

Review

Parametric Estimation of Diffusion Processes: A Review and Comparative Study

Alejandra López-Pérez , Manuel Febrero-Bande  and Wencesalo González-Manteiga 

Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Rúa de Lope Gómez de Marzoa s/n, 15705 Santiago de Compostela, Spain; manuel.febrero@usc.es (M.F.-B.); wenceslao.gonzalez@usc.es (W.G.-M.)

* Correspondence: alejandralopez.perez@usc.es

Abstract: This paper provides an in-depth review about parametric estimation methods for stationary stochastic differential equations (SDEs) driven by Wiener noise 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.

Keywords: diffusion processes; continuous-time models; parametric estimation; stochastic differential equations; SDE



Citation: López-Pérez, A.; Febrero-Bande, M.; González-Manteiga, W. Parametric Estimation of Diffusion Processes: A Review and Comparative Study. *Mathematics* **2021**, *9*, 859. <https://doi.org/10.3390/math9080859>

Academic Editor: Christos Floros

Received: 8 March 2021

Accepted: 12 April 2021

Published: 14 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Diffusion processes described by stochastic differential equations (SDE) 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 market turbulence 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 [1] used the interest rate in option pricing modeling it as a stochastic process. Subsequently, Black and Scholes [2] had an important impact on arbitrage models of the term structure of interest rates, as shown in [3–7]. 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 [8].

We consider the SDE defined in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, 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,

$$dX_t = m(X_t, \theta) dt + \sigma(X_t, \theta) dW_t, \quad \text{with } X_0 = x_0, \quad 0 \leq t \leq T, \quad (1)$$

with $X_t \in \mathbb{R}$, W_t a $\{\mathcal{F}_t\}_{t \geq 0}$ -adapted standard Wiener process and θ an unknown parameter vector such that $\theta \in \Theta \subset \mathbb{R}^d$ with d a positive integer and Θ a compact set. We assume the knowledge of the drift and diffusion functions parametric structure, $m(\cdot, \theta): \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ and $\sigma(\cdot, \theta): \mathbb{R} \times \Theta \rightarrow (0, \infty)$, respectively, where none is time dependent. Jump-diffusions and fractional Brownian motion or Lévy-driven SDEs are out of the scope of this review.

A parametric specification of (1) that encloses different interest rate models is the Chan–Karolyi–Longstaff–Sanders (CKLS) model proposed in [6], given by

$$dX_t = \kappa(\mu - X_t) dt + \sigma X_t^\gamma dW_t,$$

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 [3] and CIR [5] model, respectively—the process is tractable and admits an analytical solution.

In this article, 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. Furthermore, models based on the Vasicek process are still used in the financial market (see, e.g., [9] or [10]) and new estimation procedures are being proposed for jump-diffusions [11] or Lévy-driven processes [12].

Several studies of estimation methods for diffusion processes can be found in the literature, see [13] for a theoretical comparison or [14] for a more practical approach. In [15], a comparative study of different discretization methods and moment-based estimation is carried out, while [16] focused on simulation-based approaches. The aim of this paper 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 [17,18], Hermite polynomial expansion [19], Kalman filter [20], Markov Chain Monte Carlo [21,22] and generalized method of moments [6,23]. The procedures are provided in the companion `estsde` R package [24], implemented in C and C++ for the sake of efficiency.

The sections are organized as follows: Section 2 provides an outline of the estimation methods, Section 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 4 and conclusions are drawn in Section 5. Tabulated simulation results are deferred to Appendix A.

2. Estimation Methods

The unique strong solution X_t of the SDE in (1),

$$X_t = X_{t_0} + \int_0^t m(X_u, \theta) du + \int_0^t \sigma(X_u, \theta) dW_u,$$

exist under the assumptions that both drift $m(\cdot)$ and volatility $\sigma(\cdot)$ functions satisfy global Lipschitz continuous and growth conditions (see, e.g., [25]):

Assumption 1 (Global Lipschitz). *For all $x, y \in \mathbb{R}$ there exist a constant $C_1 < \infty$ independent of θ such that*

$$|m(x, \theta) - m(y, \theta)| + |\sigma(x, \theta) - \sigma(y, \theta)| \leq C_1|x - y|.$$

Assumption 2 (Linear growth). For all $x, y \in \mathbb{R}$ there exist a constant $C_2 < \infty$ independent of θ such that

$$|m(x, \theta)| + |\sigma(x, \theta)| \leq C_2(1 + |x|).$$

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 [26]: given an Itô process $\{X_t, 0 \leq t \leq T\}$, solution of the SDE in (1) with initial value $X_{t_0} = x_0$ and the discretization of the time interval $[0, T], 0 = t_0 < t_1 < \dots < t_n = T$, the Euler–Maruyama approximation of X is a continuous stochastic process that satisfies the iterative scheme

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

with $i = 0, 1, \dots, n - 1, t_i = i\Delta$ 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 θ . 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 3 (Bounded moments). For all $k > 0$, all order k moments of the diffusion process exist and are such that

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

The maximum likelihood (ML) estimates are yielded 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.

2.1. Exact Maximum Likelihood

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

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

where $p_\theta(\Delta, X_{t_i} | X_{t_{i-1}})$ denotes the transition density function associated to the parametric diffusion model, with unknown parameter θ . 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

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

where $\ell_n(\theta) = \log \mathcal{L}_n(\theta)$ 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., [2,3,5]. As a consequence, new procedures have been proposed in the ML framework based on different approximations of the transition density function.

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 like-

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., [18,27,28]), here we briefly detail two of them.

2.2.1. Euler Method

To estimate the model we can use an approximation scheme, such as the Euler–Maruyama [26] 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

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

where ε_{t_i} are i.i.d. $N(0, 1)$ and $\Delta = (t_{i+1} - t_i)$. Therefore, the estimation with the Euler–Maruyama method [29] 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

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

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 $\Delta \rightarrow 0$. Departures from Gaussian distribution can also increase bias since the Euler scheme increments are conditionally Gaussian.

2.2.2. 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

$$dX_t = m(X_t, \theta) dt + \sigma dW_t, \tag{3}$$

where $\sigma > 0 \in \mathbb{R}$ is assumed constant, the local linearization (LL) method developed in [17,18] is an approximation method by which the drift function $m(\cdot, \theta)$ 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 LL method is

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

where

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

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

The linear function K_t approximates the drift function $m(\cdot)$, with K_t constant in the interval $[i\Delta, (i + 1)\Delta)$. 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 $(X_{(i+1)\Delta} | X_{i\Delta})$ follows a normal distribution with mean and variance given by

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

respectively, and therefore, maximum likelihood can be used to obtain the parameter estimates. As the SDE in (3) 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 2.3). The LL 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).

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

$$\ell_n(\theta) \equiv \frac{1}{n} \sum_{i=1}^n \ln \left\{ p_{X,\theta}(\Delta, X_{i\Delta} \mid X_{(i-1)\Delta}) \right\}, \tag{4}$$

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

Ait-Sahalia [19] 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 $\{p_{X,\theta}^{(K)} \mid K \geq 0\}$ of the transition density, such that (4) is a succession of approximations of $\ell_n^{(K)}$.

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

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

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

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

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

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

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 ([19], Theorem 1) it is proved that

$$p_X^{(J)}(\Delta, x \mid x_0; \theta) \xrightarrow{J \rightarrow \infty} p_X(\Delta, x \mid x_0; \theta).$$

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 ($j = 0, \dots, 6$) are used, along with Taylor series up to order $K = 3$.

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

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

is maximized. Thus, we obtain an estimator $\hat{\theta}_n^{(J)}$ close to the exact $\hat{\theta}_n$ ([19], Theorem 2). For stationary processes, the estimator satisfies

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

where $i(\theta_0)$ is the Fisher information matrix.

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

2.4. Kalman Filter

The state-space model or dynamic linear model, introduced in [20], employs an order one vector autoregression as a state equation and assumes that we do not observe the state vector x_t directly, but a linear transformation of it with noise added, y_t . The *state-space* representation of the dynamics of y_t is given by the system of equations:

$$x_t = \Phi x_{t-1} + \Upsilon u_t + w_t, \quad w_t \sim \text{iid } N(0, \mathbf{Q}), \tag{5}$$

$$y_t = \mathbf{A}_t x_t + \Gamma u_t + v_t, \quad v_t \sim \text{iid } N(0, \mathbf{R}), \tag{6}$$

where the state vector x_t is $p \times 1$, the observed data vector y_t is $q \times 1$, the observation matrix \mathbf{A}_t is $q \times p$, 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 $\{w_t\}$ and $\{v_t\}$ are uncorrelated. Equation (5) is known as the *state equation* and Equation (6) as *observation equation*.

Let $x_{t|t-1} = \mathbb{E}\{x_t \mid y_{t-1}\}$ and $\mathbf{P}_{t|t-1} = \mathbb{E}\{(x_t - x_{t|t-1})(x_t - x_{t|t-1})'\}$, the Kalman filter equations, with initial state $x_0 \sim N(x_{0|0}, \mathbf{P}_{0|0})$, are given by

$$x_{t|t-1} = \Phi x_{t-1|t-1} + \Upsilon u_t, \tag{7}$$

$$\mathbf{P}_{t|t-1} = \Phi \mathbf{P}_{t-1|t-1} \Phi' + \mathbf{Q}, \tag{8}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{A}_t' [\mathbf{A}_t \mathbf{P}_{t|t-1} \mathbf{A}_t' + \mathbf{R}]^{-1}, \tag{9}$$

$$x_{t|t} = x_{t|t-1} + \mathbf{K}_t [y_t - \mathbf{A}_t x_{t|t-1} - \Gamma u_t], \tag{10}$$

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_t \mathbf{A}_t] \mathbf{P}_{t|t-1}. \tag{11}$$

The Kalman filter is a recursive algorithm, Equations (7) and (8) are the *time update* equations, where the state in time t is estimated with the information until time $t - 1$, and Equations (9)–(11) are the *measurement update* equations, where the new information of the estimation is incorporated and the mean squared error is minimized.

Let $\theta = (\Phi, \mathbf{Q}, \mathbf{R}, \Upsilon, \Gamma)'$ be the vector of parameters, we can use maximum likelihood under the assumption that the initial state is Gaussian and the errors $\{w_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ are uncorrelated Gaussian vectors. The Kalman filter can be set up to evaluate the likelihood function, which can be computed with the innovations $\{\varepsilon_i\}_{i=1}^n$, where $\varepsilon_t = y_t - \mathbf{A}_t x_{t|t-1} - \Gamma u_t$, and because they are independent Gaussian random vectors with zero mean and covariance matrix $\Sigma_t = \mathbf{A}_t \mathbf{P}_{t|t-1} \mathbf{A}_t' + \mathbf{R}$, we can write the likelihood, $\mathcal{L}_Y(\theta)$, as

$$\ln \mathcal{L}_Y(\theta) = \frac{1}{2} \sum_{t=1}^n \ln |\Sigma_t(\theta)| + \frac{1}{2} \sum_{t=1}^n \varepsilon_t(\theta)' \Sigma_t(\theta)^{-1} \varepsilon_t(\theta). \tag{12}$$

The log-likelihood in (12) can be maximized by numerical search procedures to obtain the estimation of θ , where the derivatives of (12) can be calculated numerically or analytically. The analytical derivatives can be obtained recursively by differentiating the Kalman

Filter recursion, see [30]. Algorithms like the EM [31] or the Newton–Raphson can be used to maximize the log-likelihood, as in [32,33], for an example of both approaches.

Under general conditions, let $\hat{\theta}_n$ be the estimator of the true parameters θ_0 obtained maximizing the innovation log-likelihood in (12). Subject to certain regularity conditions, when $n \rightarrow \infty$

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

where $i(\theta_0)$ is the asymptotic Fisher information matrix. The Kalman filter will generate strongly consistent estimates of θ_0 when the parameters lie in a compact set. More details on regularity conditions, convergence and asymptotic properties can be found in [30,34–37]. The Kalman filter can be used to estimate the parameters of the diffusion process (1) writing in state-space form the discrete version given in (2),

$$Y_{t_i} = m(X_{t_i}, \theta) + \frac{\sigma(X_{t_i}, \theta)}{\sqrt{\Delta}} \varepsilon_{t_i}, \tag{13}$$

with $Y_{t_i} = (X_{t_{i+1}} - X_{t_i})/\Delta$ and ε_{t_i} are i.i.d. $N(0, 1)$ random variables. The discretized model (13) is obtained by means of a Euler–Maruyama scheme, which makes the likelihood function in (12) equivalent to the one obtained using the Euler method (see Section 2.2.1), 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 (5) and (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 [38], which is close related to the local linearization method introduced in Section 2.2.2, see [39].

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 (2), we have

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

where $(W_{t_i} - W_{t_{i-1}})$ is a i.i.d. $N(0, \Delta)$. This discrete time approximation of the SDE can be too coarse to approximate the true transition density accurately (see [40] for the strong convergence criterion for SDE). Elerian et al. [21] and Eraker [22] 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 $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$ implies that $T(m - 1)$ data points are missing, such that $X_{t_i} = x_{t_i,0}^*, x_{t_i,1}^*, \dots, x_{t_i,m-1}^*, x_{t_i,m}^* = X_{t_{i+1}}$, as seen in Figure 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, $x_{t_i,k}^*, \dots, x_{t_i,k+M-1}^*$, 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

$$f(x_{t_i,k}^*, \dots, x_{t_i,k+M-1}^* \mid x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta) \propto \prod_{j=k-1}^{k+M-1} N[x_{t_i,j}^* + m(x_{t_i,j}^*, \theta)\Delta, \sigma^2(x_{t_i,j}^*, \theta)\Delta]. \tag{14}$$

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 (14), as suggested by [21]. The probability used to determine if the proposal value should be taken as the next item of the chain is given by

$$\alpha = \min \left\{ 1, \frac{f(w | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta) q(x_{t_i,k}^{*(p)}, \dots, x_{t_i,k+M-1}^{*(p)} | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta)}{f(x_{t_i,k}^{*(p)}, \dots, x_{t_i,k+M-1}^{*(p)} | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta) q(w | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta)} \right\},$$

where $w \sim q(x_{t_i,k}^*, \dots, x_{t_i,k+M-1}^* | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta)$ and $x_{t_i,k}^{*(p)}$ is the current value of $x_{t_i,k}^*$ at the end of the p th iteration. We then set $w = \{x_{t_i,j}^{*(p+1)}\}_{j=k}^{k+M-1}$ with probability α and $\{x_{t_i,j}^{*(p+1)}\}_{j=k}^{k+M-1} = \{x_{t_i,j}^{*(p)}\}_{j=k}^{k+M-1}$ with probability $(1 - \alpha)$.

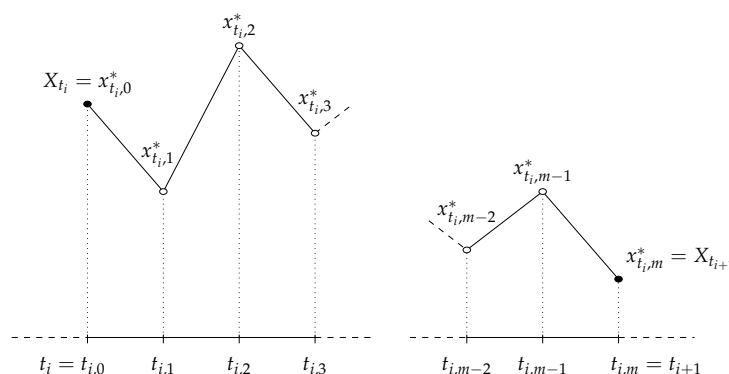


Figure 1. Augmented data in the discretization scheme: the observed data, X_{t_i} and $X_{t_{i+1}}$, is augmented by introducing $(m - 1)$ unobserved data points.

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 $\ln f(\cdot | x_{t_i,k-1}^*, x_{t_i,k+M}^*; \theta)$ and the covariance matrix is the negative of the inverse Hessian evaluated at the mode (see [21] for details regarding the gradient and Hessian matrix of the target density).

To complete one cycle of the MCMC sampler, we need to sample $\theta \sim \pi(\theta | X, X^*)$ 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

$$\mathcal{L} = \prod_{i=0}^{n-1} \left[\prod_{j=0}^M \frac{1}{\sqrt{2\pi m^2(x_{t_i,j}^*, \theta)\Delta}} \exp \left\{ -\frac{[x_{t_i,j+1}^* - x_{t_i,j}^* - m(x_{t_i,j}^*, \theta)\Delta]^2}{2\sigma^2(x_{t_i,j}^*, \theta)\Delta} \right\} \right]. \tag{15}$$

This method shows accuracy and can be extended to multi-dimensional models—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.

2.6. Generalized Method of Moments

The generalized method of moments (GMM), introduced by Hansen [23], is a special case of minimum distance estimation based in moment conditions. Let θ be the vector of parameters, we denote $f_t(\theta)$ as

$$f_t(\theta) = u_t \otimes z_t, \tag{16}$$

where u_t is an unobservable vector of disturbance terms, 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* $\mathbb{E}\{f_t(\theta)\} = 0$. Replacing the theoretical moment condition $\mathbb{E}\{f_t(\theta)\}$ with its sample counterpart,

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

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

$$Q_n(\theta) = g_n(\theta)' W_n(\theta) g_n(\theta), \tag{17}$$

where $W_n(\theta)$ is a positive semi-definite weighting matrix. Choosing $W_n(\theta) = S^{-1}(\theta)$, where $S(\theta) = \mathbb{E}\{f_t(\theta)f_t'(\theta)\}$, gives the GMM estimator of θ with the smallest asymptotic covariance matrix, see [23]. For a weighting matrix $W_n(\theta)$, the GMM estimator is

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

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

$$\mathbb{E}\{f_t(\theta)\} = 0 \quad \text{if, and only if } \theta = \theta_0,$$

to achieve the identification condition for consistency of the GMM estimator. Consistency results and conditions for asymptotic normality are given in ([41], Theorems 2.6 and 3.1). When $W(\theta) = S^{-1}(\theta)$, we have that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V),$$

where $V = (G'W(\theta)G)^{-1}$, with $G = \partial\mathbb{E}\{f_t(\theta_0)\} / \partial\theta$.

To implement the GMM for the diffusion model in (1), we can consider the discretized process

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

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

$$\mathbb{E}\{\varepsilon_{t_{i+1}} \mid \mathcal{F}_{t_i}\} = 0, \quad \mathbb{E}\{\varepsilon_{t_{i+1}}^2 \mid \mathcal{F}_{t_i}\} = \sigma(X_{t_i}, \theta)^2\Delta,$$

respectively. Due to the independence of increments property of the Wiener process, we can define (16) as the moments vector

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

therefore, we have the orthogonality condition $\mathbb{E}[f_t(\theta)] = 0$ to construct the GMM estimator of θ . 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 (17) can likely lead to optimization problems.

3. Simulation Study

In this section, a simulation study is conducted to compare the performance of the estimation methods described in this article. 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 $T \rightarrow \infty$ plays the role of the discrete-time standard asymptotic of $n \rightarrow \infty$, (iv) examine discretization bias and the impact of different sampling frequencies Δ , and (v) study the effect of volatility on the estimators performance.

3.1. Experimental Design

We consider two models, one proposed by Vasicek [3] based in the Ornstein–Uhlenbeck [42] process and the CKLS proposed by [6]. 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

$$dX_t = \kappa(\mu - X_t) dt + \sigma dW_t,$$

and the CKLS is

$$dX_t = \kappa(\mu - X_t) dt + \sigma X_t^\gamma dW_t, \quad (18)$$

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 $n = 520, 2600$, with $\Delta = 1/52$, which corresponds to weekly data on an observation window of $T = 10, 50$ 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 1 shows the eight simulated scenarios for the Vasicek model, along with unconditional 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 $\theta_1 = (\mu, \kappa, \sigma^2)' = (0.09, 0.2, 4 \times 10^{-5})'$ and $\theta_2 = (0.09, 0.2, 4 \times 10^{-4})'$, and for the low persistence (scenarios 3 and 4) the parameter are $\theta_3 = (0.09, 0.9, 1.8 \times 10^{-4})'$ and $\theta_4 = (0.09, 0.9, 1.8 \times 10^{-3})'$. 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 A1. The settings of scenarios 1 and 2 allow us to check the performance in a near unit-root case, as the

regressive coefficient of $\mathbb{E}\{X_{t+1} | X_t\}$ 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.

Table 1. Scenarios for the Monte Carlo study for the Vasicek, $dX_t = \kappa(\mu - X_t) dt + \sigma dW_t$, and CKLS, $dX_t = \kappa(\mu - X_t) dt + \sigma X_t^\gamma dW_t$, models with $\Delta = 1/52$.

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
$\text{Var}\{X_t\}$	10^{-4}	10^{-3}	10^{-4}	10^{-3}
$\mathbb{E}\{X_{(i+1)\Delta} X_{i\Delta}\}$	$3.4 \times 10^{-4} + 0.996X_t$	$3.4 \times 10^{-4} + 0.996X_t$	$0.002 + 0.98X_t$	$0.002 + 0.98X_t$
$\text{Var}\{X_{(i+1)\Delta} X_{i\Delta}\}$	7.6×10^{-7}	7.6×10^{-6}	3.4×10^{-6}	3.4×10^{-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
$\text{Var}\{X_t\}$	10^{-4}	10^{-3}	10^{-4}	10^{-3}
$\mathbb{E}\{X_{(i+1)\Delta} X_{i\Delta}\}$	$8.3 \times 10^{-3} + 0.908X_t$	$8.3 \times 10^{-3} + 0.908X_t$	$0.015 + 0.83X_t$	$0.015 + 0.83X_t$
$\text{Var}\{X_{(i+1)\Delta} X_{i\Delta}\}$	1.7×10^{-5}	1.7×10^{-4}	3.1×10^{-5}	3.1×10^{-4}
CKLS	Scenario 1	Scenario 2	Scenario 3	Scenario 4
$(\mu, \kappa, \sigma^2, \gamma)$	$(0.09, 0.2, 0.25, 1.5)$	$(0.09, 0.2, 1, 1.5)$	$(0.09, 0.9, 0.5, 1.5)$	$(0.09, 0.9, 2, 1.5)$

In regards to the scenarios for the CKLS model, a similar scheme was set, as shown in Table 1, keeping for all scenarios the same parameters for the drift function as the Vasicek. First two scenarios correspond to high persistent dependence with parameter values $\theta_1 = (\mu, \kappa, \sigma^2, \gamma)' = (0.09, 0.2, 0.25, 1.5)'$ and $\theta_2 = (0.09, 0.2, 1, 1.5)'$, and for the low persistence (scenarios 3 and 4) the parameter are $\theta_3 = (0.09, 0.9, 0.5, 1.5)'$ and $\theta_4 = (0.09, 0.9, 2, 1.5)'$. Figure A1 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 2 and 3 show the simulation study for the Vasicek model (along with Tables A1–A4, deferred to Appendix A), Table 4 shows the computational performance of the methods, while Tables 5 and 6 (and Tables A5 and A6) refer to the CKLS model. The mean of the parameter estimates for the 1000 replications of the experiment is included, as well as standard deviation (SD) and root mean squared error (RMSE) for the estimation methods in Section 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).

Table 2. Monte Carlo simulation for Vasicek model, $(\mu, \kappa, \sigma) = (0.09, 0.2, 0.00632)'$, scenario 1: low mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

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

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 1000 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, 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 2500 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 5000 iterations were executed and the first 1000 were discarded. For both models, $m = 5$ was fixed in the data augmentation step. For all methods, θ was jointly estimated. The drift parameter μ is calculated indirectly, as the drift specification was rewritten to estimate the intercept $\alpha = \kappa\mu$.

Regarding optimization, we chose to minimize the negative log-likelihood function and use the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm [24], 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 (13), 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.

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

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

Since some of the procedures are computationally expensive, we chose to integrate C and C++ code [43] in the R routines. Table 4 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.

Table 4. CPU time (in seconds) of the estimation methods per iteration, with $n = 2600$.

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++

3.3. Discussion

Tables 2 and 5 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 [44]. 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 2). 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.

Table 5. Monte Carlo simulation for CKLS model, $(\mu, \kappa, \sigma, \gamma) = (0.09, 0.2, 0.5, 1.5)'$, scenario 1: low mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

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

When volatility is increased while keeping the same drift function (Tables A1 and A5, in Appendix A), bias is moderately increased for κ and standard deviation 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 and 6) and 4 (Tables A2 and A6) 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 2) than the high persistence dependence scenarios for both models and estimation bias and standard deviation 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.

Table 6. Monte Carlo simulation for CKLS model, $(\mu, \kappa, \sigma, \gamma)' = (0.09, 0.9, 0.707, 1.5)'$, scenario 3: high mean reversion and low volatility. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

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

Scenario 6 (Table A3) 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 (note that scenario 5 is not included, as the results were very close and so too were the conclusions drawn). This is magnified in the limiting case of scenario 8 (Table A4), 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 (the performance of the methods was analogous in scenario 7, therefore results are not reported). Figure 3 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 mean reversion scenarios, while scenarios 6 and 8 display discretization bias.

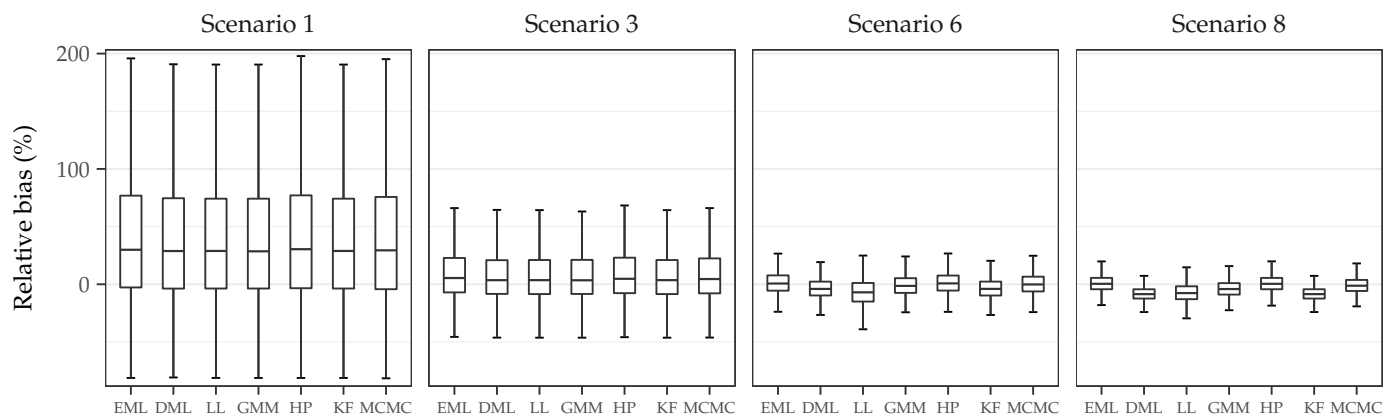


Figure 2. Relative bias (in percentage) for the drift parameter κ with weekly data ($\Delta = 1/52$) and $n = 2600$.

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 $\kappa \rightarrow 0$, mean reversion goes to zero and correlation between observations approaches one. This increases persistence, which introduces sample bias in parametric estimation [45]. Simulations show that increasing κ lowered persistence and, therefore, estimation bias was also diminished (see Figure 2). 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 $n \rightarrow \infty$. 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} [46]. 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).
- (iv) Discretization bias arises in DML and KF methods in scenarios with low sampling frequency and low persistence (see Figure 3), 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.

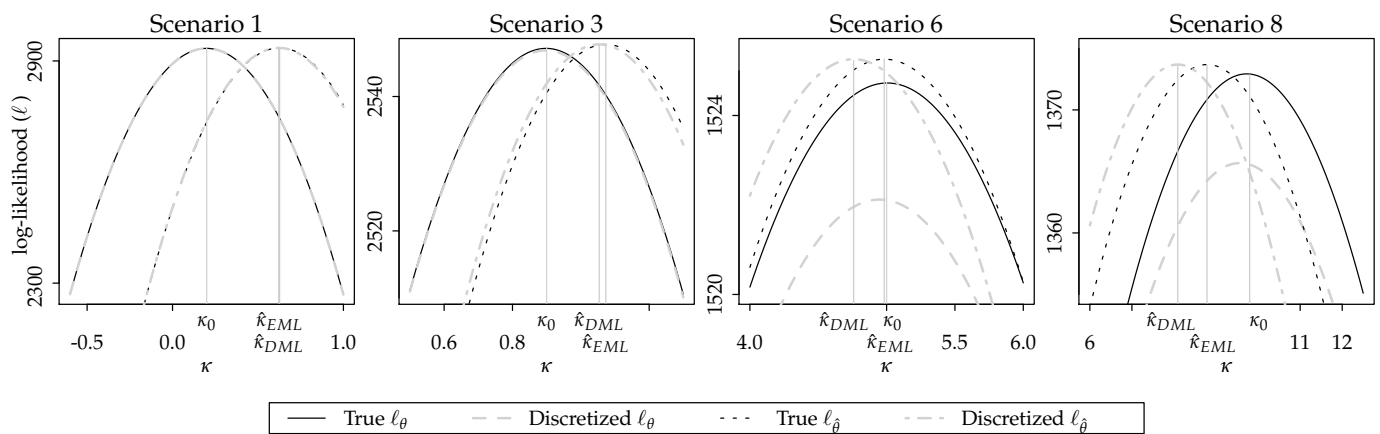


Figure 3. Vasicek model log-likelihood (ℓ) for the drift parameter κ with known (ℓ_θ) and estimated ($\ell_{\hat{\theta}}$) parameters, where κ_0 is the true value. The true density and its discretized version ($\Delta = 1/52$) is illustrated, along with the estimate obtained ($\hat{\kappa}_{EML}$ and $\hat{\kappa}_{DML}$, respectively). Note that scenarios with same κ_0 but different volatility parameter (scenarios 2, 4, 5 and 7, respectively) are not included, as the estimates were very close and figures very similar to the ones displayed.

Table 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 2). The generalized method of moments was outperformed in the majority of scenarios, notably with higher dimensions of the parameter vector θ .

Table 7. Summary of estimation procedures for SDE parameters.

Method	Authors	Asymptotic Properties	Finite Sample Performance
DML	Florens–Zmirou [29]	Asymptotically normal and consistent	Biased when Δ is large
LL	Ozaki [17] Shoji and Ozaki [18]	As ML estimators	Outperforms discrete ML and KF
HP	Ait–Sahalia [19,47]	Asymptotically normal and consistent [19]	Outperforms LL
KF	Kalman [20,48]	Asymptotically normal and consistent [30,34]	Similar to discrete ML
MCMC	Elerian et al. [21] Eraker [22]	Simulation based	Efficient but the most computationally intensive
GMM	Hansen [23] Chan et al. [6]	Asymptotically normal and consistent [23,41]	The least efficient and efficiency depends on moments

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 15th October 2001 to 30th December 2005 (sample size of $n = 1077$), see Figure 4. Numerous models have been proposed to capture the dynamics of short-term interest rate, including those by [1,3–5,49]. As these models can be nested within the unrestricted model

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

proposed by [6], we will estimate the parameters with the methods presented in the previous section. We chose a different drift parametrization of the CKLS model in (18) to provide standard errors for all estimates, so none of them are estimated indirectly.

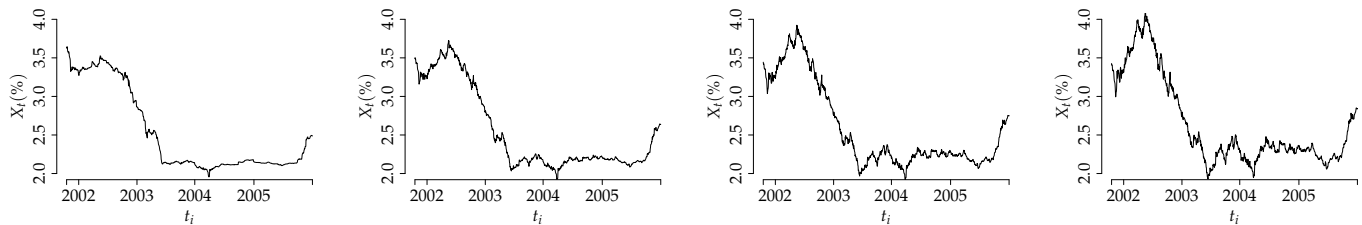


Figure 4. Euribor series. Daily evolution for the time period between 15th October 2001 and 30th December 2005. Sample size for each data set is $n = 1077$. From left to right, Euribor 3, 6, 9 and 12 months, respectively.

The results of the parameter estimation for the CKLS model are shown in Table 8, with the associated standard error in parentheses. The estimates obtained by the different approaches are considerable close, where the generalized method of moments seem to present the most noticeable disparity, mainly in the level effect parameter θ_4 . This was already noticed in [50], where different values of θ_4 were 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 5 and 6 and Tables A5 and A6). 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 [49].

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 9 shows the p -values for the goodness-of-fit test suggested by Monsalve-Cobis et al. [51]. 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 maturity.

To further analyze the fitted CKLS model, we will use a resampling procedure in the context of state space models (see Section 2.4) as it can provide insight into the validity of the model. The bootstrap technique developed in [52] 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 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 10 shows the standard errors obtained from $B = 1000$ bootstrap resamples, along with the asymptotic standard errors and parameter estimates, and Figure 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 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 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.

Table 8. Estimated parameters and standard errors (in parentheses) for the CKLS model, $dX_t = (\theta_1 - \theta_2 X_t) dt + \theta_3 X_t^{\theta_4} dW_t$, fitted to Euribor series using six estimation methods.

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 9. p -values for the Monsalve-Cobis et al. [51] goodness-of-fit test for the CKLS parametric form of the drift and diffusion functions.

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

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

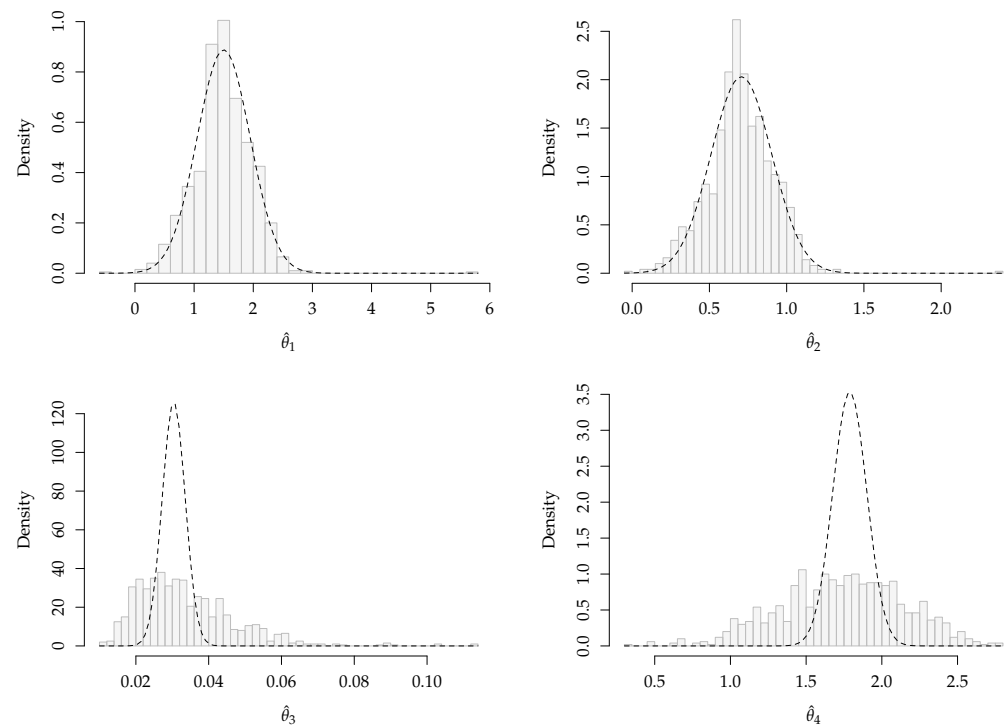


Figure 5. Bootstrap histogram of $\hat{\theta}_i$, with $i \in \{1, 2, 3, 4\}$, and asymptotic Gaussian distribution for the CKLS model, $dX_t = (\theta_1 - \theta_2 X_t) dt + \theta_3 X_t^{\theta_4} dW_t$, fitted to Euribor 3 months series.

5. Conclusions

We reviewed parametric estimation methods for univariate time-homogeneous SDEs. 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 $T = n\Delta \rightarrow \infty$, 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. All the procedures reviewed in this article are provided in the companion `estsde` R package.

Author Contributions: Conceptualization, M.F.-B., W.G.-M. and A.L.-P.; methodology, M.F.-B., W.G.-M. and A.L.-P.; software, A.L.-P.; formal analysis, A.L.-P.; investigation, A.L.-P.; writing—original draft preparation, A.L.-P.; writing—review and editing, M.F.-B., W.G.-M. and A.L.-P.; funding acquisition, M.F.-B. and W.G.-M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge support from grant MTM2016-76969-P from the Spanish Ministry of Economy and Competitiveness (cofunded with FEDER funds) and gratefully thank Spanish National Research Council for providing the computing resources of the Supercomputing Center of Galicia (CESGA).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank two anonymous referees for their insightful suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The appendix contains the tabulated results of the simulation study for the Vasicek and CKLS models. The Vasicek scenarios included here are 2, 4, 6 and 8 (see Table 1), which all of them have the same marginal distribution (see Figure A1), and scenarios 2 and 4 of the CKLS model.

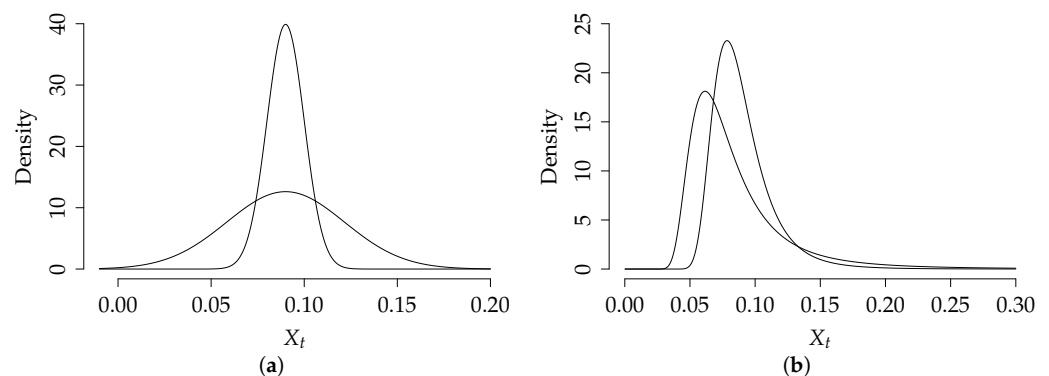


Figure A1. (a) Marginal density for Vasicek model, scenario 1 (3, 5 and 7) and 2 (4, 6 and 8) and (b) Marginal density for CKLS model, scenario 1 and 2.

Table A1. Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)' = (0.09, 0.2, 0.02)'$. Boldfaces denote the best results in terms of bias, standard deviation and RMSE.

Scenario 2		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
EML	$\hat{\mu}$	0.0938	0.0363	0.0365	0.0900	0.0145	0.0145	0.0906	0.0158	0.0158
	$\hat{\kappa}$	0.6900	0.4817	0.6871	0.2862	0.1283	0.1546	0.2972	0.1388	0.1695
	$\hat{\sigma}$	0.0200	6.27×10^{-4}	6.28×10^{-4}	0.0200	2.63×10^{-4}	2.63×10^{-4}	0.0200	6.28×10^{-4}	6.29×10^{-4}
DML	$\hat{\mu}$	0.0938	0.0366	0.0368	0.0899	0.0146	0.0146	0.0906	0.0158	0.0158
	$\hat{\kappa}$	0.6812	0.4715	0.6737	0.2838	0.1262	0.1515	0.2925	0.1345	0.1632
	$\hat{\sigma}$	0.0199	6.22×10^{-4}	6.29×10^{-4}	0.0200	2.62×10^{-4}	2.65×10^{-4}	0.0198	6.17×10^{-4}	6.51×10^{-4}
LL	$\hat{\mu}$	0.0943	0.0412	0.0415	0.0897	0.0206	0.0206	0.0892	0.0235	0.0235
	$\hat{\kappa}$	0.6781	0.4758	0.6745	0.2796	0.1302	0.1526	0.2856	0.1436	0.1672
	$\hat{\sigma}$	0.0199	6.18×10^{-4}	6.23×10^{-4}	0.0200	2.62×10^{-4}	2.64×10^{-4}	0.0198	6.21×10^{-4}	6.56×10^{-4}
HP	$\hat{\mu}$	0.0938	0.0363	0.0365	0.0900	0.0145	0.0145	0.0906	0.0158	0.0158
	$\hat{\kappa}$	0.6875	0.4784	0.6830	0.2858	0.1277	0.1538	0.2972	0.139	0.1696
	$\hat{\sigma}$	0.0200	6.27×10^{-4}	6.28×10^{-4}	0.0200	2.63×10^{-4}	2.63×10^{-4}	0.0200	6.28×10^{-4}	6.29×10^{-4}
KF	$\hat{\mu}$	0.0939	0.0370	0.0372	0.0900	0.0146	0.0146	0.0906	0.0158	0.0158
	$\hat{\kappa}$	0.6804	0.4719	0.6733	0.2836	0.1261	0.1513	0.2924	0.1346	0.1633
	$\hat{\sigma}$	0.0199	6.18×10^{-4}	6.21×10^{-4}	0.0200	2.62×10^{-4}	2.64×10^{-4}	0.0198	6.15×10^{-4}	6.38×10^{-4}
MCMC	$\hat{\mu}$	0.0941	0.0424	0.0426	0.0899	0.0146	0.0146	0.0906	0.0158	0.0158
	$\hat{\kappa}$	0.6899	0.482	0.6873	0.2848	0.1271	0.1528	0.2977	0.1389	0.1699
	$\hat{\sigma}$	0.0201	6.28×10^{-4}	6.33×10^{-4}	0.0200	2.64×10^{-4}	2.65×10^{-4}	0.0201	6.24×10^{-4}	6.26×10^{-4}
GMM	$\hat{\mu}$	0.0960	0.0578	0.0581	0.0900	0.0146	0.0146	0.0906	0.0172	0.0172
	$\hat{\kappa}$	0.7161	0.5101	0.7257	0.2856	0.1275	0.1536	0.3081	0.1462	0.1818
	$\hat{\sigma}$	0.0199	7.14×10^{-4}	7.22×10^{-4}	0.0200	2.67×10^{-4}	2.68×10^{-4}	0.0199	7.09×10^{-4}	7.20×10^{-4}

Table A2. Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)' = (0.09, 0.9, 0.0424)'$.

Scenario 4		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
EML	$\hat{\mu}$	0.0907	0.0155	0.0155	0.0900	0.0068	0.0068	0.0903	0.0074	0.0074
	$\hat{\kappa}$	1.3214	0.6079	0.7397	0.9757	0.2141	0.2271	0.9877	0.2383	0.2539
	$\hat{\sigma}$	0.0425	1.33×10^{-3}	1.34×10^{-3}	0.0425	5.59×10^{-4}	5.61×10^{-4}	0.0425	1.37×10^{-3}	1.37×10^{-3}
DML	$\hat{\mu}$	0.0906	0.0155	0.0155	0.0900	0.0068	0.0068	0.0903	0.0074	0.0074
	$\hat{\kappa}$	1.2997	0.5882	0.7112	0.9631	0.2085	0.2178	0.9465	0.2178	0.2227
	$\hat{\sigma}$	0.0420	1.31×10^{-3}	1.39×10^{-3}	0.0421	5.52×10^{-4}	6.57×10^{-4}	0.0408	1.27×10^{-3}	2.06×10^{-3}
LL	$\hat{\mu}$	0.0888	0.0190	0.0190	0.0876	0.0106	0.0109	0.0884	0.0107	0.0109
	$\hat{\kappa}$	1.2717	0.6329	0.7340	0.8910	0.2752	0.2754	0.9157	0.2766	0.2770
	$\hat{\sigma}$	0.0419	1.31×10^{-3}	1.40×10^{-3}	0.0420	5.56×10^{-4}	6.84×10^{-4}	0.0406	1.28×10^{-3}	2.21×10^{-3}
HP	$\hat{\mu}$	0.0907	0.0155	0.0155	0.0900	0.0068	0.0068	0.0903	0.0074	0.0074
	$\hat{\kappa}$	1.3207	0.6077	0.7391	0.9739	0.2131	0.2256	0.9886	0.2382	0.2542
	$\hat{\sigma}$	0.0425	1.33×10^{-3}	1.33×10^{-3}	0.0425	5.60×10^{-4}	5.61×10^{-4}	0.0425	1.37×10^{-3}	1.37×10^{-3}
KF	$\hat{\mu}$	0.0907	0.0155	0.0155	0.0900	0.0068	0.0068	0.0903	0.0074	0.0074
	$\hat{\kappa}$	1.2992	0.5886	0.7112	0.9635	0.2085	0.2180	0.9465	0.2178	0.2227
	$\hat{\sigma}$	0.0420	1.31×10^{-3}	1.39×10^{-3}	0.0421	5.52×10^{-4}	6.49×10^{-4}	0.0408	1.27×10^{-3}	2.06×10^{-3}

Table A2. Cont.

Scenario 4		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
MCMC	$\hat{\mu}$	0.0906	0.0154	0.0155	0.0900	0.0068	0.0068	0.0903	0.0074	0.0074
	$\hat{\kappa}$	1.3239	0.6069	0.7403	0.9722	0.2120	0.2240	0.9867	0.2358	0.2513
	$\hat{\sigma}$	0.0425	1.32×10^{-3}	1.33×10^{-3}	0.0424	5.57×10^{-4}	5.57×10^{-4}	0.0423	1.34×10^{-3}	1.35×10^{-3}
GMM	$\hat{\mu}$	0.0907	0.0179	0.0179	0.0900	0.0069	0.0069	0.0904	0.0090	0.0091
	$\hat{\kappa}$	1.4695	0.7586	0.9486	0.9716	0.2144	0.2260	1.0457	0.3062	0.3390
	$\hat{\sigma}$	0.0423	1.64×10^{-3}	1.65×10^{-3}	0.0423	5.72×10^{-4}	5.97×10^{-4}	0.0424	1.77×10^{-3}	1.77×10^{-3}

Table A3. Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)' = (0.09, 5, 0.1)'$.

Scenario 6		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
EML	$\hat{\mu}$	0.0902	0.0065	0.0065	0.0900	0.0029	0.0029	0.0901	0.0031	0.0031
	$\hat{\kappa}$	5.3788	1.1529	1.2136	5.0732	0.4992	0.5045	5.0939	0.6115	0.6186
	$\hat{\sigma}$	0.1002	0.0033	0.0033	0.1001	0.0014	0.0014	0.1001	0.0038	0.0038
DML	$\hat{\mu}$	0.0902	0.0065	0.0065	0.0900	0.0029	0.0029	0.0901	0.0031	0.0031
	$\hat{\kappa}$	5.0997	1.0319	1.0367	4.8317	0.4518	0.4822	4.1421	0.3959	0.9448
	$\hat{\sigma}$	0.0952	0.0030	0.0056	0.0954	0.0013	0.0048	0.0822	0.0026	0.0180
LL	$\hat{\mu}$	0.0892	0.0074	0.0074	0.0889	0.0034	0.0036	0.0891	0.0033	0.0035
	$\hat{\kappa}$	5.0206	1.2585	1.2587	4.6468	0.6269	0.7195	4.6304	0.6236	0.7249
	$\hat{\sigma}$	0.0946	0.0030	0.0061	0.0948	0.0013	0.0053	0.0810	0.0025	0.0192
HP	$\hat{\mu}$	0.0902	0.0065	0.0065	0.0900	0.0029	0.0029	0.0899	0.0033	0.0033
	$\hat{\kappa}$	5.3814	1.1493	1.2109	5.0743	0.4977	0.5032	5.1696	0.7052	0.7253
	$\hat{\sigma}$	0.1002	0.0033	0.0033	0.1001	0.0014	0.0014	0.0996	0.0041	0.0041
KF	$\hat{\mu}$	0.0902	0.0065	0.0065	0.0900	0.0029	0.0029	0.0901	0.0031	0.0031
	$\hat{\kappa}$	5.1001	1.0318	1.0367	4.8322	0.4516	0.4818	4.1418	0.3957	0.9450
	$\hat{\sigma}$	0.0952	0.0030	0.0056	0.0954	0.0013	0.0048	0.0822	0.0026	0.0180
MCMC	$\hat{\mu}$	0.0902	0.0065	0.0065	0.0900	0.0029	0.0029	0.0901	0.0031	0.0031
	$\hat{\kappa}$	5.3614	1.1402	1.1961	5.0311	0.4889	0.4899	4.9286	0.5693	0.5737
	$\hat{\sigma}$	0.0995	0.0032	0.0032	0.0992	0.0014	0.0016	0.0965	0.0034	0.0049
GMM	$\hat{\mu}$	0.0904	0.0084	0.0084	0.0900	0.0029	0.0029	0.0901	0.0041	0.0041
	$\hat{\kappa}$	5.6980	1.5548	1.7043	4.9583	0.4878	0.4895	5.0970	0.7941	0.8000
	$\hat{\sigma}$	0.1000	0.0043	0.0043	0.1000	0.0014	0.0014	0.0997	0.0047	0.0047

Table A4. Monte Carlo simulation for Vasicek model with $(\mu, \kappa, \sigma)' = (0.09, 9.8, 0.14)'$.

Scenario 8		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
EML	$\hat{\mu}$	0.0902	0.0046	0.0046	0.0900	0.0021	0.0021	0.0901	0.0023	0.0023
	$\hat{\kappa}$	10.1822	1.6165	1.6610	9.8776	0.7359	0.7400	9.9341	1.0861	1.0943
	$\hat{\sigma}$	0.1402	0.0048	0.0048	0.1401	0.0021	0.0021	0.1403	0.0064	0.0064
DML	$\hat{\mu}$	0.0902	0.0046	0.0046	0.0900	0.0021	0.0021	0.0901	0.0023	0.0023
	$\hat{\kappa}$	9.2274	1.3166	1.4357	8.9929	0.6049	1.0086	6.7332	0.4699	3.1026
	$\hat{\sigma}$	0.1275	0.0040	0.0131	0.1278	0.0017	0.0123	0.0981	0.0030	0.0421
LL	$\hat{\mu}$	0.0892	0.0050	0.0051	0.0890	0.0023	0.0025	0.0890	0.0024	0.0026
	$\hat{\kappa}$	9.4201	1.7549	1.7955	9.0671	0.8327	1.1093	8.7862	1.0441	1.4553
	$\hat{\sigma}$	0.1263	0.0039	0.0143	0.1265	0.0017	0.0136	0.0966	0.0031	0.0435
HP	$\hat{\mu}$	0.0901	0.0047	0.0047	0.0900	0.0021	0.0021	0.0898	0.0027	0.0027
	$\hat{\kappa}$	10.1842	1.6113	1.6564	9.8722	0.7334	0.7369	9.5056	1.2285	1.2632
	$\hat{\sigma}$	0.1401	0.0049	0.0049	0.1401	0.0021	0.0021	0.1339	0.0075	0.0097
KF	$\hat{\mu}$	0.0902	0.0046	0.0046	0.0900	0.0021	0.0021	0.0901	0.0023	0.0023
	$\hat{\kappa}$	9.2281	1.3174	1.4362	8.9941	0.6060	1.0083	6.7348	0.4690	3.1009
	$\hat{\sigma}$	0.1275	0.0040	0.0131	0.1278	0.0017	0.0123	0.0981	0.0030	0.0421
MCMC	$\hat{\mu}$	0.0902	0.0046	0.0046	0.0900	0.0021	0.0021	0.0901	0.0023	0.0023
	$\hat{\kappa}$	10.0552	1.5661	1.5867	9.7084	0.7076	0.7135	9.2813	0.9545	1.0864
	$\hat{\sigma}$	0.1381	0.0046	0.0050	0.1376	0.0020	0.0031	0.1303	0.0053	0.0110
GMM	$\hat{\mu}$	0.0903	0.0064	0.0064	0.0900	0.0021	0.0021	0.0901	0.0030	0.0030
	$\hat{\kappa}$	10.4527	2.1184	2.2167	9.4325	0.7327	0.8197	9.6680	1.4521	1.4581
	$\hat{\sigma}$	0.1399	0.0062	0.0062	0.1395	0.0021	0.0021	0.1384	0.0079	0.0081

Table A5. Monte Carlo simulation for CKLS model with $(\mu, \kappa, \sigma, \gamma)' = (0.09, 0.2, 1, 1.5)'$.

Scenario 2		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
DML	$\hat{\mu}$	0.0984	0.0933	0.0937	0.1002	0.0687	0.0695	0.1086	0.0974	0.0991
	$\hat{\kappa}$	0.6577	0.4887	0.6696	0.2807	0.1567	0.1762	0.2860	0.1679	0.1887
	$\hat{\sigma}$	1.0244	0.3720	0.3728	0.9944	0.1044	0.1045	0.9690	0.2392	0.2412
	$\hat{\gamma}$	1.4873	0.1372	0.1378	1.4956	0.0407	0.0409	1.4771	0.0958	0.0985
LL	$\hat{\mu}$	0.0977	0.0926	0.0929	0.0994	0.0605	0.0612	0.1102	0.1213	0.1229
	$\hat{\kappa}$	0.6650	0.4984	0.6816	0.2819	0.1580	0.1780	0.2916	0.1734	0.1961
	$\hat{\sigma}$	1.0638	0.3859	0.3911	1.0074	0.1034	0.1037	1.0247	0.2361	0.2374
	$\hat{\gamma}$	1.5038	0.1350	0.1351	1.5019	0.0395	0.0396	1.5057	0.0871	0.0873
HP	$\hat{\mu}$	0.0988	0.1091	0.1095	0.0997	0.0616	0.0623	0.1133	0.1425	0.1444
	$\hat{\kappa}$	0.6608	0.4932	0.6750	0.2812	0.1575	0.1772	0.2890	0.1724	0.1941
	$\hat{\sigma}$	1.0706	0.4013	0.4075	1.0061	0.1040	0.1042	1.0223	0.2437	0.2447
	$\hat{\gamma}$	1.5032	0.1327	0.1328	1.4999	0.0399	0.0399	1.4977	0.0899	0.0899
KF	$\hat{\mu}$	0.0980	0.0904	0.0908	0.1003	0.0692	0.0699	0.1088	0.0992	0.1010
	$\hat{\kappa}$	0.6579	0.4888	0.6698	0.2805	0.1566	0.1761	0.2861	0.1680	0.1888
	$\hat{\sigma}$	1.0241	0.3719	0.3726	0.9943	0.1042	0.1044	0.9688	0.2394	0.2414
	$\hat{\gamma}$	1.4872	0.1372	0.1378	1.4956	0.0406	0.0408	1.4770	0.0959	0.0986

Table A5. Cont.

Scenario 2		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
MCMC	$\hat{\mu}$	0.0988	0.1054	0.1057	0.1003	0.0689	0.0697	0.1109	0.1198	0.1216
	$\hat{\kappa}$	0.6651	0.4972	0.6808	0.2813	0.1576	0.1773	0.2888	0.1721	0.1936
	$\hat{\sigma}$	1.1357	0.5257	0.5429	1.0122	0.1140	0.1146	1.0606	0.2895	0.2957
	$\hat{\gamma}$	1.4998	0.1595	0.1595	1.5002	0.0430	0.0430	1.4999	0.0999	0.0999
GMM	$\hat{\mu}$	0.1003	0.1872	0.1875	0.0882	0.0218	0.0219	0.0890	0.0232	0.0233
	$\hat{\kappa}$	0.8836	0.4953	0.8442	0.4766	0.1604	0.3198	0.4813	0.1579	0.3226
	$\hat{\sigma}$	0.9625	0.4548	0.4563	0.9631	0.1923	0.1958	0.8381	0.3223	0.3606
	$\hat{\gamma}$	1.4440	0.1787	0.1873	1.4690	0.0889	0.0942	1.3820	0.1609	0.1995

Table A6. Monte Carlo simulation for CKLS model with $(\mu, \kappa, \sigma, \gamma)' = (0.09, 0.9, 1.414, 1.5)'$.

Scenario 4		$\Delta = 1/52, n = 520$			$\Delta = 1/52, n = 2600$			$\Delta = 1/12, n = 520$		
Method	$\hat{\theta}$	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
DML	$\hat{\mu}$	0.0965	0.0625	0.0628	0.0907	0.0078	0.0079	0.0909	0.0080	0.0081
	$\hat{\kappa}$	1.2752	0.6674	0.7657	0.9594	0.2535	0.2604	0.9493	0.2692	0.2737
	$\hat{\sigma}$	1.3874	0.4298	0.4307	1.3765	0.1661	0.1703	1.2486	0.3849	0.4190
	$\hat{\gamma}$	1.4762	0.1241	0.1263	1.4876	0.0488	0.0503	1.4382	0.1245	0.1390
LL	$\hat{\mu}$	0.0953	0.0421	0.0424	0.0904	0.0078	0.0078	0.0902	0.0078	0.0078
	$\hat{\kappa}$	1.2993	0.6905	0.7976	0.9759	0.2601	0.2709	1.0009	0.2939	0.3107
	$\hat{\sigma}$	1.4853	0.4496	0.4552	1.4397	0.1668	0.1688	1.4915	0.3856	0.3933
	$\hat{\gamma}$	1.5076	0.1184	0.1187	1.5085	0.0468	0.0476	1.5265	0.1048	0.1081
HP	$\hat{\mu}$	0.0955	0.0441	0.0444	0.0912	0.0150	0.0150	0.0925	0.0452	0.0453
	$\hat{\kappa}$	1.2912	0.6911	0.7941	0.9664	0.2584	0.2668	0.9838	0.2968	0.3084
	$\hat{\sigma}$	1.4823	0.4685	0.4734	1.4308	0.1770	0.1778	1.4720	0.4144	0.4185
	$\hat{\gamma}$	1.4998	0.1209	0.1209	1.5010	0.0494	0.0494	1.4989	0.1126	0.1126
KF	$\hat{\mu}$	0.0963	0.0576	0.0580	0.0907	0.0078	0.0079	0.0909	0.0080	0.0081
	$\hat{\kappa}$	1.2743	0.6663	0.7643	0.9600	0.2537	0.2607	0.9493	0.2694	0.2738
	$\hat{\sigma}$	1.3874	0.4299	0.4308	1.3757	0.1659	0.1704	1.2482	0.3846	0.4189
	$\hat{\gamma}$	1.4762	0.1240	0.1263	1.4874	0.0488	0.0504	1.4381	0.1245	0.1390
MCMC	$\hat{\mu}$	0.0961	0.0494	0.0498	0.0907	0.0078	0.0079	0.0911	0.0081	0.0082
	$\hat{\kappa}$	1.2958	0.6871	0.7930	0.9665	0.2570	0.2654	0.9799	0.2891	0.3000
	$\hat{\sigma}$	1.5353	0.5573	0.5703	1.4465	0.1839	0.1867	1.5296	0.4573	0.4717
	$\hat{\gamma}$	1.4944	0.1345	0.1347	1.5036	0.0512	0.0513	1.5040	0.1198	0.1199
GMM	$\hat{\mu}$	0.0952	0.1351	0.1352	0.0899	0.0076	0.0076	0.0903	0.0078	0.0078
	$\hat{\kappa}$	1.6327	0.6541	0.9822	1.1897	0.2754	0.3998	1.1551	0.2834	0.3812
	$\hat{\sigma}$	1.2845	0.5483	0.5635	1.3012	0.2754	0.2977	0.9994	0.3995	0.5759
	$\hat{\gamma}$	1.4259	0.1689	0.1845	1.4541	0.0905	0.1015	1.3218	0.1594	0.2391

References

- Merton, R.C. Theory of rational option pricing. *Bell J. Econ. Manag. Sci.* **1973**, *4*, 141–183. [CrossRef]
- Black, F.; Scholes, M. The pricing of options and corporate liabilities. *J. Political Econ.* **1973**, *81*, 637–654. [CrossRef]
- Vasicek, O. An equilibrium characterization of the term structure. *J. Financ. Econ.* **1977**, *5*, 177–188. doi:10.1016/0304-405x(77)90016-2. [CrossRef]
- Brennan, M.J.; Schwartz, E.S. A continuous time approach to the pricing of bonds. *J. Bank. Financ.* **1979**, *3*, 133–155. [CrossRef]
- Cox, J.C.; Ingersoll, J.E., Jr.; Ross, S.A. An intertemporal general equilibrium model of asset prices. *Econometrica* **1985**, *53*, 363–384. [CrossRef]
- Chan, K.C.; Karolyi, G.A.; Longstaff, F.A.; Sanders, A.B. An empirical comparison of alternative models of the short-term interest rate. *J. Financ.* **1992**, *47*, 1209–1227. [CrossRef]
- Ahn, D.H.; Gao, B. A parametric nonlinear model of term structure dynamics. *Rev. Financ. Stud.* **1999**, *12*, 721–762. [CrossRef]

8. Phillips, P.C.; Yu, J. Jackknifing bond option prices. *Rev. Financ. Stud.* **2005**, *18*, 707–742. [[CrossRef](#)]
9. Stübinger, J.; Endres, S. Pairs trading with a mean-reverting jump-diffusion model on high-frequency data. *Quant. Financ.* **2018**, *18*, 1735–1751. [[CrossRef](#)]
10. Endres, S.; Stübinger, J. Optimal trading strategies for Lévy-driven Ornstein-Uhlenbeck processes. *Appl. Econ.* **2019**, *51*, 3153–3169. [[CrossRef](#)]
11. Amorino, C.; Gloter, A. Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scand. J. Stat.* **2020**, *47*, 279–346. [[CrossRef](#)]
12. Endres, S.; Stübinger, J. A flexible regime switching model with pairs trading application to the S&P 500 high-frequency stock returns. *Quant. Financ.* **2019**, *19*, 1727–1740. [[CrossRef](#)]
13. Sørensen, H. Parametric inference for diffusion processes observed at discrete points in time: A survey. *Int. Stat. Rev.* **2004**, *72*, 337–354. [[CrossRef](#)]
14. Hurn, A.S.; Jeisman, J.; Lindsay, K.A. Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *J. Financ. Econom.* **2007**, *5*, 390–455. [[CrossRef](#)]
15. Shoji, I.; Ozaki, T. Comparative study of estimation methods for continuous time stochastic processes. *J. Time Ser. Anal.* **1997**, *18*, 485–506. [[CrossRef](#)]
16. Durham, G.B.; Gallant, A.R. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Stat.* **2002**, *20*, 297–338. [[CrossRef](#)]
17. Ozaki, T. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: A local linearization approach. *Stat. Sin.* **1992**, *2*, 113–135.
18. Shoji, I.; Ozaki, T. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stoch. Anal. Appl.* **1998**, *16*, 733–752. [[CrossRef](#)]
19. Ait-Sahalia, Y. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica* **2002**, *70*, 223–262. [[CrossRef](#)]
20. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
21. Elerian, O.; Chib, S.; Shephard, N. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **2001**, *69*, 959–993. [[CrossRef](#)]
22. Eraker, B. MCMC analysis of diffusion models with application to finance. *J. Bus. Econ. Stat.* **2001**, *19*, 177–191. [[CrossRef](#)]
23. Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, 1029–1054. [[CrossRef](#)]
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
25. Karatzas, I.; Shreve, S. *Brownian Motion and Stochastic Calculus*; Springer: Berlin/Heidelberg, Germany, 1998; Volume 113. [[CrossRef](#)]
26. Maruyama, G. Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo* **1955**, *4*, 48. [[CrossRef](#)]
27. Kloeden, P.E.; Platen, E. *Numerical Solution of Stochastic Differential Equations*; Springer: Berlin/Heidelberg, Germany, 1992; doi:10.1007/978-3-662-12616-5. [[CrossRef](#)]
28. Elerian, O. *A Note on the Existence of a Closed Form Conditional Transition Density for the Milstein Scheme*; Working Paper; Nuffield College, Oxford University: Oxford, UK, 1998.
29. Florens-Zmirou, D. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **1989**, *20*, 547–557. [[CrossRef](#)]
30. Caines, P.E. *Linear Stochastic Systems*; Wiley: New York, NY, USA, 1988.
31. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1977**, *39*, 1–22. [[CrossRef](#)]
32. Shumway, R.H.; Stoffer, D.S. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* **1982**, *3*, 253–264. [[CrossRef](#)]
33. Jones, R.H. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **1980**, *22*, 389–395. [[CrossRef](#)]
34. Pagan, A. Some identification and estimation results for regression models with stochastically varying coefficients. *J. Econom.* **1980**, *13*, 341–363. [[CrossRef](#)]
35. Ljung, L.; Caines, P.E. Asymptotic normality of prediction error estimators for approximate system models. *Stochastics* **1980**, *3*, 29–46. [[CrossRef](#)]
36. Hannan, E.J.; Deistler, M. *The Statistical Theory of Linear Systems*; Wiley: New York, NY, USA, 1988; doi:10.1137/1.9781611972191. [[CrossRef](#)]
37. Harvey, A.C. *Forecasting, Structural Time Series Models and the Kalman Filter*; Cambridge University Press: Cambridge, UK, 1990; doi:10.1017/cbo9781107049994. [[CrossRef](#)]
38. Jazwinski, A.H. *Stochastic Processes and Filtering Theory*; Academic Press: New York, NY, USA, 1970.
39. Singer, H. Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Itô-Taylor expansion. *J. Comput. Graph. Stat.* **2002**, *11*, 972–995. [[CrossRef](#)]
40. Kloeden, P.E.; Platen, E. Higher-order implicit strong numerical schemes for stochastic differential equations. *J. Stat. Phys.* **1992**, *66*, 283–314. [[CrossRef](#)]
41. Newey, W.K.; McFadden, D. Large sample estimation and hypothesis testing. *Handb. Econom.* **1994**, *4*, 2111–2245.
42. Uhlenbeck, G.E.; Ornstein, L.S. On the theory of the Brownian motion. *Phys. Rev.* **1930**, *36*, 823. [[CrossRef](#)]

43. Eddelbuettel, D.; François, R. Rcpp: Seamless R and C++ Integration. *J. Stat. Softw.* **2011**, *40*, 1–18. [[CrossRef](#)]
44. Yu, J.; Phillips, P.C. A Gaussian approach for continuous time models of the short-term interest rate. *Econom. J.* **2001**, *4*, 210–224. [[CrossRef](#)]
45. Ball, C.A.; Torous, W.N. Unit roots and the estimation of interest rate dynamics. *J. Empir. Financ.* **1996**, *3*, 215–238. doi:10.1016/0927-5398(95)00018-6. [[CrossRef](#)]
46. Tang, C.Y.; Chen, S.X. Parameter estimation and bias correction for diffusion processes. *J. Econom.* **2009**, *149*, 65–81. [[CrossRef](#)]
47. Aït-Sahalia, Y. Transition densities for interest rate and other nonlinear diffusions. *J. Financ.* **1999**, *54*, 1361–1395. doi:10.1111/0022-1082.00149. [[CrossRef](#)]
48. Kalman, R.E. Mathematical description of linear dynamical systems. *J. Soc. Ind. Appl. Ser. A Control* **1963**, *1*, 152–192. [[CrossRef](#)]
49. Brennan, M.J.; Schwartz, E.S. Analyzing convertible bonds. *J. Financ. Quant.* **1980**, *15*, 907–929. [[CrossRef](#)]
50. Pagan, A.R.; Hall, A.D.; Martin, V. Modeling the term structure. *Handb. Stat.* **1996**, *14*, 91–118. [[CrossRef](#)]
51. Monsalve-Cobis, A.; González-Manteiga, W.; Febrero-Bande, M. Goodness-of-fit test for interest rate models: An approach based on empirical processes. *Comput. Stat. Data Anal.* **2011**, *55*, 3073–3092. [[CrossRef](#)]
52. Stoffer, D.S.; Wall, K.D. Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Am. Stat. Assoc.* **1991**, *86*, 1024–1033. [[CrossRef](#)]