Transition density and simulated likelihood estimation for time-inhomogeneous diffusions

João Nicolau[1]

May 2010

Instituto Superior de Economia e Gestão/Universidade Técnica de Lisboa

and CEMAPRE

Office: Quelhas 406

Postal address: ISEG, Rua do Quelhas 6, 1200-781 Lisboa, Portugal

e-mail: nicolau@iseg.utl.pt

Abstract

We propose a method to estimate the transition density of a non-linear time-inhomogeneous diffusion. Expressing the transition density as a functional of a Brownian bridge, allows us to estimate the density through Monte Carlo simulations with any level of precision. We show how these transition density estimates can be effectively used to estimate the parameters of the time-inhomogeneous diffusion and the conditional moments of the process. In this paper we prove that our method is asymptotically equivalent to the maximum likelihood estimator and more reliable than the closed-form approximation approach largely used in the literature.

Running head: Estimation for time-inhomogeneous diffusions

[1] This research was supported by Fundação para a Ciência e a Tecnologia - FCT (FEDER/POCI2010).

1 Introduction

Stochastic continuous-time processes, especially diffusion processes generated by stochastic differential equations (SDE) have been widely used in financial economics. However, all models involve unknown parameters or functions, which need to be estimated from observations of the process. The estimation of diffusion processes is, thus, a crucial step in all applications and, in particular, in applied finance.

The maximum likelihood (ML) estimator for diffusion processes based on discrete observations has the usual good properties (see Yoshida, 1992, Kessler, 1997, Prakasa Rao, 1999). Unfortunately, the transition (or conditional) densities of X, required to construct the likelihood function, are usually unknown. In time-homogeneous diffusion processes several alternatives to the ML have been proposed. For a survey see, for example, Iacus (2008), Aït-Sahalia (2002) and Durham and Gallant (2002).

Estimation of time-inhomogeneous diffusions has received less attention in the literature than time-homogeneous diffusions. However, time-inhomogeneous diffusions are extremely relevant for term structure of interest rates (see a survey in Brigo and Mercurio, 2006) and, in general, to model stochastic processes that depend explicitly on time due to seasonality, economic business cycles and monetary policy, among other reasons. For example, monetary policies may not be the same during recessions and expansions or may evolve over time to reflect different economic and financial conditions. Under the influence of these "exogenous" conditions, the infinitesimal coefficients should depend explicitly on X_t and t:

$$dX_t = a\left(t, X_t; \theta\right) dt + b\left(t, X_t; \theta\right) dW_t, \qquad X_0 = \zeta.$$

As usual θ denotes the vector of unknown parameter and W the standard Wiener process. The initial condition ζ can be a constant or a random variable with the same state space of X and independent of W_t, $t \geq 0$. As in the case of time-homogeneous diffusions, the transition density of X, $p_X\left(s, x_0, t, x\right) = \partial P\left[\left. X_t \leqq x \right| X_s = x_0\right] / \partial x$ (which we admit exists) is in general unknown.

Shoji and Ozaki (1998) propose a local linearization method which approximates a nonlinear stochastic differential equation using a discretrized linear stochastic differential equation. The solution of this stochastic differential equation has known distribution density, hence the maximum likelihood method may be used.

This method produces consistent estimators but it requires that the sampling interval Δ converges to zero at an appropriate rate. Egorov et al. (2003) propose a procedure to recover the transition based on the arguments of Aït-Sahalia (2002). Using Δ fixed, they obtain a closed-form approximation of the transition and, subsequently a closed-form of likelihood function for discretely sampled time-inhomogeneous diffusions that converges, under some conditions, to the true likelihood function. Although the estimates based on their method seemed to be extremely accurate, we demonstrate that this method can fail under certain circumstances.

We propose a method to estimate the transition density of a non-linear time-inhomogeneous diffusion assuming that the sampling interval Δ is constant. Expressing the transition density as a functional of a Brownian bridge, allows us to estimate the density through Monte Carlo simulations with any level of precision. The use of a Brownian bridge approximation is also presented in Beskos and Roberts (2005), for homogeneous diffusion. Our approach extends the work of Nicolau (2002) by now considering time-inhomogeneous diffusions and conditional moments estimation. We show how these transition density estimates can be effectively used to estimate the parameters of the time-inhomogeneous diffusion and the conditional moments of the process. We prove that the proposed estimator is asymptotically equivalent to the maximum likelihood estimator and can work in situations where the method of Egorov (2003) fails.

The rest of the paper is organized as follows. In section 2 we present the main results (density estimation, conditional moments estimation and simulated maximum likelihood estimation). In section 3 we illustrate the proposed methods and compare them with the method of Egorov et al. (2003). In section 4 we conclude with a short discussion.

2 Main Results

2.1 Density Estimation

Consider a diffusion process solution to the time-inhomogeneous stochastic differential equation

$$dX_t = a\left(t, X_t\right) dt + b\left(t, X_t\right) dW_t, \qquad X_0 = \zeta$$

Under some regularity conditions, stated below, the transition density of the time-inhomogeneous process X can be written as a functional of a Brownian bridge (see Gihman and Skorohod, 1972, §13) as follows:

$$p_X(s,x_0,t,x) = \frac{1}{\sqrt{2\pi\Delta}b(t,x)} \exp\left\{-\frac{(f(t,x)-f(s,x_0))^2}{2\Delta} + \bar{A}(t,f(t,x)) - \bar{A}(s,f(s,x_0))\right\} E[\psi] \tag{1}$$

where

$$\psi = \exp\left\{\Delta \int_0^1 \bar{B}\left(s+\lambda\Delta, (1-\lambda)f(s,x_0)+\lambda f(t,x)+\sqrt{\Delta}\eta(\lambda)\right)d\lambda\right\}$$

$$f(t,x) = \int^x b^{-1}(t,\xi)\,d\xi$$

$$\eta(t) = W_t - tW_1,\ 0 \le t \le 1,\ \text{is a Brownian Bridge}$$

$$\bar{a}(t,y) = -\int_0^{g(t,y)} \frac{b'_t(t,\xi)}{b^2(t,\xi)}d\xi + \frac{a(t,g(t,y))}{b(t,g(t,y))} - \frac{1}{2}b'_y(t,g(t,y)) \tag{2}$$

$$\bar{A}(t,y) = \int_0^y \bar{a}(t,u)\,du, \qquad \bar{B}(t,y) = -\frac{1}{2}\bar{a}^2(t,y) - \int_s^y \bar{a}'_t(t,\xi)\,d\xi - \frac{1}{2}\bar{a}'_y(t,y)$$

and $g:\mathbb{R}^2 \to \mathbb{R}$ is the inverse function of f, i.e. $f(t,g(t,x)) = x$ (the derivative of h with respect to x is denoted by h'_x). In the appendix we provide an alternative proof of Gihman and Skorohod's theorem based on the ideas of Dacunha-Castelle and Florens-Zmirou (1986). The advantage of this method over other Monte Carlo methods such as in Pedersen (1995) and Santa-Clara (1995), is that, in our case, the transition density is partially known in closed form, apart from the parameters, given the infinitesimal coefficients $a(t,x)$ and $b(t,x)$. Only the expression of $\mathrm{E}[\psi]$ is unknown in a closed form. We evaluate $E[\psi]$ through Monte Carlo simulations. Aït-Sahalia (1999, page 1368), had already suggested that $E[\psi]$ could be obtained through Monte Carlo simulation. As we will see, under some mild regularity conditions, it is straightforward to compute $\mathrm{E}[\psi]$ with any level of precision. In this way, the transition density can be taken as if it was known. This is confirmed by our simulations experiments.

Note that $\psi = \psi(\omega)$ depends on the *Brownian bridge* $\{\eta(\lambda) = \eta(\lambda,\omega), 0 \le \lambda \le 1\}$, and, thus, ψ is a random variable. To estimate $E[\psi]$ we propose the estimator

$$\begin{aligned} \hat{\psi} &= \frac{1}{S}\sum_{j=1}^{S}\psi_N(\omega_j), \qquad \{\omega_j; j \geqslant 1\} \text{ are i.i.d.} \\ \psi_N(\omega) &= \exp\left\{\frac{\Delta}{N}\sum_{i=0}^{N-1}\bar{B}\left(s+\frac{i}{N}\Delta, \left(1-\frac{i}{N}\right)f(s,x_0)+\frac{i}{N}f(t,x)+\sqrt{\Delta}\eta\left(\frac{i}{N}\right)\right)\right\}. \end{aligned} \tag{3}$$

To obtain $\hat{\psi}$ by simulation we proceed as follows:

1. Fix N and S (large enough).

2. Obtain $\psi_N(\omega_j)$ for each $j = 1, ..., S$.

   (a) For each j, simulate η independently at instants i/N, $i = 0, 1, ..., N$ through the formula $\eta(i/N) = W_{i/N} - \frac{i}{N}W_1$ where $W_{i/N} = \sum_{k=0}^{i} \sqrt{1/N}\varepsilon_k$ with $\varepsilon_0 = 0$ and $\{\varepsilon_k, k = 1, ..., N\}$ is a sequence of i.i.d. random variables with $N(0,1)$ distribution.

   (b) Calculate $\psi_N(\omega)$ using formula (3)

3. Calculate $\hat{\psi} = \frac{1}{S}\sum_{j=1}^{S} \psi_N(\omega_j)$.

From the estimate of $\mathrm{E}[\psi]$, we are able to evaluate $p_X(s, x_0, t, x)$. Therefore, our estimator is the following:

$$\hat{p}_X(s, x_0, t, x) = \frac{1}{\sqrt{2\pi\Delta}b(t,x)} \exp\left\{-\frac{(x-x_0)^2}{2\Delta}\right\} \exp\left\{\bar{A}(t,x) - \bar{A}(s,x_0)\right\}\hat{\psi}. \tag{4}$$

We illustrate the estimator $\hat{p}_X$ in section 3.1. Our main results are based on the following assumptions:

A1: $b(t,x) > 0$, $a'_x(t,x)$ and $b'_x(t,x)$ exist and are bounded in any compact subset of the state space of X; the derivatives $b''_{x^2}(t,x)$, $b''_{t^2}(t,x)$ and $b'_t(t,x)$ exist.

A2: $\bar{B}(t,x)$ satisfies

$$\lim_{|x|\to\infty} \frac{1}{1+x^2} \sup_{0\le t\le T} \bar{B}(t,x) \le 0.$$

A3: $a(t,x)$ and $b(t,x)$ have continuous derivatives up to 3rd order satisfying, for $\delta > 0$

$$\lim_{|x|\to\infty} \sup_{0\le t\le T} \left\{\left|\frac{\partial\bar{B}}{\partial t}(t,x)\right| + \left|\frac{\partial\bar{B}}{\partial x}(t,x)\right| + \left|\frac{\partial^2 B}{\partial x^2}(t,x)\right|\right\} e^{-\delta x^2} = 0$$

where

$$B(t,x) = -\frac{1}{2}a^2(t,y) - \frac{1}{2}a'_y(t,y) - \int^x a'_t(t,\xi)\, d\xi. \tag{5}$$

These conditions are very weak and are easily satisfied by most stochastic differential equations used in economics and finance. For example, the infinitesimal coefficients $a(t,x) = e^t\left(e^x + x + x^2\right)$, $b(t,x) = e^t(1+x)$ satisfy the above assumptions.

We assume that the stochastic differential equation has a weak solution for all $\zeta \in \mathbb{R}$ and $\theta \in \Theta$ and that the solutions are unique in law. Conditions that ensure these conditions can be found in Rogers and Williams (1987). Sufficient conditions are the local Lipschitz and growth conditions for each $\theta \in \Theta$.

Finally, for purposes of estimation we make the additional assumption:

A4: The infinitesimal coefficients are twice continuously differentiable in $\theta \in \Theta$. The true parameter vector θ_0 belongs to the interior of Θ. Let $l_k(\theta) = \log\left(p_X\left((k-1)\Delta, X_{(k-1)\Delta}, k\Delta, X_{k\Delta}; \theta\right)\right)$ and $i_n(\theta) = diag\left(\sum_{k=1}^n \mathrm{E}\left[l_k'(\theta)\, l_k'(\theta)^T\right]\right)$ (the superscripts $'$ and T denote differentiation to θ, and transposition. We assume: $i_n^{-1}(\theta) \xrightarrow{a.s.} 0$ 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 $\left\| i_n^{1/2}(\theta_0)(\theta - \theta_0) \right\| \leq \varepsilon$ for some $\varepsilon > 0$.

This assumption guarantees the existence of the maximum likelihood estimator and its usual properties, with fixed sampling interval Δ (see Egorov et al., 2003):

Theorem 2.1 *Assume A1-A2 and $t > s$. Then $\hat{p}_X(s, x_0, t, x) \xrightarrow{p} p_X(s, x_0, t, x)$ as $N \to +\infty$ and $S \to +\infty$.*

Let $\Xi = \{(s, x_0, t, x) : (x_0, x) \in I \text{ and } p_X(s, x_0, t, x) \text{ is finite}\} \subset \mathbb{R}^4$ where I is the state space of X.

Theorem 2.2 *Under assumptions A1-A3 and $t > s$, $\hat{p}_X(s, x_0, t, x)$ converges uniformly in probability on Ξ to $p_X(s, x_0, t, x)$ when $N \to +\infty$ and $S \to +\infty$.*

These theorems are essential in order to justify how the proposed estimator $\hat{p}_X$ can be used to estimate conditional moments and to derive a simulated maximum likelihood estimator.

2.2 Conditional Moments Estimation

Uniform convergence $\hat{p}_X(s, x_0, t, x)$, established in theorem 2.2, allows us to consider conditional moments estimation according to the following result.

Theorem 2.3 *Suppose that $f(t, \xi)$ is continuous in ξ and $\int_a^b f(t, \xi)\, p_X(s, x_0, t, \xi)\, d\xi$ exists, where a and b are constants. Consider $\xi_i = a + i(b-a)/M$ and $\Delta\xi_i = (b-a)/M$. Under the conditions of theorem 2.2 we have*

$$\sum_{i=1}^{M} f(t, \xi_i)\, \hat{p}_X(s, x_0, t, \xi_i)\, \Delta\xi_i \xrightarrow{p} \int_{\xi_0}^{\xi_1} f(t, \xi)\, p_X(s, x_0, t, \xi)\, d\xi$$

uniformly on Ξ *as* $M \to \infty$, $N \to \infty$, $S \to \infty$.

Uniform convergence is the right criterion to use, since $\sum_{i=1}^{M} f(t, \xi_i) \hat{p}_X(s, x_0, t, \xi_i) \Delta\xi_i$ requires convergence at each ξ of an uncountable set of points in the interval $[\xi_0, \xi_1]$. Theorem 2.3 deals with finite limits of integration. To estimate the improper integral $\mathrm{E}[f(t, X_t)| X_s = x_0] = \int_{-\infty}^{\infty} f(t, \xi) p_X(s, x_0, t, \xi) d\xi$, assuming $\mathrm{E}[|f(t, X_t)|| X_s] < \infty$, we simply truncate the lower and upper limit values by suitable finite constants ξ_0 and ξ_1 such that $\int_{\xi_0}^{\xi_1} f(t, \xi) p_X(s, x_0, t, \xi) d\xi \approx \int_{-\infty}^{\infty} f(t, \xi) p_X(s, x_0, t, \xi) d\xi$. If the constants ξ_0 and ξ_1 are suitably defined, this procedure should not involve errors since $f(t, \xi) p_X(s, x_0, t, \xi)$ converges rapidly to zero, as $\mathrm{E}[|f(t, X_t)|| X_s]$ is finite (in other words, if ξ_0 and ξ_1 are suitably defined then $\int_{\{x: x \notin [\xi_0, \xi_1]\}} f(t, \xi) p_X(s, x_0, t, \xi) = 0$). We illustrate this method of estimation in section 3.2.

2.3 The Simulated Maximum Likelihood Estimator

We now analyze the case where the infinitesimal coefficients a and b depend on some unknown parameter vector θ. Assumption A4 guarantees the existence of the maximum likelihood estimator (MLE) and its usual properties, with fixed sampling interval Δ (hence, the asymptotics of the MLE is based on a large sample scheme, i.e. $n\Delta \to \infty$).

The simulated maximum likelihood (SML) estimator is defined as $\hat{\theta}_{n,S,N} = \arg\max \hat{L}_n(\theta)$ where

$$\hat{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log \hat{p}_X\left((i-1)\Delta, X_{(i-1)\Delta}, i\Delta, X_{i\Delta}; \theta, S, N\right).$$

From theorem 2.2 we have, in probability,

$$\begin{aligned} & \lim_{S,N \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \log \hat{p}_X\left((i-1)\Delta, X_{(i-1)\Delta}, i\Delta, X_{i\Delta}; \theta, S, N\right) \\ = \; & \frac{1}{n} \sum_{i=1}^{n} \log \lim_{S,N \to +\infty} \hat{p}_X\left((i-1)\Delta, X_{(i-1)\Delta}, i\Delta, X_{i\Delta}; \theta, S, N\right) \\ = \; & \frac{1}{n} \sum_{i=1}^{n} \log p_X\left((i-1)\Delta, X_{(i-1)\Delta}, i\Delta, X_{i\Delta}; \theta, S, N\right). \end{aligned}$$

Therefore, the optimization problem associated with the SML estimator is the same as that of the MLE. With the additional condition $\sqrt{n}/S \to 0$, the $\hat{\theta}_{n,S,N}$ estimator has the same asymptotic behavior as that of the MLE (see, Gouriéroux and Monfort, 1991).

When θ changes it is necessary to keep the same drawing $\{\varepsilon_k, k = 1, ..., N\}$, defined in the 3 steps procedure (section 2.1), in order that the difference $\left|\log \hat{p}_X(\cdot, \cdot, \cdot, \cdot; \theta) - \log \hat{p}_X\left(\cdot, \cdot, \cdot, \cdot; \theta'\right)\right|$ can only be attributed to the difference $\left|\theta - \theta'\right|$. The SML estimator is illustrated in section 3.3.

3 Numerical Experiments

3.1 Density Estimation

In this section we compare our estimator $\hat{p}_X(s, x_0, t, x)$ (equation (4)) with the estimator proposed by Egorov et al. (2003). This estimator is based on a Hermite approximation that leads to the following expression for the transition density:

$$p_X^{[m]}(s, x_0, t, x) = \frac{1}{\sqrt{\Delta} b(t, x)} p_Z^{[m]}\left(s, f(s, x_0), t, \frac{f(t, x) - f(s, x_0)}{\sqrt{\Delta}}\right)$$

where

$$p_Z(s, y_s, t, z) = \phi(z) \sum_{k=0}^{2m} \beta_k^{[m]}(s, t, y_s) H_k(z), \;\; \phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}, \;\; H_k(w) = \phi(w)^{-1} \frac{d^k \phi(w)}{dw^k}$$

and $\beta_k^{[m]}$ are the terms up to Δ^m in the expression $(k!)^{-1} \sum_{i=0}^{2m} \left(\mathcal{A}_{\theta,\tilde{y},\Delta}^i \circ H_k\right)(0, s, \Delta, \tilde{y}) \Delta^i / (i!)$ where

$$\left(\mathcal{A}_{\theta,\tilde{y},\Delta}^i \circ f\right)(z, t) = \frac{\partial\left(\mathcal{A}^{i-1} \circ f\right)}{\partial z} \frac{\bar{a}\left(t, \sqrt{\Delta} z + \tilde{y}\right)}{\sqrt{\Delta}} + \frac{1}{2\Delta} \frac{\partial^2\left(\mathcal{A}^{i-1} \circ f\right)}{\partial z^2} + \frac{\partial^2\left(\mathcal{A}^{i-1} \circ f\right)}{\partial t}.$$

To illustrate the methodology, Egorov et al. (2003) carried out extensive Monte Carlo experiments using some data generating processes. They showed that for certain values of the parameters and for certain values of Δ, the estimates $p_X^{[m]}$ for $m = 1, 2$ or 3 almost coincide with the true density p_X. We carried out similar experiments and confirmed their results. However, we also observed that a change in the data generating process (using different values θ and Δ) may have an important impact on the properties of $p_X^{[m]}(s, x_0, t, x)$. In some cases, the quality of $p_X^{[m]}(s, x_0, t, x)$ is severely compromised (for $m = 1, 2, 3$). In theory this problem can be overcome by taking $m > 3$. We discuss this issue using an example below.

Monte Carlo experiments shows that when the estimator $p_X^{[m]}(s, x_0, t, x)$ performs exceptionally well, our estimator $\hat{p}_X(s, x_0, t, x)$ can be as accurate as $p_X^{[m]}(s, x_0, t, x)$ for adequate values N and S. However, the real advantage of $\hat{p}_X(s, x_0, t, x)$ over $p_X^{[m]}(s, x_0, t, x)$ is that the former can still perform very well even when the latter shows poor quality (for $m = 1, 2$ or 3).

To analyze some of these issues we consider some time-inhomogeneous processes with known transition densities. These densities provide a benchmark for analyzing the performance of methods in analysis. We first study the diffusion process solution to the time-inhomogeneous stochastic differential equation

$$dX_t = \beta X_t dt + \sigma_0 e^{\sigma_1 t} dW_t. \tag{6}$$

To build $\hat{p}_X(s, x_0, t, x)$ we need the following functions (see equation (1)):

$$\begin{aligned} f(t,x) &= \frac{e^{-t\sigma_1}x}{\sigma_0}, \qquad g(t,x) = e^{t\sigma_1}x\sigma_0, \qquad \bar{a}(t,x) = \beta x - x\sigma_1 \\ \bar{A}(t,x) &= \frac{x^2}{2}(\beta - \sigma_1) \\ \bar{B}(t,x) &= \frac{1}{2}(\sigma_1 - \beta)\left(1 + x^2(\beta - \sigma_1)\right) \end{aligned}$$

It can be easily proved that the true density is (with $\Delta = t - s$)

$$p_X(s, x_0, t, x) = \frac{\exp\left\{-(x - m_t)^2 / (2v_t)\right\}}{\sqrt{2\pi v_t}}, \; m_t = e^{-\beta\Delta}x_0, \; v_t = \frac{\sigma_0^2\left(e^{2\sigma_1 t} - e^{2(\sigma_1 s - \beta\Delta)}\right)}{2(\beta + \alpha_1)}.$$

The SDE (6) was studied in Egorov et al. (2003) for the following parameters:

$$(\beta, \sigma_0, \sigma_1) = (1, 1, -0.001), \qquad \Delta = 1/52.$$

For this particular specification, the estimates of the transition densities based on the Hermite approximation performs very well ($p_X^{[1]}$ almost coincides with the true transition density). Our method reaches the precision of the Hermite approximation for $N \simeq 5000$. However, since the transition density based on the Euler discretization is quite good for the chosen parameters (see Egorov et al., 2003, Fig. 1), the estimation exercise does not seem particularly demanding to us. Changing the step of discretization to $\Delta = 1$ has an important impact on the results of the Monte Carlo experiment (in this case, the transition density based on the Euler discretization moves significantly away from the true density). Figure 1 presents the results. Basically, it compares $p_X^{[2]}$, $p_X^{[3]}$, $\hat{p}_X$ (for $N = 50$ and $S = 1000$) and p_X (true density) when $x_0 = 0$ (we should observe that in practical applications there is no need to fix a value as high as 1000 for S; the advantage of considering $S = 1000$ in this exercise is that the variance of estimates $\hat{p}_X(s, x_0, t, \xi_i)$ are nearly zero, so all estimates are sufficiently representative). All else being equal, that change (i.e. Δ changes to $\Delta = 1$) involves a loss of accuracy in both estimators under analysis. However, whereas our estimator can compensate this loss of accuracy with an increase in N, the method based on the Hermite approximation presents very unsatisfactory results for any expansion between $m = 1$ and $m = 3$ (the highest order considered in Egorov et al., 2003). In particular, some of these estimates become negative in some intervals of state space of X. This poor quality of $p_X^{[m]}$ for $m = 1, 2, 3$ can in theory be overcome by taking $m > 3$. However, we have not tried $m = 4$ as this expansion involves too many terms to be of interest in practical applications. To get an idea of how many terms the Hermite expansion $p_X^{[m]}$ involves in the case of equation (6), we used the command *Expand* for $p_X^{[m]}$ in the Mathematica software, and then counted how many terms were involved in the expansion, using the command *Length:* we got 19 terms for $p_X^{[1]}$, 109 for $p_X^{[2]}$ and 368 for $p_X^{[3]}$. Obviously the command *Simplify* can considerably reduce these expressions, but they are in general very long. In our method, the only cost of having more accurate estimates relates to the processing time. Table 1 gives an idea of the length of time required to obtain $\hat{p}_X(s, x_0, t, x)$ from equation 6 (we are using the GAUSS 9.0 software and an Intel Duo CPU, 2.66 GHz).

** Figure 1 HERE **

** Table 1 HERE **

Another example considered in Egorov at al. (2003) is the SDE

$$dX_t = \beta X_t dt + \sigma_0 e^{\sigma_1 t} X_t dW_t. \tag{7}$$

It can be checked that

$$\begin{aligned} f(t,x) &= \frac{e^{-t\sigma_1}\log(x)}{\sigma_0}, \qquad g(t,x) = e^{e^{t\sigma_1}x\sigma_0}, \qquad \bar{a}(t,x) = \frac{e^{-t\sigma_1}\beta}{\sigma_0} - \frac{1}{2}e^{t\sigma_1}\sigma_0 - x\sigma_1 \\ \bar{A}(t,x) &= -\frac{\sigma_1 x^2}{2} - \frac{1}{2}e^{t\sigma_1}\sigma_0 x + \frac{e^{-t\sigma_1}\beta x}{\sigma_0} \\ \bar{B}(t,x) &= -\frac{1}{2}\left(\frac{e^{-t\sigma_1}\beta}{\sigma_0} - \frac{1}{2}e^{t\sigma_1}\sigma_0 - x\sigma_1\right)^2 + \frac{\sigma_1}{2} - x\left(-\frac{1}{2}e^{t\sigma_1}\sigma_0\sigma_1 - \frac{e^{-t\sigma_1}\beta\sigma_1}{\sigma_0}\right). \end{aligned}$$

Also it can be proved that

$$p_X(s,x_0,t,x) = \frac{\exp\left\{-\left(\log(x) - m_t\right)^2/(2v_t)\right\}}{\sqrt{2\pi v_t}x}, \; v_t = \frac{\left(e^{2t\sigma_1} - e^{2s\sigma_1}\right)\sigma_0^2}{2\sigma_1}, \; m_t = \log(x_0) - \beta\Delta - \frac{1}{2}v_t.$$

The SDE (7) was studied in Egorov et al. (2003) for the following parameters:

$$(\beta, \sigma_0, \sigma_1) = (0.2, 0.25, -0.001), \qquad \Delta = 1/52.$$

The estimates of the transition densities based on the Hermite approximation once again performs exceptional well. However, changing Δ or $(\beta, \sigma_0, \sigma_1)$ may result in a severe loss of accuracy of the Hermite approximation, whereas our method continues to be highly accurate for appropriate values of N. This occurs, for example, for $\beta = 2$ (all else being equal) as can be seen in figure 2, panels A1 and A2. The Hermite approximation deteriorated further if, besides $\beta = 2$, the step of discretization increased to $\Delta = 1/20$. Figure 2, panels B1 and B2, shows that the estimates of the Hermite approximations behave erratically and assume negative values, even for the expansion $m = 3$. For example, it can be verified that for $(\beta, \sigma_0, \sigma_1) = (2, 0.25, -0.001)$ and $\Delta = 1/20$ one obtains

$$
\begin{aligned}
s &= 0,\ x_0 = 1,\ t = 1/20,\ x = 1 \Rightarrow p_X^{[1]}(0, 1, 1/20, 1) = -3.9 < 0, \\
s &= 0,\ x_0 = 1,\ t = 1/20,\ x = 0.9 \Rightarrow p_X^{[2]}(0, 1, 1/20, 0.9) = -2.1 < 0, \\
s &= 0,\ x_0 = 1,\ t = 1/20,\ x = 0.96 \Rightarrow p_X^{[3]}(0, 1, 1/20, 0.96) = -0.9 < 0.
\end{aligned}
$$

On the contrary, our method continues to be precise.

** Figure 2 HERE **

3.2 Conditional Moments Estimation

As mentioned before, $\hat{p}_X(s, x_0, t, x)$ can be used to estimate conditional moments. To illustrate this topic we consider a diffusion process solution to the time-inhomogeneous stochastic differential equation

$$dX_t = X_t \left(e^{\tau t} - \alpha \log(X_t) + \frac{\sigma^2}{2} \right) dt + \sigma X_t dW_t. \tag{8}$$

proposed by Black and Karasinski (1991) to model interest rates. The initial condition ξ It can be checked that (see Brigo and Mercurio, 2006, page 75)

$$
\begin{aligned}
\mathrm{E}\left[X_t \middle| X_s = x_0 \right] &= \int_0^\infty \xi p_X(s, x_0, t, \xi)\, d\xi \\
&= \exp\left\{ e^{-\alpha(t-s)} \ln x_0 + \int_s^t e^{-\alpha(t-u)} e^{\tau u} du + \frac{\sigma^2}{4\alpha} \left(1 - e^{-2\alpha(t-s)} \right) \right\}.
\end{aligned} \tag{9}
$$

The estimator $\hat{p}_X(s, x_0, t, \xi_i)$ is based on the following functions (see equation (1)):

$$f(t,x) = \frac{\log(x)}{\sigma}, \qquad g(t,x) = e^{x\sigma} \qquad \bar{a}(t,x) = \frac{\sigma^2/2 - x\alpha\sigma + e^{t\tau}}{\sigma} - \frac{\sigma}{2} \tag{10}$$

$$\bar{A}(t,x) = \frac{e^{t\tau}x}{\sigma} - \frac{x^2\alpha}{2}, \qquad \bar{B}(t,x) = -\frac{1}{2}\left(\frac{\sigma^2/2 - x\alpha\sigma + e^{t\tau}}{\sigma} - \frac{\sigma}{2}\right)^2 + \frac{\alpha}{2} - \frac{e^{t\tau}x\tau}{\sigma}. \tag{11}$$

To estimate (9) for each x_0 in the interval $[0.5, 1.2]$ we consider, according to theorem 2.3, the estimator

$$\sum_{i=1}^{M} \xi_i \times \hat{p}_X(s, x_0, t, \xi_i)\,\Delta\xi_i, \qquad x_0 \in [0.5, 1.2] \tag{12}$$

where $\xi_i = a + i(b-a)/M$ and $a = 0.5$ and $b = 1.7$. The values a and b were selected so that

$$\int_a^b \xi p_X(s, x_0, t, \xi)\,d\xi \simeq \int_0^\infty \xi p_X(s, x_0, t, \xi)\,d\xi.$$

To discuss the effectiveness of the estimator defined in equation (12) we set $(\tau, \alpha, \sigma) = (-0.13, 0.409, 0.344)$ and $\Delta = 1/12$. These values are the maximum likelihood estimates from the monthly observations of the U.S. Fed funds rate in the period July 1954-March 2008 (see section 3.3). The other parameters were fixed as follow: $s = 0$, $t = 1/12$, $N = 30$, $S = 1000$ (as mentioned before, there is no need to fix a value as high as 1000 for S). For this specification we take (12) as an estimator for the conditional mean (9) for different values for M.

** Figure 3 HERE **

From figure 3 we conclude that the accuracy of the estimator (12) increases as M goes from 5 to 100. At $M = 100$ the estimator (12) gives almost the true vales of the conditional mean.

3.3 SML Estimator

We now present the performance of the proposed method for parameter estimation from a set of Monte Carlo experiments based on the SDE (8). We used the expressions (10) and (11) to build the estimator $\hat{p}_X$. The transition density of X is given by

$$p_X(s, x_0, t, x) = \frac{\exp\left\{-(\log(x) - m_t)^2/(2v_t)\right\}}{\sqrt{2\pi v_t}x}, \ (m_t,\ v_t \text{ are the conditional moments})$$

Knowing the true density p_X allows us to obtain the MLE, which can then be used to assess the accuracy of the simulated maximum likelihood estimates. The data generated process used in the Monte Carlo simulation was defined according to the following lines: the step of discretization was initially fixed as $\Delta = 1/12$ and the vector $(\tau, \alpha, \sigma) = (-0.013, 0.409, 0.344)$ was calibrated according to the maximum likelihood estimates from the monthly observations of the U.S. Fed funds rate in the period July 1954-March 2008 (645 observations).

From the transition densities we simulated 2000 paths of 645 observations of X_t, first with a step size $\Delta = 1/12$ and then with $\Delta = 1/4$, using the ML estimates $(\tau, \alpha, \sigma) = (-0.013, 0.409, 0.344)$. The initial value of X was $\xi = \exp\left\{\sigma^2/(4\alpha)\right\} = 1.075$. We then estimated $\theta = (\tau, \alpha, \sigma)$ using three methods: ML, SML (we analyze $N = S = 30$ and $N = S = 60$) and Hermite Expansion ($m = 3$). In all cases, the estimation was based on the Constrained Maximum Likelihood Estimation procedure written for GAUSS with the option of switching between the BFGS, DFP, NEWTON and BHHH algorithms depending on change in function value, number of iterations, or change in line search step length - default settings.

The results of the Monte Carlo experiment are presented in tables 3 and 4. In table 2 we calculate the mean and the standard deviation of the difference between the ML estimates and the other estimates, in the cases $\Delta = 1/12$ and $\Delta = 1/4$. For example, the value -6.4E-05 in table 2, was obtained by using the expression $2000^{-1} \sum_{i=1}^{2000} \left(\hat{\tau}_{ML_i} - \hat{\tau}_{SML_i^*}\right)$, where $\hat{\tau}_{ML_i}$ and $\hat{\tau}_{SML_i^*}$ are the ith ML and SML estimate of τ, respectively.

** Table 2 HERE **

In table 3 we calculate the root mean squared error of the various estimators (times 100). For example, the value 0.454 in table 3 was obtained using the expression $\sqrt{2000^{-1} \sum_{i=1}^{2000} \left(\tau_{ML_i} - (-0.013)\right)^2} \times 100$.

** Table 3 HERE **

Some conclusions emerge from these tables:

- In the case $\Delta = 1/12$ the Hermite method is, for practical purposes, the ML estimator. However, when Δ changes to $\Delta = 1/4$ the Hermite method ($m = 3$) is completely inadequate: in 10 out of 2000 simulated paths the method of estimation could not be implemented as the transition densities assume negative values during the optimization procedure causing the algorithm of optimization to crash (for this reason, some entries in table 2 and 4 are named "N/A"). Even considering only the 1990 cases where the algorithm reaches an optimum solution, the RMSE of the estimators based on the Hermite are much higher than that of the SML estimators (about 10 times higher). The results are even worse if Δ is higher. As we discussed previously, although the Hermite method works if the problem is sufficiently undemanding, it can also be highly unstable for a certain range of the parameters and/or for high values of Δ. This can be seen again in the present Monte Carlo experiment. For example, one can verify that for $s = 0$, $x_0 = 1$, $t = 1/4$, $x = 0.6$ the transition density turns out to be negative, $p_X^{[3]}(0, 1, 1/4, 0.6) = -0.017$. Obviously, in this case, the method of estimation based on Hermite expansion could not be implemented as the transition densities assume negative values during the optimization procedure causing the optimization algorithm to crash. This poor quality of $p_X^{[3]}$ can in theory be overcome by taking $m > 3$. However, as we observed, higher expansions ($m > 3$) involves too many terms to be of interest in practical applications (after expanding the expression $p_X^{[3]}$ we count 410 terms using the *Length* command of the Mathematica software package).

- The SML estimator mimics the ML estimator. Notice that, in practical terms, the difference between the ML and SML are negligible. For example, in the case of $N = 30$, $S = 30$, $\Delta = 1/12$ the ML and SML estimates of α agree in average in the 3 decimal places which represents a difference of about 0.1% (see table 2); for example, if the ML of α is $\hat{\alpha}_{ML} = 0.4092$, a typical SML estimate may be, for example, $\hat{\alpha}_{SML^{**}} = 0.4097$. Moreover, table 3 shows that the RMSE of the SML estimator almost coincides with that of the ML estimator.

- Table 3 shows that the results of the SML and ML estimators improved, in general, when Δ changes to $\Delta = 1/4$. A simple explanation for this is that with $\Delta = 1/12$ the period of observation is $[0, 645/12] = [0, 53.75]$ whereas with $\Delta = 1/4$ the period is $[0, 645/4] = [0, 161.25]$. This effect of enlarging the period of observation increases the available information concerning the long-run behavior of the process, and thus increases the precision of the estimators. Naturally, the properties of the SML estimator could have been affected by the increase of Δ; this is not the case, however, as N is big enough. Notice that the results would be quite different if the period $[0, 645/12]$ was kept constant with $\Delta = 1/4$.

4 Conclusion

Time-inhomogeneous diffusions are extremely relevant for term structure of interest rates and, in general, to model stochastic processes that depend explicitly on time due to seasonality, economic business cycles and monetary policy, among other reasons. When the time-inhomogeneous diffusions are non-linear the MLE cannot be used in general. Probably, the best known method of estimation available in the literature to deal with such diffusions is the one proposed by Egorov et al. (2003). As shown in Egorov et al. (2003) and confirmed by us, this method, for specification of certain parameters and Δ, may provide the ML estimates. However, as shown in this paper, the method of Egorov et al. (2003), for Hermite expansion between $m = 1$ and $m = 3$, may also be completely inadequate (for example, it may render negative values for the transition densities). Although this problem may be overcome by using higher expansions ($m > 3$), the number of terms associated with such expansions increases exponentially (as mentioned, the number of terms for the transition density approximation based on the Hermite expansion for $m = 3$ is about 368 for a simple SDE as $dX_t = \beta X_t dt + \sigma_0 e^{\sigma_1 t} dW_t$). These higher Hermite expansions (say, for $m > 3$) not only involve too many terms to be of interest in practical applications, but may also raise doubts about the precision of the optimization algorithms. Our method is asymptotically equivalent to the maximum likelihood method (as $N, S \to \infty$) and the cost of having more accurate estimates is related to the processing time.

Appendix

Proof of equation (1) Consider $dY_t = a(t, Y_t)\,dt + dW_t$ and assume $P\left[\int_s^t a^2(u, Y_u)\,du < +\infty\right] = 1.$ By Girsanov's theorem we have $dP^{y_0}/dW^{y_0} = L_\Delta^{y_0}$ where $t - s = \Delta$ and

$$L_\Delta^{y_0} = \exp\left\{\int_s^t a(u, Y_u)\,dY_u - \frac{1}{2}\int_s^t a^2(u, Y_u)\,du\right\}.$$

$L_\Delta^{y_0}$ is the density of the measure P^{y_0} of the solution Y when $Y_s = y_0$ with respect to the Wiener measure W^{y_0}. Thus, for any bounded function ϕ

$$E_{P^{y_0}}[\phi(Y_t)] = E_{W^{y_0}}[\phi(Y_t)L] = E_{W^{y_0}}[E_{W^{y_0}}[\phi(Y_t)L_\Delta^{y_0}|\,Y_t]] = \int \phi(y)\,E_{W^{y_0}}[L_\Delta^{y_0}|\,Y_t = y]\,dW^{y_0}(y).$$

Thus,

$$\int \phi(y)\,p_Y(s, y_0, t, y)\,dy = \int \phi(y)\,E_{W^{y_0}}[L_\Delta^{y_0}|\,y]\frac{1}{\sqrt{2\pi\Delta}}\exp\left\{-\frac{(y-y_0)^2}{2\Delta}\right\}dy$$

and so, we can conclude

$$p_Y(s, y_0, t, y) = E_{W^{y_0}}[L_\Delta^{y_0}|\,y]\frac{1}{\sqrt{2\pi\Delta}}\exp\left\{-\frac{(y-y_0)^2}{2\Delta}\right\}.$$

Let us analyze now $E_{W^{y_0}}[L_\Delta^{y_0}|\,y]$. With $A(t, y) = \int_0^y a(t, u)\,du$ we have, by Ito's formula

$$A(t, Y_t) - A(s, y_0) = \int_s^t\int_s^{Y_u} a_t'(u, \xi)\,d\xi du + \frac{1}{2}\int_s^t a_y'(u, Y_u)\,du + \int_s^t a(u, Y_u)\,dY_u$$

Thus,

$$\begin{aligned}
L_\Delta^y &= \exp\left\{\int_s^t a(u, Y_u)\,dY_u - \frac{1}{2}\int_s^t a^2(u, Y_u)\,du\right\}\\
&= \exp\left\{A(t, Y_t) - A(s, y_0) - \int_s^t \frac{1}{2}a^2(u, Y_u)\,du - \int_s^t\int_s^{Y_u} a_t'(u, \xi)\,d\xi du - \frac{1}{2}\int_s^t a_y'(u, Y_u)\,du\right\}\\
&= \exp\left\{A(t, Y_t) - A(s, y_0) + \int_s^t B(u, Y_u)\,du\right\}\\
&= \exp\left\{A(t, Y_t) - A(s, y_0) + \Delta\int_0^1 B(s + \lambda\Delta, Y_{s+\lambda\Delta})\,d\lambda\right\}
\end{aligned}$$

where $B(t, y) = -\frac{1}{2}a^2(t, y) - \int_s^y a_t'(t, \xi)\,d\xi - \frac{1}{2}a_y'(t, y)$. We have

$$E_{W^{y_0}}[L_\Delta^y|\,Y_t = y] = \exp\{A(t, y) - A(s, y_0)\}\,E_{W^{y_0}}\left[\exp\left\{\Delta\int_0^1 B(s + \lambda\Delta, Y_{s+\lambda\Delta})\,d\lambda\right\}\middle|\,Y_t = y\right].$$

We can now check that $Y_{s+\lambda\Delta} \equiv (1-\lambda)y_0 + \lambda y + Y_{s+\lambda\Delta} - y_0 - \lambda y + \lambda y_0$ and, under the condition, $Y_t = y$ and the measure W^{y_0} one has

$$Y_{s+\lambda\Delta} = (1-\lambda)y_0 + \lambda y + \sqrt{\Delta}\left(\frac{Y_{s+\lambda\Delta} - y_0}{\sqrt{\Delta}} - \lambda\frac{Y_t - y_0}{\sqrt{\Delta}}\right) = z_u(y, y) + \sqrt{\Delta}\eta(\lambda)$$

($\{\eta(\lambda), 0 \leqq \lambda \leqq 1\}$ is the Brownian bridge). To verify that $\eta(\lambda)$ is in fact a Brownian bridge one needs to conclude that $(Y_{s+\lambda\Delta} - y_0)/\sqrt{\Delta}$ is a Wiener process under the measure W (where $dW/dP = \exp\left\{-\int_0^t a(u, Y_u)\, dY_u + 1/2 \int_0^t a^2(u, Y_u)\, du\right\}$). By Girsanov's theorem, $\tilde{W}_t = \int_0^t a(u, Y_u)\, du + W_t$ is a Wiener process under the measure W so $(Y_{s+\lambda\Delta} - y_0)/\sqrt{\Delta} = \left(\int_0^{s+\lambda\Delta} a(u, Y_u)\, du\right)/\sqrt{\Delta} = \tilde{W}_{s+\lambda\Delta}/\sqrt{\Delta}$ is also a Wiener process with variance 1 when $\lambda = 1$. To sum up,

$$\begin{aligned} E_{W^{y_0}}\left[L^y_\Delta | Y_t = y\right] &= \exp\{A(t, y) - A(s, y_0)\} \\ &\quad \times E_{W^{y_0}}\left[\exp\left\{\Delta \int_0^1 B\left(s + \lambda\Delta, (1-\lambda)y_0 + \lambda y + \sqrt{\Delta}\eta(\lambda)\right) d\lambda\right\}\right] \end{aligned}$$

$$\begin{aligned} p_Y(s, y_0, t, y) &= \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{-\frac{(y - y_0)^2}{2\Delta} + A(t, y) - A(s, y_0)\right\} \\ &\quad \times E_{W^{y_0}}\left[\exp\left\{\Delta \int_0^1 B\left(s + \lambda\Delta, (1-\lambda)y_0 + \lambda y + \sqrt{\Delta}\eta(\lambda)\right) d\lambda\right\}\right] \end{aligned}$$

where

$$A(t, y) = \int_0^y a(t, u)\, du, \qquad B(t, y) = -\frac{1}{2}a^2(t, y) - \int_s^y a'_t(t, \xi)\, d\xi - \frac{1}{2}a'_y(t, y).$$

Consider now the more general case $dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t$. Let $f(t, x) = \int^x b^{-1}(t, \xi)\, d\xi$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be the inverse function of f, i.e. $f(t, g(t, x)) = x$. By Ito's formula $Y_t = f(t, X_t)$

$$dY_t = \bar{a}(t, Y_t)\, dt + dW_t, \qquad \bar{a}(t, y) = \int_0^{g(t,y)} \frac{b'_t(t, \xi)}{b^2(t, \xi)} d\xi + \frac{a(t, g(t, y))}{b(t, g(t, y))} - \frac{1}{2}b'_x(t, g(t, y))$$

Thus

$$\begin{aligned} p_Y(s, y_0, t, y) &= \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{-\frac{(y - y_0)^2}{2\Delta} + \bar{A}(t, y) - \bar{A}(s, y_0)\right\} \\ &\quad \times E_{W^{y_0}}\left[\exp\left\{\Delta \int_0^1 \bar{B}\left(s + \lambda\Delta, (1-\lambda)y_0 + \lambda y + \sqrt{\Delta}\eta(\lambda)\right) d\lambda\right\}\right] \end{aligned}$$

where

$$\bar{A}(t, y) = \int_0^y \bar{a}(t, u)\, du, \qquad B(t, y) = -\frac{1}{2}\bar{a}^2(t, y) - \int_s^y \bar{a}'_t(t, \xi)\, d\xi - \frac{1}{2}\bar{a}'_y(t, y)$$

and by Jacobian formula

$$\begin{aligned} p_X(s, x_0, t, x) &= f'_x(t, x)\, p_Y(s, f(s, x_0), t, f(t, x)) \\ &= \frac{1}{\sqrt{2\pi\Delta}b(t, x)} \exp\left\{-\frac{(f(t, x) - f(s, x_0))^2}{2\Delta} + \bar{A}(t, f(t, x)) - \bar{A}(s, f(s, x_0))\right\} \\ &\quad \times E_{W^{y_0}}\left[\exp\left\{\Delta \int_0^1 \bar{B}\left(s + \lambda\Delta, (1-\lambda)f(s, x_0) + \lambda f(t, x) + \sqrt{\Delta}\eta(\lambda)\right) d\lambda\right\}\right]. \end{aligned}$$

Proof of theorem 2.1 By assumptions A1 and A2 the density $p_X(s,x_0,t,x)$ as it is defined in equation (1) exists and is well defined (see Gihman and Skorohod, 1972). Given the almost sure (a.s.) continuity of $\bar{B}\left(s+\lambda\Delta,(1-\lambda)f(s,x_0)+\lambda f(t,x)+\sqrt{\Delta}\eta(\lambda)\right)$ in λ, it follows that

$$\frac{\Delta}{N}\sum_{i=0}^{N-1}\bar{B}\left(s+\frac{i}{N}\Delta,\left(1-\frac{i}{N}\right)f(s,x_0)+\frac{i}{N}f(t,x)+\sqrt{\Delta}\eta\left(\frac{i}{N}\right)\right)$$

converges in probability to $\int_0^1\bar{B}\left(s+\lambda\Delta,(1-\lambda)f(s,x_0)+\lambda f(t,x)+\sqrt{\Delta}\eta(\lambda)\right)d\lambda$ as $N\to\infty$ and so, by Slutsky's theorem, $\psi_N(\omega)$ converges in probability to $\psi(\omega)$ as $N\to+\infty$. Therefore, we have in probability

$$\begin{aligned}\lim_{S\to+\infty}\left(\lim_{N\to+\infty}\hat{\psi}\right) &= \lim_{S\to+\infty}\left(\lim_{N\to+\infty}\frac{1}{S}\sum_{j=1}^{S}\psi_N(\omega_j)\right)=\lim_{S\to+\infty}\frac{1}{S}\sum_{j=1}^{S}\lim_{N\to+\infty}\psi_N(\omega_j)\\ &= \lim_{S\to+\infty}\frac{1}{S}\sum_{j=1}^{S}\psi(\omega_j)=E[\psi]\end{aligned}$$

by the law of large numbers, since the $\psi(\omega_j)$ are i.i.d. and $E[|\psi|]=E[\psi]<+\infty$. Thus, for each (s,x_0,t,x), the estimator $\hat{p}_X(s,x_0,t,x)$ converges in probability to $p_X(s,x_0,t,x)$ as $N\to+\infty$ and $S\to+\infty$.■

Proof of theorem 2.2 To guarantee that $\hat{p}_X(s,x_0,t,x)$ converges uniformly in probability on Ξ to $p_X(s,x_0,t,x)$ it is sufficient to assure that (i) $\hat{p}_X(s,x_0,t,x)$ converges in probability for each $(s,x_0,t,x)\in\Xi$ (pointwise convergence) and (ii) $\hat{p}_X(s,x_0,t,x)$ is stochastically equicontinuous (see Davidson, 1994, theorem 21.9). Theorem 2.1 guarantees (i), so it remains to prove (ii). By theorem 21.10 of Davidson (1994), a sufficient condition for $\hat{p}_X(s,x_0,t,x)$ to be stochastically equicontinuous is that $\beta=O_p(1)$ where

$$\beta=\sup_{(s^*,x_0^*,t^*,x^*)\in\Xi^*}\left\|\left(\frac{\partial\hat{p}_X(s^*,x_0^*,t^*,x^*)}{\partial s},\frac{\partial\hat{p}_X(s^*,x_0^*,t^*,x^*)}{\partial t},\frac{\partial\hat{p}_X(s^*,x_0^*,t^*,x^*)}{\partial x_0},\frac{\partial\hat{p}_X(s^*,x_0^*,t^*,x^*)}{\partial x}\right)\right\| \tag{13}$$

and $\|\cdot\|$ is a norm, Ξ^* is an open convex set containing Ξ and $(s^*,x_0^*,t^*,x^*)\in\Xi^*$ is a point on the line segment joining two arbitrary points z and z' in Ξ^*. Assumptions A1-A3 guarantee the existence of the derivatives $\partial p_X(s,x_0,t,x)/\partial s, \partial p_X(s,x_0,t,x)/\partial t, \partial p_X(s,x_0,t,x)/\partial x_0, \partial p_X(s,x_0,t,x)/\partial x$ (see Gihman and Skorohod, 1972), thus it remains to prove that $\beta=O_p(1)$. If all elements within the norm in (13) are $O_p(1)$ then $\beta=O_p(1)$. Let $\hat{p}_X(s,x_0,t,x)=\phi\hat{\psi}=\phi(s,x_0,t,x)\hat{\psi}(s,x_0,t,x)$ where

$$\phi=\phi(s,x_0,t,x)=\frac{1}{\sqrt{2\pi\Delta}b(t,x)}\exp\left\{-\frac{(f(t,x)-f(s,x_0))^2}{2\Delta}+\bar{A}(t,f(t,x))-\bar{A}(s,f(s,x_0))\right\}.$$

We have

$$\frac{\partial\hat{p}_X(s,x_0,t,x)}{\partial x}=\frac{\partial\phi}{\partial x}\hat{\psi}+\phi\frac{\partial\hat{\psi}}{\partial x}=\frac{\partial\phi}{\partial x}\hat{\psi}+\phi\frac{1}{S}\sum_{j=1}^{s}\frac{\partial\psi_N(\omega_j)}{\partial y}.$$

The term

$$\frac{\partial \phi}{\partial x} = \phi \times \left(-\frac{b'_x(t,x)}{b(t,x)} - \frac{f(t,x) - f(s,x_0)}{\Delta} f'(s,x) + \bar{A}'_x(t, f(t,x)) f'(t,x) \right)$$

is a continuous deterministic function. Therefore, it is bounded on all points on the line segment joining the points (s, x_0, t, x) and (s', x'_0, t', x') in Ξ^* such that $p_X(s, x_0, t, x)$ and $p_X(s', x'_0, t', x')$ are finite. Since $\hat{\psi} \xrightarrow{p} \mathrm{E}[\psi]$ by theorem 2.1, it follows that the term $(\partial \phi / \partial x) \hat{\psi}$ is bounded in probability. On the other hand, given the almost sure (a.s.) continuity of

$$\bar{B}\left(s + \lambda \Delta, (1-\lambda) f(s, x_0) + \lambda f(t,x) + \sqrt{\Delta}\eta(\lambda)\right)$$

and

$$\frac{\partial \left(\bar{B}\left(s + \lambda \Delta, (1-\lambda) f(s, x_0) + \lambda f(t,x) + \sqrt{\Delta}\eta(\lambda)\right)\right)}{\partial x}$$

in λ, it follows that

$$\frac{\partial \psi_N(\omega)}{\partial x} = \psi_N(\omega) \left\{ \frac{\Delta}{N} \sum_{i=0}^{N-1} \bar{B}'_x \left(s + \frac{i}{N}\Delta, \left(1 - \frac{i}{N}\right) f(s, x_0) + \frac{i}{N} f(t,x) + \sqrt{\Delta}\eta\left(\frac{i}{N}\right)\right)\right\}$$

converges in probability to $\psi(\omega)\xi(\omega)$ as $N \to +\infty$ where

$$\xi(\omega) = \Delta \int_0^1 \left(\bar{B}'_2 \left(s + \lambda \Delta, (1-\lambda) f(s, x_0) + \lambda f(t,x) + \sqrt{\Delta}\eta(\lambda)\right)\right) \lambda f'(s,x)\, d\lambda.$$

($\bar{B}'_2$ is the derivative with respect to the second argument of the function $\bar{B}$). The random variable $\xi(\omega)$ is bounded in probability since the limits of integration are finite, the integrand is continuous and η_t, $t \in [0,1]$ is a bounded process (a.s.). Using the same arguments as in the proof of theorem 2.1, one can conclude that the random quantity $S^{-1} \sum_{j=1}^{s} \partial \psi_N(\omega_j) / \partial x$ converges in probability to $E[\partial \psi_N(\omega) / \partial x]$. Similar arguments can be used to show that $\partial \hat{p}_X(s^*, x^*_0, t^*, x^*) / \partial s = O_p(1), \partial \hat{p}_X(s^*, x^*_0, t^*, x^*) / \partial t = O_p(1), \partial \hat{p}_X(s^*, x^*_0, t^*, x^*) / \partial x_0 = O_p(1)$. ■

Proof of theorem 2.3 By theorem 2.2 we can write $\hat{p}_X(s, x_0, t, \xi_i) = p_X(s, x_0, t, \xi_i) + o_p(1)$ where $\lim o_p(1) = 0$ uniformly in probability on Ξ as $N \to +\infty$ and $S \to +\infty$. Without loss of generality consider $\xi_0 = 0$ and $\xi_1 = 1$ so $\Delta \xi_i = 1/M$. We should prove that the expression (14) goes to zero in probability as

$M \to \infty$, $N \to \infty$, $S \to \infty$. We have,

$$\left|\sum_{i=1}^{M} f(t,\xi_i)\,\hat{p}_X(s,x_0,t,\xi_i)\,\Delta\xi_i - \int_{\xi_0}^{\xi_1} f(t,\xi)\,p_X(s,x_0,t,\xi)\,d\xi\right| \tag{14}$$

$$\begin{aligned}
&= \left|\frac{1}{M}\sum_{i=1}^{M} f\left(t,\frac{1}{M}\right)\hat{p}_X\left(s,x_0,t,\frac{1}{M}\right) - \int_0^1 f(t,\xi)\,p_X(s,x_0,t,\xi)\,d\xi\right| \\
&\leq \left|\frac{1}{M}\sum_{i=1}^{M} f\left(t,\frac{1}{M}\right)p_X\left(s,x_0,t,\frac{1}{M}\right) - \int_0^1 f(t,\xi)\,p_X(s,x_0,t,\xi)\,d\xi + \frac{1}{M}\sum_{i=1}^{M} o_p(1)\right| \\
&\leq \left|\frac{1}{M}\sum_{i=1}^{M} f\left(t,\frac{1}{M}\right)p_X\left(s,x_0,t,\frac{1}{M}\right) - \int_0^1 f(t,\xi)\,p_X(s,x_0,t,\xi)\,d\xi\right| + \left|\frac{1}{M}\sum_{i=1}^{M} o_p(1)\right|. \qquad (15)
\end{aligned}$$

Due to the continuity of $f(t,\xi)\,p_X(s,x_0,t,\xi)$, $\frac{1}{M}\sum_{i=1}^{M} f\left(t,\frac{1}{M}\right)p_X\left(s,x_0,t,\frac{1}{M}\right)$ is a Riemann sum. Therefore, the first term in modulus of (15) goes to zero as $M \to +\infty$. The second term goes to zero uniformly in probability on Ξ as $N \to +\infty$ and $S \to +\infty$.■

References

Aït-Sahalia, Y. (1999). Transition Densities for Interest Rate and Other Nonlinear Diffusions. Journal of Finance 54(4):1361-1395.

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. Econometrica 70(1):223-262.

Beskos, A., and G. Roberts (2005). Exact Simulation of Diffusions. Annals of Applied Probability 4, 2422-2444.

Black, F., and P. Karasinski (1991). Bond and option pricing when short rates are log-normal. Financial Analysts Journal 47:52-29.

Brigo, D., and F. Mercurio (2006). *Interest Rate Models, Theory and Practice.* Second Edition. Heidelberg: Springer Finance.

Dacunha-Castelle, D., and D. Florens-Zmirou (1986). Estimation of the Coefficient of a Diffusion From Discretely Sample Observations. Stochastic 19:263-284.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Durham, G., and A. Gallant (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. Journal of Business & Economic Statistics 20(3):297-338.

Egorov, A., L. Haitao, and X. Yuewu (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. Journal of Econometrics 114(1):107-139.

Gihman I., and A. Skorohod (1972). *Stochastic Differential Equations.* New York: Springer Verlag.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect Inference. Journal of Applied Econometrics 8:85-118.

Iacus, S. (2008). *Simulation and Inference for Stochastic Differential Equations: with R Examples.* New York: Springer.

Kessler, M. (1997). Estimation of an Ergodic Diffusion from Discrete Observations. Scandinavian Journal of Statistics 24, 211-229.

Nicolau, J. (2002). New technique for simulating the likelihood of stochastic differential equations. Econometrics Journal 5:91-203 .

Pedersen A. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equation based on Discrete Observations. Scandinavian Journal of Statistics 22:55-71.

Prakasa Rao, B. (1999). *Statistical Inference for Diffusion Type of Processes*, Oxford University Press, New York.

Rogers, L. and D. Williams (1987). *Diffusions, Markov Processes and Martingales, Vol. 2: Itô Calculus*, Wiley, Chichester.

Santa-Clara, P. (1995). Simulated Likelihood Estimation of Diffusions with an Application to the Short Term Interest Rate. Anderson Graduate School of Management, University of California at Los Angeles.

Shoji, L., and T. Osaki (1998). Estimation for Nonlinear Stochastic Differential Equations by a Local Linearizaton Method Stochastic Analysis and Applications 16(4), 733-752

Yoshida N. (1992). Estimation for Diffusion Processes from Discrete Observations. Journal of Multivariate Analysis 41(2), 220-242.

Tables

Table 1 Computational Times to obtain $\hat{p}_X(s, x_0, t, x)$

N	S	Computation Time (secs.)
100	100	$\simeq 0$
500	500	0.06
1000	1000	0.27
2000	2000	1
3000	3000	2.4
5000	5000	7.2

Table 2 Estimation results from the Monte Carlo study I

	$\hat{\tau}_{ML} - \hat{\tau}_{SML^*}$		$\hat{\tau}_{ML} - \hat{\tau}_{SML^{**}}$		$\hat{\tau}_{ML} - \hat{\tau}_{Her(3)}$	
	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$
Mean	-6.4E-05	8.98E-05	-7.6E-05	-2.35E-05	4.6E-08	N/A
SDev.	4.0E-05	2.10E-05	2.9E-05	1.78E-05	2.8E-06	N/A
	$\hat{\alpha}_{ML} - \hat{\alpha}_{SML^*}$		$\hat{\alpha}_{ML} - \hat{\alpha}_{SML^{**}}$		$\hat{\alpha}_{ML} - \hat{\alpha}_{Her(3)}$	
	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$
Mean	-6.1E-04	1.64E-03	-5.1E-04	-1.36E-05	2.5E-07	N/A
SDev.	3.9E-04	5.80E-04	3.1E-04	4.30E-04	8.9E-06	N/A
	$\hat{\sigma}_{ML} - \hat{\sigma}_{SML^*}$		$\hat{\sigma}_{ML} - \hat{\sigma}_{SML^{**}}$		$\hat{\sigma}_{ML} - \hat{\sigma}_{Her(3)}$	
	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$	$\Delta = \frac{1}{12}$	$\Delta = \frac{1}{4}$
Mean	-2.4E-05	1.90E-04	-1.6E-05	6.36E-05	-1.2E-07	N/A
SDev.	4.7E-05	6.78E-05	2.3E-05	3.17E-05	3.9E-06	N/A

$^{(*)}N = S = 30$;$^{(**)}$ $N = S = 60$; $Her(3)$: Hermite Expan., $m = 3$

Table 3 Estimation results from the Monte Carlo study II

	$\hat{\tau}_{ML}$		$\hat{\tau}_{SML^{*}}$		$\hat{\tau}_{SML^{**}}$		$\hat{\tau}_{Her(3)}$	
	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$
RMSE100	0.454	0.170	0.456	0.170	0.455	0.169	0.454	N/A
SEML100	0.449	0.169						
	$\hat{\alpha}_{ML}$		$\hat{\alpha}_{SML^{*}}$		$\hat{\alpha}_{SML^{*}}$		$\hat{\alpha}_{Her(3)}$	
	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$
RMSE100	4.966	3.309	4.996	3.263	4.987	3.305	4.966	N/A
SEML100	4.790	3.250						
	$\hat{\sigma}_{ML}$		$\hat{\sigma}_{SML^{*}}$		$\hat{\sigma}_{SML^{*}}$		$\hat{\sigma}_{Her(3)}$	
	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$	$\Delta=\frac{1}{12}$	$\Delta=\frac{1}{4}$
RMSE100	0.970	0.975	0.969	0.975	0.969	0.975	0.970	N/A
SEML100	0.967	0.975						

$^{(*)}N=S=30;^{(**)}N=S=60$; $Her(3)$: Hermite Expan., $m=3$

RMSE100 : Root mean squared error multiplied by 100

SEML100 : Std. Error of MLE multiplied by 100

Figures

Figure 1: $\hat{p}_X$ when X is governed by the SDE (6) for $(\beta, \sigma_0, \sigma_1) = (1, 1, -0.001)$, $\Delta = 1$. The estimator $\hat{p}_X$ was obtained for $N = 50$ and $S = 1000$. On the left panel we plot the densities; on the right panel we plot the difference between the true density and the approximations

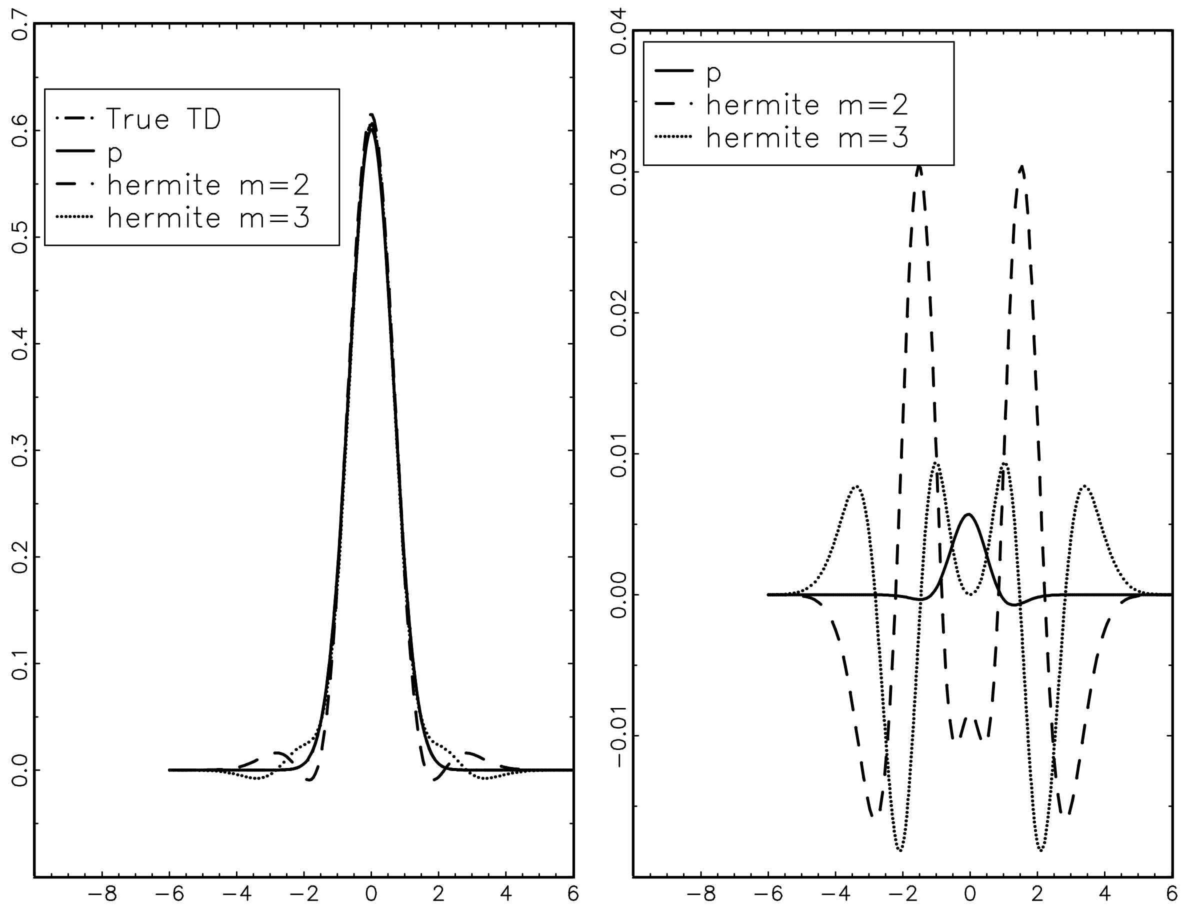

Figure 2: Assessing the performance of $p_X^{[2]}$, $p_X^{[3]}$, $\hat{p}_X$ when X is governed by the SDE (7). Key: $\cdot - \cdot\cdot$ p_X; — $\hat{p}_X$; $- -$ $p_X^{[2]}$; $\cdot\cdot\cdot\cdot$ $p_X^{[3]}$. Painels A1 and A2: $(\beta, \sigma_0, \sigma_1) = (2, 0.25, -0.001)$, $\Delta = 1/52$; $\hat{p}_X$ was obtained for $N = 60$ and $S = 1000$. Painel B1 and B2: $(\beta, \sigma_0, \sigma_1) = (2, 0.25, -0.001)$, $\Delta = 1/20$; $\hat{p}_X$ was obtained for $N = 100$ and $S = 1000$. The deviation of $p_X^{[2]}$ and $p_X^{[3]}$ from the true density is subtancial, mainly in panels B1 and B2

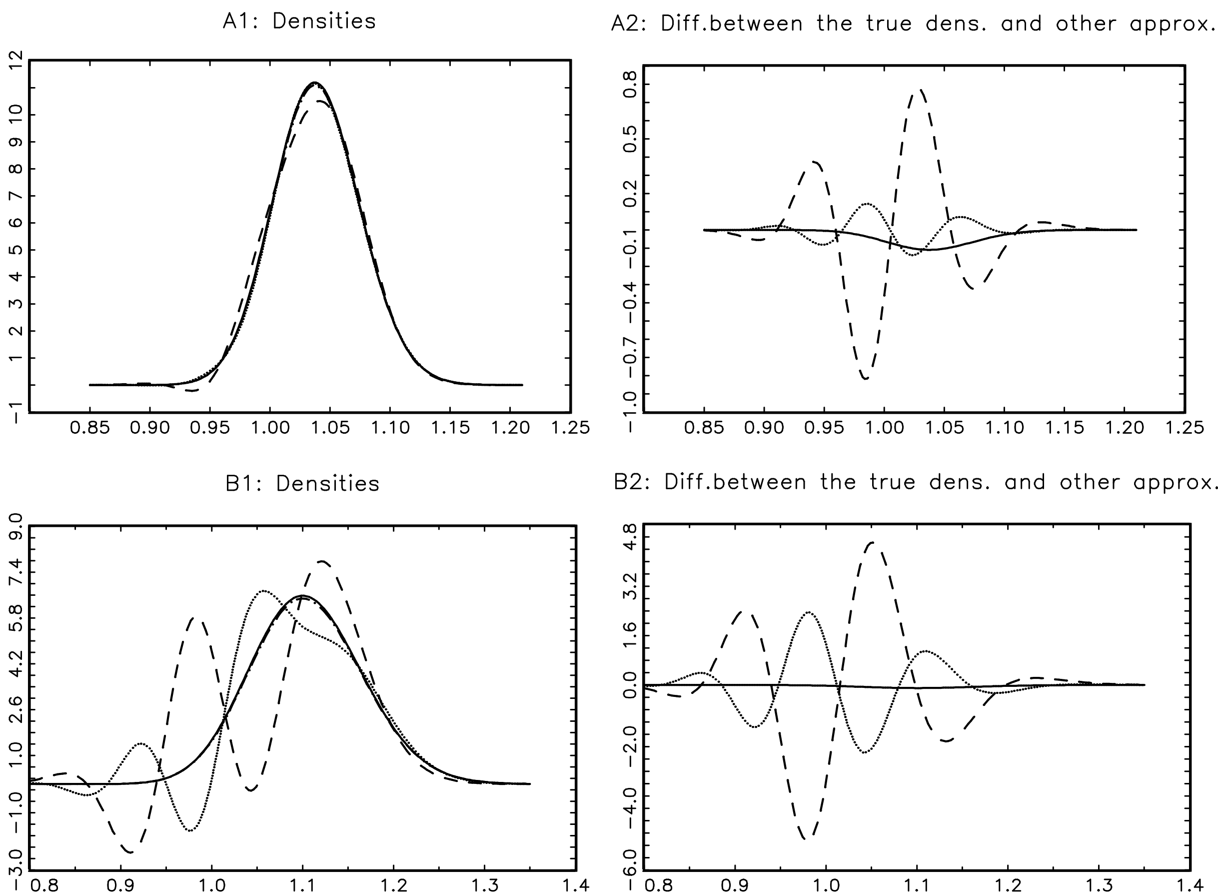

Figure 3: Comparison between the true conditional mean, equation (9) and the approximation (12), for different values of M