Quasi-maximum Likelihood Estimation of Discretely Observed Diffusions*

Xiao Huang
Kennesaw State University
November 2008

Abstract

This paper introduces quasi-maximum likelihood estimator for discretely observed diffusions when a closed-form of the transition density is unavailable. I discretize the diffusions using higher-order strong Wagner-Platen approximations and conditional normal likelihood function is used for estimation. The estimator is shown to be consistent and has asymptotic normal distribution when the order of approximation goes to infinity. Monte Carlo study shows higher-order approximations greatly improve the precision of parameter estimates. This method is applicable to diffusions without Doss transform and also yields precise estimates for untransformed diffusions. Precision of the proposed estimator is also compared with the Hermite polynomial expansion estimator in Aït-Sahalia (2002).

Key Words: Quasi-maximum likelihood estimator, diffusion, Wagner-Platen approximation.

JEL Classifications: C13.

The author thanks Eckhard Platen, Yongjun Tang for helpful comments and suggestions. All remaining errors are mine. Correspondence address: Department of Economics & Finance, Box #403, Kennesaw State University, Kennesaw, GA 30144, U.S.A. Phone: (770)423-6318. Fax: (770)499-3209. 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 modelling gene changes due to natural selection in genetics, vertical motion of the ground level in seismology, outflow from a reservoir in hydrology and asset prices in finance. When the drift and diffusion coefficient of a SDE are parametrically specified, it is crucial to obtain precise parameter estimates. A central question often poses difficulty for estimation is data are always recorded discretely while SDEs are defined in continuous time. This difficulty have inspired many researches on obtaining efficient parameter estimates based on discrete observations.

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

$$dX_t = a(X_t; \theta) \, dt + b(X_t; \theta) \, dW_t,$$

where $X_t$ is a scalar state variable, $W_t$ is a Wiener process, $a(X_t; \theta)$ and $b(X_t; \theta)$ are known parametric drift and diffusion coefficient with $p \times 1$ parameter vector $\theta$. Given a time discretization $t_0 < \cdots < t_{i-1} < t_i < \cdots < T$ and a sampling interval $\Delta = t_i - t_{i-1}$, we let $p^*(X_{t_i}|X_{t_{i-1}}; \theta)$ denote the transition density of $X_{t_i}$ given $X_{t_{i-1}}$. The sampling interval may be a very small random or nonrandom positive number but is never equal to zero in practice. If we know the transition density, maximum likelihood estimator (MLE) will be our first choice for efficient estimation. However, it is well-known that closed-form transition densities exist only for a few special SDEs and this makes MLE inapplicable to a general SDE defined in (1). If continuous data were available, Florens-Zmirou (1989) shows that estimator based on Euler approximation of (1) with a normal transition density will converge to the
true MLE as the time discretization interval goes to zero. In practice, Euler approximation may be inaccurate and discretization interval is usually larger than zero. Hence the Euler estimator inevitably introduces both approximation and discretization error. Much work has been done to improve the approximation of the unknown transition density. Elerian (1998) derives a noncentral chi-squared density based on Milstein scheme, an order 1.0 strong Wagner-Platen scheme\(^1\). Kelly et al. (2004) also use lower order Wagner-Platen approximation and transform function in estimation. Shoji and Ozaki (1998) obtain a closed-form 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 (MCMC) offer alternative approaches to parameter estimation (see, e.g., Pederson (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) propose to use Brownian bridge sampler in estimation and greatly improve the performance of SMLE; Phillips and Yu (2008) propose a two-stage realized volatility approach; Beskos et al. (2008) suggest a Simultaneous Acceptance Method (SAM) by estimating each conditional likelihood independently. However, the SAM is applicable only to a restricted class of diffusion processes. Other approaches include numerically solving Fokker-Planck equation in Lo (1988) and method-of-moments approaches in Chan

\(^1\)Wagner-Platen approximation is also called Itô-Taylor approximation in Kloeden and Platen (1999)
et al. (1992), Gallant and Tauchen (1997), Gouriéroux et al. (1993), Hansen and Scheinkman (1995), etc. See Aït-Sahalia (2007) and Hurn et al. (2007) for surveys of various estimation methods.

This paper develops quasi-maximum likelihood estimator (QMLE) for parameters in (1) based on higher order strong Wagner-Platen approximations. Due to the difficulty in obtaining closed-form transition density for higher order strong Wagner-Platen approximations, previous research is limited to lower 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 in this paper that higher order approximations can substantially improve the precision of the estimates. The idea is to first use higher order strong approximation in Kloeden and Platen (1999) to obtain a solution to (1). By treating the sum of all stochastic terms in approximations as a normal random variable, we directly apply QMLE in White (1982, 1994) to obtain parameter estimates. The conditional mean and variance used in estimation are obtained from higher order approximations. The QMLE has the following appealing features. First, its consistency and asymptotic normality is easy to establish. Second, its consistency requires the order of Wagner-Platen approximation to be infinity but not sampling interval to go to zero. Third, simulation shows single-digit order of approximation will be enough for precise estimation. Fourth, QMLE does not require Doss transformation. Namely, normalizing the diffusion coefficient, $b (X_t; \theta)$, to one is unnecessary for estimation. Simulation shows QMLE obtained from untransformed SDE is also very precise. This makes QMLE applicable to SDEs even when analytic forms of Doss transform is unavailable. Comparison of QMLE with hermite polynomial expansion method in Aït-Sahalia (2002) is discussed in Section 3.
The rest of the paper is organized as follows. Section 2 introduces strong Wagner-Platen approximations and QMLE. Section 3 presents some simulation results based on the most commonly used models in finance. Section 4 concludes.

2 The approximation and the estimator

Throughout this paper, we consider a general SDE defined in (1), where we also implicitly assume \( \theta \) is restricted in a way such that discrete observations from (1) are stationary and ergodic\(^2\). Time discrete approximation is often used to approximate the transition density when its closed-form is unavailable. Euler scheme is the simplest strong Wagner-Platen approximation and it takes the following form

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

where \( \varepsilon_{t_i} \) \( i.i.d. \) \( N(0,1) \) for all \( t_0 < t_i \leq T \). Euler scheme can be interpreted as order 0.5 strong Wagner-Platen approximation according to Kloeden and Platen (1999) and \( X_{t_i} \) in (2) is a numerical solution to the diffusion process in (1). It is found in many simulation studies that (2) gives good approximation to the underlying continuous process when both drift and diffusion coefficient are constant or nearly constant. For varying drift and diffusion coefficient, Euler scheme may not give satisfactory results. In the context of estimation, (2) can be used to derive the transition density of \( X_{t_i} \) conditioning on \( X_{t_{i-1}} \) and (2) implies a normal density function with mean \( X_{t_{i-1}} + a \left( X_{t_{i-1}}; \theta \right) \Delta \) and standard deviation (s.d.) \( b \left( X_{t_{i-1}}; \theta \right) \sqrt{\Delta} \).

\(^2\)Practically speaking, estimation and inference in this paper will not change for nonstationary and nonergodic processes. However, a different set of proof is needed for asymptotic properties of the estimator. See a discussion below Theorem 1 in Section 2.2.
Elerian (1998) uses an order 1.0 strong Wagner-Platen approximation, also called Milstein scheme, to obtain the approximate transition density. The Milstein scheme is

\[ X_{t_i} = X_{t_{i-1}} + a \left( X_{t_{i-1}}; \theta \right) \Delta + b \left( X_{t_{i-1}}; \theta \right) \sqrt{\Delta \varepsilon_t} \]

\[ + \frac{1}{2} b \left( X_{t_{i-1}}; \theta \right)' b \left( X_{t_{i-1}}; \theta \right) \left[ (\sqrt{\Delta \varepsilon_t})^2 - \Delta \right], \tag{3} \]

and the transition density associated with (3) is a noncentral chi-squared distribution. Monte Carlo simulation in Elerian (1998) shows reduced estimation bias based on Milstein scheme. However, closed-form transition density with higher order approximations is hard to obtain. An alternative way to explore time discrete approximation is proposed in Shoji and Ozaki (1998). After Doss transform, (1) becomes

\[ dY_t = a_Y (Y_t; \theta) \, dt + dW_t, \tag{4} \]

where

\[ Y = G(X) = \int^X du / b(u; \theta), \tag{5} \]

\[ a_Y (Y; \theta) = \frac{a (G^{-1}(Y; \theta); \theta)}{b (G^{-1}(Y; \theta); \theta)} - \frac{1}{2} b \left( G^{-1}(Y; \theta); \theta \right)' \cdot \]

Their local linearization method assumes \( \partial a_Y / \partial Y \) and \( \partial^2 a_Y / \partial Y^2 \) are both constant on \([t_{i-1}, t_i]\) when discretization interval is small. Upon this assumption, differentiation of \( a_Y (Y_t; \theta) \) w.r.t. \( Y \) gives

\[ a_Y (Y_{t_i}; \theta) = a_Y (Y_{t_{i-1}}; \theta) + \frac{1}{2} \frac{\partial^2 a_Y}{\partial Y^2} (t_i - t_{i-1}) + \frac{\partial a_Y}{\partial Y} (Y_{t_i} - Y_{t_{i-1}}) \]

\[ = a_Y (Y_{t_{i-1}}; \theta) - \frac{1}{2} \frac{\partial^2 a_Y}{\partial Y^2} t_{i-1} \frac{\partial a_Y}{\partial Y} Y_{t_{i-1}} + \frac{1}{2} \frac{\partial^2 a_Y}{\partial Y^2} t_i + \frac{\partial a_Y}{\partial Y} Y_{t_i}. \tag{6} \]

Conditioning on \( Y_{t_{i-1}} \), the first three terms on the r.h.s. of (6) and the coefficients for \( t_i \) and \( Y_{t_i} \) are constant and the drift becomes a linear function
of the state variable $Y$ and $t$. An explicit solution to (4) can now be obtained and it follows a normal distribution given $Y_{t_{i-1}}$, which makes MLE feasible. The key assumption in local linearization method is constant first and second derivatives of the drift on a discretization interval. While this assumption may hold for some SDEs, it may be violated in many other cases. Consider the example of $a_Y(Y_t; \theta) = -Y_t^3$ in Shoji and Ozaki (1998), where the first and second derivatives of $a_Y$ are $-3Y_t^2$ and $-6Y_t$, respectively. These two quantities can be far from constant given the size of $\Delta$ available in practice. It motivates us to relax this key assumption and possibly accommodate more variation in $Y_t$ on the discretization interval.

2.1 Wagner-Platen expansion and strong approximation

When $\frac{\partial a_Y}{\partial Y}$ and $\frac{\partial^2 a_Y}{\partial Y^2}$ in (4) are varying, they can be further expanded when approximating $Y_t$ in (4) on the interval at time $t_{i-1}$. For example, differentiating $\frac{\partial a_Y}{\partial Y}$ 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 we have

$$\frac{\partial a_Y (Y_{t_i}; \theta)}{\partial Y} = \left( \frac{\partial a_Y (Y_{t_{i-1}}; \theta)}{\partial Y} + \frac{1}{2} \frac{\partial^3 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^3} (t_i - t_{i-1}) \right) + \frac{\partial^2 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^2} (Y_{t_i} - Y_{t_{i-1}})$$

$$= \left( \frac{\partial a_Y (Y_{t_{i-1}}; \theta)}{\partial Y} - \frac{1}{2} \frac{\partial^3 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^3} t_{i-1} - \frac{\partial^2 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^2} Y_{t_{i-1}} \right)$$

$$+ \frac{1}{2} \frac{\partial^3 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^3} t_i + \frac{\partial^2 a_Y (Y_{t_{i-1}}; \theta)}{\partial Y^2} Y_{t_i}. \quad (7)$$

If $\frac{\partial^3 a_Y}{\partial Y^3}$ in (7) is not assumed to be constant when $Y$ evolves from $Y_{t_{i-1}}$ to $Y_{t_i}$, we can again differentiate it w.r.t. $Y$ in the approximation. In
theory, if we are willing to assume \( a \) 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), and it is applicable to a diffusion defined in (1) without the requirement of a Doss transform in (4). Consider the solution \( X_t \) to (1) conditioning on the observation at time \( t_{i-1} \)

\[
X_{t_i} = X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} a(X_u; \theta) \, du + \int_{t_{i-1}}^{t_i} b(X_u; \theta) \, dW_u. \tag{8}
\]

By Itô formula, we can expand \( a(X_u; \theta) \) and \( b(X_u; \theta) \) in (8) at the point \( X_{t_{i-1}} \) to have

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

where

\[
R = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^0a(X_z; \theta) \, dz du + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^1a(X_z; \theta) \, dW_z du
\]

\[
+ \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^0b(X_z; \theta) \, dz dW_u + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u} L^1b(X_z; \theta) \, dW_z dW_u,
\]

\[
L^0 = a \frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2}, \tag{9}
\]

\[
L^1 = b \frac{\partial}{\partial x}, \tag{10}
\]

and we let \( a = a(X_{t_{i-1}}; \theta) \) and \( b = b(X_{t_{i-1}}; \theta) \) to conserve space. This expansion can be continued as long as both \( a \) and \( b \) are sufficiently smooth in \( x \). For example, we can further expand \( L^0a(X_z; \theta) \) at \( X_{t_{i-1}} \) in \( R \) to obtain higher order results. The general result is summarized in Theorem 5.5.1 of Kloeden and Platen (1999).

To present a general result for approximation, a set of conditions to guarantee the existence and uniqueness of a strong solution to (1) is needed and
is listed in Appendix A. The conditions are standard in SDE literature and they also guarantee the existence of a transition density. Hereafter, we assume all conditions in Appendix A are satisfied. Next, let us introduce some notation used in Wagner-Platen expansion and detailed discussion can be found in Chapter 5 of Kloeden and Platen (1999). Let \( \alpha \) be a multi-index of length \( l \) such that
\[
\alpha = (j_1, j_2, \cdots, j_l),
\]
where
\[
j_i \in \{0, 1\} \forall i = 1, 2, \cdots, l, \quad l := l(\alpha) \in \{1, 2, \cdots, l\}.
\]
Let \( M \) be the set of all multi-indices such that
\[
M = \{(j_1, j_2, \cdots, j_l) : j_i \in \{0, 1\}, i \in \{1, 2, \cdots, l\}, \text{for } l = 1, 2, \cdots \} \cup \{v\},
\]
where \( v \) is the multi-index of length zero. For an \( \alpha \in M \) with \( l(\alpha) \geq 1 \), we let \( -\alpha \) and \( \alpha^- \) be the multi-index in \( M \) obtained by deleting the first and last element of \( \alpha \), 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 had limits such that \( I_{\alpha^-} [f(\cdot)]_{t_{i-1}t_i} \in \mathcal{H}_{(ji)}, \) for all \( t_0 < t_{i-1} < t_i \), where the multiple Itô integral \( I_{\alpha^-} [f(\cdot)]_{t_{i-1}t_i} \) is defined as
\[
I_{\alpha^-} [f(\cdot)]_{t_{i-1}t_i} = \begin{cases} 
  f(t_i) & \text{if } l = 0 \\
  \int_{t_{i-1}}^{t_i} I_{\alpha^-} [f(\cdot)]_{t_{i-1}u} \, du & \text{if } l \geq 1 \text{ and } j_l = 0 \\
  \int_{t_{i-1}}^{t_i} I_{\alpha^-} [f(\cdot)]_{t_{i-1}u} \, dW_u & \text{if } l \geq 1 \text{ and } j_l \geq 1.
\end{cases}
\]
For example, if \( \alpha = (0, 1, 1, 0) \), we have

\[
I_{(0,1,1,0)}[f(\cdot)]_{t_{i-1},t_i} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u_4} \int_{t_{i-1}}^{u_3} \int_{t_{i-1}}^{u_2} f(\cdot)du_1dW_{u_2}dW_{u_3}du_4.
\]

When \( f(t_i) = 1 \), we write \( I_\alpha[f(\cdot)]_{s,t} = I_\alpha \). As an example, let \( \alpha = (0, 0, 0) \) and we have

\[
I_{(0,0,0)}[1]_{t_{i-1},t_i} = I_{(0,0,0)} = \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{u_3} \int_{t_{i-1}}^{u_2} ds_1ds_2ds_3 = \frac{1}{3!}(t_i - t_{i-1})^3 = \frac{1}{6}\Delta^3.
\]

The researcher chooses the length \( l(\alpha) \) in Theorem 5.5.1 of Kloeden and Platen (1999) to decide how many terms to include in the Wagner-Platen expansion. Examples for \( l(\alpha) = 1 \) and \( l(\alpha) = 2 \) are

\[
X_{t_i} = X_{t_{i-1}} + aI(0) + bI(1) + R_1
\]

\[
X_{t_i} = X_{t_{i-1}} + aI(0) + bI(1) + (aa' + 0.5bb'')I(0,0) + (ab' + 0.5b^2b'')I(0,1) + ba'I(1,0) + bb'I(1,1) + R_2 \tag{13}
\]

where \( R_1 \) and \( R_2 \) correspond to remainders in Wagner-Platen expansion. For each \( \alpha \), define recursively the Itô coefficient function

\[
f_\alpha = \begin{cases} 
    f & \text{if } l = 0 \\
    L^{j_1}f_{\alpha} & \text{if } l \geq 1
\end{cases} \tag{14}
\]

where \( L^{j_1} \) is defined in (9) and (10). If we let \( f(x) \equiv x \), it is easy to verify coefficients in (12) and (13). For example, \( f(0) = a \) and \( f(0,1) = ab' + 0.5b^2b'' \).

Strong Wagner-Platen approximation can be obtained based on expansions such as (12) or (13). Define

\[
\Delta W = W_{t_i} - W_{t_{i-1}},
\]
and as an example, we have

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

Replacing all stochastic integrals in (13) with expressions similar to (15), evaluating all the coefficients at \( X_s \) and omitting the remainder \( R_2 \), 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 \( \Delta W \) and \( \Delta \), but they can be approximated (see Section 5.8 in Kloeden and Platen (1999) for discussion). However, we notice closed forms such as (15) are not needed for estimation or inference.

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

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

where \( \mathcal{A}_\gamma = \{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2\gamma \text{ or } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \} \), \( n(\alpha) \) is the number of zeros in \( \alpha \), \( f_{\alpha} \) is the coefficient function defined in (14) with \( f = x \) and \( \gamma = 0.5, 1, 1.5, \cdots \) is the order of approximation. Approximation in (16) is a special case of equation (10.6.4) in Kloeden and Platen (1999), where we let the approximation start at time \( t_{i-1} \) on the interval \([t_{i-1}, t_i]\).

The discretization interval \( \delta \) is chosen such that \( 0 < \Delta < \delta < 1 \). This is just for technical convenience so that the approximation for \( X_{t_i} \) on \([t_{i-1}, t_i + \delta]\) starts at \( X_{t_{i-1}} \). It is shown in Theorem 10.6.3 of Kloeden and Platen (1999) that the approximation in (16) converges to \( X_{t_i} \) as \( \Delta \to 0 \). We modify their result in the following lemma to show the strong approximation converges to \( X_{t_i} \) with probability 1 (w.p.1) as the approximation order increases to infinity while keeping \( \Delta \) fixed.

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 \( x \).
Lemma 1: Let $Y^\delta(t_i)$ be the order $\gamma$ strong Wagner-Platen approximation defined in (16) with $0 < \Delta < \delta < 1$. Under Assumptions in Appendix A and if the coefficient functions in (14) satisfy

$$|f_\alpha(x) - f_\alpha(y)| \leq K_1 |x - y|$$

for all $\alpha \in \mathcal{A}_\gamma$ and $x, y \in \mathbb{R}$;

$$f_{-\alpha} \in C^2 \text{ and } f_\alpha \in \mathcal{H}_\alpha$$

for all $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$; and

$$|f_\alpha(x)| \leq K_2 (1 + |x|)$$

for all $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$ and $x \in \mathbb{R}$, and the initial condition on $[t_{i-1}, t_{i-1} + \delta]$ satisfies

$$\sqrt{E \left( |X_{t_{i-1}} - Y^\delta(t_{i-1})|^2 \right)} \leq K_3 \delta^\gamma,$$

we have

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

and as $\gamma \to \infty$,

$$P \left( \lim_{\gamma \to \infty} |X_{t_i} - Y^\delta(t_i)| \right) = 0.$$

$K_1, K_2, K_3,$ and $K_4$ are positive constants independent of $\delta$ and $\Delta$.

Result (18) follows Corollary 10.6.4 in Kloeden and Platen (1999). Since (18) implies convergence in probability as $\gamma \to \infty$, (19) can be established if convergence to probability is sufficiently fast. The proof is omitted here and can be found in Lemma 1 of Huang (2008). Strong consistency is needed for a pathwise approximation of a diffusion process. Lemma 1 only requires the interval used in time discrete approximation to be less than one. This is obviously an advantage over the assumption in Pedersen (1995) which
requires the interval to go to zero for consistency of parameter estimates. Although we require \( \gamma \to \infty \) in Lemma 1, simulation shows single-digit \( \gamma \) usually gives very precise parameter estimates.

### 2.2 Quasi-maximum likelihood estimator

Based on the Wagner-Platen approximation in (16) for a fixed \( \gamma \), we can easily obtain our QMLE. Define conditional moments based on (16) as \( \mu_{t_i;\gamma} (X_{t_{i-1}}; \theta) \equiv E (Y^\delta (t_i) | X_{t_{i-1}}) \) and \( \sigma_{t_i;\gamma}^2 (X_{t_{i-1}}; \theta) \equiv Var (Y^\delta (t_i) | X_{t_{i-1}}) \) for a given \( \gamma \), and true unknown conditional moments as \( \mu_{t_i} (X_{t_{i-1}}; \theta) \equiv E (X_{t_i} | X_{t_{i-1}}) \) and \( \sigma_{t_i}^2 (X_{t_{i-1}}; \theta) \equiv Var (X_{t_i} | X_{t_{i-1}}) \). Result (19) in Lemma 1 implies that, as \( \gamma \to \infty \),

\[
\mu_{t_i;\gamma} (X_{t_{i-1}}; \theta) = \mu_{t_i} (X_{t_{i-1}}; \theta) \quad \text{w.p.1,} \tag{20}
\]

\[
\sigma_{t_i;\gamma}^2 (X_{t_{i-1}}; \theta) = \sigma_{t_i}^2 (X_{t_{i-1}}; \theta) \quad \text{w.p.1.} \tag{21}
\]

In other words, conditioning on \( X_{t_{i-1}} \), the first two moments of \( X_{t_i} \) are correctly specified w.p.1 and this further suggests quasi-maximum likelihood estimation in White (1982) will give consistent estimates for \( \theta \). Consider the Wagner-Platen expansion in (13). Conditioning on \( X_{t_{i-1}} \), the terms \( X_{t_{i-1}}, a I(0), (aa + \frac{1}{2} b^2 a'') I(0,0) \) are all constant while the terms \( b I(1), (ab' + \frac{1}{2} b^2 b') I(0,1), ba' I(0,1), bb' I(1,1) \) are all stochastic with zero mean. By taking conditional expectation on both sides of (13) and ignoring \( R_2 \), we obtain the conditional
mean and variance of $X_t$

$$E(X_t | X_{t-1}) \approx X_{t-1} + aI(0) + (aa' + 0.5b^2a'')I(0,0),$$  \hspace{1cm} (22)

$$Var(X_t | X_{t-1}) \approx b^2Var(I(1)) + (ab' + 0.5b^2b'')^2 Var(I(0,1))$$

$$+ (ba')^2 Var(I(1,0))^2 + (bb')^2 Var(I(1,1))^2$$

$$+ 2b(ab' + 0.5b^2b'')Cov(I(1), I(0,1))$$

$$+ 2b^2a'Cov(I(1), I(1,0)) + 2b^3b'Cov(I(1), I(1,1))$$

$$+ 2(ab' + 0.5b^2b'')b'Cov(I(1,0), I(1,1))$$

$$+ 2b^2a'bb'Cov(I(0,0), I(1,1)),$$ \hspace{1cm} (23)

where $Var(\cdot)$ and $Cov(\cdot)$ are variance and covariance functions and can be calculated using Lemma 5.7.2 of Kloeden and Platen (1999) if parametric forms of $a(X_t; \theta)$ and $b(X_t; \theta)$ are given. Note that integrands in all stochastic integrals in (23) become one after successive applications of (11) and (14), and the inequality in the result of Lemma 5.7.2 of Kloeden and Platen (1999) simplifies to equality. This enables us to calculate the second moments of those stochastic integrals without worrying about any closed forms such as (15). The QMLE is the parameter vector that solves

$$\max_{\theta} \bar{l}_n(\theta) = \sum_{i=1}^{n} \ln \bar{p}(X_t | X_{t-1}; \theta)$$ \hspace{1cm} (24)

with transition density

$$\bar{p}(X_t | X_{t-1}; \theta) = \left(2\pi \sigma_t^2(X_{t-1}; \theta)\right)^{-1/2} \exp \left(-\frac{(X_t - \mu_t(X_{t-1}; \theta))^2}{2\sigma_t^2(X_{t-1}; \theta)}\right).$$

When the approximation order is $\gamma$, we let

$$l_n(\theta) = \sum_{i=1}^{n} \ln p(X_t | X_{t-1}; \theta),$$ \hspace{1cm} (25)
where

\[ p(X_{ti}|X_{ti-1}; \theta) = (2\pi \sigma_{ti,\gamma}^2 (X_{ti-1}; \theta))^{-1/2} \exp \left( -\frac{(X_{ti} - \mu_{ti,\gamma} (X_{ti-1}; \theta))^2}{2\sigma_{ti,\gamma}^2 (X_{ti-1}; \theta)} \right). \]

When \( l(\alpha) = 2 \), the expressions for \( \mu_{ti,\gamma} (X_{ti-1}; \theta) \) and \( \sigma_{ti,\gamma}^2 (X_{ti-1}; \theta) \) are given in equations (22) and (23). Wagner-Platen approximations for \( X_{ti} \) when \( l(\alpha) = 3 \) and \( l(\alpha) = 4 \) are given in Appendix C. Let the estimator obtained from (24) be \( \hat{\theta}_n^{\text{QMLE}} \), and the estimator obtained from (25) be \( \hat{\theta}_n \). We will show \( \hat{\theta}_n \) converges to \( \hat{\theta}_n^{\text{QMLE}} \) and \( \hat{\theta}_n^{\text{QMLE}} \) converges to \( \theta \).

Consistency and asymptotic theory of the QMLE is discussed in detail in White (1994). We adapt the proof in Bollerslev and Wooldridge (1992) to prove the weak consistency and asymptotic normality of QMLE. Let \( e_{ti} = X_{ti} - \mu_{ti} (X_{ti-1}; \theta) \), and the score and the Hessian for \( \ln \tilde{p} (X_{ti}|X_{ti-1}; \theta) \) are

\[
\begin{align*}
    s_{ti} (\theta) &= \left( \frac{d\mu_{ti}}{d\theta} \right)^T e_{ti} / \sigma_{ti}^2 + \frac{1}{2} \left( \frac{d\sigma_{ti}^2}{d\theta} \right)^T \left( (e_{ti} - \sigma_{ti}^2) / \sigma_{ti}^2 \right), \\
    h_{ti} (\theta) &= \frac{1}{2} \frac{d\sigma_{ti}^2}{d\theta} \cdot \frac{d\sigma_{ti}^{-2}}{d\theta} \cdot \frac{\sigma_{ti}^2}{\sigma_{ti}^2} + \frac{1}{2} \frac{d\sigma_{ti}^2}{d\theta} \cdot \frac{\sigma_{ti}^{-2}}{\sigma_{ti}^2} \cdot \frac{d\sigma_{ti}^{-2}}{d\theta} \\
    &+ \frac{1}{2} \frac{d\sigma_{ti}^2}{d\theta} \cdot \frac{d\sigma_{ti}^{-2}}{d\theta} - \frac{\sigma_{ti}^2}{\sigma_{ti}^2} \cdot \frac{d\sigma_{ti}}{d\theta} \cdot \frac{d\sigma_{ti}^{-2}}{d\theta} + \left( \frac{d\mu_{ti}}{d\theta} \right)^T \frac{d\sigma_{ti}}{d\theta} \cdot \frac{d\sigma_{ti}}{d\theta} \cdot e_{ti} \\
    &- \left( \frac{d\mu_{ti}}{d\theta} \right)^T \left( \frac{d\mu_{ti}}{d\theta} \right) / \sigma_{ti}^2,
\end{align*}
\]

where the superscript \( T \) stands for matrix transpose. It is easy to verify that the score function has zero mean conditional on \( X_{ti-1} \). Next, let \( a_{ti} (\theta) = E (-h_{ti} (\theta)) \). As \( \gamma \to \infty \), under (20) and (21), we have

\[ a_{ti} (\theta) = \left( \frac{d\mu_{ti}}{d\theta} \right)^T \cdot \left( \frac{d\mu_{ti}}{d\theta} \right) / \sigma_{ti}^2 + \frac{1}{2} \left( \frac{d\sigma_{ti}^2}{d\theta} \right)^T \cdot \frac{d\sigma_{ti}^2}{d\theta} / \sigma_{ti}^4 \text{ w.p.} 1. \]

The next theorem establishes the asymptotic normality of QMLE.
Theorem 1: As $\gamma \to \infty$, we have $\hat{\theta}_n \to \hat{\theta}_n^{\text{QMLE}}$ w.p.1. Under assumptions in Appendix B and as both $\gamma \to \infty$ and $n \to \infty$, we further have
\[
(A_n^{-1}B_nA_n^{-1})^{-1/2} \frac{\sqrt{n} (\hat{\theta}_n - \theta)}{d} \to N(0, I_p),
\]
where $A_n = n^{-1} \sum_{i=1}^{n} E(a_{t_i}(\theta))$ and $B_n = n^{-1} \sum_{i=1}^{n} E(s_{t_i}(\theta)'s_{t_i}(\theta))$. Also, $\tilde{A}_n - A_n \to 0$ and $\tilde{B}_n - B_n \to 0$ in probability, where $\tilde{A}_n = n^{-1} \sum_{i=1}^{n} a_{t_i}(\hat{\theta}_n)$ and $\tilde{B}_n = n^{-1} \sum_{i=1}^{n} s_{t_i}(\hat{\theta}_n)'s_{t_i}(\hat{\theta}_n)$.

The proof is similar to the proof of Theorem 2.1 in Bollerslev and Wooldridge (1992) and is sketched in Appendix B. The asymptotic results in both Lemma 1 and Theorem 1 assume order of Wagner-Platen approximation goes to infinity but do not require the discretization interval to go to zero. In the proof of Theorem 1, we let $\gamma \to \infty$ to prove the convergence of (25) to (24) and $\hat{\theta}_n$ to $\hat{\theta}_n^{\text{QMLE}}$; when both $\gamma \to \infty$ and $\triangle \to 0$, $\hat{\theta}_n$ may converge to $\hat{\theta}_n^{\text{QMLE}}$ at a faster rate. Unlike the approach in Aït-Sahalia (2002), the method in this paper chooses to approximate a strong solution to (1), but not the transition density. This means direct proof of convergence of $\hat{\theta}_n$ to MLE is not possible. However, the convergence of $\hat{\theta}_n$ to MLE is implied by the consistency of $\hat{\theta}_n$ since both $\hat{\theta}_n$ and MLE converge to $\theta$ in probability. Extension of the Theorem 1 to nonstationary and nonergodic diffusions is possible (see Theorem 10.1 in Wooldridge (1994) for an example).

A GMM estimator is also possible based on correctly specified conditional moments. However, forms of estimation between GMM and QMLE are different and we do not pursue this direction in current paper.

An alternative way to prove the consistency of the estimator is to use the central limit theorem (CLT) for dependent processes to show the summation of all the stochastic integrals in Wagner-Platen approximation converges to a normal random variable. For example, the approximation in (13) involves several dependent terms such as $I_{(1)}$, $I_{(0,1)}$ and $I_{(1,0)}$. As the approximation or-
der increases, we expect to see more such terms to appear on the r.h.s of (13) and the summation of all these stochastic integrals can be treated as a sum of dependent random variables. Lemma 5.7.2 in Kloeden and Platen (1999) can be used to show the correlation between stochastic integrals decreases to zero as $\gamma \to \infty$. However, verifying the dependence structure among stochastic integrals is hard and we leave it for future research. Yet another way to obtain consistent estimator based on Wagner-Platen approximation is to use characteristic function. Consider the stochastic terms $I_{(1)}$, $I_{(0,1)}$, $I_{(1,0)}$ and $I_{(1,1)}$ in (13). The mean, variance and correlation of these terms can be easily calculated and by assuming normality of these terms, we can obtain the characteristic function for the summation of these stochastic integrals. The approximate density function can be obtained by numerically inverting the characteristic function and maximizing the joint density function yields the estimate. However, we note that the stochastic integrals on the r.h.s. of (13) are not only correlated but also dependent. Hence the characteristic function for a sum of dependent random variables is needed. Furthermore, it is unclear how accurate the numerical inversion of a characteristic function can be in this case. Given all these difficulties, we find it is easy to implement QMLE and its asymptotic result is also simple to establish.

In practice, discretization intervals may not be equal and may be even random. One example is the endogenous sampling process for short-term interest rate in Yu and Phillips (2001), where the process is sampled more frequently when the level of interest rate is high. This poses no difficulty in implementing our QMLE method. First, unequally spaced observations do not invalidate the conditions in (20) and (21). In fact, the results in (20) and (21) hold as long as sampling intervals are less than one. The key assumption here is the order of approximation goes to infinity. Second, consistency and
asymptotic normality of the QMLE is unaffected as long as (20) and (21) hold. Hence the assumption of constant discretization interval can be easily relaxed.

Our estimation procedure is also applicable to diffusion processes when an analytic solution to (5) is not available. Consider the preferred short-term interest rate model in Aït-Sahalia (1996):

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

which nests a wide range of models for short-term interest. The Doss transform in (5) has no analytic form. Although the value of the function, $G(X) = \int_X^X du/(\theta_5 + \theta_6 u + \theta_7 u^{\theta_8})$, can be obtained by numerical integration, this extra step poses difficulty in estimation because numerical integration has to be recomputed for each value of $\theta$ in parameter space at each $X_t$. The computation is even more difficult when the diffusion coefficient is nonlinear in parameters. Our QMLE requires no Doss transform and only needs the drift and diffusion coefficient to be smooth. This property seems to be more attractive in a multivariate setting. When the diffusion function, $b(X_t; \theta)$, is a matrix, Doss transform is not applicable to all multivariate diffusion processes. However, QMLE is applicable to diffusions without Doss transform. In simulation, we also investigate precision of QMLE when the model is untransformed. It turns out that our QMLE yields equally precise results without Doss transform.

3 Simulation results

In this section we conduct simulation experiment to study the numerical precision of QMLE based on higher order Wagner-Platen approximations. QMLE with $l(\alpha) = 3$ and $l(\alpha) = 4$ in (16) is used in simulation. Approxi-
mation with $l(\alpha) = 3$ is given in Section 5.5 of Kloeden and Platen (1999) and in Appendix C we also provide the approximation with $l(\alpha) = 4$. In order to gauge the precision of our QMLE, the following well-known models with closed-form transition density are used in simulation. The first is Vasicek model proposed in Vasicek (1977):

$$dX_t = \theta_2 (\theta_1 - X_t) \, dt + \theta_3 dW_t$$

with transition density

$$p(x_t | x_{t-1}) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x_t - \mu_t)^2}{2\sigma_t^2} \right),$$

where $\mu_t = \theta_1 + (x_{t-1} - \theta_1) \exp(-\theta_2 \Delta)$ and $\sigma_t^2 = \theta_3^2 (1 - \exp(-2\theta_2 \Delta)) / (2\theta_2)$.

The second is the CIR model in Cox, Ingersoll, and Ross (1985):

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

with transition density

$$p(x_t | x_{t-1}) = c \exp(-u - v) (v/u)^{q/2} I_q(2\sqrt{uv}),$$

where $c = 2\theta_2 / (\theta_3^2 (1 - \exp(-\theta_2 \Delta)))$, $u = c x_{t-1} \exp(-\theta_2 \Delta)$, $v = c x_t$, $q = 2\theta_2 \theta_1 / \theta_3^2 - 1$, and $I_q(2\sqrt{uv})$ is the modified Bessel function of the first kind with order $q$. The third is Black-Scholes model in Black and Scholes (1973):

$$dX_t = \theta_2 X_t dt + \theta_3 X_t dW_t$$

with transition density

$$p(x_t | x_{t-1}) = \frac{1}{x_{t-1} \sigma_t \sqrt{2\pi}} \exp \left( -\frac{(\ln(x_{t-1}) - \theta_2)^2}{2\sigma_t^2} \right).$$
Based on these closed-form densities, exact maximum likelihood estimator (MLE) can be obtained and used as a benchmark to measure the accuracy of QMLE.

In Tables 1 and 2 we simulate 1000 observations for every sample path and repeat estimation 5000 times, where \( \hat{\theta}^{(\text{MLE})} \), \( \hat{\theta}^{(\text{EUL})} \), \( \hat{\theta}^{(l=3)} \) and \( \hat{\theta}^{(l=4)} \) correspond to exact MLE, Euler estimator, QMLE with \( l(\alpha) = 3 \) and QMLE with \( l(\alpha) = 4 \), respectively. The unconditional density of the first observation is ignored in (25). The bias and standard deviation (s.d.) reported in Tables 1 and 2 are averages over 5000 replications. An alternative way to present the results is to report the average distance of between QMLE and MLE, as it is done in Table III in Aït-Sahalia (2002). This average distance between QMLE and MLE can be easily inferred from the results in Tables 1 and 2. For example, consider the estimated bias for \( \theta_1 \) in the Vasicek model in Table 1. We find \( \hat{\theta}_1^{(\text{MLE})} - \theta_1 \approx -0.0000267218 \) and the estimated bias for \( \hat{\theta}_1^{(l=3)} \) is \( \hat{\theta}_1^{(l=3)} - \theta_1 \approx -0.0000267213 \). Hence the distance between \( \hat{\theta}_1^{(l=3)} \) and MLE is equal to \( (\hat{\theta}_1^{(l=3)} - \theta_1) - (\hat{\theta}_1^{(\text{MLE})} - \theta_1) \approx 0.0000000005 \). Based on Wagner-Platen approximations in Appendix C, we can obtain the explicit expressions for \( \mu_{t_{i_t}, \gamma} \), \( \sigma^2_{t_{i_t}, \gamma} \) and a closed-form transition density. Derivations for explicit expressions are done in Mathematica. These expressions are later translated to Matlab for estimation using Tomlab/LGO, where I keep 16 decimal points for all numbers in this translation. An important tip in programming is to use the function Simplify[·] in Mathematica to simplify symbolic expressions before translating them into Matlab code. This simplification will speed up the estimation in Matlab.

The true values of the parameters used in Table 1 are the same as those used in Table III of Aït-Sahalia (2002), where \( \theta = (0.06, 0.5, 0.03) \) in the Vasicek model, \( \theta = (0.06, 0.5, 0.15) \) in the CIR model, \( \theta = (0.2, 0.3) \) in the
Black-Scholes model and $\Delta = 1/12$. Results for $\hat{\theta}^{(l=3)}$ and $\hat{\theta}^{(l=4)}$ in CIR and Black-Scholes models reported in Table 1 are obtained from the transformed models. Doss transformation for Vasicek model is not needed since the diffusion coefficient is already a constant. There are two ways to assess the numerical accuracy of QMLE in Tables 1 and 2. First, we can directly compare the estimated bias between QMLE and Euler estimator; second, since exact MLE is the most efficient estimator under certain regularity conditions, we may compare the distance between various estimators and MLE. Either way, we generally find QMLE outperforms the Euler estimator and it is usually very close to the exact MLE. The estimates for $\hat{\theta}^{(l=3)}$ and $\hat{\theta}^{(l=4)}$ in Black-Scholes model give identical results, though still outperforming the Euler estimator. The reason is after Doss transformation, $Y = \ln (X) / \theta_3$, the model has constant drift and constant diffusion coefficient and higher order approximation will not bring in more information. In this special case, higher order Wagner-Platen expansion reduces to Euler estimator applied to the transformed model.

Both Vasicek and Black-Scholes models are linear and the strength of QMLE is really in estimating nonlinear diffusions. In Table 2 we investigate specifically the precision of QMLE in the nonlinear CIR model. In addition, we also report the results obtained without Doss transformation and they are $\hat{\theta}^{(2.5,U)}$ and $\hat{\theta}^{(3.5,U)}$ in Table 2. The simulation results in Hurn et al. (2007) recommend hermite polynomial expansion method for estimation in practice and we also compare QMLE with the method in Aït-Sahalia (2002). The estimator $\hat{\theta}^{(Hermite)}$ is based on an expansion with seven hermite polynomial terms with Taylor series of order three in $\Delta$ derived on page 238 of Aït-Sahalia (2002). The data generating process (DGP) (a) is the same as that for CIR model in Table 1 and DGP (b) is selected from that used in Table 2 of
Durham and Gallant (2002) and DGP (c) is selected from Jensen and Poulsen (2002). In DGP (a), we find the bias of \( \hat{\theta}^{(l=3,U)} \) and \( \hat{\theta}^{(l=4,U)} \) is comparable to the bias of \( \hat{\theta}^{(l=3)} \) and \( \hat{\theta}^{(l=4)} \) and this is true in all three DGPs considered in Table 2. In DGP (b), we let \( \theta = (0.06, 0.5, 0.03) \) and \( \Delta = 1/12 \). We note QMLE from both transformed and untransformed models give significant improvement over the Euler estimator for \( \theta_2 \) and \( \theta_3 \). In DGP (c), we let \( \theta = (0.08, 0.24, 0.08838) \) and \( \Delta = 1/12 \). We find QMLE substantially improves the estimation precision for \( \theta_3 \). In general, we find that our QMLE gives good estimation results and outperforms the Euler estimator.

Compared with hermite polynomial method in Aït-Sahalia (2002), results in Table 2 shows QMLE yields similar numerical precision to \( \hat{\theta}^{(\text{Hermite})} \) for estimates of \( \theta_1 \) and \( \theta_2 \) in DGPs (a) and (c), and it slightly outperforms \( \hat{\theta}^{(\text{Hermite})} \) for estimate of \( \theta_3 \). In DGP (b), QMLE gives better estimation results than \( \hat{\theta}^{(\text{Hermite})} \). In Appendix C, the approximation when \( l(\alpha) = 4 \) to obtain QMLE is complicated. However, results in Tables 1 and 2 also show QMLE with \( l(\alpha) = 3 \) is enough for precise estimation for a very general class of diffusion processes.

## 4 Conclusion

This paper introduces QMLE for discretely observed diffusions. The estimator is based upon an approximated solution to (1). Instead of approximating the unknown transition density, our method chooses to approximate the strong solution of a SDE and the sum of all stochastic terms in approximation is treated as a normal random variable when constructing the likelihood function. This method applies to SDEs with arbitrary specification in drift and diffusion coefficient as long as they are sufficiently smooth. Simulation shows single-digit order of approximations gives precise estimation for the
models under consideration. The QMLE generally outperforms the Euler estimator and yields estimates that are very close to the true MLE.

It is possible to extend the current method in several directions. First, QMLE for multivariate and time-inhomogeneous diffusion processes are natural extensions. Similar to the scalar case in this paper, QMLE is expected to be applicable to multivariate diffusions with arbitrary specifications in the drift and diffusion coefficient, even if Doss transform is inapplicable. This is studied in Huang (2008). Second, it is possible to include jump components in our model by building upon the work in Bruti-Liberati and Platen (2007). Third, it is also interesting to investigate the efficiency gain by using higher order approximations in SMLE.
Appendix A

The following assumptions are adapted from Section 4.5 in Kloeden and Platen (1999) to ensure the existence and uniqueness of a strong solution to (1). These assumptions also ensure the existence of the transition density. Assumption (A2) is stronger than the Lipschitz condition in Kloeden and Platen (1999), but it ensures the existence of all derivatives used in infinite order strong Wagner-Platen approximation. Let \( \mathbb{R} \) be the real line.

(A1) \( a(x; \theta) \) and \( b(x; \theta) \) are both measurable in \( x \in \mathbb{R} \);

(A2) \( a(x; \theta) \) and \( b(x; \theta) \) are infinitely differentiable in \( x \);

(A3) For some positive constant \( K \), we have \( |a(x; \theta)|^2 \leq K(1 + |x|) \) and \( |b(x; \theta)|^2 \leq K(1 + |x|^2) \) for all \( x \in \mathbb{R} \);

Let \( \{ \mathcal{F}_t, t \geq 0 \} \) be a family of \( \sigma \)-algebras generated by \( W_t \) for all \( t \in [t_0, T] \). Then

(A4) \( X_{t_0} \) is \( \mathcal{F}_{t_0} \)-measurable with \( E(|X_{t_0}|^2) < \infty \).

Appendix B

The following regularity conditions are adapted from Bollerslev and Wooldridge (1992). Assumption (A2) implies \( \mu_{t_i}(X_{t_{i-1}}; \theta) \) and \( \sigma^2_{t_i}(X_{t_{i-1}}; \theta) \) are both continuous in \( x \) and it further implies they are measurable. Hence Assumption (B2) is identical to Condition (A.1)(ii) in Bollerslev and Wooldridge (1992). For all \( t_i \) such that \( t_0 < t_i \leq T \), we assume

(B1) \( \Theta \) is compact subset of \( \mathbb{R}^m \) and has nonempty interior. \( \theta \in \Theta \).
(B2) All integer powers of $a(x; \theta)$, $b(x; \theta)$ and derivatives of $a(x; \theta)$ and $b(x; \theta)$ in $x$ are twice differentiable w.r.t. $\theta$ for all $x \in \mathbb{R}$.

(B3) $\tilde{l}_n(\theta)$ has a unique maximizer for all $n$. In addition, let $\tilde{l}_{t_i}(\theta) = \ln \tilde{p}(X_{t_i}|X_{t_{i-1}}; \theta)$, and $\tilde{l}_{t_i}(\tilde{\theta}_n^{\text{QMLE}}) - \tilde{l}_{t_i}(\theta)$ satisfies the uniform weak law of large numbers (UWLLN) with $\theta$ being the identifiably unique maximizer of

$$n^{-1} \sum_{t=1}^{n} E(\tilde{l}_{t_i}(\tilde{\theta}_n^{\text{QMLE}}) - \tilde{l}_{t_i}(\theta)).$$

(B4) (a) $\{h_{t_i}(\theta)\}$ and $\{a_{t_i}(\theta)\}$ satisfy the WLLN.

(b) $\left\{ a_{t_i}(\tilde{\theta}_n^{\text{QMLE}}) - h_{t_i}(\theta) \right\}$ satisfies the UWLLN.

(c) $A_n = n^{-1} \sum_{i=1}^{n} E( a_{t_i}(\theta))$ is uniformly positive definite.

(B5) (a) $\{s_{t_i}(\theta)'s_{t_i}(\theta)\}$ satisfies the WLLN.

(b) $B_n = n^{-1} \sum_{i=1}^{n} E( s_{t_i}(\theta)'s_{t_i}(\theta))$ is uniformly positive definite.

(c) $B_n^{-1/2} n^{-1/2} \sum_{i=1}^{n} s_{t_i}(\theta) \overset{d}{\to} N(0, I_m)$.

(B6) (a) $\left\{ a_{t_i}(\tilde{\theta}_n^{\text{QMLE}}) - a_{t_i}(\theta) \right\}$ satisfies the UWLLN.

(b) $\left\{ s_{t_i}(\tilde{\theta}_n^{\text{QMLE}})'s_{t_i}(\tilde{\theta}_n^{\text{QMLE}}) - s_{t_i}(\theta)'s_{t_i}(\theta) \right\}$ satisfies the UWLLN.

Proof of Theorem 1: Result (19) in Lemma 1 and the continuity of (25) in $\mu_{t_i}(X_{t_{i-1}}; \theta)$ and $\sigma_{t_i}^2(X_{t_{i-1}}; \theta)$ suggest $l_n(\theta) \to \tilde{l}_n(\theta)$ w.p.1., which further implies $\tilde{\theta}_n \to \tilde{\theta}_n^{\text{QMLE}}$ w.p.1. when $\tilde{l}_n(\theta)$ has a unique maximizer under Assumption (B3). The weak consistency and asymptotic normality of $\tilde{\theta}_n^{\text{QMLE}}$ follows the proof of Theorem 2.1 in Bollerslev and Wooldridge (1992) under Assumptions (B1) to (B6). Finally, weak consistency and asymptotic normality $\tilde{\theta}_n$ in Theorem 1 follows the fact $\tilde{\theta}_n \to \tilde{\theta}_n^{\text{QMLE}}$ w.p.1. The convergence result for $\tilde{A}_n$ and $\tilde{B}_n$ in Theorem 1 follows the continuity of $A_n$ and $B_n$ in $\theta$, the weak convergence of $\tilde{\theta}_n$ to $\theta$, and Assumptions B4 (a), B5(a) and B6. $\square$
Appendix C

The following are Wagner-Platen expansions for $X_{t_i}$ at $X_{t_{i-1}}$ when $l(\alpha) = 3$ and $l(\alpha) = 4$, where the expansion for $l(\alpha) = 3$ is given in Section 5.5 in Kloeden and Platen (1999). $R_3$ and $R_4$ are remainders in expansions. By ignoring the remainders, letting $t_i - t_{i-1} = \Delta$ in all stochastic integrals appearing in the Wagner-Platen expansions and evaluating their coefficients at $X_{t_{i-1}}$, we obtain the corresponding strong Wagner-Platen approximations. Strong Wagner-Platen approximation when $l(\alpha) = 3$ is given by

$$X_{t_i} = X_{t_{i-1}} + aI_{(0)} + bI_{(1)} + \left(aa' + 0.5b^2a''\right)I_{(0,0)} + \left(ab' + 0.5b^2b''\right)I_{(0,1)} + ba'I_{(1,0)} + bb'I_{(1,1)} + [a(aa'' + (a')^2 + bb'a'' + 0.5b^2a'') + 0.5b^2(aa''' + 3a'a'') + ((b')^2 + bb'')a'' + 2bba'']I_{(0,0,0)} + [a(ab' + ab' + bb'b'' + 0.5b^2b'') + 0.5b^2(a''b' + 2a'b'') + ab'' + ((b')^2 + bb'')b'' + 2bb'b'' + 0.5b^2b^{(4)}]I_{(0,0,1)} + [a(b'a' + ba'') + 0.5b^2(b''a' + 2b'a'' + ba'')]I_{(0,1,0)} + [a((b')^2 + bb'') + 0.5b^2(b''b' + 2bb'' + bb'')]I_{(0,1,1)} + b(aa'' + (a')^2 + bb'a'' + 0.5b^2a'')I_{(1,0,0)} + b(ab' + ab' + bb'b'' + 0.5b^2b'')I_{(1,0,1)} + b(ab' + a'b)I_{(1,1,0)} + b((b')^2 + bb'')I_{(1,1,1)} + R_3.$$
Strong Wagner-Platen approximation when \( l(\alpha) = 4 \) is given by

\[
X_{t_i} = X_{t_{i-1}} + aI_{(0)} + bI_{(1)} + (aa' + 0.5b^2a'')I_{(0,0)} + (ab' + 0.5b^2b'')I_{(0,1)} + ba'I_{(1,0)} + bb'I_{(1,1)} + [a(aa'' + (a')^2 + bb'a'' + 0.5b^2a''') + 0.5b^2(3a'a'' + 3a'a'') + ((a')^2 + bb''a'' + 2bb'a''') + 0.25b^4a^{(4)}]I_{(0,0,0)} + [a(ab' + ab'' + bb'b'' + 0.5b^2b''') + 0.5b^2(ab' + 2a'b'') + ab''' + ((a')^2 + bb''b'' + 2bb'b'' + 0.5b^2b^{(4)})]I_{(0,0,1)} + [a(ab'a' + ba'') + 0.5b^2(b''a' + 2b'a'' + ba'')]I_{(0,1,0)} + [a((a')^2 + bb'') + 0.5b^2(bb' + 2bb' + bb'')I_{(0,1,1)} + b(aa'' + (a')^2 + bb'a'' + 0.5b^2a''')I_{(1,0,0)} + b(ab'' + a'b' + bb'b'' + 0.5b^2b''')I_{(1,0,1)} + b(ab' + a'b)I_{(1,1,0)} + b((a')^2 + bb')I_{(1,1,1)} + [a^3 + (a''(a')^2 + 3ab'a'b' + 4a'a'' + b(a''b' + 1.5ba^{(4)})a^2 + ((a')^3 + b(4b'a'' + 5.5ba''')a' + b(a''(b')^3 + 5.5ba''^2)(b')^2 + b(4a''b' + 4.5ba^{(4)})b' + b0.75a^{(5)}b^2 + 3.5b'a^{(3)}b + a'b^{(3)}b + 3.5(a')^2)a + b^2(0.5a''(a')^4 + 4ba^{(3)}(b')^3 + b(4a''b'' + 4.75ba^{(4)})(b')^2 + b(3.5(a''a) + 2bb^{(3)}a'' + b(7b''a'' + 1.5ba^{(5)})b' + 3.5(a')^2a'' + a'(2.5a''(b')^2 + 6ab'a') + b(2.5a''b'' + 1.75ba^{(4)})) + b^2(a''(1.25(b')^2 + 1.375a'' + 0.25bb^{(4)}) + b(a''b'' + 1.75b'a^{(4)} + 0.125ba^{(6)})]I_{(0,0,0,0)} + [a^3b'' + a^2(3a'b'' + (b')^2b'' + (a'' + 3bb''') + b((b')^2 + 1.5bb^{(4)})) + a((a')^2b' + ba'(3b'b'' + 4.5bb''') + b((b')^3b'' + (b')^2(a'' + 5.5bb''')) + bb'(4b''b' + a'' + 4.5bb^{(4)}) + b(3.5a''b'' + 4.5bb'b'' + 0.75b^2b^{(5)}))] + b^2(2(a')^2b'' + 0.5(b')^4b'' + (b')^3(0.5a'' + 4bb''') + b(b')^2(4(b')^2)
\]
\[+a^{(3)} + 4.75bb^{(4)} + a'(2b'^2b'' + b'(\pm a'' + 5bb'')) + b(2b'')^2 \\
+1.5bb^{(4)})) + bb' (4a''b'' + b(9b''b'' + 0.25a^{(4)} + 1.5bb^{(5)})) \\
+b^2(1.25a''b'' + b(b''')^2 + b'' (a'' + 2bb^{(4)} + 0.125b^2b^{(6)}))I_{(0,0,1)} \\
+[a^2 (2b'a'' + a'b' + ba^{(3)} + a((a')^2b' + ba' (a'' + b'b'' + bb'')) \\
+b (2b'^2a'' + 4bb'a'' + b (3a''b'' + ba^{(4)}))) + b^2((b')^3a'' + (a')^2b'' \\
+3.5b(b')^2a'' + bb' (4a''b'' + 2ba^{(4)} + a'(0.5(b')^2b'' + b' (2.5a'' + bb'')) \\
+b (0.5(b')^2a'' + 0.25bb^{(4)})) + b(0.5(a'')^2 + 2bb''a'' + ba''b'' \\
+0.25b^2a^{(5)}))]I_{(0,1,0)} + [a^2 (3b'b'' + bb'') + a(a'(b')^2 + bb'') \\
+b((b')^2b'' + (2(b'')^2 + b'(3b'' + 4b'''')) + b^2(b'' + b^{(4)})) \\
+b^2(0.5(b')^2b'' + (b')^2(0.5a'' + b(3b'' + 2.5b''))) + b'(3a'b'' + b(1.5(b')^2 \\
+b(3b'' + 1.75b^{(4)})))] + b(0.5a''b'' + a'b'' + bb''(1.5b'' + 1.5b''') \\
+b^2(0.5b^{(4)} + 0.25b^{(5)})))]I_{(0,1,1)} + [a^2 (b'a'' + ba'') + a((a')^2b' + 3ba'a'' \\
+b(2b'^2a'' + 3.5bb'a'' + +b (1.5a''b'' + ba^{(4)}))) + b^2((b')^3a'' \\
+0.5(a')^2b'' + 3.5b(b')^2a'' + a' (3b'a'' + 2ba^{(3)} + bb' (3a''b'' + 2ba^{(4)} \\
+b(1.5(a'')^2 + 1.75bb''a^{(3)} + 0.5ba''b'' + 0.25b^2a^{(5)}))I_{(0,1,0)} \\
+[a^2 (b'b'' + bb^{(3)} + (aa' ((b')^2 + 2bb'') + b(2b')^2b'' + b'(a'' + 3.5bb'')) \\
+b(1.5(b')^2 + 2bb^{(4)})) + b^2((b')^3b'' + (b')^2(a'' + 3.5bb'')) + b' (2.5a'b'' \\
+b(3(b'')^2 + 0.5a'' + 2bb^{(4)})) + b(1.5a'b'' + 1.5a'b''' + 2.25bb''b''') \\
+0.25b^2b^{(5)})]I_{(0,1,0)} + [a(a'(b')^2 + bb') + b(3b'a'' + ba'')] \\
+b^2(2b'^2a'' + b'(1.5a''b'' + 2.5ba''') + b(2a''b'' + 0.5a'b'' \\
+0.5ba^{(4)}))]I_{(0,1,1)} + [a((b')^3 + 4bb'b'' + b^2b^{(3)} + b^2(3.5(b')^2b'' \\
+3b'b''' + b(2(b'')^2 + 0.5bb^{(4)}))]I_{(0,1,1)} + [b((a')^3 + a^2a'' + a'(4aa'' \\
+b(4b'a'' + 2.5ba'')) + a((b')^2a'' + 3bb'a''' + b(a''b'' + ba^{(4)})) \\
+27}
\[ + b((b')^3 a'' + 3.5b(b')^2 a''' + bb'(2.5a'' b'' + 2ba^{(4)}) + b(1.5a'')^2 + 1.5bb'' a'' \\
+ 0.5ba'' b'' + 0.25b^2 a^{(5)})])I_{(1,0,0,0)} + [b((a')^2 b' + a^2 b^{(3)} + a'(3ab'' \\
+ b(3b' b'' + 2bb'')) + a((b')^2 b'' + b'(a'' + 3bb'') + b((b'')^2 + bb^{(4)}) \\
+ b(1.5a'' b'' + 2bb'' b'' + 0.25b^2 b^{(5)})])I_{(1,0,0,1)} + [b((a')^2 b' + a(2b'a'' + ba'') \\
+ a'(ab'' + b(a'' + b'b'') + 0.5b^2 b''') + b(2(b')^2 a'' + 1.5ba'' b'' + 2.5bb'a'' \\
+ 0.5b^2 a^{(4)})]I_{(1,0,1,0)} + [(ba'((b')^2 + bb'') + a(3b'b'' + bb''') + b((b')^2 b'' \\
+ b(0.5(b'')^2 + b' (3b'' + 2bb'')) + b^2 (b'' + 0.5b^{(4)}))])I_{(1,0,1,1)} \\
+ [b((a')^2 b' + 3ba' a'' + a(b'a'' + ba''') + b(2(b')^2 a'' + ba'' b'' + 2.5bb'a'') \\
+ 0.5b^2 a^{(4)})]I_{(1,1,0,0)} + [b(ab'b'' + a'((b')^2 + 2bb'') + b(b'a'' + 2(b')^2 b'' \\
+ ab''') + b^2 ((b'')^2 + 2.5bb'')) + 0.5b^3 b^{(4)}])I_{(1,1,0,1)} + [b(a'((b')^2 + bb'') \\
+ b(3b'a'' + ba'')]I_{(1,1,1,0)} + [b((b')^3 + bb'b'' + b^2 b'') ]I_{(1,1,1,1)} + R_4. \]
References


Table 1. Estimated bias and standard deviation for Vasicek, CIR and Black-Scholes models

<table>
<thead>
<tr>
<th></th>
<th>Vasicek</th>
<th></th>
<th>CIR</th>
<th></th>
<th>Black-Scholes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$dX_t = \theta_1 (\theta_2 - X_t) dt + \theta_3 dW_t$</td>
<td>$dX_t = \theta_2 (\theta_1 - X_t) dt + \theta_3 X_t^{0.5} dW_t$</td>
<td>$dX_t = \theta_2 X_t dt + \theta_3 X_t dW_t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1^{(MLE)} - \theta_1$</td>
<td>Bias</td>
<td>-0.0000267218</td>
<td>-0.00010845</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0064456224</td>
<td>0.00799336</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1^{(EUL)} - \theta_1$</td>
<td>Bias</td>
<td>-0.0000267214</td>
<td>-0.00010819</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0064455361</td>
<td>0.00802514</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1^{(l=3)} - \theta_1$</td>
<td>Bias</td>
<td>-0.0000267213</td>
<td>-0.00010906</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0064457349</td>
<td>0.00802198</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_1^{(l=4)} - \theta_1$</td>
<td>Bias</td>
<td>-0.0000267217</td>
<td>-0.00011320</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0064456897</td>
<td>0.00806676</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_2^{(MLE)} - \theta_2$</td>
<td>Bias</td>
<td>0.0484283531</td>
<td>0.05311037</td>
<td>-0.000734268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1195788793</td>
<td>0.11960684</td>
<td>0.033002163</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_2^{(EUL)} - \theta_2$</td>
<td>Bias</td>
<td>0.0355795966</td>
<td>0.04013549</td>
<td>0.001646531</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1136035832</td>
<td>0.11629094</td>
<td>0.0335298995</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_2^{(l=3)} - \theta_2$</td>
<td>Bias</td>
<td>0.0484253876</td>
<td>0.05158365</td>
<td>-0.000734271</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1195557966</td>
<td>0.11967522</td>
<td>0.0330023216</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_2^{(l=4)} - \theta_2$</td>
<td>Bias</td>
<td>0.0484283861</td>
<td>0.05140656</td>
<td>-0.000734271</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1194060203</td>
<td>0.11978155</td>
<td>0.0330023216</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_3^{(MLE)} - \theta_3$</td>
<td>Bias</td>
<td>0.0000480238</td>
<td>0.00011431</td>
<td>0.0000414685</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0006872011</td>
<td>0.00344463</td>
<td>0.0067027202</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_3^{(EUL)} - \theta_3$</td>
<td>Bias</td>
<td>-0.0006258604</td>
<td>-0.00236893</td>
<td>0.0056688116</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0006562896</td>
<td>0.00341859</td>
<td>0.0070308394</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_3^{(l=3)} - \theta_3$</td>
<td>Bias</td>
<td>0.0006478822</td>
<td>0.00010803</td>
<td>0.0000414686</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0006871695</td>
<td>0.00344402</td>
<td>0.0067027103</td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_3^{(l=4)} - \theta_3$</td>
<td>Bias</td>
<td>0.0000480253</td>
<td>0.00009957</td>
<td>0.0000414686</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0006871488</td>
<td>0.00344323</td>
<td>0.0067027103</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Bias and s.d. reported in Table 1 are averages over 5000 replications. Each estimation is based on a simulated sample path with 1000 observations. $\Delta = 1/12$ for all models in Table 1. 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 the Vasicek model, $\theta = (0.06, 0.5, 0.15)$ for the CIR model and $\theta = (N/A, 0.2, 0.3)$ for the Black-Scholes model. QMLE with $l = 3$ and $l = 4$ in Black-Scholes model are the same due to the fact that the drift and diffusion coefficients are both constant after Doss transformation.
Table 2. Estimated bias and standard deviation for CIR model

<table>
<thead>
<tr>
<th></th>
<th>DGP (a)</th>
<th>DGP (b)</th>
<th>DGP (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_1^{(MLE)} - \theta_1 )</td>
<td>Bias: -0.00010845</td>
<td>S.D.: 0.00799336</td>
<td>0.00044849</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(EUL)} - \theta_1 )</td>
<td>Bias: -0.00010819</td>
<td>S.D.: 0.00820514</td>
<td>0.00150126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(I=3)} - \theta_1 )</td>
<td>Bias: -0.00010906</td>
<td>S.D.: 0.00802198</td>
<td>0.00166959</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(I=4)} - \theta_1 )</td>
<td>Bias: -0.00011320</td>
<td>S.D.: 0.00806676</td>
<td>0.00166956</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(I=3,U)} - \theta_1 )</td>
<td>Bias: -0.00010537</td>
<td>S.D.: 0.00808349</td>
<td>0.00166953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(I=4,U)} - \theta_1 )</td>
<td>Bias: -0.00010512</td>
<td>S.D.: 0.00808601</td>
<td>0.00166953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_1^{(Hermite)} - \theta_1 )</td>
<td>Bias: 0.00009491</td>
<td>S.D.: 0.06019561</td>
<td>0.02157182</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DGP (a)</th>
<th>DGP (b)</th>
<th>DGP (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_2^{(MLE)} - \theta_2 )</td>
<td>Bias: 0.05311037</td>
<td>S.D.: 0.11960684</td>
<td>0.04439189</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(EUL)} - \theta_2 )</td>
<td>Bias: 0.04013549</td>
<td>S.D.: 0.11629094</td>
<td>0.07823301</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(I=3)} - \theta_2 )</td>
<td>Bias: 0.05158365</td>
<td>S.D.: 0.11967522</td>
<td>0.07787763</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(I=4)} - \theta_2 )</td>
<td>Bias: 0.05140565</td>
<td>S.D.: 0.11978155</td>
<td>0.07783932</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(I=3,U)} - \theta_2 )</td>
<td>Bias: 0.05177577</td>
<td>S.D.: 0.12416308</td>
<td>0.07826028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(I=4,U)} - \theta_2 )</td>
<td>Bias: 0.05171348</td>
<td>S.D.: 0.12423279</td>
<td>0.07832116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_2^{(Hermite)} - \theta_2 )</td>
<td>Bias: -0.03785141</td>
<td>S.D.: 1.10732672</td>
<td>0.54270369</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DGP (a)</th>
<th>DGP (b)</th>
<th>DGP (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_3^{(MLE)} - \theta_3 )</td>
<td>Bias: 0.00011431</td>
<td>S.D.: 0.00344463</td>
<td>0.00006320</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_3^{(EUL)} - \theta_3 )</td>
<td>Bias: -0.00236893</td>
<td>S.D.: 0.00341859</td>
<td>0.00199882</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_3^{(I=3)} - \theta_3 )</td>
<td>Bias: 0.00010803</td>
<td>S.D.: 0.00344402</td>
<td>0.00199856</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_3^{(I=4)} - \theta_3 )</td>
<td>Bias: 0.00009957</td>
<td>S.D.: 0.00344323</td>
<td>0.00199854</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_3^{(I=3,U)} - \theta_3 )</td>
<td>Bias: 0.00008891</td>
<td>S.D.: 0.00349839</td>
<td>0.00200521</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_3^{(I=4,U)} - \theta_3 )</td>
<td>Bias: -0.00007126</td>
<td>S.D.: 0.01702433</td>
<td>0.03227378</td>
</tr>
</tbody>
</table>

Notes: Bias and s.d. reported in Table 2 are averages over 5000 replications. Each estimation is based on a simulated sample path with 1000 observations. \( \Delta = 1/12 \) for all DGPs in Table 2. Estimators with superscript U are obtained from untransformed model. Estimators with superscript Hermite are obtained from the method in Ait-Sahalia (2002). DGP (a): \( \theta = (0.06, 0.5, 0.15) \). DGP (b): \( \theta = (0.06, 0.5, 0.03) \). DGP (c): \( \theta = (0.08, 0.24, 0.08838) \).