4-1-2013

Quasi-Maximum Likelihood Estimation of Multivariate Diffusions

Xiao Huang
Kennesaw State University

Quasi-maximum Likelihood Estimation of Multivariate Diffusions*

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

September 2011

Abstract

This paper introduces quasi-maximum likelihood estimator for multivariate diffusions based on discrete observations. A numerical solution to the stochastic differential equation is obtained by higher order Wagner-Platen approximation and it is used to derive the first two conditional moments. Monte Carlo simulation shows that the proposed method has good finite sample property for both normal and non-normal diffusions. In an application to estimate stochastic volatility models, we find evidence of closeness between the CEV model and the GARCH stochastic volatility model. This finding supports the discrete time GARCH modeling of market volatility.

JEL classification: C13; C32
Keywords: multivariate diffusion; quasi-maximum likelihood estimator; Wagner-Platen approximation; stochastic volatility

*I thank the editor Bruce Mizrach and two referees for many constructive comments. I also thank Yong Bao, Robert Kimmel, Stefano Mazzotta, and seminar participants at the 2009 Meetings of the Midwest Econometrics Group for helpful comments. I am also grateful to Eckhard Platen for his valuable comments and suggestions. All remaining errors are mine. Financial support from the Coles College of Business at Kennesaw State University is greatly acknowledged. Correspondence address: Department of Economics, Finance and Quantitative Analysis, M.D. #403, Kennesaw State University, Kennesaw, GA 30144, U.S.A. Phone: 1-770-423-6318. Fax: 1-770-499-3209. E-mail: xhuang3@kennesaw.edu.

1 Introduction

Diffusion processes are frequently used to model continuous time variables in many scientific fields including biology, chemistry, economics, physics, etc. Stochastic differential equations (SDEs) are probabilistic approaches to diffusions and now widely used to characterize diffusion processes. Because the closed-form transition density of a diffusion is usually hard to obtain and the sampling interval in practice is not zero, maximum likelihood estimator (MLE) based on true density is inapplicable to many parametric SDEs. Various methods have been developed to estimate the parameters in univariate parametric SDEs (see Aït-Sahalia, 2007 and Hurn *et al.*, 2007 for recent reviews). More recently, Beskos *et al.* (2009) propose a Monte Carlo MLE for discretely observed diffusions and Phillips and Yu (2009) introduce a two-stage estimator.

Although many estimation methods for univariate diffusions can be extended to multivariate cases, only a few papers specifically address the issue of estimation of multivariate SDEs. Recent developments include nonparametric method in Bianchi (2007), Markov chain Monte Carlo method in Kalogeropoulos (2007) and Golightly and Wilkinson (2008), and Hermite polynomials approximation in Aït-Sahalia (2008). The method in Aït-Sahalia (2008) offers a closed-form expansion for the transition density and yields high numerical precision for a large class of SDEs.

This paper introduces quasi-maximum likelihood estimator (QMLE) for multivariate diffusions. QMLE in previous research is based on low-order Euler approximation, which is referred to as order 0.5 strong Wagner-Platen approximation in Kloeden and Platen (1999), and it may not yield precise estimates if coefficients in the SDE are varying, nonlinear, or the sampling interval is not close to zero.[1] Within a univariate framework, transform function together with low-order approximations are used for estimation in Kelly *et al.* (2004). When the first two conditional moments are obtained from a higher-order numerical solution to a SDE, this paper shows that QMLE can be numerically very precise. Simulation also shows the proposed method is numerically robust to non-normal diffusions. Normalizing the diffusion matrix to an identity matrix is not required for QMLE, and the proposed method can be viewed as a refinement of the popular Euler method. We also compare QMLE with the method in Aït-Sahalia (2008) in both simulation and application studies.

We apply QMLE to the estimation of three stochastic volatility models: the Heston model, GARCH stochastic volatility model, and the constant elasticity of

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

volatility (CEV) model. Nelson (1990) proves the diffusion limit of GARCH model in Bollerslev (1986), bridging the gap between discreet and continuous time volatility modeling. Using QMLE and 20 years (1990-2009) of market and volatility index data, we provide empirical evidence that the market volatility under consideration can be described by a GARCH stochastic volatility model, and this finding provides additional support to the use of discrete time GARCH method for market volatility. The estimation results for two sub-periods 1990-1999 and 2000-2009 have similar conclusions, indicating the result is robust to different time periods.

This paper is an extension of the method in Huang (2010). We make several further contributions. First, we study the estimation for multivariate diffusions. Second, we use an example to show how QMLE can be used to estimate stochastic volatility models. Third, we derive the convergence result for Wagner-Platen approximation under more general assumptions and greatly expand the class of diffusions that can be estimated by QMLE.

The rest of paper is organized as follows. Section 2 presents a general parametric SDE and standard assumptions for the existence and uniqueness of its solution. Section 3 introduces the strong Wagner-Platen expansion and approximation in Kloeden and Platen (1999) and shows that the approximation converges in probability. Section 4 introduces QMLE and discusses an example of QMLE. Section 5 uses Monte Carlo simulations to study the properties of QMLE in both normal and non-normal diffusions. Section 6 applies the proposed method to the estimation of stochastic volatility models. Section 7 concludes. The proof and the approximation expressions used in simulation are deferred to the Appendix.

2 The model and assumptions

Let us consider a multivariate diffusion of the following type

$$dX_t = a(X_t;\theta)\,dt + b(X_t;\theta)\,dW_t, \tag{1}$$

where we define a $d \times 1$ parametric vector function $a : \Re^d \to \Re^d$ and a parametric $d \times m$-matrix function $b : \Re^d \to \Re^{d\times m}$, and $\Re^d$ and $\Re^{d\times m}$ are the d-dimensional and $d \times m$-dimensional Euclidean spaces, respectively. W_t is an $m \times 1$ vector Wiener process with independent components and θ is a $p \times 1$ parameter vector. The diffusion coefficient, b, can be asymmetric. The diffusion model in (1) is defined in continuous time, but data are always observed in discrete time. Consider two observations $X_{t_{q-1}}$ and X_{t_q} with $0 = t_0 < \cdots < t_{q-1} < t_q < \cdots < t_n = T$ and $q = 1, \cdots, n$, and the discretization interval is $\Delta = t_q - t_{q-1}$. The interval is assumed to be fixed, though it can be deterministic or random. We also assume a and b are constrained such that discretely observations, $\{X_{t_q}\}$, form stationary and ergodic time series.

Extension of (1) to nonstationary and nonergodic time-inhomogeneous diffusion is briefly discussed in Section 4.

The transition density $p^*\left(X_{t_q}|X_{t_{q-1}}\right)$ plays a central role in the likelihood approach to parameter estimation and it is unknown for many diffusions defined in (1). For a diffusion process at time $t_q \in [0,T]$, Kloeden and Platen (1999) suggest that a *pathwise unique strong solution* for (1) can be obtained using strong Wagner-Platen approximation method (see Platen and Heath, 2006, for a brief introduction). The solution is a functional of the initial value X_0 and the Wiener process on $[0,t_q]$, and it converges strongly to X_{t_q} as $\Delta \to 0$, which further implies the conditional first and second moments of X_{t_q} based on the numerical solution are correctly specified. This suggests that QMLE for θ in Bollerslev and Wooldridge (1992) is consistent if we replace the unknown conditional density $p^*\left(X_{t_q}|X_{t_{q-1}}\right)$ with a normal density. To apply the strong approximation method, we first need to show the existence and uniqueness of a strong solution to (1), which is guaranteed by a set of assumptions adapted from those in Section 4.5 of Kloeden and Platen (1999). Let $\{\mathscr{F}_t,\, t \geq 0\}$ be a family of σ-algebras generated by W_t for all $t \in [0,T]$ and $|\cdot|$ be the Euclidean norm.

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

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

Assumption 3. The starting value X_0 is $\mathscr{F}_0$-measurable with $E(|X_0|^2) < +\infty$.

Similar assumptions can be found in Section 5.2 in Karatzas and Shreve (1991). Infinite differentiability in Assumption 1 is stronger than the Lipschitz condition in Kloeden and Platen (1999). It allows us to construct higher-order Wagner-Platen approximations where successive differentiation of a and b w.r.t. x is needed, similar to that in Aït-Sahalia (2008). Notice that differentiability also implies that both a and b are locally Lipschitz and measurable in x. The Lipschitz condition on a and b is used to prove the uniqueness of a strong solution to (1). Linear growth bound in Assumption 2 prevents the sample path of X_t from exploding in finite time and hence is used to prove the existence of a strong solution to (1).

Above assumptions cover a large class of SDEs. Assumption 1 excludes some boundary points for certain diffusions. Consider the special yet popular univariate example of Cox, Ingersoll and Ross (CIR) model for short term nominal interest rate $dX_t = \theta_2(\theta_1 - X_t)\,dt + \theta_3\sqrt{X_t}dW_t$, which is defined on $[0,\infty)$. Although nominal interest rate is unlikely to be zero, $b = \theta_3\sqrt{X_t}$ is not differentiable at the boundary point 0, violating Assumption 1. The diffusion degenerates and in general the operators in (4) and (5) used in higher-order approximation are not defined at 0. Hence the domain that is most interesting and relevant to our study for the CIR model is $(0,\infty)$. Even on the domain $[0,\infty)$, we can still prove the uniqueness and existence of a strong solution to (1) by using the weaker Yamada condition (see Section 4.5 in Kloeden and Platen, 1999, for a discussion).

Under Assumptions 1-3, Theorem 4.5.3 in Kloeden and Platen (1999) proves the existence of a *pathwise unique strong solution* to a univariate SDE. Extension of the result to a multivariate SDE in (1) can be done by replacing the absolute values with matrix norms in the original proof. Alternatively, one may consult similar results in Stroock and Varadhan (1979) or Karatzas and Shreve (1991). Henceforth, we assume a strong solution to (1) exists and is unique.

3 Strong Wagner-Platen expansion and approximation

Consistency of QMLE is determined by correct specification of the first and second conditional moments of X_{t_q} given $X_{t_{q-1}}$. These conditional moments are obtained through strong Wagner-Platen approximations which in turn are based on strong Wagner-Platen expansions. In this section we briefly review these expansion and approximation methods in Kloeden and Platen (1999) and show that strong Wagner-Platen approximations converge to X_{t_q} in probability as $\Delta \to 0$ for any fixed order of approximation.

3.1 Wagner-Platen expansion

Strong Wagner-Platen expansions generalize the deterministic Taylor formula to processes involving Itô stochastic integral. Consider a solution to (1):

$$X_{t_q} = X_{t_{q-1}} + \int_{t_{q-1}}^{t_q} a\left(X_u;\theta\right) du + \int_{t_{q-1}}^{t_q} b\left(X_u;\theta\right) dW_u, \tag{2}$$

and its ith element is

$$X_{t_q}^i = X_{t_{q-1}}^i + \int_{t_{q-1}}^{t_q} a^i\left(X_u;\theta\right) du + \sum_{j=1}^{m} \int_{t_{q-1}}^{t_q} b^{i,j}\left(X_u;\theta\right) dW_u^j, \tag{3}$$

where a^i is the ith element of a and $b^{i,j}$ is the ijth element of b with $i = 1, \cdots, d$ and $j = 1, \cdots, m$. Through repeated use of Itô formula, both coefficients, a^i and $b^{i,j}$, can be expanded at $X_{t_{q-1}}$. For example, by using once the Itô formula in equation (3.4.6) in Kloeden and Platen (1999), we obtain the following strong Wagner-Platen expansion for $X_{t_q}^i$ given $X_{t_{q-1}}$

$$X_{t_q}^i \;=\; X_{t_{q-1}}^i + \int_{t_{q-1}}^{t_q} \Bigg(a^i(X_{t_{q-1}};\theta) + \int_{t_{q-1}}^{u} L^0 a^i(X_z;\theta) dz$$

$$
\begin{aligned}
&+\sum_{j=1}^{m}\int_{t_{q-1}}^{u}L^{j}a^{i}(X_z;\theta)dW_z^{j}\Big)du \\
&+\sum_{j=1}^{m}\int_{t_{q-1}}^{t_q}\Big(b^{i,j}(X_{t_{q-1}};\theta)+\int_{t_{q-1}}^{u}L^{0}b^{i,j}(X_z;\theta)dz \\
&+\sum_{j_1=1}^{m}\int_{t_{q-1}}^{u}L^{j_1}b^{i,j}(X_z;\theta)dW_z^{j_1}\Big)dW_u^{j} \\
=\;& X_{t_{q-1}}^{i}+a^{i}(X_{t_{q-1}};\theta)\int_{t_{q-1}}^{t_q}du+\sum_{j=1}^{m}b^{i,j}(X_{t_{q-1}};\theta)\int_{t_{q-1}}^{t_q}dW_u^{j}+R,
\end{aligned}
$$

where

$$
\begin{aligned}
R \;=\;& \int_{t_{q-1}}^{t_q}\int_{t_{q-1}}^{u}L^{0}a^{i}(X_z;\theta)\,dzdu+\sum_{j=1}^{m}\int_{t_{q-1}}^{t_q}\int_{t_{q-1}}^{u}L^{j}a^{i}(X_z;\theta)\,dW_z^{j}du \\
&+\sum_{j=1}^{m}\int_{t_{q-1}}^{t_q}\int_{t_{q-1}}^{u}L^{0}b^{i,j}(X_z;\theta)\,dzdW_u^{j} \\
&+\sum_{j=1}^{m}\sum_{j_1=1}^{m}\int_{t_{q-1}}^{t_q}\int_{t_{q-1}}^{u}L^{j_1}b^{i,j}(X_z;\theta)\,dW_z^{j_1}dW_u^{j},
\end{aligned}
$$

$$
L^{0} = \sum_{k=1}^{d}a^{k}\frac{\partial}{\partial x^{k}}+\frac{1}{2}\sum_{k,l=1}^{d}\sum_{j=1}^{m}b^{k,j}b^{l,j}\frac{\partial^2}{\partial x^{k}\partial x^{l}}, \tag{4}
$$

$$
L^{j} = \sum_{k=1}^{d}b^{k,j}\frac{\partial}{\partial x^{k}}. \tag{5}
$$

Integrands in the remainder R can be further expanded at the point $X_{t_{q-1}}$ by applying Itô formula to obtain higher order expansions. If a and b are infinitely differentiable in x, above expansion can be continued until desired precision is reached. General results for strong Wagner-Platen expansion is summarized in Theorem 5.5.1 in Kloeden and Platen (1999).

We introduce below the notation in Chapter 5 of Kloeden and Platen (1999) to derive the convergence result. Consider a multi-index α of length l such that $\alpha = (j_1, j_2, \cdots, j_l)$, where $j_i \in \{0,1,\cdots,m\}$ for $i = 1,2,\cdots,l$ and $l := l(\alpha) \in \{1,2,\cdots\}$. Let $\mathscr{M}$ be the set of all multi-indices such that

$$
\mathscr{M} = \{(j_1, j_2, \cdots, j_l) : j_i \in \{0,1,\cdots,m\}, i \in \{1,2,\cdots,l\}, \text{for } l = 1,2,\cdots\} \cup \{v\},
$$

where v is the multi-index of length zero. For an $\alpha \in \mathscr{M}$ with $l(\alpha) \geq 1$, we let $-\alpha$ and $\alpha-$ be the multi-index in $\mathscr{M}$ obtained by deleting the first and last element of α, respectively. Let $f(t)$ be a right continuous stochastic process with left hand limits for $t \geq 0$. Define the *multiple Itô integral* as

$$I_\alpha [f(\cdot)]_{t_{q-1},t_q} = \begin{cases} f(t_q) & \text{if } l = 0 \\ \int_{t_{q-1}}^{t_q} I_{\alpha-} [f(\cdot)]_{t_{q-1},u} du & \text{if } l \geq 1 \text{ and } j_l = 0 \\ \int_{t_{q-1}}^{t_q} I_{\alpha-} [f(\cdot)]_{t_{q-1},u} dW_u^{j_l} & \text{if } l \geq 1 \text{ and } j_l \geq 1. \end{cases} \tag{6}$$

For example, when $\alpha = (1,0,5,2)$, we have

$$I_\alpha [f(\cdot)]_{t_{q-1},t_q} = \int_{t_{q-1}}^{t_q} \int_{t_{q-1}}^{u_4} \int_{t_{q-1}}^{u_3} \int_{t_{q-1}}^{u_2} f(\cdot) dW_{u_1}^1 du_2 dW_{u_3}^5 dW_{u_4}^2,$$

where $W_{u_1}^1$, $W_{u_4}^2$ and $W_{u_3}^5$ correspond to the 1st, 2nd, and 5th Wiener process in W_t provided $m \geq 5$. When $\alpha = (0,0)$, we have

$$I_{(0,0)} = \int_{t_{q-1}}^{t_q} \int_{t_{q-1}}^{s_2} ds_1 ds_2 = \frac{1}{2!} (t_q - t_{q-1})^2 = \frac{1}{2} \Delta^2,$$

where $I_{(0,0)} [1]_{s,t}$ is abbreviated as $I_{(0,0)}$ if $f(t) \equiv 1$, and this abbreviation will be used throughout this paper. With this notation and the result in Theorem 5.5.1 in Kloeden and Platen (1999), let us consider a simple bivariate SDE

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \begin{pmatrix} \beta(X_t^1, X_t^2) \\ \zeta(X_t^1, X_t^2) \end{pmatrix} dt + \begin{pmatrix} \xi(X_t^1) & 0 \\ 0 & \phi(X_t^2) \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix} \tag{7}$$

with $d = 2$ and $m = 2$. When $l(\alpha) = 2$, the expansion for X_t^1 is

$$\begin{aligned} X_{t_q}^1 &= X_{t_{q-1}}^1 + \beta I_{(0)} + \xi I_{(1)} + \xi \xi' I_{(1,1)} + (\beta \xi' + 0.5 \xi^2 \xi'') I_{(0,1)} \\ &\quad + \phi \beta^{(0,1)} I_{(2,0)} + \xi \beta^{(1,0)} I_{(1,0)} \\ &\quad + (\zeta \beta^{(0,1)} + 0.5 \phi^2 \beta^{(0,2)} + \beta \beta^{(1,0)} + 0.5 \xi^2 \beta^{(2,0)}) I_{(0,0)} + R, \end{aligned} \tag{8}$$

where the coefficients for multiple Itô integrals is obtained through the coefficient function

$$f_\alpha = \begin{cases} f & \text{if } l = 0 \\ L^{j_1} f_{-\alpha} & \text{if } l \geq 1 \end{cases}, \tag{9}$$

and we let $f(\cdot) \equiv x^1$ in (8).[2] The operator L^{j_1} is defined in (4) and (5), depending on the value of j_1. For example, the coefficient for $I_{(1,1)}$ is $f_{(1,1)} = L^1 L^1 X_t^1 = \xi \xi'$.

[2]Let $\beta^{(i,j)}$ denote the ith derivative of β w.r.t. x^1 and jth derivative w.r.t. x^2. Let ξ', ξ'', and $\xi^{(r)}$ denote the 1st, 2nd, and rth derivative of ξ w.r.t. x^1 with $r \geq 3$. Similar definitions apply to ζ and ϕ in (8).

3.2 Wagner-Platen approximations and its convergence

Given an expansion such as (8) and a discretization interval $\Delta = t_q - t_{q-1}$, we can obtain a strong Wagner-Platen approximation of the Itô diffusion in (3). Define $\Delta W^j = W^j_{t_q} - W^j_{t_{q-1}}$ for $j = 1, \cdots, m$. When $j = 1$, using the following result

$$I_{(1,1)} = \int_{t_{q-1}}^{t_q} \int_{t_{q-1}}^{u} dW_z^1 dW_u^1 = 0.5((\Delta W^1)^2 - \Delta), \tag{10}$$

the term $\xi\xi' I_{(1,1)}$ in (8) can be replaced with $0.5\xi\xi'((\Delta W^1)^2 - \Delta)$, and both ξ and ξ' are evaluated at $X_{t_{q-1}}$. Replacing other integrals in (8) in a similar way and omitting R gives a strong Wagner-Platen approximation. It may be difficult to express stochastic integrals of higher multiplicity in closed forms in terms of ΔW^j and Δ, and approximation method discussed in Section 5.8 in Kloeden and Platen (1999) can be used. However, approximation to I_α is not needed and only expectations and covariances of those integrals are used in estimation. A general result of strong Wagner-Platen approximation for $X^i_{t_q}$ given $X_{t_{q-1}}$ is given by

$$Y^{i,\Delta}_{t_q} = X^i_{t_{q-1}} + \sum_{\alpha \in \mathscr{A}_\gamma \setminus \{v\}} I_\alpha \left[f_\alpha \left(X_{t_{q-1}} \right) \right]_{t_{q-1}, t_q}, \tag{11}$$

where $\mathscr{A}_\gamma = \left\{ \alpha \in \mathscr{M} : l(\alpha) + n(\alpha) \le 2\gamma \text{ or } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}$, $l(\alpha)$ is the length of α, $n(\alpha)$ is the number of zeros in α, $\mathscr{M}$ is the set of multi-indices defined earlier, and $\gamma = 0.5, 1, 1.5, \cdots$ is the order of approximation. We note that f_α is the coefficient function defined in (9) with $f = x^i$, where it is understood that the operators defined in (4) and (5) are applied to each element of x in the expansion. The approximation in (11) is essentially the *order γ strong Wagner-Platen approximation* in equation (10.6.4) in Kloeden and Platen (1999).

Let $\mathscr{H}_\alpha$ denote the sets for multi-indicies $\alpha \in \mathscr{M}$ such that $f_\alpha(x)$ is square integrable in time for $l(\alpha) > 1$, $\mathscr{B}(\mathscr{A}_\gamma) = \left\{ \alpha \in \mathscr{M} \setminus \mathscr{A}_\gamma : -\alpha \in \mathscr{A}_\gamma \right\}$, and C^2 denote the space of two times continuously differentiable functions in x.[3]

Theorem 1. Let $Y^{i,\Delta}_{t_q}$ be the *order γ strong Wagner-Platen approximation* defined in (11) with $0 < \Delta < 1$ and fixed γ. Let r be a finite positive integer. Assume $E(|X_{t_0}|^{2r}) < +\infty$ and suppose the coefficient functions in (9) have at most order r polynomial growth[4]

$$|x|^{-r} |f_\alpha(x)| \quad \le \quad K_1 \text{ if } |x| \ge 1, \tag{12}$$

[3] I would like to thank Peter E. Kloeden for clarifying some notation in Theorem 10.6.3 of Kloeden and Platen (1999).

[4] To be consistent with Assumption 1, we exclude all boundaries (0 or $\pm\infty$) for x.

$$|x|^r|f_\alpha(x)| \quad \leq \quad K_1 \text{ if } |x|<1, \tag{13}$$

for all $\alpha \in \mathscr{A}_\gamma \cup \mathscr{B}\left(\mathscr{A}_\gamma\right)$ and $x \in \Re^d$. Under Assumptions 1-3 and for all $i = 1, \cdots, d$, the approximation $Y_{t_q}^{i,\Delta}$ given $X_{t_{q-1}}$ satisfies

$$E(\left|X_{t_q}^i - Y_{t_q}^{i,\Delta}\right|) \to 0 \text{ as } \Delta \to 0 \tag{14}$$

and

$$\lim_{\Delta \to 0}(P\left|X_{t_q}^i - Y_{t_q}^{i,\Delta}\right| < \varepsilon) = 1 \text{ for every } \varepsilon > 0. \tag{15}$$

K_1 is a positive constant independent of Δ.

See Appendix for the proof. Result (14) is similar to that in Theorem 10.6.3 and Corollary 10.6.1 in Kloeden and Platen (1999), but is derived under more general assumptions. Theorem 10.6.3 in Kloeden and Platen (1999) derives the uniform (in time t) convergence of strong Wagner Platen approximation to X_t on $[0,T]$ under the following assumptions

$$|f_\alpha(x) - f_\alpha(y)| \leq K_2|x-y| \tag{16}$$

$$|f_\alpha(x)| \leq K_3(1+|x|) \tag{17}$$

and

$$\sqrt{E(\left|X_{t_{q-1}} - Y_{t_{q-1}}^{\Delta}\right|^2)} \leq K_4 \Delta^\gamma. \tag{18}$$

The assumption in (16) is a Lipschitz condition for the coefficient function. Since the coefficient function is a function of derivatives of $a(x;\theta)$ and $b(x;\theta)$, and both $a(x;\theta)$ and $b(x;\theta)$ are infinitely differentiable in x under our assumption in Section 2, the condition in (16) is automatically satisfied. Assumption (18) is satisfied since $Y_{t_{q-1}}^{i,\Delta} = X_{t_q-1}$ for each pair of observation $(X_{t_{q-1}}, X_{t_q})$. In this paper, we relax the linear growth condition in (17) to polynomial growth conditions in (12) and (13). Note that the order of polynomial in (12) and (13) can be large as long as $r < +\infty$, which covers a large class of parametric diffusions used in economics and finance. Consider an example of (7) with $\beta = \theta_2(\theta_1 - X_1)$ and $\xi = \theta_3 X_1^{1/2}$. Based on the approximation in Appendix A.2, the coefficient function when $\alpha = (0,1)$ and $\theta = (0.06, 0.5, 0.15)$ is given by

$$\begin{aligned} f_\alpha &= \beta\xi' + 0.5\xi^2\xi'' \\ &= \frac{-0.000421875}{\sqrt{X_1}} + \frac{0.0375(0.06-X_1)}{\sqrt{X_1}}, \end{aligned}$$

where $f_\alpha \to +\infty$ as $X_1 \to 0^+$. However, notice that $(X_1)^r f_\alpha < +\infty$ when $X_1 \to 0^+$ for all integer $r \geq 1$, and the assumption in (13) is satisfied. All diffusions in our simulation and application can be shown to satisfy the polynomial growth assumption in (12) and (13). In practice, there is no need to verify whether every f_α meets the linear polynomial growth assumption since given a fixed approximation order (γ) and specific parameter values (θ), as long as $a(x;\theta)$ and $b(x;\theta)$ satisfy Assumptions 1 and 2, we can always find a finite r such that f_α satisfies (12) and (13).

4 QMLE and an example

Define the true (but unknown) conditional moments as $\mu_{t_q} \equiv E\left(X_{t_q}|X_{t_{q-1}}\right)$ and $\Omega_{t_q} \equiv Var\left(X_{t_q}|X_{t_{q-1}}\right)$. Based on (11), define the approximate conditional moments as $\mu_{t_q,\Delta} \equiv E(Y_{t_q}^{\Delta}|X_{t_{q-1}})$ and $\Omega_{t_q,\Delta} \equiv Var(Y_{t_q}^{\Delta}|X_{t_{q-1}})$. Result (15) indicates $Y_{t_q}^{\Delta} \to X_{t_q}$ in probability as $\Delta \to 0$ for a fixed γ. Theorem C in Section 1.4 of Serfling (1980) then implies

$$\lim_{\Delta\to 0} \mu_{t_q,\Delta} = \mu_{t_q} \text{ and } \lim_{\Delta\to 0} \Omega_{t_q,\Delta} = \Omega_{t_q}. \tag{19}$$

Note that in order to use Theorem C in Serfling (1980), we need to show that $E|X_{t_q}| < +\infty$ and $Y_{t_q}^{\Delta}$ is uniformly integrable for every choice of Δ in $(0,1)$. The requirement $E|X_{t_i}| < +\infty$ follows Theorem 4.5.3 in Kloeden and Platen (1999). Uniform integrability of $Y_{t_q}^{\Delta}$ holds because for a fixed approximation order (γ), there are a fixed number of terms on the r.h.s. of (11), and both the coefficient function f_α and the Itô integral w.r.t. time are bounded when evaluated at $X_{t_{q-1}}$.

Hence, we conclude that the first two conditional moments are correctly specified as $\Delta \to 0$ for any fixed approximation order (γ), as it is stated in (19). The result in Bollerslev and Wooldridge (1992) then suggests that QMLE is consistent. See Bollerslev and Wooldridge (1992) for details about consistency and asymptotic normality of QMLE.

Remark 1. It is well-known that QMLE based on linear exponential family of distributions is generally less efficient than exact MLE (see Theorem 7.8 in White, 1994). However, since exact MLE is unavailable for most multivariate diffusions and it is easy to use QMLE, QMLE offers researchers an alternative and effective way for estimation. Simulation study Tables 1 and 2 suggests that the efficiency loss in QMLE compared to exact MLE is small for both normal and non-normal diffusions under consideration.

Remark 2. An alternative way for consistent estimation is to use generalized method of moment (GMM). The relationship between two stage QMLE and GMM estimator is discussed in section 5.4 of White (1994). However, we do not consider GMM in current paper because of different forms of estimation between GMM and QMLE.

Remark 3. The requirement of $\Delta \to 0$ is a commonly used in the literature to establish asymptotic results (see, e.g., Aït-Sahalia, 2008). We also note that the Euler method, which is based on order 0.5 strong approximation, is a special case of QMLE proposed in this paper. In practice, sampling interval is never zero and higher order QMLE usually outperforms Euler method when sampling interval is relatively large and the diffusion process has non-normal transition density (see Tables 3A, 3B, 3C, and other simulation results in Huang, 2010).

Next, let us consider a multivariate time-inhomogeneous diffusion when discretely sampled observations may be nonstationary and nonergodic

$$dX_t = a\left(t, X_t; \theta\right) dt + b\left(t, X_t; \theta\right) dW_t. \tag{20}$$

Wagner-Platen approximation is still applicable to (20), but the asymptotic results similar to those in Genon-Catalot and Jacod (1993) needs to be established. More study is needed.

To illustrate the use of QMLE based on the strong approximation in (11), we consider a simple univariate example in this section. Assume X_t in (1) is a univariate variable and let $l(\alpha) = 2$ in the strong approximation. Equation (11) becomes

$$\begin{aligned} Y_{t_q}^{\Delta} &= X_{t_{q-1}} + f_{(0)} I_{(0)} + f_{(1)} I_{(1)} + f_{(0,0)} I_{(0,0)} + f_{(0,1)} I_{(0,1)} + f_{(1,0)} I_{(1,0)} \\ &\quad + f_{(1,1)} I_{(1,1)}, \end{aligned} \tag{21}$$

where $f_{(0)} = a$, $f_{(1)} = b$, $f_{(0,0)} = aa' + 0.5b^2 a''$, $f_{(0,1)} = ab' + 0.5b^2 b''$, $f_{(1,0)} = ba'$, $f_{(1,1)} = bb'$. Note that all coefficient functions f_α in (21) are evaluated at the point $X_{t_{q-1}}$ so that they can be taken out of Itô integral, and all Itô integrals I_α in (21) has the integral interval $\left[t_{q-1}, t_q\right]$. Given the strong approximation in (21), taking conditional expectation and variance on both sides of (21) yields

$$\mu_{t_q,\Delta} = X_{t_{q-1}} + f_{(0)}\Delta + f_{(0,0)}\Delta^2/2, \tag{22}$$

$$\begin{aligned} \Omega_{t_q,\Delta} &= f_{(1)}^2 Var(I_{(1)}) + f_{(0,1)}^2 Var(I_{(0,1)}) + f_{(1,0)}^2 Var(I_{(1,0)}) + f_{(1,1)}^2 Var(I_{(1,1)}) \\ &\quad + 2 f_{(1)} f_{(0,1)} Cov(I_{(1)}, I_{(0,1)}) + 2 f_{(1)} f_{(1,0)} Cov(I_{(1)}, I_{(1,0)}) \\ &\quad + 2 f_{(0,1)} f_{(1,0)} Cov(I_{(0,1)}, I_{(1,0)}) \\ &= f_{(1)}^2 \Delta + f_{(0,1)}^2 \Delta^3/3 + f_{(1,0)}^2 \Delta^3/3 + f_{(1,1)}^2 \Delta^2/2 + f_{(1)} f_{(0,1)} \Delta^2 \\ &\quad + f_{(1)} f_{(1,0)} \Delta^2 + f_{(0,1)} f_{(1,0)} \Delta^3/3, \end{aligned} \tag{23}$$

where the two moments can be used with a normal density function to estimate θ. When obtaining (22), we note that $I_{(1)}$, $I_{(0,1)}$, $I_{(1,0)}$ and $I_{(1,1)}$ all have zero conditional expectation according to Lemma 5.7.1 in Kloeden and Platen (1999).

When obtaining (23), we note that conditional covariances such as $Cov(I_{(0)}, I_{(0,1)})$ or $Cov(I_{(1,0)}, I_{(1,1)})$ are all zero according to Lemma 5.7.2 in Kloeden and Platen (1999). Results such as $Cov(I_{(0,1)}, I_{(1,0)}) = \Delta^3/6$ can also be obtained using Lemma 5.7.2. QMLE can be obtained based on results in (22) and (23).

Higher order approximation in (11) and the calculation of conditional covariance among different stochastic integrals I_α in (23) may be complicated, but this computation burden can be greatly reduced once all those symbolic calculations in (22) and (23) are programmed in software such as *Mathematica* or *Maple*.

5 Monte Carlo simulation

We use Monte Carlo simulation to study the properties of QMLE in this section. Diffusions with both normal and non-normal transition densities are considered to investigate the numerical precision and robustness of QMLE.

5.1 A normal case: the O-U process

To gauge the precision of QMLE, it is desirable to have a multivariate diffusion process with closed-form transition density so that MLE can be used as a benchmark. We use the Ornstein-Uhlenbeck (O-U) model in Aït-Sahalia (2008). Notice that other multivariate diffusions may also have closed-form densities, but these diffusions require that elements of X_t are independent of each other. For example, consider the following bivariate SDE

$$\begin{pmatrix} dX_{1t} \\ dX_{2t} \end{pmatrix} = \begin{pmatrix} \theta_2(\theta_1 - X_{1t}) \\ \theta_4 X_{2t} \end{pmatrix} dt + \begin{pmatrix} \theta_3 X_{1t}^{1/2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dW_{1t} \\ dW_{2t} \end{pmatrix}. \tag{24}$$

This model is essentially two independent univariate diffusions and its transition density function is the product of a non-central chi-squared density function and a normal density function. See Huang (2010) for simulation results of similar univariate processes in (24), and we focus on (27) in this section.

Consider a bivariate O-U model

$$dX_t = \theta_2(\theta_1 - X_t)dt + \theta_3 dW_t, \tag{25}$$

where θ_1 is a 2×1 vector and both θ_2 and θ_3 are 2×2 invertible matrix. Let $\theta = (\theta_1', vec(\theta_2)', vec(\theta_3)')'$. For a pair of observation $\left(X_{t_{q-1}}, X_{t_q}\right)$, the transition density is given by

$$p^*\left(X_{t_q} | X_{t_{q-1}}; \theta\right) = (2\pi)^{-1/2} \left|\Omega_{t_q}^*\right|^{-1/2} \times \exp\left(-\frac{1}{2}(X_{t_q} - \mu_{t_q}^*)' \Omega_{t_q}^{*-1} (X_{t_q} - \mu_{t_q}^*)\right), \tag{26}$$

where $\mu^*_{t_q} = \theta_1 + \exp(-\theta_2\Delta)(X_{t_{q-1}} - \theta_1)$, $\Omega^*_{t_q} = \lambda - \exp(-\theta_2\Delta)\lambda\exp(-\theta_2'\Delta)$, $\lambda = \frac{1}{2tr(\theta_2)|\theta_2|}(|\theta_2|\,\theta_3\theta_3' + (\theta_2 - tr(\theta_2)\theta_3\theta_3'(\theta_2 - tr(\theta_2)')$, and exp denotes the matrix exponential in $\mu^*_{t_q}$ and $\Omega^*_{t_q}$. Parameters in linear SDEs such as (25) may not be uniquely identified in estimation when observations are sampled in discrete time (see Phillips, 1973). To avoid such identification problem in estimation, we impose restrictions on parameter space used in Aït-Sahalia (2008).

The example in this section is equation (48) in Aït-Sahalia (2008), where the diffusion coefficient, θ_3, is normalized to an identity matrix for estimation. This method of normalization is referred to as Doss transform or Lamperti transform. The transformed processes usually exhibit less variation and it may improve the precision of the estimates. Doss transform is applicable to a limited class of SDEs and we consider two non-normal examples in the next section where QMLE is applied without Doss transform.

For X_t in (25), let $X_t^{tr} = \theta_3^{-1}X_t$ and we have

$$dX_t^{tr} = \eta_2(\eta_1 - X_t^{tr})dt + dW_t, \tag{27}$$

where $\eta_1 = \theta_3^{-1}\theta_1$ is a 2×1 vector with elements η_1^i, $\eta_2 = \theta_3^{-1}\theta_2\theta_3$ is a 2×2 matrix with elements $\eta_2^{i,j}$ and $\eta_2^{2,1}$ is constrained to be zero for identification purposes. The transition density of (27) is given in (26). We simulate 500 observations for each sample path of (27) with 1000 replications. Table 1 reports the estimated bias and standard error of QMLE when $l(\alpha) = 3$ and $l(\alpha) = 4$ along with those of MLE. The parameter values for η_1, η_2, and $\Delta = 1/52$ used in simulation are the same as those used in Table 1 of Aït-Sahalia (2008). QMLE in this example gives precise estimates for parameters in (27) and its bias is almost indistinguishable from that of MLE. For example, the estimated bias for η_1^1 when $l(\alpha) = 4$ for QMLE is identical to that of MLE up to the seventh digit. Table 1 also show QMLE is comparable to the method in Aït-Sahalia (2008).

5.2 Two non-normal cases

QMLE is shown to yield high numerical precision for the O-U process in (27). In this section, we proceed to show that QMLE is also numerically precise and robust for non-normal multivariate diffusions. In addition, results in this section also suggest that the proposed QMLE may yield improvement over the Euler method, justifying the higher-order approach in this paper.

In the first example, we consider a nonlinear transformation of (27), which is the example in equation (49) of Aït-Sahalia (2008). Let $\tilde{X}_t^{tr} = \exp(X_t^{tr})$ and the

SDE for $\tilde{X}_t^{tr}$ is given by

$$
\begin{aligned}
d\tilde{X}_t^{tr} &= \begin{pmatrix} \tilde{X}_{1t}^{tr}(\frac{1}{2}+\eta_2^{1,1}(\eta_1^1-\ln(\tilde{X}_{1t}^{tr}))+\eta_2^{1,2}(\eta_1^2-\ln(\tilde{X}_{2t}^{tr}))) \\ \tilde{X}_{2t}^{tr}(\frac{1}{2}+\eta_2^{2,1}(\eta_1^1-\ln(\tilde{X}_{1t}^{tr}))+\eta_2^{2,2}(\eta_1^2-\ln(\tilde{X}_{2t}^{tr}))) \end{pmatrix} dt \\
&\quad + \begin{pmatrix} \tilde{X}_{1t}^{tr} & 0 \\ 0 & \tilde{X}_{2t}^{tr} \end{pmatrix} dW_t. \qquad (28)
\end{aligned}
$$

The transition density for $\tilde{X}_t^{tr}$ in (28) is obtained through Jacobian transformation based on $\tilde{X}_t^{tr}=\exp(X_t^{tr})$ and (26). The same set of parameter values is used in Table 2. Table 2 reports MLE as well as QMLE for (28) without the transform $X_t^{tr}=\ln\left(\tilde{X}_t^{tr}\right)$ and they are $\hat{\theta}_{l(\alpha)=3,U}^{QMLE}$ and $\hat{\theta}_{l(\alpha)=4,U}^{QMLE}$. It is clear from Table 2 that QMLE without transform yields precise estimates and it suggests that QMLE can be effectively applied to SDEs without normalizing the diffusion matrix. The method in Aït-Sahalia (2008) also yields very precise estimation results in Table 2.

In the second example, we consider the constant of elasticity of variance (CEV) model for stochastic volatility in Jones (2003)

$$
\begin{aligned}
dS_t &= \theta_1 S_t dt + \sqrt{1-\theta_4^2}\sqrt{V_t}S_t dW_t^1 + \theta_4\sqrt{V_t}S_t dW_t^2, \\
dV_t &= \theta_3(\theta_2 - V_t)dt + \theta_5 V_t^{\theta_6} dW_t^2, \qquad (29)
\end{aligned}
$$

where dW_t^1 and dW_t^2 are uncorrelated, and θ_4 represents the instantaneous correlation between two diffusions. Since the CEV model has no closed-form density, it is simulated using the Euler scheme on an interval of 0.0001, and the data are sampled on the interval of 0.01 in Table 3A with 1000 observations and 1000 replications. The parameter values are $\theta=(0.08,0.05,2,-0.5,2,1)$. These values are chosen to mimic the estimates for the CEV model in the next section to show QMLE can indeed yield precise estimates if these are true parameter values. The results are reported in Table 3A.

In Table 3A, QMLE $\hat{\theta}_{l(\alpha)=3}^{QMLE}$ and $\hat{\theta}_{l(\alpha)=4}^{QMLE}$ give good estimates across all parameters. The bias for θ_3 is relatively large compared to other estimates. θ_3 measures the speed of mean reversion in the V_t process. Bias in estimating speed of mean reversion is also observed in several other papers (see Yu, 2011, and references therein). In fact, this relatively large bias will occur even for MLE. See for example the bias for $\eta_2^{1,1}$ and $\eta_2^{2,2}$ in Tables 1 and 2. The method in Aït-Sahalia (2008) also yields good results, including a smaller bias for θ_3. Compared to Euler estimator, QMLE may provide some improvement. The improvement is most notable for parameters in the diffusion function $(\theta_5 V_t^{\theta_6})$. In a univariate framework, Huang (2010) also provides extensive simulation results for QMLE with non-normal diffusions and compares it with other popular estimation methods. It is interesting to observe that, for the DGP in Table 3A, the Euler method performs very well compared to both QMLEs and the method in Aït-Sahalia (2008). However, when we consider a larger sampling interval (Table 3B) and more volatile data (Table 3C), QMLE shows a clear advantage over the Euler method, justifying the higher order approach.

In Table 3B, we let $\Delta = 0.1$ and the rest of the DGP is the same as that in Table 3A. The proposed QMLE outperforms the Euler estimator for all parameters with substantial improvement for most parameter estimates. We also note that the bias of $\hat{\theta}^{QMLE}_{l(\alpha)=4}$ is smaller than the of bias of $\hat{\theta}^{QMLE}_{l(\alpha)=3}$ for 5 (out of 6) parameter estimates, offering some evidence that higher order approximation reduces the bias. The method in Aït-Sahalia (2008) does not yield reasonably good estimates and the results are not reported. The estimator in Aït-Sahalia (2008) is based on a polynomial (in Δ) expansion of the likelihood function. Given the large sampling interval $\Delta = 0.1$, a higher order expansion for the estimator in Aït-Sahalia (2008) is probably needed to make the method work for the DGP in Table 3B.

In Table 3C, we let mean variance $\theta_2 = 0.5$ while keeping the rest of DGP the same as that in Table 3A. This change increases the variance of the process by ten times. Overall, QMLE provides improvement over the Euler estimator and gives good estimation results. $\hat{\theta}^{QMLE}_{l(\alpha)=4}$ yields a smaller bias than $\hat{\theta}^{QMLE}_{l(\alpha)=3}$ does, consistent with the finding in Table 3B. The estimator in Aït-Sahalia (2008) also gives good results.

Results in Tables 3A, 3B, and 3C show that the proposed QMLE yields high numerical precision for the DGPs under consideration and may yield improvement over the low order Euler estimator. Results in Tables 3B and 3C further show that QMLE is numerically robust to large sampling intervals and to data with high volatility, which correspond to larger deviations from normality. This is consistent with the findings in Huang (2010).

6 An application to stochastic volatility models

In this section, we consider an application of QMLE to the estimation of stochastic volatility models. Stochastic volatility model is one of most important tools to study the dynamics of asset price volatility in financial econometrics. Since asset volatility plays a critical role in pricing financial derivatives, stochastic volatility model also has deep roots in mathematical finance. There are some difficulties in estimating stochastic volatility models. One is stochastic volatility itself can not be observed directly and volatility proxies must be used in certain likelihood-based estimations. The other difficulty is closed-form transition density for continuous-time

Table 1

Estimated bias and standard error for Ornstein-Uhlenbeck model in equation (27).

		$\hat{\theta}^{\text{MLE}} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=3} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=4} - \theta$		$\hat{\theta}^{\text{Aït-Sahalia}} - \theta$	
Parameter	θ	Bias	S.E.	Bias	S.E.	Bias	S.E.	Bias	S.E.
η_1^1	0	-0.0017287	0.0643	-0.0017286	0.0638	-0.0017287	0.0639	-0.0017270	0.0644
η_1^2	0	0.0009640	0.0321	0.0009645	0.0321	0.0009640	0.0321	0.0009580	0.0323
$\eta_2^{1,1}$	5	0.4700707	1.0931	0.4703992	1.0673	0.4700631	1.0673	0.4706834	1.0887
$\eta_2^{1,2}$	1	-0.0480278	2.1092	-0.0480867	1.5717	-0.0480248	1.5717	-0.0492733	1.5855
$\eta_2^{2,2}$	10	0.3844841	1.3980	0.3880180	1.4746	0.3843289	1.4735	0.3835503	1.5103

Notes: The DGP in Table 1 is the same as the one used in Table 1 of Aït-Sahalia (2008). Equation (27) is obtained by applying Doss transform to equation (25). The subscripts, $l(\alpha) = 3$ and $l(\alpha) = 4$, denote QMLEs associated with different orders of Wagner-Platen approximations in equation (11). Examples of approxiamtions when $l(\alpha) = 3$ are given in the Appendix. Results for Aït-Sahalia (2008) are obtained based on Model B13 in the Matlab code downloadable from http://www.princeton.edu/ yacine/research.

Table 2

Estimated bias and standard error for the model in equation (28).

		$\hat{\theta}^{\text{MLE}} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=3,U} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=4,U} - \theta$		$\hat{\theta}^{\text{Aït-Sahalia}} - \theta$	
Parameter	θ	Bias	S.E.	Bias	S.E.	Bias	S.E.	Bias	S.E.
η_1^1	0	-0.0017284	0.0640	-0.0014740	0.0664	-0.0014982	0.0664	-0.0017270	0.0644
η_1^2	0	0.0009640	0.0321	0.0011277	0.0331	0.0011114	0.0331	0.0009580	0.0323
$\eta_2^{1,1}$	5	0.4708671	1.0790	0.4858141	1.1076	0.4867146	1.1079	0.4706834	1.0887
$\eta_2^{1,2}$	1	-0.0434466	1.8527	-0.0488437	1.6209	-0.0496155	1.6202	-0.0492733	1.5855
$\eta_2^{2,2}$	10	0.3857980	1.4878	0.3925816	1.5191	0.3908455	1.5188	0.3835503	1.5103

Notes: The DGP used in Table 2 is the same as the one used in Table 2 of Aït-Sahalia (2008). QMLEs with subscript U are obtained without transforming diffusion coefficient to an unity matrix in equation (28). The Aït-Sahalia estimator is obtained from the transformed model and is the same as the one in Table 1.

Table 3A

Estimated bias and standard error for parameters in the non-normal CEV stochastic volatility model in equation (29) with sampling interval $\Delta = 0.01$.

		$\hat{\theta}^{\text{Euler}} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=3} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=4} - \theta$		$\hat{\theta}^{\text{Aït-Sahalia}} - \theta$	
Parameter	θ	Bias	S.E.	Bias	S.E.	Bias	S.E.	Bias	S.E.
θ_1	0.08	-0.002956	0.1457	-0.003266	0.0893	-0.003926	0.0778	-0.018964	0.0524
θ_2	0.05	0.026877	0.3333	0.025707	0.1081	0.033578	0.0996	0.011074	0.0422
θ_3	2	0.347869	8.1267	0.342211	4.8389	0.358446	3.8900	-0.073442	1.0984
θ_4	-0.5	0.001582	0.0255	0.000214	0.0246	0.002475	0.0239	0.001145	0.0228
θ_5	2	-0.117461	0.3879	-0.028194	0.5019	-0.043773	0.4435	-0.253128	0.1586
θ_6	1	-0.021576	0.0564	-0.010104	0.0635	-0.018074	0.0581	-0.046389	0.0254

Notes: Table 3A reports the simulation results for equation (29) with a sample size of 1000 and 1000 replications. The values for θ are chosen to be close to the estimates in Table 4 for the CEV model. Estimates using the method in Aït-Sahalia (2008) in Tables 3A, 3C , and 4 are obtained based on Model B4 in the Matlab code downloadable from: http://www.princeton.edu/ yacine/research.htm.

Table 3B

Estimated bias and standard error for parameters in the non-normal CEV stochastic volatility model in Equation (29) when sampling interval is large ($\Delta = 0.1$).

		$\hat{\theta}^{\text{Euler}} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=3} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=4} - \theta$	
Parameter	θ	Bias	S.E.	Bias	S.E.	Bias	S.E.
θ_1	0.08	-0.000721	0.0171	-0.000393	0.0171	-0.000352	0.0172
θ_2	0.05	0.000532	0.0068	0.001288	0.0081	0.001314	0.0118
θ_3	2	0.069004	0.4065	0.038110	0.4596	0.033794	0.4743
θ_4	-0.5	0.033404	0.0254	-0.002051	0.0271	-0.001725	0.0273
θ_5	2	-0.614957	0.2709	-0.050558	0.2790	-0.049913	0.2967
θ_6	1	-0.136136	0.0536	-0.010971	0.0465	-0.010886	0.0477

Table 3B reports the simulation results for equation (29) with a large sampling interval ($\Delta = 0.1$). The sample size is 1000 with 1000 replications. Aït-Sahalia (2008) estimates do not yield good results and are not reported (see Section 5.2 for a discussion).

Table 3C

Estimated bias and standard error for parameters in the non-normal CEV stochastic volatility model in equation (29) when volatility is large ($\theta_2 = 0.5$).

		$\hat{\theta}^{\text{Euler}} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=3} - \theta$		$\hat{\theta}^{\text{QMLE}}_{l(\alpha)=4} - \theta$		$\hat{\theta}^{\text{Aït-Sahalia}} - \theta$	
Parameter	θ	Bias	S.E.	Bias	S.E.	Bias	S.E.	Bias	S.E.
θ_1	0.08	-0.007469	0.3155	-0.0072027	0.3550	-0.0071422	0.3728	-0.007241	0.1656
θ_2	0.5	0.073503	0.9319	0.0906086	1.5494	0.0815765	0.9145	0.167230	1.3036
θ_3	2	0.380009	2.8823	0.3654666	4.4314	0.3671223	3.1195	0.166617	1.1017
θ_4	-0.5	0.002416	0.0231	-0.0016641	0.0251	-0.0016636	0.0233	0.000620	0.0228
θ_5	2	-0.033194	0.0931	-0.0093087	0.1364	-0.0092809	0.1295	-0.069374	0.0722
θ_6	1	-0.021815	0.0340	-0.0053298	0.0377	-0.0053268	0.0391	-0.046913	0.0256

Notes: Table 3C reports the simulation results for equation (29) when mean volatility (θ_2) is large. The sampling interval is the same as that in Table 3A ($\Delta = 0.01$) and the sample size is 1000 with 1000 replications.

Table 4

Estimation reults for three stochastic volatility models using S&P 500 Index and Volatility Index (VIX) data from 1/02/1990 to 12/31/2009.

	Heston		GARCH		CEV		CEV (Aït-Sahalia)	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
θ_1	0.14040	0.0277	0.07521	0.0931	0.07533	0.0240	0.00131	0.0569
θ_2	0.02337	0.0084	0.05070	0.0213	0.05072	0.0049	0.10007	0.0470
θ_3	1.59890	0.1303	1.96952	0.7274	1.96609	0.0945	1.19594	0.6338
θ_4	-0.75786	0.0097	-0.79616	0.0079	-0.79621	0.0078	-0.79394	0.0038
θ_5	0.54972	0.0170	2.29541	0.0591	2.28741	0.2134	1.80915	0.0215
θ_6	0.5	N/A	1	N/A	0.99891	0.0333	0.92346	0.0024

Notes: Table 4 reports QMLE with $l(\alpha) = 3$ for the Heston model, GARCH stochastic volatility model, and CEV model in equations (30), (31), and (29), respectively. The last two columns report the parameter estimates and s.e. obtained from the method in Aït-Sahalia (2008).

stochastic volatility only exists in some special cases, and it is usually unavailable for a stochastic volatility model with general specification. See Ghysels *et al.* (1996) and Asai *et al.* (2006) for recent surveys on various estimation methods for stochastic volatility models.

The purpose of this section is not to propose any new stochastic volatility models or compare different methods of estimating stochastic volatility. Instead, we would like to show how QMLE can be used to estimate some popular stochastic volatility models and discuss the implications of the estimates on stochastic volatility modeling.

6.1 The data and models

Since stochastic volatility is latent, we follow the approach similar to that in Jones (2003) by choosing the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) as the volatility measure for S&P 500 Index (SPX). The VIX is calculated as a weighted average of prices of SPX put and call options with different strike prices.[5] Realized volatility based on high frequency data is an alternative way to obtain a volatility proxy.

We use the daily SPX and VIX data from January 2, 1990 to December 31, 2009, a total of 5043 pairs of observations and set $\Delta = 1/252$ by assuming there are 252 trading days in each year. January 2, 1990 is the earliest date for available VIX data. For VIX, the data on March 1, 1991, January 31, 1997, and November 26, 1997 are missing and an average of the data from the two adjacent days are used. The VIX is the implied volatility scaled up by 100. We work with variance in the following stochastic volatility models, and the variance is calculated as $V_t = (VIX_t/100)^2$, where VIX_t is the VIX data at time t. Hence the state vector X_t in (1) is defined as $X_t = (S_t, V_t)'$, where S_t is the S&P 500 Index observed at time t. We emphasize that V_t in the following discussion actually represents variance, though the models are called stochastic volatility models.

The first model considered in estimation is the popular stochastic volatility model in Heston (1993), where a square-root process is used to describe the dynamics of volatility and S_t is assumed to follow a geometric Brownian motion,

$$\begin{aligned} dS_t &= \theta_1 S_t dt + \sqrt{1-\theta_4^2}\sqrt{V_t} S_t dW_t^1 + \theta_4 \sqrt{V_t} S_t dW_t^2 \\ dV_t &= \theta_3(\theta_2 - V_t)dt + \theta_5 \sqrt{V_t} dW_t^2, \end{aligned} \tag{30}$$

where dW_t^1 and dW_t^2 are uncorrelated.

[5] A more detailed description of VIX can be found at: http://www.cboe.com/micro/VIX/vixwhite.pdf

Next we consider the GARCH stochastic volatility model in Nelson (1990)

$$\begin{aligned} dS_t &= \theta_1 S_t dt + \sqrt{1-\theta_4^2}\sqrt{V_t} S_t dW_t^1 + \theta_4 \sqrt{V_t} S_t dW_t^2, \\ dV_t &= \theta_3(\theta_2 - V_t)dt + \theta_5 V_t dW_t^2, \end{aligned} \tag{31}$$

where dW_t^1 and dW_t^2 in (31) are uncorrelated.

Finally, the CEV model in (29) is also considered.

The estimation results based on $\hat{\theta}^{QMLE}_{l(\alpha)=3}$ are reported in Table 4. We discuss the results in Table 4 in the following remarks.

6.2 Remarks

Remark 1. The estimates for θ_4 in all three models are negative and statistically significant based on conventional level of significance of 5%, implying negative instantaneous correlation between SPX and VIX. Since the option market for S&P 500 Index is very liquid and active, an increase in volatility will quickly lead to a drop in SPX, giving a higher expected rate of return as investors request more premium to compensate the additional risk. Alternatively, this negative estimate for θ_4 can be explained through the instantaneous leverage effect: as asset price (SPX) drops, increased financial leverage will lead to an increase in volatility. Either way, the sign of $\hat{\theta}_4$ is consistent with both explanations and with previous findings in literature.

However, the magnitude of the correlation varies slightly across different models. For Heston model, we have $\hat{\theta}_4 \approx -0.758$, while for the GARCH stochastic volatility model and CEV model, it is $\hat{\theta}_4 \approx -0.796$. Note that the correlations are large in absolute value, suggesting a univariate modeling of SPX without the stochastic variance V_t may miss the important feedback effect from the volatility (and variance) side.

Remark 2. The estimates in both GARCH and CEV model are quite close for all parameters. In particular, we observe $\hat{\theta}_6$ in CEV model is very close to 1, the fixed exponent in the GARCH stochastic volatility model. In fact, we can not reject the null of $H_0 : \theta_6 = 1$ at all conventional levels of significance. This suggests the GARCH stochastic volatility model in (31) may be a good approximate model for the underlying SPX variance process. The model in (31) is derived in Nelson (1990) to show that the popular discrete-time GARCH model for conditional variance is indeed consistent with continuous-time specification for asset prices in finance as sampling interval shrinks to zero. Results in Table 4 suggests that model (31), a continuous-time limit of the discrete-time GARCH model, may be used as a continuous-time specification for the market index data under consideration, and it further provides support of using GARCH model for discrete time observations.

The method in Aït-Sahalia (2008) also gives similar result for θ_6 in Table 4. In fact, Aït-Sahalia and Kimmel (2007, page 444) find the estimate of θ_6 "right at the boundary value of 1" using a smaller data set and the unadjusted VIX (the one used in this paper), although they find $\hat{\theta}_6 < 1$ using adjusted VIX. In their simulation using Black-Scholes implied volatility in the CEV model, they find significant bias in $\hat{\theta}_6$ and bias is reduced by adjusting the implied volatility. Hence they also adjust VIX in the estimation. However, note that VIX is a model-free measure of volatility and is not related the Black-Scholes model.

Our simulation study in Tables 3A, 3B and 3C suggest that estimates for CEV model in Table 4 are quite reliable. To further check the robustness of the estimates in CEV model, we split the sample into two sub-periods, 1990-1999 and 2000-2009, to examine possible differences in estimates due to different samples. QMLE for these two sub-periods are reported in Table 5. Based on the estimate $\hat{\theta}_6$ in Table 5, we cannot reject the null of $H_0 : \theta_6 = 1$ for either sub-periods at conventional levels of significance. This statistical evidence also suggests that the GARCH stochastic volatility model can describe the market index reasonably well. The drift term for instantaneous return dS_t/S_t, $\hat{\theta}_1$, changes from 0.1397 during 1990-1999 to -0.0023 during 2000-2009, reflecting the huge impact on market index returns due to two market crashes in the second sub-period.

Table 5

CEV model estimates for two sub-periods using S&P 500 Index and Volatility Index (VIX) data from 1/02/1990 to 12/31/2009

	1990-1999		2000 - 2009		1990 - 2009	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
θ_1	0.13968	0.0381	-0.00229	0.0445	0.07533	0.0240
θ_2	0.03975	0.0150	0.07282	0.0216	0.05072	0.0049
θ_3	2.86361	2.1867	1.21369	0.4889	1.96609	0.0945
θ_4	-0.74922	0.0154	-0.84044	0.0066	-0.79621	0.0078
θ_5	2.20369	0.2631	2.42855	0.3380	2.28741	0.2134
θ_6	0.98489	0.0385	1.02466	0.0534	0.99891	0.0333

Notes: Table 5 reports QMLE with $l(\alpha) = 3$ for the CEV model in equation (29) for two sub-periods: 1990-1999 and 2000-2009, along with the results for the period 1990-2009.

On the other hand, results in Tables 4 and 5 also imply that the popular Heston model may be an inadequate representation for the bivariate process $(S_t, V_t)'$.

Remark 3. We also need to acknowledge the difficulty of obtaining a good proxy for volatility. As mentioned earlier, an alternative way to obtain a volatility or variance proxy is to use high frequency data to construct realized volatility or variance. Due to market microstructure noise in ultra high frequency data, it is common to sample the data every 5 or 30 minutes for the construction of realized volatility. This, however, directly violates the assumption of zero sampling interval, which is crucial for the realized volatility to converge to the quadratic variation in theory. On the other hand, implied volatility such as VIX is obtained under risk-neutral probability measure, and it is calculated across only a finite number of strike prices in practice. The theoretical results and empirical evidence in Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) nonetheless justify the use of the model-free implied volatility. In addition, VIX has arguably become the industrial standard to measure volatility. All these suggests the use of VIX as a proxy for volatility has both theoretical and empirical support. A detailed comparison of realized volatility and the model-free implied volatility such as VIX is beyond the scope of the current paper. Finally, the simulation study in Aït-Sahalia and Kimmel (2007) shows using a volatility proxy introduces little numerical noise for MLE.

7 Conclusions

In this paper we introduce QMLE for multivariate diffusion processes defined in (1). We use the higher order Wagner-Platen approximation in Kloeden and Platen (1999) to obtain the first two conditional moments of the diffusion process and compute the likelihood function based on a normal density function.

This methodology has several attractive features. First, higher order approximation and the multivariate normal density offer a closed-form density for likelihood estimation and inference. Second, QMLE only requires the drift and diffusion coefficient in (1) to be differentiable in both state variables and parameters. Once programmed, it can be conveniently applied to arbitrary multivariate diffusions as long as the parameters can be identified from discrete observations. This method covers a large class of SDEs and is easy to implement.

The application study based on market index data reveals the similarity between GARCH stochastic volatility model and the CEV model, providing further support to the GARCH modeling of volatility in discrete times.

Extensions of current work are possible. The drift and diffusion coefficient for some SDEs may not be differentiable and it is interesting to investigate whether other strong approximations such as Runge-Kutta method can be used for QMLE. Applications of this method in economics, finance, and other scientific fields are also needed to further study its property. We leave these topics for future research.

Appendix A

A.1 Proof of Theorem 1

Consider the order γ strong Wanger-Platen expansion for $X^i_{t_q}$ in equation (5.5.3) in Kloeden and Platen (1999)

$$
\begin{aligned}
X^i_{t_q} &= \sum_{\alpha\in\mathscr{A}_\gamma} I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q} + \sum_{\alpha\in\mathscr{B}_\gamma\setminus\{v\}} I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q} \\
&= X^i_{t_{q-1}} + \sum_{\alpha\in\mathscr{A}_\gamma\setminus\{v\}} I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q} + \sum_{\alpha\in\mathscr{B}_\gamma\setminus\{v\}} I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q} \quad (32)
\end{aligned}
$$

and the order γ strong Wanger-Platen approximation for $X^i_{t_q}$ in equation (11)

$$Y^{i,\Delta}_{t_q} = X^i_{t_{q-1}} + \sum_{\alpha\in\mathscr{A}_\gamma\setminus\{v\}} I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q}. \quad (33)$$

Based on (32) and (33), we have

$$Z_{t_q} = E(|X^i_{t_q} - Y^{i,\Delta}_{t_q}|^2|\mathscr{F}_{t_{q-1}}) \le K_3 \sum_{\alpha\in\mathscr{B}_\gamma\setminus\{v\}} U^\alpha_{t_q}, \quad (34)$$

where $U^\alpha_{t_q} := E(|I_\alpha\left[f_\alpha\left(X_{t_{q-1}}\right)\right]_{t_{q-1},t_q}|^2|\mathscr{F}_{t_{q-1}})$.

Using Lemma 10.8.1 in Kloeden and Platen (1999) and $t_q - t_{q-1} = \Delta$, we obtain

$$U^\alpha_{t_q} \le \begin{cases} \Delta^{2l(\alpha)-1}\int_{t_{q-1}}^{t_q} E(|f_\alpha(X_{t_{q-1}})|^2|\mathscr{F}_{t_{q-1}})du & \text{if } l(\alpha)=n(\alpha) \\ 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)-1}\int_{t_{q-1}}^{t_q} E(|f_\alpha(X_{t_{q-1}})|^2|\mathscr{F}_{t_{q-1}})du & \text{if } l(\alpha)\ne n(\alpha) \end{cases} \quad (35)$$

For the domain $(-\infty,+\infty)$ or $[0,+\infty)$, we discuss two cases: case 1 when $|x^i| \ge 1$ and case 2 when $0 \le |x^i| < 1$.

Case 1: $|x^i| \ge 1$

Without loss of generality, we assume $X^1_{t_{q-1}} = \max\{|X^1_{t_{q-1}}|,\cdots,|X^d_{t_{q-1}}|\}$. Under the rth order polynomial growth assumption in (12) and the result that Euclidean norm is less than or equal to the l_1-norm, we have

$$
\begin{aligned}
U^\alpha_{t_q} &\le \begin{cases} K_4\Delta^{2l(\alpha)-1}\int_{t_{q-1}}^{t_q} E(|X^1_{t_{q-1}}|^{2r}|\mathscr{F}_{t_{q-1}})du \\ K_4 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)-1}\int_{t_{q-1}}^{t_q} E(|X^1_{t_{q-1}}|^{2r}|\mathscr{F}_{t_{q-1}})du \end{cases} \\
&\le \begin{cases} K_4\Delta^{2l(\alpha)-1}\int_{t_{q-1}}^{t_q} (1+E(|X^1_{t_0}|^{2r}))e^{K_5 t_{q-1}}du \\ K_4 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)-1}\int_{t_{q-1}}^{t_q} (1+E(|X^1_{t_0}|^{2r}))e^{K_5 t_{q-1}}du \end{cases}
\end{aligned}
$$

$$= \begin{cases} K_4\Delta^{2l(\alpha)}(1+E(|X^1_{t_0}|^{2r}))e^{K_5 t_{q-1}} \\ K_4 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)}(1+E(|X^1_{t_0}|^{2r}))e^{K_5 t_{q-1}} \end{cases}, \quad (36)$$

where the second inequality follows Theorem 4.5.4 in Kloeden and Platen (1999) and the fact that $t_0 = 0$ and $K_5 = 2r(2r+1)K^2$, where K is the constant in Assumption 2.

Given the description of approximation order γ below (11), we can verify that $2l(\alpha) \geq 2\gamma+2 > 2\gamma+1$ when $l(\alpha) = n(\alpha)$ and $l(\alpha)+n(\alpha) \geq 2\gamma+1$ when $l(\alpha) \neq n(\alpha)$. Hence results in (36) can more compactly written as

$$U^{\alpha}_{t_q} \leq K_6(1+E(|X^1_{t_0}|^{2r}))\Delta^{2\gamma+1}e^{K_5 t_{q-1}}. \quad (37)$$

For a fixed γ, K_6 ($= K_4 4^{l(\alpha)-n(\alpha)+2}$) is a constant, and K_5 is also a constant under the assumption in (11). Since the approximation $Y^{i,\Delta}_{t_q}$ is derived for a given $X_{t_{q-1}}$, t_{q-1} is fixed in the approximation, and $U^{\alpha}_{t_q} \to 0$ as $\Delta \to 0$. Even in the case when t_{q-1} grows as the number of observation increases, $U^{\alpha}_{t_q} \to 0$ still holds as long as $\Delta \to 0$ fast enough.

Case 2: $0 \leq |x^i| < 1$ (excluding 0 if either $a(x;\theta)$ or $b(x;\theta)$ is not differentiable at the point 0).

Similar to case 1, we assume $(X^1_{t_{q-1}})^{-1} = \max\{(X^1_{t_{q-1}})^{-1}, \cdots, (X^d_{t_{q-1}})^{-1}\}$. Under the rth order polynomial growth assumption in (13), when $0 \leq x^i < 1$, (35) can be written as

$$\begin{aligned} U^{\alpha}_{t_q} &\leq \begin{cases} K_4\Delta^{2l(\alpha)-1}\int_{t_{q-1}}^{t_q} E((X^1_{t_{q-1}})^{-2r}|\mathscr{F}_{t_{q-1}})du \\ K_4 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)-1}\int_{t_{q-1}}^{t_q} E((X^1_{t_{q-1}})^{-2r}|\mathscr{F}_{t_{q-1}})du \end{cases} \\ &= \begin{cases} K_4\Delta^{2l(\alpha)}E((X^1_{t_{q-1}})^{-2r})|\mathscr{F}_{t_{q-1}}) \\ K_4 4^{l(\alpha)-n(\alpha)+2}\Delta^{l(\alpha)+n(\alpha)}E((X^1_{t_{q-1}})^{-2r})|\mathscr{F}_{t_{q-1}}) \end{cases} \end{aligned} \quad (38)$$

For the expectation $E((X^1_{t_{q-1}})^{-2r})$, let the lower bound of integration be c and $c \to 0^+$ and we have

$$\begin{aligned} E(\left(x^1\right)^{-2r}) &= \int_c^{+\infty} \left(x^1\right)^{-2r} p(x^1)dx^1 \\ &\leq (\int_c^{+\infty} \left(x^1\right)^{-4r} dx^1)^{1/2}(\int_c^{+\infty} p(x^1)^2 dx^1)^{1/2} \\ &= K_5(\frac{c^{1-4r}}{4r-1})^{1/2}, \end{aligned} \quad (39)$$

provided that $p(x^1)$ is a square integrable density function for x^1. If we do not impose the assumption of $E(\left(x^1\right)^{-2r}) < +\infty$, the r.h.s. of (39) will grow to infinity as $c \to 0^+$. However, we will still have $U_{t_q}^{\alpha} \to 0$ in (38) as long as $\Delta \to 0$ fast enough. Alternatively, we may impose an assumption on the rate at which the diffusion process approaches the boundary and require that $\Delta \to 0$ at a faster rate. Combining this analysis with the result in (37), we have $U_{t_q}^{\alpha} \to 0$ as $\Delta \to 0$. Consequently, $Z_{t_q} \to 0$ in (34) and the result in (14) holds.

The result in (15) can be obtained using Chebychev's inequality. As $\Delta \to 0$, we obtain $\mathrm{P}\left(\left|X_{t_q}^{i} - Y_{t_q}^{i,\Delta}\right| > \varepsilon\right) \leq \varepsilon^{-1} E\left(\left|X_{t_q} - Y_{t_q}^{i,\Delta}\right|\right) \to 0$which implies $Y_{t_q}^{i,\Delta} \to X_{t_q}$ in probability in (15).

A.2 Approximation expressions

Consider the SDE in (7), which nests SDEs in (27) and (28). When $l(a) = 3$, the strong Wagner-Platen approximations for $X_{t_q}^1$ and $X_{t_q}^2$ are given, respectively, by

$$
\begin{aligned}
Y_{t_q}^{1,\Delta} \quad = \quad & X_{t_{q-1}}^1 + f_{(0)}I_{(0)} + f_{(1)}I_{(1)} + f_{(1,1)}I_{(1,1)} + f_{(1,1,1)}I_{(1,1,1)} + f_{(0,1)}I_{(0,1)} \\
& + f_{(0,1,1)}I_{(0,1,1)} + f_{(2,0)}I_{(2,0)} + f_{(2,0,1)}I_{(2,0,1)} + f_{(2,2,0)}I_{(2,2,0)} + f_{(1,0)}I_{(1,0)} \\
& + f_{(1,0,1)}I_{(1,0,1)} + f_{(1,2,0)}I_{(1,2,0)} + f_{(2,1,0)}I_{(2,1,0)} + f_{(1,1,0)}I_{(1,1,0)} \\
& + f_{(0,0)}I_{(0,0)} + f_{(0,0,1)}I_{(0,0,1)} + f_{(2,0,0)}I_{(2,0,0)} + f_{(0,2,0)}I_{(0,2,0)} \\
& + f_{(1,0,0)}I_{(1,0,0)} + f_{(0,1,0)}I_{(0,1,0)} + f_{(0,0,0)}I_{(0,0,0)},
\end{aligned}
$$

where $f_{(0)} = \beta,\ f_{(1)} = \xi,\ f_{(1,1)} = \xi\xi',\ f_{(1,1,1)} = \xi(\xi'^2 + \xi\xi''),\ f_{(0,1)} = \beta\xi' + 0.5\xi^2\xi''$, $f_{(0,1,1)} = \beta(\xi'^2 + \xi\xi'') + 0.5\xi^2(3\xi'\xi'' + \xi\xi^{(3)})$, $f_{(2,0)} = \phi\beta^{(0,1)}$, $f_{(2,0,1)} = \phi\xi'\beta^{(0,1)}$, $f_{(2,2,0)} = \phi(\phi'\beta^{(0,1)} + \phi\beta^{(0,2)})$, $f_{(1,0)} = \xi\beta^{(1,0)}$, $f_{(1,0,1)} = \xi(\beta\xi'' + \xi\xi'\xi'' + 0.5\xi^2\xi^{(3)} + \xi'\beta^{(1,0)})$, $f_{(1,2,0)} = \xi\phi\beta^{(1,1)}$, $f_{(2,1,0)} = \xi\phi\beta^{(1,1)}$, $f_{(1,1,0)} = \xi(\xi'\beta^{(1,0)} + \xi\beta^{(2,0)})$, $f_{(0,0)} = \zeta\beta^{(0,1)} + 0.5\phi^2\beta^{(0.2)} + \beta\beta^{(1,0)} + 0.5\xi^2\beta^{(2,0)}$, $f_{(0,0,1)} = \beta(\beta\xi'' + \xi\xi'\xi'' + 0.5\xi^2\xi^{(3)} + \xi'\beta^{(1,0)}) + 0.5\xi^2(\xi'^2\xi'' + \xi\xi''^2 + \beta\xi^{(3)} + 2\xi\xi'\xi^{(3)} + 0.5\xi^2\xi^{(4)} + 2\xi''\beta^{(1,0)} + \xi'\beta^{(2,0)}) + \zeta\xi'\beta^{(0,1)} + 0.5\phi^2\xi'\beta^{(0,1)}$, $f_{(2,0,0)} = \phi(\zeta\beta^{(0,2)} + \beta^{(0,1)}\zeta^{(0,1)} + \phi\phi'\beta^{(0,2)} + 0.5\phi^2\beta^{(0,3)} + \beta^{(0,1)}\beta^{(1,0)} + \beta\beta^{(1,1)} + 0.5\xi^2\beta^{(2,1)})$, $f_{(0,2,0)} = \zeta(\phi'\beta^{(0,1)} + \phi\beta^{(0,2)}) + 0.5\phi^2(\phi''\beta^{(0,1)} + 2\phi'\beta^{(0,2)} + \phi\beta^{(0,3)}) + \beta\phi\beta^{(1,1)} + 0.5\xi^2\phi\beta^{(2,1)}$, $f_{(1,0,0)} = \xi(\beta^{(1,0)2} + \beta^{(0,1)}\zeta^{(1,0)} + \zeta\beta^{(1,1)} + 0.5\phi^2\beta^{(1,2)} + \beta\beta^{(2,0)} + \xi\xi'\beta^{(2,0)} + 0.5\xi^2\beta^{(3,0)})$, $f_{(0,1,0)} = \zeta\xi\beta^{(1,1)} + 0.5\xi\phi^2\beta^{(1,2)} + \beta(\xi'\beta^{(1,0)} + \xi\beta^{(2,0)}) + 0.5\xi^2(\xi''\beta^{(1,0)} + 2\xi'\beta^{(2.0)} + \xi\beta^{(3,0)})$, $f_{(0,0,0)} = \zeta(\beta^{(0,1)}\zeta^{(0,1)} + \zeta\beta^{(0,2)} + \phi\phi'\beta^{(0,2)} + 0.5\phi^2\beta^{(0,3)} + \beta^{(0,1)}\beta^{(1,0)} + \beta\beta^{(1,1)} + 0.5\xi^2\beta^{(2,1)}) + 0.5\phi^2(\phi'^2\beta^{(0,2)} + \phi\phi''\beta^{(0,2)} + 2\zeta^{(0,1)}\beta^{(0,2)} + \beta^{(0,1)}\zeta^{(0,2)} + \zeta\beta^{(0,3)} + 2\phi\phi'\beta^{(0,3)} + 0.5\phi^2\beta^{(0,4)} + \beta^{(0,2)}\beta^{(1,0)} + 2\beta^{(0,1)}\beta^{(1,1)} + \beta\beta^{(1,2)} + 0.5\xi^2\beta^{(2,2)}) + \beta(\beta^{(1,0)2} + \beta^{(0,1)}\zeta^{(1,0)} + \zeta\beta^{(1,1)} + 0.5\phi^2\beta^{(1,2)} + \beta\beta^{(2,0)} + \xi\xi'\beta^{(2,0)} + 0.5\xi^2\beta^{(3,0)}) + 0.5\xi^2(2\zeta^{(1,0)}\beta^{(1,1)} +

$\xi'^2\beta^{(2,0)}+\xi\xi''\beta^{(2,0)}+3\beta^{(1,0)}\beta^{(2,0)}+\beta^{(0,1)}\zeta^{(2,0)}+\zeta\beta^{(2,1)}+0.5\phi^2\beta^{(2,2)}+\beta\beta^{(3,0)}+2\xi\xi'\beta^{(3,0)}+0.5\xi^2\beta^{(4,0)})$, and

$$\begin{aligned}Y_{t_q}^{2,\Delta} \;=\; & X_{t_{q-1}}^2+f_{(0)}I_{(0)}+f_{(2)}I_{(2)}+f_{(1,2)}I_{(1,2)}+f_{(1,1,2)}I_{(1,1,2)}+f_{(0,2)}I_{(0,2)}\\ & +f_{(0,1,2)}I_{(0,1,2)}+f_{(2,0,2)}I_{(2,0,2)}+f_{(2,0)}I_{(2,0)}+f_{(2,2,0)}I_{(2,2,0)}\\ & +f_{(1,0,2)}I_{(1,0,2)}+f_{(1,0)}I_{(1,0)}+f_{(2,1,0)}I_{(2,1,0)}+f_{(1,2,0)}I_{(1,2,0)}+f_{0,0,2)}I_{(0,0,2)}\\ & +f_{(1,1,0)}I_{(1,1,0)}+f_{(0,0)}I_{(0,0)}+f_{(2,0,0)}I_{(2,0,0)}+f_{(0,2,0)}I_{(0,2,0)}\\ & +f_{(1,0,0)}I_{(1,0,0)}+f_{(0,1,0)}I_{(0,1,0)}+f_{(0,0,0)}I_{(0,0,0)},\end{aligned}$$

where $f_{(0)}=\zeta$, $f_{(2)}=\xi$, $f_{(1,2)}=\xi\xi'$, $f_{(1,1,2)}=\xi(\xi'^2+\xi\xi'')$, $f_{(0,2)}=\beta\xi'+0.5\xi^2\xi''$, $f_{(0,1,2)}=\beta(\xi'^2+\xi\xi'')+0.5\xi^2(3\xi'\xi''+\xi\xi^{(3)})$, $f_{(2,0,2)}=\xi\xi'\beta^{(0,1)}$, $f_{(2,0)}=\xi\zeta^{(0,1)}$, $f_{(2,2,0)}=\xi^2\zeta^{(0,2)}$, $f_{(1,0,2)}=\xi(\beta\xi''+\xi\xi'\xi''+0.5\xi^2\xi^{(3)}+\xi'\beta^{(1,0)})$, $f_{(1,0)}=\xi\zeta^{(1,0)}$, $f_{(2,1,0)}=\xi^2\zeta^{(1,1)}$, $f_{(1,2,0)}=\xi(\xi'\zeta^{(0,1)}+\xi\zeta^{(1,1)})$, $I_{(0,0,2)}=\zeta\xi'\beta^{(0,1)}+0.5\xi^2\xi'\beta^{(0,2)}+\beta(\beta\xi''+\xi\xi'\xi''+0.5\xi^2\xi^{(3)}+\xi'\beta^{(1,0)})+0.5\xi^2(\xi'^2\xi''+\xi\xi''^2+\beta\xi^{(3)}+2\xi\xi'\xi^{(3)}+0.5\xi^2\xi^{(4)}+2\xi''\beta^{(1,0)}+\xi'\beta^{(2,0)})$, $f_{(1,1,0)}=\xi(\xi'\zeta^{(1,0)}+\xi\zeta^{(2,0)})$, $f_{(0,0)}=\zeta\zeta^{(0,1)}+0.5\xi^2\zeta^{(0.2)}+\beta\zeta^{(1,0)}+0.5\xi^2\zeta^{(2,0)}$, $f_{(2,0,0)}=\xi(\zeta^{(0,1)2}+\zeta\zeta^{(0,2)}+0.5\xi^2\zeta^{(0,3)}+\beta^{(0,1)}\zeta^{(1,0)}+\beta\zeta^{(1,1)}+0.5\xi^2\zeta^{(2,1)})$, $f_{(0,2,0)}=\zeta\xi\zeta^{(0.2)}+0.5\xi^3\zeta^{(0,3)}+\beta(\xi'\zeta^{(0,1)}+\xi\zeta^{(1,1)})+0.5\xi^2(\xi''\zeta^{(0,1)}+2\xi'\zeta^{(1,1)}+\xi\zeta^{(2,1)})$, $f_{(1,0,0)}=\xi(\xi\xi'\zeta^{(0,2)}+\zeta^{(0,1)}\zeta^{(1,0)}+\beta^{(1,0)}\zeta^{(1,0)}+\zeta\zeta^{(1,1)}+0.5\xi^2\zeta^{(1,2)}+\beta\zeta^{(2,0)}+\xi\xi'\zeta^{(2,0)}+0.5\xi^2\zeta^{(3,0)})$, $f_{(0,1,0)}=\zeta\xi\zeta^{(1,1)}+0.5\xi^3\zeta^{(1,2)}+\beta(\xi'\zeta^{(1,0)}+\xi\zeta^{(2,0)})+0.5\xi^2(\xi''\zeta^{(1,0)}+2\xi'\zeta^{(2,0)}+\xi\zeta^{(3,0)})$, $f_{(0,0,0)}=\zeta(\zeta^{(0,1)2}+\zeta\zeta^{(0.2)}+0.5\xi^2\zeta^{(0,3)}+\beta^{(0.1)}\zeta^{(1,0)}+\beta\zeta^{(1,1)}+0.5\xi^2\zeta^{(2,1)})+0.5\xi^2(3\zeta^{(0,1)}\zeta^{(0,2)}+\zeta\zeta^{(0.3)}+0.5\xi^2\zeta^{(0,4)}+\beta^{(0,2)}\zeta^{(1,0)}+2\beta^{(0,1)}\zeta^{(1,1)}+\beta\zeta^{(1,2)}+0.5\xi^2\zeta^{(2,2)})+\beta(\xi\xi'\zeta^{(0,2)}+\zeta^{(0,1)}\zeta^{(1,0)}+\beta^{(1,0)}\zeta^{(1,0)}+\zeta\zeta^{(1,1)}+0.5\xi^2\zeta^{(1,2)}+\beta\zeta^{(2,0)}+\xi\xi'\zeta^{(2,0)}+0.5\xi^2\zeta^{(3,0)})+0.5\xi^2(\xi'^2\zeta^{(0,2)}+\xi\xi''\zeta^{(0,2)}+2\zeta^{(1,0)}\zeta^{(1,1)}+2\xi\xi'\zeta^{(1,2)}+\zeta^{(1,0)}\beta^{(2,0)}+\xi'^2\zeta^{(2,0)}+\xi\xi''\zeta^{(2,0)}+\zeta^{(0,1)}\zeta^{(2,0)}+2\beta^{(1,0)}\xi^{(2,0)}+\zeta\zeta^{(2,1)}+0.5\xi^2\zeta^{(2,2)}+\beta\zeta^{(3,0)}+2\xi\xi'\zeta^{(3,0)}+0.5\xi^2\zeta^{(4,0)})$. Expressions for $l(\alpha)=4$ are available from the author upon request.

References

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,* Vol. 3, 261–327, New York, NY, USA: Cambridge University Press.

Aït-Sahalia, Y. (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906-937.

Aït-Sahalia, Y. and R. Kimmel (2007): "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal Financial Economics*, 83, 413-452.

Asai, M., M. McAleer and J. Yu (2006): "Multivariate Stochastic Volatility: A Review," *Econometric Reviews*, 25, 145-175.

Beskos, A., O. Papaspillopoulos and G. Roberts (2009): "Monte-Carlo Maximum Likelihood Estimation for Discretely Observed diffusion Processes," *Annals of Statistics*, 37, 223-245.

Bianchi, A. (2007): "Nonparametric Trend Coefficient Estimation for Multidimensional Diffusions," *Comptes Rendus Mathematique*, 345, 101-105.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

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.

Britten-Jones, M. and A. Neuberger (2000): "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839-866.

Bruti-Liberati, N. and E. Platen (2007): "Strong Approximations of Stochastic Differential Equations with Jumps," *Journal of Computational and Applied Mathematics*, 205, 982-1001.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

Genon-Catalot, V. and J. Jacod (1993): "On the Estimation of the Diffusion Coefficient for Multi-dimensional Diffusion Process," *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29, 119-151.

Ghysels, E., A. C. Harvey and E. Renault (1996): "Stochastic Volatility," *Handbook of Statistics*, Vol 14. Maddala, G. S., C. R. Rao (eds). Elsevier Science B. V.

Golightly, A. and D. J. Wilkinson (2008): "Bayesian Inference for Nonlinear Multivariate Diffusions Models Observed with Error," *Computational Statistics & Data Analysis*, 52, 1674-1693.

Heston, S. (1993): "A Closed-form Solution for Option with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-343.

Huang, X. (2010), "Quasi-maximum Likelihood Estimation of Discretely Observed Diffusions," *Econometrics Journal*, 14, 241-256.

Hurn, A. S., J. I. Jeisman and K. A. Lindsay (2007): "Seeing the Wood for the Trees: A Critical Evaluation of the Methods to Estimate the Parameters of Stochastic Differential Equations," *Journal of Financial Econometrics*, 5, 390-455.

Jiang, G. J. and Y. S. Tian (2005): "The Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies*, 18, 1305-1342.

Jones, C. S. (2003), "The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets," *Journal of Econometrics* 116, 181-224.

Kalogeropoulos, K. (2007): "Likelihood-based Inference for a Class of Multivariate Diffusions with Unobserved Paths," *Journal of Statistical Planning and Inference*, 137, 3092-3102.

Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*. New York: Springer.

Kelly, L., E. Platen and M. Sørensen (2004): "Estimation for Discretely Observed Diffusions Using Transform Functions," *Journal of Applied Probability*, 41A, 99-118.

Kloeden, P. and E. Platen (1999), *Numerical Solution of Stochastic Differential Equations*, New York, NY, USA: Springer, corrected 3rd printing.

Nelson, D. B. (1990): "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7-38.

Phillips, P. C. B. (1973): "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351-362.

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, 139-150.

Platen, E. and D. Heath (2006): *Benchmark Approach to Quantitative Finance*, NY, USA: Springer.

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

Stroock, D. W. and S. R. S. Varadhan (1979): *Multidimensional Diffusion Process*, New York, NY, USA: Springer.

White, H. (1994): *Estimation, Inference and Specification Analysis*, New York, NY, USA: Cambridge University Press.

Wooldridge, J. M. (1994): Estimation and Inference for Dependent Processes. In *Handbook of Econometrics* Vol. IV, Engle, R. F. and D. L. MacFadden (eds.), San Diego, CA, USA: Elsevier Science B.V.

Yu, J. (2011), "Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models," *Journal of Econometrics*, 161, 228-245.