$.

Table 6.3 depicts the rate of rejection under the null hypothesis, where the empirical regression process-based proposal of Monsalve-Cobis et al. (2011), both for the Kolmogorov-Smirnov and the Cramér-von Mises statistics, and the covariance distance test of Section 6.3 attained the nominal level $\alpha = 0.05$ under the null hypothesis. However, when analyzing the empirical power in Table 6.4, the curse of dimensionality arises as the dimension p increases for both the Kolmogorov-Smirnov and the Cramér-von Mises criteria, where power drastically decreases for higher order p. Except for dimension $p = 1$ when both tests are equivalent and produce similar results, the covariance distance test is more powerful, circumventing the curse of dimensionality.

	Kolmogorov-Smirnov				**Cramér-von Mises**				**Covariance distance**			
p n:	100	250	500	1000	100	250	500	1000	100	250	500	1000
1	0.045	0.054	0.039	0.057	0.056	0.049	0.037	0.053	0.044	0.041	0.042	0.048
2	0.060	0.055	0.062	0.053	**0.066**	0.046	0.056	0.049	0.052	0.048	0.041	0.046
3	0.052	0.039	0.044	0.042	0.051	0.040	0.039	0.047	0.050	0.048	0.042	0.045
4	0.045	0.034	0.055	0.051	0.039	0.042	0.050	0.039	0.049	0.047	0.039	0.043
5	0.047	0.041	0.053	0.054	0.042	0.046	0.054	0.043	0.054	0.045	0.038	0.046

Table 6.3: Empirical levels under the null hypothesis, with $\alpha = 0.05$, testing the CAR(p) process with order $p \in \{1, 2, 3, 4, 5\}$. Bolfaced the rates of rejection outside a 95% confidence interval for the nominal level α.

	Kolmogorov-Smirnov				**Cramér-von Mises**				**Covariance distance**			
p n:	100	250	500	1000	100	250	500	1000	100	250	500	1000
1	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.942	1.000	1.000	1.000
2	0.743	0.999	1.000	1.000	0.811	1.000	1.000	1.000	0.926	1.000	1.000	1.000
3	0.204	0.447	0.702	0.949	0.194	0.499	0.804	0.986	0.921	1.000	1.000	1.000
4	0.131	0.247	0.458	0.722	0.127	0.280	0.514	0.823	0.917	1.000	1.000	1.000
5	0.105	0.165	0.272	0.472	0.100	0.174	0.324	0.571	0.911	1.000	1.000	1.000

Table 6.4: Empirical power under the alternative hypothesis, with $\alpha = 0.05$, testing the CAR(p) process with order $p \in \{1, 2, 3, 4, 5\}$.

6.5 Real data application

We apply the methodology reviewed in Section 6.2 and Section 6.3 to interest rate models involving treasury securities from the secondary market rates. As in Chapter 5, we use daily market yields on U.S. Treasury securities with different maturities, ranging from 1 month to 30 years, obtained from the Federal Reserve Bank of St. Louis. Treasury securities (T-bills, T-notes, and T-bonds) are financial contracts issued by the U.S. government with price $P(t, \tau)$ at time t that yields a know amount on the maturity date τ, where $P(t, \tau)$ is determined by the rate evolution. The data consist of 999 daily observations ($\Delta = 1/252$), from January 2, 2016 to December 31, 2019, for 11 maturities (see Figure 5.11).

As in the previous section, we consider the empirical regression (ER) process-based test of Monsalve-Cobis et al. (2011) and its asymptotically distribution-free (ADF) transformation (Chen et al., 2015), for both Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) test statistics; the nonparametric (NP) proposal of Gao and Casas (2008); and the test based on distance covariance (DCOV) introduced in Section 6.3. We apply the procedures to test the parametric specification of the diffusion function of CKLS interest rate model (Chan et al., 1992) given by

$$\mathrm{d}X_t = \kappa(\mu - X_t) + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

that is, the null hypothesis $\mathcal{H}_0\colon \sigma(X_t, \boldsymbol{\theta}) = \sigma X_t^{\gamma}$. Table 6.5 reports the p-values for the goodness-of-fit tests, which were obtained based on $1\ 000$ bootstrap resamples. The empirical p-values are large for longer maturities (7, 10, 20 and 30 years), hence, the samples show no significant evidences against the CKLS diffusion function for any sensible significance level. Conversely, the p-values lead to strong rejections of the null hypothesis for short maturities (1,3,6 and 12 months), implying that the model is inadequate to explain the volatility of the T-bill series. The 2, 3 and 5 years maturities show the most noticeable disparity among the different procedures. The deterministic parametric specification of the CKLS diffusion function, where many known models can be nested (as Vasicek, 1977, with $\gamma = 0$ or the Cox et al., 1985 model with $\gamma = 1/2$), is not able to explain the dynamics of the treasury data for short maturities. Therefore, capturing the volatility behavior may require a more intricate form, such as stochastic volatility.

Under the sample period considered, central banks from different economies have acted in unprecedented scale. Regarding the U.S., the Federal Reserve has increased rates several times since 2015 and reduced the rate three times in 2019, for a total of 75 basis points. Short maturities of treasury yield are tied close to Fed policy, as shown in Figure 6.1, where the evolution of the one year bond yield moves closely with the Federal fund rate. Therefore, the

	ER		ADF			
Maturity	**KS**	**CvM**	**KS**	**CvM**	**NP**	**DCOV**
1 month	< 0.001	< 0.001	$< 10^{-6}$	$< 10^{-6}$	< 0.001	0.006
3 month	0.002	0.014	5.9×10^{-4}	0.002	0.013	0.003
6 month	0.003	0.017	0.008	0.038	0.001	0.002
12 month	< 0.001	< 0.001	$< 10^{-6}$	$< 10^{-6}$	< 0.001	0.002
2 year	0.007	0.013	2×10^{-4}	3.2×10^{-4}	0.001	0.056
3 year	0.456	0.662	0.004	0.008	0.875	0.164
5 year	0.557	0.680	0.078	0.170	0.052	0.428
7 year	0.847	0.926	0.447	0.590	0.280	0.532
10 year	0.906	0.967	0.969	0.922	0.987	0.469
20 year	0.981	0.983	0.380	0.400	0.998	0.436
30 year	0.549	0.847	0.614	0.805	0.999	0.488

Table 6.5: p-values for the goodness-of-fit test for the CKLS parametric form of the diffusion function.

CKLS model might fail to adequately account structural shifts of the process due to intervention of the Federal Reserve and extending the diffusion function to incorporate unexpected shocks could provide a superior empirical fit.

To provide more insight into the rejection of the null hypothesis for shorter maturities (T-bill), we compare the parametric and non parametric estimation of the volatility function for all maturities. The parametric $\sigma_{\hat{\boldsymbol{\theta}}}(x)$ and kernel estimation $\sigma_{n,h}(x)$ of the diffusion function are shown in Figure 5.14 of Chapter 5, along 95% confidence bands, where a Gaussian kernel was used, $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$, with a cross-validation bandwidth.

The poor fit of the parametric diffusion function for T-bills, almost constant, may be due to the noisy nature of the series or the presence of jumps and structural shocks. In the parametric specification of the CKLS model, the parameter γ is called the "level effect", as it determines the sensitivity of the variance with respect to the interest rate level. The estimated value of γ is close to zero for the shorter yields and decreases with maturity, reaching negative values. In this regard, the models may be unstationary, as $\gamma < 0$ does not guarantee stationarity. The kernel estimation for longer maturities, very close to the parametric estimation, also shows an inverse relationship between the interest rate level and the volatility, which is rather unexpected. However, the null hypothesis was not rejected. As already discussed, the sample period considered was subject of structural changes in markets. The interest rates on some long maturities bonds fell bellow short-term debt, as shown in Figure 6.1, where the spread between the 1 year treasury note and the 10 year bond inverted in 2019. Long-term bond yields evolved in the direction of shorter terms, with the 30 year bond falling bellow 2%,

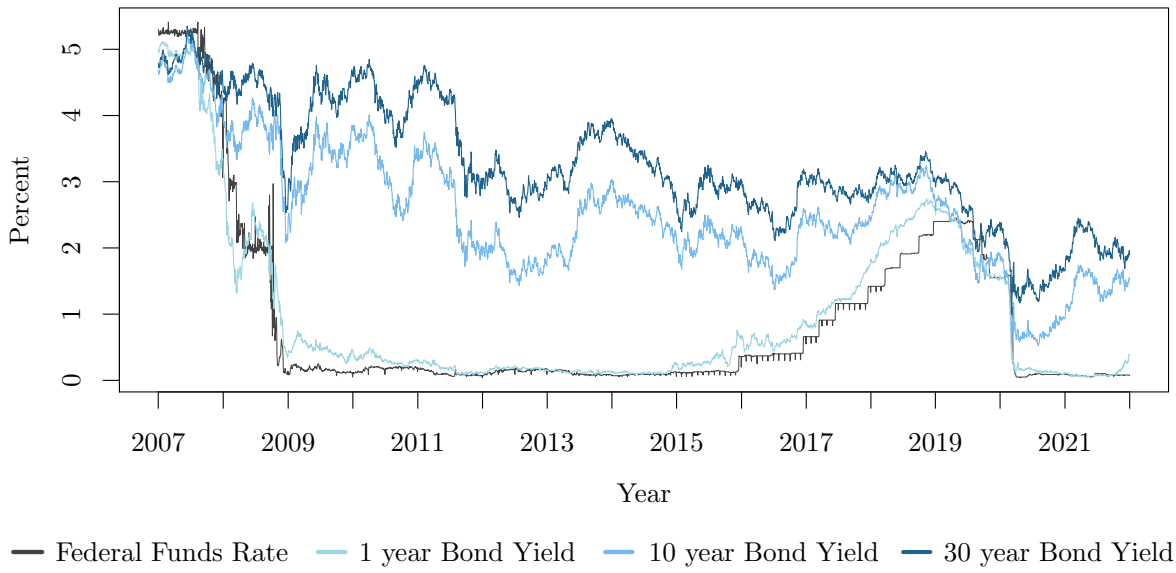


Figure 6.1: U.S. federal funds rate and U.S. treasury yields (1, 10 and 30 year bond yields).

reaching its lowest level up to that date. This implied a flattening yield curve for the sample period, actually inverting in 2019. Higher yields were pushed down by several factors other than market expectations about interest rate evolution, so that the inversion of the yield curve may be a result of structural market changes, which could account for the negative level effect in long-term bonds observed in the sample.

6.6 Conclusions

We reviewed goodness-of-fit test for diffusion models, collecting the proposals of the last 25 years, many of which are extension of more classical proposals to the context of continuous-time stochastic processes. An application of correlation distance ideas to develop a test for the parametric specification of diffusion models was also provided. Empirical results showed that the performance of the procedures, in terms of size and power, are close, calibrating adequately the null hypothesis and detecting a variety of alternatives. However, when dimension increases, some tests lose power. In this scenario, the correlation distance proposal outperforms others, avoiding the curse of dimensionality. An application of the tests to interest rate models involving U.S. daily Treasury securities was also undertaken, finding empirical support to the CKLS diffusion function for long maturities. As the maturity range was reduced, the p-values became small, indicating that the model was mis-specified for T-bill (1,3,6 and 12 month maturities) series.

Chapter 7

Goodness-of-fit test for diffusion models with stochastic volatility

Given the importance of continuous-time stochastic volatility models to describe the dynamics of interest rates, we propose a goodness-of-fit test for the parametric form of the drift and diffusion functions, based on a marked empirical process of the residuals. The test statistics are constructed using a continuous functional (Kolmogorov-Smirnov and Cramér-von Mises) over the empirical processes. In order to evaluate the proposed tests, we implement a simulation study where a bootstrap method is considered for the calibration of the tests. An application of the procedures to real data is provided, namely the Euribor time series analyzed in Chapters 3 and 4, where the estimation of one and two-factor diffusion models was addressed.

The contents of this chapter are collected in López-Pérez et al. (2022b).

Contents

7.1 Introduction

Throughout this dissertation it has been discussed that, in the last five decades, continuous-time models have proven to be an essential part of the financial econometrics field. A large body of literature for the term structure of interest rates is written in continuous-time and different specifications have been proposed, such as the time-homogeneous diffusion given by a stochastic differential equation driven by a Wiener process W_t,

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t,$$

defined on a complete probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}\big)$. In order to determine if the model is appropriate for a given time series, the parametric form of both drift and volatility functions can be tested. For this purpose, there exist several proposals for continuous-time model specification, as discussed in Chapter 6. However, the empirical evidence obtained from the goodness-of-fit tests for the one-factor model proved unsatisfactory empirical fits and suggested that more flexible specifications for the volatility function were needed to capture the dynamics of returns of interest rates. This conclusion was also reached in Chapters 3 and 6, when testing the deterministic volatility function with the Euribor and treasury bonds data (see Sections 3.4 and 6.5, respectively). Therefore, the literature has moved towards two-factor formulations, allowing the volatility function to incorporate a source of random variation, leading to a continuous-time stochastic volatility (SV) model, such as

$$\mathrm{d}X_t = m_1(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma_t \nu_1(X_t, \boldsymbol{\theta})\,\mathrm{d}W_{1,t}, \tag{7.1}$$

$$\mathrm{d}g(\sigma_t) = m_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}t + \nu_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}W_{2,t}, \tag{7.2}$$

where the functions g, m_1, ν_1, m_2 and ν_2 are sufficiently smooth and satisfy growth conditions to obtain existence and uniqueness for the stochastic differential equation solution, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ is an unknown parameter vector. As discussed in Chapter 4, several parametrizations of (7.1)-(7.2) have been examined, see Hull and White (1987), Heston (1993), Andersen and Lund (1997a), Gallant and Tauchen (1998), Eraker (2001), or Christoffersen et al. (2009), among others.

Given the importance of the volatility in the financial market—a measure of risk that impacts in portfolio selection, option pricing, risk management or hedging—its misspecification could lead to serious consequences. To overcome this, goodness-of-fit test should be used to study the adequacy of the proposed model towards the dynamics of the volatility. Some recent literature have addressed this issue, Lin et al. (2013, 2016) proposed a test based on the deviations between the empirical characteristic function and the parametric counterpart; Lin et al. (2014) considered a Bickel-Rosenblatt type test; Zu (2015) used the L^2-distance to measure the discrepancy between the kernel and parametric deconvolution density estimator of an integrated volatility density; Vetter (2015) test was based on a Kolmogorov-Smirnov statistic; Bull (2017) proposed a wavelet-based test; Ebner et al. (2018) used Fourier methods; Christensen et al. (2019) proposal was based on the empirical distribution function; and Li et al. (2022) proposed a test for the integrated volatility of volatility.

In the present chapter, we propose goodness-of-fit tests based on empirical processes, extending the methodology proposed by Monsalve-Cobis et al. (2011) to diffusion models with stochastic volatility, following the ideas suggested in González-Manteiga et al. (2017). Many goodness-of-fit test in the literature for continuous-time models, such as (7.1)–(7.2), focus on the stationary distribution of the volatility σ_t^2, however, we test that the diffusion and drift functions belong to a certain parametric family, that is,

$$\mathcal{H}_{0\nu}\colon \nu_1 \in \{\nu_1(\cdot, \boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\},$$
$$\mathcal{H}_{0m}\colon m_1 \in \{m_1(\cdot, \boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\},$$

respectively. To construct the test statistic we use integrated regression models, an approach discussed in Stute (1997), where the study of a marked empirical process based on residuals was introduced and, subsequently, extended to time series in Koul and Stute (1999). The empirical regression processes-based goodness-of-fit tests have been studied by other authors, see Diebolt (1995) for a nonlinear parametric regression function or Diebolt and Zuber (1999, 2001), for an extension to nonlinear and heteroscedastic regression.

The rest of the chapter is structured as follows. In Section 7.2, the goodness-of-fit tests for the drift and diffusion functions are introduced, while in Section 7.3 a simulation study of the proposed tests is implemented. Real data application to interest rate series is presented in Section 7.4, where the Euribor time series of Chapters 3 and 4 is analyzed, as in those chapters the estimation of the one and two-factor CKLS model was discussed. Lastly, conclusions are drawn in Section 7.5.

7.2 A goodness-of-fit test for diffusion processes

In this section, two goodness-of-fit test are introduced. Based on the methodology developed by Stute (1997) and extending the goodness-of-fit test presented in Monsalve-Cobis et al. (2011), we propose a test for the parametric form of the drift and diffusion functions of the continuous-time stochastic volatility models in (7.1)–(7.2). The test for the drift function is based on the integrated regression function of the process, while the test for diffusion function relies on the integrated volatility function. The test statistics are based on a distance of the resulting residual marked empirical processes from their expected zero mean, measured by Kolmogorov-Smirnov and Cramér-von Mises functionals. In both tests, the distribution of the statistic is approximated by bootstrap techniques, which rely on the Kalman filter algorithm introduced in Chapter 4 (see Section 4.2.1) and is discussed in this section.

7.2.1 Test for the volatility function

The goodness-of-fit test for the parametric form of the volatility function $\nu_1(\cdot)$ in (7.1)–(7.2) under the assumption that the null hypothesis

$$\mathcal{H}_{0\nu}\colon \nu_1 \in \{\nu_1(\cdot,\boldsymbol{\theta})\colon \boldsymbol{\theta}\in\Theta\}$$

holds, is based on the integrated conditional variance function

$$V_0(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \nu_1^2(u,v)\,\mathrm{d}F_{\boldsymbol{\theta}}(u,v) = \mathbb{E}\big[\nu_1^2(X_t,\boldsymbol{\theta})\mathbb{1}_{\{X_t\leqslant x,\sigma_t^2\leqslant y\}}\big],$$

with $x,y\in\mathbb{R}$ and where F is the stationary distribution of $\{X_t,\sigma_t^2\}$ and $\mathbb{1}_{\{\cdot\}}$ the indicator function. Considering an empirical estimator of $V(x,y)$,

$$V_{0n}(x,y) = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_{\{X_{t_i}\leqslant x,\sigma_{t_i}^2\leqslant y\}}\left(\frac{X_{t_{i+1}}-X_{t_i}}{\Delta} - m_1(X_{t_i},\hat{\boldsymbol{\theta}})\right)^2,$$

and assuming that $\hat{\boldsymbol{\theta}}$ is a root-n consistent estimator of the true parameter $\boldsymbol{\theta}$, the goodness-of-fit test is based on the empirical process

$$R_n^{\nu}(x,y) = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathbb{1}_{\{X_{t_i}\leqslant x,\hat{\sigma}_{t_i}^2\leqslant y\}}\left[\left(\frac{X_{t_{i+1}}-X_{t_i}}{\Delta} - m_1(X_{t_i},\hat{\boldsymbol{\theta}})\right)^2 - \frac{\hat{\sigma}_{t_i}^2\nu_1^2(X_{t_i},\hat{\boldsymbol{\theta}})}{\Delta}\right],$$

with $x,y\in\mathbb{R}$ and $\hat{\sigma}_t^2$ an estimate of the volatility. A continuous functional $\Psi(\cdot)$ of the empirical process can be considered to define the test statistic $U_n = \Psi(R_n^{\nu})$. The null hypothesis is

rejected if $U_n > c_{1-\alpha}$, where $c_{1-\alpha}$ is the critical value for the α-level test,

$$\mathbb{P}\left(\Psi(R_n^\nu) > c_{1-\alpha} \mid \mathcal{H}_{0\nu}\right) = \alpha.$$

The critical value $c_{1-\alpha}$ can be determined by approximating the distribution of the process R_n^ν using bootstrap techniques (Stute et al., 1998a). Let $c^*_{1-\alpha}$ denote the bootstrap approximated critical value $c_{1-\alpha}$, so that $\mathbb{P}^*\left(\Psi(R_n^{\nu*}) > c^*_{1-\alpha} \mid \mathcal{H}_{0\nu}\right) = \alpha$, where $\mathbb{P}^*$ is the probability measure generated by the bootstrap sample and the bootstrap counterpart of the empirical process $R_n^\nu(\cdot)$ is given by

$$R_n^{\nu*}(x,y) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \mathbb{1}_{\{X^*_{t_i} \leqslant x, \hat{\sigma}^{*2}_{t_i} \leqslant y\}} \left[\left(\frac{X^*_{t_{i+1}} - X^*_{t_i}}{\Delta} - m_1(X^*_{t_i}, \hat{\boldsymbol{\theta}}^*) \right)^2 - \frac{\hat{\sigma}^{*2}_{t_i} \nu_1^2(X^*_{t_i}, \hat{\boldsymbol{\theta}}^*)}{\Delta} \right],$$

with $\hat{\boldsymbol{\theta}}^*$ an estimator obtained from the bootstrapped sample $\{(X^*_{t_i}, \sigma^{*2}_{t_i})\}$ (the procedure to obtain the resamples is defined in Section 7.2.3). Bootstrap replicates of the statistic $U_n^{*j} = \Psi\left(R_n^{\nu*}(x,y)\right)$, for $j = 1, \ldots, B$, are obtained and, using Monte Carlo techniques, the critical value is approximated with the statistic of order $\lceil B(1-\alpha) \rceil$ from the B bootstrap replicates, that is, $c^*_{1-\alpha} = U_n^{*\lceil B(1-\alpha) \rceil}$ (see Algorithm 4 for a summary of the bootstrap procedure to approximate the critical value $c_{1-\alpha}$). The null hypothesis $\mathcal{H}_{0\nu}$ is rejected if $U_n > c^*_{1-\alpha}$.

The functional $\Psi(\cdot)$ can take the form of the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) criteria, so that the statistics can be expressed as

$$U_n^{KS} := \sup_{x,y} \left| R_n^\nu(x,y) \right|,$$
$$U_n^{CvM} := \int\int_{\mathbb{R}^2} \left(R_n^\nu(x,y) \right)^2 F_n(\mathrm{d}x, \mathrm{d}y),$$

respectively, where F_n is the empirical distribution of $\{X_{t_i}, \hat{\sigma}^2_{t_i}\}_{i=0}^{n-1}$. The empirical p-value is estimated with the proportion of the B bootstrap replicates U_n^{*j} exceeding U_n, that is, $\sharp\{U_n^{*j} > U_n\}/B$.

7.2.2 Test for the drift function

Aiming to test if the parametric form of the drift function $m_1(\cdot)$ in (7.1)–(7.2) belongs to a certain parametric family, we establish the null hypothesis

$$\mathcal{H}_{0m} \colon m_1 \in \{m_1(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}.$$

We propose a test based on the integrated regression function $I(x) = \mathbb{E}\left[Y \mathbb{1}_{\{X \leqslant x\}}\right] = \int_{-\infty}^{x} m(y)\,\mathrm{d}F(y)$, where F is the marginal distribution function of X. An empirical estimator of the integrated regression function for the model (7.1)–(7.2) is given by

$$I_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{X_{t_i} \leqslant x\}} \frac{X_{t_{i+1}} - X_{t_i}}{\Delta},$$

and an estimator under the null hypothesis $\mathcal{H}_{0m}$ is

$$I_{0n}(x) = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{\{X_{t_i} \leqslant x\}} m_1(X_{t_i}, \hat{\boldsymbol{\theta}}).$$

The goodness-of-fit test compares the estimated integrated regression function $I_n(\cdot)$ with the estimation obtained under the null hypothesis, that is, $I_{0n}(\cdot)$. Therefore, the test is defined by the process

$$R_n^m(x) = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} \mathbb{1}_{\{X_{t_i} \leqslant x\}} \left(\frac{X_{t_{i+1}} - X_{t_i}}{\Delta} - m_1(X_{t_i}, \hat{\boldsymbol{\theta}}) \right),$$

with $x \in \mathbb{R}$ and where $\hat{\boldsymbol{\theta}}$ is a root-n consistent estimator of the true parameter vector $\boldsymbol{\theta}$. As in the previous test, a continuous functional $\Psi(\cdot)$ can be considered to define the statistic $T_n = \Psi(R_n^m)$, such as the Kolmogorov-Smirnov

$$T_n^{KS} := \sup_x \left| R_n^m(x) \right|,$$

and Cramér-von Mises

$$T_n^{CvM} := \int_{\mathbb{R}} R_n^m(x)^2 F_n(\mathrm{d}x),$$

criteria, with F_n the empirical distribution function of $\{X_{t_i}\}_{i=0}^{n-1}$. The null hypothesis is rejected if $T_n > c_{1-\alpha}$. Again, the critical value $c_{1-\alpha}$ can be approximated by its bootstrap counterpart $c_{1-\alpha}^*$, such that $\mathbb{P}^*\left(\Psi(R_n^{m*}) > c_{1-\alpha}^* \mid \mathcal{H}_{0m}\right) = \alpha$ with

$$R_n^{m*}(x) = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} \mathbb{1}_{\{X_{t_i} \leqslant x\}} \left(u_{t_i}^* - m_1(X_{t_i}, \hat{\boldsymbol{\theta}}^*) \right),$$

where $u_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ and $u_{t_i}^* = m_1(X_{t_i}, \hat{\boldsymbol{\theta}}) + e_{t_i}^*$ is obtained sampling the innovations e_{t_i} with the algorithm introduced in Section 7.2.3, and the bootstrap estimator $\hat{\boldsymbol{\theta}}^*$ is calculated from the bootstrap sample $\{(X_{t_i}^*, \sigma_{t_i}^{*2})\}$. As explained in the previous test, the empirical p-value is the proportion of the B bootstrap replicates T_n^{*j} exceeding T_n, $\sharp\{T_n^{*j} > T_n\}/B$.

7.2.3 Bootstrap resampling procedure

The estimation of $\boldsymbol{\theta}$ and the bootstrap sample $\{(X^*_{t_i}, \sigma^{*2}_{t_i})\}$ can be obtained with the Kalman filter algorithm (Shumway and Stoffer, 2000). Given the filtering equations (4.13)–(4.14) introduced in Chapter 4, the bootstrap resample algorithm for model (7.1)–(7.2) can be implemented as follows. First, calculate the residuals from the linear regression

$$\begin{aligned} e_{t_i} &= \frac{X_{t_i} - m_1(X_{t_i}, \boldsymbol{\theta})}{\sqrt{\Delta}} \\ &= \sigma_{t_i} \nu_1(X_{t_i}, \boldsymbol{\theta}) \varepsilon_{1,t_i}, \end{aligned}$$

second, define $y_{t_i} = \log e^2_{t_i}$ and $h_{t_i} = \log \sigma^2_{t_i}$ to linearize the space equation, therefore we have the following space-state equations,

$$\begin{aligned} y_{t_i} &= h_{t_i} + \log \nu_1(X_{t_i}, \boldsymbol{\theta})^2 + v_{t_i}, & v_{t_i} &\sim \log \chi^2_1, \\ h_{t_i} &= m_2(h_{t_i}, \boldsymbol{\theta}) + w_{t_i}, & w_{t_i} &\sim \mathcal{N}(0, \sigma^2_w), \end{aligned}$$

with $w_{t_i} = \nu_2(h_{t_i}, \boldsymbol{\theta})\sqrt{\Delta}\, \varepsilon_{2,t_i}$, $\varepsilon_{2,t_i} \sim \mathcal{N}(0,1)$ and $\sigma^2_w = \Delta \nu_2(h_{t_i}, \boldsymbol{\theta})^2$. Given the filtering equations (4.13)–(4.14) we have

$$y_{t_i} = h_{t_i|t_{i-1}} + \log \nu_1(X_{t_{i-1}}, \boldsymbol{\theta})^2 + \pi_{0,t_i} \epsilon_{0,t_i} + \pi_{1,t_i}(\epsilon_{1,t_i} + \mu_1).$$

Let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimator of the true parameter $\boldsymbol{\theta}$,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\Theta} \ln \mathcal{L}_Y(\boldsymbol{\theta}),$$

by means of the Kalman filter algorithm, and ϵ_{j,t_i}, with $\mathbb{V}\mathrm{ar}\left[\epsilon_{j,t_i}\right] = \hat{\Sigma}_{j,t_i}$, the innovations and the innovations variance obtained by running the filter under $\hat{\boldsymbol{\theta}}$. To implement the bootstrap algorithm we consider the standardized innovations

$$\tilde{\epsilon}_{j,t_i} = \hat{\Sigma}^{-1/2}_{j,t_i} \epsilon_{j,t_i}, \qquad \text{for } j = 0, 1,$$

and sample with replacement from $\{\tilde{\epsilon}_{j,t_i}\}_{i=0}^{n-1}$ to obtain the bootstrap sample of standardized innovations $\{\tilde{\epsilon}^*_{j,t_i}\}_{i=0}^{n-1}$. We then generate the bootstrap sample $\{(y^*_{t_i}, h^*_{t_i})\}_{i=0}^{n-1}$ with

$$\begin{aligned} h^*_{t_{i+1}} &= m_2(h_{t_i|t_{i-1}}, \hat{\boldsymbol{\theta}}) + \hat{\pi}_{0,t_i} K_{0,t_i} \hat{\Sigma}^{1/2}_{0,t_i} \tilde{\epsilon}^*_{0,t_i} + \hat{\pi}_{1,t_i} K_{1,t_i} \hat{\Sigma}^{1/2}_{1,t_i} \tilde{\epsilon}^*_{1,t_i}, \\ y^*_{t_i} &= h_{t_i|t_{i-1}} + \log \nu_1(X_{t_i}, \hat{\boldsymbol{\theta}})^2 + \hat{\pi}_{1,t_i} \hat{\mu}_1 + \hat{\pi}_{0,t_i} \hat{\Sigma}^{1/2}_{0,t_i} \tilde{\epsilon}^*_{0,t_i} + \hat{\pi}_{1,t_i} \hat{\Sigma}^{1/2}_{1,t_i} \tilde{\epsilon}^*_{1,t_i}, \end{aligned} \tag{7.3}$$

where X_{t_i} remains fixed and the bootstrapped dataset is given by $\exp(y^*_{t_i}) = e^{*2}_t$. Algorithm 4 summarizes the bootstrap procedure to approximate the critical value $c_{1-\alpha}$.

Algorithm 4 (Bootstrap resampling procedure)**.** The critical value $c_{1-\alpha}$ can be approximated with the following bootstrap procedure:

1. Obtain the maximum likelihood estimator of the true parameter $\boldsymbol{\theta}$ by means of the Kalman filter algorithm introduced in (4.15), that is, $\hat{\boldsymbol{\theta}} = \arg\max_{\Theta} \ln \mathcal{L}_Y(\boldsymbol{\theta})$.
2. Construct the bootstrap sample $\{(y^*_{t_i}, \sigma^{*2}_{t_i})\}_{i=0}^{n-1}$ as in (7.3).
3. Estimate the parameter vector $\hat{\boldsymbol{\theta}}^*$ from the bootstrap resample $\{(y^*_{t_i}, \sigma^{*2}_{t_i})\}_{i=0}^{n-1}$ obtained in Step 2.
4. Compute the bootstrap version of the process $R^{\nu *}_n(x, y)$ or $R^{m*}_n(x)$, for $x, y \in \mathbb{R}$.
5. Determine $U^*_n = \Psi\big(R^{\nu *}_n(x, y)\big)$ or $T^*_n = \Psi\big(R^{m*}_n(x)\big)$.
6. Repeat B times the previous Steps 2–5 to obtain $j = 1, \ldots, B$, bootstrap replicates U^{*j}_n or T^{*j}_n.
7. Approximate the critical value $\hat{c}^*_{1-\alpha} = U^{*\lceil B(1-\alpha)\rceil}_n$ or $\hat{c}^*_{1-\alpha} = T^{*\lceil B(1-\alpha)\rceil}_n$.

7.3 Simulation study

In this Section, a simulation study to illustrate the finite sample properties of the tests was conducted under different settings, for both the drift and volatility goodness-of-fit tests. As to our knowledge, the tests available in the literature for continuous-time stochastic volatility models do not test the same null hypothesis as our test, that is, the parametric form of the diffusion function, we are not including a comparison with other procedures.

7.3.1 Drift test

The performance of the goodness-of-fit test for the parametric form of the drift function introduced in Section 7.2.2, regarding size and power, is illustrated by means of a simulation study. We test the null hypothesis that the drift function $m_1(\cdot)$ of the SDE in (7.1)–(7.2) belongs to a certain parametric family,

$$\mathcal{H}_{0m} \colon m_1 \in \{m_1(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}.$$

We consider the CKLS-OU model as the null hypothesis and to evaluate the power of the test we use a series of alternative models indexed by the parameter ρ of the non-linear function $\rho_m(X_t) = \rho(1 - X_t^\rho)$,

$$\begin{aligned} \mathrm{d}X_t &= \left[\kappa(\mu - X_t) + \rho(1 - X_t^\rho)\right] \mathrm{d}t + \sigma_t X_t^\gamma \, \mathrm{d}W_{1,t}, \\ \mathrm{d}\ln\sigma_t^2 &= (\theta_0 - \theta_1 \log\sigma_t^2) \, \mathrm{d}t + \xi \, \mathrm{d}W_{2,t}, \end{aligned}$$

with $\rho \in \{0, 0.07, 0.09, 0.10, 0.125, 0.15\}$, where $\rho = 0$ under the null hypothesis. The parameter vector for the data generating process (DGP) is given by $\boldsymbol{\theta} = (\mu, \kappa, \gamma, \theta_0, \theta_1, \xi)^\top = (0.06, 0.6, 1.5, -0.7, 0.1, 0.4)^\top$. The process was generated with weekly frequency ($\Delta = 1/52$) for different sample sizes $n \in \{500, 1000, 1500, 2000\}$ and the first thousand observations were discarded as a burn-in period. The rate of rejection ($\hat{\alpha}$) is calculated based on 1000 Monte Carlo replicates and $B = 1000$ bootstrap resamples (see Algorithm 4) for the Kolmogorov-Smirnov ($\hat{\alpha}_{\text{KS}}$) and Cramér-von Mises ($\hat{\alpha}_{\text{CvM}}$) criteria.

Table 7.1 shows the size (first row) and power of the goodness-of-fit test for the drift function $m_1(\cdot)$ for the null hypothesis that the drift function follows the CKLS-OU parametric form, that is, $\mathcal{H}_{0m}\colon m_1(X_t, \boldsymbol{\theta}) = \kappa(\mu - X_t)$, with $\alpha = 0.05$. Regarding the size, the tests are well calibrated as both rejection rates are very close to the nominal level α. The behavior of the power, on the other hand, shows an increase with the sample size, as expected, and the higher the value of ρ, the further we depart from the null hypothesis, obtaining higher rejection rates.

	$\hat{\alpha}_{\text{KS}}$				$\hat{\alpha}_{\text{CvM}}$			
DGP	500	1000	1500	2000	500	1000	1500	2000
$\rho = 0$	0.047	0.043	0.045	0.044	0.054	0.051	0.060	0.056
$\rho = 0.07$	0.102	0.186	0.203	0.301	0.169	0.267	0.298	0.356
$\rho = 0.09$	0.301	0.305	0.356	0.456	0.314	0.364	0.448	0.508
$\rho = 0.10$	0.365	0.456	0.481	0.526	0.441	0.528	0.560	0.606
$\rho = 0.125$	0.523	0.618	0.669	0.703	0.531	0.636	0.682	0.747
$\rho = 0.15$	0.790	0.848	0.869	0.901	0.785	0.864	0.907	0.923

Table 7.1: Size ($\rho = 0$) and power simulation for the CKLS-OU model drift test, with $\alpha = 0.05$, for the null hypothesis $\mathcal{H}_0\colon m_1(X_t, \boldsymbol{\theta}) = \kappa(\mu - X_t)$, under different alternative scenarios $m_1(X_t, \boldsymbol{\theta}) = \left(\kappa(\mu - X_t) + \rho(1 - X_t^\rho)\right)$, for $\rho \in \{0, 0.07, 0.09, 0.10, 0.125, 0.15\}$.

7.3.2 Volatility test

The study of the finite sample properties of the goodness-of-fit test for the parametric form of the volatility function introduced in Section 7.2.1 is accomplish with a simulations study, testing both simple and composite null hypotheses. We test the null hypothesis that the diffusion function $\nu_1(\cdot)$ of the continuous-time model in (7.1)–(7.2) belongs to a certain parametric family, that is,

$$\mathcal{H}_{0\nu}\colon \nu_1 \in \{\nu_1(\cdot, \boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\},$$

for the composite hypothesis. We consider three different models under the null hypothesis, which are described in Table 7.2, and, to assess the performance of the power of the test, we take the CKLS-OU model and create a series of alternative scenarios by adding a non-linear function to the diffusion function $\nu_1(\cdot)$

$$\begin{aligned} \mathrm{d}X_t &= \kappa(\mu - X_t)\,\mathrm{d}t + \left[\sigma_t X_t^{\gamma} + \rho(1 - X_t^{\rho})\right]\mathrm{d}W_{1,t}, \\ \mathrm{d}\ln\sigma_t^2 &= (\theta_0 - \theta_1 \log\sigma_t^2)\,\mathrm{d}t + \xi\,\mathrm{d}W_{2,t}, \end{aligned}$$

with $\rho \in \{0.007, 0.01, 0.02, 0.04\}$, where $\rho = 0$ under the null hypothesis (scenario *CKLS-OU* in Table 7.2). The processes were generated with weekly frequency ($\Delta = 1/52$) for sample sizes $n \in \{500, 1000, 1500, 2000\}$ and the first thousand observations were discarded. The rate of rejection ($\hat{\alpha}$) is calculated based on 1000 Monte Carlo replicates and $B = 1000$ bootstrap resamples (see Algorithm 4) for the Kolmogorov-Smirnov ($\hat{\alpha}_{\text{KS}}$) and Cramér-von Mises ($\hat{\alpha}_{\text{CvM}}$) criteria. The size and power of the test for the simple null hypothesis will be evaluated using the CKLS-OU model ($\nu_1(X_t, \boldsymbol{\theta}) = X_t^{\gamma}$), testing the null hypothesis $\mathcal{H}_{0\nu}\colon \nu_1(X_t, \boldsymbol{\theta}) = X_t^{1.5}$ for a set of values $\gamma \in \{1.5, 1.25, 1.0\}$.

Scenario	Model	Parameters
OU-OU	$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_{1,t}$ $\mathrm{d}\ln\sigma_t^2 = (\theta_0 - \theta_1\log\sigma_t^2)\,\mathrm{d}t + \xi\,\mathrm{d}W_{2,t}$	$\boldsymbol{\theta} = (\mu, \kappa, \gamma, \theta_0, \theta_1, \xi)^{\top}$ $= (0.06, 0.6, -0.7, 0.1, 0.4)^{\top}$
CKLS-null	$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma_t X_t^{\gamma}\,\mathrm{d}W_{1,t}$ $\mathrm{d}\ln\sigma_t^2 = \xi\,\mathrm{d}W_{2,t}$	$\boldsymbol{\theta} = (\mu, \kappa, \gamma, \xi)^{\top}$ $= (0.06, 0.6, 1.5, 0.4)^{\top}$
CKLS-OU	$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma_t X_t^{\gamma}\,\mathrm{d}W_{1,t}$ $\mathrm{d}\ln\sigma_t^2 = (\theta_0 - \theta_1\log\sigma_t^2)\,\mathrm{d}t + \xi\,\mathrm{d}W_{2,t}$	$\boldsymbol{\theta} = (\mu, \kappa, \gamma, \theta_0, \theta_1, \xi)^{\top}$ $= (0.06, 0.6, 1.5, -0.7, 0.1, 0.4)^{\top}$

Table 7.2: Scenarios under the composite null hypothesis.

Table 7.3 shows the empirical size and power for simple and composite hypotheses, the

later under the null hypotheses for the scenarios in Table 7.2. Regarding the simple hypothesis test, the size (first row) is close to the nominal level $\alpha = 0.05$, with a slight over rejection for the smallest sample size scenario, and the power increases with sample size and shows that the test is capable of discriminating between models with different values of γ. Focusing on the composite null hypothesis, the estimated sizes (first three rows) remain close the true nominal level, although the smallest sample sizes show some small deviations, and the power increases both with the sample size and the value of ρ, which controls the level of noise added to the diffusion function.

		$\hat{\alpha}_{\text{KS}}$				$\hat{\alpha}_{\text{CvM}}$			
$\mathcal{H}_0$	DGP	500	1000	1500	2000	500	1000	1500	2000
Simple hypothesis									
$\gamma = 1.50$	CKLS-OU	**0.089**	0.046	0.041	0.049	0.040	0.044	0.062	0.052
$\gamma = 1.50$	$\gamma = 1.25$	0.454	0.613	0.708	0.808	0.462	0.602	0.767	0.818
$\gamma = 1.50$	$\gamma = 1.00$	0.734	0.884	0.957	0.986	0.649	0.873	0.974	0.988
Composite hypothesis									
$\nu_1(X_t, \boldsymbol{\theta}) = 1$	OU-OU	0.037	0.046	0.046	0.058	0.045	0.051	0.032	0.064
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	CKLS-null	**0.033**	0.048	0.038	0.057	**0.035**	0.051	0.034	0.063
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	CKLS-OU	0.040	0.047	0.041	0.047	0.044	0.053	0.034	0.041
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	$\rho = 0.007$	0.125	0.177	0.206	0.331	0.099	0.148	0.190	0.303
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	$\rho = 0.01$	0.364	0.468	0.641	0.657	0.304	0.423	0.578	0.619
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	$\rho = 0.02$	0.490	0.644	0.710	0.796	0.415	0.581	0.674	0.701
$\nu_1(X_t, \boldsymbol{\theta}) = X_t^\gamma$	$\rho = 0.04$	0.527	0.766	0.870	0.951	0.491	0.620	0.714	0.819

Table 7.3: Size and power simulation for the volatility function tests, with $\alpha = 0.05$, considering simple and composite null hypotheses. Under $\mathcal{H}_0$, rejection rates are boldfaced if they lie outside a 95%-confidence interval for the nominal level $\alpha = 0.05$.

7.4 Real data applications

In this section, we consider the Euribor (Euro Interbank Offered Rate) interest rate series corresponding to four maturities (three, six, nine and twelve months), from Chapters 3 and 4 (see Figure 3.4). The four datasets expand from October 15$^{\text{th}}$ 2001 to December 30$^{\text{th}}$ 2005 (sample size of $n = 1\,077$). We fit the CKLS with stochastic volatility in 7.2 (CKLS-OU), as different models can be nested within this unrestricted model, and test the goodness of fit of the model in terms of the parametric form of the volatility function. See Section 4.4 in Chapter 4 for a discussion on the estimation of the two-factor CKLS model.

Regarding the goodness-of-fit test, Table 7.4 shows the p-values for the parametric form of the volatility function, that is, the null hypothesis $\mathcal{H}_{0\nu}\colon \nu_1(X_t, \boldsymbol{\theta}) = X_t^{\gamma}$, both for the Kolmogorov-Smirnov and Cramér-von Mises statistics, which exhibit minor discrepancies. In Chapter 3 the same datasets were used to fit a CKLS model with deterministic volatility function and the null hypothesis for the parametric form of the diffusion function was strongly rejected for the four maturities. However, when considering a more flexible diffusion function with stochastic volatility, we do not reject the null hypothesis, suggesting that a model that incorporates stochastic volatility may adequately explain the dynamics of the series. Rejecting a deterministic volatility function in favor of a stochastic function indicates that the volatility evolution is not exclusively tied to the level of the short rate, but rather the process is governed by dynamic factors. This was discussed in the financial literature (see Aït-Sahalia, 1996a; Brenner et al., 1996; Andersen and Lund, 1997a; Koedijk et al., 1997; Gallant and Tauchen, 1998), where less restrictive models were proposed, including the two-factor model or even multi-factor models of the short rate (Andersen and Lund, 1997b).

Maturity	3 months	6 months	9 months	12 months
p-value Kolmogorov-Smirnov	0.464	0.266	0.319	0.189
p-value Cramér-von Mises	0.550	0.215	0.422	0.113

Table 7.4: p-values for the goodness-of-fit test for the CKLS-OU parametric form of the diffusion function.

7.5 Conclusions

We proposed goodness-of-fit tests for the parametric form of the drift and volatility functions for two-factor continuous-time stochastic volatility models, based on a residual marked empirical process. The tests showed great power through several alternative hypotheses and was well calibrated under null hypotheses and its implementation, regarding the computation of the test statistic and the bootstrap resampling scheme, is quite straightforward. The application to real data demonstrated that the incorporation of stochastic volatility to diffusion models does capture the features of interest rate time series, unlike deterministic volatility functions, suggesting that the volatility depends on an additional factor that varies independently of the short rate level.

Chapter 8

Specification test for diffusion models as functional time series

High-frequency financial data can be collected as a sequence of curves over time; for example, as intra-day prices. The Functional Data Analysis framework provides a suitable tool to extract the information contained in the shape of the daily paths, often unavailable from classical statistical methods. In this chapter, a novel goodness-of-fit test for autoregressive Hilbertian (ARH) models, with unknown and general order, is proposed. The test imposes just the Hilbert-Schmidt assumption on the functional form of the autocorrelation operator, and the test statistic is formulated in terms of a Cramér-von Mises norm. A wild bootstrap resampling procedure is used for calibration, such that the finite sample behavior of the test, regarding power and size, is checked via a simulation study. Furthermore, a new specification test for diffusion models, such as Ornstein-Uhlenbeck processes, is also provided and illustrated with an application to intra-day currency exchange rates. In particular, a two-stage methodology is proffered: firstly, we test if the functional samples and their past values are related via an ARH(1) model; secondly, under linearity, a functional F-test is performed.

The contents of this chapter are collected in Álvarez-Liébana et al. (2022).

Contents

8.1 Introduction

Motivated by the recent availability of high-frequency data in finance, in this chapter a twofold contribution in the field of temporally correlated functional data and diffusion processes is provided. On the one hand, up to our knowledge, no proposals of goodness-of-fit tests for autoregressive Hilbertian (ARH) processes, for unknown order and against unspecified alternatives, can be found. On the other hand, few papers were devoted to specification tests for diffusion models in a functional data analysis setup. The main challenge and novelty of this work lies in covering both gaps. Regarding the former, we develop a composite null hypothesis test for functional time series in terms of a Cramér-von Mises statistic, such that no assumptions concerning alternatives are required. The latter is addressed by a characterization of Ornstein-Uhlenbeck processes as ARH models, providing a novel specification test in a high-dimensional setup, which could be extended to several diffusion processes, as long as they can be characterized as temporally correlated functional data.

Modeling the relationship between functional random variables is one of the main topics in this context, where the Functional Linear Model with Functional Response (FLMFR) $\mathcal{Y} = m(\mathcal{X}) + \mathcal{E}$, with m a linear operator and $\mathcal{E}$ a functional error, is likely the best known parametric model. Within a Hilbertian framework, the operator m is typically assumed to be a Hilbert-Schmidt integral operator between L^2 spaces. Several authors have contributed in the goodness-of-fit for regression models field, most of them keeping integrated regression-based methodologies proposed by Stute (1997). This framework was extended to functional

setups in García-Portugués et al. (2014), Cuesta-Albertos et al. (2019) and García-Portugués et al. (2021). As a particular case of linear models, functional time series were widely studied, mainly focusing on ARH processes, in which functional regressors are given by their own past values. Early results on the parametric moment-based estimation were proposed by Bosq (2000), on the estimation of ARH(1) models,

$$\mathcal{X}_n(\cdot) = \Gamma\left(\mathcal{X}_{n-1}\right)(\cdot) + \mathcal{E}_n(\cdot),$$

with $n \in \mathbb{Z}$. Recently, Álvarez-Liébana et al. (2017) proposed a plug-in estimator in terms of the eigenelements of the autocovariance operator involved, and Ruiz-Medina and Álvarez-Liébana (2019a) exploits the Hilbert space structure with an extension to the Banach space context.

Notwithstanding the growing interest in functional time series, up to our knowledge, no proposals of goodness-of-fit tests in the ARH(z) setup, with unspecified alternatives, can be found in the literature. This lack of approaches probably reflects the fact that, besides the goodness-of-fit test recently proposed in García-Portugués et al. (2021), there are no proposals of tests in the FLMFR setup neither, against unspecified alternatives (see Kokoszka et al., 2008 and Patilea et al., 2016 for testing no effects hypothesis). One of the most crucial aspects working with functional time series is checking whether just a single model can be used, not rupturing the underlying stochastic structure. As early contributions, Laukaitis and Rackauskas (2002) analyzed the functional residuals for change-point detection, whereas Berkes et al. (2009) focused on testing the assumption of a common functional mean of i.i.d. samples. We also refer to Gabrys and Kokoszka (2007), for testing i.i.d. functional random variables. Horváth et al. (2010) proposed a methodology to check if the autocorrelation operator involved remains unchanged with time. Another aspect to ensure meaningful inference relies in testing the assumption of stationarity, within a functional data framework. With this purpose, KPSS tests were formalized in Horváth et al. (2014).

In this work, we provide a goodness-of-fit test for ARH(z) processes, under the null hypothesis

$$\mathcal{H}_0\colon\ \mathcal{X}_n \text{ and } (\mathcal{X}_{n-1}, \ldots, \mathcal{X}_{n-z}) \text{ are linearly related,}$$

i.e.,

$$\mathcal{X}_n(t) = \sum_{r=1}^{z} \Gamma_r(\mathcal{X}_{n-r})(t) + \mathcal{E}_n(t),$$

against an unspecified alternative hypothesis, where $\{\Gamma_r\}_{r=1}^{z}$, with $z \geqslant 1$, is a set of linear

operators Γ_ρ given by

$$\Gamma_\rho\left(\mathcal{X}\right)(t) = \int_0^h \rho(s,t)\mathcal{X}(s)\,\mathrm{d}s,$$

with

$$\int_0^h \int_0^h \rho^2(s,t)\,\mathrm{d}s\,\mathrm{d}t < \infty.$$

The methodology here proposed is based on reinterpreting an ARH(z) process as a particular FLMFR and the characterization of $\mathcal{H}_0$ in terms of the integral regression operator derived, projected into finite-dimensional functional directions, in keeping with the work by García-Portugués et al. (2021). The deviation of data from $\mathcal{H}_0$ is measured by a Cramér-von Mises norm and the test is calibrated via wild bootstrap. The novelty of contribution is twofold: *(i)* a new goodness-of-fit test for ARH(z) processes, given a positive integer order z; *(ii)* a sequential procedure to determine the order of an ARH model. As the former contribution has no competitors, we compare the latter against the multistage order detection procedure in Kokoszka and Reimherr (2013). Note that their iterative methodology was developed just considering ARH(w) alternatives, with $w > z$, in contrast with our unspecified alternative hypothesis.

Due to the key role of diffusion processes in finance, a framework for model specification of diffusion models from a functional perspective was also developed. In a classical context (see Chapter 6), diffusion models for financial data are commonly used at daily, weekly or monthly frequency. However, high-frequency data constitutes time series at a very fine resolution, and using daily data would imply discarding large fractions of observations. It is in this high-frequency scenario where dealing with diffusion models using a functional data approach allows to model the realized daily trajectories, analyzing the patterns of the curves across the days. Since Vasicek (1977) model was used to capture the dynamics of interest rates, the Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930) is one of the main diffusion models. In this chapter, as an alternative for univariate and multivariate settings, a novel two-stage specification test for OU processes as a functional process is proposed: in the first step, whether the underlying stochastic differential equation can be characterized via ARH(1) process is verified, based on an ARH(1) characterization of OU processes; secondly, under linearity, a functional F-test is implemented.

The rest of this chapter is organized as follows. In Section 8.2, a brief introduction to functional data, FLMFR and ARH(z) models, is provided. A goodness-of-fit test for the referred autoregressive models is detailed in Section 8.3, and a simulation study is undertaken in Section 8.3.4, comparing it with Kokoszka and Reimherr (2013) proposal. Section 8.4 proposes a new specification test for diffusion models. The performance of the specification

test is also depicted by means of a wide simulation study, while in Section 8.5, a real data application to daily currency exchange rates curves is addressed. Section 8.6 concludes the chapter and theoretical details are relegated to Appendix C.

8.2 Background on FLMFR and functional time series

Some of the main notions on FLMFR and functional time series are introduced in this section.

8.2.1 Functional basis

Whilst Hilbert spaces are the common option, it is well worth canvassing the alternatives which could be adopted. The most general context is given by a metric space endowed with a distance. In contrast with metric spaces, useful when no information about curves are available but rather abstract since a norm cannot be guaranteed, a Banach space $\mathbb{B}, \|\cdot\|_{\mathbb{B}}$ may be adopted in place, say $\mathcal{C}([0,1])$. Unless otherwise explicitly mentioned, we consider separable Hilbert spaces, with inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}}$. This common choice is not arbitrary since, under separability, the existence of a countable functional basis is guaranteed. In what follows, $\{\Psi_j\}_{j=1}^{\infty}$ and $\{\Phi_k\}_{k=1}^{\infty}$ are orthonormal functional bases. From separability, any $\mathcal{X} \in \mathbb{H}_1$ and $\mathcal{Y} \in \mathbb{H}_2$ can be represented as

$$\mathcal{X} = \sum_{j=1}^{\infty} x_j \Psi_j \qquad \text{and} \qquad \mathcal{Y} = \sum_{k=1}^{\infty} y_k \Phi_k,$$

with $x_j = \langle \mathcal{X}, \Psi_j \rangle_{\mathbb{H}_1}$ and $y_k = \langle \mathcal{Y}, \Phi_k \rangle_{\mathbb{H}_2}$, for each $j, k \geqslant 1$.

Since sparsity effects usually appear, dimension reduction is essential in this context. This sparsity can be understood in a wide sense, such as referring to the sparsity of the model (Aneiros and Vieu, 2014), the sparsity of data (Vieu, 2018) or related to the discretization grid (Yao et al., 2005). We can further distinguish among fixed bases (e.g., B-splines or Fourier bases), flexible but usually require a larger number of elements, and data-driven functional bases. More parsimonious representations can be achieved with the latter ones, performing a dimension reduction explicitly based on the covariance structure, being the most popular choice the Functional Principal Components (FPC) bases, given as the eigenfunctions of the empirical autocovariance operator

$$C_n^{\mathcal{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i \otimes \mathcal{X}_i.$$

Henceforth, for given finite cut-off levels $p, q \geqslant 1$, we define the (p, q)-truncated bases expansions as

$$\mathcal{X}^{(p)} = \sum_{j=1}^{p} x_j \Psi_j,$$
$$\mathcal{Y}^{(q)} = \sum_{k=1}^{q} y_k \Phi_k,$$

with $(x_1, \ldots, x_p) \in \mathbb{R}^p$ and $(y_1, \ldots, y_q) \in \mathbb{R}^q$, such that

$$C_n^{\mathcal{X}}(\Psi_j) = \lambda_j^{\Psi} \Psi_j,$$
$$C_n^{\mathcal{Y}}(\Phi_k) = \lambda_k^{\Phi} \Phi_k,$$

for each $j = 1, \ldots, p$ and $k = 1, \ldots, q$.

8.2.2 The FLMFR

In this chapter, we assume that $\mathcal{X} \in \mathbb{H}_1 = L^2([a, b])$ and $\mathcal{Y} \in \mathbb{H}_2 = L^2([c, d])$ are centered functional random variables, such that

$$\langle f, g \rangle_{L^2([a,b])} = \int_a^b f(t) g(t) \, \mathrm{d}t.$$

We will characterize a functional time series as a particular linear model, and so, the following FLMFR setting will be introduced:

$$\mathcal{Y} = m_\beta(\mathcal{X}) + \mathcal{E}, \qquad \text{with } \mathbb{E}[\mathcal{E} \mid \mathcal{X}] = 0, \tag{8.1}$$

where $m_\beta(x) = \mathbb{E}[\mathcal{Y} \mid \mathcal{X} = x]$ is a Hilbert-Schmidt integral operator, that is, m_β admits an integral representation given by a bivariate and square-integrable kernel $\beta \in \mathbb{H}_1 \otimes \mathbb{H}_2$. In the aftermath of the compactness, β can be decomposed as follows,

$$m_\beta(\mathcal{X})(\cdot) = \int_a^b \beta(s, \cdot) \mathcal{X}(s) \, \mathrm{d}s,$$
$$\beta = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \beta_{jk} (\Psi_j \otimes \Phi_k),$$

with $\int\int \beta^2(s, t) < \infty$ and where $\beta_{jk} = \langle \beta, \Psi_j \otimes \Phi_k \rangle_{\mathbb{H}_1 \otimes \mathbb{H}_2}$, for each $j, k \geqslant 1$. Adopting a (p, q)-truncated component-wise orthonormal expansion projected into $\{\Psi_j\}_{j=1}^p$, $\{\Phi_k\}_{k=1}^q$

and $\{\Psi_j \otimes \Phi_k\}_{j,k=1}^{p,q}$, we have

$$\begin{aligned} \mathcal{Y}^{(q)} &= \sum_{k=1}^{q} y_k \Phi_k, \\ y_k &= \sum_{j=1}^{p}\sum_{\ell=1}^{p} b_{\ell,k} x_j \langle \Psi_j, \Psi_\ell \rangle_{\mathbb{H}_1} + e_k, \qquad k = 1, \ldots, q, \\ \mathcal{E}^{(q)} &= \sum_{k=1}^{q} e_k \Phi_k, \end{aligned} \tag{8.2}$$

where $\langle \Psi_j, \Psi_\ell \rangle_{\mathbb{H}_1} = \delta_{j\ell}$, for each $j, \ell \geqslant 1$. Now, given a centered sample $\{(\mathcal{X}_i, \mathcal{Y}_i)\}_{i=1}^{n}$, equations (8.1)–(8.2) may be expressed as

$$\begin{aligned} \mathbf{Y}_q &= \mathbf{X}_p \mathbf{B}_{p,q} + \mathbf{E}_q, \\ \mathcal{Y}_i &= \langle\langle \mathcal{X}_i, \beta \rangle\rangle + \mathcal{E}_i, \\ \langle\langle \mathcal{X}, \beta \rangle\rangle(t) &\coloneqq \langle \mathcal{X}, \beta(\cdot, t) \rangle_{\mathbb{H}_1}, \end{aligned} \tag{8.3}$$

where $\mathbf{Y}_q$ and $\mathbf{E}_q$ are the $n \times q$ matrices with the coefficients of $\{\mathcal{Y}_i\}_{i=1}^{n}$ and $\{\mathcal{E}_i\}_{i=1}^{n}$ on $\{\Phi_k\}_{k=1}^{q}$, respectively, $\mathbf{X}_p$ is the $n \times p$ matrix with the coefficients of $\{\mathcal{X}_i\}_{i=1}^{n}$ on $\{\Psi_j\}_{j=1}^{p}$ and $\mathbf{B}_{p,q}$ is the $p \times q$ matrix with the coefficients (to be estimated) of β on $\{\Psi_j \otimes \Phi_k\}_{j,k=1}^{p,q}$.

From now on, we consider the hybrid linearly-constrained FPC Regression (FPCR-L1S) estimator recently formulated in García-Portugués et al. (2021), defined as follows,

$$\begin{aligned} \hat{\mathbf{Y}}_q &= \widetilde{\mathbf{X}}_{\widetilde{p}} \hat{\mathbf{B}}_{\widetilde{p},q}^{(\lambda),C} = \mathbf{H}_C^{(\lambda)} \mathbf{Y}_q = \left[\widetilde{\mathbf{X}}_{\widetilde{p}} \left(\widetilde{\mathbf{X}}_{\widetilde{p}}^\top \widetilde{\mathbf{X}}_{\widetilde{p}} \right)^{-1} \widetilde{\mathbf{X}}_{\widetilde{p}}^\top \right] \mathbf{Y}_q, \\ \hat{\mathbf{B}}_{\widetilde{p},q}^{(\lambda),C} &= \left(\widetilde{\mathbf{X}}_{\widetilde{p}}^\top \widetilde{\mathbf{X}}_{\widetilde{p}} \right)^{-1} \widetilde{\mathbf{X}}_{\widetilde{p}}^\top \mathbf{Y}_q. \end{aligned} \tag{8.4}$$

The estimated $\hat{\mathbf{Y}}_q$ in (8.4) are computed in four steps:

(i) Orthonormal bases $\{\Psi_j\}_{j=1}^{p}$ and $\{\Phi_k\}_{k=1}^{q}$ are defined as the empirical FPC of $\mathcal{X}$ and $\mathcal{Y}$, respectively (in what follows, $\{\Psi_j\}_{j=1}^{p}$ and $\{\Phi_k\}_{k=1}^{q}$ are the eigenfunctions of the empirical covariance operators $C_n^{\mathcal{X}}$ and $C_n^{\mathcal{Y}}$, respectively—see Section 8.2.1—);

(ii) Initial values (p, q) are fixed with regard to a certain proportion of Explained Variance EV_p and EV_q (e.g., $\mathrm{EV}_p = \mathrm{EV}_q \geqslant 0.99$), where $\mathrm{EV}_p = \sum_{j=1}^{p} \lambda_j^{\Psi}$ and $\mathrm{EV}_q = \sum_{k=1}^{q} \lambda_k^{\Phi}$, being $\{\lambda_j^{\Psi}\}_{j=1}^{p}$ and $\{\lambda_k^{\Phi}\}_{k=1}^{q}$ the associated eigenvalues;

(iii) Variable selection is implemented based on a LASSO (L1) regularization, considering the no null rows of $\hat{\mathbf{B}}_{p,q}^{(\lambda)} = \underset{\mathbf{B}_{p,q}}{\arg\min} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left\| (\mathbf{Y}_q)_i - (\mathbf{X}_p \mathbf{B}_{p,q})_i \right\|^2 + \lambda \sum_{j=1}^{p} \left\| (\mathbf{B}_{p,q})_j \right\|_2 \right\}$,

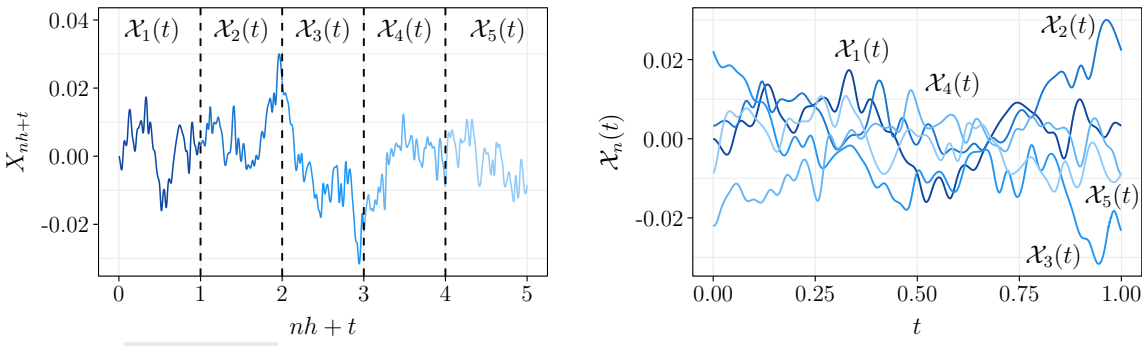
being $(\mathbf{Y}_q)_i$ the i-th row of $\mathbf{Y}_q$;

(iv) A FPCR estimation is performed just using the set of $\widetilde{p} \leqslant p$ predictors selected, not necessarily consecutive FPC, whose coefficients are denoted as $\widetilde{\mathbf{X}}_{\widetilde{p}}$, and therefore, having a hat matrix $\mathbf{H}_C^{(\lambda)}$ at our disposal, which will be crucial within the bootstrap algorithm.

Hence, the FPCR-L1-selected (FPCR-L1S) estimator in (8.4) allows to seize the advantages of both estimation paradigms. This estimator critically depends on the penalty parameter λ. Whilst leave-one-out cross-validation $\hat{\lambda}_{CV}$ is an optimal choice for estimating β, we adopt the so-called one standard error rule (see Friedman et al., 2010), denoted as $\hat{\lambda}_{1SE}$ (a variance reduction is achieved to obtain a more biased estimator for test calibration). The estimator in (8.4) could be extended to nonorthonormal bases, just replacing $\mathbf{X}_p$ by $\check{\mathbf{X}}_p = \mathbf{X}_p\boldsymbol{\Psi}$ in (8.3), where $\boldsymbol{\Psi} = \left(\langle \Psi_j, \Psi_j \rangle_{\mathbb{H}_1}\right)_{i,j=1,\ldots,p}$.

8.2.3 Functional time series: autoregressive Hilbertian processes

Since the underlying idea of the methodology proposed in Section 8.3 revolves around characterizing ARH processes as a particular case of FLMFR, firstly some preliminary elements on these models from a time-domain perspective are introduced. Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{X_t\}_{t\in\mathbb{R}}$ be now a continuous-time zero-mean stochastic process. In keeping with Bosq (2000), we split the paths as $\mathcal{X}_n(t) = X_{nh+t}$, with $t \in [0, h]$, and $\mathcal{X}_n \in \mathbb{H} = L^2([0, h])$, for each $n \in \mathbb{Z}$, constituting an infinite-dimensional discrete-time process. This representation (see Figure 8.1) is especially fruitful if $\{X_t\}_{t\in\mathbb{R}}$ displays a seasonal component either it will be forecasted over $[0, h]$.



(a) Continuous path of the OU process $\{X_t\}_{t\in\mathbb{R}^+}$, splitted into $[0, h]$, with $h = 1$, such that $\mathcal{X}_i(t) = X_{ih+t}$.

(b) OU process characterized as a set of ARH(1) trajectories $\{\mathcal{X}_n(t) : t \in [0, 1]\}_{i=1,\ldots,n}$, with $n = 5$.

Figure 8.1: Stochastic process $\{X_t\}_{t\in\mathbb{R}^+}$ characterized and splitted as an ARH(1) process.

We say that $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ is a zero-mean autoregressive Hilbertian process of order one, valued in $\mathbb{H} = L^2([0,h])$, denoted as ARH(1), if the following state equation is satisfied

$$\mathcal{X}_n(t) = \Gamma\left(\mathcal{X}_{n-1}\right)(t) + \mathcal{E}_n(t), \qquad n \in \mathbb{Z}, \quad t \in [0,h], \tag{8.5}$$

with $\mathcal{X}_n,\ \mathcal{E}_n \in \mathbb{H} = L^2([0,h])$ and where $\Gamma \equiv \Gamma_\rho$ denotes the linear autocorrelation operator, with $\|\Gamma\|_{\mathcal{L}(\mathbb{H})} = \sup_{\|\mathcal{X}\|_{\mathbb{H}}\leqslant 1} \|\Gamma(\mathcal{X})\|_{\mathbb{H}}$, and the error $\{\mathcal{E}_n\}_{n\in\mathbb{Z}}$ is assumed to be a strong-white noise, with $\sigma^2_{\mathcal{E}} = \mathbb{E}\left[\|\mathcal{E}_n\|^2_{\mathbb{H}}\right] < \infty$ and i.i.d. components, and where $\mathcal{L} = \left\{\Gamma_\rho(\mathcal{X})(t) = \int_0^h \rho(s,t)\mathcal{X}(s)\,\mathrm{d}s\colon \rho \in L^2([0,h]\times[0,h])\right\}$. The subsequent assumptions 10–11 will be considered.

Assumption 10. $\left\|\Gamma^k\right\|_{\mathcal{L}(\mathbb{H})} < 1$, for any $k \geqslant k_0$, and for some $k_0 \geqslant 1$, where Γ^k denotes the composition operator $\Gamma \overbrace{\circ\cdots\circ}^{k} \Gamma$, leading to $\sum_{n=0}^{\infty} \|\Gamma^n\|_{\mathcal{L}(\mathbb{H})} < \infty$.

Assumption 11. The autocorrelation operator is given by

$$\Gamma_\rho(\mathcal{X})(t) = \int_0^h \rho(s,t)\mathcal{X}(s)\,\mathrm{d}s,$$

that is, a Hilbert-Schmidt integral operator, with $\rho = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \rho_{jk}\Psi_j \otimes \Psi_k$, such that $\int\int \rho^2(s,t)\,\mathrm{d}s\,\mathrm{d}t < \infty$ and $\rho_{jk} = \langle \rho(\Psi_j), \Psi_k\rangle_{\mathbb{H}}$, for each $j,k = 1,\ldots,\infty$.

Remark that the autocovariance operator $C := \mathbb{E}\left[\mathcal{X}_n \otimes \mathcal{X}_n\right]$ is a self-adjoint, trace and positive operator, for each $n \in \mathbb{Z}$. As a result, it admits a diagonal FPC-based decomposition $C = \sum_{j=1}^{\infty} C_j \Psi_j \otimes \Psi_j$, being $\{\Psi_j\}_{j=1}^{\infty}$ the theoretical eigenfunctions of C, associated with eigenvalues $C_1 \geqslant \ldots \geqslant C_j \geqslant \ldots > 0$. From Bosq (2000, Theorem 3.1), Assumption 10 is required for ensuring an unique stationary solution. Since $C^{-1} = \sum_{j=1}^{\infty} \frac{1}{C_j}\Psi_j \otimes \Psi_j$, under the trace property, C cannot be inverted, and thus, $\Gamma := DC^{-1}$ should be estimated. Note that Assumption 11 does not imply that DC^{-1} could be diagonally decomposed, since the symmetry of the cross-covariance operator $D := \mathbb{E}\left[\mathcal{X}_n \otimes \mathcal{X}_{n+1}\right]$ is not imposed.

Remark 1. As a sideways contribution, in Appendix C a strongly consistent estimator of Γ is provided, under weaker conditions that those ones in Bosq (2000), improving the decay rate of convergence of the associated predictor. The estimation under the compactness of D was achieved in Álvarez-Liébana et al. (2017) and Ruiz-Medina and Álvarez-Liébana (2019b), on the Hilbert-Schmidt and trace norms.

8.3 A goodness-of-fit test for functional autoregressive processes

In this section, a new goodness-of-fit test for ARH processes against an unspecified alternative is proposed. Firstly, the ARH(1) process in Section 8.2.3 is extended to ARH(z) models, with $z \geqslant 1$. Secondly, the ARH(z) processes are characterized as a particular case of FLMFR, under the setting established in Section 8.2.2. Lastly, the goodness-of-fit test proposal for ARH(z) processes is detailed, from a FLMFR perspective, and a simulation study is deployed.

The proposed goodness-of-fit test for ARH(z) models can also be used as a sequential procedure to determine the autoregressive order z. As, up to our knowledge, no other goodness-of-fit tests have been proposed for ARH(z) processes, the performance of the proposed test used as a sequential procedure for order selection is compared against the proposal by Kokoszka and Reimherr (2013). The authors proposed a multistage procedure to determine the order of an ARH model, using an iterative process considering ARH($z+1$) alternatives, that is, the ARH structure is assumed, in contrast to the unspecified alternative hypothesis of the test proposed here.

8.3.1 ARH(z) processes: state equation

The Markovianess of the ARH(1) model in (8.5) allows us to easily generalized it by including more lagged functional regressors, i.e., $\mathcal{X}_n$ is inferred from $(\mathcal{X}_{n-1}, \ldots, \mathcal{X}_{n-z})$. Formally, we say that $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ is a zero-mean autoregressive Hilbertian process of order $z \geqslant 1$, valued in $\mathbb{H} = L^2([0,h])$ and denoted as ARH(z), if the following state equation is satisfied,

$$\mathcal{X}_n(t) = \sum_{r=1}^{z} \Gamma_r\left(\mathcal{X}_{n-r}\right)(t) + \mathcal{E}_n(t), \qquad \mathcal{X}_n, \mathcal{E}_n \in \mathbb{H} = L^2([0,h]), \tag{8.6}$$

with $n \in \mathbb{Z}$, $t \in [0,h]$ and order $z \geqslant 1$, and where $\{\Gamma_r\}_{r=1}^{z}$ are bounded linear operators in $\mathcal{L}(\mathbb{H})$, and $\{\mathcal{E}_n\}_{n\in\mathbb{Z}}$ is a $\mathbb{H}$-valued strong white noise, such that $\mathbb{P}\left(\Gamma_z\left(\mathcal{X}_n\right) \neq 0\right) > 0$ is implicitly assumed, for each $n \in \mathbb{Z}$, for the identifiability of z.

ARH(1) estimation results can be effortlessly extended to this context since ARH(z) process in (8.6) can be reinterpreted a particular multivariate ARH(1) process, valued in $\mathbb{H}^z := \prod_{r=1}^{z} \mathbb{H}$, constituting likewise a separable Hilbert space (see Lemma 1 in Appendix C). In this way, the following proposition allows us to characterize the ARH(z) in (8.6) as a $\mathbb{H}^z$-valued stationary ARH(1) process, endowed with the norm $\langle\cdot,\cdot\rangle_{\mathbb{H}^z}$. See proofs of Propositions 1–2 below in Appendix C.

Proposition 1. Let $\mathcal{X} \coloneqq \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ be a zero-mean ARH(z) process valued in $\mathbb{H} = L^2([0,h])$, with $z \geqslant 1$. The ARH(z) model in (8.6) can be reinterpreted as

$$\overline{\mathcal{X}}_n = \begin{pmatrix} \mathcal{X}_n \\ \vdots \\ \mathcal{X}_{n-z+1} \end{pmatrix} \in \mathbb{H}^z,$$

$$\overline{\Gamma} = \begin{pmatrix} \Gamma_1 & \Gamma_2 & \dots & \Gamma_{z-1} & \Gamma_z \\ Id_{\mathbb{H}} & 0_{\mathbb{H}} & \dots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ 0_{\mathbb{H}} & Id_{\mathbb{H}} & \dots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\mathbb{H}} & 0_{\mathbb{H}} & \dots & Id_{\mathbb{H}} & 0_{\mathbb{H}} \end{pmatrix}, \tag{8.7}$$

$$\overline{\mathcal{E}}_n = (\mathcal{E}_n, \mathbf{0}, \dots, \mathbf{0})^\top \in \mathbb{H}^z,$$

where $\mathbb{H}^z$ is the Cartesian product of z copies of $\mathbb{H}$. Thus, $\overline{\mathcal{X}}_n = \overline{\Gamma}\left(\overline{\mathcal{X}}_{n-1}\right) + \overline{\mathcal{E}}_n$, constitutes a $\mathbb{H}^z$-valued ARH(1) process, for each $n \in \mathbb{Z}$, with $\overline{\mathcal{E}}_n$ a $\mathbb{H}^z$-valued strong white noise. Furthermore, the joint operator $\overline{\Gamma}$ is also a bounded linear operator in $\mathbb{H}^z$. Note that in (8.7), $Id_{\mathbb{H}}$ and $0_{\mathbb{H}}$ denote the identity and null operators, respectively, and $\mathbf{0}$ the null element on $\mathbb{H}$.

Concerning Assumption 10, a sufficient condition for the existence of an unique stationary solution of (8.6) is provided in Proposition 2.

Proposition 2. Let $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ be a zero-mean ARH(z) process with $z \geqslant 1$, as explicitly defined in (8.6)–(8.7). If $\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)} < 1$, for any $k \geqslant k_0$ and for some $k_0 \geqslant 1$, then the state equation established in (8.6) has a unique stationary solution given by

$$\mathcal{X}_n = \sum_{j=0}^{\infty} \Pi_1 \left(\overline{\Gamma}^j \left(\overline{\mathcal{E}}_{n-j}\right)\right),$$

for each $n \in \mathbb{Z}$, where $\Pi_r \colon \left(\mathcal{X}_{(1)}, \dots, \mathcal{X}_{(z)}\right) \mapsto \mathcal{X}_{(r)}$, with $r = 1, \dots, z$.

Remark 2. As proved in Lemma 2 in Appendix C, Assumption 10 on $\overline{\Gamma}$ also implies that Assumption 10 is held for each Γ_r, with $r = 1, \dots, z$ (i.e., stationarity condition in Proposition 2 leads to $\sum_{n=0}^{\infty} \|\Gamma_r^n\|_{\mathcal{L}(\mathbb{H})} < \infty$). Remark also that, from the definition of norm $\|\cdot\|_{\mathbb{H}^z}$, Assumption 11 on $\overline{\Gamma}$ is verified as long as Assumption 11 is satisfied for each Γ_r, for any $r = 1, \dots, z$. Note that now $\overline{C} \coloneqq \mathbb{E}\left[\overline{\mathcal{X}}_n \otimes_{\mathbb{H}^z} \overline{\mathcal{X}}_n\right]$ is the autocovariance operator, where $\otimes_{\mathbb{H}^z}$

denotes the tensor product on $\mathbb{H}^z$, given by

$$\overline{C}(\overline{\mathcal{X}}) = \sum_{j=1}^{\infty} \overline{C}_j \overline{\Psi}_j \left\langle \overline{\Psi}_j, \overline{\mathcal{X}} \right\rangle_{\mathbb{H}^z},$$

$$\left\langle \overline{C}(\overline{\mathcal{X}}), \overline{\mathcal{Y}} \right\rangle_{\mathbb{H}^z} = \sum_{j=1}^{\infty} \overline{C}_j \left\langle \overline{\Psi}_j, \overline{\mathcal{X}} \right\rangle_{\mathbb{H}^z} \left\langle \overline{\Psi}_j, \overline{\mathcal{Y}} \right\rangle_{\mathbb{H}^z},$$

with $\overline{\mathcal{X}},\ \overline{\mathcal{Y}} \in \mathbb{H}^z$ and where $\left\{\overline{\Psi}_j\right\}_{j=1}^{\infty}$ and $\left\{\overline{C}_j\right\}_{j=1}^{\infty}$ denote their $\mathbb{H}^z$-valued FPC and their eigenvalues, respectively. The joint operator $\overline{C}$ can be reinterpreted as $\overline{C} = (C_{r,s})_{r=1,\ldots,z}^{s=1,\ldots,z}$, where $C_{r,s} = \mathbb{E}\left[\mathcal{X}_{n,(r)} \otimes \mathcal{X}_{n,(s)}\right]$, being $\mathcal{X}_{n,(r)}$ the r-th element of $\overline{\mathcal{X}}_n$, for each $n \in \mathbb{Z}$ and $r = 1, \ldots, s$. Specifically,

$$C_{r,s} = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} C_{j,k}^{(r,s)} \Psi_j \otimes \Psi_k,$$

with $C_{r,s}\left(\mathcal{X}\right)(\cdot) = \langle c_{r,s}(\cdot, u), \mathcal{X} \rangle_{\mathbb{H}}$ and

$$c_{r,s}(t,u) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} C_{j,k}^{(r,s)} \Psi_j(t) \Psi_k(u),$$

for each $r, s = 1, \ldots, z$, and $t, u \in [0, h]$.

8.3.2 ARH(z) processes characterized as FLMFR

Notwithstanding a strongly consistent estimator of Γ in the ARH setting is provided in Appendix C, its expression may be intricate, and the lack of an explicit hat matrix implies a considerably increase in the computational cost. For these reasons, an alternative characterization of the ARH(z) model in (8.6), based on properties detailed in Section 8.3.1, is now suggested, re-expressed it as a FLMFR. As commented, from Remark 2, under Assumption 11 on each $\{\Gamma_r\}_{r=1}^{z}$, we have

$$\begin{aligned} \Gamma_r\left(\mathcal{X}\right)(t) &= \int_0^h \rho_r(s,t) \mathcal{X}(s)\, \mathrm{d}s, \\ \rho_r &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \rho_{jk}^{(r)} \Psi_j \otimes \Psi_k, \end{aligned} \tag{8.8}$$

with $\int_0^h \int_0^h \rho_r^2(s,t)\, \mathrm{d}s\, \mathrm{d}t < \infty$. Given a zero-mean ARH(z) process in (8.6), we define $\widetilde{\mathcal{X}}_n(s) = \sum_{r=1}^{z} \mathcal{X}_{n-r}\left(sz - (r-1)\right) \mathbb{1}_r(s)$, for each $s \in [0, h]$, where $\mathbb{1}_r(\cdot) := \mathbb{1}_{\left\{\frac{(r-1)h}{z}, \frac{rh}{z}\right\}}(\cdot)$ denotes the indicator function in $\left[\frac{(r-1)h}{z}, \frac{rh}{z}\right]$, for each $r = 1, \ldots, z$. In the same way, a change of

variable $x \coloneqq (s + r - 1)/z$ yields

$$\begin{aligned}\Gamma_r\left(\mathcal{X}_{n-r}\right)(t) &= \int_0^h \rho_r(s,t)\mathcal{X}_{n-r}(s)\,\mathrm{d}s \\ &= z\int_{\frac{(r-1)h}{z}}^{\frac{rh}{z}} \rho_r(xz-(r-1),t)\mathcal{X}_{n-r}(xz-(r-1))\,\mathrm{d}x,\end{aligned}$$

for each $r = 1, \ldots, z$ and $t \in [0, h]$, and then,

$$\begin{aligned}\sum_{r=1}^{z}\Gamma_r\left(\mathcal{X}_{n-r}\right)(t) &= \sum_{r=1}^{z} z\int_{\frac{(r-1)h}{z}}^{\frac{rh}{z}} \rho_r(xz-(r-1),t)\mathcal{X}_{n-r}(xz-(r-1))\,\mathrm{d}x \\ &= \int_0^h \left(z\sum_{r=1}^{z}\rho_r(sz-(r-1),t)\mathbb{1}_r(s)\right)\left(\sum_{r=1}^{z}\mathcal{X}_{n-r}(sz-(r-1))\mathbb{1}_r(s)\right)\mathrm{d}s,\end{aligned}$$

such that

$$\begin{aligned}\sum_{r=1}^{z}\Gamma_r\left(\mathcal{X}_{n-r}\right)(t) &= \int_0^h \widetilde{\rho}(s,t)\widetilde{\mathcal{X}}_n(s)\,\mathrm{d}s, \\ \widetilde{\rho}(s,t) &= z\sum_{r=1}^{z}\rho_r(sz-(r-1),t)\mathbb{1}_r(s),\end{aligned} \tag{8.9}$$

and $\widetilde{\mathcal{X}}_n(s) = \sum_{r=1}^{z}\mathcal{X}_{n-r}(sz-(r-1))\mathbb{1}_r(s)$, where we get, under Assumption 11 on $\{\Gamma_r\}_{r=1}^{z}$, the inequality

$$\int_0^h\int_0^h \widetilde{\rho}^2(s,t)\,\mathrm{d}s\,\mathrm{d}t \leqslant z\sum_{r=1}^{z}\int_0^h\int_0^h \rho_r^2(s,t)\,\mathrm{d}s\,\mathrm{d}t < \infty,$$

as long as $z < \infty$. Hence, we are under the FLMFR setting detailed in Section 8.2.2, under Assumptions 10–11, which ensure us the stationarity. Given a sample of a zero-mean ARH(z) process $\{\mathcal{X}_i\}_{i=0}^{n-1+z}$ valued in $\mathbb{H} = L^2([0,h])$, we then have, for each $i = 1, \ldots, n$,

$$\begin{aligned}\widetilde{\mathcal{Y}}_i &= \widetilde{\Gamma}\left(\widetilde{\mathcal{X}}_i\right) + \widetilde{\mathcal{E}}_i, \qquad \mathbb{E}\left[\widetilde{\mathcal{E}}_i \mid \widetilde{\mathcal{X}}_i\right] = 0, \\ \widetilde{\Gamma}(x) &= \mathbb{E}\left[\widetilde{\mathcal{Y}} \mid \widetilde{\mathcal{X}} = x\right] : \mathbb{H} \mapsto \mathbb{H},\end{aligned} \tag{8.10}$$

where

$$\begin{aligned}\widetilde{\mathcal{X}}_i &= \sum_{r=1}^{z}\mathcal{X}_{z-r+(i-1)}(sz-(r-1))\mathbb{1}_r(s), \\ \widetilde{\mathcal{Y}}_i &\coloneqq \mathcal{X}_{(i-1)+z} \\ \widetilde{\mathcal{E}}_i &\coloneqq \mathcal{E}_{(i-1)+z},\end{aligned}$$

for each $i = 1, \ldots, n$, all of them centered. From (8.8)–(8.10), we have the Hilbert-Schmidt operator

$$\widetilde{\Gamma}\left(\mathcal{X}\right)(\cdot) = \int_0^h \widetilde{\rho}(s, \cdot)\mathcal{X}(s)\,\mathrm{d}s \tag{8.11}$$

with

$$\widetilde{\rho} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \widetilde{\rho}_{jk}\Psi_j \otimes \Psi_k,$$

where $\widetilde{\mathbf{\Gamma}}_{\widetilde{p},q} = (\widetilde{\rho}_{jk})_{j=1,\ldots,\widetilde{p}}^{k=1,\ldots,q}$ is the $\widetilde{p} \times q$ matrix of the coefficients of $\widetilde{\rho}_{jk} = \langle \widetilde{\rho}, \Psi_j \otimes \Psi_k \rangle_{\mathbb{H}^z}$, being $\widetilde{\rho}$ defined in (8.9). Now, $\{\Psi_j \otimes \Psi_k\}_{j,k=1}^{\widetilde{p},q}$ is constituted by the truncated versions of the same functional basis $\{\Psi_j\}_{j=1}^{\infty}$, although different cut-off levels $(\widetilde{p}, q)$ might be required. Note that, though $\mathcal{X}_i$ and $\mathcal{E}_i$ are correlated, we have $\mathbb{E}\big[\widetilde{\mathcal{E}}_i \mid \widetilde{\mathcal{X}}_i\big] = 0$ for each $i = 1, \ldots, n$, since $(\mathcal{X}_{n-1}, \ldots, \mathcal{X}_{n-z})$ and $\mathcal{E}_n$ are uncorrelated for each $n \in \mathbb{Z}$. Since the previous z-lagged values are used, $\{\mathcal{X}_i\}_{i=0}^{n-1+z}$ are required for computing $\{\widetilde{\mathcal{X}}_i\}_{i=1}^{n}$.

Remark 3. Concerning to the case where $\widetilde{\Gamma}$ is furthermore compact, Assumption 11 could be modified, since $\widetilde{\Gamma}$ would admit a diagonal spectral decomposition in terms of an infinite sequence of real-valued AR(1) state equations, after projection the fully functional trajectories into the set of eigenfunctions of autocovariance operator (see, e.g., Salmerón and Ruiz-Medina, 2009 and Ruiz-Medina and Salmerón, 2010). In that case, we could extend in a direct way the asymptotic results by Koul and Stute (1999) for the real-valued case, such that considering ARH processes in terms of generalized processes (see Gelfand and Vilenkin, 1964), the weak-convergence of the projected empirical processes to a two-parameter generalized continuous Gaussian process may be achieved. This process would admit in distribution sense a representation, in terms of two-parameter generalized Brownian motion.

8.3.3 A goodness-of-fit test for ARH(z) models

As already exposed in the introduction, the guiding thread of this chapter is verifying whether the relationship among functional random variables $\{\widetilde{\mathcal{Y}}_i = \mathcal{X}_{i-1+z}\}_{i=1}^{n}$ and their z-lagged values $\{(\mathcal{X}_{z-1+(i-1)}, \mathcal{X}_{z-2+(i-1)}, \ldots, \mathcal{X}_{(i-1)})\}_{i=1}^{n}$ can be linearly related. For this purpose, a goodness-of-fit test for the ARH(z) model in (8.6) is now formulated, based on its characterization as a FLMFR. Particularly, for a given $z \geqslant 1$, we want to test the composite null hypothesis

$$\mathcal{H}_0\colon \mathcal{X}_n \text{ and } (\mathcal{X}_{n-1}, \ldots, \mathcal{X}_{n-z}) \text{ are linearly related}, \quad n \in \mathbb{Z}, \tag{8.12}$$

against an unspecified alternative. Note that, from Proposition 1, the null hypothesis $\mathcal{H}_0$ in (8.12) is equivalent to

$$\mathcal{H}_0 \colon \overline{\mathcal{X}}_n = \overline{\Gamma}\left(\overline{\mathcal{X}}_{n-1}\right) + \overline{\mathcal{E}}_n, \qquad \overline{\mathcal{X}} \in \mathbb{H}^z, \quad n \in \mathbb{Z}, \tag{8.13}$$

where

$$\overline{\Gamma}\left(\overline{\mathcal{X}}\right)(t) := \overline{\Gamma}_{\overline{\rho}}\left(\overline{\mathcal{X}}\right)(t) = \left\langle \overline{\rho}(\cdot, t), \overline{\mathcal{X}}(\cdot) \right\rangle_{\mathbb{H}^z},$$

for some unknown squared-integrable $\overline{\rho} \in \mathbb{H}^z \otimes_{\mathbb{H}^z} \mathbb{H}^z$, where $\mathbb{H}^z$ is a separable Hilbert space, as proved in Lemma 1 in Appendix C. Furthermore, from Section 8.3.2, $\mathcal{H}_0$ in (8.12)–(8.13) is also equivalent to

$$\mathcal{H}_0 \colon \widetilde{\Gamma} \in \mathcal{L} = \left\{ \langle\langle \cdot, \widetilde{\rho} \rangle\rangle \colon \widetilde{\rho} \in \mathbb{H} \otimes \mathbb{H} \right\}, \tag{8.14}$$

against the alternative hypothesis $\mathcal{H}_1 \colon \mathbb{P}\left(\widetilde{\Gamma} \notin \mathcal{L}\right) > 0$, with

$$\widetilde{\mathcal{Y}} = \widetilde{\Gamma}\left(\widetilde{\mathcal{X}}\right) + \widetilde{\mathcal{E}},$$
$$\widetilde{\Gamma}\left(\widetilde{\mathcal{X}}\right)(t) := \widetilde{\Gamma}_{\widetilde{\rho}}\left(\widetilde{\mathcal{X}}\right)(t) = \int_0^h \widetilde{\rho}(s,t)\widetilde{\mathcal{X}}(s)\,\mathrm{d}s,$$

where $\widetilde{\mathcal{X}}_i = \sum_{r=1}^{z} \mathcal{X}_{z-r+(i-1)}(sz-(r-1))\mathbb{1}_r(s)$, $\widetilde{\mathcal{Y}}_i := \mathcal{X}_{i-1+z}$ and $\widetilde{\mathcal{E}}_i := \mathcal{E}_{i-1+z}$ are centered functional random variables, for each $i = 1, \ldots, n$, being $\widetilde{\Gamma}$ a Hilbert-Schmidt integral operator. In what follows, we will consider the characterization of null hypothesis $\mathcal{H}_0$ in (8.14) from a FLMFR perspective, based on developments in (8.8)–(8.11).

From García-Portugués et al. (2021, Lemmas 1–2), the null hypothesis in (8.14) can be also characterized in terms of the projections of the functional paths as follows

$$\mathcal{H}_0 \colon \ \mathbb{E}\left[\langle \widetilde{\mathcal{Y}} - \langle\langle \widetilde{\mathcal{X}}, \widetilde{\rho} \rangle\rangle, \gamma_{\widetilde{\mathcal{Y}}} \rangle_{\mathbb{H}} \mathbb{1}_{\left\{\langle \widetilde{\mathcal{X}}, \gamma_{\widetilde{\mathcal{X}}} \rangle_{\mathbb{H}} \leqslant u\right\}}\right] = 0,$$

for almost every $u \in \mathbb{R}$, and for each direction $\gamma_{\widetilde{\mathcal{X}}}, \gamma_{\widetilde{\mathcal{Y}}} \in \mathbb{S}_{\mathbb{H}}$, where $\mathbb{S}_{\mathbb{H}} = \{\mathcal{X} \in \mathbb{H} \colon \ \|\mathcal{X}\|_{\mathbb{H}} = 1\}$ is the functional analogue of the Euclidean sphere. Extending the ideas of Escanciano (2006), deviations from $\mathcal{H}_0$ can be detected by computing a residual marked empirical process

$$R_n\left(u, \gamma_{\widetilde{\mathcal{X}}}, \gamma_{\widetilde{\mathcal{Y}}}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \langle \widetilde{\mathcal{E}}_i, \gamma_{\widetilde{\mathcal{Y}}} \rangle_{\mathbb{H}} \mathbb{1}_{\left\{\langle \widetilde{\mathcal{X}}, \gamma_{\widetilde{\mathcal{X}}} \rangle_{\mathbb{H}} \leqslant u\right\}}, \quad u \in \mathbb{R}, \quad \gamma_{\widetilde{\mathcal{X}}}, \gamma_{\widetilde{\mathcal{Y}}} \in \mathbb{S}_{\mathbb{H}},$$

where marks are given by the projected errors $\left\{\langle \widetilde{\mathcal{E}}_i, \gamma_{\widetilde{\mathcal{Y}}} \rangle_{\mathbb{H}} = \langle \widetilde{\mathcal{Y}}_i - \langle\langle \widetilde{\mathcal{X}}_i, \hat{\rho} \rangle\rangle, \gamma_{\widetilde{\mathcal{Y}}} \rangle_{\mathbb{H}}\right\}_{i=1}^{n}$ and jumps are determined by the projected regressors $\left\{\langle \widetilde{\mathcal{X}}_i, \gamma_{\widetilde{\mathcal{X}}} \rangle_{\mathbb{H}}\right\}_{i=1}^{n}$, where $\hat{\rho}$ denotes some estimator of $\widetilde{\rho}$ (see equations (8.8)–(8.11)). To measure how close the empirical process is to

zero, a projected Cramér-von Mises (PCvM) statistic, on the norm on $\mathbb{S}_{\mathbb{H}} \times \mathbb{S}_{\mathbb{H}} \times \mathbb{R}$, will be adopted as

$$\text{PCvM}_n = \int_{\mathbb{S}_{\mathbb{H}} \times \mathbb{S}_{\mathbb{H}} \times \mathbb{R}} \left[R_n\left(u, \gamma_{\widetilde{\mathcal{X}}}, \gamma_{\widetilde{\mathcal{Y}}}\right)\right]^2 F_{n,\gamma_{\widetilde{\mathcal{X}}}}(\mathrm{d}u)\omega_{\widetilde{\mathcal{X}}}(\mathrm{d}\gamma_{\widetilde{\mathcal{X}}})\omega_{\widetilde{\mathcal{Y}}}(\mathrm{d}\gamma_{\widetilde{\mathcal{Y}}}),$$

where $F_{n,\gamma_{\widetilde{\mathcal{X}}}}$ is the empirical cumulative distribution function (ecdf) of projected regressors, and $\omega_{\widetilde{\mathcal{X}}}$ and $\omega_{\widetilde{\mathcal{Y}}}$ are suitable measures on $\mathbb{S}_{\mathbb{H}}$. The infinite dimension of functional spheres makes this functional hard to handle, for that reason, the projected Cramér-von Mises statistic will be expressed in terms of finite-dimensional directions $\gamma_{\widetilde{\mathcal{X}}}^{(\widetilde{p})}$ and $\gamma_{\widetilde{\mathcal{Y}}}^{(q)}$, as well as $\widetilde{\mathcal{E}}_i^{(q)}$ and $\widetilde{\mathcal{X}}^{(\widetilde{p})}$, just adopting a $(\widetilde{p}, \text{q})$-truncated expansions. After some algebra developments, an easily computable statistic was obtained as

$$\text{PCvM}_{n,\widetilde{p},q} = \frac{1}{n^2} \frac{2\pi^{\widetilde{p}/2+q/2-1}}{q\Gamma(\widetilde{p}/2)\Gamma(q/2)} \text{Tr}\left[\widetilde{\mathbf{E}}_q^\top \mathbf{A}_\bullet \widetilde{\mathbf{E}}_q\right], \tag{8.15}$$

where

$$\left(\widetilde{\mathbf{E}}_q\right)_{i=1,\ldots,n}^{k=1,\ldots,q} = \hat{e}_{i,k} = \langle \widetilde{\mathcal{E}}_i, \Psi_k \rangle_{\mathbb{H}},$$

being $\text{Tr}\left[\cdot\right]$ the trace operator, $\Gamma\left(\cdot\right)$ the Gamma function and $\widetilde{\mathbf{E}}_q$ the error coefficients, that is, the projected estimators of $\widetilde{\mathcal{E}}_i := \mathcal{E}_{i-1+z}$, for each $i = 1, \ldots, n$. The matrix $\mathbf{A}_\bullet = (\mathbf{A}_\bullet)_{ij} = \left(\sum_{r=1}^n A_{ijr}\right)_{ij}$ is obtained as surface areas of particular spherical regions, which explicit expression can be given as (see García-Portugués et al., 2014 and García-Portugués et al., 2021)

$$A_{ijr} = A_{ijr}^{\measuredangle} \frac{\pi^{\widetilde{p}/2-1}}{\Gamma(\widetilde{p}/2)},$$

with

$$A_{ijr}^{\measuredangle} := \begin{cases} 2\pi, & \boldsymbol{x}_{i,\widetilde{p}} = \boldsymbol{x}_{j,\widetilde{p}} = \boldsymbol{x}_{r,\widetilde{p}}, \\ \pi, & \boldsymbol{x}_{i,\widetilde{p}} \neq \boldsymbol{x}_{j,\widetilde{p}} \text{ , and } \boldsymbol{x}_{i,\widetilde{p}} = \boldsymbol{x}_{r,\widetilde{p}} \text{ or } \boldsymbol{x}_{j,p} = \boldsymbol{x}_{i,\widetilde{p}}, \\ \pi - \cos^{-1}\left(\frac{\left(\boldsymbol{x}_{i,\widetilde{p}} - \boldsymbol{x}_{r,\widetilde{p}}\right)^\top \left(\boldsymbol{x}_{j,\widetilde{p}} - \boldsymbol{x}_{r,\widetilde{p}}\right)}{\left\|\boldsymbol{x}_{i,\widetilde{p}} - \boldsymbol{x}_{r,\widetilde{p}}\right\| \left\|\boldsymbol{x}_{j,\widetilde{p}} - \boldsymbol{x}_{r,\widetilde{p}}\right\|}\right), \end{cases}$$

where $\boldsymbol{x}_{i,\widetilde{p}} = \left(\langle \mathcal{X}_i, \Psi_1 \rangle, \ldots, \langle \mathcal{X}_i, \Psi_{\widetilde{p}} \rangle\right)^\top$. Geometrical arguments $\left(A_{ijr}\right)_{i,j,r=1,\ldots,n}$ depend exclusively on the covariates, and then, they only needs to be computed once in the testing procedure. The goodness-of-fit test for ARH(z) models proposed is then detailed within the next algorithm.

Algorithm 5 (GoF test for ARH(z) model in practice)**.** Let $\{\mathcal{X}_i\}_{i=0}^{n-1+z}$ be a sample of a centered ARH(z) model, within the setting above detailed, and under Assumptions 10–11. We

proceed as follows:

(i) Construct the functional random variables

$$\widetilde{\mathcal{X}}_i(s) := \sum_{r=1}^{z} \mathcal{X}_{z-r+(i-1)}(sz-(r-1))\mathbb{1}_r(s),$$
$$\widetilde{\mathcal{Y}}_i(t) := \mathcal{X}_{(i-1)+z}(t),$$
$$\widetilde{\mathcal{E}}_i(t) := \mathcal{E}_{(i-1)+z}(t),$$

all of them centered, with $\widetilde{\mathcal{Y}}_i = \widetilde{\Gamma}\left(\widetilde{\mathcal{X}}_i\right) + \widetilde{\mathcal{E}}_i$, for any $i = 1, \ldots, n$, and $s, t \in [0, h]$, constituting a particular case of FLMFR.

(ii) Compute the FPC of $\left\{\widetilde{\mathcal{X}}_i\right\}_{i=1}^n$ and $\left\{\widetilde{\mathcal{Y}}_i\right\}_{i=1}^n$, and choose initial cut-off levels (p, q) as the minimum number of FPC required for capturing a proportion of EV (e.g., EV_p = EV_q = 0.99). After that, we obtain the p- and q-truncated FPC scores $\widetilde{\mathbf{X}}_p$ and $\widetilde{\mathbf{Y}}_q$ of $\left\{\widetilde{\mathcal{X}}_i\right\}_{i=1}^n$ and $\left\{\widetilde{\mathcal{Y}}_i\right\}_{i=1}^n$, respectively.

(iii) Compute the FPCR-L1S estimator $\widehat{\widetilde{\mathbf{B}}}_{\widetilde{p},q}^{(\lambda),C}$ of $\widetilde{\mathbf{B}}_{\widetilde{p},q}$, where $\widetilde{\mathbf{B}}_{\widetilde{p},q} = \left(\widetilde{\beta}_{jk}\right)_{j=1,\ldots,\widetilde{p}}^{j=1,\ldots,q}$ is the $\widetilde{p} \times q$ matrix of the coefficients of $\widetilde{\beta}_{jk} = \langle \widetilde{\rho}, \Psi_j \otimes \Psi_k \rangle_{\mathbb{H}^z}$, and the associated coefficients of residuals $\widetilde{\mathbf{e}}_{i,q} = \widetilde{\mathbf{Y}}_{i,q} - \widetilde{\mathbf{X}}_{i,\widetilde{p}} \widehat{\widetilde{\mathbf{B}}}_{\widetilde{p},q}^{(\lambda),C}$, for each $i = 1, \ldots, n$. Note that $\widetilde{p}$ out of p FPC coefficients are automatically selected. We compute the statistic $\text{PCvM}_{n,\widetilde{p},q}$ in (8.15).

(iv) Implement a golden-section bootstrap resampling procedure, by simulating independent zero-mean and unit-variance random variables, setting bootstrapped errors as $\mathbf{e}_{i,q}^{*b}$ and $\widetilde{\mathbf{Y}}_{i,q}^{*b} := \widetilde{\mathbf{X}}_{i,\widetilde{p}} \widehat{\widetilde{\mathbf{B}}}_{\widetilde{p},q}^{(\lambda),C} + \mathbf{e}_{i,q}^{*b}$, for each $i = 1, \ldots, n$ and $b = 1, \ldots, B$, being B the number of replicates.

(v) After centering them, from the bootstrap sample, recompute the FPCR-L1S estimator in (8.4), obtaining the bootstrapped residuals, and then, the bootstrapped statistic $\text{PCvM}_{n,\widetilde{p},q}^{*b}$ from (8.15).

(vi) Estimate the p-value as $\#\left\{\text{PCvM}_{n,\widetilde{p},q} \leqslant \text{PCvM}_{n,\widetilde{p},q}^{*b}\right\}/B$.

8.3.4 Numerical results: a comparative study

The performance of the test, regarding power and size, is now compared to the multi-stage testing procedure in Kokoszka and Reimherr (2013), henceforth abbreviated as KR, to determine the order of an ARH process. Note that their approach was developed just under ARH alternatives, in contrast with our unspecified alternatives. The following common settings

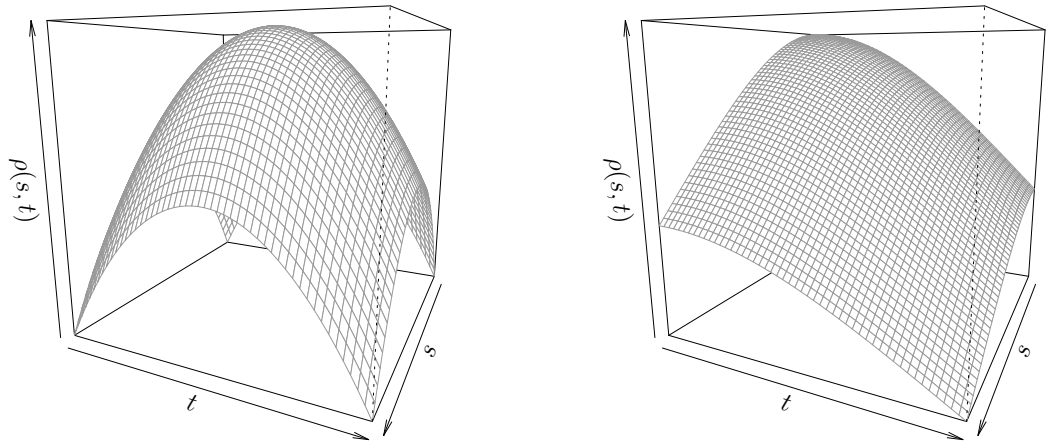
will be used through scenarios described in Table 8.1: functional trajectories $\{\mathcal{X}_i\}_{i=1}^n$ and errors $\{\mathcal{E}_i\}_{i=1}^n$ simulated as standard Brownian bridges, as shown in Figure 8.1, are valued in 101 equispaced points in $[0,1]$, with $1\,000$ Monte Carlo replicates, $\mathrm{B} = 1\,500$ bootstrap resamples. The initial values were simulated as standard Brownian bridges and a burn-in period of 200 functional observations was used. Our test was run following Algorithm 5, using initial cut-off levels (p,q) for ensuring us an explained variance of $\mathrm{EV}_p = \mathrm{EV}_q = 0.995$. Subsequently, the FPCR-L1S estimator in (8.4) was implemented by using the so-called one standard error rule λ_{1SE}, such that a L1-based variable selection is achieved. Note that, as discussed in Section C.2 in Appendix C, the variable selection is addressed focusing on the testing procedure, not on the estimation of the autocorrelation operator.

Notation	**Model**	**Parameters**
ARH(0)	$\mathcal{X}_n(t) = \mathcal{E}_n(t)$	None
ARH(1)	$\mathcal{X}_n(t) = \Gamma_1(\mathcal{X}_{n-1})(t) + \mathcal{E}_n(t)$ $\rho_1(s,t) = c_1(2-(2s-1)^2-(2t-1)^2)$	$\lVert\rho_1\rVert_{L^2([0,1]^2)} = 0.7$
ARH(2)	$\mathcal{X}_n(t) = \Gamma_1(\mathcal{X}_{n-1})(t) + \Gamma_2(\mathcal{X}_{n-2})(t) + \mathcal{E}_n(t)$ $\rho_i(s,t) = c_{2,i}e^{-(t^2+s^2)/2}$	$\lVert\rho_1\rVert_{L^2([0,1]^2)} = 0.5$ $\lVert\rho_2\rVert_{L^2([0,1]^2)} = 0.3$
Non linear, quadratic (NLQ)	$\mathcal{X}_n(t) = \Gamma_1\left(\mathcal{X}_{n-1}^2\right)(t) + \mathcal{E}_n(t)$ $\rho_1(s,t) = c_{2,1}e^{-(t^2+s^2)/2}$	$\lVert\rho_1\rVert_{L^2([0,1]^2)} = 0.5$
Non linear, square root (NLS)	$\mathcal{X}_n(t) = \Gamma_1\left(\lvert\mathcal{X}_{n-1}\rvert^{0.5}\right)(t) + \mathcal{E}_n(t)$ $\rho_1(s,t) = c_{2,1}e^{-(t^2+s^2)/2}$	$\lvert\rho_1\rvert_{L^2([0,1]^2)} = 0.5$

Table 8.1: Summary of null and alternative hypotheses.

As summarized in Table 8.1, different ARH(z) models are generated, with $z \in \{0,1,2\}$, and two nonlinear models, where $\Gamma(\mathcal{X})(\cdot) = \int_0^1 \rho(s,\cdot)\mathcal{X}(s)\,\mathrm{d}s$ and ρ are parabolic (see Figure 8.2a) or Gaussian (see Figure 8.2b) kernels, commonly used in the literature (see Gabrys et al., 2010 and Hörmann and Kokoszka, 2010). In Table 8.1, $\mathcal{X}^2$ and $\lvert\mathcal{X}\rvert^{0.5}$ denote the pointwise exponentiation and square root operations, respectively, to the discretized trajectories. Similar conditions on $\lVert\rho\rVert_{L^2([0,1]^2)}$ proposed in Kokoszka and Reimherr (2013) are imposed, such that Assumptions 10–11 are held, and $(c_1, c_{2,1}, c_{2,2}) = (0.500568, 0.669502, 0.401701)$ are set, so that the L^2 norm of ρ is approximately the value displayed in the last column of Table 8.1. Table 8.2 shows the empirical rejection rates of testing $\mathcal{H}_0$ whether $\mathcal{X}_n$ is an ARH process of order z, with $z \in \{0,1\}$, comparing the performance of the proposed CvM test and the KR test, for scenarios in which data is generated following an ARH(z) model, with $z \in \{0,1,2\}$, and two nonlinear models. As remarked, the proposed CvM test does not spec-

ify an alternative hypothesis, while the KR proposal assumes a specific alternative, that is, it tests an ARH(z) against an ARH($z+1$) model. Nominal level $\alpha = 0.05$ and sample sizes $n \in \{150, 250, 350, 500\}$ are considered.



(a) Parabolic kernel for the ARH(1) simulation, where $\rho(s,t) = c_1(2-(2s-1)^2-(2t-1)^2)$.

(b) Gaussian kernel for the ARH(2) simulation, where $\rho(s,t) = c_{2,1}\exp\left(-(t^2+s^2)/2\right)$.

Figure 8.2: Gaussian and parabolic kernels used in Table 8.1.

		CvM					**KR**			
Data	$\mathcal{H}_0$	150	250	350	500	$\mathcal{H}_1$	150	250	350	500
ARH(0)	ARH(0)	0.042	0.042	0.047	0.041	ARH(1)	0.045	0.056	0.051	0.054
	ARH(1)	0.045	0.042	0.047	0.040	ARH(2)	0.056	0.055	0.061	0.050
ARH(1)	ARH(0)	1.000	1.000	1.000	1.000	ARH(1)	1.000	1.000	1.000	1.000
	ARH(1)	**0.030**	0.047	0.058	0.041	ARH(2)	0.056	0.053	0.063	0.042
ARH(2)	ARH(0)	1.000	1.000	1.000	1.000	ARH(1)	1.000	1.000	1.000	1.000
	ARH(1)	0.367	0.596	0.632	0.651	ARH(2)	0.954	1.000	1.000	1.000
NLQ	ARH(0)	0.541	0.818	0.949	0.998	ARH(1)	0.406	0.652	0.842	0.972
	ARH(1)	0.535	0.808	0.946	0.984	ARH(2)	0.059	0.047	0.058	0.048
NLS	ARH(0)	0.203	0.358	0.580	0.795	ARH(1)	0.164	0.235	0.311	0.426
	ARH(1)	0.198	0.355	0.597	0.797	ARH(2)	0.056	0.045	0.069	0.053

Table 8.2: Empirical size and power of CvM and KR tests for testing the order of ARH processes with $n \in \{150, 250, 350, 500\}$. *Under* $\mathcal{H}_0$*, rejection rates are boldfaced if they lie outside a 95%-confidence interval for* $\alpha = 0.05$.

According to scenarios displayed in Table 8.2, regarding the size under null hypotheses (first, second and forth row), CvM and KR tests seem to be well calibrated, though both

show a slight over-rejection in the ARH(1) scenario when indeed testing ARH(1) as $\mathcal{H}_0$. As observed, size of tests are also close to the nominal level $\alpha = 0.05$, even when small sample size are considered. Concerning powers under ARH alternatives (third, fifth and sixth rows), the KR test is more powerful, as expected since the KR test was specifically designed for these scenarios, rejecting ARH(z) in favor of ARH($z+1$). However, since KR test assumes an ARH structure even under alternatives, it fails against nonlinear alternatives (eighth, ninth and tenth rows), showing a poor power, in contrast with the high empirical power exhibits by the CvM test in most of scenarios. As opposed to the KR test, the proposed test shows an increasing power inasmuch as increasing the sample size.

8.4 Specification test for diffusion models as functional time series

The methodology introduced in Section 8.3 can be further extended to diffusion processes, where the common framework is one dimensional, as usually one path of the process is considered. However, in finance, and motivated by the wide set of scenarios in which high-frequency data are available, multivariate goodness-of-fit tests could be unsuitable for such data, and thus, functional data setup has been exploited to take advantage of the shapes of curves (Müller et al., 2011). As a sideways contribution, a new specification test for stochastic diffusion models from a high-dimensional data perspective is proposed, namely for Ornstein-Uhlenbeck processes, which will be characterized as a particular ARH(1) process.

8.4.1 Ornstein-Uhlenbeck process as a particular ARH(1)

The relevance of OU processes is not limited to finance, since they have a long history in physics, introduced by Langevin (1908), formalized as $\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, with $t \in \mathbb{R}^+$, and $\kappa,\ \sigma > 0$. This process is often employed under its integral representation, via the Fourier method of separation of variables,

$$X_t = X_0 e^{-\kappa t} + \mu\left(1 - e^{-\kappa t}\right) + \sigma \int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W_s, \qquad t \in \mathbb{R}^+, \tag{8.16}$$

with X_0 the initial condition at $t_0 = 0$. With the same philosophy of the splitted representation adopted in Section 8.3 (see Figure 8.1), and according to the proposal by Álvarez-Liébana et al. (2016), the OU process in (8.16) can be characterized as an ARH(1) process. Let $\mathbb{H}$ be a separable Hilbert space given by $\mathbb{H} = L^2\left([0,h], \mathcal{B}_{[0,h]}, \lambda + \delta_{(h)}\right)$, where

$\mathcal{B}_{[0,h]}$ is the σ-algebra generated by subintervals $[0,h]$, λ denotes the Lebesgue measure and $\delta_{(h)}(s) = \delta(s-h)$ is the Dirac measure at h. The associated norm is defined as $\|\mathcal{X}\|_{\mathbb{H}} = \sqrt{\int_0^h \mathcal{X}^2(s)\,\mathrm{d}\left(\lambda+\delta_{(h)}\right)} = \sqrt{\int_0^h \mathcal{X}^2(s)\,\mathrm{d}s + \mathcal{X}^2(h)}$. Remark that $\|\cdot\|_{\mathbb{H}}$ directly establishes equivalent classes of functions such that $\mathcal{X} \underset{\lambda+\delta_{(h)}}{\sim} \mathcal{Y}$ if and only if $\lambda\left(\{s\colon \mathcal{X}(s) \neq \mathcal{Y}(s)\}\right) = 0$ and $\mathcal{X}(h) = \mathcal{Y}(h)$. In this setting, the OU process in (8.16) can be splitted as follows, for each $n \in \mathbb{Z}$,

$$\mathcal{X}_n(t) = X_{nh+t} = X_0 e^{-\kappa(nh+t)} + \mu\left(1 - e^{-\kappa(nh+t)}\right) + \sigma\int_0^{nh+t} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s,$$

with $t \in [0,h]$, and then,

$$e^{-\kappa t}\mathcal{X}_{n-1}(h) = e^{-\kappa t}X_{nh} = X_0 e^{-\kappa(nh+t)} + \mu\left(e^{-\kappa t} - e^{-\kappa(nh+t)}\right) + \sigma\int_0^{nh} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s.$$

The set of paths $\{\mathcal{X}_n(t)\}_{n\in\mathbb{Z}} = \{X_{nh+t}\colon t \in [0,h]\}_{n\in\mathbb{Z}}$ can be expressed as follows,

$$e^{-\kappa t}\left(\mathcal{X}_{n-1}(h) - \mu\right) = \left(\mathcal{X}_n(t) - \mu\right) - \sigma\int_{nh}^{nh+t} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s.$$

From now on, $\{X_t\}_{t\in\mathbb{R}^+}$ will be centered respect to its long term mean ($\mu = 0$), and therefore, from Álvarez-Liébana et al. (2016), the OU process in (8.16) can be characterized as a zero-mean stationary ARH(1) process $\{\mathcal{X}_n(t) \coloneqq X_{nh+t},\ t \in [0,h]\}_{n\in\mathbb{Z}}$ (see Figures 8.1a–8.1b), given by

$$\mathcal{X}_n(t) = e^{-\kappa t}\mathcal{X}_{n-1}(h) + \sigma\int_{nh}^{nh+t} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s = \Gamma_\kappa\left(\mathcal{X}_{n-1}\right)(t) + \mathcal{E}_n(t), \tag{8.17}$$

with $n \in \mathbb{Z}$ and where $\left\{\mathcal{E}_n(t) \coloneqq \sigma\int_{nh}^{nh+t} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s\right\}_{n\in\mathbb{Z}}$ constitutes a $\mathbb{H}$-valued strong white noise and Γ_κ is a bounded linear operator, for each $\kappa > 0$ (Álvarez-Liébana et al., 2016). Note that $\|\Gamma_\kappa(\mathcal{X})\|_{\mathbb{H}}^2 = \int_0^h |\Gamma_\kappa(\mathcal{X})(t)|^2\,\mathrm{d}t + |\Gamma_\kappa(\mathcal{X})(h)|^2$, for each $\mathcal{X} \in \mathbb{H}$, $\left\|\Gamma_\kappa^k\right\|_{\mathcal{L}(\mathbb{H})} < 1$, for each $k \geqslant k_0$ and for some $k_0 \geqslant 1$ (Álvarez-Liébana et al., 2016, Lemma 1).

8.4.2 A two-steps specification test for the Ornstein-Uhlenbeck model

As explained, the ARH(1) characterization of an OU process provided in (8.17) will allow us to propose a two-stage methodology in this subsection with the aim of developing a specification test for diffusion processes. In brief, the underlying idea will be, firstly, test whether a diffusion process $\{X_t\}_{t\in\mathbb{R}^+}$, splitted and characterized as $\{\mathcal{X}_n(t) \coloneqq X_{nh+t},\ t \in [0,h]\}_{n\in\mathbb{Z}}$,

constitutes an ARH(1) process (via Algorithm 5), i.e., we will test whether, for each $n \in \mathbb{Z}$,

$$\mathcal{H}_0^{(1)}\colon \mathcal{X}_n \text{ and } \mathcal{X}_{n-1} \text{ are linearly related via } \Gamma \in \mathcal{L}, \tag{8.18}$$

that is, $\mathcal{X}_n(t) = \Gamma(\mathcal{X}_{n-1})(t) + \mathcal{E}_n(t)$. After the characterization of an ARH(1) process as a FLMFR model in (8.14), the second stage will be to check the parametric form of the linear operator Γ, with the null hypothesis

$$\mathcal{H}_0^{(2)}\colon \Gamma(\mathcal{X})(t) \coloneqq \Gamma_\kappa(\mathcal{X})(t) = e^{-\kappa t}\mathcal{X}(h),$$

for each $\mathcal{X} \in \mathbb{H} = L^2\left([0,h], \mathcal{B}_{[0,h]}, \lambda + \delta_{(h)}\right)$, with

$$\mathcal{X}_n(t) = X_{nh+t} = X_0 e^{-\kappa(nh+t)} + \sigma \int_0^{nh+t} e^{-\kappa(nh+t-s)}\,\mathrm{d}W_s,$$

for each $n \in \mathbb{Z}$, via a functional F-test, as long as the linear null hypothesis $\mathcal{H}_0^{(1)}$ in (8.18) is not rejected, against the alternative of an unspecified FLMFR.

As referred, a F-test will be implemented. In keeping with the classical F-statistic, the functional version proposed by Shen and Faraway (2004) has been implemented, based on the residual sum of squared norm (RSSN) of functional errors (Cuevas et al., 2004),

$$\begin{aligned} \mathrm{RSSN}_n &= \sum_{i=0}^{n-1} \left\| \mathcal{Y}_i - \hat{\mathcal{Y}}_i \right\|_{\mathbb{H}}^2, \\ \mathcal{Y}_i &= \Gamma(\mathcal{X}_i) + \mathcal{E}_i. \end{aligned} \tag{8.19}$$

Thus, the F-statistic is given by

$$F_n = \frac{\mathrm{RSSN}_n^{\mathrm{OU}} - \mathrm{RSSN}_n^{\mathrm{FLMFR}}}{\mathrm{RSSN}_n^{\mathrm{FLMFR}}}, \tag{8.20}$$

where $\mathrm{RSSN}_n^{\mathrm{OU}}$ denotes the RSSN of functional errors under the assumption that the specification of the linear operator is defined as $\Gamma(\mathcal{X})(t) \coloneqq \Gamma_\kappa(\mathcal{X})(t) = e^{-\kappa t}\mathcal{X}(h)$, i.e., under null hypothesis $\mathcal{H}_0^{(2)}$, whereas $\mathrm{RSSN}_n^{\mathrm{FLMFR}}$ denotes the RSSN of functional errors of an unrestricted FLMFR, whose estimator is given by the FPCR-L1S estimator provided in Section 8.2.2. In the former case, and estimator of $\Gamma_\kappa(\mathcal{X})(t) = e^{-\kappa t}\mathcal{X}(h)$ is required, in terms of the maximum likelihood sample-dependent estimator $\hat{\kappa}_n = -(\int_0^{nh} X_t\,\mathrm{d}X_t)/(\int_0^{nh} X_t^2\,\mathrm{d}t)$ proposed in Álvarez-Liébana et al. (2016), whose strong-consistency was therein proved. Henceforth, $\hat{\Gamma}_\kappa(\mathcal{X})(t) \coloneqq \Gamma_{\hat{\kappa}_n}(\mathcal{X})(t) = e^{-\hat{\kappa}_n t}\mathcal{X}(h)$ is considered. In Algorithm 6, a summary of

the testing procedure is presented, via Bonferroni correction to counteract the multiple tests.

Algorithm 6 (Two-stage OU specification test)**.** Let $\{X_t\}_{t\in[0,T]}$ be a stochastic processes on a given filtered space $\left(\Omega, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}\right)$ where X_t is $\mathcal{F}_t$-measurable, for each $t \in [0,T]$, given by a parametric time-homogeneous one dimensional SDE under global Lipschitz and linear growth assumptions and null long term mean. The procedure is as follows.

Stage 1 ($\mathcal{H}_0^{(1)}$): testing if $\{X_t\}_{t\in[0,T]}$ can be reinterpreted as a functional ARH(1) model.

1. Split $\{X_t\}_{t\in[0,T]}$ into a set of functional paths $\{\mathcal{X}_i(t) \coloneqq X_{ih+t},\ t\in[0,h]\}_{i=0,1,\ldots,n-1}$ valued in n subintervals, where $T = nh$ and $\mathcal{X}_i \in L^2\left([0,h], \mathcal{B}_{[0,h]}, \lambda + \delta_{(h)}\right)$, for each $i = 0, 1, \ldots, n-1$.

2. Construct the functional random variables $\widetilde{\mathcal{X}}_i(s) \coloneqq \mathcal{X}_{i-1}(s)$, $\widetilde{\mathcal{Y}}_i(t) \coloneqq \mathcal{X}_i(t)$ and $\widetilde{\mathcal{E}}_i(t) \coloneqq \mathcal{E}_i(t)$, all of them centered, with $\widetilde{\mathcal{Y}}_i = \widetilde{\Gamma}\left(\widetilde{\mathcal{X}}_i\right) + \widetilde{\mathcal{E}}_i$, for any $i = 1, \ldots, n-1$, and $s, t \in [0,h]$.

3. Obtain the FPC of $\left\{\widetilde{\mathcal{X}}_i\right\}_{i=1}^{n-1}$ and $\left\{\widetilde{\mathcal{Y}}_i\right\}_{i=1}^{n-1}$, choose initial (p,q) required for $\text{EV}_p = \text{EV}_q = 0.99$, and achieve their p- and q-truncated FPC scores $\widetilde{\mathbf{X}}_p$ and $\widetilde{\mathbf{Y}}_q$, respectively. We then compute the FPCR-L1S estimator $\widehat{\widetilde{\mathbf{B}}}_{\widetilde{p},q}^{(\lambda),C}$ and the associated residuals. As detailed in Algorithm 5, we compute the statistic $\text{PCvM}_{n,\widetilde{p},q}$ in (8.15), the bootstrapped errors $\left\{\mathbf{e}_{i,q}^{*b}\right\}_{i=1,\ldots,n-1}^{b=1,\ldots,B}$, where B is the number of replicates, and $\text{PCvM}_{n,\widetilde{p},q}^{*b}$ the bootstrapped statistic.

4. Estimate the p-value as $p^{(1)} \coloneqq \#\left\{\text{PCvM}_{n,\widetilde{p},q} \leqslant \text{PCvM}_{n,\widetilde{p},q}^{*b}\right\}/B$. If $\mathcal{H}_0^{(1)}$ in (8.18) is rejected, the procedure stops. Otherwise, go to stage 2.

Stage 2 ($\mathcal{H}_0^{(2)}$): testing if $\{X_t\}_{t\in[0,T]}$, characterized as an ARH(1) model (under linearity given by FLMFR), constitutes an OU process via F-test, against the alternative of an unspecified FLMFR.

5. If $\mathcal{H}_0^{(1)}$ in (8.18) is not rejected, calculate the F-statistic as in (8.20), where RSSN_n is computed as (8.19), with $\text{RSSN}_n^{\text{OU}} = \sum_{i=1}^{n-1}\left\|\widetilde{\mathcal{Y}}_i(t) - e^{-\hat{\kappa}_n t}\widetilde{\mathcal{X}}_i(h)\right\|_{\mathbb{H}}^2$.

6. After computing F_n, to calibrate the test, we implement a parametric bootstrap, by simulating a set of OU processes $\left\{X_t^{*b}\right\}_{t\in[0,T]}^{b=1,\ldots,B}$, driven by $\mathrm{d}X_t^{*b} = -\hat{\kappa}_n X_t^{*b}\,\mathrm{d}t + \hat{\sigma}_n\,\mathrm{d}W_t^{*b}$, for each $b = 1, \ldots, B$, where B is the number of replicates, $\hat{\kappa}_n$ is the detailed estimator of κ, and $\hat{\sigma}_n$ a root-n consistent estimator of σ, e.g.,

$$\hat{\sigma}_n = \arg\min_{\sigma_0} \frac{1}{N-1}\sum_{i=0}^{N-1}\left(\ln\left(\sigma_0^2\right) + \frac{\left(X_{t_{i+1}} - X_{t_i}\right)^2}{\sigma_0^2\Delta}\right)$$

(see Corradi and White, 1999), where $\{t_0, \ldots, t_N\}$ are the grid points in which $[0, T]$ is discretized, with $T = nh$, and Δ the discretization step.

7. With the bootstrapped $\left\{X_t^{*b}\right\}_{t\in[0,T]}^{b=1,\ldots,B}$, we recompute

$$F_n^{*b} = \frac{\text{RSSN}_n^{OU,*b} - \text{RSSN}_n^{FLMFR,*b}}{\text{RSSN}_n^{FLMFR,*b}},$$

such that an estimator $\hat{\kappa}_n^{*b}$ is required to be recomputed, as well as a FPCR-L1S estimator for $\Gamma_{\hat{\kappa}_n}$, for each $b = 1, \ldots, B$.

8. Estimate the p-value as $p^{(2)} := \#\left\{F_n \leqslant F_n^{*b}\right\}/B$.

Remark 4 (Bonferroni correction: stage 1 + stage 2)**.** As we are conducting multiple tests, given the corresponding p-values $p^{(1)}$ and $p^{(2)}$ for testing the hypotheses $\mathcal{H}_0^{(1)}$ and $\mathcal{H}_0^{(2)}$, the Bonferroni correction is used to set an upper bound on the significance level α (Miller, 1981), by rejecting $\mathcal{H}_0 = \{\mathcal{H}_0^{(1)}, \mathcal{H}_0^{(2)}\}$ if any p-value $p^{(1)}$ or $p^{(2)}$ is less than $\alpha/2$. As referred, the second stage in Algorithm 6 is omitted if $\mathcal{H}_0^{(1)}$ is rejected, respect to $\alpha/2$.

8.4.3 Simulation study

The finite sample properties of the specification test is illustrated with $1\,000$ Monte Carlo replicates and $B = 1\,000$ bootstrap resamples, with $n \in \{50, 150, 250, 350, 500, 1000, 5000\}$, such that trajectories $\{\mathcal{X}_i(t) := X_{ih+t}\}_{i=0}^{n-1}$ are valued in 101 equispaced points in $[0, 1]$. The FPCR-L1S estimator was used with $\text{EV}_p = \text{EV}_q = 0.995$. As reflected in Table 8.3, several scenarios have been simulated. Under $\mathcal{H}_0$, we simulate sample paths for the centered OU model (see Table 8.3). On the other hand, concerning the power, we simulate different interest rate models, as the CKLS model (Chan et al., 1992), the Inverse Feller (IF) model (Ahn and Gao, 1999) and the Aït-Sahalia (AS) model (Aït-Sahalia, 1996c). Similar parameters of the CKLS model to those ones proposed in Hong and Li (2005), to examine different persistent dependence and volatility, were tested. The IF and AS models are given by $\mathrm{d}X_t = X_t(\kappa - (\sigma^2 - \kappa\mu)X_t)\,\mathrm{d}t + \sigma X_t^{3/2}\,\mathrm{d}W_t$ and $\mathrm{d}X_t = (\tau_{-1}X_t^{-1} + \tau_0 + \tau_1 X_t + \tau_2 X_t^2)\,\mathrm{d}t + \sigma X_t^{3/2}\,\mathrm{d}W_t$, with similar set of parameters as proposed in Ahn and Gao (1999) and Aït-Sahalia (1999), respectively (see Table 8.3). For testing deviations from the null hypothesis, we also consider the radial OU process[1], given by $\mathrm{d}X_t = (\lambda X_t^{-1} - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, which corresponds with the OU if $\lambda = 0$. A model without drift (denoted as null model) has been also tested, given by $\mathrm{d}X_t = \sigma\,\mathrm{d}W_t$, with $\sigma \in \{0.1, 0.5\}$.

[1]The drift function is non-Lipschitz and unbounded, however, there exist a unique weak SDE solution.

Notation	Description	Model	Parameters and scenarios
$\mathcal{H}_0$	OU (centered)	$\mathrm{d}X_t = -\kappa X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$	$\kappa \in \{0.2, 0.5, 0.8\}$ $\sigma^2 \in \{0.008, 0.05, 0.15, 0.50, 0.75\}$
$\mathcal{H}_{1,Null}$	Null	$\mathrm{d}X_t = \sigma\,\mathrm{d}W_t$	S1: $\sigma = 0.1$ S2: $\sigma = 0.5$
$\mathcal{H}_{1,IF}$	Inverse-Feller	$\mathrm{d}X_t = X_t(\kappa - (\sigma^2 - \kappa\mu)X_t)\,\mathrm{d}t + \sigma X_t^{3/2}\,\mathrm{d}W_t$	$(\kappa, \mu, \sigma^2) = (0.364, 0.08, 1.6384)$
$\mathcal{H}_{1,AS}$	Aït-Sahalia	$\mathrm{d}X_t = (\tau_{-1}X_t^{-1} + \tau_0 + \tau_1 X_t + \tau_2 X_t^2)\,\mathrm{d}t + \sigma X_t^{3/2}\,\mathrm{d}W_t$	$(\tau_{-1}, \tau_0, \tau_1, \tau_2, \sigma) = (0.00107, -0.0517, 0.877, -4.604, 0.8)$
$\mathcal{H}_{1,CKLS}$	CKLS	$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t$	S1: $(\mu, \kappa, \sigma, \gamma) = (0.09, 0.9, 0.5, 1.5)$ S2: $(\mu, \kappa, \sigma, \gamma) = (0.09, 0.2, 1.5, 1.5)$ S3: $(\mu, \kappa, \sigma, \gamma) = (0.09, 0.2, 3, 1.5)$
$\mathcal{H}_{1,ROU}$	Radial OU	$\mathrm{d}X_t = (\lambda X_t^{-1} - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$	S1: $(\lambda, \kappa, \sigma) = (0.05, 0.1, 0.5)$ S2: $(\lambda, \kappa, \sigma) = (0.075, 0.1, 0.5)$ S3: $(\lambda, \kappa, \sigma) = (0.1, 0.1, 0.5)$ S4: $(\lambda, \kappa, \sigma) = (0.125, 0.1, 0.5)$

Table 8.3: Summary of simulated scenarios.

Tables 8.4–8.5 display the empirical sizes and powers, respectively, with $\alpha = 0.05$ as nominal level. In the former table, regarding the calibration of test, the empirical sizes are close to α through all values of σ and κ. As expected, scenarios with $n = 50$ exhibit over-rejection for larger κ values, owing to κ represents the speed of reversion at which trajectories are rearranged around μ, and then, greater values of κ lead to more similar paths to be discriminated, since $\lim_{t\to\infty} \mathbb{V}\mathrm{ar}\,[X_t] = \sigma^2/(2\kappa)$ constitutes the long term variance. This behavior is overcome as n increases, as just 2 of the 90 scenarios (2.22% of cases) with $n \in \{150, 250, 350, 500, 1000, 5000\}$ display slightly over-rejections (rejection rates are bold-faced if they lie outside a 95%-confidence interval for $\alpha = 0.05$). Note that sample size in the functional data framework refers to the number of curves, the actual number of observations in the simulation reaches $500\,000$ for the $n = 5\,000$ scenario, as the curves are evaluated in 101 equispaced points.

With respect to empirical power, displayed in Table 8.5, deviations from the null by augmenting λ in the radial OU show the increasing power of the test as n increases. The null model also exhibits high empirical powers, just like the nonlinear drift alternatives, specially the IF model, displaying great empirical powers even for small sample sizes. Concerning the CKLS model, all the scenarios provide increasing rejection rates, such that the higher the volatility parameter, the more empirical power. A slightly poor performance is obtained for

the more complex AS model, although empirical powers seem to tend to one as n increases.

		n						
σ^2	κ	50	150	250	350	500	1000	5000
0.008	0.2	0.048	0.043	0.049	0.053	0.045	0.051	0.058
	0.5	0.047	0.053	0.048	0.046	0.050	0.049	0.052
	0.8	**0.078**	0.048	0.051	0.056	0.042	0.058	0.054
0.05	0.2	0.043	0.051	0.054	0.048	0.053	0.052	0.061
	0.5	0.051	0.057	0.055	0.055	0.061	0.048	0.054
	0.8	**0.102**	0.060	0.052	**0.075**	0.057	0.053	0.044
0.15	0.2	0.050	0.045	0.048	0.053	0.046	0.051	0.060
	0.5	0.045	0.053	0.053	0.046	0.050	0.050	0.048
	0.8	**0.078**	0.051	0.052	0.056	0.043	0.056	0.056
0.50	0.2	0.051	0.046	0.053	0.046	0.051	0.052	0.058
	0.5	0.050	0.056	0.059	0.054	0.053	0.051	0.061
	0.8	**0.096**	**0.065**	0.057	0.059	0.056	0.059	0.053
0.75	0.2	0.046	0.044	0.049	0.052	0.047	0.051	0.053
	0.5	0.048	0.054	0.050	0.047	0.051	0.048	0.052
	0.8	**0.080**	0.058	0.053	0.054	0.041	0.057	0.050

Table 8.4: Size simulation. Empirical rejection rates under the null hypothesis. The rejection rates are boldfaced if they lie outside a 95%-confidence interval for $\alpha = 0.05$.

	n						
Scenario	50	150	250	350	500	1000	5000
$\mathcal{H}^{S1}_{1,ROU}$	0.043	0.145	0.241	0.340	0.481	0.818	0.945
$\mathcal{H}^{S2}_{1,ROU}$	0.060	0.232	0.390	0.530	0.744	0.956	0.989
$\mathcal{H}^{S3}_{1,ROU}$	0.070	0.280	0.487	0.722	0.900	0.976	1.000
$\mathcal{H}^{S4}_{1,ROU}$	0.065	0.276	0.586	0.807	0.945	0.986	1.000
$\mathcal{H}^{S1}_{1,Null}$	0.097	0.322	0.583	0.741	0.907	1.000	1.000
$\mathcal{H}^{S2}_{1,Null}$	0.108	0.317	0.566	0.765	0.919	1.000	1.000
$\mathcal{H}_{1,IF}$	0.455	0.750	0.836	0.870	0.928	0.976	1.000
$\mathcal{H}_{1,AS}$	0.139	0.370	0.499	0.611	0.756	0.958	1.000
$\mathcal{H}^{S1}_{1,CKLS}$	0.179	0.277	0.360	0.488	0.667	0.763	0.886
$\mathcal{H}^{S2}_{1,CKLS}$	0.319	0.635	0.727	0.738	0.811	0.895	0.994
$\mathcal{H}^{S3}_{1,CKLS}$	0.690	0.933	0.985	0.993	1.000	1.000	1.000

Table 8.5: Power simulation. Empirical rejection rates under the alternative hypothesis, from scenarios displayed in Table 8.3.

8.5 Real data applications

Motivated by the extensive use of diffusion processes in finance, commonly used to model currency exchange rates (Ball and Roma, 1994) and intra-day patterns in the foreign exchange market (Andersen and Bollerslev, 1997), we now illustrate a real data application of the proposed OU specification test, in the context of high-frequency financial data (Reboredo et al., 2012). Unlike univariate frameworks, where a vast time window is required for getting properly sample sizes, and thus, certain properties essential for long horizon asymptotics (e.g., ergodicity) would not be achieved, we consider a finite time observation window, where the functional data scheme allows to capture the dynamic of the process (Müller et al., 2011). It is also noteworthy that, due to the high-frequency scheme, observations are subject to market microstructure noise (Aït-Sahalia et al., 2005), interacting with the sampling frequency which could lead to the rejection of the null hypothesis.

The three datasets considered consist on currency pair rates of Euro-British pound (EURGBP), Euro-US Dollar (EURUSD) and British pound-US Dollar (GBPUSD), determined in the foreign exchange market, such that data was recorded every 5 minutes from January 1, 2019 to December 31, 2019. Figure 8.3 shows the exchange rates (left) with $73\,440$ observations, and the splitted paths (right) $\{\mathcal{X}_i(t)\}_{i=1}^n$ featuring $n = 255$ daily curves. The daily curves are valued in $\mathbb{H} = L^2\left([0,1], \mathcal{B}_{[0,1]}, \lambda + \delta_{(1)}\right)$, where the interval $[0,1]$ accounts for a 1-day window, discretized in 288 equispaced grid points. Table 8.6 shows the empirical p-values for the three datasets, with $B = 1\,000$ bootstrap resamples, at each of the two-stages defined in Algorithm 6. In this way, we test at stage one the null hypothesis that the daily curves $\{\mathcal{X}_i(t)\}_{i=1}^n$ constitute an ARH(1) process,

$$\mathcal{H}_0^{(1)}\colon \mathcal{X}_n \text{ and } \mathcal{X}_{n-1} \text{ are linearly related via } \Gamma \in \mathcal{L},$$

with $\mathcal{X}_n(t) = \Gamma(\mathcal{X}_{n-1})(t) + \mathcal{E}_n(t)$. At stage two, a F-test is implemented to test the parametric form of the OU process,

$$\mathcal{H}_0^{(2)}\colon \Gamma(\mathcal{X})(t) \coloneqq \Gamma_\kappa(\mathcal{X})(t) = e^{-\kappa t}\mathcal{X}(h).$$

Regarding the p-values from Table 8.6, the null hypothesis of OU process in the EURGBP series is rejected, at a 5% level, since it does not seem to follow an stationary ARH(1) process (stage one) when the annual trajectory is considered as daily curves. However, it is not rejected for the EURUSD and GBPUSD exchange rates, where a simple model with constant volatility function as the OU seems to capture the dependence of the series. The nature of the

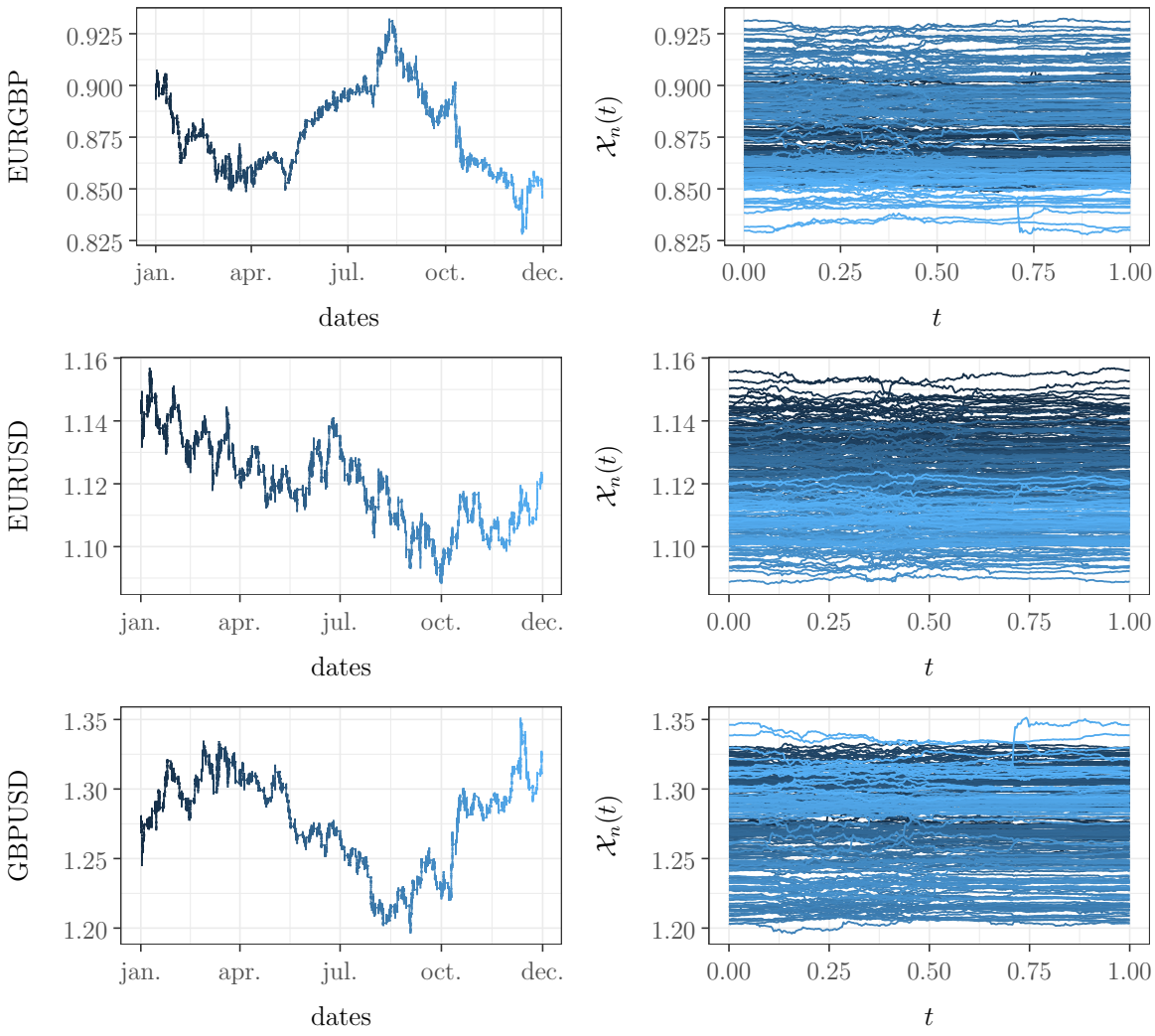


Figure 8.3: From top to bottom: EURGBP, EURUSD and GBPUSD exchange rates throughout 2019. On the left, the observed stochastic process; on the right, daily splitted paths.

sampling mechanism could generate a different result, as classical data analysis for diffusion processes deals with monthly, weekly or daily data, at most. As the sampling frequency is not dictated by the data, usually a large fraction of data is discarded, not without loss of information. In our data example, sampling daily would mean to retain only 255 observations, from a total of 73 440, working with one observation a day instead of the whole daily trajectory. Therefore, when dealing with intra-day data, conducting the study using a functional framework takes advantage of the information retained in the shape of the curve, treating the daily curves as a statistical object, instead of a collection of individual observations.

Two-stages test	EURGBP	EURUSD	GBPUSD
Stage 1: ARH(1) GoF test	0.020	0.339	0.091
Stage 2: F-test	0.349	0.253	0.510

Table 8.6: p-values under the null hypothesis that the currency exchange rates series follow a centered OU process.

8.6 Conclusions

In this chapter a novel goodness-of-fit test for ARH(z) models, against an unspecified alternative, was proposed based on the their characterization as FLMFR. No competitors, against an unrestricted alternative, could be found, but empirical results under different alternative were compared with Kokoszka and Reimherr (2013) procedure (KR) on determining the order z. Outcomes showed that the proposal is well calibrated under the null hypothesis and exhibited a great power through several alternative scenarios, being competitive even against nonlinear alternatives where the KR test cannot reach a suitable power.

On the other hand, a novel specification test for the Vasicek model was proposed, firstly characterizing it as an ARH(1) model, and secondly, testing the parametric form, under linearity, via functional F-test calibrated by a parametric bootstrap. The finite sample behaviour was also illustrated, generating the data under parametric families of diffusion processes, providing the empirical evidence that the two-stages specification test is well-calibrated and detects different alternative models widely used in finance. As traditional data analysis ignores the information contained in the shape of the daily curves, the functional data approach could be more suitable when working with intra-day or high-frequency data, as it can extract the additional information of the curves. The classical diffusion process framework regards the whole observation sequence as one sample path, ignoring the pattern of the process across days. Working in a functional data setting, we take the advantage of the information retained in the shape of the curve, treating the daily curves as a statistical object, instead of a collection of individual observations. The proposed tests are illustrated with an application to daily currency exchange rates curves, with 5-minute intervals, where it was concluded that the Vasicek model does capture the underlying dynamic changes of the daily curves for the EURUSD and GBPUSD pairs.

Future applications could be developed, since the referred specification test may be extended to other diffusion processes as long as they may be characterized as ARH(z) models. The tests here derived may be also extended to Hilbertian moving-average processes and ARH processes with exogenous variables.

Chapter 9

Conclusions and discussion

This dissertation was dedicated to the development of new goodness-of-fit tests for diffusion models and the study of estimation methods, aiming to model the empirical behavior of interest rates. In this chapter, the results obtained through the dissertation chapters are summarized, along with open questions requiring further research.

Contents

9.1 Results and discussion

As Chapters 1 and 2 provided an introduction and theoretical background, respectively, this section includes a discussion of Chapters 3 to 8.

Chapter 3: Parametric estimation of diffusion models

As discretization and estimation bias can hinder the test procedures, the first objective of this dissertation is to provide a discussion on estimation methods. The first study addressed the parametric estimation of diffusion processes with deterministic volatility function. The parameters of continuous-time diffusion processes are subject to estimation bias, as data are highly persistent, and discretization bias, as data are discretely sampled despite the continuous-time nature of the model. Chapter 3 assessed the role of persistence and the impact of sampling frequency on the estimation of the drift and diffusion functions. A simulation study

under different settings was discussed, finding that the estimation bias was large for the drift parameter that controls the speed of mean reversion, with higher persistence increasing this bias. Regarding the discretization bias, it was discerned in low persistence scenarios with coarse discretization frequency. Unlike discrete-time series, where the bias and variance of estimators is controlled by the sample size n, in continuous-time models sampled at discrete time points bias and variance in the drift parameter estimation is dominated by the total observation time $T = n\Delta$. Thus, estimation bias and variance are reduced as $T \to \infty$, rather than when only $n \to \infty$. The estimation procedures reviewed provided similar performances, with the generalized method of moments being the least efficient and the simulation-based methods providing a highly model-dependent implementation. As a contribution, the estimation methods used in this chapter were implemented in a R package, `estsde`.

Chapter 4: Parametric estimation of diffusion models with stochastic volatility

Chapter 4 extends the estimation study to diffusion models with stochastic volatility. This two-factor models provide a more flexible specification but at the expense of more intricate estimation procedures, as the volatility is non observable. To study the estimation of this latent variable, a Monte Carlo simulation study was carried out for different estimation methods, extending some of the procedures reviewed in Chapter 3 for one-factor models to two-factor diffusion processes. The close performance of the procedures, together with the computational demanding implementation of simulation methods, lead to conclude that the Kalman filter was a computational efficient estimation method to use in goodness-of-fit testing procedures. Furthermore, the space-state model structure allows to easily implement a bootstrap procedure to calibrate the tests.

Chapter 5: Nonparametric estimation of diffusion models

As Chapters 3 and 4 focused on parametric estimation, Chapter 5 addressed the non parametric estimation of the different components of the diffusion model, namely the drift, diffusion and stationary density estimation. The common used bandwidth in the econometric field is obtained using the rule of thumb, so this chapter deployed a simulation study to compare the performance of various bandwidths under different settings. The rule of thumb bandwidth showed a similar performance to the cross-validation bandwidth when estimating the marginal density function. However, this bandwidth was outperformed in the drift and diffusion functions estimation. Using cross-validation bandwidths or using a pilot bandwidth with a bimodal kernel function satisfying $K(0) = 0$—proposed in De Brabanter et al. (2018)

for a random design setting with correlated errors—yielded lower MISE. The level of persistence hinders the estimation of the components, as discussed in Chapter 3 for the parametric estimation, which may indicate that taking the mean reversion speed into account when computing the smoothing parameter could provide a bandwidth selection criteria adapted for econometric diffusion models. This is an open research line and, though this problem has been addressed in the literature, not many proposals have been studied.

Chapter 6: Goodness-of-fit test for diffusion models

With the estimation study groundwork laid in Chapters 3–5, the goodness-of-fit of diffusion models was addressed hereinafter. Chapter 6 starts collecting the developments of the last 25 years on goodness-of-fit, implementing the methods to illustrate their finite sample behavior. A new goodness-of-fit test for the parametric specification of diffusion models is then proposed, based on the ideas of distance correlation (Székely et al., 2007), and its performance is compared to the reviewed procedures, analyzing the effect of the curse of dimensionality. Empirical results showed that the performance of the procedures, in terms of size and power, are close, calibrating adequately the null hypothesis and detecting a variety of alternatives. However, when dimension increases, some tests lose power. In this scenario is where the correlation distance proposal outperforms others, avoiding the curse of dimensionality. An application of the tests to interest rate models involving U.S. daily treasury securities with different maturities, ranging from 1 month to 30 years, was also undertaken, finding empirical support to the CKLS diffusion function for long maturities. As the maturity range was reduced, the p-values became small, indicating that the model was mis-specified for T-bill (1,3,6 and 12 month maturities) series.

Chapter 7: Goodness-of-fit test for diffusion models with stochastic volatility

In Chapter 7, a goodness-of-fit test for diffusion models with stochastic volatility was proposed, testing the parametric form of the drift and diffusion functions. The test for the drift function is based on the integrated regression function of the process, while the test for diffusion function relies on the integrated volatility function. The test statistics are based on a distance of the resulting residual marked empirical processes from their expected zero mean, measured by Kolmogorov-Smirnov and Cramér-von Mises functionals. In both tests, the distribution of the test statistic is approximated by bootstrap techniques, which rely on the Kalman filter algorithm introduced in Chapter 4. A simulation study illustrated the finite sample properties of the proposed tests, which did not include a comparison with other pro-

cedures as, to the best of our knowledge, the tests available in the literature for continuous-time stochastic volatility models do not test the same null hypothesis as our test, that is, the parametric form of the diffusion function. The tests showed great power through several alternative hypotheses and was well calibrated under null hypotheses and its implementation, regarding the computation of the test statistic and the bootstrap resampling scheme, is quite straightforward. The application to the Euribor series demonstrated that the incorporation of stochastic volatility to diffusion models does capture the features of interest rate time series, unlike deterministic volatility functions, suggesting that the volatility depends on an additional factor that varies independently of the short rate level.

Chapter 8: Specification test for diffusion models as functional time series

Lastly, Chapter 8 deals with diffusion models in a different framework, proposing a novel specification test for diffusion models as functional time series. In this chapter, the diffusion model is used in a high-frequency setting, as these data can be collected as a sequence of curves over time, like intra-day prices. In the previous chapters of this dissertation, the whole observation sequence of the diffusion process was regarded as one sample path, ignoring the pattern of the process across days. Working in a functional data setting, we take advantage of the information retained in the shape of the curve, treating the daily curves as a statistical object, instead of a collection of individual observations. The contribution of this chapter is twofold: a goodness-of-fit test for autoregressive Hilbertian (ARH) time series and a specification test for diffusion models characterized as ARH models.

Regarding the goodness-of-fit test for functional time series, the proposed test extends the ideas in García-Portugués et al. (2021) to ARH(z) processes, with $z \in \mathbb{Z}$, testing a composite null hypothesis for functional time series in terms of a Cramér-von Mises statistic, such that no assumptions concerning alternatives are required. The novelty of this contribution is twofold: *(i)* a new goodness-of-fit test for ARH(z) processes; *(ii)* a sequential procedure to determine the order z of an ARH(z) model. As the former contribution has no competitors, the latter procedure was compared against the multistage order detection procedure in Kokoszka and Reimherr (2013). The simulation study showed that the proposal was well calibrated under the null hypothesis and exhibited a great power through several alternative scenarios, being competitive even against nonlinear alternatives where the Kokoszka and Reimherr (2013) test cannot reach a suitable power.

Extending the goodness-of-fit test for functional time series to diffusion processes, a new specification test for diffusion models from a high-dimensional data perspective was proposed, namely for Ornstein-Uhlenbeck processes, which are characterized as a particular

ARH(1) process. The test is a two-stages methodology where firstly, the Ornstein-Uhlenbeck model is characterized as an ARH(1), and secondly, under linearity, the parametric form of the model is tested by means of a functional F-test calibrated by a parametric bootstrap. The finite sample behavior, studied under different parametric families of diffusion processes, provided empirical evidence that the two-stages specification test was well calibrated and detected different alternative models widely used in finance. The application of the test to daily currency exchange rates curves, with 5-minute intervals, showed that a simple model with constant volatility function as the Ornstein-Uhlenbeck process did capture the underlying dynamic changes of the daily curves for the EURUSD and GBPUSD pairs.

9.2 Further research

The specification test proposed in Chapter 8 for diffusion processes as functional time series was focused on the Ornstein-Uhlenbeck process. Even though it can be extended to other specifications of diffusion models as long as they can be characterized as temporally correlated functional data, the question of whether there exist other diffusion models that can be characterized as ARH(1) remains open. Furthermore, the development of a fully functional procedure, as well as the theoretical formulation of the asymptotic distribution of the proposed test statistic is deferred to future research.

The availability of inference tools for diffusion processes as functional time series opens the possibility of analyzing the term structure of interest rates from a functional perspective. The functional framework is a natural setting to model yield curves, which show the relationship between the yields (interest rates) of a financial asset and their respective maturities. The yield curve has a track record of predicting recessions over the past 50 years, as inverted curves (long-term bond yields lower than short-term yields) have appeared 5 to 16 months before each of the last seven recessions, giving markets an early warning signal before the onset of an economic crisis. The main information of the yield curves is retained in its shape, which is an fundamental economic indicator of future rates and inflation expectations. The analysis of the yield curves has been developed in classical contexts where the shape is ignored. Treating the curves as an statistical object instead of a collection of individual observations opens a bridge between economic theory, that relates different types of yield curves to the economic situation and prospects, with statistical modeling.

The European Central Bank estimates zero coupon yield curves for the eurozone. See Figure 9.1, with daily yield curves in gray from 2006 to 2009, which expands before and after the outbreak of the economic crisis of 2008. The functional data analysis framework

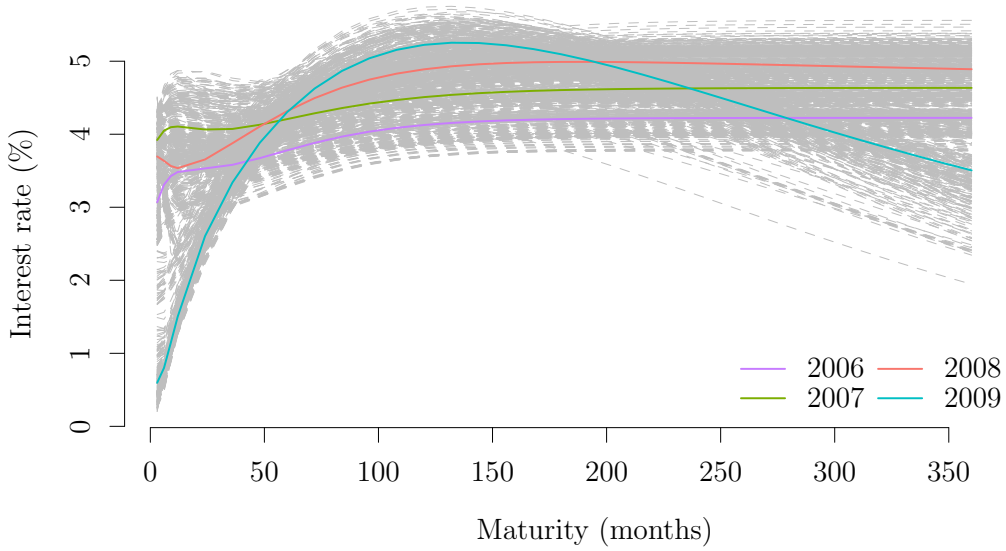


Figure 9.1: Daily yield curves (dashed gray) from 2006 to 2009 and mean curve (solid color) for each year.

allows us to calculate the mean curve for each year. Figure 9.1 shows the mean curve for each year (in solid color). For the year 2006, the mean has a somewhat flatter shape than what it is considered a normal shape, but it has a positive slope and the interest rate level is not very high. In 2007 the curve is almost flat with a relatively high level, so that, on average, the interest rate hardly fluctuates for different maturities. Flat curves indicate uncertainty in the economy and the statistical analysis matches the economic theory, as 2007 is the year prior to the crisis outburst. The 2008 mean has a negative slope for short-term maturities, so that the interest rate for very short-term maturities is higher than for medium-term maturities. Finally, the 2009 mean shows an unusual shape, where medium-term interest rates are higher than short and long-term and with very low short-term interest rates, which would indicate expectations of an economic slowdown. The analysis of the mean for the years 2006 to 2009 seems to comply with economic theory and reflect the financial situation before and after the 2008 financial crisis.

The prospect of extending the use of a functional setting for modeling yield curves is supported by the increasing availability of goodness-of-fit test tools, as the proposed specification test in Chapter 8. For that matter, further research needs to be done.

Another research line related to the functional framework, is the study of stock pinning by testing the local volatility surface, that is, to test whether the volatility $\mathcal{X}(t) = \sigma(t, S_t)$, with S_t the underlying asset value, influences the instantaneous volatility Y. This will allow us to assess the stock pinning phenomenon, which occurs when the stock prices are at or

near option strike prices minutes before the option is due to expire. For this purpose, the volatility $\mathcal{X}(t)$ can be considered as functional data, so that a goodness-of-fit test for the null hypothesis of the functional linear model with scalar response could be performed,

$$\mathcal{H}_0 \colon Y = \langle \mathcal{X}, \beta \rangle + \varepsilon = \int \mathcal{X}(t)\beta(t)\,\mathrm{d}t + \varepsilon.$$

This is equivalent to testing

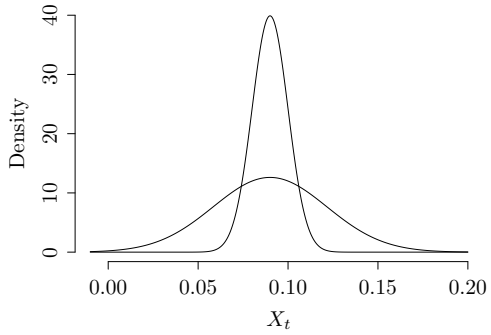
$$\begin{cases} \mathcal{H}_0 \colon m \in \{\langle \cdot, \beta \rangle \colon \beta \in \mathbb{H}\}, \\ \mathcal{H}_a \colon \mathbb{P}\big(m \notin \{\langle \cdot, \beta \rangle \colon \beta \in \mathbb{H}\}\big) > 0, \end{cases}$$

where $m(\mathcal{X}) = \mathbb{E}\left[\mathcal{X} \mid Y\right] = \langle \mathcal{X}, \beta \rangle$ and with $\mathbb{H} = L^2[0, T]$ the Hilbert space of integrable square functions. Therefore, we have the random variables $(\mathcal{X}, Y)$ in $\mathbb{H} \times \mathbb{R}$ and the sample $\{\mathcal{X}_i, Y_i\}_{i=1}^n$, where Y is the instantaneous volatility and $\mathcal{X}$ the volatility curves. This analysis of stock pinning and the study of the local volatility surface from a functional data analysis perspective remains open.

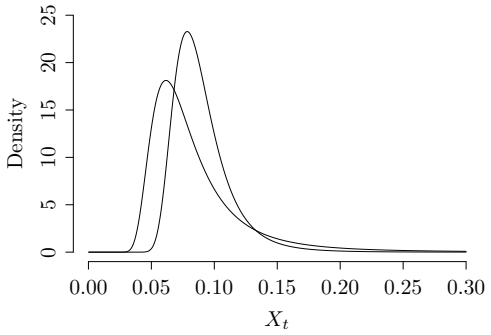
Appendix A

Supplementary material for Chapter 3

This appendix contains the supplementary material for Chapter 3, whose contents appear on López-Pérez et al. (2021). The appendix is divided into two sections, which contain the tabulated results of the simulation study for the Vasicek (Section A.1) and CKLS (Section A.2) models. The Vasicek scenarios included here are 2, 4, 6 and 8 (see Table 3.1), which all of them have the same marginal distribution (see Figure A.1a). Appendix A.2 provides implementation details for the Hermite polynomial expansion procedure (see Section 3.2.3) for the CKLS model, as well as simulation results for scenarios 2 and 4 of the CKLS model (see Table 3.1).



(a) Marginal density for Vasicek model, scenario 1 (3, 5 and 7) and 2 (4, 6 and 8).



(b) Marginal density for CKLS model, scenario 1 and 2.

Figure A.1: Marginal densities for the Vasicek and CKLS models.

A.1 Vasicek model simulations

In this section, the tabulated results for the Vasicek model are included: scenario 2 (Table A.1), low mean reversion and high volatility; scenario 4 (Table A.2), high mean reversion

and high volatility; and scenario 6 (Table A.3) and 8 (Table A.4), with very high mean reversion parameters, increased by a factor of 25 and 49, respectively, from scenarios 1 and 2, keeping the same marginal density (see Figure A.1a).

Scenario 2	θ	**EML**	**DML**	**LL**	**HP**	**KF**	**MCMC**	**GMM**
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	$\mathbf{0.0938}$	0.0938	0.0943	0.0938	0.0939	0.0941	0.0960
	$\hat{\kappa}$	0.6900	0.6812	$\mathbf{0.6781}$	0.6875	0.6804	0.6899	0.7161
	$\hat{\sigma}$	0.0200	0.0199	0.0199	$\mathbf{0.0200}$	0.0199	0.0201	0.0199
SD	$\hat{\mu}$	$\mathbf{0.0363}$	0.0366	0.0412	0.0363	0.0370	0.0424	0.0578
	$\hat{\kappa}$	0.4817	$\mathbf{0.4715}$	0.4758	0.4784	0.4719	0.482	0.5101
	$\hat{\sigma}$	6.27×10^{-4}	6.22×10^{-4}	6.18×10^{-4}	6.27×10^{-4}	$\mathbf{6.18\times10^{-4}}$	6.28×10^{-4}	7.14×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0365}$	0.0368	0.0415	0.0365	0.0372	0.0426	0.0581
	$\hat{\kappa}$	0.6871	0.6737	0.6745	0.6830	$\mathbf{0.6733}$	0.6873	0.7257
	$\hat{\sigma}$	6.28×10^{-4}	6.29×10^{-4}	6.23×10^{-4}	6.28×10^{-4}	$\mathbf{6.21\times10^{-4}}$	6.33×10^{-4}	7.22×10^{-4}
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	$\mathbf{0.0900}$	0.0899	0.0897	0.0900	0.0900	0.0899	0.0900
	$\hat{\kappa}$	0.2862	0.2838	$\mathbf{0.2796}$	0.2858	0.2836	0.2848	0.2856
	$\hat{\sigma}$	$\mathbf{0.0200}$	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200
SD	$\hat{\mu}$	$\mathbf{0.0145}$	0.0146	0.0206	0.0145	0.0146	0.0146	0.0146
	$\hat{\kappa}$	0.1283	0.1262	0.1302	0.1277	$\mathbf{0.1261}$	0.1271	0.1275
	$\hat{\sigma}$	2.63×10^{-4}	2.62×10^{-4}	2.62×10^{-4}	2.63×10^{-4}	$\mathbf{2.62\times10^{-4}}$	2.64×10^{-4}	2.67×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0145}$	0.0146	0.0206	0.0145	0.0146	0.0146	0.0146
	$\hat{\kappa}$	0.1546	0.1515	0.1526	0.1538	$\mathbf{0.1513}$	0.1528	0.1536
	$\hat{\sigma}$	$\mathbf{2.63\times10^{-4}}$	2.65×10^{-4}	2.64×10^{-4}	2.63×10^{-4}	2.64×10^{-4}	2.65×10^{-4}	2.68×10^{-4}
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	$\mathbf{0.0906}$	0.0906	0.0892	0.0906	0.0906	0.0906	0.0906
	$\hat{\kappa}$	0.2972	0.2925	$\mathbf{0.2856}$	0.2972	0.2924	0.2977	0.3081
	$\hat{\sigma}$	$\mathbf{0.0200}$	0.0198	0.0198	0.0200	0.0198	0.0201	0.0199
SD	$\hat{\mu}$	$\mathbf{0.0158}$	0.0158	0.0235	0.0158	0.0158	0.0158	0.0172
	$\hat{\kappa}$	0.1388	$\mathbf{0.1345}$	0.1436	0.139	0.1346	0.1389	0.1462
	$\hat{\sigma}$	6.28×10^{-4}	6.17×10^{-4}	6.21×10^{-4}	6.28×10^{-4}	$\mathbf{6.15\times10^{-4}}$	6.24×10^{-4}	7.09×10^{-4}
RMSE	$\hat{\mu}$	$\mathbf{0.0158}$	0.0158	0.0235	0.0158	0.0158	0.0158	0.0172
	$\hat{\kappa}$	0.1695	$\mathbf{0.1632}$	0.1672	0.1696	0.1633	0.1699	0.1818
	$\hat{\sigma}$	6.29×10^{-4}	6.51×10^{-4}	6.56×10^{-4}	6.29×10^{-4}	6.38×10^{-4}	$\mathbf{6.26\times10^{-4}}$	7.20×10^{-4}

Table A.1: Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)^\top = (0.09, 0.2, 0.02)^\top$, scenario 2: low mean reversion and high volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

Scenario 4	θ	EML	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	0.0907	0.0906	0.0888	0.0907	0.0907	**0.0906**	0.0907
	$\hat{\kappa}$	1.3214	1.2997	**1.2717**	1.3207	1.2992	1.3239	1.4695
	$\hat{\sigma}$	0.0425	0.0420	0.0419	0.0425	0.0420	**0.0425**	0.0423
SD	$\hat{\mu}$	0.0155	0.0155	0.0190	0.0155	0.0155	**0.0154**	0.0179
	$\hat{\kappa}$	0.6079	**0.5882**	0.6329	0.6077	0.5886	0.6069	0.7586
	$\hat{\sigma}$	1.33×10^{-3}	1.31×10^{-3}	$\mathbf{1.31\times10^{-3}}$	1.33×10^{-3}	1.31×10^{-3}	1.32×10^{-3}	1.64×10^{-3}
RMSE	$\hat{\mu}$	0.0155	0.0155	0.0190	0.0155	0.0155	**0.0155**	0.0179
	$\hat{\kappa}$	0.7397	0.7112	0.7340	0.7391	**0.7112**	0.7403	0.9486
	$\hat{\sigma}$	1.34×10^{-3}	1.39×10^{-3}	1.40×10^{-3}	1.33×10^{-3}	1.39×10^{-3}	$\mathbf{1.33\times10^{-3}}$	1.65×10^{-3}
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	**0.0900**	0.0900	0.0876	0.0900	0.0900	0.0900	0.0900
	$\hat{\kappa}$	0.9757	0.9631	**0.8910**	0.9739	0.9635	0.9722	0.9716
	$\hat{\sigma}$	0.0425	0.0421	0.0420	0.0425	0.0421	**0.0424**	0.0423
SD	$\hat{\mu}$	**0.0068**	0.0068	0.0106	0.0068	0.0068	0.0068	0.0069
	$\hat{\kappa}$	0.2141	**0.2085**	0.2752	0.2131	0.2085	0.2120	0.2144
	$\hat{\sigma}$	5.59×10^{-4}	5.52×10^{-4}	5.56×10^{-4}	5.60×10^{-4}	$\mathbf{5.52\times10^{-4}}$	5.57×10^{-4}	5.72×10^{-4}
RMSE	$\hat{\mu}$	**0.0068**	0.0068	0.0109	0.0068	0.0068	0.0068	0.0069
	$\hat{\kappa}$	0.2271	**0.2178**	0.2754	0.2256	0.2180	0.2240	0.2260
	$\hat{\sigma}$	5.61×10^{-4}	6.57×10^{-4}	6.84×10^{-4}	5.61×10^{-4}	6.49×10^{-4}	$\mathbf{5.57\times10^{-4}}$	5.97×10^{-4}
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	**0.0903**	0.0903	0.0884	0.0903	0.0903	0.0903	0.0904
	$\hat{\kappa}$	0.9877	0.9465	**0.9157**	0.9886	0.9465	0.9867	1.0457
	$\hat{\sigma}$	0.0425	0.0408	0.0406	0.0425	0.0408	0.0423	**0.0424**
SD	$\hat{\mu}$	**0.0074**	0.0074	0.0107	0.0074	0.0074	0.0074	0.0090
	$\hat{\kappa}$	0.2383	0.2178	0.2766	0.2382	**0.2178**	0.2358	0.3062
	$\hat{\sigma}$	1.37×10^{-3}	1.27×10^{-3}	1.28×10^{-3}	1.37×10^{-3}	$\mathbf{1.27\times10^{-3}}$	1.34×10^{-3}	1.77×10^{-3}
RMSE	$\hat{\mu}$	**0.0074**	0.0074	0.0109	0.0074	0.0074	0.0074	0.0091
	$\hat{\kappa}$	0.2539	0.2227	0.2770	0.2542	**0.2227**	0.2513	0.3390
	$\hat{\sigma}$	1.37×10^{-3}	2.06×10^{-3}	2.21×10^{-3}	1.37×10^{-3}	2.06×10^{-3}	$\mathbf{1.35\times10^{-3}}$	1.77×10^{-3}

Table A.2: Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)^\top = (0.09, 0.9, 0.0424)^\top$, scenario 4: high mean reversion and high volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

Scenario 6	θ	EML	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	**0.0902**	0.0902	0.0892	0.0902	0.0902	0.0902	0.0904
	$\hat{\kappa}$	5.3788	5.0997	**5.0206**	5.3814	5.1001	5.3614	5.6980
	$\hat{\sigma}$	0.1002	0.0952	0.0946	0.1002	0.0952	0.0995	**0.1000**
SD	$\hat{\mu}$	**0.0065**	0.0065	0.0074	0.0065	0.0065	0.0065	0.0084
	$\hat{\kappa}$	1.1529	1.0319	1.2585	1.1493	**1.0318**	1.1402	1.5548
	$\hat{\sigma}$	0.0033	0.0030	0.0030	0.0033	**0.0030**	0.0032	0.0043
RMSE	$\hat{\mu}$	**0.0065**	0.0065	0.0074	0.0065	0.0065	0.0065	0.0084
	$\hat{\kappa}$	1.2136	**1.0367**	1.2587	1.2109	1.0367	1.1961	1.7043
	$\hat{\sigma}$	0.0033	0.0056	0.0061	0.0033	0.0056	**0.0032**	0.0043
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	**0.0900**	0.0900	0.0889	0.0900	0.0900	0.0900	0.0900
	$\hat{\kappa}$	5.0732	4.8317	4.6468	5.0743	4.8322	**5.0311**	4.9583
	$\hat{\sigma}$	0.1001	0.0954	0.0948	0.1001	0.0954	0.0992	**0.1000**
SD	$\hat{\mu}$	**0.0029**	0.0029	0.0034	0.0029	0.0029	0.0029	0.0029
	$\hat{\kappa}$	0.4992	0.4518	0.6269	0.4977	**0.4516**	0.4889	0.4878
	$\hat{\sigma}$	0.0014	0.0013	0.0013	0.0014	**0.0013**	0.0014	0.0014
RMSE	$\hat{\mu}$	**0.0029**	0.0029	0.0036	0.0029	0.0029	0.0029	0.0029
	$\hat{\kappa}$	0.5045	0.4822	0.7195	0.5032	**0.4818**	0.4899	0.4895
	$\hat{\sigma}$	0.0014	0.0048	0.0053	0.0014	0.0048	0.0016	**0.0014**
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	**0.0901**	0.0901	0.0891	0.0899	0.0901	0.0901	0.0901
	$\hat{\kappa}$	5.0939	4.1421	4.6304	5.1696	4.1418	**4.9286**	5.0970
	$\hat{\sigma}$	**0.1001**	0.0822	0.0810	0.0996	0.0822	0.0965	0.0997
SD	$\hat{\mu}$	**0.0031**	0.0031	0.0033	0.0033	0.0031	0.0031	0.0041
	$\hat{\kappa}$	0.6115	0.3959	0.6236	0.7052	**0.3957**	0.5693	0.7941
	$\hat{\sigma}$	0.0038	**0.0026**	0.0025	0.0041	0.0026	0.0034	0.0047
RMSE	$\hat{\mu}$	**0.0031**	0.0031	0.0035	0.0033	0.0031	0.0031	0.0041
	$\hat{\kappa}$	0.6186	0.9448	0.7249	0.7253	0.9450	**0.5737**	0.8000
	$\hat{\sigma}$	**0.0038**	0.0180	0.0192	0.0041	0.0180	0.0049	0.0047

Table A.3: Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)^\top = (0.09, 5, 0.1)^\top$. *Boldfaces denote the best results in terms of bias, standard deviation and RMSE.*

Scenario 8	θ	EML	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$								
Mean	$\hat{\mu}$	0.0902	0.0902	0.0892	**0.0901**	0.0902	0.0902	0.0903
	$\hat{\kappa}$	10.1822	9.2274	9.4201	10.1842	9.2281	**10.0552**	10.4527
	$\hat{\sigma}$	0.1402	0.1275	0.1263	**0.1401**	0.1275	0.1381	0.1399
SD	$\hat{\mu}$	**0.0046**	0.0046	0.0050	0.0047	0.0046	0.0046	0.0064
	$\hat{\kappa}$	1.6165	**1.3166**	1.7549	1.6113	1.3174	1.5661	2.1184
	$\hat{\sigma}$	0.0048	**0.0040**	0.0039	0.0049	0.0040	0.0046	0.0062
RMSE	$\hat{\mu}$	**0.0046**	0.0046	0.0051	0.0047	0.0046	0.0046	0.0064
	$\hat{\kappa}$	1.6610	**1.4357**	1.7955	1.6564	1.4362	1.5867	2.2167
	$\hat{\sigma}$	**0.0048**	0.0131	0.0143	0.0049	0.0131	0.0050	0.0062
Weekly sampling ($\Delta = 1/52$) and $n = 2600$								
Mean	$\hat{\mu}$	**0.0900**	0.0900	0.0890	0.0900	0.0900	0.0900	0.0900
	$\hat{\kappa}$	9.8776	8.9929	9.0671	**9.8722**	8.9941	9.7084	9.4325
	$\hat{\sigma}$	0.1401	0.1278	0.1265	**0.1401**	0.1278	0.1376	0.1395
SD	$\hat{\mu}$	**0.0021**	0.0021	0.0023	0.0021	0.0021	0.0021	0.0021
	$\hat{\kappa}$	0.7359	**0.6049**	0.8327	0.7334	0.6060	0.7076	0.7327
	$\hat{\sigma}$	0.0021	**0.0017**	0.0017	0.0021	0.0017	0.0020	0.0021
RMSE	$\hat{\mu}$	**0.0021**	0.0021	0.0025	0.0021	0.0021	0.0021	0.0021
	$\hat{\kappa}$	0.7400	1.0086	1.1093	0.7369	1.0083	**0.7135**	0.8197
	$\hat{\sigma}$	**0.0021**	0.0123	0.0136	0.0021	0.0123	0.0031	0.0021
Monthly sampling ($\Delta = 1/12$) and $n = 520$								
Mean	$\hat{\mu}$	**0.0901**	0.0901	0.0890	0.0898	0.0901	0.0901	0.0901
	$\hat{\kappa}$	**9.9341**	6.7332	8.7862	9.5056	6.7348	9.2813	9.6680
	$\hat{\sigma}$	**0.1403**	0.0981	0.0966	0.1339	0.0981	0.1303	0.1384
SD	$\hat{\mu}$	**0.0023**	0.0023	0.0024	0.0027	0.0023	0.0023	0.0030
	$\hat{\kappa}$	1.0861	0.4699	1.0441	1.2285	**0.4690**	0.9545	1.4521
	$\hat{\sigma}$	0.0064	0.0030	0.0031	0.0075	**0.0030**	0.0053	0.0079
RMSE	$\hat{\mu}$	**0.0023**	0.0023	0.0026	0.0027	0.0023	0.0023	0.0030
	$\hat{\kappa}$	1.0943	3.1026	1.4553	1.2632	3.1009	**1.0864**	1.4581
	$\hat{\sigma}$	**0.0064**	0.0421	0.0435	0.0097	0.0421	0.0110	0.0081

Table A.4: Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)^\top = (0.09, 9.8, 0.14)^\top$. *Boldfaces denote the best results in terms of bias, standard deviation and RMSE.*

A.2 CKLS model simulations

We provide here implementation details for the MCMC implementation and the Hermite polynomial expansion procedure (Aït-Sahalia, 2002).

Regarding the MCMC implementation for the CKLS model, we can pluggin the drift $m(\cdot)$ and diffusion $\sigma(\cdot)$ functions so that the likelihood function (3.14) can be expressed as

$$\mathcal{L}(\boldsymbol{\theta}) \propto \prod_{i=0}^{n-1} \left[\prod_{j=0}^{M} \frac{1}{\sigma x_{t_i,j}^{*\gamma}} \exp\left(- \frac{\left[x_{t_i,j+1}^{*} - x_{t_i,j}^{*} - (\alpha - \beta x_{t_i,j}^{*})\Delta\right]^2}{2\,\sigma^2\, x_{t_i,j}^{*2\gamma}\,\Delta} \right) \right]. \tag{A.1}$$

Defining $\boldsymbol{y}$ to be the vector obtained by stacking $(x_{t_i,j+1}^{*} - x_{t_i,j}^{*})/(x_{t_i,j}^{*\gamma})$ and $\boldsymbol{Z}$ as the matrix obtained by stacking $\begin{pmatrix} x_{t_i,j}^{*\gamma}\sqrt{\Delta} & x_{t_i,j}^{*\gamma}\sqrt{\Delta} \end{pmatrix}$ for all i and j, equation (A.1) can be rewritten as the likelihood function corresponding to a linear regression of $\boldsymbol{y}$ on $\boldsymbol{Z}$, such that

$$\begin{aligned} (\alpha, \beta \mid \sigma, X, X^*) &\sim \mathcal{N}\left(\bar{\boldsymbol{\theta}},\ \sigma^2(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\right), \\ (\sigma^{-2} \mid X, X^*) &\sim IG\left(\frac{mn-2}{2}, \frac{2}{(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^\top (\boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{\theta}})}\right), \end{aligned}$$

where $\bar{\boldsymbol{\theta}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top \boldsymbol{y})$ and $IG(\cdot)$ denotes an inverted gamma density. To estimate γ, Eraker (2001) notes that the second derivative of the log posterior, $p(\gamma \mid \alpha, \beta, \sigma, X, X^*)$, with respect to γ is

$$-2 \sum_{i=0}^{n-1} \left[\sum_{j=0}^{M} \left(\frac{\left(x_{t_i,j+1}^{*} - x_{t_i,j}^{*} - (\alpha - \beta x_{t_i,j}^{*})\Delta\right)^2}{x_{t_i,j}^{*\gamma}} \right)^2 \log(x_{t_i,j}^{*})^2 \right],$$

which is negative, therefore the density is log-concave. For sampling γ we can use a Metropolis-Hastings procedure with a normal proposal density based on a second-order Taylor expansion of the log-posterior, such that

$$\gamma^{(p)} \sim \mathcal{N}\left(\gamma^{(p-1)} - \frac{\partial \log \mathcal{L}(\gamma^{(p-1)})/\partial (X, X^*)}{\partial^2 \log \mathcal{L}(\gamma^{(p-1)})/\partial (X, X^*)^2},\ -\left(\frac{\partial^2 \log \mathcal{L}\left(\gamma^{(p-1)}\right)}{\partial (X, X^*)^2} \right)^{-1} \right).$$

Concerning the Hermite polynomial expansion procedure, the Lamperti transform for the CKLS process,

$$\begin{aligned} \mathrm{d}X_t &= m(X_t; \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t; \boldsymbol{\theta})\,\mathrm{d}W_t \\ &= (\theta_1 - \theta_2 X_t)\,\mathrm{d}t + \theta_3 X_t^{\theta_4}\,\mathrm{d}W_t, \end{aligned}$$

with $\theta_4 > 1$, is given by

$$U_t = \psi(X_t;\boldsymbol{\theta}) = \int^{X_t} \frac{1}{\sigma(s;\boldsymbol{\theta})}\,\mathrm{d}s = \int_0^{X_t} \frac{1}{\theta_3 s^{\theta_4}}\,\mathrm{d}s = \frac{X_t^{1-\theta_4}}{\theta_3(\theta_4-1)}.$$

Applying Itô's formula to the new process U,

$$\begin{aligned}
\mathrm{d}U_t &= m_U(U_t;\boldsymbol{\theta})\,\mathrm{d}t + \mathrm{d}W_t \\
&= \left(\frac{m(\psi^{-1}(u;\boldsymbol{\theta});\boldsymbol{\theta})}{\sigma(\psi^{-1}(u;\boldsymbol{\theta});\boldsymbol{\theta})} - \frac{1}{2}\frac{\partial\sigma}{\partial x}\left(\psi^{-1}(u;\boldsymbol{\theta});\boldsymbol{\theta}\right)\right)\mathrm{d}t + \mathrm{d}W_t \\
&= \left(\frac{\theta_4}{2(\theta_4-1)U_t} - \theta_2(\theta_4-1)U_t + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}U_t^{\theta_4/(\theta_4-1)}\right)\mathrm{d}t + \mathrm{d}W_t,
\end{aligned}$$

we obtain a process with unitary diffusion coefficient. The drift of the process U_t is

$$m_U(u_0;\boldsymbol{\theta}) = \frac{\theta_4}{2(\theta_4-1)u_0} - \theta_2(\theta_4-1)u_0 + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}u_0^{\theta_4/(\theta_4-1)},$$

and the partial derivatives are given by

$$\begin{aligned}
m_U^{[1]} &= \frac{-\theta_4}{2(\theta_4-1)u_0^2} - \theta_2(\theta_4-1) + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}\frac{\theta_4}{(\theta_4-1)}u_0^{\frac{\theta_4}{(\theta_4-1)}-1}, \\
m_U^{[2]} &= \frac{\theta_4}{(\theta_4-1)u_0^3} + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}\frac{\theta_4}{(\theta_4-1)^2}u_0^{\frac{\theta_4}{(\theta_4-1)}-2}, \\
m_U^{[3]} &= \frac{-3\theta_4}{(\theta_4-1)u_0^4} + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}\frac{\theta_4(2-\theta_4)}{(\theta_4-1)^3}u_0^{\frac{\theta_4}{(\theta_4-1)}-3}, \\
m_U^{[4]} &= \frac{12\theta_4}{(\theta_4-1)u_0^5} + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)}\frac{\theta_4(2\theta_4^2-7\theta_4+6)}{(\theta_4-1)^4}u_0^{\frac{\theta_4}{(\theta_4-1)}-4}, \\
m_U^{[5]} &= \frac{-60\theta_4}{(\theta_4-1)u_0^6} + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)} \\
&\quad\times\frac{\theta_4(-6\theta_4^3+29\theta_4^2-46\theta_4+24)}{(\theta_4-1)^5}u_0^{\frac{\theta_4}{(\theta_4-1)}-5}, \\
m_U^{[6]} &= \frac{360\theta_4}{(\theta_4-1)u_0^7} + \theta_1\theta_3^{1/(\theta_4-1)}(\theta_4-1)^{\theta_4/(\theta_4-1)} \\
&\quad\times\frac{\theta_4(24\theta_4^4-146\theta_4^3+329\theta_4^2-326\theta_4+120)}{(\theta_4-1)^6}u_0^{\frac{\theta_4}{(\theta_4-1)}-6},
\end{aligned}$$

where $m_U^{[k]} = \partial^k m_U(u_0;\boldsymbol{\theta})/\partial u_0^k$.

Scenario 2	θ	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$							
Mean	$\hat{\mu}$	0.0984	**0.0977**	0.0988	0.0980	0.0988	0.1003
	$\hat{\kappa}$	**0.6577**	0.6650	0.6608	0.6579	0.6651	0.8836
	$\hat{\sigma}$	1.0244	1.0638	1.0706	**1.0241**	1.1357	0.9625
	$\hat{\gamma}$	1.4873	1.5038	1.5032	1.4872	**1.4998**	1.4440
SD	$\hat{\mu}$	0.0933	0.0926	0.1091	**0.0904**	0.1054	0.1872
	$\hat{\kappa}$	**0.4887**	0.4984	0.4932	0.4888	0.4972	0.4953
	$\hat{\sigma}$	0.3720	0.3859	0.4013	**0.3719**	0.5257	0.4548
	$\hat{\gamma}$	0.1372	0.1350	**0.1327**	0.1372	0.1595	0.1787
RMSE	$\hat{\mu}$	0.0937	0.0929	0.1095	**0.0908**	0.1057	0.1875
	$\hat{\kappa}$	**0.6696**	0.6816	0.6750	0.6698	0.6808	0.8442
	$\hat{\sigma}$	0.3728	0.3911	0.4075	**0.3726**	0.5429	0.4563
	$\hat{\gamma}$	0.1378	0.1351	**0.1328**	0.1378	0.1595	0.1873
Weekly sampling ($\Delta = 1/52$) and $n = 2600$							
Mean	$\hat{\mu}$	0.1002	0.0994	0.0997	0.1003	0.1003	**0.0882**
	$\hat{\kappa}$	0.2807	0.2819	0.2812	**0.2805**	0.2813	0.4766
	$\hat{\sigma}$	**0.9944**	1.0074	1.0061	0.9943	1.0122	0.9631
	$\hat{\gamma}$	1.4956	1.5019	**1.4999**	1.4956	1.5002	1.4690
SD	$\hat{\mu}$	0.0687	0.0605	0.0616	0.0692	0.0689	**0.0218**
	$\hat{\kappa}$	0.1567	0.1580	0.1575	**0.1566**	0.1576	0.1604
	$\hat{\sigma}$	0.1044	**0.1034**	0.1040	0.1042	0.1140	0.1923
	$\hat{\gamma}$	0.0407	**0.0395**	0.0399	0.0406	0.0430	0.0889
RMSE	$\hat{\mu}$	0.0695	0.0612	0.0623	0.0699	0.0697	**0.0219**
	$\hat{\kappa}$	0.1762	0.1780	0.1772	**0.1761**	0.1773	0.3198
	$\hat{\sigma}$	0.1045	**0.1037**	0.1042	0.1044	0.1146	0.1958
	$\hat{\gamma}$	0.0409	**0.0396**	0.0399	0.0408	0.0430	0.0942
Monthly sampling ($\Delta = 1/12$) and $n = 520$							
Mean	$\hat{\mu}$	0.1086	0.1102	0.1133	0.1088	0.1109	**0.0890**
	$\hat{\kappa}$	**0.2860**	0.2916	0.2890	0.2861	0.2888	0.4813
	$\hat{\sigma}$	0.9690	1.0247	**1.0223**	0.9688	1.0606	0.8381
	$\hat{\gamma}$	1.4771	1.5057	1.4977	1.4770	**1.4999**	1.3820
SD	$\hat{\mu}$	0.0974	0.1213	0.1425	0.0992	0.1198	**0.0232**
	$\hat{\kappa}$	0.1679	0.1734	0.1724	0.1680	0.1721	**0.1579**
	$\hat{\sigma}$	0.2392	**0.2361**	0.2437	0.2394	0.2895	0.3223
	$\hat{\gamma}$	0.0958	**0.0871**	0.0899	0.0959	0.0999	0.1609
RMSE	$\hat{\mu}$	0.0991	0.1229	0.1444	0.1010	0.1216	**0.0233**
	$\hat{\kappa}$	**0.1887**	0.1961	0.1941	0.1888	0.1936	0.3226
	$\hat{\sigma}$	0.2412	**0.2374**	0.2447	0.2414	0.2957	0.3606
	$\hat{\gamma}$	0.0985	**0.0873**	0.0899	0.0986	0.0999	0.1995

Table A.5: Monte Carlo simulation for CKLS model with $(\mu, \kappa, \sigma, \gamma)^\top = (0.09, 0.2, 1, 1.5)^\top$, scenario 2: low mean reversion and high volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

Scenario 4	θ	DML	LL	HP	KF	MCMC	GMM
Weekly sampling ($\Delta = 1/52$) and $n = 520$							
Mean	$\hat{\mu}$	0.0965	0.0953	0.0955	0.0963	0.0961	**0.0952**
	$\hat{\kappa}$	1.2752	1.2993	1.2912	**1.2743**	1.2958	1.6327
	$\hat{\sigma}$	1.3874	1.4853	1.4823	**1.3874**	1.5353	1.2845
	$\hat{\gamma}$	1.4762	1.5076	**1.4998**	1.4762	1.4944	1.4259
SD	$\hat{\mu}$	0.0625	**0.0421**	0.0441	0.0576	0.0494	0.1351
	$\hat{\kappa}$	0.6674	0.6905	0.6911	0.6663	0.6871	**0.6541**
	$\hat{\sigma}$	**0.4298**	0.4496	0.4685	0.4299	0.5573	0.5483
	$\hat{\gamma}$	0.1241	**0.1184**	0.1209	0.1240	0.1345	0.1689
RMSE	$\hat{\mu}$	0.0628	**0.0424**	0.0444	0.0580	0.0498	0.1352
	$\hat{\kappa}$	0.7657	0.7976	0.7941	**0.7643**	0.7930	0.9822
	$\hat{\sigma}$	**0.4307**	0.4552	0.4734	0.4308	0.5703	0.5635
	$\hat{\gamma}$	0.1263	**0.1187**	0.1209	0.1263	0.1347	0.1845
Weekly sampling ($\Delta = 1/52$) and $n = 2600$							
Mean	$\hat{\mu}$	0.0907	0.0904	0.0912	0.0907	0.0907	**0.0899**
	$\hat{\kappa}$	**0.9594**	0.9759	0.9664	0.9600	0.9665	1.1897
	$\hat{\sigma}$	1.3765	1.4397	**1.4308**	1.3757	1.4465	1.3012
	$\hat{\gamma}$	1.4876	1.5085	**1.5010**	1.4874	1.5036	1.4541
SD	$\hat{\mu}$	0.0078	0.0078	0.0150	0.0078	0.0078	**0.0076**
	$\hat{\kappa}$	**0.2535**	0.2601	0.2584	0.2537	0.2570	0.2754
	$\hat{\sigma}$	0.1661	0.1668	0.1770	**0.1659**	0.1839	0.2754
	$\hat{\gamma}$	0.0488	**0.0468**	0.0494	0.0488	0.0512	0.0905
RMSE	$\hat{\mu}$	0.0079	0.0078	0.0150	0.0079	0.0079	**0.0076**
	$\hat{\kappa}$	**0.2604**	0.2709	0.2668	0.2607	0.2654	0.3998
	$\hat{\sigma}$	0.1703	**0.1688**	0.1778	0.1704	0.1867	0.2977
	$\hat{\gamma}$	0.0503	**0.0476**	0.0494	0.0504	0.0513	0.1015
Monthly sampling ($\Delta = 1/12$) and $n = 520$							
Mean	$\hat{\mu}$	0.0909	**0.0902**	0.0925	0.0909	0.0911	0.0903
	$\hat{\kappa}$	**0.9493**	1.0009	0.9838	0.9493	0.9799	1.1551
	$\hat{\sigma}$	1.2486	1.4915	**1.4720**	1.2482	1.5296	0.9994
	$\hat{\gamma}$	1.4382	1.5265	**1.4989**	1.4381	1.5040	1.3218
SD	$\hat{\mu}$	0.0080	**0.0078**	0.0452	0.0080	0.0081	0.0078
	$\hat{\kappa}$	**0.2692**	0.2939	0.2968	0.2694	0.2891	0.2834
	$\hat{\sigma}$	0.3849	0.3856	0.4144	**0.3846**	0.4573	0.3995
	$\hat{\gamma}$	0.1245	**0.1048**	0.1126	0.1245	0.1198	0.1594
RMSE	$\hat{\mu}$	0.0081	**0.0078**	0.0453	0.0081	0.0082	0.0078
	$\hat{\kappa}$	**0.2737**	0.3107	0.3084	0.2738	0.3000	0.3812
	$\hat{\sigma}$	0.4190	**0.3933**	0.4185	0.4189	0.4717	0.5759
	$\hat{\gamma}$	0.1390	**0.1081**	0.1126	0.1390	0.1199	0.2391

Table A.6: Monte Carlo simulation for the CKLS model with $(\mu, \kappa, \sigma, \gamma)^\top = (0.09, 0.9, 1.414, 1.5)^\top$*, scenario 4: high mean reversion and high volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.*

Appendix B

The estsde R package

Type Package

Title Parametric Estimation of SDE

Version 0.0.0.9000

Author A. Lopez-Perez

Maintainer A. Lopez-Perez <alejandralopez.perez@usc.es>

Description Provides different parametric estimation methods for stochastic differential equations (SDE), namely the Vasicek and CKLS models, as well as functions to simulate paths of the SDEs.

License GPL-2

Encoding UTF-8

RoxygenNote 7.1.1

Roxygen list(old_usage = TRUE)

Depends R (>= 3.5.0), Rcpp

LinkingTo Rcpp, RcppArmadillo, nloptr

Imports stats, nloptr, nlme, numDeriv

R topics documented:

`est.CKLS.DML` *ML estimation for the CKLS model (Euler method)*

Description

Parametric estimation for the CKLS model using maximum likelihood and the discretized version of the model, obtained with the Euler-Maruyama method. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.DML(X, Delta = deltat(X), par = NULL)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

Examples

```
x <- rCKLS(360, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.DML(x)
```

`est.CKLS.GMM` *Generalized method of moments estimator for the CKLS model*

Description

Parametric estimation for the CKLS model using the Generalized Method of Moments. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.GMM(X, Delta = deltat(X), par = NULL, maxiter = 25)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.
`maxiter`	an integer, the maximum number of iterations.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, pages 1029–1054.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance, 47(3):1209–1227.

Examples

```
x <- rCKLS(360, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.GMM(x)
```

`est.CKLS.HP` *ML estimation for the CKLS model (Hermite polynomial expansion)*

Description

Parametric estimation for the CKLS model using maximum likelihood and the discretized version of the model, obtained with the Ait-Sahalia Hermite polynomial expansion method. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.HP(X, Delta = deltat(X), par = NULL)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. Econometrica, 70(1):223–262.

Examples

```
x <- rCKLS(360, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.HP(x)
```

`est.CKLS.KF` *ML estimation for the CKLS model (Kalman filter)*

Description

Parametric estimation for the CKLS model using the Kalman filter algorithm, where the discretized version of the model is obtained with the Euler-Maruyama method. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.KF(X, Delta = deltat(X), par = NULL, mu0 = 0, Sigma0 = 1)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.
`mu0`	a single numeric, the initial mean. Defaults to zero.
`Sigma0`	a single numeric, the initial variance. Defaults to one.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

Examples

```
x <- rCKLS(360, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.KF(x)
```

---

est.CKLS.LL *ML estimation for the CKLS model (local linearization)*

Description

Parametric estimation for the CKLS model using maximum likelihood and the discretized version of the model, obtained with the local linearization method. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.LL(X, Delta = deltat(X), par = NULL)
```

Arguments

X	a numeric vector, the sample path of the SDE.
Delta	a single numeric, the time step between two consecutive observations.
par	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. Statistica Sinica, pages 113–135.

Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. Stochastic Analysis and Applications, 16(4):733–752.

Examples

```
x <- rCKLS(360, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.LL(x)
```

`est.CKLS.MCMC` *MCMC estimation for the CKLS model*

Description

Parametric estimation for the CKLS model using Markov Chain Monte Carlo and involving data augmentation, as proposed in Elerian et al. (2001) and Eraker (2001). The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
est.CKLS.MCMC(X, Delta = deltat(X), par = NULL,
              niter = 4000, burn_in = 1000)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension four indicating initial values of the parameters. Defaults to NULL, fits a linear model using generalized least squares with AR1 correlation and a power variance heteroscedasticity structure.

`niter` an integer, number of iterations.

`burn_in` an integer indicating the number of initial iterations to be discarded.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. Econometrica, 69(4):959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. Journal of Business & Economic Statistics, 19(2):177–191.

Examples

```
set.seed(987)
x <- rCKLS(480, 1/12, 0.09, 0.08, 0.9, 1.2, 1.5)
est.CKLS.MCMC(x)
```

`est.VAS.DML` *ML estimation for the Vasicek model (Euler method)*

Description

Parametric estimation for the Vasicek model using maximum likelihood and the discretized version of the model, obtained with the Euler-Maruyama method. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.DML(X, Delta = deltat(X), par = NULL)
```

Arguments

X	a numeric vector, the sample path of the SDE.
Delta	a single numeric, the time step between two consecutive observations.
par	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.DML(x)
```

---

est.VAS.EML *ML estimation for the Vasicek model*

Description

Parametric estimation for the Vasicek model using (exact) maximum likelihood. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.EML(X, Delta = deltat(X), par = NULL)
```

Arguments

X	a numeric vector, the sample path of the SDE.
Delta	a single numeric, the time step between two consecutive observations.
par	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.EML(x)
```

`est.VAS.GMM` *Generalized method of moments estimator for the Vasicek model*

Description

Parametric estimation for the Vasicek model using the Generalized Method of Moments. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.GMM(X, Delta = deltat(X), par = NULL, maxiter = 25)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.
`maxiter`	an integer, the maximum number of iterations.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, pages 1029–1054.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.GMM(x)
```

`est.VAS.HP` *ML estimation for the Vasicek model (Hermite polynomial expansion)*

Description

Parametric estimation for the Vasicek model using maximum likelihood and the discretized version of the model, obtained with the Ait-Sahalia Hermite polynomial expansion method. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.HP(X, Delta = deltat(X), par = NULL)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. Econometrica, 70(1):223–262.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.HP(x)
```

`est.VAS.KF` *ML estimation for the Vasicek model (Kalman filter)*

Description

Parametric estimation for the Vasicek model using the Kalman filter algorithm, where the discretized version of the model is obtained with the Euler-Maruyama method. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.KF(X, Delta = deltat(X), par = NULL, mu0 = 0, Sigma0 = 1)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.
`mu0`	a single numeric, the initial mean. Defaults to zero.
`Sigma0`	a single numeric, the initial variance. Defaults to one.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.KF(x)
```

`est.VAS.LL` *ML estimation for the Vasicek model (local linearization)*

Description

Parametric estimation for the Vasicek model using maximum likelihood and the discretized version of the model, obtained with the local linearization method. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.LL(X, Delta = deltat(X), par = NULL)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. Statistica Sinica, pages 113–135.

Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. Stochastic Analysis and Applications, 16(4):733–752.

Examples

```
x <- rVAS(360, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.LL(x)
```

`est.VAS.MCMC` *MCMC estimation for the Vasicek model*

Description

Parametric estimation for the Vasicek model using Markov Chain Monte Carlo and involving data augmentation, as proposed in Elerian et al. (2001) and Eraker (2001). The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
est.VAS.MCMC(X, Delta = deltat(X), par = NULL, niter = 2000, burn_in = 500)
```

Arguments

`X`	a numeric vector, the sample path of the SDE.
`Delta`	a single numeric, the time step between two consecutive observations.
`par`	a numeric vector with dimension three indicating initial values of the parameters. Defaults to NULL, fits a linear model as an initial guess.
`niter`	an integer, number of iterations.
`burn_in`	an integer indicating the number of initial iterations to be discarded.

Value

A list containing a matrix with the estimated coefficients and the associated standard errors.

References

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. Econometrica, 69(4):959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. Journal of Business & Economic Statistics, 19(2):177–191.

Examples

```
set.seed(987)
x <- rVAS(480, 1/12, 0, 0.08, 0.9, 0.1)
est.VAS.MCMC(x)
```

`rCKLS` *Simulation of the CKLS model*

Description

`rCKLS` simulates a single path of the CKLS process using the Milstein scheme. The parametric form of the CKLS model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t.$$

Usage

```
rCKLS(n, Delta, X0, alpha, kappa, sigma, gamma)
```

Arguments

n	an integer indicating the sample size.
Delta	a single numeric, the time step between two consecutive observations.
X0	a single numeric, initial value of the process.
alpha	a single numeric, drift parameter.
kappa	a single numeric, rate of mean reversion.
sigma	a single numeric, volatility parameter.
gamma	a single numeric, proportional volatility exponent.

Value

A ts object, a numeric vector with the sample path of the SDE.

References

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance, 47(3):1209–1227.

Examples

```
x <- rCKLS(360, Delta = 1/12, X0 = 0.09, alpha = 0.08,
           kappa = 0.9, sigma = 1.2, gamma = 1.5)
plot(x)
```

---

rVAS *Simulation of the Vasicek model*

Description

rVAS simulates a single path of the Vasicek process using its known conditional distribution. The parametric form of the Vasicek model used here is given by

$$\mathrm{d}X_t = (\alpha - \kappa X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t.$$

Usage

```
rVAS(n, Delta, X0, alpha, kappa, sigma)
```

Arguments

n	an integer indicating the sample size.
Delta	a single numeric, the time step between two consecutive observations.
X0	a single numeric, initial value of the process.
alpha	a single numeric, drift parameter.
kappa	a single numeric, rate of mean reversion.
sigma	a single numeric, volatility parameter.

Value

A ts object, a numeric vector with the sample path of the SDE.

Examples

```
x <- rVAS(360, Delta = 1/12, X0 = 0.09, alpha = 0.08, kappa = 0.9, sigma = 0.1)
plot(x)
```

Appendix C

Supplementary material for Chapter 8

This appendix is provided as supplementary material for Chapter 8. It pretends to be self-contained and, consequently, some details also contained in the chapter are here sketched. Section C.1 is devoted to discussions on the choice of functional bases in the functional data analysis framework, often neglected in the literature. Section C.2 supplies details about our hybrid estimator and the choice of the penalty parameter involved in it. Section C.3 provides the main details about the theoretical ARH framework proposed, as well as the main assumptions required. Proofs of main theoretical results are also provided.

C.1 Functional bases: component-wise perspective

Sparsity effects usually appear in Functional Data Analysis (FDA) setup under many guises, marked by the infinite dimension of the samples. This sparsity can be understood in a wide sense, such as referring to the sparsity of the model (Aneiros and Vieu, 2014), the sparsity of data related to the topological effects of small ball probabilities (Vieu, 2018) or related to the discretization grid (Yao et al., 2005). Because of this, as argued in Kirichenko and Nikitin (2014), dimension reduction is essential in the FDA setup.

Let $\{\Psi_j\}_{j=1}^{\infty}$ and $\{\Phi_k\}_{k=1}^{\infty}$ be functional bases in Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$. Under the assumption of separability of Hilbert spaces, any elements $\mathcal{X} \in \mathbb{H}_1$ and $\mathcal{Y} \in \mathbb{H}_2$ can be represented as $\mathcal{X} = \sum_{j=1}^{\infty} x_j \Psi_j$ and $\mathcal{Y} = \sum_{k=1}^{\infty} y_k \Phi_k$, with $x_j = \langle \mathcal{X}, \Psi_j \rangle_{\mathbb{H}_1}$ and $y_k = \langle \mathcal{Y}, \Phi_k \rangle_{\mathbb{H}_2}$, for any $j, k \geqslant 1$. We can further distinguish among fixed (deterministic) bases (e.g., B-splines bases, wavelet bases or Fourier bases), flexible but usually require a larger number of elements to adequately represent a functional sample, and data-driven orthogonal functional bases. More parsimonious representations can be achieved by considering the latter ones, performing a dimension reduction explicitly based on the covariance structure, being

the most popular choice the Functional Principal Components (FPC) of $\{\mathcal{X}_i\}_{i=1}^n$, $\left\{\hat{\Psi}_j\right\}_{j=1}^n$, emerging as the solution to the eigenequation

$$\int C_n(s,t)\hat{\Psi}_j(s)\,\mathrm{d}s = \lambda_{j,n}\hat{\Psi}_j(t),$$

with $C_n(s,t) = \frac{1}{n}\sum_{i=1}^n \mathcal{X}_i(s)\mathcal{X}_i(t)$; that is, as the eigenfunctions of the empirical autocovariance operator $C_n = \frac{1}{n}\sum_{i=1}^n \mathcal{X}_i \otimes \mathcal{X}_i$, denoting by $\otimes$ the tensorial product on a Hilbert space $\mathbb{H}$, such that $(\mathcal{X}\otimes\mathcal{Y})(\mathcal{Z})$, for each $\mathcal{X},\ \mathcal{Y},\ \mathcal{Z} \in \mathbb{H}$. Some sparse FPCA proposals, involving wavelet denoising, can be also found (see Roislien and Winje, 2013). Nonetheless, even though FPC is the most optimal choice for representing our curves, it may not be the best option for making inference.

Since then, Partial Least Squares (PLS), extended to the FDA setting as Functional PLS (FPLS) by Preda and Saporta (2005), focuses on the predictive performance instead of achieving an accurate representation, such that "best" functional directions are selected, but bearing in mind simultaneously both input and output variables. Nonlinear dimension reduction techniques, in the manifolds-valued FDA setting, are recently developed in Dai and Müller (2018).

To handle the infinite-dimensional nature, (p,q)-truncated bases $\{\Psi_j\}_{j=1}^p$ and $\{\Phi_k\}_{k=1}^q$ are usually adopted, corresponding to the first p and q elements of $\{\Psi_j\}_{j=1}^\infty$ and $\{\Phi_k\}_{k=1}^\infty$, in $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively:

$$\mathcal{X}^{(p)} = \sum_{j=1}^p x_j\Psi_j,$$
$$\mathcal{Y}^{(q)} = \sum_{k=1}^q y_k\Phi_k,$$

with $(x_1,\ldots,x_p)\in\mathbb{R}^p$ and $(y_1,\ldots,y_q)\in\mathbb{R}^q$. Cut-off levels (p,q) are commonly selected by capturing a minimum amount of the explained variance (see the variable selection proposal developed by Berrendero et al., 2016 in the FDA classification context, and the proposal by García-Portugués et al., 2021 in the FLMFR setting).

C.2 FLMFR estimation

We will assume that $\mathcal{X}\in\mathbb{H}_1 = L^2([a,b])$, $\mathcal{Y}\in\mathbb{H}_2 = L^2([c,d])$, both already centered. Let us detail the FLMFR setting in García-Portugués et al. (2021), given by

$$\mathcal{Y} = m_\beta(\mathcal{X}) + \mathcal{E}, \qquad \text{with } \mathbb{E}[\mathcal{E}\mid\mathcal{X}] = 0, \tag{C.1}$$

where $m_\beta(x) = \mathbb{E}\left[\mathcal{Y} \mid \mathcal{X} = x\right]$ is a Hilbert-Schmidt integral operator, that is, $m \equiv m_\beta$ admits an integral representation given by a bivariate and square-integrable kernel $\beta \in \mathbb{H}_1 \otimes \mathbb{H}_2$. In the aftermath of the compactness, β can be decomposed as follows,

$$
\begin{aligned}
m_\beta\left(\mathcal{X}\right)(\cdot) &= \int_a^b \beta(s,\cdot)\mathcal{X}(s)\,\mathrm{d}s, \\
\beta &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \beta_{jk}\left(\Psi_j \otimes \Phi_k\right),
\end{aligned}
$$

with $\int\int \beta^2(s,t) < \infty$ and where $\beta_{jk} = \langle \beta, \Psi_j \otimes \Phi_k \rangle_{\mathbb{H}_1 \otimes \mathbb{H}_2}$, for each $j, k \geqslant 1$. Adopting (p,q)-truncated component-wise expansions, projected into $\{\Psi_j\}_{j=1}^p$, $\{\Phi_k\}_{k=1}^q$ and $\{\Psi_j \otimes \Phi_k\}_{j,k=1}^{p,q}$, respectively, we get

$$
\begin{aligned}
\mathcal{Y}^{(q)} &= \sum_{k=1}^{q} y_k \Phi_k, \\
y_k &= \sum_{j=1}^{p}\sum_{\ell=1}^{p} b_{\ell,k} x_j \langle \Psi_j, \Psi_\ell \rangle_{\mathbb{H}_1} + e_k, \qquad k = 1, \ldots, q, \\
\mathcal{E}^{(q)} &= \sum_{k=1}^{q} e_k \Phi_k,
\end{aligned}
\tag{C.2}
$$

where $\langle \Psi_j, \Psi_\ell \rangle_{\mathbb{H}_1} = \delta_{j\ell}$, for each $j, \ell \geqslant 1$. Now, given a centered sample $\{(\mathcal{X}_i, \mathcal{Y}_i)\}_{i=1}^n$, the sample version of (C.1)–(C.2) can be expressed as:

$$
\begin{aligned}
\mathbf{Y}_q &= \mathbf{X}_p \mathbf{B}_{p,q} + \mathbf{E}_q, \\
\mathcal{Y}_i &= \langle\langle \mathcal{X}_i, \beta \rangle\rangle + \mathcal{E}_i, \\
\langle\langle \mathcal{X}, \beta \rangle\rangle(t) &:= \langle \mathcal{X}, \beta(\cdot, t) \rangle_{\mathbb{H}_1},
\end{aligned}
$$

where $\mathbf{Y}_q$ and $\mathbf{E}_q$ are the $n \times q$ matrices with the coefficients of $\{\mathcal{Y}_i\}_{i=1}^n$ and $\{\mathcal{E}_i\}_{i=1}^n$, respectively, on $\{\Phi_k\}_{k=1}^q$, $\mathbf{X}_p$ is the $n \times p$ matrix with the coefficients of $\{\mathcal{X}_i\}_{i=1}^n$ on $\{\Psi_j\}_{j=1}^p$ and $\mathbf{B}_{p,q}$ is the $p \times q$ matrix with the coefficients (to be estimated) of β on $\{\Psi_j \otimes \Phi_k\}_{j,k=1}^{p,q}$.

From now on, we consider the hybrid linearly-constrained FPC Regression estimator (FPCR-L1S estimator), formulated in García-Portugués et al. (2021) as follows:

$$
\begin{aligned}
\hat{\mathbf{Y}}_q &= \widetilde{\mathbf{X}}_{\widetilde{p}} \hat{\mathbf{B}}_{\widetilde{p},q}^{(\lambda),C} = \mathbf{H}_C^{(\lambda)} \mathbf{Y}_q = \left[\widetilde{\mathbf{X}}_{\widetilde{p}} \left(\widetilde{\mathbf{X}}_{\widetilde{p}}^\top \widetilde{\mathbf{X}}_{\widetilde{p}}\right)^{-1} \widetilde{\mathbf{X}}_{\widetilde{p}}^\top\right] \mathbf{Y}_q, \\
\hat{\mathbf{B}}_{\widetilde{p},q}^{(\lambda),C} &= \left(\widetilde{\mathbf{X}}_{\widetilde{p}}^\top \widetilde{\mathbf{X}}_{\widetilde{p}}\right)^{-1} \widetilde{\mathbf{X}}_{\widetilde{p}}^\top \mathbf{Y}_q.
\end{aligned}
\tag{C.3}
$$

The estimated $\hat{\mathbf{Y}}_q$ in (C.3) are computed in three steps: *(i)* initial values (p, q) are fixed with regard to a certain EV_p and EV_q; *(ii)* we implement a variable selection by considering the no null rows of $\hat{\mathbf{B}}_{p,q}^{(\lambda)} = \underset{\mathbf{B}_{p,q}}{\arg\min} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left\| (\mathbf{Y}_q)_i - (\mathbf{X}_p \mathbf{B}_{p,q})_i \right\|^2 + \lambda \sum_{j=1}^{p} \left\| (\mathbf{B}_{p,q})_j \right\|_2 \right\}$, being $(\mathbf{Y}_q)_i$ the i-th row of $\mathbf{Y}_q$; *(iii)* we perform a FPCR estimation just using the $\widetilde{p} \leqslant p$ predictors selected, whose coefficients are $\widetilde{\mathbf{X}}_{\widetilde{p}}$, having a hat matrix $\mathbf{H}_C^{(\lambda)}$ at our disposal. Hence, the estimator in (C.3) allows us to seize the advantages of both estimation paradigms.

Not many FLMFR estimators can be found in the literature (in comparison with FLMSR), but the simplest and best known one is the Functional Principal Component Regression (FPCR) estimator, early proposed in Ramsay and Silverman (2005), although innovative approaches have been recently developed (see, e.g., Benatia et al., 2017 and Imaizumi and Kato, 2018). The main criticism against the FPCR estimator, entailed by the least-squares estimator of the truncated model given by $\hat{\mathbf{B}}_{p,q} = \left(\mathbf{X}_p^\top \mathbf{X}_p\right)^{-1} \mathbf{X}_p^\top \mathbf{Y}_q$, is its critical dependence on (p, q), where $\mathbf{X}^\top$ denotes the transpose of matrix $\mathbf{X}$. Data-driven selection could be considered by extending to the FLMFR setup the predictive cross-validation criterion (a likely high computational cost befall) or the generalized cross-validation procedure, but the challenge of estimating the degrees of freedom remains open in the FLMFR context.

An initial but simple selection of (p, q) can be fixed as the minimum number of components associated with a proportion of Explained Variance (EV_p and EV_q, respectively). As shown in García-Portugués et al. (2021, Section 2.4), while a weak dependency on q seems to manifest, large values of p could lead to an overfitted problem when $(\mathcal{X}, \mathcal{Y})$ are so regular, or when a suitable choice of (p, q) for representing $(\mathcal{X}, \mathcal{Y})$ might not be so for estimating β: many components could be not informative due its sparsity and noise would be incorporated.

An elastic net-based regularized estimator can be adopted in place, in terms of a penalty parameter $\lambda \geqslant 0$, such as ridge (FPCR-L2) and LASSO (FPCR-L1) regression. The first one performs a global penalization and provides an estimator in terms of a hat matrix (which can be reused for the re-estimations within a bootstrap algorithm), whilst the second one accomplishes a row wise penalization that effectively zeros full rows, achieving a variable selection but notably increasing the bootstrapping cost. We here consider the hybrid linearly-constrained FPCR estimator (henceforth denoted as FPCR-L1S estimator), formulated in García-Portugués et al. (2021), and already described.

Remark 5 (Choice of the penalty parameter λ)**.** The penalty parameter λ could be data-driven selected by achieving a leave-one-out cross-validation in a efficiently way, as implemented in Friedman et al. (2010) and usually denoted as $\hat{\lambda}_{CV}$. Whilst $\hat{\lambda}_{CV}$ is an optimal choice for estimating β, Friedman et al. (2010) proposed the so-called one standard error rule, denoted in the GOF folklore as $\hat{\lambda}_{1SE}$ (a variance reduction is achieved to obtain a more biased estimator),

being the best choice for the test calibration.

C.3 ARH(z) processes: theoretical framework

Some preliminary elements on functional autoregressive processes from a time-domain perspective are here discussed.

C.3.1 ARH(1) processes: theoretical framework

Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{X_t\}_{t\in\mathbb{R}}$ be now a continuous-time zero mean stochastic process. In keeping with Bosq (2000), we split the paths as $\mathcal{X}_n(t) = X_{nh+t}$, with $t \in [0, h]$, and $\mathcal{X}_n \in \mathbb{H} = L^2([0, h])$, for each $n \in \mathbb{Z}$, constituting an infinite dimensional discrete time process. This representation is especially fruitful if $\{X_t\}_{t\in\mathbb{R}}$ displays a seasonal component either it will be forecasted over $[0, h]$.

We say that $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ is a zero-mean autoregressive Hilbertian process of order one, valued in $\mathbb{H} = L^2([0, h])$, denoted as ARH(1), if the following state equation is satisfied

$$\mathcal{X}_n(t) = \Gamma\left(\mathcal{X}_{n-1}\right)(t) + \mathcal{E}_n(t), \qquad n \in \mathbb{Z}, \quad t \in [0, h],$$

with $\mathcal{X}_n,\ \mathcal{E}_n \in \mathbb{H} = L^2([0, h])$ and where Γ denotes the linear autocorrelation operator, with $\|\Gamma\|_{\mathcal{L}(\mathbb{H})} = \sup_{\|\mathcal{X}\|_{\mathbb{H}}\leqslant 1} \|\Gamma(\mathcal{X})\|_{\mathbb{H}}$, and $\{\mathcal{E}_n\}_{n\in\mathbb{Z}}$ is assumed to be a strong white noise, with $\sigma^2_{\mathcal{E}} = \mathbb{E}\left[\|\mathcal{E}_n\|^2_{\mathbb{H}}\right] < \infty$ and iid components. Henceforward, we consider the subsequent assumptions:

Assumption 12. $\left\|\Gamma^k\right\|_{\mathcal{L}(\mathbb{H})} < 1$, for any $k \geqslant k_0$, and for some $k_0 \geqslant 1$, where Γ^k denotes the composition operator $\Gamma \overbrace{\cdots}^{k} \Gamma$, leading to $\sum_{n=0}^{\infty} \left\|\Gamma^n\right\|_{\mathcal{L}(\mathbb{H})} < \infty$.

Assumption 13. The autocorrelation Hilbert–-Schmidt integral operator Γ is given by

$$\Gamma(\mathcal{X})(t) \equiv \Gamma_\rho(\mathcal{X})(t) = \int_0^h \rho(s,t)\mathcal{X}(s)\,\mathrm{d}s,$$
$$\rho = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \rho_{jk}\Psi_j \otimes \Psi_k,$$

with $\int\int \rho^2(s,t)\,\mathrm{d}s\,\mathrm{d}t < \infty$ and $\rho_{jk} = \langle \rho(\Psi_j), \Psi_k\rangle_{\mathbb{H}}$, for each $j, k = 1, \ldots, \infty$.

Remark 6. From Bosq (2000, Theorem 3.1), Assumption 12 is required for ensuring an unique stationary solution to ARH(1) process. By definition, C is a trace, self-adjoint and

positive autocovariance operator, and then, C can be also given in terms of a bivariate squared integrable kernel $c(s,t)$ such that $c(s,t) = \sum_{j=1}^{\infty} C_j \Psi_j(s)\Psi_j(t)$ and $C(\mathcal{X})(t) := \langle \mathcal{X}, c(\cdot,t)\rangle_{\mathbb{H}}$. Since $C^{-1} = \sum_{j=1}^{\infty} \frac{1}{C_j}\Psi_j \otimes \Psi_j$, with the trace property, C cannot be theoretically inverted, and thus, $\Gamma := DC^{-1}$ should be estimated. Note that Assumption 13 does not imply that DC^{-1} could be diagonally decomposed, since the symmetry of the cross-covariance operator $D := \mathbb{E}[\mathcal{X}_n \otimes \mathcal{X}_{n+1}]$ is not imposed. The estimation under the compactness of D has been achieved in (Álvarez-Liébana et al., 2017) and (Ruiz-Medina and Álvarez-Liébana, 2019b), providing a strongly consistent estimator of Γ in the Hilbert-Schmidt and trace operator norms, respectively, as well as from a Bayesian perspective in (Ruiz-Medina and Álvarez-Liébana, 2019a).

C.3.2 ARH(1) estimator and plug-in predictor

The follow theorem provides an strongly consistent estimator for the ARH(1) process. Note that weaker assumptions on the decay rate of eigenvalues are here considered, in comparison with those ones proposed in Bosq (2000)

Theorem C.1. *Let suppose that $\|\mathcal{X}_0\|_{\mathbb{H}} < \infty$ a.s., assume that the autocovariance operator $C := \mathbb{E}[\mathcal{X}_n \otimes \mathcal{X}_n]$ is a strictly positive operator and define $a_1 = 2\sqrt{2}(C_1 - C_2)^{-1}$ and $a_j = 2\sqrt{2}\max\{(C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\}$, for any $j \geqslant 2$. Given a n-dependent parameter p_n with $\hat{C}_{p_n} > 0$ a.s., $\lim_{n\to\infty} p_n = \infty$, but $\lim_{n\to\infty} p_n/n = 0$ and $\Lambda_{p_n} = \sup_{1\leqslant j\leqslant p_n} (C_j - C_{j+1})^{-1}$ such that $C_{p_n}^{-1}\sqrt{p_n}\Lambda_{k_n} = o\big(\sqrt{n/\ln(n)}\big)$, under Assumptions 12–13 and*

$$\sum_{j=1}^{p_n} \left\| \hat{\Psi}_j - \operatorname{sgn}\left(\langle \hat{\Psi}_j, \Psi_j\rangle_{\mathbb{H}}\right)\Psi_j \right\|_{\mathbb{H}} \to 0$$

a.s., if condition $C_{p_n}^{-1} p_n \sum_{j=1}^{p_n} a_j = o\big(\sqrt{n/\ln(n)}\big)$ is held, we have

$$\left\| \hat{\Gamma}_{p_n} - \Gamma \right\|_{\mathcal{L}(\mathbb{H})} \to 0 \quad \textit{a.s.}$$

Proof of Theorem C.1.

Firstly, let us prove the following preliminary result. Let us suppose $\|\mathcal{X}_0\|_{\mathbb{H}} < \infty$ a.s. Under strictly positiveness assumption on C, we define $a_1 = 2\sqrt{2}(C_1 - C_2)^{-1}$ and $a_j = 2\sqrt{2}\max\{(C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\}$, for any $j \geqslant 2$. Given a n-dependent parameter p_n with $\hat{C}_{p_n} > 0$ a.s. and $\lim_{n\to\infty} p_n = \infty$, but $\lim_{n\to\infty} p_n/n = 0$, and $\Lambda_{p_n} = \sup_{1\leqslant j\leqslant p_n} (C_j - C_{j+1})^{-1}$

such that $C_{p_n}^{-1}\sqrt{p_n}\Lambda_{k_n} = o\left(\sqrt{n/\ln(n)}\right)$, under Assumptions 12–13 and

$$\sum_{j=1}^{p_n}\left\|\hat{\Psi}_j - \operatorname{sgn}\left(\langle\hat{\Psi}_j,\Psi_j\rangle_{\mathbb{H}}\right)\Psi_j\right\|_{\mathbb{H}} \to 0$$

a.s., we will demonstrate that

$$\mathbb{P}\left(\left\|\hat{\Gamma}_{p_n}-\Gamma\right\|_{\mathcal{L}(\mathbb{H})} \geqslant \eta\right) \leqslant M\exp\left(-n\eta^2/Q_n\right),$$

where

$$Q_n = \mathcal{O}\left(\left(C_{p_n}^{-1}p_n\sum_{j=1}^{p_n}a_j\right)^2\right), \quad n\to\infty,$$

for any $\eta>0$, being M a finite constant.

On the one hand, since $\sup_{\mathcal{X}\in\mathbb{H},\ \|\mathcal{X}\|_{\mathbb{H}}\leqslant 1}\left\|\Gamma(\mathcal{X}) - \sum_{j=1}^{k}\langle\Gamma(\mathcal{X}),\Psi_j\rangle_{\mathbb{H}}\Psi_j\right\|_{\mathbb{H}} \to 0$, as $k\to\infty$, from Ruiz-Medina and Álvarez-Liébana (2019c, Lemma 9), if

$$\sum_{j=1}^{p_n}\left\|\hat{\Psi}_j - \operatorname{sgn}\left(\langle\hat{\Psi}_j,\Psi_j\rangle_{\mathbb{H}}\right)\Psi_j\right\|_{\mathbb{H}} \longrightarrow 0 \quad \text{a.s.},$$

as $n\to\infty$, we have $\sup_{\mathcal{X}\in\mathbb{H},\ \|\mathcal{X}\|_{\mathbb{H}}\leqslant 1}\left\|\Gamma(\mathcal{X}) - \sum_{j=1}^{p_n}\langle\Gamma(\mathcal{X}),\hat{\Psi}_j\rangle_{\mathbb{H}}\hat{\Psi}_j\right\|_{\mathbb{H}} \to 0.$

On the other hand, under Assumptions 12–13, if $\|\mathcal{X}_0\|_{\mathbb{H}}<\infty$ is held and eigenvalues of C verify that $\Lambda_{p_n} = \sup_{1\leqslant j\leqslant p_n}(C_j - C_{j+1})^{-1}$ such that $C_{p_n}^{-1}\sqrt{p_n}\Lambda_{k_n} = o\big(\sqrt{n/\ln(n)}\big)$, then from Ruiz-Medina and Álvarez-Liébana (2019c, Lemma 8),

$$\sup_{j=1,\ldots,p_n}\left\|\hat{\Psi}_j - \operatorname{sgn}\left(\langle\hat{\Psi}_j,\Psi_j\rangle_{\mathbb{H}}\right)\Psi_j\right\|_{\mathbb{H}} \to 0 \quad \text{a.s.},$$

as $n\to\infty$. As exposed in Ruiz-Medina and Álvarez-Liébana (2019c, Theorem 1), these asymptotic results and the conditions imposed ensure us the upper bound in probability

$$\mathbb{P}\left(\left\|\hat{\Gamma}_{p_n}-\Gamma\right\|_{\mathcal{L}(\mathbb{H})} \geqslant \eta\right) \leqslant M\exp\left(-n\eta^2/Q_n\right), \tag{C.4}$$

where

$$Q_n = \mathcal{O}\Big(\left(C_{p_n}^{-1}p_n\sum_{j=1}^{p_n}a_j\right)^2\Big), \quad n\to\infty,$$

for any $\eta > 0$, being M a finite constant. Furthermore, if $C_{p_n}^{-1} p_n \sum_{j=1}^{p_n} a_j = o\big(\sqrt{n/\ln(n)}\big)$, then the almost surely convergence of the estimator is directly achieved from the asymptotics established in (C.4). □

C.3.3 ARH(z) processes: theoretical framework

The Markovian nature of the ARH(1) model allows us to extend it by including more lagged functional regressors, i.e., $\mathcal{X}_n$ is inferred from $(\mathcal{X}_{n-1}, \ldots, \mathcal{X}_{n-z})$. Formally, we say that $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ is a zero mean autoregressive Hilbertian process of order $z \geqslant 1$, valued in $\mathbb{H} = L^2([0,h])$ and denoted as ARH(z), if the following state equation is satisfied

$$\mathcal{X}_n(t) = \sum_{r=1}^{z} \Gamma_r\left(\mathcal{X}_{n-r}\right)(t) + \mathcal{E}_n(t), \qquad \mathcal{X}_n, \mathcal{E}_n \in \mathbb{H} = L^2([0,h]), \tag{C.5}$$

with $n \in \mathbb{Z}$, $t \in [0,h]$ and order $z \geqslant 1$, and where $\{\Gamma_r\}_{r=1}^{z}$ are bounded linear operators in $\mathcal{L}(\mathbb{H})$, and $\{\mathcal{E}_n\}_{n\in\mathbb{Z}}$ is a $\mathbb{H}$-valued strong white noise, such that $\mathbb{P}\left(\Gamma_z\left(\mathcal{X}_n\right) \neq 0\right) > 0$ is implicitly assumed, for each $n \in \mathbb{Z}$, for the identifiability of z. Let us characterize the ARH(z) process in (C.5) as a particular multivariate ARH(1) process valued in $\mathbb{H}^z := \prod_{r=1}^{z} \mathbb{H}$, constituting likewise a separable Hilbert space (see Lemma 1 below).

Lemma 1. Given a finite set of separable Hilbert spaces $\{\mathbb{H}_r\}_{r=1}^{z}$, let $\mathbb{H}^z := \prod_{r=1}^{z} \mathbb{H}_r$ be its Cartesian product. Then, $\mathbb{H}^z$ becomes a separable Hilbert space equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}^z}$ given as $\big\langle \big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(z)}\big), \big(\mathcal{Y}_{(1)}, \ldots, \mathcal{Y}_{(z)}\big)\big\rangle_{\mathbb{H}^z} := \sum_{r=1}^{z} \big\langle \mathcal{X}_{(r)}, \mathcal{Y}_{(r)}\big\rangle_{\mathbb{H}_r}$, where $\big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(z)}\big) \in \mathbb{H}^z$ and $\big(\mathcal{Y}_{(1)}, \ldots, \mathcal{Y}_{(z)}\big) \in \mathbb{H}^z$, with $\mathcal{X}_{(r)}, \mathcal{Y}_{(r)} \in \mathbb{H}_r$ for each $r = 1, \ldots, z$.

Proof of Lemma 1. Firstly, let us prove that $\mathbb{H}^s = \prod_{r=1}^{s} \mathbb{H}_r$ becomes a normed space equipped with the norm

$$\big\|\big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big)\big\|_{\mathbb{H}^s} = \sqrt{\sum_{r=1}^{s} \big\|\mathcal{X}_{(r)}\big\|_{\mathbb{H}_r}^2}, \qquad \big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big) \in \prod_{r=1}^{s} \mathbb{H}_r.$$

Since $\{\mathbb{H}_r\}_{r=1}^{s}$ is a set of Hilbert spaces, in particular, $\mathbb{H}_r$ is a complete normed space, and then, $\big\|\mathcal{X}_{(r)}\big\|_{\mathbb{H}_r} \geqslant 0$, for each $\mathcal{X}_{(r)} \in \mathbb{H}_r$ and each $r = 1, \ldots, s$. Thus, $\big\|\big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big)\big\|_{\mathbb{H}^s} \geqslant 0$, such that $\big\|\big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big)\big\|_{\mathbb{H}^s} = 0$ if and only $\big\|\mathcal{X}_{(r)}\big\|_{\mathbb{H}_r}^2 = 0$, for any $r = 1, \ldots, s$, which happens if and only $\big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big)$ is the null element on $\prod_{r=1}^{s} \mathbb{H}_r$.

Let us now consider the elements $\widetilde{\mathcal{X}} = \big(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\big)$ and $\widetilde{\mathcal{Y}} = \big(\mathcal{Y}_{(1)}, \ldots, \mathcal{Y}_{(s)}\big)$, with $\widetilde{\mathcal{X}},\ \widetilde{\mathcal{Y}} \in \prod_{r=1}^{s} \mathbb{H}_r$, such that $\widetilde{\mathcal{X}} + \widetilde{\mathcal{Y}} = \big(\mathcal{X}_{(1)} + \mathcal{Y}_{(1)}, \ldots, \mathcal{X}_{(s)} + \mathcal{Y}_{(s)}\big) \in \prod_{r=1}^{s} \mathbb{H}_r$. Since

triangle inequality is held for each norm, we have

$$\begin{aligned}\left\|\widetilde{\mathcal{X}}+\widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}^2 &= \sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2 \leqslant \sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}\left(\left\|\mathcal{X}_{(r)}\right\|_{\mathbb{H}_r}+\left\|\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}\right)\\ &\leqslant \sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}\left\|\mathcal{X}_{(r)}\right\|_{\mathbb{H}_r}+\sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}\left\|\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}.\end{aligned}$$

Applying the Cauchy-Schwarz's inequality,

$$\begin{aligned}\left\|\widetilde{\mathcal{X}}+\widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}^2 &\leqslant \sqrt{\sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2}\sqrt{\sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}\right\|_{\mathbb{H}_r}^2}+\sqrt{\sum_{r=1}^{s}\left\|\mathcal{X}_{(r)}+\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2}\sqrt{\sum_{r=1}^{s}\left\|\mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2}\\ &= \left\|\widetilde{\mathcal{X}}+\widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}\left(\left\|\widetilde{\mathcal{X}}\right\|_{\mathbb{H}^s}+\left\|\widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}\right).\end{aligned}$$

In the general case that $\widetilde{\mathcal{X}}+\widetilde{\mathcal{Y}}\neq 0$, we can divide both sides of equation (C.6) by $\left\|\widetilde{\mathcal{X}}+\widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}$, obtaining the triangle inequality. Otherwise, the triangle property is directly achieved. Lastly, it can be noted that $\left\|c\widetilde{\mathcal{X}}\right\|_{\mathbb{H}^s}=\sqrt{\sum_{r=1}^{s}\left\|c\mathcal{X}_{(r)}\right\|_{\mathbb{H}_r}^2}=|c|\left\|\widetilde{\mathcal{X}}\right\|_{\mathbb{H}^s}$, for each $c\in\mathbb{R}$ and $\widetilde{\mathcal{X}}\in\mathbb{H}^s$. Hence, $(\mathbb{H}^s,\|\cdot\|_{\mathbb{H}^s})$ is a normed space.

Concerning completeness, consider a Cauchy sequence $\left\{\left(\mathcal{X}_{(1),j},\ldots,\mathcal{X}_{(s),j}\right),\ j\in\mathbb{N}\right\}$ on $\mathbb{H}^s$, under the norm $\|\cdot\|_{\mathbb{H}^s}$. That is, for any $\varepsilon>0$, there is a positive integer N such that, for each $m_1,m_2>N$,

$$\left\|\left(\mathcal{X}_{(1),m_1},\ldots,\mathcal{X}_{(s),m_1}\right)-\left(\mathcal{X}_{(1),m_2},\ldots,\mathcal{X}_{(s),m_2}\right)\right\|_{\mathbb{H}^s}=\sqrt{\sum_{r=1}^{s}\left\|\mathcal{X}_{(r),m_1}-\mathcal{X}_{(r),m_2}\right\|_{\mathbb{H}_r}^2}<\varepsilon,$$

leading to $\left\|\mathcal{X}_{(r),m_1}-\mathcal{X}_{(r),m_2}\right\|_{\mathbb{H}_r}<\varepsilon$, for each $r=1,\ldots,s$ and $m_1,m_2>N$. Then, $\left\{\mathcal{X}_{(r),j},\ j\in\mathbb{N}\right\}$ is a Cauchy sequence, for each $r=1,\ldots,s$, Since $\left\{\left(\mathbb{H}_r,\|\cdot\|_{\mathbb{H}_r}\right)\right\}_{r=1,\ldots,s}$ is a set of Hilbert (and therefore, complete) spaces, there exists a set $\left\{\mathcal{Y}_{(r)}\right\}_{r=1,\ldots,s}$ such that $\lim_{j\to\infty}\mathcal{X}_{(r),j}=\mathcal{Y}_{(r)}$, and so, $\lim_{j\to\infty}\left(\mathcal{X}_{(1),j},\ldots,\mathcal{X}_{(s),j}\right)=\left(\mathcal{Y}_{(1)},\ldots,\mathcal{Y}_{(s)}\right)$. Thus, any Cauchy sequence defined on $(\mathbb{H}^s,\|\cdot\|_{\mathbb{H}^s})$ has a limit on such space, and therefore, $(\mathbb{H}^s,\|\cdot\|_{\mathbb{H}^s})$ is a complete normed space.

On the other hand, concerning the separability of $(\mathbb{H}^s,\|\cdot\|_{\mathbb{H}^s})$, we must find a countable set D which is dense in $\mathbb{H}^s$, under the norm $\|\cdot\|_{\mathbb{H}^s}$. Since $\left(\mathbb{H}_r,\|\cdot\|_{\mathbb{H}_r}\right)$ is a separable Hilbert space, for each $r=1,\ldots,s$, there exist countable sets $\{C_r\}_{r=1}^{s}$ in a manner that $\overline{C_r}^{\|\cdot\|_{\mathbb{H}_r}}=\mathbb{H}_r$, for each $r=1,\ldots,s$. Since $D=C_1\times\ldots\times C_s$ is a countable subset in $\mathbb{H}^s$, as long as s is finite, it is enough to prove that D is a dense subset in $\mathbb{H}^s$, under the norm $\|\cdot\|_{\mathbb{H}^s}$. Let us consider,

for each $\widetilde{\mathcal{X}} \in \mathbb{H}^s$ and $\varepsilon > 0$, an open ball $\mathcal{B}_{\|\cdot\|_{\mathbb{H}^s}}\left(\widetilde{\mathcal{X}}, \varepsilon\right) = \left\{\widetilde{\mathcal{Y}} \in \mathbb{H}^s \colon \left\|\widetilde{\mathcal{X}} - \widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s} < \varepsilon\right\}$.

We know that, since C_r is dense in $\mathbb{H}_r$, for each $r = 1, \ldots, s$, $\mathcal{B}_{\|\cdot\|_{\mathbb{H}_r}}\left(\mathcal{X}_{(r)}, \frac{\sqrt{s}}{s}\varepsilon\right) \cap C_r \neq \varnothing$, for any $\varepsilon > 0$, and therefore, there exists an element $\mathcal{Z}_{(r)} \in \mathbb{H}_r$ such that, for each $r = 1, \ldots, s$,

$$\mathcal{Z}_{(r)} \in \mathcal{B}_{\|\cdot\|_{\mathbb{H}_r}}\left(\mathcal{X}_{(r)}, \frac{\sqrt{s}}{s}\varepsilon\right) \cap C_r, \qquad \left\|\mathcal{X}_{(r)} - \mathcal{Z}_{(r)}\right\|_{\mathbb{H}_r} < \frac{\sqrt{s}}{s}\varepsilon.$$

From equation (C.3.3), being $\left(\mathcal{Z}_{(1)}, \ldots, \mathcal{Z}_{(s)}\right) \in D$, then

$$\left\|\widetilde{\mathcal{X}} - \left(\mathcal{Z}_{(1)}, \ldots, \mathcal{Z}_{(s)}\right)\right\|_{\mathbb{H}^s} = \sqrt{\sum_{r=1}^{s} \left\|\mathcal{X}_{(r)} - \mathcal{Z}_{(r)}\right\|_{\mathbb{H}_r}^2} \leqslant \sqrt{\sum_{r=1}^{s} \frac{s}{s^2}\varepsilon^2} = \varepsilon,$$

which directly implies that, for each $\widetilde{\mathcal{X}} \in \mathbb{H}^s$ and $\varepsilon > 0$, there exists $\left(\mathcal{Z}_{(1)}, \ldots, \mathcal{Z}_{(s)}\right) \in \mathbb{H}^s$ verifying $\left(\mathcal{Z}_{(1)}, \ldots, \mathcal{Z}_{(s)}\right) \in \mathcal{B}_{\|\cdot\|_{\mathbb{H}^s}}\left(\widetilde{\mathcal{X}}, \varepsilon\right) \cap D$. The separability is then proved.

Lastly, we need to assess that the inner product defined in Lemma 1 is given by applying the parallelogram law to the norm $\|\cdot\|_{\mathbb{H}^s}$:

$$\begin{aligned}
\left\langle \left(\mathcal{X}_{(1)}, \ldots, \mathcal{X}_{(s)}\right), \left(\mathcal{Y}_{(1)}, \ldots, \mathcal{Y}_{(s)}\right) \right\rangle_{\mathbb{H}^s} &= \frac{\left\|\widetilde{\mathcal{X}} + \widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}^2 - \left\|\widetilde{\mathcal{X}} - \widetilde{\mathcal{Y}}\right\|_{\mathbb{H}^s}^2}{4} \\
&= \sum_{r=1}^{s} \frac{\left\|\mathcal{X}_{(r)} + \mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2 - \left\|\mathcal{X}_{(r)} - \mathcal{Y}_{(r)}\right\|_{\mathbb{H}_r}^2}{4} \\
&= \sum_{r=1}^{s} \left\langle \mathcal{X}_{(r)}, \mathcal{Y}_{(r)} \right\rangle_{\mathbb{H}_r}.
\end{aligned}$$

The proof is then completed. □

In this way, the following proposition allows us to characterize the ARH(z) in (C.5) as a $\mathbb{H}^z$-valued ARH(1) process, endowed with the norm $\langle \cdot, \cdot \rangle_{\mathbb{H}^z}$ defined in Lemma 1 above.

Proposition 3. Let $\mathcal{X} := \{\mathcal{X}_n\}_{n \in \mathbb{Z}}$ be a zero mean ARH(z) process valued in $\mathbb{H} = L^2([0, h])$, with $z \geqslant 1$. The ARH(z) model in (C.5) can be reinterpreted as

$$\overline{\mathcal{X}}_n = \begin{pmatrix} \mathcal{X}_n \\ \vdots \\ \mathcal{X}_{n-z+1} \end{pmatrix} \in \mathbb{H}^z, \quad \overline{\Gamma} = \begin{pmatrix} \Gamma_1 & \Gamma_2 & \ldots & \Gamma_{z-1} & \Gamma_z \\ Id_{\mathbb{H}} & 0_{\mathbb{H}} & \ldots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ 0_{\mathbb{H}} & Id_{\mathbb{H}} & \ldots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\mathbb{H}} & 0_{\mathbb{H}} & \ldots & Id_{\mathbb{H}} & 0_{\mathbb{H}} \end{pmatrix}, \tag{C.7}$$

and $\overline{\mathcal{E}}_n = (\mathcal{E}_n, \mathbf{0}, \ldots, \mathbf{0})^\top \in \mathbb{H}^z$, where $\mathbb{H}^z$ is the Cartesian product of z copies of $\mathbb{H}$. Thus, $\overline{\mathcal{X}}_n = \overline{\Gamma}\left(\overline{\mathcal{X}}_{n-1}\right) + \overline{\mathcal{E}}_n$, constitutes a $\mathbb{H}^z$-valued ARH(1) process, for each $n \in \mathbb{Z}$, with $\overline{\mathcal{E}}_n$ a $\mathbb{H}^z$-valued strong white noise. Furthermore, the joint operator $\overline{\Gamma}$ is also a bounded linear operator in $\mathbb{H}^z$. In (C.7), $Id_{\mathbb{H}}$ and $0_{\mathbb{H}}$ denote the identity and null operators, respectively, and $\mathbf{0}$ the null element on $\mathbb{H}$.

Proof of Proposition 3. Firstly, from Lemma 1, $\mathbb{H}^z := \prod_{r=1}^{z} \mathbb{H}$ is a separable Hilbert space. Secondly, by definition, it is clear that $\left\{\overline{\mathcal{E}}_n\right\}_{n\in\mathbb{Z}}$, with $\overline{\mathcal{E}}_n = (\mathcal{E}_n, \mathbf{0}, \ldots, \mathbf{0})^\top \in \mathbb{H}^z$, is also a $\mathbb{H}^z$-valued strong white noise.

Note that from the matrix definition of $\overline{\Gamma}$ provided by equation (C.7)

$$\overline{\Gamma}\left(\overline{\mathcal{X}}\right) = \begin{pmatrix} \Gamma_1 & \Gamma_2 & \ldots & \Gamma_{z-1} & \Gamma_z \\ Id_{\mathbb{H}} & 0_{\mathbb{H}} & \ldots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ 0_{\mathbb{H}} & Id_{\mathbb{H}} & \ldots & 0_{\mathbb{H}} & 0_{\mathbb{H}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{\mathbb{H}} & 0_{\mathbb{H}} & \ldots & Id_{\mathbb{H}} & 0_{\mathbb{H}} \end{pmatrix} \begin{pmatrix} \mathcal{X}_n \\ \mathcal{X}_{n-1} \\ \mathcal{X}_{n-2} \\ \vdots \\ \mathcal{X}_{n-z+1} \end{pmatrix} = \begin{pmatrix} \sum_{r=1}^{z} \Gamma_r\left(\mathcal{X}_{n-r+1}\right) \\ \mathcal{X}_n \\ \mathcal{X}_{n-1} \\ \vdots \\ \mathcal{X}_{n-z+2} \end{pmatrix} \in \mathbb{H}^z.$$

Since each Γ_r is a linear bounded operator, for any $r = 1, \ldots, z$, from the norm $\|\cdot\|_{\mathbb{H}^z}$ defined in Lemma 1 and the triangle inequality,

$$\begin{aligned} \left\|\overline{\Gamma}\right\|_{\mathcal{L}(\mathbb{H}^z)} &= \sup_{\|\overline{x}\|_{\mathbb{H}^z}\leqslant 1} \left\|\overline{\Gamma}\left(\overline{\mathcal{X}}\right)\right\|_{\mathbb{H}^z} = \sup_{\|\overline{x}\|_{\mathbb{H}^z}\leqslant 1} \sqrt{\left\|\sum_{r=1}^{z} \Gamma_r\left(\mathcal{X}_{n-r+1}\right)\right\|_{\mathbb{H}}^2 + \sum_{r=0}^{z-2} \left\|\mathcal{X}_{n-r}\right\|_{\mathbb{H}}^2} \\ &\leqslant \sup_{\|\overline{x}\|_{\mathbb{H}^z}\leqslant 1} \sqrt{\left(\sum_{r=1}^{z} \left\|\Gamma_r\left(\mathcal{X}_{n-r+1}\right)\right\|_{\mathbb{H}}\right)^2 + \sum_{r=1}^{z-1} \left\|\mathcal{X}_{n-r}\right\|_{\mathbb{H}}^2} < \infty, \end{aligned}$$

with $\|\mathcal{X}_n\|_{\mathbb{H}}^2 < \infty$, for each $n \in \mathbb{Z}$; that is, $\overline{\Gamma} \in \mathcal{L}\left(\mathbb{H}^z\right)$. Concerning the weak stationarity, note that

$$\mathbb{E}\left[\overline{\mathcal{X}}_n\right] = \left(\mathbb{E}\left[\mathcal{X}_n\right], \mathbb{E}\left[\mathcal{X}_{n-1}\right], \ldots, \mathbb{E}\left[\mathcal{X}_{n-z+1}\right]\right)^\top = (\mathbf{0}, \ldots, \mathbf{0})^\top, \quad n \in \mathbb{Z},$$

not depending on n. Furthermore, given $\overline{\mathcal{Y}} \in \mathbb{H}^z$, since the inner product and expectation are both linear functionals, then

$$\begin{aligned} \left\langle C_{\overline{\mathcal{X}}_n, \overline{\mathcal{X}}_m}\left(\overline{\mathcal{X}}\right), \overline{\mathcal{Y}}\right\rangle_{\mathbb{H}^z} &= \left\langle \mathbb{E}\left[\overline{\mathcal{X}}_n \otimes_{\mathbb{H}^z} \overline{\mathcal{X}}_m\right]\left(\overline{\mathcal{X}}\right), \overline{\mathcal{Y}}\right\rangle_{\mathbb{H}^z} = \sum_{r=1}^{z} \left\langle C_{\mathcal{X}_{n-r+1}, \mathcal{X}_{m-r+1}}\left(\mathcal{Y}_{(r)}\right), \mathcal{Y}_{(r)}\right\rangle_{\mathbb{H}} \\ &= \sum_{r=1}^{z} \left\langle C_{\mathcal{X}_0, \mathcal{X}_{m-n}}\left(\mathcal{Y}_{(r)}\right), \mathcal{Y}_{(r)}\right\rangle_{\mathbb{H}}, \end{aligned}$$

i.e., $\overline{C}_{\overline{\mathcal{X}}_n,\overline{\mathcal{X}}_m}$ only depends on $m-n$, and so, the weak stationarity is achieved. □

The sufficient condition for the existence of an unique stationary solution of a zero mean ARH(z) process, with $z \geqslant 1$, given as (C.5)–(C.7), is now provided in Proposition 4.

Proposition 4. Let $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ be a zero mean ARH(z) process, with $z \geqslant 1$, as explicitly defined in (C.5)–(C.7). If $\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)} < 1$, for any $k \geqslant k_0$, and for some $k_0 \geqslant 1$, then the state equation established in (C.5) has a unique stationary solution given by $\mathcal{X}_n = \sum_{j=0}^{\infty} \Pi_1\Big(\overline{\Gamma}^j\left(\overline{\mathcal{E}}_{n-j}\right)\Big)$, for each $n \in \mathbb{Z}$, where $\Pi_r\colon \left(\mathcal{X}_{(1)},\ldots,\mathcal{X}_{(z)}\right) \mapsto \mathcal{X}_{(r)}$ is the projector operator, for each $r = 1,\ldots,z$.

Proof of Proposition 4. On the one hand, from Lemma 1 and Proposition 3, the stochastic process $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ explicitly defined in (C.5)–(C.7) constitutes also a zero mean multidimensional ARH(1) process into a separable Hilbert space $\mathbb{H}^z := \prod_{r=1}^{z}\mathbb{H}_r$. Thus, as exposed in Remark 6, as long as Assumption 12 is held for $\overline{\Gamma}$ (that is, $\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)} < 1$, for any $k \geqslant k_0$, and for some $k_0 \geqslant 1$), the ARH(1) process $\overline{\mathcal{X}} = \left\{\overline{\mathcal{X}}_n\right\}_{n\in\mathbb{Z}}$ admits an unique stationary solution

$$\overline{\mathcal{X}}_n(t) = \sum_{j=0}^{\infty}\overline{\Gamma}^j\left(\overline{\mathcal{E}}_{n-j}\right)(t), \quad n \in \mathbb{Z}, \quad \overline{\Gamma} \in \mathcal{L}(\mathbb{H}^z).$$

On the other hand, from Markovian representation in (C.5)–(C.7),

$$\begin{aligned}\mathcal{X}_n = \Pi_1\left(\overline{\mathcal{X}}_n\right) &= \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}\overline{\mathcal{X}}_n \\ &= \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}\begin{pmatrix} \mathcal{X}_n \\ \mathcal{X}_{n-1} \\ \vdots \\ \mathcal{X}_{n-z+1} \end{pmatrix}, \quad n \in \mathbb{Z},\end{aligned}$$

such that, since Π_1 is a (linear) projection operator, the existence is proved as follows:

$$\mathcal{X}_n(t) = \Pi_1\left(\overline{\mathcal{X}}_n(t)\right) = \Pi_1\left(\sum_{j=0}^{\infty}\overline{\Gamma}^j\left(\overline{\mathcal{E}}_{n-j}\right)(t)\right) = \sum_{j=0}^{\infty}\left(\Pi_1\overline{\Gamma}^j\right)\left(\overline{\mathcal{E}}_{n-j}\right)(t).$$

Concerning the uniqueness, let us suppose that there exists an alternative stationary solution $\mathcal{X}_{n,1}(t)$. Applying (C.5)–(C.7), we could construct its associated $\mathbb{H}^z$-valued trajectory $\overline{\mathcal{X}}_{n,1}$, which would be a stationary solution of (C.7). However, under Assumption 12 and from

Proposition 3, the uniqueness of the multidimensional ARH(1) in (C.7) is derived, and thus, $\overline{\mathcal{X}}_{n,1} = \overline{\mathcal{X}}_n$ a.s. is verified. Hence,

$$\mathcal{X}_{n,1} = \Pi_1\left(\overline{\mathcal{X}}_{n,1}\right) = \Pi_1\left(\overline{\mathcal{X}}_n\right) = \mathcal{X}_n \quad a.s.$$

□

As proved below in Lemma 2, Assumption 12 on Γ implies that Assumption 12 is held for Γ_r, for each $r = 1, \ldots, z$ (i.e., stationarity condition in Proposition 4 above leads to $\sum_{n=0}^{\infty} \|\Gamma_r^n\|_{\mathcal{L}(\mathbb{H})} < \infty$).

Lemma 2. Let $\mathcal{X} = \{\mathcal{X}_n\}_{n\in\mathbb{Z}}$ be a zero mean ARH(z) process, with $z \geqslant 1$, as explicitly defined in (C.5) and reinterpreted as a $\mathbb{H}^z$-valued ARH(1) processes given in (C.7). If $\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)} < 1$, for any $k \geqslant k_0$, and for some $k_0 \geqslant 1$ (i.e., $\sum_{n=0}^{\infty} \left\|\overline{\Gamma}^n\right\|_{\mathcal{L}(\mathbb{H}^z)} < \infty$), then $\sum_{n=0}^{\infty} \|\Gamma_r^n\|_{\mathcal{L}(\mathbb{H})} < \infty$, for each $r = 1, \ldots, z$.

Proof of Lemma 2. Let us consider the projection operators $\Pi_i : \boldsymbol{x} \longmapsto (0, \ldots, x_i, \ldots, 0)$, where $\boldsymbol{x} = (x_1, \ldots, x_z) \in \mathbb{H}^z$, for each $i = 1, \ldots, z$, given by

$$\Pi_i(\boldsymbol{x}) = \begin{pmatrix} 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 1 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_z \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ x_i \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{x} \in \mathbb{H}^z,$$

where $\|\Pi_i\|_{\mathcal{L}(\mathbb{H}^z)} = \sup_{\|\boldsymbol{x}\|_{\mathbb{H}^z}<1} \|\Pi_i(\boldsymbol{x})\|_{\mathbb{H}^z} = \sup_{\|\boldsymbol{x}\|_{\mathbb{H}^z}<1} \|(0, \ldots, x_i, \ldots, 0)\|_{\mathbb{H}^z}$. From Lemma 1, we directly get that if $\|\boldsymbol{x}\|_{\mathbb{H}^z} < 1$, we have $\|x_i\|_{\mathbb{H}^z} < 1$ for each $i = 1, \ldots, z$, and then, $\|\Pi_i\|_{\mathcal{L}(\mathbb{H}^z)} < 1$ is directly achieved.

On the other hand, from the linearity of $\overline{\Gamma}$ derived in Proposition 3, the set of linear operators $\{\Gamma_r\}_{r=1}^z$ can be restated as

$$\Pi_1\overline{\Gamma}\Pi_i(\boldsymbol{x}) = \Pi_1\overline{\Gamma}\begin{pmatrix} 0 \\ \vdots \\ x_i \\ \vdots \\ 0 \end{pmatrix} = \Pi_1\begin{pmatrix} \Gamma_i(x_i) \\ \vdots \\ x_i \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \dashrightarrow \text{first row} \\ \\ \longrightarrow (i+1)\text{-th row} \\ \\ \dashrightarrow z\text{-th row} \end{matrix}, \quad \text{if } i < z,$$

and

$$\Pi_1\overline{\Gamma}\Pi_z\left(\boldsymbol{x}\right)=\Pi_1\overline{\Gamma}\begin{pmatrix}0\\ \vdots\\ 0\\ \vdots\\ x_z\end{pmatrix}=\Pi_1\begin{pmatrix}\Gamma_z\left(x_z\right)\\ \vdots\\ 0\\ \vdots\\ 0\end{pmatrix},\quad \text{if } i=z,$$

and therefore, $\Gamma_i\left(x_i\right)=\Pi_1\overline{\Gamma}\Pi_i\left(\boldsymbol{x}\right)$, for each $i=1,\ldots,z$. Furthermore, it is trivial to obtain that, for some $k\geqslant 1$, $\Gamma_i^k\left(x_i\right)=\left(\Pi_1\overline{\Gamma}\Pi_i\right)^k\left(\boldsymbol{x}\right)=\Pi_1\overline{\Gamma}^k\Pi_i\left(\boldsymbol{x}\right)$, for each $i=1,\ldots,z$. Thus, if $\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)}<1$, for any $k\geqslant k_0$, and for some $k_0\geqslant 1$ (i.e., Assumption 12 over $\overline{\Gamma}$), then

$$\begin{aligned}\left\|\Gamma_i^k\right\|_{\mathcal{L}(\mathbb{H})} &= \sup_{\|x\|_{\mathbb{H}}<1}\left\|\Gamma_i^k(x_i)\right\|_{\mathbb{H}}=\sup_{\|x\|_{\mathbb{H}}<1}\left\|\left(\Pi_1\overline{\Gamma}^k\Pi_i\right)\left(0,\ldots,x_i,\ldots,0\right)^\top\right\|_{\mathbb{H}}\\ &\leqslant \sup_{\|\boldsymbol{x}\|_{\mathbb{H}^z}<1}\left\|\left(\Pi_1\overline{\Gamma}^k\Pi_i\right)\left(\boldsymbol{x}\right)\right\|_{\mathbb{H}^z}\leqslant\left\|\Pi_1\right\|_{\mathcal{L}(\mathbb{H}^z)}\left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)}\left\|\Pi_i\right\|_{\mathcal{L}(\mathbb{H}^z)}\\ &< \left\|\overline{\Gamma}^k\right\|_{\mathcal{L}(\mathbb{H}^z)}<1,\end{aligned}$$

with $i=1,\ldots,z$. Hence, Assumption 12 over $\overline{\Gamma}$ implies Assumption 12 is held for each $\{\Gamma_r\}_{r=1}^z$. □

Declaration for publications
incorporated in the thesis

This appendix provides information regarding the published and unpublished articles included in this dissertation. The sections are devoted to those chapters reproducing the contents of published papers (Chapter 3) and to those whose contents have been submitted to a journal (Chapters 4, 6, 7 and 8) but not yet published.

Published contents of Chapter 3

The contents of Chapter 3 appear on López-Pérez et al. (2021), published as López-Pérez, A.[†], Febrero-Bande, M.[†], and González-Manteiga, W.[†] (2021). Parametric estimation of diffusion processes: A review and comparative study. *Mathematics*, 9(8):859, doi: `https://doi.org/10.3390/math9080859`, ISSN 2227-7390.

[†] Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Rúa de Lope Gómez de Marzoa s/n, 15705, Santiago de Compostela, Spain.

Contribution:

Conceptualization, methodology, software development and implementation, formal analysis, investigation, writing (original draft preparation), review and editing.

Impact factor:

Chapter 3 was published in a journal with impact factor 2.592 (2021 Journal Citation Reports), a Cite Score index of 2.9 and a Q1 JCR category rank in Mathematics.

Journal permissions:

Information about permissions can be found in the publisher's website `https://www.mdpi.com/authors/rights`, where the following statement is placed: "*For all articles published in MDPI journals, copyright is retained by the authors. Articles are licensed under an open access Creative Commons CC BY 4.0 license, meaning that anyone may download and read the paper for free. In addition, the article may be reused and quoted provided that the original published version is cited. These conditions allow for maximum use and exposure of the work, while ensuring that the authors receive proper credit.*"

Submitted contents of Chapters 4 and 7

The contents of Chapters 4 and 7 have been submitted to a journal, collected in the manuscript López-Pérez et al. (2022b) as López-Pérez, A.[†], Febrero-Bande, M.[†], and González-Manteiga, W.[†] (2022). Estimation and specification test for diffusion models with stochastic volatility. *Submitted manuscript*, arXiv:2208.08415.

[†] Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Rúa de Lope Gómez de Marzoa s/n, 15705, Santiago de Compostela, Spain.

Contribution:

Conceptualization, methodology, software development and implementation, formal analysis, investigation, writing (original draft preparation), review and editing.

Submitted contents of Chapter 6

The contents of Chapter 6 have been submitted to a journal, collected in the manuscript López-Pérez et al. (2022a) as López-Pérez, A.[†], Febrero-Bande, M.[†], and González-Manteiga, W.[†] (2022). A comparative review of specification tests for diffusion models. *Submitted manuscript*, arXiv:2208.08420.

[†] Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Rúa de Lope Gómez de Marzoa s/n, 15705, Santiago de Compostela, Spain.

Contribution:
Conceptualization, methodology, software development and implementation, formal analysis, investigation, writing (original draft preparation), review and editing.

Submitted contents of Chapter 8

The contents of Chapter 8 have been submitted to a journal, collected in the manuscript Álvarez-Liébana et al. (2022) as Álvarez-Liébana, J.[‡], López-Pérez, A.[†], Febrero-Bande, M.[†], and González-Manteiga, W.[†] (2022). A goodness-of-fit test for functional time series with applications to Ornstein-Uhlenbeck processes. *Submitted manuscript*, arXiv:2206.12821.

[†] Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Rúa de Lope Gómez de Marzoa s/n, 15705, Santiago de Compostela, Spain.
[‡] Department of Statistics and Data Science. Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, Ciudad Universitaria, Avenida Puerta de Hierro, s/n, 28040, Madrid, Spain.

Contribution:
Conceptualization, methodology, software development and implementation, formal analysis, investigation, writing (original draft preparation), review and editing.

Resumen en castellano

Los modelos de difusión descritos por ecuaciones diferenciales estocásticas son una extensión natural de modelos deterministas con ecuaciones diferenciales ordinarias. Los procesos de difusión son ampliamente utilizados en la modelización de fenómenos dinámicos en tiempo continuo y su aplicación ha demandado el desarrollo de nuevas metodologías estadísticas. Estos modelos han sido ampliamente utilizados en el campo de la matemática financiera para el análisis de precios de activos financieros, tipos de cambio o tipos de interés. El uso de modelos en tiempo continuo para explicar las dinámicas de los tipos de interés fue extendido a partir el trabajo pionero de Merton (1975) y, desde entonces, comprender y modelar la estructura temporal de los tipos de interés representa uno de los principales retos de la investigación financiera. Los tipos de interés son series que varían de forma aleatoria respecto al tiempo y son un elemento central en el ámbito financiero, ya que su evolución afecta a distintas variables y agentes económicos, desde la valoración de activos financieros hasta la toma de decisiones políticas en el ámbito económico.

El principal objetivo de esta tesis es la inferencia estadística en modelos en tiempo continuo, proponiendo nuevos tests de bondad de ajuste para modelos de difusión y abarcando el estudio de la estimación de los procesos, ya que impacta en la implementación de los contrastes. A continuación se detallan los capítulos que componen la tesis, indicando las principales contribuciones realizadas en cada uno de ellos.

Capítulo 1: Introducción

En el Capítulo 1 se introducen los modelos de difusión descritos por ecuaciones diferenciales estocásticas, cuya tractabilidad ha convertido a los modelos en tiempo continuo en una pieza central de la teoría de fijación de precios de activos. Dado un espacio de probabilidad dotado de filtración $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geqslant 0}, \mathbb{P})$, se define la ecuación diferencial estocástica

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t, \qquad t \in [0, T],$$

con $X_0 = x_0 \in \mathbb{R}$, W_t un proceso de Wiener adaptado a $\{\mathcal{F}\}_{t\geqslant 0}$ y $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, con $d \in \mathbb{Z}^+$, un vector de parámetros desconocidos. X_t es un proceso de Itô, donde la función $m(\cdot)$ es el *drift* o tendencia determinista, que describe los cambios de X_t en tiempo t, y las perturbaciones aleatorias son introducidas por la función de difusión o volatilidad $\sigma(\cdot)$.

El paradigma en tiempo continuo ha demostrado ser una herramienta especialmente útil, pero su naturaleza conlleva ciertas dificultades que complican la identificación y estimación de los modelos de difusión, ya que las observaciones están disponibles en tiempo discreto. La discretización de los modelos en tiempo continuo conlleva la introducción de un sesgo de discretización que, conjuntamente con el sesgo muestral, complica la estimación de los modelos de difusión y puede deteriorar el calibrado de los test de bondad de ajuste. Además, la persistencia que muestran las series de tipo de interés agudiza el problema de la estimación.

El objetivo de esta tesis es el desarrollo de nuevos contrastes de bondad de ajuste para modelos de difusión en tiempo continuo, considerando procesos de difusión con volatilidad determinista y estocástica, así como procesos de Itô como series de tiempo funcionales. Estas propuestas permitirán estudiar el ajuste de la especificación del modelo de difusión respecto las dinámicas de los tipos de interés. Además, dado el impacto de la estimación, donde los factores latentes y la naturaleza en tiempo continuo dificulta el proceso de estimación de los modelos, este problema será abordado. Por ello, antes de la contribución metodológica de los contrastes de bondad de ajuste, se debatirá sobre distintas implementaciones de métodos de estimación, desde el punto de vista paramétrico y no-paramétrico.

Capítulo 2: Procesos de difusión y la estructura temporal de los tipos de interés

El Capítulo 2 define los conceptos fundamentales sobre los que se construyen los modelos de difusión. Aunque esta tesis se centra en modelar el comportamiento empírico de los tipos de interés desde una perspectiva estadística, la formulación de las ecuaciones diferenciales estocásticas utilizan el cálculo estocástico y, por ello, se realiza una breve introducción al mismo para poder definir de forma rigurosa los procesos de difusión en tiempo continuo, objeto de estudio en este documento. Además, también se introducen las principales familias paramétricas de modelos de difusión utilizadas en el campo econométrico. El modelo propuesto en Chan et al. (1992),

$$\mathrm{d}X_t = (\alpha + \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

denotado por las siglas CKLS, establece una parametrización general donde diferentes especificaciones paramétricas pueden ser anidadas. Este modelo será objeto de estudio en las aplicaciones a datos reales a lo largo del documento de tesis, analizando su ajuste para diferentes

series de tiempo de tipos de interés, en su versión con función de volatilidad determinista y con volatilidad estocástica.

Capítulo 3: Estimación paramétrica de modelos de difusión

Una vez introducidos los conceptos básicos sobre los procesos de difusión y la importancia de la modelización de los tipos de interés, el Capítulo 3 abre el primer bloque de la tesis, dedicado a la estimación de los modelos de difusión. Este capítulo se centra en la estimación paramétrica de los modelos con volatilidad determinista, comparando distintos métodos de estimación para evaluar el impacto en la estimación de la persistencia y del sesgo de discretización. Los métodos de máxima verosimilitud (Ozaki, 1992; Shoji y Ozaki, 1998), expansión de los polinomios de Hermite (Aït-Sahalia, 2002), filtro de Kalman (1960), métodos de Markov Chain Monte Carlo (Elerian et al., 2001; Eraker, 2001) y el método generalizado de los momentos (Hansen, 1982), han sido implementados en un paquete de R (`estsde`, incluido en el Apéndice B) y comparados en un estudio de simulación.

En el estudio de simulación se evalúan los distintos métodos bajo diferentes escenarios, evaluando el impacto de la discretización y el sesgo de estimación introducido por la persistencia de las series de tiempo de tipos de interés. De este estudio se concluye que el sesgo de estimación surge principalmente en el parámetro que controla la velocidad de reversión a la media, donde a mayor persistencia mayor sesgo, mientras que el sesgo debido a la discretización se aprecia en escenarios con baja persistencia y una mayor frecuencia de discretización (valores altos de $\Delta = t_i - t_{i-1}$). A diferencia de las series en tiempo discreto, donde el sesgo y la varianza de los estimadores está controlado por el tamaño muestral n, en series en tiempo continuo con observaciones tomadas en tiempo discreto el sesgo y la varianza en la estimación de los parámetros del drift están dominados por el tiempo total de observación $T = \Delta n$. Esto hace que el sesgo y la varianza se reduzcan cuando $T \to \infty$, a diferencia de las series de tiempo clásicas que se reducen cuando solamente $n \to \infty$. Respecto a los métodos de estimación, los resultados son similares, siendo el método generalizado de los momentos el menos eficiente y los métodos basados en simulación los más computacionalmente exigentes y cuya implementación cambia con cada especificación paramétrica.

El capítulo finaliza con una aplicación de las metodologías a datos de euríbor, estimando la especificación paramétrica general del modelo CKLS para los distintos vencimientos. Las estimaciones de los parámetros obtenidas con los distintos métodos no muestran discrepancias significativas y en todas las series la volatilidad incrementa a medida que aumenta el tipo de interés. La especificación de la función drift no se rechaza pero la función de volatilidad determinista no logra capturar las dinámicas de los tipos de interés de la serie del euríbor.

Capítulo 4: Estimación paramétrica de modelos de difusión con volatilidad estocástica

En el Capítulo 4 se extiende la revisión de métodos de estimación a los modelos de difusión con volatilidad estocástica,

$$\begin{aligned} \mathrm{d}X_t &= m_1(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma_t \nu_1(X_t, \boldsymbol{\theta})\,\mathrm{d}W_{1,t}, \\ \mathrm{d}g(\sigma_t) &= m_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}t + \nu_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}W_{2,t}, \end{aligned}$$

realizando un estudio de simulación para evaluar la estimación de los parámetros cuando una de las variables es no observable (σ_t). En el estudio de Monte Carlo se extienden métodos de estimación introducidos en el capítulo anterior a los modelos con volatilidad estocástica y se incluye el filtro de partículas o Bayesiano. Los métodos considerados son el filtro de Kalman adaptado por Shumway y Stoffer (2000), los métodos de Markov Chain Monte Carlo de Shephard (1993) y Jacquier et al. (1994), el filtro de partículas de Liu y West (2001) y los métodos de la transformada de Fourier de Cuchiero y Teichmann (2015) y Merkle et al. (2020). Debido a que las estimaciones obtenidas con los métodos son muy similares, se concluye que el filtro de Kalman es un método computacionalmente eficiente para utilizar en los contrastes de bondad de ajuste y, además, escribir el modelo en la estructura de espacio de estados permite implementar procedimientos bootstrap para el calibrado de los test.

En el estudio de datos reales se extiende el análisis de los datos del euríbor del Capítulo 3, considerando el modelo CKLS con volatilidad estocástica y estimando la trayectoria no observable de la volatilidad. Posteriormente, en el Capítulo 7 se contrastará la especificación de este modelo para los datos de euríbor.

Capítulo 5: Estimación no-paramétrica de modelos de difusión

El Capítulo 5 finaliza el bloque dedicado a la estimación, centrándose en la estimación no-paramétrica de las funciones drift y volatilidad de los modelos de difusión, así como la densidad marginal o estacionaria. El análisis está centrado en métodos de suavizado y se han realizado estudios de simulación para comparar distintas ventanas. Para mitigar el impacto de la persistencia en la estimación, se ha utilizado el criterio de selección de ventana propuesto en De Brabanter et al. (2018), desarrollado para modelos de diseño aleatorio con correlación en los errores. Las conclusiones extraídas de la comparación de los criterios de selección de ventana difieren entre la estimación de la densidad y de las funciones drift y volatilidad. La regla del pulgar, el principal criterio utilizado en el ámbito econométrico, obtiene resultados similares a la validación cruzada en la estimación de la densidad marginal. Sin embargo, esta

ventana no resulta adecuada para las funciones drift y volatilidad. El MISE obtenido utilizando ventanas de validación cruzada o considerando una ventana piloto con una función kernel bimodal que satisfaga $K(0) = 0$, como en De Brabanter et al. (2018), es menor. Mayores niveles de persistencia empeoran la estimación, como ya se estudió en el Capítulo 3 en el caso de la estimación paramétrica, lo que podría indicar que tener el cuenta la velocidad de la reversión a la media en el cálculo de el parámetro de suavizado podría resultar en un criterio de selección de ventana adaptado para los modelos de difusión econométricos. Esta línea de investigación está abierta y apenas ha sido considerada en la literatura de los modelos de difusión.

Como aplicación a datos reales, se analizan once series de tiempo correspondientes a diferentes vencimientos de bonos del tesoro estadounidenses, estimando la densidad marginal y las funciones drift y volatilidad, comparando la estimación no-paramétrica con la estimación paramétrica del modelo CKLS. Se concluye que para vencimientos largos ambas estimaciones son próximas, mientras que para las letras del tesoro (vencimientos de 1, 3, 6 y 12 meses) difieren considerablemente. Esta serie de datos es analizada en más profundidad en el siguiente capítulo mediante el uso de diferentes tests de bondad de ajuste.

Capítulo 6: Contrastes de bondad de ajuste para modelos de difusión

El Capítulo 6 abre el segundo bloque de la tesis, dedicado a los test de bondad de ajuste. El capítulo comienza realizando una revisión sobre los desarrollos en contrastes de bondad de ajuste para modelos de difusión para, posteriormente, proponer un nuevo test de bondad de ajuste para la especificación paramétrica de los modelos de difusión basado en las ideas de correlación de distancias (Székely et al., 2007). Dado el modelo de difusión discretizado,

$$Y_{t_i} = m(X_{t_i}) + \frac{\sigma(X_{t_i})}{\sqrt{\Delta}}\varepsilon_{t_i},$$

con $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ y $\varepsilon_{t_i} \sim \mathcal{N}(0,1)$, para $i = 0, \ldots, n-1$, definida en el intervalo $[0,1]$ con $\Delta = 1/n$. Bajo estacionaridad y ergodicidad, $\varepsilon_{t_i} = \big(Y_{t_i} - m(X_{t_i})\big)/\big(\sigma(X_{t_i})/\sqrt{\Delta}\big)$ son una secuencia i.i.d. independiente del proceso estocástico $\{X_{t_i}\}_{i=0}^{n-1}$ con $\mathbb{E}\left[\varepsilon_{t_i} \,\middle|\, \mathcal{F}_t\right] = 0$. Por ello, la especificación paramétrica del drift y de la volatilidad del modelo de difusión se pueden testar contrastando la independencia de las variables $\{\varepsilon_{t_i}, X_{t_i}\}_{i=0}^{n-1}$, considerando la hipótesis nula $\mathcal{H}_0 \colon \varphi_{\varepsilon,X} = \varphi_\varepsilon \varphi_X$, donde φ denota la función característica. El estadístico de

contraste utilizado viene dado por

$$T_n = n\frac{\mathcal{V}_n^2(\varepsilon_{t_i}, X_{t_i})}{S_2},$$

donde la distancia de covarianzas se define como $\mathcal{V}_n^2(\varepsilon_t, X_t; \mu) = \|\varphi_{\varepsilon,X}(t,s) - \varphi_\varepsilon(t)\varphi_X(s)\|_\mu^2$ y $S_2 = n^{-2}\sum_{k,l=1}^n |\varepsilon_{t_k} - \varepsilon_{t_l}| n^{-2}\sum_{k,l=1}^n |X_{t_k} - X_{t_l}|$. Esta nueva propuesta es comparada bajo diferentes escenarios con diferentes metodologías de bondad de ajuste, además de analizar el efecto de la maldición de la dimensionalidad. Los resultados empíricos muestran resultados similares de los métodos en términos de calibrado y potencia, alcanzando el nivel nominal bajo la hipótesis nula y detectando diversas hipótesis alternativas. Sin embargo, cuando la dimensión incrementa, los tests pierden potencia. Es en este escenario donde el test propuesto basado en la correlación de distancias supera al resto, evitando el desastre de la dimensionalidad. Este efecto se comprueba considerando procesos autorregresivos en tiempo continuo, CAR(p), para diferentes órdenes $p \in \{1, 2, 3, 4, 5\}$.

El capítulo finaliza con una aplicación a datos reales, en concreto se contrastan series de tipos de interés, con datos de bonos del tesoro estadounidenses con distintos vencimientos, de 1 mes a 30 años. En este análisis se han encontrado evidencias empíricas sobre la especificación paramétrica de la función de difusión del modelo CKLS, $\sigma(X_t, \boldsymbol{\theta}) = \sigma X_t^\gamma$, para vencimientos a largo plazo. Cuando el vencimiento disminuye, los p-valores son cercanos a cero, indicando que para las letras del tesoro (vencimientos de 1, 3, 6 y 12 meses) la especificación del CKLS se rechaza. Las conclusiones extraídas concuerdan con el análisis realizado en el Capítulo 5, donde la estimación paramétrica del modelo CKLS y la estimación no-paramétrica de la función de volatilidad diferían para las letras del tesoro, mientras que para vencimientos a más largo plazo se aproximaban.

Capítulo 7: Contraste de bondad de ajuste para modelos de difusión con volatilidad estocástica

Basado en las ideas de González-Manteiga et al. (2017), en el Capítulo 7 se propone un test de bondad de ajuste para modelos de difusión con volatilidad estocástica, para contrastar la especificación de las funciones drift $m_1(\cdot)$ y volatilidad $\nu_1(\cdot)$,

$$\mathcal{H}_{0\nu}\colon \nu_1 \in \{\nu_1(\cdot, \boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\},$$
$$\mathcal{H}_{0m}\colon m_1 \in \{m_1(\cdot, \boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\}.$$

El test para la función drift está basado en la función de regresión integrada del procesos,

mientras que el test para la función de difusión se basa en la función de volatilidad integrada. Los estadísticos están basados en la distancia del proceso empírico resultante a su esperanza esperada de cero, medida por los funcionales de Kolmogorov-Smirnov y Cramér-von Mises. En ambos tests, la distribución del estadístico se aproxima mediante metodología bootstrap, utilizando el filtro de Kalman del Capítulo 4. El estudio de simulación del test propuesto no incluye una comparativa con otros contrastes de bondad de ajuste ya que hasta la fecha, según nuestro conocimiento, no hay ningún contraste para la especificación de la función de volatilidad para modelos de difusión con volatilidad estocástica.

El estudio de Monte Carlo para muestras finitas ilustra el buen calibrado de los tests para las funciones drift y volatilidad bajo hipótesis simples y compuestas, así como la potencia bajo diferentes alternativas. Además, el uso del filtro de Kalman para la estimación y el calibrado del test ofrece una implementación que simplifica el procedimiento de contraste.

Como aplicación a datos reales, se utilizan las series de tiempo del euríbor del Capítulo 3—donde se contrastaba y rechazaba la función de volatilidad determinista—para testar la especificación de la función de difusión con volatilidad estocástica. La incorporación de un parámetro estocástico permite capturar las dinámicas de los tipos de interés, de modo que la volatilidad depende de un factor adicional que varía de forma independiente al nivel del tipo de interés.

Capítulo 8: Test de especificación para modelos de difusión como series de tiempo funcionales

El último capítulo dedicado a los test de bondad de ajuste para modelos de difusión considera un contexto de alta frecuencia, ya que los datos financieros pueden ser recogidos como una secuencia de curvas a través del tiempo (p.e., precios intra-diarios). En los capítulos anteriores, la secuencia de observaciones del proceso de difusión se consideraba como una sola trayectoria, ignorando el patrón diario del proceso. Trabajando en un contexto funcional se puede retener la información contenida en la forma de las curvas, tratando las curvas diarias como un objeto estadístico en lugar de una colección individual de observaciones. En el Capítulo 8 se realizan dos contribuciones: un test de bondad de ajuste para series de tiempo autorregresivas hilbertianas (ARH) y un test de especificación para modelos de difusión caracterizados como ARH.

Respecto al contraste de bondad de ajuste para series de tiempo funcionales, el test propuesto extiende las ideas de García-Portugués et al. (2021) a los procesos ARH(z), con $z \in \mathbb{Z}$, contrastando una hipótesis nula compuesta mediante un estadístico de Cramér-von Mises. La novedad de esta contribución es doble, por un lado permite contrastar los modelos ARH(z)

y, por otro, permite realizar un proceso secuencial para determinar el orden autorregresivo z de un modelo ARH(z). La primera contribución no tiene, según nuestro conocimiento, ningún competidor, por lo que se realiza un estudio comparativo del procedimiento multi-etapa de detección del orden autorregresivo con la propuesta de Kokoszka y Reimherr (2013). El estudio muestra el buen comportamiento de la propuesta bajo las hipótesis nula y alternativa, siendo competitivo incluso bajo alternativas no lineales, donde el test de Kokoszka y Reimherr (2013) pierde potencia.

Extendiendo el test de bondad de ajuste para series de tiempo funcionales a procesos de difusión, se propone un nuevo test de especificación para modelos de difusión desde una perspectiva de alta frecuencia. En concreto, se testan los procesos de Ornstein-Uhlenbeck caracterizándolos como un proceso ARH(1) con una metodología en dos etapas donde, primero, se contrasta si el proceso es un ARH(1) y, posteriormente, bajo linealidad, se testa la forma paramétrica del operador mediante un F test funcional calibrado utilizando bootstrap paramétrico. Esta metodología en dos etapas es evaluada bajo diferentes escenarios en un amplio estudio de simulación y se aplica a datos reales de tipos de cambio, con observaciones 5 minutales, donde se concluye que un modelo como el Ornstein-Uhlenbeck puede capturar las dinámicas de las curvas diarias de los pares EURUSD y GBPUSD, mientras que para el tipo de cambio EURGBP se rechaza la hipótesis nula en la primera etapa, es decir, el modelo ARH(1) no parece adecuado para explicar la evolución temporal de las curvas diarias.

Capítulo 9: Conclusiones y discusión

En el Capítulo 9 se recopilan las principales conclusiones extraídas en cada capítulo y se debaten líneas de investigación abiertas, como la propuesta de un parámetro de suavizado para la función de regresión que considere la persistencia del proceso (Capítulo 5); la extensión de la metodología propuesta en el Capítulo 8 para otras especificaciones paramétricas, estudiando la posibilidad de reescribir los modelos de difusión como ARH(1); la propuesta de un modelo autorregresivo funcional para modelizar la estructura temporal de los tipos de interés; o el estudio del fenómeno de *stock pinning* desde la perspectiva funcional.

Apéndices

Los Apéndices A y C contienen material suplementario para los Capítulos 3 y 8, respectivamente. El Apéndice B recopila la documentación del paquete de R `estsde`, que recoge las principales implementaciones computaciones de esta tesis.

Resumo en galego

Os modelos de difusión descritos por ecuacións diferenciais estocásticas son unha extensión natural de modelos deterministas con ecuacións diferenciais ordinarias. Os procesos de difusión son amplamente utilizados na modelización de fenómenos dinámicos en tempo continuo e a súa aplicación demandou o desenvolvemento de novas metodoloxías estatísticas. Estes modelos foron amplamente empregados no campo da matemática financeira para a análise de prezos de activos financeiros, tipos de cambio ou tipos de interese. O uso de modelos en tempo continuo para explicar as dinámicas dos tipos de interese estendeuse a partir do traballo pioneiro de Merton (1975) e, dende ese momento, comprender e modelar a estrutura temporal dos tipos de interese representa un dos principais retos da investigación financeira. Os tipos de interese son series que varían de forma aleatoria respecto ao tempo e son un elemento central no ámbito financeiro, xa que a súa evolución afecta a distintas variables e axentes económicos, desde a valoración de activos financeiros ata a toma de decisións políticas no ámbito económico.

O principal obxectivo desta tese é a inferencia estatística en modelos en tempo continuo, propoñendo novos tests de bondade de axuste para modelos de difusión e abarcando o estudo da estimación dos procesos, xa que impacta na implementación dos contrastes. A continuación detállanse os capítulos que compoñen a tese, indicando as principais contribucións realizadas en cada un deles.

Capítulo 1: Introdución

No Capítulo 1 introdúcense os modelos de difusión descritos por ecuacións diferenciais estocásticas, cuxa tractabilidad converteu aos modelos en tempo continuo nunha peza central da teoría de fixación de prezos de activos. Dado un espazo de probabilidade dotado de filtración $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geqslant 0}, \mathbb{P})$, defínese a ecuación diferencial estocástica

$$\mathrm{d}X_t = m(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma(X_t, \boldsymbol{\theta})\,\mathrm{d}W_t, \qquad t \in [0, T],$$

con $X_0 = x_0 \in \mathbb{R}$, W_t un proceso de Wiener adaptado a $\{\mathcal{F}\}_{t\geqslant 0}$ e $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, con $d \in \mathbb{Z}^+$, un vector de parámetros descoñecidos. X_t é un proceso de Itô, onde a función $m(\cdot)$ é o *drift* ou tendencia determinista, que describe os cambios de X_t en tempo t, e as perturbacións aleatorias son introducidas pola función de difusión ou volatilidade $\sigma(\cdot)$.

A paradigma en tempo continuo demostrou ser unha ferramenta especialmente útil, pero a súa natureza conleva certas dificultades que complican a identificación e estimación dos modelos de difusión, xa que as observacións están dispoñibles en tempo discreto. A discretización dos modelos en tempo continuo conleva a introdución dun nesgo de discretización que, conxuntamente co nesgo mostral, complica a estimación dos modelos de difusión e pode deteriorar o calibrado dos test de bondade de axuste. Ademais, a persistencia que mostran as series de tipo de interese agudiza o problema da estimación.

O obxectivo desta tese é o desenvolvemento de novos contrastes de bondade de axuste para modelos de difusión en tempo continuo, considerando procesos de difusión con volatilidade determinista e estocástica, así como procesos de Itô como series de tempo funcionais. Estas propostas permitirán estudar o axuste da especificación do modelo de difusión respecto as dinámicas dos tipos de interese. Ademais, dado o impacto da estimación, onde os factores latentes e a natureza en tempo continuo dificulta o proceso de estimación dos modelos, este problema será abordado. Por iso, antes da contribución metodolóxica dos contrastes de bondade de axuste, debaterase sobre distintas implementacións de métodos de estimación, desde o punto de vista paramétrico e non-paramétrico.

Capítulo 2: Procesos de difusión e a estrutura temporal dos tipos de interese

O Capítulo 2 define os conceptos fundamentais sobre os que se constrúen os modelos de difusión. Aínda que esta tese céntrase en modelar o comportamento empírico dos tipos de interese desde unha perspectiva estatística, a formulación das ecuacións diferenciais estocásticas utilizan o cálculo estocástico e, por iso, realízase unha breve introdución ao mesmo para poder definir de forma rigorosa os procesos de difusión en tempo continuo, obxecto de estudo neste documento. Ademais, tamén se introducen as principais familias paramétricas de modelos de difusión utilizadas no campo econométrico. O modelo proposto en Chan et al. (1992),

$$\mathrm{d}X_t = (\alpha + \kappa X_t)\,\mathrm{d}t + \sigma X_t^{\gamma}\,\mathrm{d}W_t,$$

denotado polas siglas CKLS, establece unha parametrización xeral onde diferentes especificacións paramétricas poden ser aniñadas. Este modelo será obxecto de estudo nas aplicacións a datos reais ao longo do documento de tese, analizando o seu axuste para diferentes series

de tempo de tipos de interese, na súa versión con función de volatilidade determinista e con volatilidade estocástica.

Capítulo 3: Estimación paramétrica de modelos de difusión

Unha vez introducidos os conceptos básicos sobre os procesos de difusión e a importancia da modelización dos tipos de interese, o Capítulo 3 abre o primeiro bloque da tese, dedicado á estimación dos modelos de difusión. Este capítulo céntrase na estimación paramétrica dos modelos con volatilidade determinista, comparando distintos métodos de estimación para avaliar o impacto na estimación da persistencia e do nesgo de discretización. Os métodos de máxima verosimilitud (Ozaki, 1992; Shoji e Ozaki, 1998), expansión dos polinomios de Hermite (Aït-Sahalia, 2002), filtro de Kalman (1960), métodos de Markov Chain Monte Carlo (Elerian et al., 2001; Eraker, 2001) e o método xeneralizado dos momentos (Hansen, 1982), foron implementados nun paquete de R (`estsde`, incluído no Apéndice B) e comparados nun estudo de simulación.

No estudo de simulación avalíanse os distintos métodos baixo diferentes escenarios, avaliando o impacto da discretización e o nesgo de estimación introducido pola persistencia das series de tempo de tipos de interese. Deste estudo conclúese que o nesgo de estimación xorde principalmente no parámetro que controla a velocidade de reversión á media, onde a maior persistencia maior nesgo, mentres que o nesgo debido á discretización apréciase en escenarios con baixa persistencia e unha maior frecuencia de discretización (valores altos de $\Delta = t_i - t_{i-1}$). A diferenza das series en tempo discreto, onde o nesgo e a varianza dos estimadores está controlado polo tamaño mostral n, en series en tempo continuo con observacións tomadas en tempo discreto o nesgo e a varianza na estimación dos parámetros do drift están dominados polo tempo total de observación $T = \Delta n$. Isto fai que o nesgo e a varianza redúzanse cando $T \to \infty$, a diferenza das series de tempo clásicas que se reducen cando soamente $n \to \infty$. Respecto a os métodos de estimación, os resultados son similares, sendo o método xeneralizado dos momentos o menos eficiente e os métodos baseados en simulación os máis computacionalmente esixentes e cuxa implementación cambia con cada especificación paramétrica.

O capítulo remata cunha aplicación das metodoloxías a datos de euríbor, estimando a especificación paramétrica xeral do modelo CKLS para os distintos vencementos. As estimacións dos parámetros obtidas cos distintos métodos non mostran discrepancias significativas e en todas as series a volatilidade incrementa a medida que aumenta o tipo de interese. A especificación da función drift non se rexeita pero a función de volatilidade determinista non logra capturar as dinámicas dos tipos de interese da serie do euríbor.

Capítulo 4: Estimación paramétrica de modelos de difusión con volatilidade estocástica

No Capítulo 4 esténdese a revisión de métodos de estimación aos modelos de difusión con volatilidade estocástica,

$$\begin{aligned} \mathrm{d}X_t &= m_1(X_t, \boldsymbol{\theta})\,\mathrm{d}t + \sigma_t \nu_1(X_t, \boldsymbol{\theta})\,\mathrm{d}W_{1,t}, \\ \mathrm{d}g(\sigma_t) &= m_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}t + \nu_2(g(\sigma_t), \boldsymbol{\theta})\,\mathrm{d}W_{2,t}, \end{aligned}$$

realizando un estudo de simulación para avaliar a estimación dos parámetros cando unha das variables é non observable (σ_t). No estudo de Monte Carlo esténdense métodos de estimación introducidos no capítulo anterior aos modelos con volatilidade estocástica e inclúese o filtro de partículas ou Bayesiano. Os métodos considerados son o filtro de Kalman adaptado por Shumway e Stoffer (2000), os métodos de Markov Chain Monte Carlo de Shephard (1993) e Jacquier et al. (1994), o filtro de partículas de Liu e West (2001) e os métodos da transformada de Fourier de Cuchiero e Teichmann (2015) e Merkle et al. (2020). Debido a que as estimacións obtidas cos métodos son moi similares, conclúese que o filtro de Kalman é un método computacionalmente eficiente para utilizar nos contrastes de bondade de axuste e, ademais, escribir o modelo na estrutura de espazo de estados permite implementar procedementos bootstrap para o calibrado dos test.

No estudo de datos reais esténdese a análise dos datos do euríbor do Capítulo 3, considerando o modelo CKLS con volatilidade estocástica e estimando a traxectoria non observable da volatilidade. Posteriormente, no Capítulo 7 contrastarase a especificación deste modelo para os datos do euríbor.

Capítulo 5: Estimación non-paramétrica de modelos de difusión

O Capítulo 5 finaliza o bloque dedicado á estimación, centrándose na estimación non-paramétrica das funcións drift e volatilidade dos modelos de difusión, así como a densidade marxinal ou estacionaria. A análise está centrada en métodos de suavizado e realizáronse estudos de simulación para comparar distintas xanelas. Para mitigar o impacto da persistencia na estimación, utilizouse o criterio de selección de xanela proposto en De Brabanter et al. (2018), desenvolvido para modelos de deseño aleatorio con correlación nos erros. As conclusións extraídas da comparación dos criterios de selección de xanela difiren entre a estimación da densidade e das funcións drift e volatilidade. A regra do polgar, o principal criterio utilizado no ámbito econométrico, obtén resultados similares á validación cruzada na estimación da densidade marxinal. Con todo, esta xanela non resulta adecuada para as funcións drift e

volatilidade. O MISE obtido utilizando xanelas de validación cruzada ou considerando unha xanela piloto cunha función kernel bimodal que satisfaga $K(0) = 0$, como en De Brabanter et al. (2018), é menor. Maiores niveis de persistencia empeoran a estimación, como xa se estudou no Capítulo 3 no caso da estimación paramétrica, o que podería indicar que ter o conta a velocidade da reversión á media no cálculo do parámetro de suavizado podería resultar nun criterio de selección de xanela adaptado para os modelos de difusión econométricos. Esta liña de investigación está aberta e apenas foi considerada na literatura dos modelos de difusión.

Como aplicación a datos reais, analízanse once series de tempo correspondentes a diferentes vencementos de bonos do tesouro estadounidenses, estimando a densidade marxinal e as funcións drift e volatilidade, comparando a estimación non-paramétrica coa estimación paramétrica do modelo CKLS. Conclúese que para vencementos longos ambas estimacións son próximas, mentres que para as letras do tesouro (vencementos de 1, 3, 6 e 12 meses) difiren considerablemente. Esta serie de datos é analizada en máis profundidade no seguinte capítulo mediante o uso de diferentes tests de bondade de axuste.

Capítulo 6: Contrastes de bondade de axuste para modelos de difusión

O Capítulo 6 abre o segundo bloque da tese, dedicado aos test de bondade de axuste. O capítulo comeza realizando unha revisión sobre os desenvolvementos en contrastes de bondade de axuste para modelos de difusión para, posteriormente, propoñer un novo test de bondade de axuste para a especificación paramétrica dos modelos de difusión baseado nas ideas de correlación de distancias (Székely et al., 2007). Dado o modelo de difusión discretizado,

$$Y_{t_i} = m(X_{t_i}) + \frac{\sigma(X_{t_i})}{\sqrt{\Delta}}\varepsilon_{t_i},$$

con $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ y $\varepsilon_{t_i} \sim \mathcal{N}(0,1)$, para $i = 0, \ldots, n-1$, definida no intervalo $[0,1]$ con $\Delta = 1/n$. Baixo estacionaridade e ergodicidade, $\varepsilon_{t_i} = \big(Y_{t_i} - m(X_{t_i})\big)/\big(\sigma(X_{t_i})/\sqrt{\Delta}\big)$ son unha secuencia i.i.d. independente do proceso estocástico $\{X_{t_i}\}_{i=0}^{n-1}$ con $\mathbb{E}\left[\varepsilon_{t_i} \,\middle|\, \mathcal{F}_t\right] = 0$. Por iso, a especificación paramétrica do drift e da volatilidade do modelo de difusión pódense testar contrastando a independencia das variables $\{\varepsilon_{t_i}, X_{t_i}\}_{i=0}^{n-1}$, considerando a hipótese nula $\mathcal{H}_0\colon \varphi_{\varepsilon,X} = \varphi_\varepsilon\varphi_X$, onde φ denota a función característica. O estadístico de contraste utilizado ven dado por

$$T_n = n\frac{\mathcal{V}_n^2(\varepsilon_{t_i}, X_{t_i})}{S_2},$$

onde a distancia de covarianzas defínese como $\mathcal{V}_n^2(\varepsilon_t, X_t; \mu) = \|\varphi_{\varepsilon,X}(t,s) - \varphi_\varepsilon(t)\varphi_X(s)\|_\mu^2$ e $S_2 = n^{-2}\sum_{k,l=1}^n |\varepsilon_{t_k} - \varepsilon_{t_l}| n^{-2}\sum_{k,l=1}^n |X_{t_k} - X_{t_l}|$. Esta nova proposta é comparada baixo

diferentes escenarios con diferentes metodoloxías de bondade de axuste, ademais de analizar o efecto da maldición da dimensionalidad. Os resultados empíricos mostran resultados similares dos métodos en termos de calibrado e potencia, alcanzando o nivel nominal baixo a hipótese nula e detectando diversas hipóteses alternativas. Con todo, cando a dimensión incrementa, os tests perden potencia. É neste escenario onde o test proposto baseado na correlación de distancias supera ao resto, evitando o desastre da dimensionalidad. Este efecto compróbase considerando procesos autorregresivos en tempo continuo, CAR(p), para diferentes ordes $p \in \{1, 2, 3, 4, 5\}$.

O capítulo finaliza cunha aplicación a datos reais, en concreto contrástanse series de tipos de interese, con datos de bonos do tesouro estadounidenses con distintos vencementos, de 1 mes a 30 anos. Nesta análise atopáronse evidencias empíricas sobre a especificación paramétrica da función de difusión do modelo CKLS, $\sigma(X_t, \boldsymbol{\theta}) = \sigma X_t^{\gamma}$, para vencementos a longo prazo. Cando o vencemento diminúe, os p-valores son próximos a cero, indicando que para as letras do tesouro (vencementos de 1, 3, 6 e 12 meses) a especificación do CKLS rexéitase. As conclusións extraídas concordan coa análise realizada no Capítulo 5, onde a estimación paramétrica do modelo CKLS e a estimación non-paramétrica da función de volatilidade diferían para as letras do tesouro, mentres que para vencementos a máis longo prazo aproximábanse.

Capítulo 7: Contraste de bondade de axuste para modelos de difusión con volatilidade estocástica

Baseado nas ideas de González-Manteiga et al. (2017), no Capítulo 7 proponse un test de bondade de axuste para modelos de difusión con volatilidade estocástica, para contrastar a especificación das funcións drift $m_1(\cdot)$ e volatilidade $\nu_1(\cdot)$,

$$\mathcal{H}_{0\nu} \colon \nu_1 \in \{\nu_1(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\},$$
$$\mathcal{H}_{0m} \colon m_1 \in \{m_1(\cdot, \boldsymbol{\theta}) \colon \boldsymbol{\theta} \in \Theta\}.$$

O test para a función drift está baseado na función de regresión integrada do procesos, mentres que o test para a función de difusión baséase na función de volatilidade integrada. Os estatísticos están baseados na distancia do proceso empírico resultante á súa esperanza esperada de cero, medida polos funcionais de Kolmogorov-Smirnov e Cramér-von Mises. En ambos os tests, a distribución do estatístico aproxímase mediante metodoloxía bootstrap, utilizando o filtro de Kalman do Capítulo 4. O estudo de simulación do test proposto non inclúe unha comparativa con outros contrastes de bondade de axuste xa que ata a data, segundo o noso coñecemento, non hai ningún contraste para a especificación da función de volatilidade

para modelos de difusión con volatilidade estocástica.

O estudo de Monte Carlo para mostras finitas ilustra o bo calibrado dos tests para as funcións drift e volatilidade baixo hipóteses simples e compostas, así como a potencia baixo diferentes alternativas. Ademais, o uso do filtro de Kalman para a estimación e o calibrado do test ofrece unha implementación que simplifica o procedemento de contraste.

Como aplicación a datos reais, utilízanse as series de tempo do euríbor do Capítulo 3—onde se contrastaba e rexeitaba a función de volatilidade determinista—para testar a especificación da función de difusión con volatilidade estocástica. A incorporación dun parámetro estocástico permite capturar as dinámicas dos tipos de interese, de modo que a volatilidade depende dun factor adicional que varía de forma independente ao nivel do tipo de interese.

Capítulo 8: Test de especificación para modelos de difusión como series de tempo funcionais

O último capítulo dedicado aos test de bondade de axuste para modelos de difusión considera un contexto de alta frecuencia, xa que os datos financeiros poden ser recollidos como unha secuencia de curvas a través do tempo (p.e., prezos intra-diarios). Nos capítulos anteriores, a secuencia de observacións do proceso de difusión considerábase como unha soa traxectoria, ignorando o patrón diario do proceso. Traballando nun contexto funcional pódese reter a información contida na forma das curvas, tratando as curvas diarias como un obxecto estatístico en lugar dunha colección individual de observacións. No Capítulo 8 realízanse dúas contribucións: un test de bondade de axuste para series de tempo autorregresivas hilbertianas (ARH) e un test de especificación para modelos de difusión caracterizados como ARH.

Respecto ao contraste de bondade de axuste para series de tempo funcionais, o test proposto estende as ideas de García-Portugués et al. (2021) aos procesos ARH(z), con $z \in \mathbb{Z}$, contrastando unha hipótese nula composta mediante un estatístico de Cramér-von Mises. A novidade desta contribución é dobre, por unha banda permite contrastar os modelos ARH(z) e, por outro, permite realizar un proceso secuencial para determinar a orde autorregresiva z dun modelo ARH(z). A primeira contribución non ten, segundo o noso coñecemento, ningún competidor, polo que se realiza un estudo comparativo do procedemento multi-etapa de detección da orde autorregresivo coa proposta de Kokoszka e Reimherr (2013). O estudo mostra o bo comportamento da proposta baixo as hipótese nula e alternativa, sendo competitivo mesmo baixo alternativas non lineais, onde o test de Kokoszka e Reimherr (2013) perde potencia.

Estendendo o test de bondade de axuste para series de tempo funcionais a procesos de difusión, proponse un novo test de especificación para modelos de difusión desde unha pers-

pectiva de alta frecuencia. En concreto, se testan os procesos de Ornstein-Uhlenbeck caracterizándoos como un proceso ARH(1) cunha metodoloxía en dúas etapas onde, primeiro, contrástase se o proceso é un ARH(1) e, posteriormente, baixo linealidade, se testa a forma paramétrica do operador mediante un F test funcional calibrado empregando bootstrap paramétrico. Esta metodoloxía en dúas etapas é avaliada baixo diferentes escenarios nun amplo estudo de simulación e aplícase a datos reais de tipos de cambio, con observacións 5 minutales, onde se conclúe que un modelo como o Ornstein-Uhlenbeck pode capturar as dinámicas das curvas diarias dos pares EURUSD e GBPUSD, mentres que para o tipo de cambio EURGBP rexéitase a hipótese nula na primeira etapa, é dicir, o modelo ARH(1) non parece adecuado para explicar a evolución temporal das curvas diarias.

Capítulo 9: Conclusións e discusión

No Capítulo 9 recompílanse as principais conclusións extraídas en cada capítulo e debátense liñas de investigación abertas, como a proposta dun parámetro de suavizado para a función de regresión que considere a persistencia do proceso (Capítulo 5); a extensión da metodoloxía proposta no Capítulo 8 para outras especificacións paramétricas, estudando a posibilidade de reescribir os modelos de difusión como ARH(1); a proposta dun modelo autorregresivo funcional para modelizar a estrutura temporal dos tipos de interese; ou o estudo do fenómeno de *stock pinning* desde a perspectiva funcional.

Apéndices

Os Apéndices A e C conteñen material suplementario para os Capítulos 3 e 8, respectivamente. O Apéndice B recompila a documentación do paquete de R `estsde`, que recolle as principais implementacións computacións desta tese.

References

Ahn, D.-H. and Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *The Review of Financial Studies*, 12(4):721–762.

Aït-Sahalia, Y. (1996a). Do interest rates really follow continuous-time Markov diffusions? Working paper, Graduate School of Business, University of Chicago.

Aït-Sahalia, Y. (1996b). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64(3):527–560.

Aït-Sahalia, Y. (1996c). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9(2):385–426.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *The Journal of Finance*, 54(4):1361–1395.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.

Aït-Sahalia, Y., Fan, J., and Peng, H. (2009). Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association*, 104(487):1102–1116.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., Li, C., and Li, C. X. (2020). Maximum likelihood estimation of latent Markov models using closed-form approximations. *Journal of Econometrics*.

Aït-Sahalia, Y., Mykland, P., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18(2):351–416.

Aït-Sahalia, Y. and Park, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *Journal of Econometrics*, 169(2):279–292.

Álvarez-Liébana, J., Bosq, D., and Ruiz-Medina, M. D. (2016). Consistency of the plug-in functional predictor of the Ornstein–Uhlenbeck process in Hilbert and Banach spaces. *Statistics & Probability Letters*, 117:12–22.

Álvarez-Liébana, J., Bosq, D., and Ruiz-Medina, M. D. (2017). Asymptotic properties of a component-wise ARH(1) plug-in predictor. *Journal of Multivariate Analysis*, 155:12–34.

Álvarez-Liébana, J., López-Pérez, A., Febrero-Bande, M., and González-Manteiga, W. (2022). A goodness-of-fit test for functional time series with applications to Ornstein-Uhlenbeck processes. Submitted manuscript, arXiv:2206.12821.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115–158.

Andersen, T. G. and Lund, J. (1997a). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77(2):343–377.

Andersen, T. G. and Lund, J. (1997b). Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve. Working paper, L. Kellogg Graduate School of Management, Northwestern University.

Andersen, T. G. and Sørensen, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business & Economic Statistics*, 14(3):328–352.

Aneiros, G. and Vieu, P. (2014). Variable selection in infinite-dimensional problems. *Statistics & Probability Letters*, 94:12–20.

Arapis, M. and Gao, J. (2006). Empirical comparisons in short-term interest rate models using nonparametric methods. *Journal of Financial Econometrics*, 4(2):310–345.

Artigas, J. C. and Tsay, R. S. (2004). Efficient estimation of stochastic diffusion models with leverage effects and jumps. Working paper, Graduate School of Business, University of Chicago.

Ball, C. and Roma, A. (1994). Target zone modelling and estimation for European Monetary System exchange rates. *Journal of Empirical Finance*, 1(3-4):385–420.

Ball, C. A. and Torous, W. N. (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, 3(2):215–238.

Bandi, F. M. and Phillips, P. C. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283.

Bass, R. F. (2011). *Stochastic processes*, volume 33. Cambridge University Press.

Bates, D. S. (2006). Maximum likelihood estimation of latent affine processes. *The Review of Financial Studies*, 19(3):909–965.

Benatia, D., Carrasco, M., and Florens, J. P. (2017). Functional linear regression with functional response. *Journal of Econometrics*, 201(2):269–291.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.

Berrendero, J. R., Cuevas, A., and Torrecilla, J. L. (2016). Variable selection in functional data classification: a maxima-hunting proposal. *Statistica Sinica*, 26:619–638.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics. Springer, New York.

Brennan, M. J. and Schwartz, E. S. (1979). A continuous time approach to the pricing of bonds. *Journal of Banking & Finance*, 3(2):133–155.

Brennan, M. J. and Schwartz, E. S. (1980). Analyzing convertible bonds. *Journal of Financial and Quantitative analysis*, 15(4):907–929.

Brenner, R. J., Harjes, R. H., and Kroner, K. F. (1996). Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis*, 31(1):85–107.

Brites, N. M. and Braumann, C. A. (2020). Stochastic differential equations harvesting policies: Allee effects, logistic-like growth and profit optimization. *Applied stochastic models in business and industry*, 36(5):825–835.

Broto, C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys*, 18(5):613–649.

Brown, S. J. and Dybvig, P. H. (1986). The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates. *The Journal of Finance*, 41(3):617–630.

Brugière, P. (1991). Estimation de la variance d'un processus de diffusion dans le cas multidimensionnel. *Comptes Rendus de l'Académie des Sciences*, 312(13):999–1004.

Brugière, P. (1993). Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus de diffusion multidimensionnelle. In *Annales de l'IHP Probabilités et statistiques*, volume 29, pages 357–389.

Bull, A. D. (2017). Semimartingale detection and goodness-of-fit tests. *The Annals of Statistics*, 45(3):1254–1283.

Caines, P. E. (1988). *Linear Stochastic Systems*. Wiley, New York.

Carvalho, C. M., Johannes, M. S., Lopes, H. F., and Polson, N. G. (2010). Particle learning and smoothing. *Statistical Science*, 25(1):88–106.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209–1227.

Chen, B. and Hong, Y. (2010). Characteristic function–based testing for multifactor continuous-time Markov models via nonparametric regression. *Econometric Theory*, 26(4):1115–1179.

Chen, Q., Hu, M., and Song, X. (2019). A nonparametric specification test for the volatility functions of diffusion processes. *Econometric Reviews*, 38(5):557–576.

Chen, Q., Zheng, X., and Pan, Z. (2015). Asymptotically distribution-free tests for the volatility function of a diffusion. *Journal of Econometrics*, 184(1):124–144.

Chen, S. X. and Gao, J. (2011). Simultaneous specification testing of mean and variance structures in nonlinear time series regression. *Econometric Theory*, 27(4):792–843.

Chen, S. X., Gao, J., and Tang, C. Y. (2008). A test for model specification of diffusion processes. *The Annals of Statistics*, 36(1):167–198.

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.

Christensen, K., Thyrsgaard, M., and Veliyev, B. (2019). The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing. *Journal of Econometrics*, 212(2):556–583.

Christoffersen, P., Heston, S., and Jacobs, K. (2009). The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932.

Constantinides, G. M. and Ingersoll, J. E. (1982). Optimal bond trading with personal tax: Implications for bond prices and estimated tax brackets and yield curves. *The Journal of Finance*, 37(2):349–352.

Corradi, V. and Swanson, N. R. (2005). Bootstrap specification tests for diffusion processes. *Journal of Econometrics*, 124(1):117–148.

Corradi, V. and White, H. (1999). Specification tests for the variance of a diffusion. *Journal of Time Series Analysis*, 20(3):253–270.

Courtadon, G. (1982). The pricing of options on default-free bonds. *Journal of Financial and Quantitative Analysis*, 17(1):75–100.

Cox, J. (1975). Notes on option pricing I: Constant elasticity of variance diffusions. Working paper, Stanford University.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1980). An analysis of variable rate loan contracts. *The Journal of Finance*, 35(2):389–403.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica*, 53(2):363–384.

Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1-2):145–166.

Cuchiero, C. and Teichmann, J. (2015). Fourier transform methods for pathwise covariance estimation in the presence of jumps. *Stochastic Processes and their Applications*, 125(1):116–160.

Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M., and González-Manteiga, W. (2019). Smoothing splines estimators for functional linear regression. *Annals of Statistics*, 47(1):439–467.

Cuevas, A., Febrero-Bande, M., and Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.

Dai, X. and Müller, H. G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Annals of Statistics*, 46(6B):3334–3361.

Danielsson, J. and Richard, J.-F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8(S1):S153–S173.

Davis, R. A., Matsui, M., Mikosch, T., and Wan, P. (2018). Applications of distance correlation to time series. *Bernoulli*, 24(4A):3087–3116.

De Brabanter, K., Cao, F., Gijbels, I., and Opsomer, J. (2018). Local polynomial regression with correlated errors in random design and unknown correlation structure. *Biometrika*, 105(3):681–690.

Dehling, H., Matsui, M., Mikosch, T., Samorodnitsky, G., and Tafakori, L. (2020). Distance covariance for discretized stochastic processes. *Bernoulli*, 26(4):2758–2789.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–22.

Dette, H. and Podolskij, M. (2008). Testing the parametric form of the volatility in continuous time diffusion models: A stochastic process approach. *Journal of Econometrics*, 143(1):56–73.

Dette, H., Podolskij, M., and Vetter, M. (2006). Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scandinavian Journal of Statistics*, 33(2):259–278.

Dette, H. and und Wilkau, C. v. L. (2003). On a test for a parametric form of volatility in continuous time financial models. *Finance and Stochastics*, 7(3):363–384.

Diebolt, J. (1995). A nonparametric test for the regression function: Asymptotic theory. *Journal of Statistical Planning and Inference*, 44(1):1–17.

Diebolt, J. and Zuber, J. (1999). Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statistics & Probability Letters*, 42(1):53–60.

Diebolt, J. and Zuber, J. (2001). On testing the goodness-of-fit of nonlinear heteroscedastic regression models. *Communications in Statistics-Simulation and Computation*, 30(1):195–216.

Dothan, L. U. (1978). On the term structure of interest rates. *Journal of Financial Economics*, 6(1):59–69.

Dufour, J.-M. and Valéry, P. (2009). Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. *Journal of Econometrics*, 150(2):193–206.

Durham, G. B. (2006). Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics*, 133(1):273–305.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–338.

Durrett, R. (2012). *Essentials of stochastic processes*. Springer.

Ebner, B., Klar, B., and Meintanis, S. G. (2018). Fourier inference for stochastic volatility models with heavy-tailed innovations. *Statistical Papers*, 59(3):1043–1060.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Elerian, O. (1998). A note on the existence of a closed form conditional transition density for the Milstein scheme. Working paper, Nuffield College, Oxford University.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, 20(4):317–337.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, 13(4):965–992.

Fan, J. and Zhang, C. (2003). A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, 98(461):118–134.

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, 29(1):153–193.

Ferreira, E. and Gil-Bazo, J. (2004). Beyond single-factor affine term structure models. *Journal of Financial Econometrics*, 2(4):565–591.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30(4):790–804.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, 105(491):1113–1125.

Gabrys, R. and Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, 102(480):1338–1348.

Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association*, 93(441):10–24.

Gao, J. and Casas, I. (2008). Specification testing in discretized diffusion models: Theory and practice. *Journal of Econometrics*, 147(1):131–140.

Gao, J. and King, M. (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory*, 20(5):844–882.

García-Portugués, E., Álvarez-Liébana, J., Álvarez-Pérez, G., and González-Manteiga, W. (2021). A goodness-of-fit test for the functional linear model with functional response. *Scandinavian Journal of Statistics*, 48(2):502–528.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational & Graphical Statistics*, 23(3):761–778.

Gelfand, I. M. and Vilenkin, N. Y. (1964). *Generalized Functions*. Academic Press. New York.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l'IHP Probabilités et statistiques*, volume 29, pages 119–151. Gauthier-Villars.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scandinavian Journal of Statistics*, 21(3):192–221.

Genon-Catalot, V., Laredo, C., and Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, 19:317–335.

Gibbons, M. R. and Ramaswamy, K. (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. *The Review of Financial Studies*, 6(3):619–658.

González-Manteiga, W., Zubelli, J. P., Monsalve-Cobis, A., and Febrero-Bande, M. (2017). Goodness–of–fit test for stochastic volatility models. In *From Statistics to Mathematical Finance*, pages 89–104. Springer.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F-radar and Signal Processing*, volume 140, pages 107–113. IET.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054.

Harrison, J. M. and Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381–408.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge university press.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Hille, E. and Phillips, R. S. (1957). *Functional Analysis and Semi-groups*. American Mathematical Society, Providence, RI.

Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79(1):135–163.

Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *The Review of Financial Studies*, 18(1):37–84.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics*, 38(3):1845–1884.

Horváth, L., Husková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, 101(2):352–367.

Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179:66–82.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300.

Hurn, A., Lindsay, K., and McClelland, A. (2013). A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. *Journal of Econometrics*, 172(1):106–126.

Hurn, A. S., Jeisman, J., and Lindsay, K. A. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics*, 5(3):390–455.

Imaizumi, M. and Kato, K. (2018). PCA–based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis*, 163:15–36.

Jacobsen, M. (2002). Optimality and small δ-optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Jiang, G. J. and Knight, J. L. (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory*, 13(5):615–645.

Jimenez, J., Biscay, R., and Ozaki, T. (2005). Inference methods for discretely observed continuous-time stochastic volatility models: A commented overview. *Asia-Pacific Financial Markets*, 12(2):109–141.

Johannes, M. and Polson, N. (2010). MCMC methods for continuous-time financial econometrics. In *Handbook of Financial Econometrics: Applications*, pages 1–72. Elsevier.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, 22(3):389–395.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(2):152–192.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., and Chopin, N. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

Karatzas, I. and Shreve, S. (1998). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, New York.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.

Khmaladze, E. V. (1979). The use of ω^2 tests for testing parametric hypotheses. *Theory of Probability & Its Applications*, 24(2):283–301.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications*, 26(2):240–257.

Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, 32(3):995–1034.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

Kirichenko, A. and Nikitin, Y. (2014). Precise small deviations in l^2 of some Gaussian processes appearing in the regression context. *Central European Journal of Mathematics*, 12(11):1674–1686.

Kloeden, P. E. and Platen, E. (1992a). Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of Statistical Physics*, 66(1-2):283–314.

Kloeden, P. E. and Platen, E. (1992b). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag Berlin Heidelberg.

Koedijk, K. G., Nissen, F. G., Schotman, P. C., and Wolff, C. C. (1997). The dynamics of short-term interest rate volatility reconsidered. *Review of Finance*, 1(1):105–130.

Kokoszka, P., Maslova, I., Sojka, J., and Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics*, 36(2):207–222.

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, 34(1):116–129.

Kotecha, J. H. and Djuric, P. M. (2003). Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612.

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Annals of Statistics*, 27(1):204–236.

Kristensen, D. (2011). Semi-nonparametric estimation and misspecification testing of diffusion models. *Journal of Econometrics*, 164(2):382–403.

Kroll, M. (2021). Asymptotic behaviour of the empirical distance covariance for dependent data. *Journal of Theoretical Probability*, 35:1–21.

Kutoyants, Y. A. (2010). On the goodness-of-fit testing for ergodic diffusion processes. *Journal of Nonparametric Statistics*, 22(4):529–543.

Langevin, P. (1908). On the theory of brownian motion. *Comptes Rendus de l'Académie des Sciences*, 146:530–533.

Laukaitis, A. and Rackauskas, A. (2002). Functional Data Analysis of payment systems. *Nonlinear Analysis: Modelling and Control*, 7(2):53–68.

Lavergne, P. and Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, 143(1):103–122.

Lee, S. (2006). The Bickel–Rosenblatt test for diffusion processes. *Statistics & Probability Letters*, 76(14):1494–1502.

Li, F. (2007). Testing the parametric specification of the diffusion function in a diffusion process. *Econometric Theory*, 23(2):221–250.

Li, Y., Liu, G., and Zhang, Z. (2022). Volatility of volatility: Estimation and tests based on noisy high frequency data with jumps. *Journal of Econometrics*, 229(2):422–451.

Lin, L.-C., Lee, S., and Guo, M. (2013). Goodness-of-fit test for stochastic volatility models. *Journal of Multivariate Analysis*, 116:473–498.

Lin, L.-C., Lee, S., and Guo, M. (2014). The Bickel–Rosenblatt test for continuous time stochastic volatility models. *Test*, 23(1):195–218.

Lin, L.-C., Lee, S., and Guo, M. (2016). Goodness-of-fit test for the SVM based on noisy observations. *Statistica Sinica*, 26(3):1305–1329.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, volume 793. John Wiley & Sons.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, pages 197–223. Springer.

Ljung, L. and Caines, P. E. (1980). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics*, 3(1-4):29–46.

Lopes, H. F. and Tsay, R. S. (2011). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168–209.

López-Pérez, A., Febrero-Bande, M., and González-Manteiga, W. (2021). Parametric estimation of diffusion processes: A review and comparative study. *Mathematics*, 9(8):859.

López-Pérez, A., Febrero-Bande, M., and González-Manteiga, W. (2022a). A comparative review of specification tests for diffusion models. Submitted manuscript, arXiv:2208.08420.

López-Pérez, A., Febrero-Bande, M., and González-Manteiga, W. (2022b). Estimation and specification test for diffusion models with stochastic volatility. Submitted manuscript, arXiv:2208.08415.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48.

Masuda, H., Negri, I., and Nishiyama, Y. (2011). Goodness-of-fit test for ergodic diffusions by discrete-time observations: An innovation martingale approach. *Journal of Nonparametric Statistics*, 23(2):237–254.

Matsui, M., Mikosch, T., and Samorodnitsky, G. (2017). Distance covariance for stochastic processes. *Probability and Mathematical Statistics*, 37(2):355–372.

Mattia, M. and Del Giudice, P. (2002). Population dynamics of interacting spiking neurons. *Physical Review E*, 66(5):051917.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2):239–265.

Merkle, M., Saporito, Y. F., and Targino, R. S. (2020). Bayesian approach for parameter estimation of continuous-time stochastic volatility models using fourier transform methods. *Statistics & Probability Letters*, 156:108600.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183.

Merton, R. C. (1975). An asymptotic theory of growth under uncertainty. *The Review of Economic Studies*, 42(3):375–393.

Miller, R. G. (1981). *Simultaneous Statistical Inference*. Springer-Verlag, New York.

Milshtein, G. N. (1979). A method of second-order accuracy integration of stochastic differential equations. *Theory of Probability & Its Applications*, 23(2):396–401.

Monsalve-Cobis, A., González-Manteiga, W., and Febrero-Bande, M. (2011). Goodness-of-fit test for interest rate models: An approach based on empirical processes. *Computational Statistics & Data Analysis*, 55(12):3073–3092.

Müller, H.-G., Sen, R., and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, 165(2):233–245.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Negri, I. and Nishiyama, Y. (2009). Goodness of fit test for ergodic diffusion processes. *Annals of the Institute of Statistical Mathematics*, 61(4):919–928.

Negri, I. and Nishiyama, Y. (2010). Goodness of fit test for ergodic diffusions by tick time sample scheme. *Statistical Inference for Stochastic Processes*, 13(1):81–95.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Statistica Sinica*, pages 113–135.

Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients. *Journal of Econometrics*, 13(3):341–363.

Pagan, A. R., Hall, A. D., and Martin, V. (1996). Modeling the term structure. *Handbook of Statistics*, 14:91–118.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Patilea, V., Sánchez-Sellero, C., and Saumard, M. (2016). Testing the predictor effect on a functional response. *Journal of the American Statistical Association*, 111(516):1684–1695.

Phillips, P. C. and Yu, J. (2005). Jackknifing bond option prices. *The Review of Financial Studies*, 18(2):707–742.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Podolskij, M. and Ziggel, D. (2008). A range-based test for the parametric form of the volatility in diffusion models. *CREATES Research Paper*, 22.

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 48(1):149–158.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *The Review of Financial Studies*, 11(3):449–487.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York.

Reboredo, J. C., Matías, J. M., and Garcia-Rubio, R. (2012). Nonlinearity in forecasting of high-frequency stock returns. *Computational Economics*, 40(3):245–264.

Rendleman, R. J. and Bartter, B. J. (1980). The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis*, 15(1):11–24.

Roislien, J. and Winje, B. (2013). Feature extraction across individual time series observations with spikes using wavelet principal component analysis. *Statistics in Medicine*, 32(21):3660–3669.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63(1):289–306.

Ruiz-Medina, M. D. and Álvarez-Liébana, J. (2019a). Classical and Bayesian componentwise predictors for non-compact correlated ARH(1) processes. *Revstat*, 17(3):265–296.

Ruiz-Medina, M. D. and Álvarez-Liébana, J. (2019b). A note on strong-consistency of componentwise ARH(1) predictors. *Statistics & Probability Letters*, 145:224–228.

Ruiz-Medina, M. D. and Álvarez-Liébana, J. (2019c). Strongly consistent autoregressive predictors in abstract Banach spaces. *Journal of Multivariate Analysis*, 170:186–201.

Ruiz-Medina, M. D. and Salmerón, R. (2010). Functional maximum likelihood estimation of ARH(p) models. *Stochastic Environmental Research and Risk Assessment*, 24:131–146.

Salmerón, R. and Ruiz-Medina, M. D. (2009). Multispectral decomposition of functional autoregressive models. *Stochastic Environmental Research and Risk Assessment*, 23:289–297.

Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87(2):271–301.

Sapp, T. R. (2009). Estimating continuous-time stochastic volatility models of the short-term interest rate: a comparison of the generalized method of moments and the Kalman filter. *Review of Quantitative Finance and Accounting*, 33(4):303–326.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons.

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica*, 14:1239–1257.

Shephard, N. (1993). Fitting nonlinear time-series models with applications to stochastic variance models. *Journal of Applied Econometrics*, 8(S1):S135–S152.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3):653–667.

Shoji, I. and Ozaki, T. (1997). Comparative study of estimation methods for continuous time stochastic processes. *Journal of Time Series Analysis*, 18(5):485–506.

Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4):733–752.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and its Applications*, volume 3. Springer.

Singer, H. (2002). Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *Journal of Computational & Graphical Statistics*, 11(4):972–995.

Song, Z. (2011). A martingale approach for testing diffusion models based on infinitesimal operator. *Journal of Econometrics*, 162(2):189–212.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72(3):337–354.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, 52(5):1973–2002.

Steele, J. M. (2001). *Stochastic Calculus and Financial Applications*, volume 1. Springer.

Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *The Review of Financial Studies*, 4(4):727–752.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, 86(416):1024–1033.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8(6):1348–1360.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25(2):613–641.

Stute, W., González-Manteiga, W., and Quindimil, M. P. (1998a). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441):141–149.

Stute, W., Quindimil, M. P., González-Manteiga, W., and Koul, H. (2006). Model checks of higher order time series. *Statistics & Probability Letters*, 76(13):1385–1396.

Stute, W., Thies, S., and Zhu, L.-X. (1998b). Model checks for regression: An innovation process approach. *The Annals of Statistics*, 26(5):1916–1934.

Stute, W., Xu, W., and Zhu, L. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika*, 95(2):451–467.

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1236–1265.

Székely, G. J. and Rizzo, M. L. (2012). On the uniqueness of distance covariance. *Statistics & Probability Letters*, 82(12):2278–2282.

Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1):65–81.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag, New York.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, 36(5):823.

Van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*, volume 1. Elsevier.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Vetter, M. (2015). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. *Bernoulli*, 21(4):2393–2418.

Vetzal, K. R. (1997). Stochastic volatility, movements in short term interest rates, and bond option values. *Journal of Banking & Finance*, 21(2):169–196.

Vieu, P. (2018). On dimension reduction models for functional data. *Statistics & Probability Letters*, 136:134–138.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372.

West, M. (1992). *Modelling with Mixtures*. Oxford University Press: Oxford.

Xia, Y. (2009). Model checking in regression via dimension reduction. *Biometrika*, 96(1):133–148.

Yao, F., Müller, H. G., and Wang, J. (2005). Functional Data Analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Yu, J. and Phillips, P. C. (2001). A gaussian approach for continuous time models of the short-term interest rate. *The Econometrics Journal*, 4(2):210–224.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2):263–289.

Zheng, X. (2009). Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canadian Journal of Statistics*, 37(2):282–300.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, 33(3):438–457.

Zu, Y. (2015). Nonparametric specification tests for stochastic volatility models based on volatility density. *Journal of Econometrics*, 187(1):323–344.

Due to their analytical tractability, continuous-time models have become a centerpiece in the financial literature. The goal of this thesis is the development of new goodness-of-fit test for continuous-time diffusion models, considering stochastic differential equations with deterministic and stochastic volatility and Itô diffusions as functional time series. Notwithstanding the importance of goodness-of-fit tools, latent factors and a continuous-time setting with observations occurring at discrete time points challenge the estimation of the models. Therefore, the estimation problem is addressed, as it hinders the goodness-of-fit procedures, discussing the intricacies of different estimation implementations prior to the methodological contribution of the test procedures.