7-2011

Quasi-maximum Likelihood Estimation of Discretely Observed Diffusions

Xiao Huang
Kennesaw State University, xhuang3@kennesaw.edu

Quasi-maximum Likelihood Estimation of Discretely Observed Diffusions*

Xiao Huang
Department of Economics, Finance and Quantitative Analysis
Kennesaw State University

June 2010

Abstract

This paper introduces quasi-maximum likelihood estimator for discretely observed diffusions when a closed-form transition density is unavailable. Higher order Wagner-Platen strong approximation is used to derive the first two conditional moments and a normal density function is used in estimation. Simulation study shows that the proposed estimator has high numerical precision and good numerical robustness. This method is applicable to a large class of diffusions.

Keywords Quasi-maximum likelihood estimator, diffusions, Wagner-Platen approximation.

*The author thanks a co-editor and three referees for outstanding comments and suggestions. Thanks also go to Eckhard Platen, Yongjun Tang, and seminar participants at the SNDE 17th annual symposium for helpful comments. All remaining errors are mine. E-mail: xhuang3@kennesaw.edu

1 Introduction

Diffusion processes have been widely used in many research fields to model continuous time phenomena and they are usually characterized by stochastic differential equations (SDEs). Examples include modeling gene changes due to natural selection in genetics, vertical motion of the ground level in seismology, outflow from a reservoir in hydrology, asset prices in finance, etc. When the drift and diffusion coefficient of a SDE are parametrically specified, it is crucial to obtain precise parameter estimates. A major difficulty in estimation is data are always recorded discretely while SDEs are defined in continuous time. This has inspired many researches on obtaining good parameter estimates based on discrete observations.

In this paper, we consider the estimation of a scalar, time-homogeneous diffusion characterized by the following SDE

$$dX_t = a\left(X_t;\theta\right)dt + b\left(X_t;\theta\right)dW_t, \tag{1.1}$$

where X_t is an observed scalar variable, W_t is a Wiener process, $a\left(X_t;\theta\right)$ and $b\left(X_t;\theta\right)$ are parametric drift and diffusion coefficient with $p\times 1$ parameter vector θ. Given a time discretization $t_0(=0) < \cdots < t_{i-1} < t_i < \cdots < t_n\,(=T)$ and a sampling interval $\Delta = t_i - t_{i-1}$, we let $p\left(X_{t_i}|X_{t_{i-1}};\theta\right)$ denote the transition density of X_{t_i} given $X_{t_{i-1}}$. If p is known, maximum likelihood estimator (MLE) will be the first choice for efficient estimation. However, closed-form transition densities exist only for a few special SDEs and this makes MLE inapplicable to a general SDE in (1.1). Florens-Zmirou (1989) shows that estimator based on Euler approximation to (1.1) with a normal transition density will converge to the true MLE as the time discretization interval goes to zero. In practice, the discretization interval is usually larger than zero, and Euler estimator inevitably introduces approximation error. Much work has been done to improve the approximation to p. Shoji and Ozaki (1998) obtain a closed-form approximation to the transition density by using local linearization. Hermite polynomial expansion is used in Aït-Sahalia (2002) to approximate the transition density. The estimator in Aït-Sahalia (2002) yields high numerical precision and is shown in Hurn et al. (2007) to outperform many existing estimation methods from the perspective of speed/accuracy trade-off .

Simulated MLE (SMLE) and Markov chain Monte Carlo offer alternative approaches to estimation (see, e.g., Pedersen (1995), Brandt and Santa-Clara (2002), Eraker (2001), Elerian et al. (2001)). These simulation-based methods can also achieve high numerical precision but the computation cost is high. More recently, Durham and Gallant (2002) use Brownian bridge sampler to improve SMLE; Phillips and Yu (2009) propose a two-stage realized volatility approach; Beskos et al. (2009) suggest a Simultaneous Acceptance Method (SAM) by estimating each conditional likelihood independently. However, SAM is applicable only to a restricted class of diffusions. Other approaches include numerically solving Fokker-Planck equation in Lo (1988), estimation functions based on low order Wagner-Platen approximation in Kelly et al. (2004), and method-of-moments approaches in Chan et al. (1992), Gallant and Tauchen

(1997), Gouriéroux et al. (1993), Hansen and Scheinkman (1995), etc.[1] See Fan (2005), Aït-Sahalia (2007), and Hurn et al. (2007) for surveys on various estimation methods.

This paper develops quasi-maximum likelihood estimator (QMLE) by using higher order strong Wagner-Platen approximations. Due to the difficulty in obtaining closed-form approximate transition density in higher order approximations, previous research is limited to low order approximations such as Euler and Milstein schemes, and the estimates are often less precise compared to the results in Aït-Sahalia (2002) and Durham and Gallant (2002). We show that higher order approximations can improve estimation. The idea is to derive the first two conditional moments based on a strong numerical solution to (1.1), and use the QMLE in Bollerslev and Wooldridge (1992) for estimation. QMLE has the following appealing features. First, its consistency and asymptotic normality is easy to establish. Second, simulation shows that order three or four approximation will often be enough for precise estimation, and the estimator is also numerically robust. Third, our approach does not require (1.1) to be first transformed such that $b(X;\theta) = 1$, in contrast to some other existing techniques. Simulation shows QMLE obtained from untransformed SDE is also very precise. In a closely related paper, Kessler (1997) approximates the conditional moments directly while our approach is based on a numerical solution to the SDE.

The rest of the paper is organized as follows. Section 2 introduces strong Wagner-Platen approximations and QMLE. Section 3 presents simulation evidence for numerical precision and robustness of QMLE. Section 4 concludes. An example of QMLE is given in the Appendix.

2 The approximation and the estimator

Throughout this paper, we consider a general SDE defined in (1.1), where we assume discrete observations are stationary and ergodic. Extension to nonstationary and nonergodic processes is possible (see Section 2.2 for a discussion). Euler scheme is the simplest strong Wagner-Platen approximation and it takes the following form

$$X_{t_i} = X_{t_{i-1}} + a\left(X_{t_{i-1}};\theta\right)\Delta + b\left(X_{t_{i-1}};\theta\right)\sqrt{\Delta}\varepsilon_{t_i}, \tag{2.2}$$

where $\varepsilon_{t_i} \overset{i.i.d.}{\sim} N(0,1)$ for all $t_0 < t_i \leq t_n$. Equation (2.2) is an order 0.5 strong Wagner-Platen approximation and X_{t_i} in (2.2) is a numerical solution to (1.1). Euler scheme implies a conditional normal distribution with mean $X_{t_{i-1}} + a\left(X_{t_{i-1}};\theta\right)\Delta$ and standard deviation $b\left(X_{t_{i-1}};\theta\right)\sqrt{\Delta}$. Elerian (1998) uses an order 1.0 Milstein scheme approximation to obtain a closed-form density. However, closed-form approximate transition density with higher order approximations is hard to derive.

[1]Wagner-Platen approximation is also called Itô-Taylor approximation in Kloeden and Platen (1999)

An alternative way to explore time discrete approximation is proposed in Shoji and Ozaki (1998). After transforming $b(X;\theta)$ to 1, (1.1) becomes

$$dY_t = a_Y\left(Y_t;\theta\right)dt + dW_t, \tag{2.3}$$

where

$$\begin{aligned} Y &\equiv G\left(X\right) = \int^X du/b\left(u;\theta\right), \qquad (2.4)\\ a_Y\left(Y;\theta\right) &= \frac{a\left(G^{-1}\left(Y;\theta\right);\theta\right)}{b\left(G^{-1}\left(Y;\theta\right);\theta\right)} - \frac{1}{2}\frac{\partial b\left(G^{-1}\left(Y;\theta\right);\theta\right)}{\partial X}. \end{aligned}$$

We may linearize $a_Y\left(Y_{t_i};\theta\right)$ around the point $Y_{t_{i-1}}$ to obtain

$$\begin{aligned} a_Y\left(Y_{t_i};\theta\right) &\approx a_Y\left(Y_{t_{i-1}};\theta\right) + \frac{1}{2}\frac{\partial^2 a_Y}{\partial Y^2}\left(t_i - t_{i-1}\right) + \frac{\partial a_Y}{\partial Y}\left(Y_{t_i} - Y_{t_{i-1}}\right)\\ &= a_Y\left(Y_{t_{i-1}};\theta\right) - \frac{1}{2}\frac{\partial^2 a_Y}{\partial Y^2}t_{i-1} - \frac{\partial a_Y}{\partial Y}Y_{t_{i-1}} + \frac{1}{2}\frac{\partial^2 a_Y}{\partial Y^2}t_i + \frac{\partial a_Y}{\partial Y}Y_{t_i}. \qquad (2.5) \end{aligned}$$

Given $Y_{t_{i-1}}$, the first three terms on the r.h.s. of (2.5) and the coefficients for t_i and Y_{t_i} are constant, and the drift becomes a linear function of Y_{t_i} and t_i. An explicit solution to (2.3) can be obtained, and it follows a conditional normal distribution, making MLE feasible. We note that the values of the first and second derivative of the drift in Equation (2.5) may also change as Y_t evolves from t_{i-1} to t_i, and this observation motivates us to accommodate these changes to possibly improve estimation.

2.1 Wagner-Platen expansion and strong approximation

When $\partial a_Y/\partial Y$ and $\partial^2 a_Y/\partial Y^2$ in (2.5) are varying on the interval $[t_{i-1}, t_i]$, they can be further expanded when we approximate Y_t in (2.3). Differentiating $\partial a_Y/\partial Y$ gives

$$d\left(\frac{\partial a_Y}{\partial Y}\right) = \frac{1}{2}\frac{\partial^3 a_Y}{\partial Y^3}dt + \frac{\partial^2 a_Y}{\partial Y^2}dY,$$

and in discrete time, omitting θ, we have

$$\begin{aligned} \frac{\partial a_Y\left(Y_{t_i}\right)}{\partial Y} \approx\ & \frac{\partial a_Y\left(Y_{t_{i-1}}\right)}{\partial Y} - \frac{1}{2}\frac{\partial^3 a_Y\left(Y_{t_{i-1}}\right)}{\partial Y^3}t_{i-1} - \frac{\partial^2 a_Y\left(Y_{t_{i-1}}\right)}{\partial Y^2}Y_{t_{i-1}}\\ & + \frac{1}{2}\frac{\partial^3 a_Y\left(Y_{t_{i-1}}\right)}{\partial Y^3}t_i + \frac{\partial^2 a_Y\left(Y_{t_{i-1}}\right)}{\partial Y^2}Y_{t_i}. \qquad (2.6) \end{aligned}$$

If $\partial^3 a_Y/\partial Y^3$ in (2.6) also varies when Y evolves from $Y_{t_{i-1}}$ to Y_{t_i}, we can again differentiate it w.r.t. Y in approximation. In theory, if we assume a_Y is infinitely differentiable in Y, above differentiation can be continued until desired precision in approximation is reached. This way of expanding diffusion process is analogous to Taylor series expansion and is referred to as Wagner-Platen expansion in Kloeden and Platen (1999) (henceforth KP). It is applicable to a diffusion defined in (1.1) without normalizing $b(X;\theta)$ to one. Consider the solution X_t to (1.1) conditioning on $X_{t_{i-1}}$

$$X_{t_i} = X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} a\left(X_u;\theta\right) du + \int_{t_{i-1}}^{t_i} b\left(X_u;\theta\right) dW_u. \tag{2.7}$$

By Itô formula, we expand $a\left(X_u;\theta\right)$ and $b\left(X_u;\theta\right)$ in (2.7) at $X_{t_{i-1}}$ to have

$$X_{t_i} = X_{t_{i-1}} + a\left(X_{t_{i-1}};\theta\right) \int_{t_{i-1}}^{t_i} du + b\left(X_{t_{i-1}};\theta\right) \int_{t_{i-1}}^{t_i} dW_u + R,$$

where

$$\begin{aligned} R &= \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^0 a\left(X_z;\theta\right) dz du + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^1 a\left(X_z;\theta\right) dW_z du \\ &\quad + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^0 b\left(X_z;\theta\right) dz dW_u + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^1 b\left(X_z;\theta\right) dW_z dW_u, \\ L^0 &= a\frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2}, \text{ and } L^1 = b\frac{\partial}{\partial x}. \end{aligned} \tag{2.8}$$

We let $a = a\left(X_{t_{i-1}};\theta\right)$ and $b = b\left(X_{t_{i-1}};\theta\right)$ to conserve space. This expansion can be continued as long as both a and b are smooth in x. For example, we can further expand $L^0 a\left(X_z;\theta\right)$ at $X_{t_{i-1}}$ in R to obtain higher order results. The general result is summarized in Theorem 5.5.1 of KP.

The following assumptions are adapted from Section 4.5 in KP to guarantee the existence and uniqueness of a strong solution to (1.1). Let $\Re$ be the real line and $\{\mathcal{F}_t, t \geq 0\}$ be a family of σ-algebras generated by W_t for all $t \in [t_0, T]$. For all x in a compact set in $\Re$, we assume

Assumption 1. $a(x;\theta)$ and $b\left(x;\theta\right)$ are infinitely differentiable in x;

Assumption 2. For some positive constant K, we have $\left|a\left(x;\theta\right)\right|^2 \leq K^2(1 + |x|^2)$ and $\left|b\left(x;\theta\right)\right|^2 \leq K^2(1 + |x|^2)$;

Assumption 3. X_{t_0} is $\mathcal{F}_{t_0}$-measurable with $E(|X_{t_0}|^2) < \infty$.

Assumption 1 is stronger than the Lipschitz condition in KP. It offers the possibility of establishing consistency using infinite order approximation with a fixed Δ, though we assume $\Delta \to 0$ and a fixed approximation order in the paper. It also helps to establish the global Lipschitz condition in (2.14).

Next, we introduce the Wagner-Platen expansion (detailed discussion can be found in Chapter 5 of KP). Let α be a multi-index of length l such that $\alpha = (j_1, j_2, \cdots, j_l)$, $j_i \in \{0, 1\}$ for $i = 1, 2, \cdots, l$ and $l := l\left(\alpha\right) \in \{1, 2, \cdots, l\}$. Let $\mathcal{M}$ be the set of all multi-indices such that $\mathcal{M} = \{(j_1, j_2, \cdots, j_l) : j_i \in \{0, 1\}, i \in \{1, 2, \cdots, l\}, \text{ for } l = 1, 2, \cdots\} \cup \{v\}$, where v is the multi-index of length zero. For an $\alpha \in \mathcal{M}$ with $l\left(\alpha\right) \geq 1$, we let $-\alpha$ and $\alpha-$ be the multi-index in $\mathcal{M}$ obtained by deleting the first and last element of α, respectively. We define a sequence of sets for adapted right continuous stochastic processes $f(t)$ with left hand limits: let $\mathcal{H}_v$ be the totality of all processes such that $|f(t)| < \infty$, $\mathcal{H}_{(0)}$ be the totality of all processes such that $\int_{t_0}^{t} |f(u)|\, du < \infty$, $\mathcal{H}_{(1)}$ be the totality of all processes such that $\int_{t_0}^{t} |f(u)|^2 \, du < \infty$, and $\mathcal{H}_\alpha$ be the totality of adapted right continuous processes with left hand limits such that $I_{\alpha-}[f(\cdot)]_{t_{i-1},\cdot} \in \mathcal{H}_{(j_l)}$ for all $t_0 \leq t \leq T$ and $l(\alpha) \geq 2$, and the multiple Itô integral $I_\alpha [f(\cdot)]_{t_{i-1},t_i}$ is defined as

$$I_\alpha [f(\cdot)]_{t_{i-1},t_i} = \begin{cases} f(t_i) & \text{if } l = 0, \\ \int_{t_{i-1}}^{t_i} I_{\alpha-}[f(\cdot)]_{t_{i-1},u} \, du & \text{if } l \geq 1 \text{ and } j_l = 0, \\ \int_{t_{i-1}}^{t_i} I_{\alpha-}[f(\cdot)]_{t_{i-1},u} \, dW_u & \text{if } l \geq 1 \text{ and } j_l = 1, \end{cases} \tag{2.9}$$

where $t_0 \leq t_{i-1} < t_i \leq T$. For example, if $\alpha = (0,1,1,0)$, we have

$$I_{(0,1,1,0)} [f(\cdot)]_{t_{i-1},t_i} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u_4} \int_{t_{i-1}}^{u_3} \int_{t_{i-1}}^{u_2} f(\cdot) du_1 dW_{u_2} dW_{u_3} du_4.$$

We write $I_\alpha [f(\cdot)]_{t_{i-1},t_i}$ as I_α when $f(t_i) = 1$. As an example, we have

$$I_{(0,0,0)} [1]_{t_{i-1},t_i} = I_{(0,0,0)} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u_3} \int_{t_{i-1}}^{u_2} ds_1 ds_2 ds_3 = \frac{1}{3!} (t_i - t_{i-1})^3 = \frac{1}{6} \Delta^3.$$

The researcher chooses the length $l(\alpha)$ in Theorem 5.5.1 of KP to decide how many terms to include in the Wagner-Platen expansion. For example, the expansion for $l(\alpha) = 2$ is

$$\begin{aligned} X_{t_i} &= X_{t_{i-1}} + aI_{(0)} + bI_{(1)} + \left(aa' + 0.5b^2 a''\right) I_{(0,0)} \\ &\quad + \left(ab' + 0.5b^2 b''\right) I_{(0,1)} + ba' I_{(1,0)} + bb' I_{(1,1)} + R, \end{aligned} \tag{2.10}$$

where R is the remainder in expansion.[2] For each α, we define recursively the Itô coefficient function

$$f_\alpha = \begin{cases} f & \text{if } l = 0, \\ L^{j_1} f_{-\alpha} & \text{if } l \geq 1, \end{cases} \tag{2.11}$$

where L^{j_1} is defined in (2.8). If we let $f(x) \equiv x$, it is easy to verify the coefficients in (2.10). For example, $f_{(0)} = a$ and $f_{(0,1)} = ab' + 0.5b^2 b''$.

Strong Wagner-Platen approximation can be obtained based on expansions such as (2.10). Let $\Delta W = W_{t_i} - W_{t_{i-1}}$, and an example of I_α is

$$I_{(1,1)} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u_1} dW_{u_2} dW_{u_1} = 0.5((\Delta W)^2 - \Delta). \tag{2.12}$$

Replacing all stochastic integrals in (2.10) with expressions similar to (2.12), evaluating all the coefficients at $X_{t_{i-1}}$, we obtain a strong Wagner-Platen approximation when $l(\alpha) = 2$. For stochastic integrals with higher multiplicity, it is not always possible to derive a closed form in terms of ΔW and Δ, but

[2] We let a', a'', and a''' be the first, second and third derivative of $a(X;\theta)$ w.r.t. X, respectively. Let $a^{(r)}$ be the rth derivative of $a(X;\theta)$ w.r.t. X when $r \geq 4$. The same notation applies to derivatives of $b(X;\theta)$ w.r.t. X.

they can be approximated (see Section 5.8 in KP). However, we note that closed forms such as (2.12) are not needed for estimation.

A general form of strong Wagner-Platen approximation is given by

$$Y^{\Delta}(t_i) = X_{t_{i-1}} + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} I_\alpha \left[f_\alpha \left(X_{t_{i-1}} \right) \right]_{t_{i-1}, t_i}, \tag{2.13}$$

where $\mathcal{A}_\gamma = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2\gamma \text{ or } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}$, $n(\alpha)$ is the number of zeros in α, f_α is the coefficient function defined in (2.11) with $f \equiv x$ and $\gamma = 0.5, 1, 1.5, \cdots$ is the approximation order. Approximation in (2.13) is a special case of Equation (10.6.4) in KP, where we let $Y^{\Delta}(t_{i-1}) = X_{t_{i-1}}$.

Let $\mathcal{H}_\alpha$ denote the sets for multi-indices $\alpha \in \mathcal{M}$ such that $f_\alpha(x)$ is square integrable in time t for $l(\alpha) > 1$, $\mathcal{B}(\mathcal{A}_\gamma) = \{\alpha \in \mathcal{M} \setminus \mathcal{A}_\gamma : -\alpha \in \mathcal{A}_\gamma\}$, and C^2 denote the space of two times continuously differentiable functions in x.

Theorem 2.1 *Let $Y^{\Delta}(t_i)$ be the order γ strong Wagner-Platen approximation defined in* (2.13) *with $t_0 \leq t_i \leq T$ and $0 < \Delta < 1$. Under Assumptions 1 to 3, if the coefficient functions in* (2.11) *satisfy*

$$|f_\alpha(x) - f_\alpha(y)| \leq K_1 |x - y| \tag{2.14}$$

for all $\alpha \in \mathcal{A}_\gamma$ and x, y in a compact set in $\Re$; $f_{-\alpha} \in C^2$ and $f_\alpha \in \mathcal{H}_\alpha$ for all $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$; $|f_\alpha(x)| \leq K_2(1 + |x|)$ for all $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$ and x in a compact set in $\Re$; and the initial condition at t_{i-1} satisfies

$$\sqrt{E\left(\left| X_{t_{i-1}} - Y^{\Delta}(t_{i-1}) \right|^2 \right)} \leq K_3 \Delta^\gamma. \tag{2.15}$$

Then for all i and every fixed γ, we have

$$E\left(\left| X_{t_i} - Y^{\Delta}(t_i) \right| \right) \leq K_4 \Delta^\gamma \tag{2.16}$$

and

$$\lim_{\Delta \to 0} P(|Y^{\Delta}(t_i) - X_{t_i}| < \varepsilon) = 1 \text{ for every } \varepsilon > 0. \tag{2.17}$$

K_1, K_2, K_3, and K_4 are positive constants and independent of Δ.

Result (2.16) follows Corollary 10.6.4 in KP and Equation (2.17) is simply the result that convergence in the rth mean implies convergence in probability. In the Lipschitz condition of (2.14), we restrict the domain of $f_\alpha(x)$ to be a compact set in $\Re$, which rules out SDEs with explosive, unbounded solutions. The assumption is practically relevant since many observed data are bounded. Assumption 1 implies $f_\alpha(x)$ is also infinitely differentiable in x, which further implies that it is locally Lipschitz. A locally Lipschitz function with a compact domain is globally Lipschitz, which gives the condition in (2.14). $Y^{\Delta}(t_{i-1})$ in (2.15) is the first approximation on $[t_{i-1}, t_i]$. By letting $Y^{\Delta}(t_{i-1}) = X_{t_{i-1}}$, (2.15) is always satisfied.

Result (2.17) is obtained by assuming $\Delta \to 0$ with a fixed γ. Note that the same result can also be obtained by letting $\gamma \to \infty$ with Δ fixed in (2.16), provided K_4 is bounded. However, it was pointed out by a referee that, just like not all smooth function can be approximated by Taylor series expansion, it is possible that K_4 may explode as $\gamma \to \infty$. It is for this reason that we adopt the assumption of $\Delta \to 0$ with a fixed γ in this paper.

2.2 Quasi-maximum likelihood estimator

Based on the approximation in (2.13) for a fixed γ, we can proceed to obtain QMLE. Define conditional moments based on (2.13) as $\mu_{t_i,\Delta} \equiv E\left(Y^{\Delta}(t_i)\,|X_{t_{i-1}}\right)$ and $\sigma^2_{t_i,\Delta} \equiv Var\left(Y^{\Delta}(t_i)\,|X_{t_{i-1}}\right)$, and the true conditional moments are $\mu_{t_i} \equiv E\left(X_{t_i}|X_{t_{i-1}}\right)$ and $\sigma^2_{t_i} \equiv Var\left(X_{t_i}|X_{t_{i-1}}\right)$. We note that $E|X_{t_i}| < \infty$ (a consequence of Theorem 4.5.3 in KP) and $Y^{\Delta}(t_i)$ is uniformly integrable for every choice of Δ in $(0,1)$, which follows that there are a fixed number of terms on the r.h.s. of (2.13), each term is bounded when evaluated at $X_{t_{i-1}}$, and multiple Itô integrals w.r.t. time in $E(Y^{\Delta}(t_i)\,|X_{t_{i-1}})$ are all bounded. Theorem C in Section 1.4 of Serfling (1980) in conjunction with (2.17) then implies that

$$\lim_{\Delta\to 0}\mu_{t_i,\Delta} = \mu_{t_i}, \quad \lim_{\Delta\to 0}\sigma^2_{t_i,\Delta} = \sigma^2_{t_i}. \tag{2.18}$$

Hence the first two conditional moments of X_{t_i} are correctly specified as $\Delta \to 0$ and QMLE in Bollerslev and Wooldridge (1992) can be used for estimation. Details about consistency, asymptotic normality and identification of QMLE can be found in Bollerslev and Wooldridge (1992). Extension of the asymptotic results to nonstationary and nonergodic process is also possible (see Theorem 10.1 in Wooldridge (1994)).

2.3 Remarks

Remark 2.1 This paper differs from Kessler (1997) in several aspects. First, Taylor series expansion of inverse and log of the second moment is used to prove asymptotic properties (see p. 215 in Kessler (1997)) while we use the approach in Bollerslev and Wooldridge (1992). Second, this paper provides simulation results. Third, while Kessler (1997) approximates conditional moments directly, we approximate the unique strong solution to a SDE. Our approach also permits the possibility of establishing consistency using $\gamma \to \infty$ with a fixed Δ.

Remark 2.2 QMLE is generally less efficient than exact MLE (see, e.g., White (1994)). Simulation results in Tables 1, 2, and 3 nonetheless suggest that, at least for the models and parameter settings used in the simulation study, the efficiency loss of QMLE is small for both normal and non-normal data.

Remark 2.3 Expressions for conditional mean and variance used in QMLE can be easily derived using software such as *Mathematica* even for higher order approximations. Because QMLE provides a closed-form approach, it is also computationally efficient.

Remark 2.4 QMLE is applicable to a large class of parametric SDEs. Assumptions 1 and 3 are satisfied by many SDEs. For Assumption 2, consider the CIR model in Section 3.1 with positive parameters. It can be shown that the required K for $a(x;\theta) = \theta_2(\theta_1 - x)$ and $b(x;\theta) = \theta_3\sqrt{x}$ are $K \geq \max(\theta_1\theta_2, \theta_2)$ and $K \geq \theta_3/\sqrt{2}$, respectively. Assumption 2 is satisfied as long as $K \geq \max(\theta_1\theta_2, \theta_2, \theta_3/\sqrt{2})$. Note that $\theta_3\sqrt{x}$ is not differentiable at 0, but this can be resolved by replacing Assumption 2 with a weaker Yamada condition (see Theorem 3.2 in Chapter IV of Ikeda and Watanabe (1989)). If we assume the domain of $a(x;\theta)$ and $b(x;\theta)$ is compact, this assumption is easily satisfied. QMLE is also applicable to diffusions when an analytic solution to (2.4) is unavailable such as the example in Aït-Sahalia (1996):

$$dX_t = (\theta_1 + \theta_2 X_t + \theta_3 X_t^2 + \theta_4/X_t)dt + (\theta_5 + \theta_6 X_t + \theta_7 X_t^{\theta_8})dW_t.$$

This property becomes more attractive in multivariate diffusions where the transform in (2.4) is not applicable to all diffusion matrix b. QMLE can also be used when sampling intervals are unequal or random as in Yu and Phillips (2001).

Remark 2.5 Efficiency might be improved by using GMM with more moment conditions such as the third and fourth moments. Deriving higher order moments involving multiple Itô integrals is complicated. Even they are obtained, GMM is still less efficient than MLE. Compared to the GMM estimator in Hansen and Scheinkman (1995), QMLE is simple because only the first two conditional moments are needed.

3 Simulation results

In this section, we show QMLE yields high numerical precision when data are close to normal and numerically robust when data are non-normal.

3.1 Numerical precision

QMLEs with $l(\alpha) = 3$ and $l(\alpha) = 4$ in (2.13) are used. We use $l(\alpha)$ instead of the approximation order γ just for simplicity reasons. Approximation with $l(\alpha) = 3$ is given in the Appendix, and the result for $l(\alpha) = 4$ is available from the author upon request. In order to gauge the efficiency loss of QMLE, the following models with closed-form transition density are used: the Ornstein–Uhlenbeck (OU) process $dX_t = \theta_2(\theta_1 - X_t)dt + \theta_3 dW_t$, the CIR model in Cox et al. (1985) $dX_t = \theta_2(\theta_1 - X_t)dt + \theta_3\sqrt{X_t}dW_t$, and the Black-Scholes (BS) model in Black and Scholes (1973) $dX_t = \theta_2 X_t dt + \theta_3 X_t dW_t$. Exact MLE can be obtained for these processes.

In Tables 1 and 2, $\hat{\theta}^{(\mathrm{MLE})}$, $\hat{\theta}^{(\mathrm{EUL})}$, $\hat{\theta}^{(l=3)}$ and $\hat{\theta}^{(l=4)}$ correspond to exact MLE, Euler estimator, QMLE with $l(\alpha) = 3$ and $l(\alpha) = 4$, respectively. Instead of the average bias, we can report the average distance between QMLE and MLE, similar to Table III in Aït-Sahalia (2002). This average distance can be easily inferred from Tables 1 and 2. Take θ_1 in the OU model for example. We find $\hat{\theta}_1^{(\mathrm{MLE})} - \theta_1 \approx -0.0000267218$ and $\hat{\theta}_1^{(l=3)} - \theta_1 \approx -0.0000267213$. Hence the distance between $\hat{\theta}_1^{(l=3)}$ and MLE is equal to $|(\hat{\theta}_1^{(l=3)} - \theta_1) - (\hat{\theta}_1^{(\mathrm{MLE})} - \theta_1)| \approx 0.0000000005$.

Parameter values in Table 1 are selected from Table III in Aït-Sahalia (2002). Results for $\hat{\theta}^{(l=3)}$ and $\hat{\theta}^{(l=4)}$ in the CIR and BS models are obtained from transforming b to one. Table 1 suggests that QMLE outperforms Euler estimator in some cases and it is usually very close to the exact MLE. The estimates for $\hat{\theta}^{(l=3)}$ and $\hat{\theta}^{(l=4)}$ in BS model give identical results because after the transform, $Y = \ln(X)/\theta_3$, the model has constant drift and diffusion coefficient, and higher order terms in approximations are equal to zero.

In Table 2 we investigate specifically the precision of QMLE in the CIR model. Untransformed QMLEs, $\hat{\theta}^{(3,\mathrm{U})}$ and $\hat{\theta}^{(4,\mathrm{U})}$, and the estimator using all seven Hermite coefficients on page 238 of Aït-Sahalia (2002) are also reported. The data generating process (DGP) (a) is the same as that in Table 1. DGP (b) and (c) are selected from Durham and Gallant (2002) and Jensen and Poulsen (2002), respectively. In Table 2, we find that QMLE offers improvement over Euler estimator.

Higher order Hermite expansion yields negative approximate density for all DGPs in Table 2. All negative densities are replaced with a small positive number ($eps \approx 2.22 \times 10^{-16}$ in *Matlab*). After this modification, $\hat{\theta}^{(\mathrm{Hermite})}$ is also precise. $\hat{\theta}^{(\mathrm{Hermite})}$ performs poorly in DGP (b) because it is sensitive to starting values in simulated data. After discarding the first 200 observations in DGP (b), the biases of $\hat{\theta}^{(\mathrm{Hermite})}$ for $(\theta_1, \theta_2, \theta_3)$ are about -5.76E-05, 0.05, and 3.80E-05, respectively.

3.2 Numerical robustness

In the previous section, QMLE is shown to have high numerical precision and little efficiency loss. The purpose of this section is twofold: first, we show that QMLE continues to work reasonably well for certain nonlinear and non-normal diffusions; second, higher order approximations in QMLE provide improvement over the Euler scheme when sampling interval is large. Both findings reveal good numerical robustness of QMLE.[3]

In Table 3, we continue to work with the CIR model but let $\theta = (0.05, 0.3, 0.15)$. If the process takes the value at the long-run mean, $\theta_1 (= 0.05)$, the selected value for θ implies a density with a skewness of 1.7 and a kurtosis of 7.5, an enough deviation from normality for robustness test purposes. We consider a sample size of 1000 with 1000 replications and four different sampling intervals. A reasonably large bound of [0.01,10] is imposed in estimation. Table 3 reports average bias of different estimators. $\hat{\theta}^{(\mathrm{PDE})}$ is obtained from numerically solving the Fokker-Planck equation using finite difference method. Details of $\hat{\theta}^{(\mathrm{PDE})}$ can be found in Hurn et al. (2007).

[3] Discussion in the section is motivated by a referee's comment on numerical robustness of various existing estimators. I am very grateful to his/her strong insight on this subject.

In Table 4, the following DGP in Shoji and Ozaki (1998) is used

$$dX_t = (\theta_1 + \theta_2 X_t + \theta_3 X_t^2 + \theta_4 X_t^3)dt + \theta_5 X_t^{\theta_6} dW_t, \tag{3.19}$$

where $\theta = (6, -11, 6, -1, 1, 0.5)$. Since there is no closed-form transition density for (3.19) , we simulate data at an interval of 0.005 based on Euler scheme, and use four different sampling intervals with 1000 observations and 1000 replications. Reasonably large bounds are imposed in estimation: [0.01, 50] or [-50, -0.01] for θ_1 to θ_3; [-10, -0.01] or [0.01, 10] for θ_4 and θ_5; [0, 1] for θ_6. Neither $\hat{\theta}^{(\mathrm{EUL})}$ nor QMLE hits any bound in estimation. By construction, the simulated data are already in favor of the Euler estimator, yet this does not rescue the Euler scheme in estimation.

Remark 3.1 In Table 3, QMLE $\hat{\theta}^{(\mathrm{l=3,U})}$ provides some improvement over $\hat{\theta}^{(\mathrm{EUL})}$, but $\hat{\theta}^{(\mathrm{EUL})}$ is generally better than $\hat{\theta}^{(\mathrm{l=4,U})}$. In Table 4, there is a very clear pattern that higher order QMLEs outperform $\hat{\theta}^{(\mathrm{EUL})}$ as sampling interval increases. Results in Tables 1 to 4 find some, but not overwhelming evidence that higher order QMLEs improve estimation. One interesting observation in Tables 1 and 2 is the distance between QMLE and MLE, measured by difference in biases, is usually smaller than the distance between $\hat{\theta}^{(\mathrm{EUL})}$ and MLE, offering additional evidence that higher order estimators might improve efficiency. We also note that Fan and Zhang (2005) find higher order nonparametric estimators reduces bias but increases variances.

Remark 3.2 Hermite polynomials produce negative densities for DGPs in Tables 3 and 4, and all negative densities are replaced with a small positive number in estimation. The corresponding log-likelihood function tends to have a much larger value on the boundary than in the neighborhood of θ. To prevent estimates from taking values on the boundary, all estimates in Tables 3 and 4 are obtained from a local search algorithm instead of a global algorithm with multiple searches. After these modifications, $\hat{\theta}^{(\mathrm{Hermite})}$ is also precise.

Remark 3.3 $\hat{\theta}^{(\mathrm{PDE})}$ requires recursively solving a tri-diagonal system and uses several *for*-loops in *Matlab*, and the optimization is extremely slow. Instead, the results for $\hat{\theta}^{(\mathrm{PDE})}$ in Table 3 are obtained using the C++ code in Hurn et al. (2007). Whenever the finite difference method produces a nonpositive density, it is replaced with a small positive number (1.0E-15). After this modification, $\hat{\theta}^{(\mathrm{PDE})}$ gives good results in Table 3. However, $\hat{\theta}^{(\mathrm{PDE})}$ fails to obtain sensible estimates in Table 4 and is not reported. This is due to the particular way data are generated. Note that in estimation we let the unit of discretization of state space be $\Delta x = 0.001$ and that of time be $\Delta/10$, respectively. When $\Delta = 0.05$, $\Delta/10 = 0.005$, which is exactly equal to the interval used in simulating data in (3.19) . This poses difficulty for the finite difference method because the data are normal on this interval (0.005), but the model we try to fit with is non-normal. Even if we find a way to make the method work, computation cost remains a big concern. For a sample of 1000 observations from (3.19) and on a

Athlon 2.91 GHz desktop with 4 Gb RAM, the computation time in *Matlab* for $\hat{\theta}^{(l=3,\mathrm{U})}$, $\hat{\theta}^{(l=4,\mathrm{U})}$, and $\hat{\theta}^{(\mathrm{Hermite})}$ (with seven Hermite polynomial coefficients) is about 4 seconds, 11 seconds, and 16 seconds, respectively. Using C++ code, the computation time for $\hat{\theta}^{(\mathrm{PDE})}$ is about 5000 seconds without producing a sensible result. To possibly improve estimation, a finer discretization in both x and time is needed, which will further increase the computation cost. Similar results are found in Table 5 of Hurn et al. (2007), where $\hat{\theta}^{(\mathrm{Hermite})}$ is more than 3000 times faster than $\hat{\theta}^{(\mathrm{PDE})}$ in a four-parameter model. For a six-parameter model in (3.19) , we conjecture that other estimators will be chosen over $\hat{\theta}^{(\mathrm{PDE})}$ because of computation cost.

Our limited simulation results suggest that the proposed QMLE works better when both the drift and diffusion coefficient are (highly) nonlinear.

4 Conclusion

This paper introduces QMLE for discretely observed diffusions. The estimator is based on higher order numerical solutions to a SDE, and it is applicable to a large class of parametric diffusions. Simulation study reveals some good finite sample properties. In summary, QMLE is conceptually simple, computationally efficient, and numerically precise and robust. Extending the current method to multivariate diffusions is straightforward, and QMLE works without transforming diffusion matrix to an identity matrix (see Huang (2010)). It will be interesting to investigate the efficiency gain by using higher order approximations in SMLE. We leave it for future research.

References

[1] Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies 9*, 385–426.

[2] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach. *Econometrica 70*, 223–62.

[3] Aït-Sahalia, Y. (2007). Estimating continuous-time models using discretely sampled data. In R. Blundell, W. K. Newey, and P. Torsten (Eds.), *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Volume 3*, 261–327. Cambridge: Cambridge University Press.

[4] Beskos, A., O. Papaspiliopoulos, and G. Roberts (2009). Monte-Carlo maximum likelihood estimation for discretely observed diffusion processes. *Annals of Statistics 37*, 223–45.

[5] Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 637–54.

[6] Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*, 143–72.

[7] Brandt, M. W. and P. Santa-Clara (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics 63*, 161–210.

[8] Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*, 1209–28.

[9] Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

[10] Durham, G. B. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics 20*, 297–316.

[11] Elerian, O. (1998). A note on the existence of a closed form conditional transition density for the Milstein scheme. Working paper, OxFord University.

[12] Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69*, 959–93.

[13] Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics 19*, 177–91.

[14] Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics (with discussion). *Statistical Science 20*, 316–40.

[15] Fan, J. and C. Zhang (2003). A re-examination of diffusion estimations with applications to financial model validation. *Journal of the American Statistical Association 98*, 118–34.

[16] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics 20*, 547–57.

[17] Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, 85–118.

[18] Gallant, A. R. and G. Tauchen (1997). Estimation of continuous-time models for stock returns and interest rates. *Macroeconomic Dynamics 1*, 135–68.

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

[20] Huang, X. (2010). Quasi-maximum likelihood estimation of multivariate diffusions. Working paper, Kennesaw State University.

[21] Hurn, A. S., J. I. Jeisman, and K. A. Lindsay (2007). Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics 5*, 390–455.

[22] Ikeda, N. and S. Watanabe. (1989). *Stochastic Differential Equations and Diffusion Processes* (2nd ed.). New York: North-Holland.

[23] Jensen, B. and R. Poulsen (2002). Transition densities of diffusion processes: numerical comparison of approximation techniques. *Journal of Derivatives 9*, 18–32.

[24] Kelly, L., E. Platen, and M. Sørensen (2004). Estimation for discretely observed diffusions using transform functions. *Journal of Applied Probability 41*, 99–118.

[25] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics 24*, 211–29.

[26] Kloeden, P. and E. Platen (1999). *Numerical Solution of Stochastic Differential Equations* (corrected 3rd printing). New York: Springer-Verlag.

[27] Lo, A. W. (1988). Maximum likelihood estimation of generalized Ito processes with discretely sampled data. *Econometric Theory 24*, 231–47.

[28] Pedersen, A. R. (1995). A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics 22*, 55–71.

[29] Phillips, P. C. B. and J. Yu (2009). A two-stage realized volatility approach to estimation of diffusion processes with discrete data. *Journal of Econometrics 150*, 139–50.

[30] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

[31] Shoji, I. and T. Ozaki (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications 16*, 733–52.

[32] Yu, J. and P. C. B. Phillips (2001). A Gaussian approach for continuous time models of the short-term interest rate. *Econometrics Journal 4*, 210–24.

[33] White, H. (1994). *Estimation, Inference and Specification Analysis.* Cambridge: Cambridge University Press.

[34] Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2639–738. Amsterdam: North-Holland.

Appendix A: Proof of Theorem 2.1

Proof of Theorem 2.1: Equation (2.16) is a special case of the uniform convergence result in Corollary 10.6.4 in KP. Using Chebychev's inequality, as $\Delta \to 0$, we obtain

$$P\left(\left|X_{t_i} - Y^{\Delta}(t_i)\right| > \varepsilon\right) \leq \varepsilon^{-1} E\left(\left|X_{t_i} - Y^{\Delta}(t_i)\right|\right) \leq \varepsilon^{-1} K_4 \Delta^{\gamma} \to 0$$

which implies $Y^{\Delta}(t_i) \to X_{t_i}$ in probability in (2.17). □

Appendix B: An example of QMLE

Recall f_α is the coefficient function defined in (2.11). When evaluated at $X_{t_{i-1}}$, f_α can be taken outside the integral. The following approximation when $l(\alpha)$ = 3 is obtained from Section 5.5 in KP

$$\begin{aligned} Y^{\Delta}(t_i) &= X_{t_{i-1}} + f_{(0)}I_{(0)} + f_{(1)}I_{(1)} + f_{(0,0)}I_{(0,0)} + f_{(0,1)}I_{(0,1)} + f_{(1,0)}I_{(1,0)} \\ &\quad + f_{(1,1)}I_{(1,1)} + f_{(0,0,0)}I_{(0,0,0)} + f_{(0,0,1)}I_{(0,0,1)} + f_{(0,1,0)}I_{(0,1,0)} \\ &\quad + f_{(0,1,1)}I_{(0,1,1)} + f_{(1,0,0)}I_{(1,0,0)} + f_{(1,0,1)}I_{(1,0,1)} + f_{(1,1,1)}I_{(1,1,1)}, \end{aligned}$$

where $f_{(0)} = a$, $f_{(1)} = b$, $f_{(0,0)} = aa' + 0.5b^2a''$, $f_{(0,1)} = ab' + 0.5b^2b''$, $f_{(1,0)} = ba'$, $f_{(1,1)} = bb'$,

$$\begin{aligned} f_{(0,0,0)} &= a(aa'' + (a')^2 + bb'a'' + 0.5b^2a''') \\ &\quad + 0.5b^2(aa''' + 3a'a'' + ((b')^2 + bb'')a'' + 2bb'a''') + 0.25b^4a^{(4)}, \\ f_{(0,0,1)} &= a\left(a'b' + ab'' + bb'b'' + 0.5b^2b'''\right) \\ &\quad + 0.5b^2(a''b' + 2a'b'' + ab''' + ((b')^2 + bb'')b'' + 2bb'b'' + 0.5b^2b^{(4)}), \\ f_{(0,1,0)} &= a\left(b'a' + ba''\right) + 0.5b^2\left(b''a' + 2b'a'' + ba'''\right), \\ f_{(0,1,1)} &= a((b')^2 + bb'') + 0.5b^2\left(b''b' + 2bb'' + bb'''\right), \\ f_{(1,0,0)} &= b(aa'' + (a')^2 + bb'a'' + 0.5b^2a'''), \\ f_{(1,0,1)} &= b\left(ab'' + a'b' + bb'b'' + 0.5b^2b'''\right), \\ f_{(1,1,0)} &= b\left(a'b' + a''b\right), \\ f_{(1,1,1)} &= b((b')^2 + bb''). \end{aligned}$$

The conditional expectation and variance of $Y^{\Delta}(t_i)$ are

$$
\begin{aligned}
\mu_{t_i,\Delta} &= X_{t_{i-1}} + f_{(0)}\Delta + f_{(0,0)}\Delta^2/2 + f_{(0,0,0)}\Delta^3/6, \\
\sigma^2_{t_i,\Delta} &= f^2_{(1)}\Delta + (f_{(0,1)}f_{(1)} + f_{(1)}f_{(1,0)} + f^2_{(1,1)}/2)\Delta^2 + (f^2_{(0,1)}/3 + f_{(0,0,1)}f_{(1)}/3 \\
&\quad + f_{(0,1,0)}f_{(1)}/3 + f_{(0,1)}f_{(1,0)}/3 + f^2_{(1,0)}/3 + f_{(1)}f_{(1,0,0)}/3 + f_{(0,1,1)}f_{(1,1)}/3 \\
&\quad + f_{(1,0,1)}f_{(1,1)}/3 + f_{(1,1)}f_{(1,1,0)}/3 + f^2_{(1,1,1)}/6)\Delta^3 + (f_{(0,0,1)}f_{(0,1)}/4 \\
&\quad + f_{(0,1)}f_{(0,1,0)}/6 + f^2_{(0,1,1)}/12 + f_{(0,0,1)}f_{(1,0)}/12 + f_{(0,1,0)}f_{(1,0)}/6 \\
&\quad + f_{(0,1)}f_{(1,0,0)}/12 + f_{(1,0)}f_{(1,0,0)}/4 + f_{(0,1,1)}f_{(1,0,1)}/12 + f^2_{(1,0,1)}/12 \\
&\quad + f_{(0,1,1)}f_{(1,1,0)}/12 + f_{(1,0,1)}f_{(1,1,0)}/12 + f^2_{(1,1,0)}/12)\Delta^4 + (f^2_{(0,0,1)}/20 \\
&\quad + f_{(0,0,1)}f_{(0,1,0)}/20 + f^2_{(0,1,0)}/30 + f_{(0,0,1)}f_{(1,0,0)}/60 + f_{(0,1,0)}f_{(1,0,0)}/20 \\
&\quad + f^2_{(1,0,0)}/20)\Delta^5,
\end{aligned}
$$

where terms such as $Var(I_{(1,0)})$ and $Cov\left(I_{(1,0,1)}, I_{(1,1)}\right)$ are calculated using Lemma 5.7.2 in KP. Finally, a normal density function is used for estimation.

Table 1: Estimated bias and standard error for OU, CIR, and BS models.

		OU model	CIR model	BS model
$\hat{\theta}_1^{(\mathrm{MLE})} - \theta_1$	Bias	-0.0000267218	-0.00010845	N/A
	S.E.	0.0064456224	0.00799336	N/A
$\hat{\theta}_1^{(\mathrm{EUL})} - \theta_1$	Bias	-0.0000267214	-0.00010819	N/A
	S.E.	0.0064455361	0.00802514	N/A
$\hat{\theta}_1^{(l=3)} - \theta_1$	Bias	-0.0000267213	-0.00010906	N/A
	S.E.	0.0064457349	0.00802198	N/A
$\hat{\theta}_1^{(l=4)} - \theta_1$	Bias	-0.0000267217	-0.00011320	N/A
	S.E.	0.0064456897	0.00806676	N/A
$\hat{\theta}_2^{(\mathrm{MLE})} - \theta_2$	Bias	0.0484283531	0.05311037	-0.00007343
	S.E.	0.1195788793	0.11960684	0.03300022
$\hat{\theta}_2^{(\mathrm{EUL})} - \theta_2$	Bias	0.0355759566	0.04013549	0.00164653
	S.E.	0.1136035832	0.11629094	0.03352989
$\hat{\theta}_2^{(l=3)} - \theta_2$	Bias	0.0484253876	0.05158365	-0.00007343
	S.E.	0.1195557966	0.11967522	0.03300232
$\hat{\theta}_2^{(l=4)} - \theta_2$	Bias	0.0484283861	0.05140656	-0.00007343
	S.E.	0.1194060203	0.11978155	0.03300232
$\hat{\theta}_3^{(\mathrm{MLE})} - \theta_3$	Bias	0.0000480238	0.00011431	0.00004147
	S.E.	0.0006872011	0.00344463	0.00670272
$\hat{\theta}_3^{(\mathrm{EUL})} - \theta_3$	Bias	-0.0006258604	-0.00236893	0.00566881
	S.E.	0.0006562896	0.00341859	0.00703084
$\hat{\theta}_3^{(l=3)} - \theta_3$	Bias	0.0000478822	0.00010803	0.00004147
	S.E.	0.0006871695	0.00344402	0.00670271
$\hat{\theta}_3^{(l=4)} - \theta_3$	Bias	0.0000480253	0.00009957	0.00004147
	S.E.	0.0006871488	0.00344323	0.00670271

Note: Bias and s.e. reported in Table 1 are averages over 5000 replications with a sample size of 1000 and $\Delta = 1/12$. Parameter values used Table 1 are the same as those in Table III of Aït-Sahalia (2002). We let $\theta = (0.06, 0.5, 0.03)$ for OU model, $\theta = (0.06, 0.5, 0.15)$ for CIR model, and $\theta = (N/A, 0.2, 0.3)$ for BS model. QMLEs with $l = 3$ and $l = 4$ in BS model are the same because both drift and diffusion coefficient are constant after transforming the diffusion coefficient to one.

Table 2: Estimated bias and standard error for CIR model.

		DGP (a)	DGP (b)	DGP (c)
$\hat{\theta}_1^{(\mathrm{MLE})} - \theta_1$	Bias	-0.00010845	-0.00000835	0.00044849
	S.E.	0.00799336	0.00166959	0.01086030
$\hat{\theta}_1^{(\mathrm{EUL})} - \theta_1$	Bias	-0.00010819	-0.00000641	0.00044838
	S.E.	0.00802514	0.00166950	0.01086525
$\hat{\theta}_1^{(l=3)} - \theta_1$	Bias	-0.00010906	-0.00000833	0.00044797
	S.E.	0.00802198	0.00166956	0.01085883
$\hat{\theta}_1^{(l=4)} - \theta_1$	Bias	-0.00011320	-0.00000832	0.00044799
	S.E.	0.00806676	0.00166956	0.01085778
$\hat{\theta}_1^{(l=3,\mathrm{U})} - \theta_1$	Bias	-0.00010537	-0.00000685	0.00044099
	S.E.	0.00808349	0.00166954	0.01088451
$\hat{\theta}_1^{(l=4,\mathrm{U})} - \theta_1$	Bias	-0.00010512	-0.00000685	0.00044098
	S.E.	0.00808601	0.00166954	0.01088446
$\hat{\theta}_1^{(\mathrm{Hermite})} - \theta_1$	Bias	-0.00011272	-0.00031291	0.00044853
	S.E.	0.00808862	0.00268678	0.01085836
$\hat{\theta}_2^{(\mathrm{MLE})} - \theta_2$	Bias	0.05311037	0.00150107	0.04439189
	S.E.	0.11960684	0.02001636	0.07823301
$\hat{\theta}_2^{(\mathrm{EUL})} - \theta_2$	Bias	0.04013549	-0.00882021	0.04086283
	S.E.	0.11629094	0.01912200	0.07615430
$\hat{\theta}_2^{(l=3)} - \theta_2$	Bias	0.05158365	0.00150104	0.04437409
	S.E.	0.11967522	0.01993963	0.07787763
$\hat{\theta}_2^{(l=4)} - \theta_2$	Bias	0.05140656	0.00150127	0.04437507
	S.E.	0.11978155	0.01993980	0.07783932
$\hat{\theta}_2^{(l=3,\mathrm{U})} - \theta_2$	Bias	0.05177577	0.00152367	0.04432115
	S.E.	0.12416308	0.01994102	0.07826028
$\hat{\theta}_2^{(l=4,\mathrm{U})} - \theta_2$	Bias	0.05171348	0.00152528	0.04432130
	S.E.	0.12423279	0.01994075	0.07832116
$\hat{\theta}_2^{(\mathrm{Hermite})} - \theta_2$	Bias	0.05269981	-0.15188206	0.04439189
	S.E.	0.11914530	0.00670097	0.07772617
$\hat{\theta}_3^{(\mathrm{MLE})} - \theta_3$	Bias	0.00011431	-0.00001566	0.00006320
	S.E.	0.00344463	0.00067055	0.00199882
$\hat{\theta}_3^{(\mathrm{EUL})} - \theta_3$	Bias	-0.00236893	-0.00065034	-0.00088656
	S.E.	0.00341859	0.00065642	0.00196946
$\hat{\theta}_3^{(l=3)} - \theta_3$	Bias	0.00010803	-0.00001574	0.00006326
	S.E.	0.00344402	0.00067053	0.00199856
$\hat{\theta}_3^{(l=4)} - \theta_3$	Bias	0.00009957	-0.00001566	0.00006334
	S.E.	0.00344323	0.00067053	0.00199854
$\hat{\theta}_3^{(l=3,\mathrm{U})} - \theta_3$	Bias	0.00008891	-0.00001616	0.00006536
	S.E.	0.00349839	0.00067077	0.00200521
$\hat{\theta}_3^{(l=4,\mathrm{U})} - \theta_3$	Bias	0.00008640	-0.00001607	0.00006540
	S.E.	0.00349792	0.00067077	0.00200525
$\hat{\theta}_3^{(\mathrm{Hermite})} - \theta_3$	Bias	0.00011274	0.00266855	0.00006320
	S.E.	0.00344853	0.00065758	0.00200082

Note: Bias and s.e. reported in Table 2 are averages over 5000 replications with a sample size of 1000 and $\Delta = 1/12$. QMLEs with superscript U are obtained from untransformed model. DGP (a): $\theta = (0.06, 0.5, 0.15)$. DGP (b): $\theta = (0.06, 0.5, 0.03)$. DGP (c): $\theta = (0.08, 0.24, 0.08838)$.

Table 3: Estimated bias for CIR model in Section 3.2.

Sampling		$dX_t = \theta_2(\theta_1 - X_t)dt + \theta_3 X_t^{0.5} dW_t$		
interval	Bias	$\theta_1 = 0.05$	$\theta_2 = 0.3$	$\theta_3 = 0.15$
$\Delta = 0.05$	$\hat{\theta}^{(\mathrm{MLE})} - \theta$	0.000581	0.080117	0.000161
	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	0.000690	0.076252	0.000219
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	0.000810	0.072305	-0.000001
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.000897	0.077676	-0.000171
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	0.001219	0.078511	0.000394
	$\hat{\theta}^{(\mathrm{PDE})} - \theta$	0.000119	0.150129	-0.000970
$\Delta = 0.1$	$\hat{\theta}^{(\mathrm{MLE})} - \theta$	0.000172	0.041170	0.000196
	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	0.004525	0.036714	0.000435
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	0.000567	0.012868	-0.000203
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.000673	0.046490	-0.000667
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	0.001168	0.024282	0.000809
	$\hat{\theta}^{(\mathrm{PDE})} - \theta$	-0.000263	0.097011	-0.001051
$\Delta = 0.15$	$\hat{\theta}^{(\mathrm{MLE})} - \theta$	0.000231	0.023558	0.000267
	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	0.000232	0.014127	0.000603
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	0.000531	0.019219	-0.000521
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.001046	0.029633	-0.001198
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	0.002367	0.000857	0.001525
	$\hat{\theta}^{(\mathrm{PDE})} - \theta$	0.000368	0.071206	-0.001030
$\Delta = 0.2$	$\hat{\theta}^{(\mathrm{MLE})} - \theta$	0.000095	0.018427	0.000260
	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	0.001964	0.005736	0.000835
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	0.000490	0.012888	-0.000710
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.001207	0.024207	-0.001653
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	0.003693	-0.007201	0.001950
	$\hat{\theta}^{(\mathrm{PDE})} - \theta$	0.000387	0.063130	-0.001124

Note: All biases are averages of 1000 replication for a sample of 1000 observations. QMLEs with superscript U are obtained without transforming the diffusion coefficient to one.

Table 4: Estimated bias for the model in Equation (3.19).

Sampling interval	Bias	$dX_t = (\theta_1 + \theta_2 X_t + \theta_3 X_t^2 + \theta_4 X_t^3)dt + \theta_5 X_t^{\theta_6} dW_t$					
		$\theta_1 = 6$	$\theta_2 = -11$	$\theta_3 = 6$	$\theta_4 = -1$	$\theta_5 = 1$	$\theta_6 = 0.5$
$\Delta = 0.05$	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	-0.5143	1.2670	-0.8066	0.1299	-0.0396	0.0049
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	0.4113	-0.6454	0.2946	-0.0545	0.0059	-0.0044
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.5674	-1.0410	0.5768	-0.1116	0.0077	-0.0032
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	0.7068	-1.2189	0.7238	-0.1563	0.0041	-0.0051
$\Delta = 0.1$	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	-1.3575	3.0268	-1.8406	0.3085	-0.0731	0.0110
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	-0.1419	0.6002	-0.4991	0.0940	-0.0036	0.0010
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	0.2454	-0.2705	0.0417	-0.0025	0.0077	-0.0135
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	-0.0967	0.4727	-0.3364	-0.0156	-0.0090	0.0040
$\Delta = 0.15$	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	-1.9770	4.2979	-2.5724	0.4317	-0.0979	0.0094
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	-0.8969	2.2662	-1.5219	0.2762	-0.0212	0.0194
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	-0.4304	1.2441	-0.9217	0.1739	0.0086	-0.0477
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	-1.2150	2.8853	-1.6679	0.2592	-0.0110	0.0152
$\Delta = 0.2$	$\hat{\theta}^{(\mathrm{EUL})} - \theta$	-2.4530	5.2669	-3.1276	0.5246	-0.1169	0.0029
	$\hat{\theta}^{(l=3,\mathrm{U})} - \theta$	-1.3886	3.7118	-2.4003	0.4241	-0.0390	0.0471
	$\hat{\theta}^{(l=4,\mathrm{U})} - \theta$	-1.1086	2.7116	-1.8218	0.3333	0.0109	-0.0838
	$\hat{\theta}^{(\mathrm{Hermite})} - \theta$	-1.9597	4.2395	-2.5136	0.3716	-0.0125	0.0190

Note: Bias in Table 4 is the average over 1000 replications for a sample size of 1000. QMLEs with superscript U are obtained without transforming the diffusion coefficient to one.