Journal of Econometrics 114 (2003) 107 – 139

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Maximum likelihood estimation of time-inhomogeneous diffusions

Alexei V. Egorov[a], Haitao Li[b,*], Yuewu Xu[c]

[a] Department of Economics, Cornell University, Ithaca, NY 14853, USA
[b] Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, USA
[c] Yale School of Management, Yale University, New Haven, CT 06520, USA

Received 22 March 2000; received in revised form 19 August 2002; accepted 13 September 2002

Abstract

We extend the maximum likelihood estimation method of Ait-Sahalia (Econometrica 70 (2002) 223) for time-homogeneous diffusions to time-inhomogeneous ones. We derive a closed-form approximation of the likelihood function for discretely sampled time-inhomogeneous diffusions, and prove that this approximation converges to the true likelihood function and yields consistent parameter estimates. Monte Carlo simulations for several financial models reveal that our method largely outperforms other widely used numerical procedures in approximating the likelihood function. Furthermore, parameter estimates produced by our method are very close to the parameter estimates obtained by maximizing the true likelihood function, and superior to estimates obtained from the Euler approximation.

JEL classification: C0; C1; C4; G0

Keywords: Maximum likelihood estimation; Time-inhomogeneous diffusion; Transition density; Hermite expansion

1. Introduction

Continuous-time diffusion processes have been widely used by financial economists to model the stochastic behavior of different economic variables, such as interest rates, exchange rates, and stock prices. The famous option pricing model of Black and Scholes (1973), and term structure models of Vasicek (1977), Cox et al. (1985), and Heath et al. (1992), all assume that the underlying state variables follow diffusions.

* Corresponding author. Tel.: +1-607-255-4961; fax: +1-607-254-4590.
E-mail address: hl70@cornell.edu (Haitao Li).

PII: S0304-4076(02)00221-X

Given the importance of diffusion models in modern finance theories, and the availability of high-quality data on prices of many financial assets, developing rigorous statistical methods to estimate and test such models has become a major subject of the emerging field of financial econometrics.[1] However, while finance theories are written in continuous-time framework, data in empirical work are always sampled discretely. As shown by Lo (1988), estimating the discretized version of a continuous-time model can result in inconsistent estimates. Different methods have been developed in the literature for estimating diffusions without a continuous record of observations. These include the non-parametric methods of Ait-Sahalia (1996a, b) and Stanton (1997), the generalized method of moments of Hansen and Scheinkman (1995) and Conley et al. (1997), and the simulation methods of Duffie and Singleton (1993), to cite just a few examples.

In classical statistics, if we are confident about the parametric specification of the model that is under investigation, maximum likelihood is the method of choice, given the "nice" properties of the maximum likelihood estimators. However, this method has not been widely used in estimating diffusions because closed-form solutions for the transition density of diffusion processes, and hence, the likelihood functions are generally unknown,[2] and obtaining the transition density requires the forward Kolmogorov equation to be solved numerically (e.g., Lo, 1988). In addition, when looking for the maximum value of the log-likelihood function, the forward Kolmogorov equation has to be solved each time when the value of the parameter changes infinitesimally as part of the numerical maximization algorithm. As a result, it is not trivial to implement this method in practice.

The seminal papers of Ait-Sahalia (1999, 2002) make the maximum likelihood estimation of diffusions practically feasible by providing a closed-form approximation of the transition density of the one-dimensional time-homogeneous diffusion process,

$$\mathrm{d}X_t = \mu(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\theta)\,\mathrm{d}W_t, \tag{1}$$

where W_t is the standard Brownian Motion, and θ is a vector of model parameters that need to be estimated. The closed-form approximation of the transition density of the diffusion in Eq. (1) allows one to construct a sequence of likelihood functions that are in closed form and converge to the true likelihood function. Ait-Sahalia (2002) shows that the estimator obtained from maximizing this sequence converges to the true maximum likelihood estimator. This important result greatly reduces the numerical burden of using other computationally intensive approaches, and makes it very convenient to estimate and test any one-dimensional diffusion process in empirical studies. Schaumburg (2000) develops a method for maximum likelihood estimation of discretely sampled Levy processes, which uses approximation structure similar to that of Ait-Sahalia (2002).

While Ait-Sahalia (2002) considers only time-homogeneous diffusions, there are reasons to believe that the underlying data generating process for many economic variables might change over time, due to changes in business cycles, monetary policy,

[1] See Melino (1994), Tauchen (1997), Campbell et al. (1997), and Sundaresan (2001) for reviews of some of the recent literature.

[2] See Wong (1964) for a small list of exceptions.

and general macroeconomic conditions. For example, it is a well-known fact that the volatility of interest rates at all maturities increases on Federal Open Market Committee (FOMC) meeting days, and around releases of key macroeconomic information, especially non-farm payroll employment and the consumer price index (e.g., Piazzesi, 2000, and the references therein). The well-documented calendar effects also show that stock prices behave differently on different days of the week, month, and year (e.g., Jacobs and Levy, 1988). The dynamics of certain asset prices, such as fixed-income securities and options, change over time, as the maturities of the contracts approach.

One possible approach to capture the time-dependent behavior of asset prices given in the above examples is to model the drift and diffusion terms in Eq. (1) as functions of not only X_t and θ, but also time t, as in the following equation:

$$\mathrm{d}X_t = \mu(X_t, t; \theta)\,\mathrm{d}t + \sigma(X_t, t; \theta)\,\mathrm{d}W_t. \tag{2}$$

In fact, time-inhomogeneous models of option pricing and term structure of interest rates have been developed in the finance literature. For example, to capture the "smiles" observed in the implied volatility from option prices, Rubinstein (1994), Derman and Kani (1994), and Dupire (1994) model stock return volatility as a deterministic function of stock price and time, and develop different techniques for pricing options on such assets. Hull and White (1990, 1993), Black et al. (1990), and Black and Karasinski (1991) also develop time-inhomogeneous term structure models where the short rate follows a process like that in Eq. (2).

Given that the theory is silent on whether time-homogeneous diffusion is the right process for describing financial data except for its simplicity, and that there is evidence of time-inhomogeneity in many asset prices, we should let the data decide whether time-inhomogeneity is an important feature that finance theories should seriously consider. [3] Therefore, we need to develop econometric methods to estimate and test time-inhomogeneous diffusion models. [4]

In this paper, we extend Ait-Sahalia's (2002) results for time-homogeneous diffusions to the important class of time-inhomogeneous ones. We build on Ait-Sahalia's (2002) insight by conducting two transformations of the original process such that the transition density of the transformed process is close to a normal distribution. Then we approximate the transition density by a linear combination of Hermite polynomials. The difference from Ait-Sahalia's (2002) results is that we have to explicitly take into account the time-varying feature of the drift and diffusion coefficients of the process. We show that under certain regularity conditions, our method produces parameter estimates that converge to the true parameter values. We show by numerical simulations that (i) our method is significantly better in approximating the likelihood function than the Euler scheme, Monte Carlo simulations, the binomial model of Nelson

[3] This idea has been expressed by Ait-Sahalia (1996a, pp. 528–529) in his justification for using the non-parametric method. "In the absence of any theoretical rationale for adopting one particular specification of the diffusion over the other, the question must ultimately be decided by taking the model to the data".

[4] Like any parametric approach, the maximum likelihood method of Ait-Sahalia (2002) and its extension in this paper rely on the important assumption that the underlying model is correctly specified. To test model misspecification, we can use the non-parametric tests developed in Ait-Sahalia (1996b) and Hong and Li (2002).

and Ramaswamy (1990), and the Crank–Nicholson scheme for the forward Kolmogorov equation, and (ii) parameter estimates produced by our method are very close to the parameters obtained by maximizing the true likelihood function and superior to the parameter estimates obtained from the Euler approximation.

The rest of this paper is organized as follows. Section 2 provides the main mathematical results about the maximum likelihood estimation of time-inhomogeneous diffusions. Section 3 studies the performance of our method through numerical simulations. Section 4 concludes the paper.

2. Maximum likelihood estimation via Hermite approximation

Consider the general time-inhomogeneous diffusion process in Eq. (2), where μ and σ not only depend on X_t and θ, but also on time t. Let $p_X(x,t;\theta)$ be the unconditional density of X_t at time t, and $p_X(x,t|x_s,s;\theta)$ the transition density of X_t given $X_s = x_s$. Suppose we have discretely sampled observations of X_t on dates $t_1,t_2,\dots,t_n$, where $t_i - t_{i-1} = h$, with h being fixed. Due to the Markovian nature of the diffusion process, the log-likelihood function of these observations equals $L_n(\theta) = \sum_{i=1}^{n} \log p_X(X_{t_i},t_i|X_{t_{i-1}},t_{i-1};\theta)$.[5] The maximum likelihood estimator of model parameters can be obtained by maximizing the log-likelihood function just constructed:

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

The difficulty in carrying out the above maximum likelihood procedure has been widely recognized due to the work of Lo (1988): the transition density $p_X(x,t|x_s,s;\theta)$ rarely has a closed-form solution, and in most cases $p_X(x,t|x_s,s;\theta)$ has to be obtained numerically, which makes the procedure almost impossible to implement in practice.

In what follows we extend the important results of Ait-Sahalia (2002) for time-homogeneous diffusions to general time-inhomogeneous diffusions. The main task is to obtain a closed-form approximation of the transition density $p_X(x,t|x_s,s)$ and hence the log-likelihood function $L_n(\theta)$.

Our approach is to construct a sequence of functions $p_X^{(J)}(x,t|x_s,s;\theta)$ that closely approximates $p_X(x,t|x_s,s;\theta)$. This is done by expanding a transformation of $p_X(x,t|x_s,s;\theta)$ with respect to the Hermite base of the Hilbert space $L^2(\mathrm{P})$, where P is the standard normal distribution. In Section 2.3 we show that $p_X^{(J)}(x,t|x_s,s;\theta)$ can be obtained in closed form as a function of the drift and diffusion parameters μ and σ. This approximation to $p_X(x,t|x_s,s;\theta)$ leads naturally to a sequence of log-likelihood functions $L_n^{(J)}(\theta)$ that approximate $L_n(\theta)$:

$$L_n^{(J)}(\theta) = \sum_{i=1}^{n} \log\left(p_X^{(J)}(X_{t_i},t_i|X_{t_{i-1}},t_{i-1};\theta) \vee \frac{\varepsilon}{J}\right), \tag{3}$$

[5] We ignore initial density $p_X(x_0,t_0)$ in the definition of likelihood function for the simplicity of exposition. Retaining it does not change our results. In practice omitting initial density in the likelihood function can be justified when sample is large, the initial observation x_0 is not very unlikely, or x_0 is known.

where ε is a small positive constant. The approximation can be made arbitrarily precise with a big enough J. Because $p_X^{(J)}$ may turn negative or zero, we truncate $p_X^{(J)}(X_{t_i}, t_i|X_{t_{i-1}}, t_{i-1}; \theta)$ from below by ε/J.[6] This bound does not depend on sample size and disappears when $J \to \infty$.

2.1. Two transformations

Two transformations of the original process X_t are needed before such an approximation can be obtained. The purpose of these two transformations, as explained in Ait-Sahalia (2002), is to make the transition density of the transformed process close to a normal distribution, so that the standard Hermite expansion can be applied to such distributions.

Let Θ be a compact set in $\mathbf{R}^\kappa$ for some fixed integer κ, and D_X denote the domain of the diffusion process X_t. We assume that D_X is either $(-\infty, \infty)$ or $(0, \infty)$, which is usual for diffusion processes used in finance. We modify the regularity conditions of Ait-Sahalia (2002) to suit time-inhomogeneous diffusions.

Assumption 1 (Smoothness of the Coefficients). Functions $\mu(x, t; \theta)$ and $\sigma(x, t; \theta)$ are infinitely differentiable in $t \in [0, \infty)$ and $x \in D_X$, and twice continuously differentiable in $\theta \in \Theta$.

Assumption 2 (Non-Degeneracy of the Diffusion). 1. If $D_X = (-\infty, \infty)$, then $\sigma(x, t; \theta)$ is non-degenerate: for any positive T there exists a constant c_T such that $\sigma(x, t; \theta) \geqslant c_T > 0$ for all $(x, t, \theta) \in D_X \times [0, T] \times \Theta$.

2. If $D_X = (0, +\infty)$, then $\sigma(x, t; \theta)$ is non-degenerate away from $x = 0$, but can be locally degenerate at $x=0$: (i) for any positive T and ξ, there exists a constant $c_{T,\xi}$ such that $\sigma(x, t; \theta) \geqslant c_{T,\xi} > 0$ for all $(x, t, \theta) \in [\xi, +\infty) \times [0, T] \times \Theta$, and (ii) if $\sigma(0, t; \theta) = 0$ for some positive t, then for any positive T there exist positive constants ξ_0, ω, ρ such that $\sigma(x, t; \theta) \geqslant \omega x^\rho$ for all $(x, t, \theta) \in (0, \xi_0] \times [0, T] \times \Theta$.

The first transformation standardizes the variance of the density so that it has unit variance. Using map $\varphi : D_X \times [0, \infty) \times \Theta \to R$

$$(x, t, \theta) \mapsto y = \varphi(x; t, \theta) \equiv \int^x \frac{\mathrm{d}w}{\sigma(w, t; \theta)}, \tag{4}$$

the first transformation of X_t is defined as

$$Y_t = \varphi(X_t; t, \theta). \tag{5}$$

Because σ is positive on the interior of the domain D_X, φ is strictly increasing in x for any given t and θ. Therefore, it is always invertible. We denote this

[6] One should choose ε small enough to guarantee that for all θ for which $p_X^{(J)}(\theta)$ turns negative or zero, the log-likelihood function $L_n^{(J)}(\theta)$ is a sufficiently large negative number, so that when maximizing of log-likelihood function $L_n^{(J)}(\theta)$ such parameters θ will be discarded. Of course in practice, ε should be such that $\log \varepsilon$ is computer representable. So in our simulations below, we chose $\varepsilon = \exp(-10^{25})$.

inverse by φ^{-1} and the domain of Y_t by D_Y. Applying Itô's formula to Y_t, we get that Y_t follows

$$\mathrm{d}Y_t = \mu_Y(Y_t, t; \theta)\,\mathrm{d}t + \mathrm{d}W_t, \tag{6}$$

where

$$\begin{aligned}
\mu_Y(y,t;\theta) &= \left[\frac{\partial\varphi}{\partial x}\,\mu(x,t;\theta) + \frac{\partial\varphi}{\partial t} + \frac{1}{2}\,\frac{\partial^2\varphi}{\partial x^2}\,\sigma^2(x,t;\theta)\right] \\
&= \frac{\mu(\varphi^{-1}(y;t,\theta),t;\theta)}{\sigma(\varphi^{-1}(y;t,\theta),t;\theta)} + \frac{\partial\varphi}{\partial t}\,(\varphi^{-1}(y;t,\theta),t;\theta) \\
&\quad - \frac{1}{2}\,\frac{\partial\sigma}{\partial x}\,(\varphi^{-1}(y;t,\theta),t;\theta).
\end{aligned}$$

The second transformation is a linear map that transforms Y_t into another process Z_t. It is defined as $\psi : D_Y^2 \times \mathbf{R}_+ \to \mathbf{R}$, such that

$$(y, y_s, h) \mapsto z = \psi(y; y_s, h) \equiv \frac{y - y_s}{\sqrt{h}}.$$

The inverse of ψ is

$$y = \psi^{-1}(z; y_s, h) = \sqrt{h}z + y_s,$$

where h is the fixed sampling interval. Let $t = s + h$. We define process Z_t with domain denoted by D_Z as

$$Z_t = \psi(Y_t; y_s, h) = \frac{Y_t - y_s}{\sqrt{h}}.$$

Let $p_Y(y,t|y_s,s;\theta)$ be the transition density of Y_t given $Y_s = y_s$, and $p_Z(z,t|y_s,s;\theta)$ be the transition density of Z_t given $Y_s = y_s$. The transition densities of p_X, p_Y, and p_Z are related in the following ways:

$$\begin{aligned}
p_Z(z,t|y_s,s;\theta) &= \sqrt{h}\,p_Y(\sqrt{h}z + y_s|y_s,s;\theta), \\
p_Y(y,t|y_s,s;\theta) &= \frac{1}{\sqrt{h}}\,p_Z\left(\frac{y-y_s}{\sqrt{h}}, t|y_s,s;\theta\right)
\end{aligned}$$

and

$$\begin{aligned}
p_Y(y,t|y_s,s;\theta) &= \sigma(\varphi^{-1}(y;t,\theta),t;\theta)\,p_X(\varphi^{-1}(y;t,\theta),t|\varphi^{-1}(y_s;s,\theta),s;\theta), \\
p_X(x,t|x_s,s;\theta) &= \frac{1}{\sigma(x,t;\theta)}\,p_Y(\varphi(x;s,\theta),t|\varphi(x_s;s,\theta),s;\theta).
\end{aligned}$$

Thus, if the transition density p_Z or its approximation is known, then the approximation for p_X is obtained naturally. Next, we will show how to obtain such approximations.

2.2. Approximating the likelihood function

Observe that for a Hilbert space $L^2(\mathrm{P})$ with measure P, density ϕ and inner product $\langle\cdot,\cdot\rangle$, a "nice" density $g(w)$ on the real line can be represented by an expansion with respect to an orthogonal base $\{H_1, H_2, \ldots\}$. That is,

$$g = \sum_{k=0}^{\infty} \frac{\langle g, H_k\rangle}{\langle H_k, H_k\rangle} H_k, \quad \text{where } \langle g, H_k\rangle = \int_{-\infty}^{\infty} g(w) H_k(w) \phi(w)\,\mathrm{d}w.$$

For any density $p(w)$ on $\mathbf{R}$, if we expand the ratio $p(w)/\phi(w)$ according to the above relation, we get

$$\frac{p(w)}{\phi(w)} = \sum_{k=0}^{\infty} \frac{\langle p/\phi, H_k\rangle}{\langle H_k, H_k\rangle} H_k,$$

which simply yields

$$p(w) = \phi(w) \sum_{k=0}^{\infty} \beta_k H_k, \tag{7}$$

where the coefficients are given as

$$\beta_k = \frac{\langle p/\phi, H_k\rangle}{\langle H_k, H_k\rangle} = \frac{\int_{-\infty}^{\infty} p(w) H_k(w)\,\mathrm{d}w}{\int_{-\infty}^{\infty} H_k(w) H_k(w) \phi(w)\,\mathrm{d}w}. \tag{8}$$

If the probability measure P is the standard normal distribution, a natural choice for the orthogonal base $\{H_0, H_1, \ldots\}$ would be the Hermite base, as in Ait-Sahalia (2002). While it is entirely feasible to consider other bases, we will focus our discussion on the Hermite base in this paper. The Hermite polynomials, defined as

$$H_k(w) \equiv \phi(w)^{-1} \frac{\mathrm{d}^k}{\mathrm{d}w^k} \phi(w), \quad k \geqslant 0, \tag{9}$$

where ϕ denotes the standard normal density $\phi(w) = 1/\sqrt{2\pi}\mathrm{e}^{-w^2/2}$, have the following properties.

Proposition 1. *The Hermite polynomials H_k satisfy*:

(1) $(k+1)H_k(w) = (\mathrm{d}/\mathrm{d}w)H_{k+1}(w)$ *for all w in $\mathbf{R}$ and every non-negative integer k.*

(2) $\int_{-\infty}^{\infty} H_i(w) H_k(w) \phi(w)\,\mathrm{d}w = \begin{cases} i! & if\ k = i, \\ 0 & if\ k \neq i. \end{cases}$

(3) *There exists a constant ϖ such that for all z in $\mathbf{R}$ and every integer k*

$$|H_k(z)| \leqslant \frac{\varpi\sqrt{k!}}{\sqrt[4]{k}} \left(1 + \frac{z^{5/2}}{2^{5/4}}\right) \mathrm{e}^{z^2/4}. \tag{10}$$

Proof. See Sansone (1991, p. 304, 308) and Stone (1927, Theorem II).

Hermite polynomials are easily computed using their definition (9). In particular, the first seven Hermite polynomials are $H_0(w) = 1$, $H_1(w) = -w$, $H_2(w) = w^2 - 1$, $H_3(w) = -w^3 + 3w$, $H_4(w) = w^4 - 6w^2 + 3$, $H_5(w) = -w^5 + 10w^3 - 15w$, $H_6(w) = w^6 - 15w^4 + 45w^2 - 15$.

Specializing (7) to the Hermite base, we obtain the Hermite expansion of transition density $p_Z(z, t|y_s, s)$:

$$p_Z(z, t|y_s, s; \theta) = \phi(z) \sum_{k=0}^{\infty} \beta_k(t, y_s, s; \theta) H_k(z), \tag{11}$$

where coefficients β_k equal

$$\beta_k(t, y_s, s) = \frac{1}{k!} \int_{-\infty}^{\infty} H_k(z) p_Z(z, t|y_s, s)\, \mathrm{d}z. \tag{12}$$

Let $p_Z^{(J)}$ denote the partial sum of the Hermite expansion (11) of p_Z:

$$p_Z^{(J)}(z, t|y_s, s) \equiv \phi(z) \sum_{k=0}^{J} \beta_k(t, y_s, s) H_k(z). \tag{13}$$

The corresponding approximations for p_Y and p_X are

$$p_Y^{(J)}(y, t|y_s, s; \theta) \equiv \frac{1}{\sqrt{h}}\, p_Z^{(J)}\left(\frac{y - y_s}{\sqrt{h}}, t|y_s, s; \theta\right), \tag{14}$$

$$p_X^{(J)}(x, t|x_s, s; \theta) \equiv \frac{1}{\sigma(x, t)}\, p_Y^{(J)}(\varphi(x; t, \theta), t|\varphi(x_s; s, \theta), s; \theta). \tag{15}$$

For the rest of the paper, we assume that transformation (4) is such that the domain of Y_t is $D_Y = (y_-, \infty)$ where y_- can be either 0 or $-\infty$. This assumption and the similar above assumption on D_X simplify our exposition. Moreover, for most, if not all, financial models the domains D_X andD_Y are either $(-\infty, \infty)$ or $(0, \infty)$. Let

$$g(y, t; \theta) \equiv -\frac{1}{2}\left[\mu_Y(y, t; \theta)^2 + \frac{\partial}{\partial y}\mu_Y(y, t; \theta)\right] - \int_{y_s}^{y} \frac{\partial}{\partial t}\mu_Y(y, t; \theta)\, \mathrm{d}y. \tag{16}$$

The next assumption puts restrictions on behavior of the drift μ_Y near the boundaries of D_Y. When μ_Y does not explicitly depend on t, this assumption becomes the corresponding assumption of Ait-Sahalia (2002) for time-homogeneous diffusion processes.

Assumption 3 (Boundary Behavior). For all $(t, \theta) \in [0, \infty) \times \Theta$, μ_Y and its derivatives with respect to y, t and θ have at most polynomial growth in y near the boundaries of D_Y.[7]

1. $\lim_{y \to \infty \text{ or } y_-} g(y, t; \theta) < +\infty$ for all $(t, \theta) \in [0, \infty) \times \Theta$.

[7] Function $f(y)$ is said to have at most polynomial growth in y, when $y \to \infty$ ($y \to -\infty$, $y \to 0$) if there exists a non-negative integer p such that $f(y)y^{-p}$ ($f(y)y^{-p}$, $f(y)y^{p}$)) is bounded in some vicinity of $y = \infty$ ($y = -\infty, y = 0$).

2. Moreover, there exists a positive constant Γ such that for any positive T, there exists a positive constant η such that (i) $\mu_Y(y,t;\theta) \leqslant \Gamma y$ for all $(t,\theta,y) \in [0,T] \times \Theta \times [\eta,\infty)$, and (ii) $\mu_Y(y,t;\theta) \geqslant \Gamma y$ for all $(t,\theta,y) \in [0,T] \times \Theta \times (-\infty,\eta]$.
3. If $y_- = 0$ then for any positive T there exist constants k, α and a positive constant η_0 such that for any $(t,\theta,y) \in [0,T] \times \Theta \times (0,\eta_0)$, $\mu_Y(y,t;\theta) \geqslant k/y^\alpha$ with either $\alpha = 1$, $k \geqslant 1$, or $\alpha > 1$, $k > 0$.

Proposition 2. *Let Assumptions* 1–3 *hold. For any distribution of initial value* Y_0, *there is a weak solution* $\{Y_t\}_{t=0}^{\infty}$ *to the stochastic differential equation* (6). *This solution is unique in probability law. With probability one,* Y_t *cannot attain the boundaries of* D_Y *in a finite time. Let constant* $K = 0$, *if* $\mu_Y \leqslant 0$ *near* $y = +\infty$ *and* $\mu_Y \geqslant 0$, *near* $y = -\infty$, *and* $K = \Gamma$ *otherwise. For any positive* T *and any* $\tilde{y} \in D_Y$, *there exist* $C_0 = C_0(T,\tilde{y})$, $C_1 = C_1(T,\tilde{y})$ *such that inequalities*

$$0 \leqslant p_Z(z, s+h|\tilde{y}, s;\theta) \leqslant C_0 \mathrm{e}^{-z^2(1-hK)/2 + C_1|z|}, \tag{17}$$

$$\left|\frac{\partial p_Z}{\partial z}(z, s+h|\tilde{y}, s;\theta)\right| \leqslant C_0 \mathrm{e}^{-z^2(1-hK)/2 + C_1|z|} \tag{18}$$

hold for all $(z,s,\theta,h) \in D_Z \times [0,T] \times \Theta \times (0, T-s]$. *If* $y_- = 0$, p_Z *vanishes as* y *tends to zero.*

Proof. See Appendix A.

Recall that X_t and Y_t are related through transformation (3). It follows immediately that under Assumptions 1–3, for any distribution of initial value X_0 stochastic differential equation (2) admits a weak solution $\{X_t\}_{t=0}^{\infty}$ which is unique in probability law, and that with probability one, X_t cannot attain the boundaries of D_X in a finite time.

The following theorem shows that the approximation $p_X^{(J)}(x,t|x_s,s)$ converges to the true transition density uniformly when the number of terms included in the expansion, J, becomes large.

Theorem 1. *Under Assumptions* 1–3, *there exists a* $\bar{h} > 0$ *such that for every* $(h,\theta,s,x,x_s) \in (0,\bar{h}) \times \Theta \times [0,\infty) \times D_X^2$,

$$p_X^{(J)}(x, s+h|x_s, s;\theta) \to p_X(x, s+h|x_s, s;\theta) \quad \text{as } J \to \infty.$$

The convergence is uniform in $\theta \in \Theta$, x_s *in any compact subset of* D_X, *and* $s \in [0,T]$ *for any positive finite* T. If σ *is non-degenerate, then the convergence is further uniform in* x *over the entire domain* D_X. *If* σ *is degenerate at zero, then the convergence is uniform in* x *in each interval of the form* $[\varepsilon, +\infty)$ *where* $\varepsilon > 0$. *If* $\mu_Y \leqslant 0$ *near the right boundary* ($y = +\infty$) *and* $\mu_Y \geqslant 0$ *near the left boundary* (*either* $y = 0$ *or* $y = -\infty$), *then* $\bar{h} = \infty$.

Proof. See Appendix A.

Based on the above approximation for the transition density, we can construct an approximation (3) for the true log-likelihood function. We call the parameter estimates obtained from the true likelihood function the "true maximum likelihood estimators".

Next, we show that parameter estimates obtained by maximizing this approximated likelihood function, $\hat{\theta}_n^{(J)} \equiv \arg\max_\theta L_n^{(J)}(\theta)$, converge to the true maximum likelihood estimator when the number of terms included in the expansion, J, becomes large. Before that, we introduce the following definitions and assumptions. Let $l_k(\theta) \equiv \log[p_X(X_{t_k}, t_k | X_{t_{k-1}}, t_{k-1}; \theta)]$, and the superscripts $'$ and T denote differentiation with respect to θ, and transposition. Let $H_n(\theta) \equiv -\sum_{k=1}^n l_k''(\theta)$, $i_n(\theta) \equiv diag \sum_{k=1}^n E_\theta[l_k'(\theta) l_k'(\theta)^T]$ and $G_n(\theta) \equiv i_n^{-1/2}(\theta) H_n(\theta) i_n^{-1/2}(\theta)$. Denote P_θ the probability measure induced by the transition density $p_X(x, t|\tilde{x}, s; \theta)$, $\|\cdot\|_p$ a norm in $L^2(\mathrm{P}_\theta)$,and $\mathrm{N}(0, \mathbf{I}_\kappa)$ a Gaussian random vector in $\mathbf{R}^\kappa$ with zero mean and identity $\kappa \times \kappa$ variance–covariance matrix $\mathbf{I}_\kappa$.

Assumption 4 (Identification). The true parameter vector θ_0 belongs to the interior of Θ, $i_n^{-1}(\theta) \xrightarrow{\text{a.s.}} 0$ uniformly in $\theta \in \Theta$, and $i_n^{-1/2}(\theta_0) \sum_{k=1}^n l_{k''}'(\theta) i_n^{-1/2}(\theta_0)$ is uniformly bounded in probability for all $\theta \in \Theta$, such that $\|i_n^{1/2}(\theta_0)(\theta - \theta_0)\| \leqslant \varepsilon$ for some $\varepsilon > 0$, where $\|\cdot\|$ is the Euclidean norm in $\mathbf{R}^\kappa$.

Assumption 5. There exists a (possibly random) matrix $G(\theta)$, almost surely finite and positive definite, such that $G_n(\theta) \xrightarrow{\text{p}} G(\theta)$ uniformly over compact subsets of Θ.

Under Assumptions 1–5 the log-likelihood ratio has locally asymptotically mixed normal (LAMN) structure of the log-likelihood ratio.[8] LAMN structure is a generalization to non-ergodic processes, of the model of Fisher–Rao, and, in particular, it implies locally asymptotically quadratic (LAQ) structure of the log-likelihood ratio (e.g., Basawa and Scott, 1983; Jeganathan, 1995).

Now we are ready to prove the next important result.

Theorem 2. *Let Assumptions* 1–5 *be satisfied with the additional constraint on k added to Assumption* 3(3): *if* $\alpha = 1$ *and in Assumption* 2(2) $\rho \in (0,1)$, *then* $k \geqslant 2\rho/(1-\rho)$. *Choose any* $h \in (0, \bar{h})$, *where* $\bar{h}$ *is defined in Theorem* 1.

(1) *Fix the sample size n, then as* $J \to \infty$, $\hat{\theta}_n^{(J)} \xrightarrow{\text{p}} \hat{\theta}_n$ *under* P_{θ_0}.
(2) *As* $n \to \infty$, *there exists a sequence* $J_n \to \infty$ *such that*

$$\hat{\theta}_n^{(J_n)} \xrightarrow{\text{p}} \theta_0 \text{ under } \mathrm{P}_{\theta_0}, \text{ and } i_n^{1/2}(\theta_0)(\hat{\theta}_n^{(J_n)} - \theta_0) \xrightarrow{\text{d}} G^{-1/2}(\theta_0) \times \mathrm{N}(0, \mathbf{I}_\kappa),$$

where $\mathrm{N}(0, \mathbf{I}_\kappa)$ *is distributed independently of* $G(\theta_0)$.

Proof. See Appendix A.

[8] See the proof of Theorem 2(2) in Appendix A, which uses Theorem 1 of Basawa and Scott (1983, p. 34).

2.3. How to implement the approximation in practice

Despite the fact that approximation (13) has nice theoretical properties, as demonstrated by Theorems 1 and 2, its use will be quite limited if the approximation cannot be evaluated easily in practice. To carry out the approximation, we need a method to evaluate the coefficients β_k. Fortunately, β_k can be evaluated in a closed form with arbitrary precision.

Denote $\mathrm{E}_s[\cdot] = \mathrm{E}[\cdot \,|\, Y_s = y_s, s; \theta]$. Definition of β_k given in (8) and Proposition 1(2) imply that

$$\beta_k(s+h, y_s, s; \theta) = \frac{1}{k!}\int_{-\infty}^{\infty} H_k(z)\, p_Z(z, s+h | y_s, s; \theta)\, \mathrm{d}z = \frac{1}{k!}\,\mathrm{E}_s[H_k(z)]. \tag{19}$$

The expectation in expression (19) for β_k can be evaluated using a variant of Taylor expansion given below in Proposition 3.

Definition. Let U_t be a time-inhomogeneous diffusion in $\mathbf{R}$. The infinitesimal generator $\mathscr{L}$ of U_t is defined by

$$(\mathscr{L} \circ \chi)(u, t) = \lim_{\tau \searrow 0} \frac{E[\chi(U_{t+\tau}, t+\tau) | U_t = u] - \chi(u, t)}{\tau}, \quad u \in \mathbf{R}. \tag{20}$$

The set of functions $\chi : \mathbf{R} \times [0, \infty) \to \mathbf{R}$ such that limit (20) exists at all $(u, t) \in \mathbf{R} \times [0, \infty)$ is denoted by $\mathscr{D}(\mathscr{L})$ and called the domain of infinitesimal generator $\mathscr{L}$.

Denote $\mathscr{D}(\mathscr{L}^i)$ the domain of operator $\mathscr{L}^i = \mathscr{L} \circ \mathscr{L} \circ \cdots \circ \mathscr{L}$ (i times). Denote $C_0^{\infty}(\mathbf{R})$ the space of infinitely differentiable functions $f : \mathbf{R} \to \mathbf{R}$ such that $z^i \mathrm{d}^j f(z)/\mathrm{d}z^j \to 0$ as $z \to \pm\infty$ for any non-negative i, j.

Proposition 3 (A Variant of Taylor Expansion). *Let Assumptions* 1–3 *be satisfied and* $\mathscr{A}_{\theta, \tilde{y}, h}$ *denotes the infinitesimal generator of the process* Z_t *for any fixed* $(\theta, \tilde{y}, h) \in \Theta \times D_Y \times (0, \infty)$. *Let* $f(z) \in C_0^{\infty}(\mathbf{R})$. *Then for any* $i = 1, 2, \ldots,$ $f \in \mathscr{D}(\mathscr{A}^i_{\theta, \tilde{y}, h})$ *and for all* $(z, t) \in D_Z \times [0, \infty)$,

$$(\mathscr{A}^i_{\theta, \tilde{y}, h} \circ f)(z, t) = \frac{\partial(\mathscr{A}^{i-1}_{\theta, \tilde{y}, h} \circ f)}{\partial z}\mu_Z + \frac{\sigma_Z^2}{2}\frac{\partial^2(\mathscr{A}^{i-1}_{\theta, \tilde{y}, h} \circ f)}{\partial z^2} + \frac{\partial(\mathscr{A}^{i-1}_{\theta, \tilde{y}, h} \circ f)}{\partial t}, \tag{21}$$

where $\mu_Z(z, t; h, y_s, \theta) \equiv \mu_Y(\sqrt{h}z + \tilde{y}, t; \theta)/\sqrt{h}$ *and* $\sigma_Z^2 \equiv 1/h$. *Moreover, for any positive constant* h *and any non-negative integer* I,

$$\mathrm{E}_s[f(Z_{s+h})] = f(0) + \sum_{i=1}^{I} (\mathscr{A}^i_{\theta, \tilde{y}, h} \circ f)(0, s)\frac{h^i}{i!} + \mathrm{E}_s[R_I(f, \mathscr{A}_{\theta, \tilde{y}, h}, Z_\xi, \xi; s, \theta, \tilde{y}, h)],$$

where $R_I(f, \mathscr{A}_{\theta, \tilde{y}, h}, Z_\xi, \xi; s, \theta, \tilde{y}, h) = (\mathscr{A}^{I+1}_{\theta, \tilde{y}, h} \circ f)(Z_\xi, \xi)(\xi - s)^{I+1}/(I+1)!$ *for some* $\xi \in (s, s+h)$.

Proof. See Appendix A.

Note that $H_k \notin C_0^\infty(\mathbf{R})$. Let $\{H_{k,j}(z) \equiv \frac{1}{2}\,\mathrm{e}^j(\cosh j + \cosh z)^{-1}H_k(z),\ z \in \mathbf{R}\}_{k,j=0}^\infty$. Since $H_{k,j} \in C_0^\infty(\mathbf{R})$, Proposition 3 applies to $H_{k,j}$. Moreover, as $j \to \infty$, $H_{k,j}(z) \to H_k(z)$ uniformly on any compact subset of $\mathbf{R}$. The same is true for any derivative of $H_{k,j}$.

If $\|R_I(H_{k,j}, \mathscr{A}_{\theta,\tilde{y},h}, Z_\xi, \xi; s, \theta, \tilde{y}, h)\|_p \to 0$ as $I \to \infty$, then taking large j, we get approximation

$$\beta_k(s+h,\tilde{y},s;\theta) \approx \frac{1}{k!}\,\mathrm{E}_s[H_{k,j}(Z_{s+h})] \approx \frac{1}{k!}\sum_{i=0}^{I}(\mathscr{A}^i_{\theta,\tilde{y},h}\circ H_{k,j})(0,s;h,\tilde{y},\theta)\frac{h^i}{i!}$$

$$\approx \frac{1}{k!}\sum_{i=0}^{I}(\mathscr{A}^i_{\theta,\tilde{y},h}\circ H_k)(0,s;h,\tilde{y},\theta)\,\frac{h^i}{i!}, \tag{22}$$

where $(\mathscr{A}^i_{\theta,\tilde{y},h}\circ H_k)(0,s;h,\tilde{y},\theta)$ is a polynomial in $\sqrt{h}$ with the smallest power of $\sqrt{h}$ equal to $k+i$. We follow Ait-Sahalia (2002) and Jensen and Poulsen (2000) in taking approximation $\beta_k^{[m]}$ for β_k as follows: for any non-negative m, we take $I=2m$ and leave in (22) only terms up to h^m. It is easy to show that (i) $\beta_k^{[m]}=0$ for all $k>2m$, and (ii) the corresponding feasible estimator is consistent if $\|R_I(H_{k,j}, \mathscr{A}_{\theta,\tilde{y},h}, Z_\xi, \xi; s, \theta, \tilde{y}, h)\|_p \to 0$ as $I \to \infty$. Since $A^i \circ H_k$ in the above expression can be computed iteratively using (21), approximation $\beta_k^{[m]}$ can be obtained in a mechanical fashion. The above procedure can be conveniently implemented using Mathematica or any software with symbolic differentiation capabilities. In particular, choosing $m=3$, we obtain coefficients

$$\begin{aligned}\beta_1^{[3]} &= -h^{1/2}\zeta - \tfrac{1}{4}h^{3/2}(2\zeta_{0,1}+2\zeta\zeta_{1,0}+\zeta_{2,0}) - \tfrac{1}{24}h^{5/2}(4\zeta_{0,2}+4\zeta_{0,1}\zeta_{1,0}+4\zeta\zeta_{1,0}^2\\ &\quad + 8\zeta\zeta_{1,1}+4\zeta^2\zeta_{2,0}+6\zeta_{1,0}\zeta_{2,0}+4\zeta_{2,1}+4\zeta\zeta_{3,0}+\zeta_{4,0}),\end{aligned}$$

$$\begin{aligned}\beta_2^{[3]} &= \tfrac{1}{2}h(\zeta^2+\zeta_{1,0}) + \tfrac{1}{12}h^2(6\zeta\zeta_{0,1}+6\zeta^2\zeta_{1,0}+4\zeta_{1,0}^2+4\zeta_{1,1}+7\zeta\zeta_{2,0}+2\zeta_{3,0})\\ &\quad + \frac{1}{96}h^3(12\zeta_{0,1}^2+16\zeta\zeta_{0,2}+40\zeta\zeta_{0,1}\zeta_{1,0}+28\zeta^2\zeta_{1,0}^2+16\zeta_{1,0}^3+32\zeta^2\zeta_{1,1}\\ &\quad + 40\zeta_{1,0}\zeta_{1,1}+12\zeta_{1,2}+16\zeta^3\zeta_{2,0}+24\zeta_{0,1}\zeta_{2,0}+88\zeta\zeta_{1,0}\zeta_{2,0}+21\zeta_{2,0}^2\\ &\quad + 40\zeta\zeta_{2,1}+28\zeta^2\zeta_{3,0}+32\zeta_{1,0}\zeta_{3,0}+12\zeta_{3,1}+16\zeta\zeta_{4,0}+3\zeta_{5,0}),\end{aligned}$$

$$\begin{aligned}\beta_3^{[3]} &= -\tfrac{1}{6}h^{3/2}(\zeta^3+3\zeta\zeta_{1,0}+\zeta_{2,0}) - \tfrac{1}{48}h^{5/2}(12\zeta^2\zeta_{0,1}+12\zeta^3\zeta_{1,0}+12\zeta_{0,1}\zeta_{1,0}\\ &\quad + 28\zeta\zeta_{1,0}^2+16\zeta\zeta_{1,1}+22\zeta^2\zeta_{2,0}+24\zeta_{1,0}\zeta_{2,0}+6\zeta_{2,1}+14\zeta\zeta_{3,0}+3\zeta_{4,0}),\end{aligned}$$

$$\begin{aligned}\beta_4^{(3)} &= \tfrac{1}{24}h^2(\zeta^4+6\zeta^2\zeta_{1,0}+3\zeta_{1,0}^2+4\zeta\zeta_{2,0}+\zeta_{3,0}) + \tfrac{1}{240}h^3(20\zeta^3\zeta_{0,1}+20\zeta^4\zeta_{1,0}\\ &\quad + 60\zeta\zeta_{0,1}\zeta_{1,0}+100\zeta^2\zeta_{1,0}^2+40\zeta_{1,0}^3+40\zeta^2\zeta_{1,1}+40\zeta_{1,0}\zeta_{1,1}+50\zeta^3\zeta_{2,0}\\ &\quad + 20\zeta_{0,1}\zeta_{2,0}+180\zeta\zeta_{1,0}\zeta_{2,0}+34\zeta_{2,0}^2+30\zeta\zeta_{2,1}+50\zeta^2\zeta_{3,0}+52\zeta_{1,0}\zeta_{3,0}\\ &\quad + 8\zeta_{3,1}+23\zeta\zeta_{4,0}+4\zeta_{5,0}),\end{aligned}$$

$$\beta_5^{[3]} = -\tfrac{1}{120}\, h^{5/2}(\zeta^5 + 10\zeta^3\zeta_{1,0} + 10\zeta^2\zeta_{2,0} + 10\zeta_{1,0}\zeta_{2,0} + 15\zeta\zeta_{1,0}^2 + 5\zeta\zeta_{3,0} + \zeta_{4,0}),$$

$$\beta_6^{[3]} = \tfrac{1}{720}\, h^3(\zeta^6 + 15\zeta^4\zeta_{1,0} + 20\zeta^3\zeta_{2,0} + 10\zeta_{2,0}^2 + 15\zeta_{1,0}^3 + 15\zeta_{1,0}\zeta_{3,0} + 45\zeta^2\zeta_{1,0}^2$$
$$+ 15\zeta^2\zeta_{3,0} + 60\zeta\zeta_{1,0}\zeta_{2,0} + 6\zeta\zeta_{4,0} + \zeta_{5,0}),$$

$$\beta_0^{[3]} = 1 \text{ and } \beta_k^{[3]} = 0 \quad \text{for } k > 6,$$

where $\zeta \equiv \mu_Y(y_s, s)$ and $\zeta_{i,j} \equiv \partial^{i+j}\mu_Y(y,s)/\partial y^i \partial s^j|_{y=y_s}$ for all i and j. We get $\beta_k^{[2]}$, if we drop terms with $h^{5/2}$ and h^3 in the above formulas. If we additionally drop terms with $h^{3/2}$ and h^2 then we obtain expressions for $\beta_k^{[1]}$. When μ_Y does not depend on t, all coefficients reduce to those of Ait-Sahalia (2002) for time-homogeneous diffusions.

Taking $\beta_k^{[m]}$ instead of β_k in (13), and recalling that all $\beta_k^{[m]} = 0$ for $k > 2m$, we get the closed form approximation $p_Z^{[m]}$ for p_Z:

$$p_Z(z,t|y_s,s;\theta) \approx p_Z^{[m]}(z,t|y_s,s;\theta) \equiv \phi(z)\sum_{k=0}^{2m} \beta_k^{[m]}(t,y_s,s;\theta)H_k(z). \tag{23}$$

The corresponding approximation $p_X^{[m]}$ for p_X is defined as

$$p_X(x,s+h|x_s,s;\theta) \approx p_X^{[m]}(x,t|x_s,s;\theta)$$
$$\equiv \frac{1}{\sigma(x,t)\sqrt{h}}\, p_Z^{[m]}\left(\frac{\varphi(x,s+h,\theta)-y_s}{\sqrt{h}}, s+h|y_s,s;\theta\right).$$

Finally, using $p_X^{[m]}$ instead of $p_X^{(J)}$, we get a feasible estimator $\hat\theta^{[m]}$ for θ_0. This feasible estimator is consistent if Theorems 1 and 2 hold for $p_X^{[m]}$. From (22) it follows that for this end it is enough that, as $I, j \to \infty$, $\|R_I(H_{k,j}, \mathscr{A}_{\theta,\tilde y,h}, Z_\xi, \xi; s, \theta, \tilde y, h)\|_\theta \to 0$ uniformly over the same sets as in Theorem 1.

If μ_Y does not depend explicitly on t, we can use the results of Ait-Sahalia (1999, 2002), which rely on Hansen and Scheinkman (1995) and Hansen et al. (1995). Note that X_t is still a time-inhomogeneous diffusion if the diffusion coefficient σ depends explicitly on t, since, according to (4), the transformation $X_t = \varphi^{-1}(Y_t; t, \theta)$ depends on σ. The next corollary shows that in that case, the feasible estimator is consistent for a broad class of diffusions.

Corollary 1. *Let Assumptions* 1–5 *hold with* $\alpha > 1$ *in Assumption* 3(3) *and* $\mu_Y = \mu_Y(y,\theta)$. *Then the feasible estimator* $\hat\theta^{[m]}$ *is consistent.*

Proof. See Appendix A.

In general, however, for time-inhomogeneous diffusions the drift coefficient μ_Y does depend explicitly on time t. To proceed further we make the following definitions. Let $\tilde Z_t$ be the (time-homogeneous) diffusion process which differs from Z_t only by the drift term $\mu_{\tilde Z} \equiv \mu_Z(z,s;h,y_s,\theta)$; $\tilde{\mathscr{A}}_{\theta,\tilde y,h}$ the infinitesimal generator of $\tilde Z_t$; $\mathscr{B}_{\theta,\tilde y,h} \equiv \mathscr{A}_{\theta,\tilde y,h} - \tilde{\mathscr{A}}_{\theta,\tilde y,h}$; and $\upsilon(y,t;s,\theta) \equiv \mu_Y(y,t;\theta) - \mu_Y(y,s;\theta)$. If $\tilde Z_t$ is stationary, then denote Q_θ the stationary probability measure associated with the infinitesimal generator

$\tilde{\mathscr{A}}_{\theta,\tilde{y},h}$. This decomposition of operator $\mathscr{A}_{\theta,\tilde{y},h}$ provides an effective tool for studying consistency of feasible estimators, especially for some particular functional forms, by imposing conditions that make both $\tilde{\mathscr{A}}_{\theta,\tilde{y},h}$ and $\mathscr{B}_{\theta,\tilde{y},h}$ "nice".

Proposition 4. *Let Assumptions* 1–5 *hold with* $\alpha > 1$ *in Assumption* 3(3). *Then* $\tilde{Z}_t$ *is stationary. Let* Q_θ *be absolutely continuous with respect to* P_θ. *The feasible estimator* $\hat{\theta}^{[m]}$ *is consistent if for any compact subset* $A_Y \subset D_Y$ *and any positive* T *there exists a constant* $C = C(A_Y, T)$ *such that* $|\partial^{i+j}\upsilon(y,t;s,\theta)/\partial y^i \partial t^j| \leqslant C^{i+j}$ *for all* $(y,t,s,\theta) \in A_Y \times [0,T]^2 \times \Theta$ *and all non-negative integers* i,j.

Proof. See Appendix A.

3. Performance of the Hermite approximation

In this section, we study the Hermite approximation through numerical simulations. For several time-inhomogeneous models with known analytical transition densities, we show that our method provides much faster (in terms of CPU time) and more accurate approximation of the true transition density than several other popular numerical techniques. We also show that the parameter estimators obtained by maximizing the likelihood function approximated using Hermite expansions with just a few first terms, behave very similarly to the true maximum likelihood estimators and outperform the estimators obtained from the Euler approximation. This suggests that the superior performance of Ait-Sahalia's (2002) method for time-homogeneous diffusions documented in Jensen and Poulsen (1999) and Ait-Sahalia (1999, 2002) also extends to time-inhomogeneous models.

3.1. Transition density of several time-inhomogeneous models

In our numerical simulations, we consider several concrete time-inhomogeneous diffusions whose transition densities are known in closed form. Although we confine ourselves to these four models when testing our method, one could equally use other specifications.

1. The Extended Ornstein–Uhlenbeck model (EOU hereafter)

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \sigma_0 \mathrm{e}^{\sigma_1 t}\,\mathrm{d}W_t$$

with Gaussian transition density

$$p_X(x, s+h|x_s, s;\theta) = \mathrm{N}\left(\mathrm{e}^{-ah}x_s, \frac{\sigma_0^2}{2(a+\sigma_1)}(\mathrm{e}^{2\sigma_1(s+h)} - \mathrm{e}^{2(\sigma_1 s - ah)})\right)$$

and Hermite approximation

$$p_X^{[m]}(x, s+h|x_s, s;\theta) = \frac{\phi(z)\mathrm{e}^{-\sigma_1(s+h)}}{\sigma_0\sqrt{h}} \sum_{k=0}^{2m} \beta_k^{[m]} H_k(z), \quad z = \frac{\mathrm{e}^{-\sigma_1 h}x - x_s}{\sigma_0 \mathrm{e}^{\sigma_1 s}\sqrt{h}}. \tag{24}$$

2. The Extended Black–Scholes model (EBS)

$$\mathrm{d}X_t = aX_t\mathrm{d}t + \sigma_0 \mathrm{e}^{\sigma_1 t} X_t \mathrm{d}W_t$$

with Gaussian transition density $p_X(x, s+h|x_s, s; \theta) = \mathrm{N}(\log x_s + ah - \frac{1}{2}V, V)x^{-1}$, where $V = \frac{1}{2}\sigma_0^2\sigma_1^{-1}(\mathrm{e}^{2\sigma_1(s+h)} - \mathrm{e}^{2\sigma_1 s})$ and Hermite approximation

$$p_X^{[m]}(x, s+h|x_s, s; \theta) = \frac{\phi(z)\mathrm{e}^{-\sigma_1(s+h)}}{\sigma_0 x\sqrt{h}} \sum_{k=0}^{2m} \beta_k^{[m]} H_k(z), \quad z = \frac{\mathrm{e}^{-\sigma_1 h}\log x - \log x_s}{\sigma_0 \mathrm{e}^{\sigma_1 s}\sqrt{h}}.$$

3. The Hull–White model (HW)

$$\mathrm{d}X_t = a(b + ct - X_t)\mathrm{d}t + \sigma_0 \mathrm{e}^{\sigma_1 t}\mathrm{d}W_t$$

with Gaussian transition density $p_X(x, s+h|x_s, s; \theta) = \mathrm{N}(M, V)$ with mean and variance

$$M = \mathrm{e}^{-ah}x_s + ch + (1 - \mathrm{e}^{-ah})\left(b + cs - \frac{c}{a}\right),$$

$$V = \frac{\sigma_0^2}{2(a+\sigma_1)}(\mathrm{e}^{2\sigma_1(s+h)} - \mathrm{e}^{2(\sigma_1 s - ah)})$$

and Hermite approximation (24).

4. The Extended Cox–Ingersoll–Ross model (ECIR)

$$\mathrm{d}X_t = a(b_t - X_t)\,\mathrm{d}t + \sigma_0 \mathrm{e}^{\sigma_1 t}\sqrt{X_t}\,\mathrm{d}W_t$$

where $b_t = (\sigma_0^2 d/4a)\mathrm{e}^{2\sigma_1 t}$ and d is a positive constant, with transition density (see Maghsoodi, 1996)

$$p_X(x, s+h|x_s, s; \theta) = \tfrac{1}{2}\, G\mathrm{e}^{-1/2(\lambda + Gx)}\left(\frac{Gx}{\lambda}\right)^{(d-2)/4} I_{d/2-1}(\sqrt{\lambda G x})$$

where $\lambda = x_s \upsilon$, $G = \mathrm{e}^{ah}\upsilon$, $\upsilon = (8\sigma_1/\sigma_0^2)\mathrm{e}^{-ah}(\mathrm{e}^{2\sigma_1(s+h)} - \mathrm{e}^{2\sigma_1 s})$, and $I_\varrho(\cdot)$ is the modified Bessel function of the first kind of order ϱ. Its Hermite approximation is

$$p_X^{[m]}(x, s+h|x_s, s; \theta) = \frac{\phi(z)\mathrm{e}^{-\sigma_1(s+h)}}{\sigma_0\sqrt{hx}} \sum_{k=0}^{2m} \beta_k^{[m]} H_k(z), \quad z = 2\,\frac{\mathrm{e}^{-\sigma_1 h}\sqrt{x} - \sqrt{x_s}}{\sigma_0 \mathrm{e}^{\sigma_1 s}\sqrt{h}}.$$

3.2. *Accuracy of transition density approximation*

The analytical solution for the transition density for each of the four above models provides a clear benchmark for studying the performance of the Hermite approximation. We compare the speed[9] and accuracy of our method with some widely used numerical procedures for density approximation. Specifically, we consider the Euler scheme,

[9] All computations are conducted using Fortran95 on a cluster of 8 Dell PowerEdge servers, each with four Intel Pentium® III Xeon 550 MHz processors.

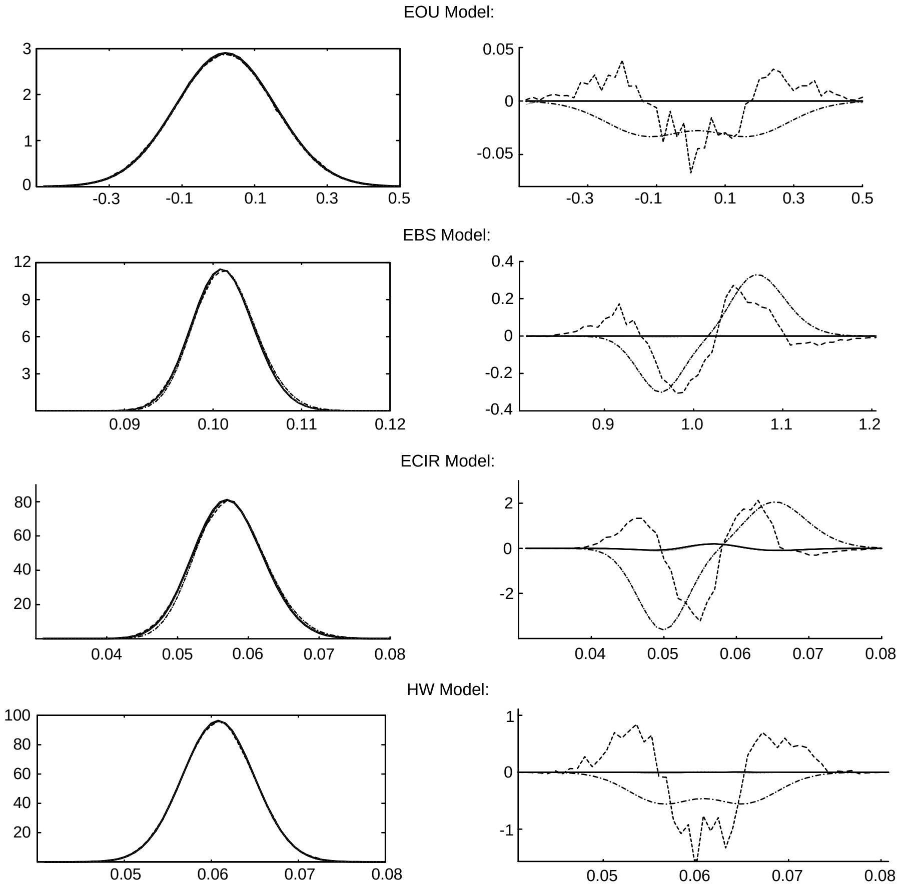

Fig. 1. Comparison of true transition density and its approximations. On the left panel, for each model we plot the true transition density $p_X(x,h|x_0,0;\theta)$ (thick solid line), its Hermite approximation (solid line), as well as approximations by the Euler scheme (dash–dot line), Monte Carlo simulations with the Sobol's quasi-random number generator (dash–dash line), the binomial model of Nelson and Ramaswamy (1990) (solid gray line), and the Crank–Nicholson scheme for the forward Kolmogorov equation (dot–dot line). The right panel plots deviations of these approximations from the true transition density using the same line style for each approximation method. CPU time can be read from the rightmost point of the corresponding curve in Fig. 2. For each model, the true parameter values θ and the initial value x_0 of the diffusion process X_t are given in Table 1.

Monte Carlo simulations, the binomial model of Nelson and Ramaswamy (1990), and the Crank–Nicholson scheme for the forward Kolmogorov equation. In Appendix B, we provide details on how each method is implemented.

For each of the four models, Fig. 1 plots the true transition density and its approximations using the Hermite and the other four approaches. Since these approximations are close to the true density, we also plot their deviations from the true density. From

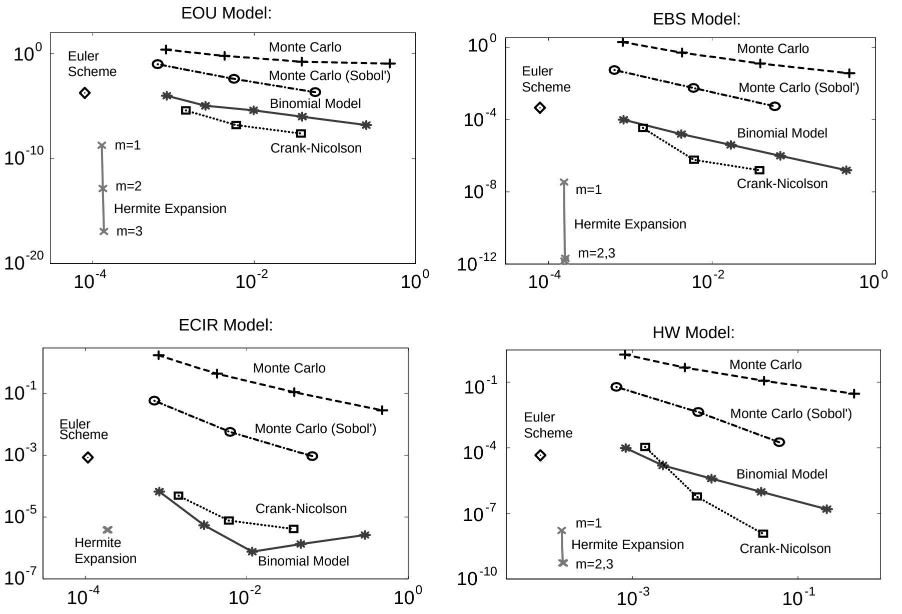

Fig. 2. Accuracy and speed of transition density approximation. The horizontal axis plots CPU time for computation of each approximation of the true transition density. The vertical axis plots the average error over 2000 simulated weakly observations, $\sum_{i=1}^{2000} \varepsilon_{(i+1)/52,i/52}(x_0)/2000$. We consider the approximation of the true transition density by Hermite approach ($\times$), the Crank–Nicholson scheme for the forward Kolmogorov equation ($\square$), the binomial model of Nelson and Ramaswamy (1990) ($*$), the Euler scheme ($\diamond$), Monte Carlo simulations with pseudo-random number generator ($+$), and Monte Carlo simulations with the Sobol's quasi-random number generator ($\circ$) for the EOU, EBS, ECIR and HW models. For each model, the true parameter values θ and the initial value x_0 of the diffusion process X_t are given in Table 1. See Appendix B for further details.

all the approximations shown in Fig. 1, the Hermite approach is the most accurate and, except for the Euler scheme, by far the least time-consuming. Specifically, all the approximations plotted in Fig. 1 correspond to the rightmost points on the respective graphs in Fig. 2, i.e. they are the most time-consuming among those depicted in Fig. 2 for each method.

We introduce as a formal measure for the approximation error of the transition density, the normalized version of the mean squared distance between the true transition density $p_X(x,t|x_s,s)$ and its approximation $\tilde{p}_X(x,t|x_s,s)$, as considered in Ait-Sahalia (1996b): [10]

$$\varepsilon_{t,s}(x_s) = \frac{\int_{D_X} (\tilde{p}_X(x,t|x_s,s) - p_X(x,t|x_s,s))^2 \, p_X(x,t|x_s,s)\,\mathrm{d}x}{\int_{D_X} (p_X(x,t|x_s,s))^2 \, p_X(x,t|x_s,s)\,\mathrm{d}x}.$$

[10] While we only present the results using $\varepsilon_{t,s}$ as a measure of the accuracy of transition density approximation below, we also considered other measures and obtained very similar findings.

This measure puts more weight on the discrepancies between the approximated and the true density for the values that the true process is more likely to take. It has the nice property that its possible values range from 0 for zero discrepancy to 1 for maximum discrepancy between the true density and its approximation.

Fig. 2 summarizes the performance of our method and the other four methods in terms of the accuracy of approximation and the speed of computation. The approximation errors are the averages over 2,000 simulated weekly observations with the same initial value x_0 for each time period, that is $\frac{1}{2000}\sum_{i=1}^{2000}\varepsilon_{(i+1)h,ih}(x_0)$, where $h=\frac{1}{52}$. It is clearly seen from the graphs that our method in general delivers more accurate approximation than the other four methods for a comparable computational time. The next best performing methods are the Crank–Nicholson scheme for the forward Kolmogorov equation and the binomial model of Nelson and Ramaswamy (1990). Monte Carlo simulations using the pseudo-random generator perform the worst, while using the Sobol's quasi-random generator provides significant improvement. Note that while Monte Carlo simulations perform poorly in our setting, they will likely outperform the other three methods for multi-dimensional models, especially in terms of speed. [11]

For each of the four models, Fig. 3 plots the approximation error $\varepsilon_{t+h,t}(x)$ for Hermite expansion of order $m=3$ as a function of time t and initial value x of the diffusion process X_t, using simulated weekly observations ($h=\frac{1}{52}$). The approximation errors are extremely small for all the four models over a wide range of initial values and over time. This evidence further reveals the accuracy of our method on a global basis. In particular, Fig. 3 can be useful in assessing the approximation error along different sample paths of X_t, since each sample path of X_t is just a curve on the plain (x,t).

3.3. Accuracy of parameter estimation

Having considered the accuracy of our method for transition density approximation, in this section we investigate its performance for parameter estimation.

Using the true transition density, we simulate 10,000 paths of 2,000 weekly observations of X_t for each of the four time-inhomogeneous models. Table 1 provides the true parameter values θ and the initial value x_0 of the diffusion process X_t for each model. For each sample path we obtain parameter estimates by maximizing: the true likelihood function ($\hat{\theta}$); the likelihood function approximated by the Euler scheme ($\hat{\theta}_{\mathrm{ea}}$); and by Hermite expansions of order $m=1,2,$ and 3 ($\hat{\theta}^{[1]},\hat{\theta}^{[2]}$, and $\hat{\theta}^{[3]}$). The results are summarized in Tables 1, 2, and Fig. 4.

From Table 1, we conclude that average values of the parameter estimates obtained by the Hermite approach are significantly closer to the corresponding average values of the true maximum likelihood estimates than average values of the parameter estimates

[11] For the ECIR model, we do not see a clear pattern of convergence of Hermite expansion with different orders in the graph, although numerical values of the approximation error are smaller for higher order expansions. This could be due to the fact that numerical approximations are involved in computing the true transition density of the ECIR model.

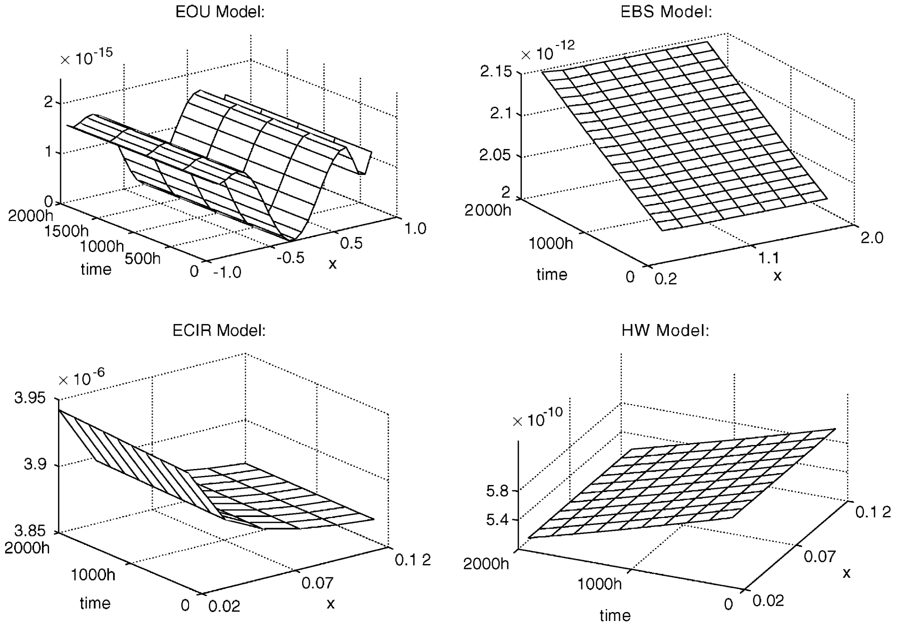

Fig. 3. Distribution of approximation errors of the Hermite approach. For the Hermite expansion of order $m=3$, this figure plots approximation error $\varepsilon_{t+h,t}(x)$ as a function of x and time t, for simulated weakly observations ($h=1/52$). For each model, the true parameter values θ are given in Table 1.

obtained using the Euler approximation. The standard deviations of $\hat{\theta}^{[1]}-\hat{\theta}, \hat{\theta}^{[2]}-\hat{\theta}$, $\hat{\theta}^{[3]}-\hat{\theta}$, $\hat{\theta}_{\mathrm{ea}}-\hat{\theta}$ are in general much smaller than the standard deviation of $\hat{\theta}$ itself. Table 2 shows that the correlation structure of parameter estimates obtained from the Hermite approach resemble that of the true maximum likelihood estimators.

We also examine histograms of all parameter estimates. For all four models, the Hermite approach produces histograms of parameter estimates that are very close to the corresponding histograms of the true maximum likelihood estimators. To illustrate this closeness, Fig. 4 presents histograms of $\hat{\theta}$ and $\hat{\theta}^{[3]}$ for the EOU model over 10,000 simulations of weekly observations.

Thus, numerical results suggest that just a few first terms in the Hermite expansion provide a close approximation to the true maximum likelihood method, and generally do so significantly better then the Euler approximation.

4. Conclusion

In this paper, we extend the maximum likelihood estimation method of Ait-Sahalia (2002) for time-homogeneous diffusions to time-inhomogeneous ones. We derive a closed-form approximation of the likelihood function for discretely sampled

Table 1
Comparison of parameter estimates

θ	$\hat\theta-\theta$	$\hat\theta^{[3]}-\hat\theta$	$\hat\theta^{[2]}-\hat\theta$	$\hat\theta^{[1]}-\hat\theta$	$\hat\theta_{ea}-\hat\theta$
EOU	($x_0=0,\ a=1,\ \sigma_0=1,\ \sigma_1=-0.001$)				
a	0.1055 (0.364)	−0.00002 (0.00008)	0.0001 (0.002)	−0.0132 (0.018)	0.0394 (0.372)
σ_0	−0.0002 (0.045)	−0.0000004 (0.0001)	−0.000003 (0.0002)	0.00006 (0.0004)	−0.0125 (0.004)
σ_1	0.0001 (0.004)	0.0000001 (0.00001)	0.0000001 (0.00002)	−0.0000003 (0.00003)	0.0002 (0.0002)
EBS	($x_0=1,\ a=0.2,\ \sigma_0=0.25,\ \sigma_1=-0.001$)				
a	0.0052 (0.007)	0.000003 (0.0004)	0.000005 (0.0004)	−0.00004 (0.003)	−0.0640 (0.012)
σ_0	−0.0002 (0.008)	0.000002 (0.0001)	0.000002 (0.0001)	−0.000004 (0.0001)	−0.0007 (0.001)
σ_1	0.00002 (0.001)	−0.0000004 (0.00001)	−0.0000004 (0.00001)	−0.0007 (0.0001)	0.00001 (0.0002)
ECIR	($x_0=0.05625,\ d=5,\ a=0.5,\ \sigma_0=0.15,\ \sigma_1=0.001$)				
a	0.0200 (0.11)	−0.00001 (0.0003)	−0.000005 (0.0003)	−0.0007 (0.0038)	−0.0033 (0.0023)
σ_0	−0.0005 (0.005)	0.0004 (0.0001)	0.0004 (0.00002)	0.0004 (0.0002)	−0.0001 (0.001)
σ_1	0.0001 (0.001)	−0.0000004 (0.00002)	−0.0000004 (0.00002)	−0.000001 (0.0001)	−0.00001 (0.0002)
HW	($x_0=0.06,\ a=0.5,\ b=0.06,\ c=0.001,\ \sigma_0=0.03,\ \sigma_1=0.001$)				
a	0.0132 (0.065)	−0.0037 (0.015)	−0.0004 (0.039)	−0.0850 (0.041)	−0.1609 (0.025)
b	0.0024 (0.0238)	−0.0002 (0.0026)	0.0001 (0.0040)	−0.0066 (0.0057)	−0.0475 (0.0086)
c	−0.0001 (0.001)	−0.00002 (0.00001)	−0.00001 (0.0001)	0.0003 (0.0003)	0.0018 (0.0004)
σ_0	0.00001 (0.001)	0.00001 (0.0001)	0.00001 (0.0001)	0.0007 (0.0003)	0.0002 (0.0001)
σ_1	0.00003 (0.001)	0.000004 (0.0001)	−0.00001 (0.0003)	−0.0009 (0.0003)	−0.0001 (0.0001)

Note: 10,000 paths of 2,000 weekly observations of the diffusion process X_t were simulated from its true transition density for the EOU, EBS, ECIR and HW models. For each model, this table provides the true parameter values θ and the initial value x_0 of X_t. For each sample path we obtain the true maximum likelihood estimator $\hat\theta$ as well as the maximum likelihood estimators for the Euler approximation ($\hat\theta_{ea}$) and for the Hermite expansion of order $m=1,2,3(\hat\theta^{[1]},\hat\theta^{[2]},\hat\theta^{[3]})$. This table also shows average values and standard deviations (in brackets) for $\hat\theta-\theta$, $\hat\theta^{[3]}-\hat\theta$, $\hat\theta^{[2]}-\hat\theta$, $\hat\theta^{[1]}-\hat\theta$, $\hat\theta_{ea}-\hat\theta$.

Table 2
Correlation of parameter estimates

	θ_1	θ_2	True MLE	Hermite			Euler
				$m=3$	$m=2$	$m=1$	
EOU	a	σ_0	0.062	0.062	0.062	0.062	0.029
	a	σ_1	0.005	0.005	0.005	0.006	−0.018
	σ_0	σ_1	−0.867	−0.867	−0.867	−0.867	−0.869
EBS	a	σ_0	−0.382	−0.383	−0.383	−0.373	0.113
	a	σ_1	0.362	0.363	0.363	0.354	−0.093
	σ_0	σ_1	−0.871	−0.871	−0.871	−0.871	−0.874
ECIR	a	σ_0	0.057	0.074	0.074	0.073	0.067
	a	σ_1	0.005	0.004	0.004	0.005	−0.009
	σ_0	σ_1	−0.866	−0.865	−0.865	−0.865	−0.865
HW	a	b	0.572	0.574	0.556	0.602	0.619
	a	c	−0.510	−0.511	−0.495	−0.536	−0.548
	a	σ_0	0.034	0.056	−0.050	0.172	−0.097
	a	σ_1	−0.019	−0.038	0.057	−0.137	0.090
	b	c	−0.906	−0.906	−0.907	−0.911	−0.916
	b	σ_0	0.016	0.024	−0.001	0.053	−0.040
	b	σ_1	0.004	−0.001	0.021	−0.026	0.038
	c	σ_0	−0.013	−0.022	0.001	−0.047	0.025
	c	σ_1	−0.004	0.001	−0.019	0.023	−0.021
	σ_0	σ_1	−0.869	−0.868	−0.874	−0.859	−0.870

Note: 10,000 paths of 2,000 weekly observations of X_t were simulated from its true transition density for the EOU, EBS, ECIR and HW models, and parameter estimates were obtained for each sample path. The table provides correlation of parameter estimates obtained by maximizing the true likelihood function, its Euler approximation and its approximation by our method with Hermite expansion of order $m = 1, 2, 3$. The correlation matrix for the true maximum likelihood estimators is also given for reference. For each model, the true parameter values θ and the initial value x_0 of the diffusion process X_t are given in Table 1.

time-inhomogeneous diffusions which converges to the true likelihood function. We also prove that maximizing the approximated likelihood function yields parameter estimates that converge to the true model parameters. Monte Carlo simulations for several financial models reveal that our method largely outperforms other widely used numerical procedures in approximating the likelihood function. Furthermore, parameter estimates produced by our method are very close to the true maximum likelihood estimators, and superior to estimates obtained from the Euler approximation.

Time inhomogeneity is an important feature of the data-generating processes for many important economic variables. Our method can be applied by empirical researchers to estimate and test some specific time-inhomogeneous models. Beyond their methodological contribution to financial econometrics literature, our results can also be applied to derivative pricing when the underlying asset prices follow time-inhomogeneous diffusions. This, of course, is left for future research.

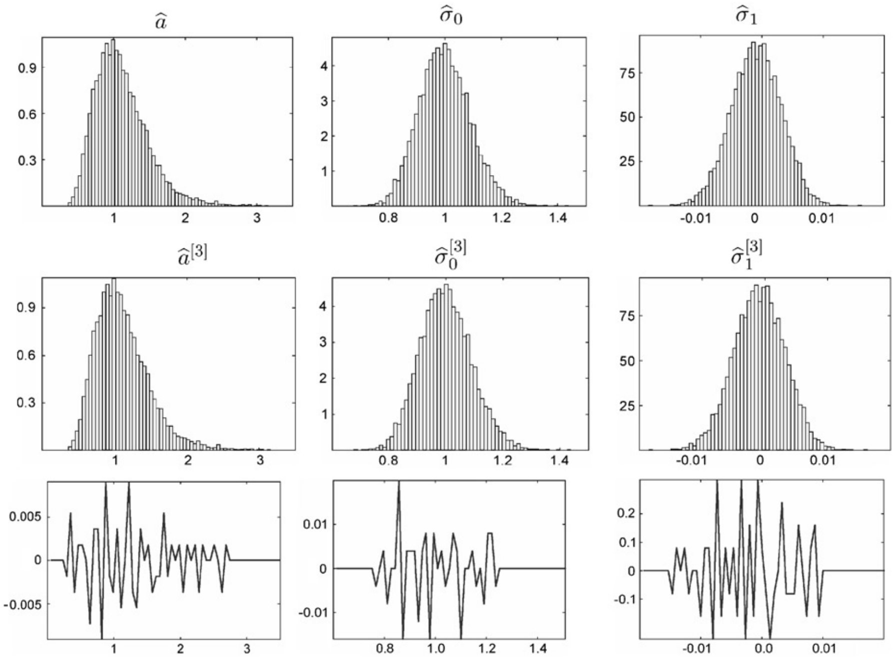

Fig. 4. Distribution of $\hat{\theta}$, and $\hat{\theta}^{[3]}$ for the EOU model. For 10,000 simulated paths of 2,000 weekly observations of EOU model with $x_0 = 0$ and parameters $a = 1$, $\sigma_0 = 1$, $\sigma_1 = -0.001$, this figure plots histograms for the true maximum likelihood estimates $\hat{a}$, $\hat{\sigma}_0$, $\hat{\sigma}_1$, and the parameter estimates $\hat{a}^{[3]}$, $\hat{\sigma}_0^{[3]}$, $\hat{\sigma}_1^{[3]}$ obtained by our method, as well as the difference between histograms for $\hat{a}$ and $\hat{a}^{[3]}$, $\hat{\sigma}_0$ and $\hat{\sigma}_0^{[3]}$, $\hat{\sigma}_1$ and $\hat{\sigma}_1^{[3]}$.

Acknowledgements

We would like to thank the two anonymous referees, the associate editor, and the co-editor, Professor Cheng Hsiao, for their comments and suggestions that greatly improved the paper.

Appendix A. Proofs of the main results

Proof of Proposition 2. We first prove that if Y_t exists then it cannot attain with probability one the boundaries of its domain D_Y in a finite time. Take any positive T and $\theta \in \Theta$. Let $\bar{m}_Y(y;T,\theta)$ and $\underline{m}_Y(y;T,\theta)$ denote the maximum and the minimum of $\mu_Y(y,t;\theta)$ over $t \in [0,T]$, and $\bar{Y}_{T,t}$ and $\underline{Y}_{T,t}$ be diffusion processes with the unit diffusion coefficient and the drift coefficient $\bar{m}_Y(y;T;\theta) + \varepsilon/T$ and $\underline{m}_Y(y;T;\theta) - \varepsilon/T$, respectively, where ε is a positive constant. Let $\bar{Y}_{T,t}$ and $\underline{Y}_{T,t}$ satisfy the same initial conditions as Y_t. Lemma 2 of Gihman and Skorohod (1972, Chapter 4.16) implies that $\bar{Y}_{T,t} > Y_{T,t} > \underline{Y}_{T,t}$ for any $t \in (0,T]$. Under Assumptions 1–3, Proposition 1 of Ait-Sahalia (2002) applies to $\bar{Y}_{T,t}$ and $\underline{Y}_{T,t}$. In particular, with probability one, $\bar{Y}_{T,t}$

and $\underline{Y}_{T,t}$ do not attain the boundaries of D_Y (that is, 0 and ∞, or $-\infty$ and ∞) at any $t \in (0, T]$. Since $\bar{Y}_{T,t} > Y_t > \underline{Y}_t$ for any $t \in (0, T]$, and T can be made arbitrarily large, we conclude that Y_t does not attain with probability one the boundaries of D_Y in a finite time.

Existence of a weak solution of (2), as well as applicability of the Girsanov theorem follows from modification of Proposition 5.3.6 of Karatzas and Shreve (1991) which holds under more general condition given in their Corollary 3.5.13. This condition in turn holds under our Assumptions 1–3 since Y_t cannot explode with probability one in a finite time. Uniqueness in a probability law follows from Proposition 5.3.10 of Karatzas and Shreve (1991).

Let W_t be a standard Brownian motion, $B_{s+uh}^{y_s \to y} \equiv y_s(1-u) + yu + (W_{s+uh} - uW_{s+h})$ be a Brownian Bridge from y_s to y on $t \in [s, s+h]$. Fix a positive T, any $\tilde{y} \in D_Y$ and assume below that both s and $s+h$ belong to interval $[0, T]$ and that $y_s = \tilde{y}$. Let $\bar{\mu}_Y(y; T)$ and $\underline{\mu}_Y(y; T)$ denote the maximum and minimum of $\mu_Y(y, t; \theta)$ over (t, θ) in the compact $[0, T] \times \Theta$.

We first consider $D_Y = (-\infty, \infty)$. We partition it as $D_Y = (-\infty, -\eta) \cup [-\eta, \eta] \cup (\eta, \infty)$ where η is defined in Assumption 3(2). We take $\eta = \eta(T, y_s)$ large enough that the interval $(-\eta, \eta)$ contains y_s. Depending on value of y, we get the following cases: (i) $-\eta < y_s < \eta \leqslant y$, (ii) $-\eta < y_s \leqslant y < \eta$, (iii) $-\eta < y \leqslant y_s < \eta$, and (iv) $y \leqslant -\eta < y_s < \eta$. Recall that Girsanov theorem applies. Moreover, it can be written in the form given in Chapter 3.13 of Gihman and Skorohod (similar arguments apply):

$$p_Y(y, s+h | y_s, s; \theta) = \frac{1}{\sqrt{2\pi h}} \mathrm{e}^{-(y-y_s)^2/2h + \int_{y_s}^{y} \mu_Y(s+h, u; \theta)\, \mathrm{d}u} \mathrm{E}_s[\mathrm{e}^{h \int_0^1 g(B_{s+uh}^{y_s \to y}, s+uh; \theta)\, \mathrm{d}u}], \tag{A.1}$$

where g is defined in (16).

Using Assumption 3(2) we get for case (i):

$$\begin{aligned}\int_{y_s}^{y} \mu_Y(s+h, u; \theta)\, \mathrm{d}u &\leqslant \int_{y_s}^{y} \bar{\mu}_Y(u; T)\, \mathrm{d}u = \int_{y_s}^{\eta} \bar{\mu}_Y(u; T)\, \mathrm{d}u + \int_{\eta}^{y} \bar{\mu}_Y(u; T)\, \mathrm{d}u \\ &\leqslant (\eta - y_s) \max_{u \in [y_s, \eta]} \bar{\mu}_Y(u; T) \\ &\quad + \frac{K}{2}((y - y_s)^2 + 2|y - y_s| + |y_s|).\end{aligned}$$

Proceeding similarly for the other three cases, we get that for some $\tilde{C}_1 = \tilde{C}_1(T, y_s) \geqslant 0$ and $\tilde{C}_2 = \tilde{C}_2(T, y_s) \geqslant 0$,

$$\int_{y_s}^{y} \mu_Y(s+h, u; \theta)\, \mathrm{d}u \leqslant \frac{K}{2}(y - y_s)^2 + \tilde{C}_1 |y - y_s| + \tilde{C}_2. \tag{A.2}$$

Under Assumptions 1 and 3, there exists a constant C_g such that the last term in (A.1) is bounded from above by $\mathrm{e}^{C_g h}$ uniformly in $(y, y_s, s, \theta) \in D_Y^2 \times [0, T] \times \Theta$. This bound along with (A.2) gives the upper bound for (A.1). From (A.1) it also follows that p_Y is strictly positive. Substituting $z = h^{-1/2}(y - y_s)$ and $p_Z(z, t|y_s, s;\theta) = h^{1/2} p_Y(h^{1/2} z + y_s|y_s, s;\theta)$ produces the desired inequality (17) for p_Z.

We differentiate (A.1) to obtain $(\partial/\partial y)p_Y$:

$$\frac{\partial p_Y}{\partial y} = \left(\mu_Y(s+h, y;\theta) - \frac{y-y_s}{h}\right) p_Y + \frac{1}{\sqrt{2\pi h}}$$

$$\mathrm{e}^{-(y-y_s)^2/2h+\int_{y_s}^{y}\mu_Y(s+h,u;\theta)\,\mathrm{d}u}\mathrm{E}_s[G\mathrm{e}^{h\int_0^1 g\,\mathrm{d}u}], \tag{A.3}$$

where $p_Y \equiv p_Y(y, s+h|y_s, s;\theta)$, $g \equiv g(B_{s+uh}^{y_s\to y}, s+uh;\theta)$, $G \equiv h\int_0^1(\partial/\partial y)g(B_{s+uh}^{y_s\to y}, s+uh;\theta)\,\mathrm{d}u$. Note that we can differentiate under expectation operator, since we show below that the last term in (A.3) is uniformly bounded.

By Assumption 3(2) μ_Y can grow at most polynomially when $y\to\infty$. This implies that function G can grow at most polynomially in y, and hence, at most exponentially y when $y\to\infty$.[12] Together with Assumption 3(1) this implies that there exist positive $a = a(T, y_s)$, $C_a = C_a(T, y_s)$ and $C_g = C_g(T, y_s)$ such that for any $y\in(\eta,\infty)$

$$|\mathrm{E}_s[G\mathrm{e}^{h\int_0^1 g\,\mathrm{d}u}]| \leqslant \mathrm{E}_s\left[C_a\int_0^1 u\mathrm{e}^{|y_s(1-u)+yu+(W_{s+uh}-uW_{s+h})|a+C_g h}\,\mathrm{d}u\right]$$

$$\leqslant C_a\int_0^1 u\mathrm{e}^{(|y_s|+|y-y_s|u)a+C_g h}\mathrm{E}_s[\mathrm{e}^{|W_{s+uh}-uW_{s+h}|}]\,\mathrm{d}u \leqslant \tilde{C}_a\mathrm{e}^{a|y-y_s|},$$

where $\tilde{C}_a \equiv C_a\mathrm{e}^{a|y_s|+(C_g+\frac{1}{8})T}$, since $\mathrm{E}_s[\mathrm{e}^{|W_{s+uh}-uW_{s+h}|}] = \mathrm{e}^{hu(1-u)/2}$. Applying similar arguments to $\mu_Y(s+h, y;\theta) - (y-y_s)/h$, we see that both terms in (A.3) are uniformly bounded. Changing variable from y to z, we conclude that (18) holds for $\partial p_Z/\partial z$.

Let $D_Y = (0,\infty)$. Then the expectation operator $\mathrm{E}_s[\,\cdot\,]$ is additionally conditioned by $h<\tau$, where τ is the first time when $B_{s+uh}^{y_s\to y}$ hits zero. Under Assumption 3(2,3) $D_Y = (0,\eta_0)\cup[\eta_0,\eta]\cup[\eta,\infty)$. We choose $\eta_0 = \eta_0(T, y_s)$ small enough and $\eta = \eta(T, y_s)$ large enough that interval (η_0, η) contains y_s. There are four different cases depending on location of y: (i) $\eta_0 < y_s < \eta \leqslant y$, (ii) $\eta_0 < y_s \leqslant y < \eta$, (iii) $\eta_0 < y < y_s < \eta$, (iv) $y \leqslant \eta_0 < y_s < \eta$. Formally changing $-\eta$ for η_0 in the above arguments, we conclude that (17) and (18) hold for all $y\in(\eta_0,\infty)$. We find for case (iv)

$$\int_{y_s}^{y}\mu_Y(s+h,u;\theta)\,\mathrm{d}u \leqslant -\min_{u\in[\eta_0,y_s]}\underline{\mu}_Y(u;T)(y_s-\eta_0)$$

$$-\int_{y}^{\eta_0}\mu_Y(s+h,u;\theta)\,\mathrm{d}u. \tag{A.4}$$

From (A.4) with small enough positive η_0 and Assumption 3(3) it follows that $\mathrm{e}^{\int_{y_s}^{y}\mu_Y(s+h,u;\theta)\,\mathrm{d}u}$ decreases at least exponentially in $1/y$ as $y\searrow 0$ if $\alpha>1$ and $k>0$,

[12] Function $f(y)$ is said to grow at most exponentially in y when $y\to\infty(y\to-\infty,\ y\to 0)$ if there exists a non-negative constant a such that $f(y)\mathrm{e}^{-ay}$ $(f(y)\mathrm{e}^{ay},\ f(y)\mathrm{e}^{a/|y|})$ is bounded in some vicinity of $y=\infty$ $(y=-\infty,\ y=0)$.

and at least as y^k as $y \searrow 0$ if $\alpha = 1$ and $k \geqslant 1$. Thus, $\lim_{y\to 0} p_Z = 0$. Both terms in (A.3) are uniformly bounded, since by definition (16) of g and Assumption 3, G is at most polynomially growing when $y \searrow 0$. This implies that $\partial p_Z/\partial z$ is also uniformly bounded. Moreover, p_Z is strictly positive (see the arguments above). We conclude that inequalities (17), (18) hold for all $z \in D_z$. $\square$

Proof of Theorem 1. Let $K = 0$, if $\mu_Y \leqslant 0$ near $y = +\infty$ and $\mu_Y \geqslant 0$, near $y = -\infty$, and $K = \Gamma$ otherwise. Fix a positive constant T. If $K = \Gamma$, then assume that $\bar{h} < 1/2\Gamma$, otherwise $\bar{h} = \infty$. Fix $h \in (0, \bar{h})$ and T_0 such that $T_0 > T + h$. We assume throughout the proof that $s \in [0, T]$. Let A_X be a compact subset of D_X, and $A_{Y,s}$ be a family of compact sets indexed by s such that $\{y = \varphi(x; s, \theta) | x \in A_X,\ \theta \in \Theta\} \subset A_{Y,s} \subset D_Y$.

We first prove the convergence of $p_Z^{(J)}$. Note that $D_Z = (\underline{z}, \bar{z})$, where $\bar{z} = \infty$, and $\underline{z} = -\infty$ for $D_Y = (-\infty, \infty)$ and $\underline{z} = -y_s/\sqrt{h}$ for $D_Y = (0, \infty)$. In the last case we define $p_Z(z, t | y_s, s; \theta) \equiv 0$ for $z \in (-\infty, -y_s/\sqrt{h}]$. We integrate (12) by parts, and then use Propositions 1(1) and 2 to get that $\beta_k(t, y_s, s) = 1/(k+1) \int_{-\infty}^{\infty} H_{k+1}(w) \partial p_Z(w, t | y_s, s; \theta)/\partial w \, dw$. We use below in the proof shortcut β_k for $\beta_k(t, y_s, s)$. Note that

$$\phi(z) \sum_{k=0}^{J} |\beta_k H_k(z)| < \varpi \left(1 + \frac{z^{5/2}}{2^{5/4}}\right) e^{-z^2/4} \sum_{k=0}^{J} \frac{\sqrt{k!}}{\sqrt[4]{k}} |\beta_k| \frac{\sqrt{k+1}}{\sqrt{k+1}}$$

$$\leqslant \frac{1}{2} \varpi \left(1 + \frac{z^{5/2}}{2^{5/4}}\right) e^{z^2/4} \left(\sum_{k=0}^{J} \frac{1}{(k+1)\sqrt{k}} + \sum_{k=0}^{J} (k+1)! \beta_k^2 \right).$$

Bound (10) on Hermite polynomials and the upper bound (18) on $\partial p_Z/\partial z$ in Proposition 2 imply that

$$\sum_{k=0}^{J} (k+1)! \beta_k^2 \leqslant C_0(T_0, y) \varpi^2 \int_{-\infty}^{\infty} \left(1 + \frac{w^{5/2}}{2^{5/4}}\right) e^{-w^2(1-2hK)/2 + 2|w| C_1(T_0, y_s)} \, dw \sum_{k=0}^{J} \frac{1}{(k+1)!}.$$

Since (i) $1 - 2hK > 0$ by above choice of h and $\bar{h}$, (ii) $e^{-z^2/2}(1 + (z^{5/2}/2^{5/4}) e^{z^2/4})$ is bounded on D_Z, and (iii) the series $\sum_{k=1}^{\infty} 1/(k+1)\sqrt{k}$ and $\sum_{k=0}^{\infty} 1/(k+1)!$ converge, we conclude that there exists $\hat{C}_0 = \hat{C}_0(T_0, y_s)$ such that $\phi(z) \sum_{k=0}^{J} |\beta_k H_k(z)| < \hat{C}_0$ for any $(z, \theta, s, y_s) \in D_Z \times \Theta \times [0, T_0] \times A_{Y,s}$.

We further get that if σ is non-degenerate, then

$$\frac{\phi(z)}{\sigma(x, s+h; \theta)\sqrt{h}} \sum_{k=0}^{J} |\beta_k H_k(z)| < \frac{\hat{C}_0}{\sigma(x, s+h; \theta)\sqrt{h}} < \frac{\hat{C}_0}{c_{T_0}\sqrt{h}}, \tag{A.5}$$

where c_{T_0} is defined in Assumption 2(1) and $z = \varphi(x; s+h, \theta) - y_s/\sqrt{h}$. Recall that $p_X^{(J)}$ is defined by (13)–(15). Inequality (A.5) says that $p_X^{(J)}$ absolutely converges (and thus converges) for any $(x, x_s, \theta, s) \in D_X^2 \times \Theta \times [0; T_0]$. By (A.5) this convergence is uniform in $(x, x_s, \theta, s) \in D_X \times A_X \times \Theta \times [0; T_0]$ for any positive T_0 and $h \in (0, \bar{h})$. If σ is degenerate, Assumption 2(2) similarly implies that convergence in x is uniform only over $[\varepsilon, \infty)$ for any positive constant ε.

Finally, $p_Z^{(J)}(z, t|y_s, s; \theta)$ converges to $p_Z(z, t|y_s, s; \theta)$ as $J \to \infty$, since

$$\frac{1}{k!}\int_{-\infty}^{\infty} H_k(z)\left(p_Z - \phi(z)\sum_{j=0}^{\infty}\beta_j H_j(z)\right)\mathrm{d}z$$

$$=\frac{1}{k!}\int_{-\infty}^{\infty}(H_k(z)p_Z - \phi(z)\beta_k H_k^2(z))\,\mathrm{d}z$$

$$=\frac{1}{k!}\int_{-\infty}^{\infty} H_k(z)p_Z\,\mathrm{d}z - \beta_k = 0, \quad k = 0, 1, 2, \ldots,$$

where $p_Z = p_Z(z, t|y_s, s; \theta)$. Hence, $p_X^{(J)}(x, t|x_s, s; \theta)$ also converges to $p_X(x, t|x_s, s; \theta)$. □

Proof of Theorem 2. (1) We assume below that $\theta_0 \in \Theta$, $h \in (0, \bar{h})$, $T > nh$, where n is a fixed sample size with observation interval h, and that $s \in [0, T - h]$. Let $R_{X,J}(x, s + h|x_s; s, \Theta) \equiv \sup_{\theta \in \Theta}|p_X(x, s + h|x_s, s; \theta) - p_X^{(J)}(x, s + h|x_s, s; \theta)|$.

We first show that

$$\lim_{J\to\infty}\Pr[R_{X,J}(x, s + h|x_s; s, \Theta) > \varepsilon|\theta = \theta_0] = 0 \quad \text{for any } \varepsilon > 0. \tag{A.6}$$

Using Chebyshev's inequality, we prove (A.6) by demonstrating that for any $\delta > 0$ there exists $\tilde{J}$ such that for any $J \geqslant \tilde{J}$,

$$\mathrm{E}_{\theta_0}[R_{X,J}^m(X_{s+h}, s + h|X_s, s; \Theta)] < \delta, \quad m = 1, 2. \tag{A.7}$$

Since the left-hand side of (A.7) equals to $\int_{x_s \in D_X}\int_{x \in D_X} R_{X,J}^m(x, s + h|x_s, s; \Theta)p_X(x, s + h|x_s, s; \theta_0)\,\mathrm{d}x\,\mathrm{d}\Pr[X_s \leqslant x_s|\theta = \theta_0]$, and $\int_{x_s \in D_X}\mathrm{d}\Pr[X_s \leqslant x_s|\theta = \theta_0] = 1$, inequality (A.7) holds if for any $y_s \in D_Y$,

$$\int_{x \in D_X} R_{X,J}^m(x, s + h|x_s, s; \Theta)p_X(x, s + h|x_s, s; \theta_0)\,\mathrm{d}x < \delta, \quad m = 1, 2. \tag{A.8}$$

If diffusion X_t is non-degenerate, then Theorem 1 implies that $R_{X,J}(x, s + h|x_s, s; \Theta)$ can be made arbitrarily small by picking a large enough J and thus for this case (A.8) is proved.

Now, let diffusion X_t be locally degenerate. We change the variable of integration in (A.8) from x to y and denote the integrand by $\Psi = \Psi(y; y_s s, h, \Theta, \theta_0)$. We use below shortcut $p_Y = p_Y(y, s + h|y_s, s; \theta_0)$. Assumption 2(2) and Theorem 1 imply that for any $\varepsilon > 0$ there exists an integer $\tilde{J}$ such that for all $J > \tilde{J}$ and some constants $\xi_0, \omega, c_{T,\xi_0} > 0$, and $\rho > 0$

$$\int_{y \in D_Y}\Psi\,\mathrm{d}y \leqslant \varepsilon c_{T,\xi_0}^{-1} + \varepsilon\int_{\varphi^{-1}(y; s+h, \theta) \in (0, \xi_0)} p_Y \times \sup_{\theta \in \Theta, s \in [0,T)}\sigma^{-m}(\varphi^{-1}(y; s + h, \theta), s; \theta)\,\mathrm{d}y. \tag{A.9}$$

Let $\rho \in (0,1)$. From Assumption 2(2) it follows that $y = \int_0^x \sigma^{-1}(u,t;\theta)\,\mathrm{d}u \leqslant \omega^{-1}(1-\rho)^{-1}x^{1-\rho}$, and $\sigma^{-1}(x,s;\theta) \leqslant \omega^{-1}x^{-\rho} \leqslant \omega^{-1}(\omega(1-\rho)y)^{-\rho/(1-\rho)}$ for any $x \in (0,\xi_0)$. Choosing ξ_0 such that $\omega^{-1}(1-\rho)^{-1}\xi_0^{1-\rho} < \eta_0$ where η_0 is defined in Assumption 3(3), one gets further from (A.9) that for a positive constant C and a positive $\hat{C} = \hat{C}(T, y_s)$,

$$\int_{y \in D_Y} \Psi\,\mathrm{d}y \leqslant \varepsilon c_{\xi_0}^{-1} + \varepsilon C \int_{y \in (0,\eta_0)} y^{-m\rho/(1-\rho)} p_Y\,\mathrm{d}y$$

$$\leqslant \varepsilon c_{\xi_0}^{-1} + \varepsilon C\hat{C} \int_{y \in (0,\eta_0)} y^{-m\rho/(1-\rho)} \mathrm{e}^{-k(s+h,\theta)\int_y^{\eta_0} w^{-\alpha(s+h,\theta)}\mathrm{d}w}\,\mathrm{d}y. \tag{A.10}$$

Recall that $m = 1, 2$, and that for $0 < \rho < 1$ either $\alpha > 1$ and $k > 0$, or $\alpha = 1$ and $k \geqslant 2\rho/(1-\rho)$. Thus, the integral on the right-hand side of inequality (A.10) converges. Similar arguments show that if $\rho \geqslant 1$ then the integral on the right-hand side of inequality (A.9) converges. Thus, inequality (A.8) holds, and the proof of (A.6) is finished.

Given (A.6) and continuity of logarithm, it follows that for any positive ε,

$$\log(p_X^{(J)}(X_{t_i}, t_i | X_{t_{i-1}}, t_{i-1}; \theta) \vee (\varepsilon/J)) \xrightarrow{\mathrm{p}} \log p_X(X_{t_i}, t_i | X_{t_{i-1}}, t_{i-1}; \theta)$$

uniformly in θ as $J \to \infty$. Therefore, for any fixed sample size n we have that $L_n^{(J)}(\theta) \xrightarrow{\mathrm{p}} L_n(\theta)$ uniformly in θ as $J \to \infty$. Since $L_n^{(J)}(\theta), L_n(\theta)$ and their derivatives are continuous in θ for all n and J, it follows that as $J \to \infty$, $\hat{\theta}_n^{(J)} \xrightarrow{\mathrm{p}} \theta_0$ under P_{θ_0}.

(2) We apply Taylor's formula to the score function:

$$L_n'(\hat{\theta}_n) - L_n'(\theta_0) = -L_n''(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) = H_n(\tilde{\theta}_n) i_n^{-1/2}(\theta_0) i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0).$$

Substitute $L_n'(\hat{\theta}_n) = 0$ and rearrange terms to obtain

$$i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) = \{i_n^{1/2}(\theta_0) H_n(\tilde{\theta}_n) i_n^{1/2}(\theta_0)\} i_n^{-1/2}(\theta_0) L_n'(\theta_0).$$

Finally, we apply Taylor's formula to $H_n(\tilde{\theta}_n)$, using Assumption 4 and definition of G_n:

$$i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) = i_n^{1/2}(\theta_0) H_n(\theta_0) L_n'(\theta_0) + \mathrm{o_p}(1)$$

$$= G_n^{-1}(\theta_0) i_n^{-1/2}(\theta_0) L_n'(\theta_0) + \mathrm{o_p}(1).$$

Under Assumptions 1–5, Theorem 1 of Basawa and Scott (1983, p. 34) implies that the log-likelihood ratio has LAMN structure, and that

$$(i_n^{-1/2}(\theta_0) L_n'(\theta_0), G_n(\theta_0)) \xrightarrow{\mathrm{d}} (G^{1/2}(\theta_0) \times \mathrm{N}(0, \mathbf{I}_\kappa), G(\theta_0)) \quad \text{under } \mathrm{P}_{\theta_0},$$

where $\mathrm{N}(0,\mathbf{I}_\kappa)$ is distributed independently of $G(\theta_0)$. Continuous Mapping Theorem (e.g., Hall and Heyde, 1980, Theorem A.3, p. 276) implies that $i_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{\mathrm{d}} G^{-1/2}(\theta_0) \times \mathrm{N}(0, \mathbf{I}_\kappa)$ under P_{θ_0}. By part (1) of Theorem 2, for any fixed n, $\hat{\theta}_n^{(J)}$ converges to $\hat{\theta}_n$ in probability under P_{θ_0}. This means that for any fixed $n, \varepsilon, \delta > 0$ there exists a large enough J_n such that $\Pr[|\hat{\theta}_n^{(J)} - \hat{\theta}_n| > \varepsilon i_n^{-1/2}(\theta_0) | \theta = \theta_0] \leqslant \delta$ componentwise for all $J \geqslant J_n$, i.e. $i_n^{1/2}(\theta_0)(\hat{\theta}_n^{(J)} - \hat{\theta}_n) = \mathrm{o_p}(1)$. Thus, $i_n^{1/2}(\theta_0)(\hat{\theta}_n^{(J_n)} - \theta_0) \xrightarrow{\mathrm{d}} G^{-1/2}(\theta_0) \times \mathrm{N}(0, \mathbf{I}_\kappa)$, under P_{θ_0}, and by Assumption 4, $\hat{\theta}_n^{(J_n)} \xrightarrow{\mathrm{p}} \theta_0$ under P_{θ_0}. □

Proof of Proposition 3. Let us drop subscripts h, y_s, θ for brevity. Note first that $C_0^\infty(\mathbf{R}) \subset \mathscr{D}(\mathscr{A})$ and for $f \in C_0^\infty(\mathbf{R})$,

$$(\mathscr{A} \circ f)(z,t) = \frac{\partial f}{\partial z}\, \mu_Z + \frac{\sigma_Z}{2}\, \frac{\partial^2 f}{\partial z^2} + \frac{\partial f}{\partial t}.$$

Since μ_Y is infinitely differentiable in y and t, and has at most polynomial growth near infinity boundaries, it follows that $\mathscr{A}^i \circ f \in C_0^\infty(\mathbf{R}) \subset \mathscr{D}(\mathscr{A})$ for any $i = 1,2,\ldots$ Hence, (i) formula (21) for $\mathscr{A}^i \circ f$ holds, (ii) the process $(\mathscr{A}^i \circ f)(Z_t,t) - (\mathscr{A}^i \circ f)(0,s) - \int_s^t [(\mathscr{A}^{i+1} \circ f)(Z_{\tau_1},\tau_1)]\,\mathrm{d}\tau_1$ is a local martingale for any $i = 0,1,2,\ldots,$ that is

$$\mathrm{E}_s[(\mathscr{A}^i \circ f)(Z_t,t)] = (\mathscr{A}^i \circ f)(0,s) + \mathrm{E}_s \int_s^t (\mathscr{A}^{i+1} \circ f)(Z_{\tau_1},\tau_1)\mathrm{d}\tau_1,$$
$$i = 0,1,2,\ldots\ . \tag{A.11}$$

Substituting out $\mathscr{A}^{i+1}$ in (A.11) successively for $i = 0,1,\ldots,m$, we obtain

$$\mathrm{E}_s[f(Z_{s+h})] = \sum_{i=0}^{m} (\mathscr{A}^i \circ f)(0,s)\frac{h^i}{i!} + \mathrm{E}_s \int_s^{s+h} \int_s^{\tau_1}$$
$$\cdots \int_s^{\tau_{m-1}} (\mathscr{A}^{m+1} \circ f)(Z_{\tau_m},\tau_m)\,\mathrm{d}\tau_m \cdots \mathrm{d}\tau_2\,\mathrm{d}\tau_1$$

for any positive h. Because $\mathscr{A}^{m+1} \circ f$ is continuous in the second argument, the last term can be rewritten as $\mathrm{E}_s[(\mathscr{A}^{m+1} \circ f)(Z_\xi,\xi)](\xi - s)^{m+1}/(m+1)!$ for some $\xi \in [s,s+h]$. □

Proof of Corollary 1. Proposition 4 of Ait-Sahalia (2002) applies. In particular, $\beta^{[m]} \to \beta$ as $m \to \infty$. By (15), (14), (13), (23) and Theorem 1, $p_X^{[m]} - p_X^{(2m)} \to 0$ as $m \to \infty$. Moreover, all the above convergence is uniform over the same sets as the convergence in Theorem 1. Applying Theorem 2 to $p_X^{[m]}$ instead of $p_X^{(J)}$ finishes the proof. □

Proof of Proposition 4. We have $(\tilde{\mathscr{A}}_{\theta,\tilde{y},h} \circ f)(z;s) = h^{-1/2}\mu_Y(\sqrt{h}z + \tilde{y},s;\theta)\partial f/\partial z + (2h)^{-1}\partial^2 f/\partial z^2$ and $(\mathscr{B}_{\theta,\tilde{y},h} \circ g)(z,t;s) = \upsilon(t;s,\theta)\partial g/\partial z + \partial g/\partial t$ for appropriate f,g. Then there exist functions $b_{r,i,k}(z,t;s,h,\tilde{y},\theta)$ such that $(\mathscr{B}^i_{\theta,\tilde{y},h} \circ H_k)(z,t;s) = \sum_{r=0}^k b_{r,i,k}(z,t;s,h,\tilde{y},\theta)H_k^{(r)}(z)$, where $H_k^{(r)}(z) \equiv \mathrm{d}^r H_k(z)/\mathrm{d}z^r$. Let $r_{I,j} \equiv R_{I-1}(H_{k,j},\mathscr{A}_{\theta,\tilde{y},h},Z_\xi,\xi;s,\theta,\tilde{y},h)$. Rearranging terms we get that

$$r_I \equiv \lim_{j\to\infty} r_{I,j} = \sum_{r=0}^{k}\sum_{i=0}^{I} b_{r,i,k}(z,t;s,h,\tilde{y},\theta)$$
$$\frac{h^i}{i!}\,(\tilde{\mathscr{A}}^{I-i}_{\theta,\tilde{y},h} \circ H_k^{(r)})(Z_\xi;s)\,\frac{h^{I-i}}{(I-i)!}. \tag{A.12}$$

The series $\sum_{l=0}^\infty \|(\tilde{\mathscr{A}}^l_{\theta,\tilde{y},h} \circ H_k^{(r)})(Z_\xi;s)\|_p h^l l!^{-1}$ converges (and the process $\tilde{Z}_t$ is stationary), since Proposition 4 of Ait-Sahalia (2002) applies to $\tilde{Z}_t$, and Q_θ is absolutely continuous with respect to P_θ. Recall that for any compact subset $A_Y \subset D_Y$ and any positive T there exists a constant $C = C(A_Y,T)$ such that $|\partial^{i+j}\upsilon(y,t;s,\theta)/\partial y^i\partial t^j| \leqslant C^{i+j}$ for all $(y,t,s,\theta) \in A_Y \times [0,T]^2 \times \Theta$ and all non-negative integers i,j. By recursive structure of $b_{r,i,k}$ it is straightforward to check that this condition implies that $b_{r,i,k}(z,t;s,h,\tilde{y},\theta) = O((kC)^i)$ as $i \to \infty$. Hence, $\lim_{i\to\infty} |b_{r,i,k}(z,t;s,h,\tilde{y},\theta)| h^i i!^{-1} = 0$. By (A.12), $\lim_{I\to\infty} \|r_I\|_p = 0$, since the series $\sum_{l=0}^{\infty} \|(\tilde{\mathscr{A}}^l_{\theta,\tilde{y},h} \circ H_k^{(r)})(Z_\xi;s)\|_p h^l l!^{-1}$ converges and $\lim_{i\to\infty} |b_{r,i,k}(z,t;s,h,\tilde{y},\theta)| h^i i!^{-1} = 0$. Under Assumptions 1–3 and by definition of $H_{k,j}$, we get that (i) both $r_{I,j}$ and p_Z exponentially decreases as z approaches the boundaries of D_Z and (ii) $r_{I,j}$ approximates r_I uniformly on any compact subset of D_Z. Hence, Lebesgue's dominated convergence theorem applies: $\lim_{j\to\infty} \|r_{I,j}\|_p = \|r_I\|_p$. Letting I tend to infinity, we get that $\lim_{I,j\to\infty} \|r_{I,j}\|_p = 0$. Note that all the above limits are uniform in $(y,t,s,\theta) \in A_Y \times [0,T]^2 \times \Theta$ and in h over any compact subset of $(0,\infty)$. The consistency of the feasible estimator now follows by the same arguments as in the proof of Corollary 1. □

Appendix B. Alternative approximation methods

B.1. Euler scheme

If time interval h is small, then diffusion equation (2) can be approximated by the Euler scheme as

$$X_{s+h} = x_s + \mu(x_s,s;\theta)h + \sigma(x_s,s;\theta)W_h. \tag{B.1}$$

Transition density for (B.1) is Gaussian

$$p_X^{\mathrm{EA}}(x,s+h|x_s,s;\theta) = \mathrm{N}(x_s + \mu(x_s,s;\theta)h, \sigma^2(x_s,s;\theta)h).$$

Similar to our method, the Euler scheme delivers closed-form approximation for the transition density and for the likelihood function. This makes parameter estimates easy to obtain. The Euler scheme is locally first-order accurate in h (e.g., Kloeden et al., 1994). Unfortunately, for financial data, the time interval h between observations is often not small enough for the Euler scheme to provide an adequate approximation.

B.2. Monte Carlo simulation

In order to approximate the transition density p_X by Monte Carlo simulation, we discretize further the time interval $[s,s+h]$ by taking $s_i = ih/N + s$, $i = 0,1,2,\ldots,N$ for some positive integer N. For $i = 1,2,\ldots,N$, we simulate

$$X_{s_i} = X_{s_{i-1}} + \mu(X_{s_{i-1}},s_{i-1};\theta)\frac{h}{N} + \sigma(X_{s_{i-1}},s_{i-1};\theta)\sqrt{\frac{h}{N}}\,\varepsilon_i \tag{B.2}$$

if the Euler scheme is used, and

$$\begin{aligned} X_{s_i} &= X_{s_{i-1}} + \mu(X_{s_{i-1}},s_{i-1};\theta)\frac{h}{N} + \sigma(X_{s_{i-1}},s_{i-1};\theta)\sqrt{\frac{h}{N}}\varepsilon_i \\ &\quad + \frac{1}{2}\sigma^2(X_{s_{i-1}},s_{i-1};\theta)\frac{h}{N}(\varepsilon_i^2 - 1) \end{aligned} \tag{B.3}$$

if the Milstein scheme is used. Values for ε_i are drawn independently from standard normal distribution. Transition density p_X is approximated through the histogram for X_{s_N} after performing K simulations. While the Euler scheme is locally first-order accurate in h/N, the Milstein scheme's local order of accuracy is $(h/N)^2$. Higher accuracy schemes are also available in the literature (e.g., Kloeden et al., 1994), but the computational time increases significantly, and it might be faster to use the Euler or the Milstein scheme with smaller time steps instead.

We use two different approaches to simulate the white noise ε_i. We first get approximations of uniformly distributed random sequences using either Sobol's sequence of quasi-random numbers or pseudo-random number generator *ran*3 (Press et al., 1992). Then we apply the Box–Muller method to get approximations for white noise ε_i. Note that Monte Carlo simulations produce an asymptotic error of order $K^{-1/2}$, if one uses a pseudo-random number generator. Using Sobol' numbers increases asymptotic accuracy of Monte Carlo simulations (e.g., Niederreiter, 1992).

The results of Monte Carlo simulations with pseudo-random generator and the Milstein scheme for (N,K) equal to $(1,8),(2,32),(4,128),(8,512)$, as well as Monte Carlo simulations with Sobol's sequence of quasi-random numbers and the Euler scheme for (N,K) equal to $(1,100),(1,1000),(1,10000)$ are summarized in Fig. 2. Histograms are constructed by dividing on 50 equal segment of interval $(-0.5,0.5)$ for the EOU model, $(0.85,1.15)$ for the EBS model, $(0.03,0.08)$ for the ECIR model, and $(0.04,0.08)$ for the HW model.

B.3. Binomial model

Another approximation method we use is the binomial model developed by Nelson and Ramaswamy (1990) for time-inhomogeneous diffusions. Time interval h is divided into N equal time steps of length $\delta = h/N$. It is convenient to construct the binomial tree for Y_t. Let $y_0^0 = Y_s$, then for any $i = 0,1,2,\dots,N$ and any $j = 0,1,\dots,i+1$, a typical building block for the binomial tree consists of node $(s+i\delta, y_i^j)$, and two branches originating at this node and ending in its two successor nodes $(s+(i+1)\delta, y_{i+1}^j)$ and $(s+(i+1)\delta, y_{i+1}^{j+1})$. Let $q_i^j = \Pr(Y_{s+(i+1)\delta} = y_{i+1}^{j+1} | Y_{s+i\delta} = y_i^j)$ and assume that $\Pr(Y_{s+(i+1)\delta} = y_{i+1}^j | Y_{s+i\delta} = y_i^j) = 1 - q_i^j$. Such a binomial tree is called recombining or computationally simple, since any two adjacent nodes $(s+i\delta, y_i^j)$ and $(s+i\delta, y_i^{j+1})$ have one common successor node $(s+(i+1)\delta, y_{i+1}^{j+1})$. To guarantee recombining, we follow Nelson and Ramaswamy (1990) and take [13]

$$y_{i+1}^j = y_i^j - \sigma(y_i^j, s+i\delta;\theta)\sqrt{h}, \qquad y_{i+1}^{j+1} = y_i^j - \sigma(y_i^j, s+i\delta;\theta)\sqrt{h},$$

$$q_i^j = \frac{h\mu(y_i^j, s+i\delta;\theta) + y_i^j - y_{i+1}^j}{y_{i+1}^{j+1} - y_{i+1}^j}.$$

[13] When σ hits zero, this algorithm needs to be slightly modified to insure its uniformity [see Nelson and Ramaswamy (1990) on how this should be done].

For approximation of the true transition density $p_X(x,s+h|x_0,s;\theta)$, the inverse of transformation (4) is applied to $y_N^0, y_N^1, \ldots, y_N^{N+1}$ as well as their probabilities are computed. Such binomial model weakly converges to the true diffusion X_t for diffusion models commonly used in finance (see Nelson and Ramaswamy, 1990).

For the four models considered in this paper, Fig. 2 shows accuracy and computational time for approximation of transition densities by binomial model with $h=1/52$ and $N=20,50,100,200,500$.

B.4. Numerical solution to the forward Kolmogorov equation

Finally, we obtain the approximation $p_X^{\mathrm{CN}}(x,s+h|x_s,s;\theta)$ by numerically solving the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu(x,t;\theta)p) + \frac{1}{2}\frac{\partial}{\partial x^2}(\sigma^2(x,t;\theta)p) \tag{B.4}$$

using the Crank–Nicholson finite difference scheme (e.g., Strikwerda, 1989). Because this method cannot accommodate the initial condition $p_X(x,s|x_s,s;\theta)=\delta(x-x_s)$, where $\delta(\cdot)$ is the Dirac-δ function, we use an approximate initial condition $p_X(x,\hat{s}|x_s,s;\theta) \approx p_X^{\mathrm{EA}}(x,\hat{s}|x_s,s;\theta)$ at some $\hat{s}$ such that $s<\hat{s} \ll s+h$, where p_X^{EA} is transition density for the process X_t^{EA} which is determined by the Euler scheme

$$X_{\hat{s}}^{\mathrm{EA}} = x_s + \mu(x_s,s;\theta)(\hat{s}-s) + \sigma(x_s,s;\theta)W_{\hat{s}-s}.$$

In order to apply the above method, boundary conditions for p_X should also be specified. Let $\hat{x}_+$ be such that $\hat{x}_+ > x_s$ and the true transition density is approximately zero at $\hat{x}_+$. Take $\hat{x}_- = 0$ when $D_X=(0,\infty)$. If $D_X=(-\infty,\infty)$ then pick some $\hat{x}_-$ such that $\hat{x}_- < x_s$ and the true transition density is approximately zero at $\hat{x}_-$. Note that under assumptions of Proposition 2, the process X_t does not attain boundaries of its domain in finite time with probability one, and thus $\hat{x}_+$ and $\hat{x}_-$ exist. We then impose $p_X(\hat{x}_\pm, s=h|x_s,s;\theta)=0$ as boundary condition for p_X.

Let $I(J)$ refer to the number of equal intervals of length $\varkappa(\tau)$ on which the interval $[\hat{x}_-,\hat{x}_+]$ ($[\hat{s},T]$) is divided. The Crank–Nicholson scheme is then specified by the following tridiagonal system of linear equations for $p_i^j \equiv p_X^{\mathrm{CN}}(x^i,s^{j+1}|x_s,s^j;\theta)$

$$\begin{aligned}(-\alpha_i^j - \beta_i^j)p_{i-1}^{j+1} &+ (1+\lambda_i^j)p_i^{j+1} + (\alpha_i^j - \beta_i^j)p_{i+1}^{j+1} \\ &= (\alpha_i^j + \beta_i^j)p_{i-1}^j + (1-\lambda_i^j)p_i^j + (\beta_i^j - \alpha_i^j)p_{i+1}^j, \quad i=1,2,\ldots,I-1,\end{aligned}$$

where $\alpha_i^j = \tau\mu(x^i,s^j;\theta)/4\varkappa$, $\beta_i^j = \tau\sigma^2(x^i,s^j;\theta)/4\varkappa^2$, $\lambda_i^j = 2\beta_i^j - (\tau/2)\partial\mu(x^i,s^j;\theta)/\partial x + (\tau/4)\partial^2\sigma^2(x^i,s^j;\theta)/\partial x^2$, and $\varkappa = (\hat{x}_+ - \hat{x}_-)/I$, $\tau=(T-\hat{s})/J$.

The above equation is solved successively for $j=1,2,\ldots,J$, using the boundary conditions $p_0^j = p_I^j = 0$, for $j=1,2,\ldots,J$, and the initial condition $p_i^1 = p_X^{\mathrm{EA}}(x^i,\hat{s}|x_s,s;\theta)$. The Crank–Nicholson scheme is locally second-order accurate in both $\varkappa$ and τ, unconditionally stable, and thus convergent (e.g., Strikwerda, 1989).

For the four models considered in this paper, Fig. 2 shows accuracy and computational time for approximation of transition densities by the Crank–Nicholson scheme with $h=1/52$, $\hat{s}=s+0.01h$, $T=h$ and (I,J) equal to $(50,20),(100,40),(250,100)$. We choose $(\hat{x}_-,\hat{x}_+)$ equal to $(-0.5,0.5)$ for the EOU model, $(0.85,1.15)$ for the EBS model, $(0.03, 0.08)$ for the ECIR model, and $(0.04, 0.08)$ for the HW model. At these boundary points $\hat{x}_-$ and $\hat{x}_+$ the true transition density for x_0 and θ defined in Table 1 is approximately zero.

References

Ait-Sahalia, Y., 1996a. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Ait-Sahalia, Y., 1996b. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Ait-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

Ait-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely sampled diffusions: a closed-form approach. Econometrica 70, 223–262.

Basawa, I.V., Scott, D.J., 1983. Asymptotic Optimal Inference for Non-Ergodic Models. In: Lecture Notes in Statistics, Vol. 17. Springer, New York.

Black, F., Karasinski, P., 1991. Bond and option pricing when short rates are lognormal. Financial Analysts Journal 47, 52–59.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Black, F., Derman, E., Toy, W., 1990. A one-factor model of interest rates and its application to treasury bond options. Financial Analysts Journal 46, 33–39.

Campbell, J., Lo, A., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–578.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

Derman, E., Kani, I., 1994. The Volatility Smile and its Implied Tree, Quantitative Strategies Research Notes. Goldman Sachs, New York.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Dupire, B., 1994. Pricing with a smile. Risk 7, 18–20.

Gihman, I.I., Skorohod, A.V., 1972. Stochastic Differential Equations. Springer, Heidelberg.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.

Hansen, L.P., Scheinkman, J.A., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63, 767–804.

Hansen, L.P., Scheinkman, J.A., Touzi, N., 1995. Identification of scalar diffusions using eigenvectors. Journal of Econometrics 86, 1–32.

Heath, D.C., Jarrow, R.A., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuation. Econometrica 60, 77–105.

Hong, Y., Li, H., 2002. Nonparametric specification testing of continuous-time models with application to spot interest rates. Working Paper, Cornell University.

Hull, J., White, A., 1990. Pricing interest-rate derivative securities. Review of Financial Studies 3, 573–592.

Hull, J., White, A., 1993. One-factor interest-rate models and valuation of interest-rate derivative securities. Journal of Financial and Quantitative Analysis 28, 235–254.

Jacobs, B., Levy, K., 1988. Calendar anomalies: abnormal returns at calendar turning points. Financial Analysts Journal 44, 28–39.

Jeganathan, P., 1995. Some aspects of asymptotic theory with applications to time series models. Econometric Theory 11, 818–887.

Jensen, B., Poulsen, R., 1999. A comparison of approximation techniques for transition densities of diffusion processes. Manuscript, Aarhus University.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.
Kloeden, P.E., Platen, E., Schurz, H., 1994. Numerical Solution of SDE through Computer Experiments. Springer, New York.
Lo, A.W., 1988. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econometric Theory 4, 231–247.
Maghsoodi, Y., 1996. Solution of the extended CIR term structure and bond optional valuation. Mathematical Finance 6, 89–109.
Melino, A., 1994. Estimation of continuous-time models in finance. In: Sims, C. (Ed.), Advances in Econometrics: Sixth World Congress, Vol. 2. Cambridge University Press, Cambridge.
Nelson, D.B., Ramaswamy, K., 1990. Simple binomial processes as diffusion approximations in financial markets. Review of Financial Studies 3, 393–430.
Niederreiter, N., 1992. Random number generation and quasi-Monte Carlo methods. SIAM, Philadelphia.
Piazzesi, M., 2000. An econometric model of the yield curve with macroeconomic jump effects. Working Paper, UCLA.
Press, W., Teukolsky, S., Vetterling, W.T., Flannery, B., 1992. Numerical Recipes in Fortran. Cambridge University Press, Cambridge.
Rubinstein, M., 1994. Implied binomial trees. Journal of Finance 49, 771–818.
Sansone, G., 1991. Orthogonal Polynomials. Dover, New York.
Schaumburg, E., 2000. Maximum likelihood estimation of jump processes with applications to finance. Manuscript, Princeton University.
Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973–2002.
Stone, M.H., 1927. Development in Hermite polynomials. Annals of Mathematics 29, 1–13.
Strikwerda, J., 1989. Finite-Difference Schemes and Partial Differential Equations. Wadsworth, Belmont.
Sundaresan, S., 2001. Continuous-time methods in finance: a review and an assessment. Journal of Finance 55, 1569–1622.
Tauchen, G., 1997. New minimum chi-square methods in empirical finance. In: Kreps, D., Wallis, K. (Eds.), Advances in Econometrics: Seventh World Congress. Cambridge University Press, Cambridge.
Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.
Wong, E., 1964. The construction of a class of stationary Markov processes. In: Bellman, R. (Ed.), Stochastic Processes in Mathematical Physics and Engineering, Proceedings of Symposia in Applied Mathematics, Vol. 16. American Mathematical Society, Providence, RI, pp. 264–276.