Density Approximations for Multivariate Diffusions via an
Itô-Taylor Expansion Approach

Nian Yang*    Nan Chen†    Xiangwei Wan‡

October 14, 2017

Abstract

In this paper we develop a new Itô-Taylor expansion approach to deriving analytical approximations to the transition densities of multivariate diffusions. The obtained approximations can thereby be used to carry out the maximum likelihood estimation for the diffusions with discretely sampled data. Different from the existing density expansion methodologies that are usually model-specific, this new approach is universally applicable for a wide spectrum of models, particularly time-inhomogeneous non-affine irreducible multivariate diffusions. We manage to show that the expansion converges to the true probability density under some regularity conditions. The method enjoys a significant computational efficiency over the other alternatives in the literature because only differentiation operations are required. Extensive numerical experiments on a variety of models demonstrate the accuracy and efficiency of the estimators based on the approximate densities yielded from our approach.

*Department of Finance and Insurance, School of Business, Nanjing University, Nanjing, China. Email: yangnian@nju.edu.cn
†Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. Email: nchen@se.cuhk.edu.hk
‡Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China. Email: xwwan@sjtu.edu.cn
1 Introduction

The transition probability densities constitute the essential inputs when we intend to apply the maximum likelihood estimation (MLE) method to diffusion models for the purpose of parameter estimation on the basis of discretely observed data (see, e.g., Lo, 1988). However, except a handful of simple examples, closed-form expressions of the transition density of a general multivariate diffusion process is not available. To overcome this technical difficulty, an active research line, initiated by the seminal contribution of Aït-Sahalia (2002; 2008), aims to constructing tractable approximations to the transition density.

In the same spirit of this literature, we propose a new Itô-Taylor expansion based approach in the current paper to deriving density approximations for multivariate time-inhomogeneous diffusion processes, from which a new type of approximate maximum likelihood estimators can be obtained. The Itô-Taylor expansion, a stochastic analogy to the Taylor expansion in the sense of classical calculus, originates from iterated applications of the Itô-Dynkin formula on a “smooth function” of diffusion processes (see, e.g., p.133 of Dynkin, 1965; Kallenberg, 2002). Note that we can represent the process’ transition density as a conditional expectation of the Dirac delta function of the diffusion. Approximating the delta function by a sequence of smooth functions enables us to apply the Itô-Taylor expansion to the expectation to yield density approximations.

The proposed new approach contributes to the literature in three aspects. First, it provides a unified treatment on various types of diffusions, nesting univariate and multivariate, reducible and irreducible, time-homogeneous and time-inhomogeneous,
affine and non-affine diffusions. This is in a sharp contrast to the existing literature that typically develop model-specific methods to cope with different types of diffusions. Second, we manage to show that the obtained approximations will converge to the true probability densities of the underlying diffusion processes under some regularity conditions as the observational time interval shrinks. The numerical experiments in the paper further demonstrate the accuracy of the maximum likelihood estimators based on our density approximations for a wide range of diffusion models. Third, the implementation of our approach is computationally cheap because only differentiation operations are involved. The procedure is so mechanical that we can easily utilize some symbolic computation programs to facilitate the calculation. In this way, we avoid solving the Kolmogorov partial differential equations (PDE) for approximating transition densities as required by the conventional approaches (see, e.g., Aït-Sahalia, 2008 and Choi, 2013).

As noted previously, diffusion density approximation and the related MLE have been well investigated in the literature. A fruitful breakthrough was made by Aït-Sahalia (2002), who managed to use a sequence of mutually orthogonal Hermite polynomials to obtain accurate approximations to the transition density of a univariate diffusion process. This work triggers a substantial body of discussion on how to search for diffusion density approximations for the purpose of implementing the MLE method. Just to name a few, Schaumburg (2001) incorporated a jump component in the process dynamic; Egorov et al. (2003) extended the result to univariate time-inhomogeneous diffusions; Bakshi and Ju (2005) provided some refinements to Aït-Sahalia (2002); Lee et al. (2014) obtained a general expression of the coefficients of the Hermite expansion of Aït-Sahalia (2002) and rearranged it to get a new approximation.

One crucial step in the method proposed in Aït-Sahalia (2002) is to apply the Lamperti transform to unitize the process volatility. However, not every multivariate process is amenable to such a transformation. This obstacle prevents us from extend-
ing the method to irreducible diffusion processes. To tackle it, Aït-Sahalia (2008) suggested a new expansion of the probability densities of multivariate diffusions on the basis of a double series in both the time and state variables, where the coefficients are determined recursively through solving the Kolmogorov PDEs. Yu (2007) and Choi (2013; 2015) investigated the extensions of Aït-Sahalia’s method in the cases of multivariate jump diffusions and multivariate time-inhomogeneous diffusions, respectively.

Beyond the aforementioned research line initiated by the seminal work of Aït-Sahalia (2002; 2008), alternative attempts have also been made to obtain closed-form approximations of transition densities in multivariate models, including the saddle-point approximation proposed by Aït-Sahalia and Yu (2006), the orthonormal polynomials expansion of Filipović et al. (2013) for affine jump-diffusion models, and the Malliavin calculus theoretic approach in Li (2013) and Li and Chen (2016) for time-homogeneous models.

From this brief literature review, we can see that many of the existing approximate MLEs for multivariate diffusions are model-specific, exploiting the special structures of one certain class of processes. Taking an innovative path relative to these literatures, our approach aims to establish a versatile expansion that approximate accurately to the transition probability densities for a general family of diffusion models.

Our research is also related to a rich body of literature about the Itô-Taylor expansion; see Milstein (1975), Kessler (1997), Stanton (1997), Fan and Zhang (2003), Aït-Sahalia and Mykland (2003), Kristensen and Mele (2011), Uchida and Yoshida (2012), Xiu (2014), and Li and Li (2015) for a variety of applications of this expansion in moment computing and option pricing. It is worthwhile to mention that all of the above papers just directly apply the Itô-Taylor expansion to expand the conditional expectation of smooth functions. But what we encounter in this paper is challenging: the irregularity of the Dirac delta function may cause divergence of our approximations. As a major theoretical contribution, we are the first ones to propose the idea
of approximating the true transition density by a special Itô-Taylor expansion and manage to prove its convergence.

The rest of the paper is organized as follows. In Section 2, we develop our Itô-Taylor expansion to approximate the transition density of a general multivariate diffusion process, and the convergence of the expansion to the true transition density is proved. Section 3 presents some convergence results about the resulting approximate maximum likelihood estimators. Section 4 contains numerical evidence of the performance of the approximate transition densities and the approximate maximum likelihood estimators under various diffusions. Technical assumptions, lemmas, and proofs are collected in the Appendix.

For the convenience of reference, we would like to define here some notations that will be used throughout the paper. Let $D \subset \mathbb{R}^m$ be the domain of state variables and denote $D^c$ as a compact subset of $D$. Let $\| \cdot \|$ be the Euclidean norm. Denote $\top$ to be the transpose operation on matrices or vectors. Let $h = (h_1, h_2, \cdots, h_m)$ be an index vector with nonnegative integer components and $|h| := \sum_{i=1}^{m} h_i$. Let $e_i$ be a special index vector, in which the $i$-th component is 1, and the others are 0. Define $\partial_t := \partial / \partial t$ to be the partial derivative with respect to the time variable, and $\partial_h := \partial^{\mid h\mid} / (\partial x_1^{h_1} \cdots \partial x_m^{h_m})$ to be the partial derivatives with respect to the state variable $x := (x_1, x_2, \cdots, x_m)^\top \in D$. For example, $\partial_{e_i} = \partial / \partial x_i$ and $\partial_{e_i+e_j} = \partial^2 / (\partial x_i \partial x_j)$.

### 2 The Itô-Taylor Expansion of the Transition Density of Multivariate Diffusions

#### 2.1 The Model

Consider a multivariate diffusion process

$$dX(t) = \mu(t, X(t); \theta)dt + \sigma(t, X(t); \theta)dW(t),$$

(1)
where $X(t)$ is an $m \times 1$ vector of state variables in the domain $D \subset \mathbb{R}^m$, $\{W(t), t \geq 0\}$ is a $d$-dimensional standard Brownian motion, $\mu(t, X(t); \theta)$ and $\sigma(t, X(t); \theta)$ are an $m \times 1$ drift vector and an $m \times d$ volatility (or dispersion) matrix, respectively. The explicit forms of both are known. We emphasize that each component of the drift vector and volatility matrix is a function dependent on the time variable $t$. Therefore the models considered in this paper are allowed to be time-inhomogeneous. The unknown parameter $\theta$ belongs to a compact set $\Theta \subset \mathbb{R}^K$. Furthermore, we assume that all the four assumptions in Appendix A.1 hold in the remaining part of the paper.

We intend to develop a maximum likelihood estimator for $\theta$ in this paper. Introduce more notations to make the subsequent discussion more precise. Fix two time points $t' > t$. Let $p(t', x'|t, x; \theta)$ be the conditional transition density function of the process driven by the stochastic differential equation (SDE) in (1); that is,

$$
P[X(t') \in dx'|X(t) = x] = p(t', x'|t, x; \theta)dx'.
$$

Assume that we have a sequence of observations on the state variables at a discrete time grid $\{t = t_i : i = 0, 1, \ldots, n\}$. By the Markovian setting of (1), the log-likelihood function\(^1\) is given by

$$\ell_n(\theta) := \sum_{i=1}^{n} \ln p(t_i, X(t_i)|t_{i-1}, X(t_{i-1}); \theta). \quad (2)$$

Then, the maximizer

$$\hat{\theta}_n := \arg \max_{\theta \in \Theta} \ell_n(\theta)$$

is the MLE of parameter $\theta$.

However, it is well known that a major technical difficulty with the MLE method for diffusions resides in the fact that closed-form expressions of $p(t', x'|t, x; \theta)$ is unavailable for most cases. As noted in the introduction, we shall first proposes an

\(^1\)As in A"it-Sahalia (2002; 2008), we ignore the unconditional density of the first observation $(t_0, X(t_0))$.\)
approximation to the function \( p \) via an Itô-Taylor expansion approach and then apply it to derive the MLE of \( \theta \).

### 2.2 Heuristic Idea of the Itô-Taylor Expansion

To facilitate the exposition, we would like to present the heuristic idea behind our Itô-Taylor expansion in this subsection. Consider any sufficiently smooth function \( G(s, y) \). By the Itô formula, we have

\[
G(s, X(s)) = G(t, X(t)) + \int_t^s \partial_u G(u, X(u))du + \sum_{i=1}^m \int_t^s \mu_i(u, X(u); \theta) \partial_{e_i} G(u, X(u))du \\
+ \frac{1}{2} \sum_{i,j=1}^m \int_t^s \nu_{ij}(u, X(u); \theta) \partial_{e_i+e_j} G(u, X(u))du \\
+ \sum_{i,j=1}^m \int_t^s \partial_{e_i} G(u, X(u)) \sigma_{ij}(u, X(u); \theta) dW_j(u).
\]

Here \( \nu_{ij} \) is the \((i, j)\)-element of the variance-covariance (or diffusion) matrix of the diffusion \( X \) defined below:

\[
\nu(t, x_0; \theta) := \sigma(t, x_0; \theta) \sigma(t, x_0; \theta)^\top.
\]

Let \( \mathbb{E}^{t,x}[\cdot] \) denote expectation conditional on \( X(t) = x \). Taking expectations on both sides of the above equality, we have

\[
\mathbb{E}^{t,x}[G(s, X(s))] = G(t, x) + \int_t^s \mathbb{E}^{t,x}[(\partial_u + \mathcal{L})G(u, X(u))]du, \tag{3}
\]

where \( \mathcal{L} \) is the infinitesimal generator of process \( (1) \) such that

\[
(\mathcal{L}G)(u, y) = \sum_{i=1}^m \mu_i(u, y; \theta) \partial_{e_i} G(u, y) + \frac{1}{2} \sum_{i,j=1}^m \nu_{ij}(u, y; \theta) \partial_{e_i+e_j} G(u, y), \tag{4}
\]

for any \( u \in (t, s) \) and \( y \in \mathbb{R}^m \). We may continue to apply the above idea to expand \( \mathbb{E}^{t,x}[(\partial_u + \mathcal{L})G(u, X(u))] \), treating \( (\partial_u + \mathcal{L})G(u, X(u)) \) as a new function on the process \( X \). This will lead to

\[
\mathbb{E}^{t,x}[(\partial_u + \mathcal{L})G(u, X(u))] = (\partial_t + \mathcal{L})G(t, x) + \int_t^u \mathbb{E}^{t,x}[(\partial_{u_2} + \mathcal{L})^2 G(u_2, X(u_2))]du_2.
\]

(5)
Substituting (5) back into (3), we have
\[
E^{t,x}[G(s, X(s))] = G(t, x) + (\partial_t + \mathcal{L})G(t, x) \cdot (s - t)
+ E^{t,x} \left[ \int_t^s du_1 \int_t^{u_1} (\partial_{u_2} + \mathcal{L})^2 G(u_2, X(u_2)) du_2 \right].
\]
In this way, repeatedly applying the expansion for \(J\) times yields
\[
E^{t,x}[G(s, X(s))] = \sum_{N=0}^J \frac{(s - t)^N}{N!} (\partial_t + \mathcal{L})^N G(t, x) + R_J,
\]
where the remainder term \(R_J\) is given by
\[
R_J = E^{t,x} \left[ \int_t^s du_1 \int_t^{u_1} \cdots \int_t^{u_J} (\partial_{u_{J+1}} + \mathcal{L})^{J+1} G(u_{J+1}, X(u_{J+1})) du_{J+1} \right].
\]

Now, we turn to apply the Itô-Taylor expansion (6) to approximate the transition density \(p\). Intuitively, the density function admits the following expression:
\[
p(t', x'|t, x; \theta) = E^{t,x}[\delta_{x'}(X(t'))],
\]
where \(\delta_{x'}(\cdot)\) is the Dirac delta function centered at \(x'\). The function \(\delta_{x'}(\cdot)\) does not have any derivatives in the classical sense. To circumvent this obstacle to applying the above expansion, we need to “mollify” the irregularity of \(\delta_{x'}(\cdot)\) by introducing a sequence of smooth functions to approximate it. More precisely, given \(\mu_0 \in \mathbb{R}^m\) and \(\nu_0 \in \mathbb{R}^{m \times m}\), define a function \(G\) such that, for \(t \leq s < t'\) and \(y \in \mathbb{R}^m\),
\[
G_{t',x'}(s, y) := \frac{1}{(2\pi(t' - s))^{m/2} \det(\nu_0)^{1/2}} \exp \left( -\frac{(x' - y - (t' - s)\mu_0)^\top \nu_0^{-1}(x' - y - (t' - s)\mu_0)}{2(t' - s)} \right).
\]
In other words, if we fix \(t', s\) and \(x'\), \(G_{t',x'}(s, \cdot)\) is the probability density function of a multivariate normal distribution with mean \(x' - (t' - s)\mu_0\) and covariance matrix \((t' - s)\nu_0\).

It is apparent to see that, as \(s \to t'\), the function \(G_{t',x'}\) converges to the Dirac delta function \(\delta_{x'}\) in the following sense:
\[
\lim_{s \uparrow t'} E^{t,x}[G_{t',x'}(s, X(s))] = E^{t,x}[\delta_{x'}(X(t'))] = p(t', x'|t, x; \theta).
\]
Since the function $G_{t',x'}$ is infinitely differentiable, we now can invoke the Itô-Taylor expansion (6) to expand the conditional expectation $E^{t,x}[G_{t',x'}(s, X(s))]$ on the left-hand side of (9). That is,

$$E^{t,x}[G_{t',x'}(s, X(s))] \approx \sum_{N=0}^{J} \frac{(s - t)^N}{N!} (\partial_t + \mathcal{L})^N G_{t',x'}(t, x),$$

if we skip the residual term $\mathcal{R}_J$. From (9), letting $s$ tend to $t'$ in the above expression results in our density approximation

$$p(t', x'|t, x; \theta) \approx \sum_{N=0}^{J} \frac{(t' - t)^N}{N!} (\partial_t + \mathcal{L})^N G_{t',x'}(t, x) =: p^{(J)}(t', x'|t, x; \theta) \quad (10)$$

for any positive integer $J$. From now on, we will refer to $p^{(J)}$ as the $J$-th order Itô-Taylor expansion of the density $p$.

### 2.3 The Itô-Taylor Expansion of the Transition Density

Noting that the derivation in Section 2.2 is heuristic, in this section we shall establish the convergence property of our Itô-Taylor density approximation rigorously. It turns out that not all the choices of $\mu_0$ and $\nu_0$ will warrant a convergent expansion. Therefore we just focus on one of such choices that $\mu_0 = 0$ and $\nu_0 = \nu(t, x; \theta)$ (the diffusion matrix of SDE (1) at time $t$ when $X_t = x$).

From (10), it is easy to see that we construct the approximation $p^{(J)}$ via repeatedly applying the differential operator $\mathcal{L}$ on a known function $G_{t',x'}$. The analytical form of $p^{(J)}$ is thus achievable. Define some more notations to facilitate the result presentation. Let $\phi(y; \Sigma)$ be the density of the $m$-dimensional normal distribution with mean 0 and covariance matrix $\Sigma$, i.e.,

$$\phi(y; \Sigma) = \frac{1}{(2\pi)^{m/2} \det(\Sigma)^{1/2}} \exp \left( -\frac{y^T \Sigma^{-1} y}{2} \right). \tag{11}$$

For an $m$-dimensional nonnegative integer vector $h = (h_1, \ldots, h_m)$, define $H_h(y; \nu_0)$ to be the corresponding multivariate Hermite polynomial associated with this normal
density; that is,
\[ H_h(y; \Sigma) := (-1)^{|h|} \phi^{-1}(y; \Sigma) \partial_h \phi(y; \Sigma). \]  
\hspace{1cm} (12)

Willink (2005) provides a recursive approach to computing these multivariate Hermite polynomials. Namely, for any vector \( h \) with \( h_j \geq 0, \ j = 1, \cdots, m \), we have
\[ H_{h+e_k}(y; \Sigma) = \left( \sum_{j=1}^{m} \Sigma^{(kj)} y_j \right) H_h(y; \Sigma) - \sum_{j=1}^{m} \Sigma^{(kj)} h_j H_{h-e_j}(y; \Sigma), \]  
\hspace{1cm} (13)

where \( \Sigma^{(kj)} \) is the \((k, j)\)-element of the matrix \( \Sigma^{-1} \). In addition, \( H_0(y; \Sigma) = 1 \) and \( H_h(y; \Sigma) = 0 \) for any \( h \) with \( \min\{h_1, \cdots, h_m\} < 0 \). With the help of the Hermite polynomials, we can show

**Theorem 1.** Set \( \mu_0 = 0 \) and \( \nu_0 = \nu(t, x; \theta) \). For \( J \geq 1 \), there exists a sequence of functions \( \{w_{N,h}\} \) such that
\[ p^{(J)}(t', x'|t, x; \theta) = q(t', x'|t, x) \left( 1 + \sum_{N=1}^{J} \sum_{1 \leq |h| \leq 3N/2} \frac{w_{N,h}(t, x)H_h(z; \nu_0)}{N!} \Delta^{N-|h|/2} \right), \]  
\hspace{1cm} (14)

where \( H_h(z; \nu_0) \) are the Hermite polynomials defined through (13) with

\[ \Delta = t' - t, \quad z = \frac{x' - x}{\sqrt{\Delta}}, \]

and
\[ q(t', x'|t, x) := \frac{1}{(2\pi \Delta)^{m/2} \det(\nu_0)^{1/2}} \exp \left( -\frac{(x' - x)^\top \nu_0^{-1}(x' - x)}{2\Delta} \right). \]  
\hspace{1cm} (15)

Theorem 1 characterizes the structure of \( p^{(J)} \) clearly: it can be expressed in terms of a linear combination of Hermite polynomials. We can analytically determine the coefficient functions \( w \) in a recursive fashion. More specifically, fixed \((t, x)\), introduce a function sequence \( \{w_{N,h}(s, y) := w_{N,h}(s, y; t, x, \theta) : s \geq 0, y \in D\} \) indexed by a positive integer \( N \) and an \( m \)-dimensional vector \( h = (h_1, \cdots, h_m) \) with \( h_j \geq -2 \) for
all \( j = 1, \cdots, m \). For any \( N \geq 1 \), define \( w_{N,h}(s,y) \equiv 0 \) if either \( \min\{h_1, \cdots, h_m\} < 0 \),
either \( h = 0 \), or \(|h| > 2N\). When \( N = 1 \),
\[
\begin{aligned}
w_{1,e_i}(s,y) &= \mu_i(s,y;\theta), \quad i = 1, \cdots, m; \\
w_{1,2e_i}(s,y) &= \frac{1}{2} (\nu_{ii}(s,y;\theta) - \nu_{ii}(t,x;\theta)), \quad i = 1, \cdots, m; \\
w_{1,e_i+e_j}(s,y) &= \nu_{ij}(s,y;\theta) - \nu_{ij}(t,x;\theta), \quad i \neq j, \quad i, j = 1, \cdots, m,
\end{aligned}
\]
where \( \mu_i(\cdot,\cdot;\theta) \) is the \( i \)-component of the drift vector \( \mu \) and \( \nu_{ij}(\cdot,\cdot;\theta) \) is the \((i,j)\)-
element of the diffusion matrix \( \nu \). When \( N > 1 \) and all the components in \( h \) are
nonnegative, \( 0 < |h| \leq 2N \), define recursively
\[
w_{N,h}(s,y) = (\partial_s + \mathcal{L}) w_{N-1,h}(s,y) \\
+ \sum_{i,j=1}^{m} \nu_{ij}(s,y;\theta) \partial_{e_j} w_{N-1,h-e_i}(s,y) + \sum_{i=1}^{m} \mu_i(s,y;\theta) w_{N-1,h-e_i}(s,y) \\
+ \frac{1}{2} \sum_{i,j=1}^{m} (\nu_{ij}(s,y;\theta) - \nu_{ij}(t,x;\theta)) w_{N-1,h-e_i-e_j}(s,y),
\]
where the infinitesimal generator \( \mathcal{L} \) is given in (4) acting on the state variable \( y \).

Note that only differentiation operations are involved in the computation of coefficient functions \( w \) in (17). Unlike the other existing density expansion methods, it avoids the complicated integrations in the step of solving Kolmogorov PDEs. Therefore, the implementation of our expansion is computationally convenient. We can even simply use some symbolic computation programs such as Mathematica to accomplish the derivation.

The next theorem establishes one of the main outcomes in this paper. Under some regularity conditions (cf. Appendix A.1), we manage to show that the expansion \( p^{(J)} \) uniformly converges to the true transition density \( p \).

**Theorem 2.** Suppose that Assumptions A.1-A.4 hold. Define \( p^{(J)}(t', x'|t, x; \theta) \) through (14). Then, given any positive integer \( J > 2m - 1 \) and compact subset \( D^c \subset D \),
\[
\lim_{t' \to t} \sup_{(t,x,x',\theta) \in [0,T] \times D^c \times D \times \Theta} \left| p^{(J)}(t', x'|t, x; \theta) - p(t', x'|t, x; \theta) \right| = 0.
\]

This theorem provides a theoretical guarantee of the accuracy of our Itô-Taylor expansion based approximation. It shows the convergence to the true transition density $p$ holds as the observational time interval $t' - t$ shrinks to 0, which ensures that the resulted approximate maximum likelihood estimator should also converge to the unknown true maximum likelihood estimator. By the technical assumptions A.1-A.4, we can see that the structural features of a diffusion model, whether it is univariate or multivariate, reducible or irreducible, affine or non-affine, time-homogeneous or not, does not play a major role in the proof of convergences. In a sharp contrast, the previous literatures, including Aït-Sahalia (2002; 2008), Egorov et al. (2003), Choi (2013; 2015), Li (2013), and Filipović et al. (2013), suggest different methods to obtain approximations to the densities of processes of different types.

The proof of Theorem 2 is deferred to Appendix A.3. The key observation is that the coefficients $w_{N,h}(t,x)$ are all zeros for $|h| > 3N/2$, by which we build up a tight upper bound estimation on the approximation error and ensure the convergence of our density approximation. This observation is not true for any other choice of $\nu_0$ than $\nu_0 = \nu(t,x; \theta)$. One may refer to Lemmas 1 and 2, Example 1, and the proofs in Appendix A.3 for more details.

3 Approximate Maximum Likelihood Estimator

In this section, we shall use the Itô-Taylor expansion $p^{(J)}$ in (14) as an approximate to the true but unknown transition density $p$ to compute an approximate MLE. Given a set of discretely observed data \( \{X(t_0), X(t_1), \ldots, X(t_n)\} \) with \( t_i - t_{i-1} = \Delta, i = \)
1, \ldots, n, replace \( p \) in (2) with its approximation \( p^{(J)} \)

\[
\ell_n^{(J)}(\theta) := \sum_{i=1}^{n} \ln p^{(J)}(t_i, X(t_i)|t_{i-1}, X(t_{i-1}); \theta)
\]

and solve for \( \hat{\theta}_n^{(J)} := \arg \max_{\theta \in \Theta} \ell_n^{(J)}(\theta) \).

Assume that the drift vector \( \mu(t, x; \theta) \) and the diffusion matrix \( \nu(t, x; \theta) \) are infinitely continuous differentiable with respect to \( \theta \in \Theta \). Suppose the log-likelihood function \( \ell_n(\theta) \) has a unique maximizer \( \hat{\theta}_n \in \Theta \), which is the true MLE of parameter \( \theta \). With the help of the result of the uniform convergence established in Theorem 2, we can show that \( \hat{\theta}_n^{(J)} \) converges in probability to \( \hat{\theta}_n \), as \( \Delta \to 0 \).

**Theorem 3.** Fix the sample size \( n \) and \( J > 2m - 1 \). Denote the true value of the parameter vector to be \( \theta_0 \). Under Assumptions A.1-A.4, we have

\[
\hat{\theta}_n^{(J)} - \hat{\theta}_n \to 0
\]

in \( \mathbb{P}_{\theta_0} \)-probability as \( \Delta \to 0 \).

Theorem 3 constitutes a very useful step towards establishing the asymptotic consistency of our approximate MLE \( \hat{\theta}_n^{(J)} \). Specifically, we can make the speed, at which \( \hat{\theta}_n^{(J)} - \hat{\theta}_n \) converges to zero, arbitrarily small for any given sample size \( n \) by taking \( \Delta \to 0 \) sufficiently fast. In general, if we happen to know that the true MLE \( \hat{\theta}_n \) converges to \( \theta_0 \) as \( n \to \infty \), we can choose a sequence \( \Delta_n \to 0 \) such that the same asymptotic property also holds for \( \hat{\theta}_n^{(J)} \) (the approximate MLE corresponding to the observation time intervals \( \Delta_n \)).

A¨ıt-Sahalia (2002) and Chang and Chen (2011) investigated the asymptotic properties of the true and approximate MLEs for univariate diffusions. To our best knowledge, the corresponding results for multivariate diffusions are challenging and still

\( p^{(J)} \) may be negative in our expansion. To make the computation of logarithm feasible, we truncate the approximate density at a sufficiently small positive number when implementing the expansion in the numerical examples. More precisely, we take \( \ln(\cdot) \) on \( \max\{p^{(J)}, \varepsilon/J\} \) for some fixed small \( \varepsilon > 0 \). Similar procedures are also used in A¨ıt-Sahalia (2002) and Egorov et al. (2003).
open in the literature. We have provided some supportive evidences in numerical experiments. A thorough theoretic investigation is beyond the scope of the current paper. We leave it for future research.

4 Numerical Experiments

Below we undertake some numerical experiments to examine the performance of our Itô-Taylor expansion based approximate densities and the associated approximate MLE. Six different types of models are considered: the univariate Ornstein-Uhlenbeck (OU) model, the Cox-Ingersoll-Ross (CIR) model, the bivariate Ornstein-Uhlenbeck (BOU) model, the Heston model, the time-inhomogeneous bivariate Ornstein-Uhlenbeck model (BOUI), and the non-affine GARCH model. The purpose of taking so many processes is to check out how the approach will perform under a variety of modeling choices. The first two are univariate diffusions. The third and fourth models are multivariate reducible and irreducible diffusions, respectively. All the first four models are time-homogenous with time-independent affine coefficients. In contrast, the BOUI model is not time-homogenous and the last one is non affine.

The assessments on the efficiency of our proposed approach mainly consist of two parts: one is about density approximation and the other is about MLE approximation. The subsequent contents in this section are thus organized as follows. In Section 4.1, we provide more detailed modeling information about these six processes. In Section 4.2, we illustrate through numerical experiments that the approximate densities stemmed from the Itô-Taylor expansion converges to the true density in a very fast manner. In Section 4.3, we provide Monte Carlo evidence to show the accuracy and efficiency of our approximate MLE.
4.1 Models

Model 1. **Ornstein-Uhlenbeck (OU) Model.**

\[ dX(t) = \kappa(\alpha - X(t))dt + \sigma dW(t). \]

The OU process was first used by Vasicek (1977) to model the short term interest rate. Its true transition density \( p(t', x'| t, x) \) is normally distributed with mean

\[ \alpha + (x - \alpha)e^{-\kappa(t'-t)} \]

and variance

\[ \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t'-t)}). \]

Model 2. **Cox-Ingersoll-Ross (CIR) Model.**

\[ dX(t) = \kappa(\alpha - X(t))dt + \sigma \sqrt{X(t)} dW(t). \]

It can be shown that \( X(t) \) remains nonnegative almost surely for all \( t \geq 0 \). In addition, the process has a tendency of reverting to its long-run mean \( \alpha \). For these two important features, the CIR model is widely used in the literature to describe the movements of the short term interest rates (Cox et al., 1985) or equity volatilities (Heston, 1993). The true transition density of this model is given by

\[ p(t', x'| t, x) = \frac{e^{\kappa(t'-t)}}{2c(t' - t)} \left( \frac{x'e^{\kappa(t'-t)}}{x} \right)^{d/4} \exp \left( -\frac{x + x'e^{\kappa(t'-t)}}{2c(t' - t)} \right) I_{d/2-1} \left( \frac{\sqrt{xx'e^{-\kappa(t'-t)}}}{c(t' - t)} \right), \]

where

\[ c(t) = \frac{\sigma^2}{4\kappa}(e^{\kappa t} - 1), \quad d = \frac{4\kappa \theta}{\sigma^2}, \]

and

\[ I_{\gamma}(x) = \sum_{k=0}^{+\infty} \frac{(x/2)^{2k+\gamma}}{k!\Gamma(k + \gamma + 1)} \]

is the modified Bessel function of the first kind.

\[ dX(t) = \kappa (\alpha - X(t))dt + dW(t), \]

where \( X(t) = (X_1(t), X_2(t))^\top, \alpha = (\alpha_1, \alpha_2)^\top, \) and \( W(t) = (W_1(t), W_2(t))^\top, t \geq 0 \) is a 2-dimensional standard Brownian motion. \( \kappa \) is a 2 \times 2 matrix:

\[ \kappa = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}. \]

The BOU diffusion is one of the few multivariate processes with explicitly known transition densities. Assume that \( \kappa \) has full rank. The future state \( X(t') \), conditional on the current state \( X(t) = x \), is bivariate-normally distributed (Aït-Sahalia, 2008).

The mean of the distribution is

\[ \alpha + e^{-\kappa(t'-t)}(x - \alpha) \]

and the covariance matrix is \( \lambda - e^{-\kappa(t'-t)}\lambda e^{-\kappa^\top(t'-t)} \), where

\[ \lambda = \frac{1}{2\text{tr}(\kappa)\det(\kappa)} (\det(\kappa)I_d + (\kappa - \text{tr}(\kappa)I_d)\kappa - \text{tr}(\kappa)I_d)^\top \]

with \( I_d \) being a 2 \times 2 identity matrix.


\[ dX(t) = \kappa(\alpha + \beta t - X(t))dt + dW(t), \]

is obtained if we add a deterministic term \( \beta t \) on the drift coefficient of Model 3, where \( \beta = (\beta_1, \beta_2)^\top. \)

Similar to the BOU model, the BOUI model’s transition density is explicitly known. Under it, the distribution of a future state \( X(t') \), conditional on the current state \( X(t) = x \), is also a normal with the same covariance matrix as the BOU model. But its mean is given by

\[ \alpha + \beta t + e^{-\kappa(t'-t)}(x - \alpha - \beta t) + e^{-\kappa(t'-t)} \int_0^{t'-t} e^{\kappa u} \kappa \beta du. \]
Model 5. Consider the following model
\[
\frac{d}{dt}\begin{pmatrix} S(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \mu S(t) \\ \kappa(\alpha - Y(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{(1-\rho^2)Y(t)}S(t) & \rho\sqrt{Y(t)}S(t) \\ 0 & \sigma Y^\beta(t) \end{pmatrix} dW(t),
\]
where \(W(t) = (W_1(t), W_2(t))^\top\) is a 2-dimensional standard Brownian motion, \(\mu, \kappa, \alpha, \rho,\) and \(\beta\) are all constants, \(\beta \geq 1/2\). Express the dynamic of \(S(t)\) in terms of \(X(t) = \ln(S(t))\). We have
\[
\frac{d}{dt}\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \mu - Y(t)/2 \\ \kappa(\alpha - Y(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{(1-\rho^2)Y(t)} & \rho\sqrt{Y(t)} \\ 0 & \sigma Y^\beta(t) \end{pmatrix} dW(t). \tag{19}
\]
This class of models nests several important stochastic volatility processes that are widely used in describing asset price dynamics. When we take \(\beta = 1/2\) in (19), we have the model proposed by Heston (1993); when \(\beta = 1\), it will be identical as the continuous-time GARCH model (cf. Nelson, 1990; Duan, 1995). The Heston model is affine and irreducible, while the GARCH model is an example of non-affine processes.

4.2 Density Approximation

This subsection examines the accuracy of the expansion \(p^{(J)}\) as an approximation to the true transition probability density of the underlying process. For the convenience of comparison, we consider the first four models out of the above six, i.e., OU, CIR, BOU, and BOUI, because the densities of all of them are explicitly known. We use the maximum absolute error between the \(J\)-th order Itô-Taylor expansion and the true density as a measure of approximation error. As noted in Section 2.3, only differentiation operations are involved in the expansion formula (14). We use Mathematica in the numerical experiments to compute the expansion coefficients \(w_{N,h}\).

We take the value of \(\kappa\) in both the BOU and BOUI models to be the same as what were used in Aït-Sahalia (2008) and Choi (2013), respectively. Such a choice guarantees that the eigenvalues of the matrix \(\kappa\) are real positive numbers, which is a necessary restriction to make sure that these parameters are identifiable in these two
continuous models with discretely observed data as discussed in Aït-Sahalia (2008). One may refer to, for example, Pedersen (1995), Hansen and Sargent (1983), and Kessler and Rahbek (2004) for more discussions on the identification problem of model specification.

Figure 1 displays the approximation errors of the Itô-Taylor expansion under these four models. Two general patterns arise in the experiment outcomes. First, for a fixed number of terms $J$, the error of our density approximation decreases as the observational time interval $t' - t$ shrinks. When we change $\Delta := t' - t$ from $1/12$ to $1/252$, i.e., the observation frequency from monthly to daily, the maximum absolute error of the Itô-Taylor expansions reduced very significantly. Second, when we fix the observation frequency $\Delta$, the expansion with a larger $J$ will lead to a smaller approximation error. Both patterns corroborate the theoretical statements in Theorem 2.

Through Figure 2, we intend to compare the maximum absolute error of the Itô-Taylor expansion proposed in this paper with those ones given by Aït-Sahalia (2008) and Choi (2013). As the order of the Itô-Taylor expansion increases, the performance of our Itô-Taylor expansion exceeds that of the 3rd-order density expansion of Aït-Sahalia (2008) or the 2nd-order density expansion of Choi (2013) in terms of the maximum absolute error.\(^3\)

\[\text{Figure 1 about here.}\]

Through Figure 2, we intend to compare the maximum absolute error of the Itô-Taylor expansion proposed in this paper with those ones given by Aït-Sahalia (2008) and Choi (2013). As the order of the Itô-Taylor expansion increases, the performance of our Itô-Taylor expansion exceeds that of the 3rd-order density expansion of Aït-Sahalia (2008) or the 2nd-order density expansion of Choi (2013) in terms of the maximum absolute error.\(^3\)

\[\text{Figure 2 about here.}\]

### 4.3 Monte Carlo Evidences for the Approximate MLE

In this subsection, we shall provide Monte Carlo evidences for the BOU, BOUI, Heston, and GARCH models to investigate the performance of the approximate MLE

\(^3\)Prof. Yacine Aït-Sahalia provides a 3rd-order density approximate formula for the OU, CIR and BOU model on his website [https://www.princeton.edu/~yacine/](https://www.princeton.edu/~yacine/). We thank Professor Seungmoon Choi for sharing with us his 2nd-order density approximate formulas for the BOUI model.
resulted from the Itô-Taylor expansion. Due to the space limit and the similarity in the result pattern, we focus only on these four multivariate diffusions here.

Recall that $n$ is the number of observations in each path. In light of the following decomposition
\[
\hat{\theta}_n^{(J)} - \theta_0 = (\hat{\theta}_n^{(J)} - \hat{\theta}_n) + (\hat{\theta}_n - \theta_0),
\]
we identify two sources of errors contributing to the estimation error of our Itô-Taylor expansion based MLE. The first one is $\hat{\theta}_n - \theta_0$, the discrepancy between the true MLE and the true parameter value. It measures the error caused intrinsically by the maximum likelihood method, having nothing to do with the proposed approximation. The other one is $\hat{\theta}_n^{(J)} - \hat{\theta}_n$, which is affected by the accuracy of our density approximations. Since the true densities are explicitly known under the BOU and BOUI models, we can compute the true MLEs in the experiments so as to discern quantitatively the impacts of these two types of errors. We tabulate the means and standard deviations of these errors in Tables 1 and 2.

Table 1 contains the estimation results for the BOU model. It shows that $\hat{\theta}_n^{(J)} - \hat{\theta}_n$, the difference between our approximate MLE and the true MLE, decreases rapidly as $J$ increases. In addition, this difference is dominated (at least one order of magnitude) by $\hat{\theta}_n - \theta_0$, the difference between the true MLE and the true parameter values. Hence, for the purpose of estimating $\theta_0$, our Itô-Taylor expansion based estimator $\hat{\theta}_n^{(J)}$ with a relatively small $J$ (say, $J = 4$) can be used as a meaningful substitute for the (generally incomputable) MLE $\hat{\theta}_n$. Note that the BOU process is an example of the reducible multivariate diffusions, amenable to the Lamperti transform. We also report the estimation results from the 3rd-order expansion of Aït-Sahalia (2008) in Table 1. The comparison shows that the accuracy of our estimator when taking $J = 7$ outperforms his estimators for most of the parameters.

[Table 1 about here.]

Table 2 compares the performance of a variety of MLEs under a time-inhomogeneous
model BOUI, including the MLE derived from the true process density, the approximate MLE developed in Choi (2013), and the approximate MLE based on the Itô-Taylor expansion. As we can see, the accuracy of the MLE yielded by the 4th-order Itô-Taylor expansion exceeds that of the approximate MLE using the 2nd-order Choi’s expansion. Like what Table 1 reveals, the approximation error caused by the Itô-Taylor expansion (cf. the last column in Table 2) is dominated by the sampling error of the maximum likelihood method (cf. the second column in Table 2).

[Table 2 about here.]

We consider the Heston and GARCH models in Table 3, demonstrating the performance of our method for irreducible and nonaffine diffusion processes. Neither of these two models admits closed-form expressions for their transition densities. Hence, we compute the difference between the true parameter values and the Itô-Taylor MLE to assess the method’s accuracy. Aït-Sahalia (2008) expanded the log-likelihood functions of such processes from their accompanying Kolmogorov equations to develop approximate MLEs. We also include the estimation results from his method\(^4\) in Table 3 for the purpose of comparison. With \(J = 4\), the biases of our estimators under both the Heston and GARCH models are very small relative to the true parameter values. The accuracy of our results are comparable to that of the 2nd order approximation of Aït-Sahalia (2008).

[Table 3 about here.]

Finally, we examine the standard deviation of our estimators with finite samples. In theory, when a process is known to be stationary, its true MLE should have a local asymptotic normal structure, that is,

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(0, i(\theta_0)^{-1}),
\]

\(^4\)The 2nd order density approximation formulas for the Heston and GARCH models are also available on Yacine Aït-Sahalia’s website.
as \( n \to \infty \) with \( \Delta \) fixed, where \( i(\theta_0) \) is the Fisher’s information matrix defined as

\[
i(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \ln p(X(t_1)|X(t_0); \theta) \right) \left( \frac{\partial}{\partial \theta} \ln p(X(t_1)|X(t_0); \theta) \right)^\top \right].
\]

In Table 4, we take certain parameters under which the BOU model is stationary. Using its explicit density function and our approximations, we compute the asymptotic standard deviation (ASD) of the true MLE and the finite-sample standard deviation (FSSD) of the approximate MLE respectively. The table demonstrates that the finite-sample standard deviations of our estimators are very close to the efficient asymptotic standard deviations, for different number of observations \( (n = 500, 2000, 5000) \). Moreover, the rate they decrease is the same as the order of \( \sqrt{n} \), consistent with what predicted by the local asymptotic normal structure.

[Table 4 about here.]

We carry out the same experiments for some other non-stationary diffusions. The results in Table 5 show that the standard deviations of our estimators for the BOUI, Heston and GARCH models all decrease at a rate of \( \sqrt{n} \), similar as that in the stationary case.

[Table 5 about here.]

5 Conclusion

This paper constructs a closed-form approximation of the transition densities for multivariate diffusions using the Itô-Taylor expansion, in which only differentiation is involved. We manage to prove that it will converge to the true density under a set of very mild technical conditions. In this way, the method unifies the treatment for a wide range of models, especially time-inhomogeneous irreducible non-affine multivariate diffusions. Numerical experiments illustrate the computational efficiency and accuracy of the approximate MLE resulted from our Itô-Taylor density expansion.
Acknowledgement

The authors are grateful to Prof. Dacheng Xiu of the University of Chicago, Prof. George J. Jiang of Washington State University, for their valuable suggestions. Part of the research of the second author is supported by Hong Kong Research Grant Council under GRF Grant No. CUHK1420114. The third author thanks the National Natural Science Foundation of China (NSFC) for the financial support under Grants No. 71401103 and 71421002.

A The Assumptions and Proofs

A.1 The Assumptions

We need the following technical conditions in our convergence analysis.

Assumption A.1. Let $D = \prod_{i=1}^{m} (x_i, \bar{x}_i)$ be the domain of diffusion $X$ defined in (1). It is possible that $x_i = -\infty$ and/or $\bar{x}_i = +\infty$. Moreover, the boundary of $D$ is unattainable for the process $X$.

The literature has developed a systematic approach to testing whether or not a boundary is attainable for a diffusion process through its drift and volatility coefficients. For instance, Karatzas and Shreve (1991) and Aït-Sahalia (2002) consider the issue of unattainability for univariate diffusions; Friedman (1976) discusses about multivariate diffusions in Chapter 11.

The Itô-Taylor expansion involves repeated differentiation of the drift $\mu(t, x; \theta)$ and the diffusion matrix $\nu(t, x; \theta)$, so we need,

Assumption A.2. All the components of $\mu(t, x; \theta)$ and $\nu(t, x; \theta)$ are infinitely differentiable in $(t, x)$, at any $(t, x; \theta) \in [0, +\infty) \times D \times \Theta$. 

22
The differentiability of \((\mu, \nu)\), or consequentially \((\mu, \sigma)\), implies they are locally Lipschitz continuous. This assumption thus ensures that the solution to SDE (1) is strongly unique in the sense of Definition 5.2.3 in Karatzas and Shreve (1991). However, it is not sufficient to guarantee the existence of a transition density for \(X\) in (1). Therefore, to ensure the existence of a transition density, we impose two classical conditions below (see, e.g., Friedman, 1964, Chapter 1, Theorem 10; Friedman, 1975, Chapter 6, Theorem 4.5).

**Assumption A.3.** The diffusion matrix \(\nu(t, x; \theta)\) is uniformly positive definite; that is, there exists a positive constant \(c_0\) such that \(\xi^\top \nu(t, x; \theta) \xi \geq c_0 \xi^\top \xi\) for any nonzero vector \(\xi \in \mathbb{R}^m\) and \((t, x, \theta) \in [0, +\infty) \times D \times \Theta\).

**Assumption A.4.** \(\mu(t, x; \theta)\) and \(\nu(t, x; \theta)\) are bounded and their derivatives exhibit at most polynomial growth in \(x\) for \((t, x, \theta) \in [0, +\infty) \times D \times \Theta\).

The above two assumptions are sufficient conditions for the existence. It is in general difficult (if not impossible) to relax the two conditions. A theoretical relaxation should be considered case-by-case.

### A.2 Technical Lemmas

**Lemma 1.** Fix \((t', x')\) and \((t, x)\).\(^{5}\) For \(N \geq 1\),

\[
(\partial_s + \mathcal{L})^N q(t', x'|s, y) = \sum_{|h|=1}^{2N} w_{N,h}(s, y) \partial_h q(t', x'|s, y),
\]

where the operator \(\mathcal{L}\) is defined by (4) acting on the state variable \(y\), the multivariate normal density \(q(t', x'|s, y)\) is defined as follows

\[
q(t', x'|s, y) := \frac{1}{(2\pi(t' - s))^{m/2} \det(\nu_0)^{1/2}} \exp\left(-\frac{(x' - y)^\top \nu_0^{-1}(x' - y)}{2(t' - s)}\right),
\]

and the coefficient function \(w_{N,h}(s, y)\) is defined by (16) and (17).

\(^{5}\)Thus, \(\nu_0 = \nu(t, x; \theta)\) is also fixed.
Proof of Lemma 1. We use mathematical induction to prove this lemma. For simplicity, we omit the arguments in the functions $q(t', x'|s, y)$, $\mu_i(s, y; \theta)$, $\nu_{ij}(s, y; \theta)$, and $w_{N,h}(s, y)$ without confusion hereafter.

For $N = 1$, by the definition of $q(t', x'|s, y)$, we have

$$\partial_s q = -L_0 q,$$  \hspace{1cm} (21)

where $L_0$ is a differential operator acting on $y$ defined by

$$L_0 = \frac{1}{2} \sum_{i,j=1}^{m} \nu_{0ij} \partial e_i + \nu_{ij} \partial e_j.$$

Then we have

$$(\partial_s + L)q = (-L_0 + L)q = \sum_{i=1}^{m} \mu_i \partial e_i q + \frac{1}{2} \sum_{i,j=1}^{m} (\nu_{ij} - \nu_{0ij}) \partial e_i + \nu_{ij} \partial e_j q.$$

Recalling the definition of $w_{N,h}$ in (16), (20) holds for $N = 1$.

Next, assume that (20) holds for $N$. Then, for $N + 1$, we have

$$(\partial_s + L)^{N+1} q = (\partial_s + L) \left( \sum_{|h|=1}^{2N} w_{N,h} \partial_h q \right) = \sum_{|h|=1}^{2N} \left( \partial_s (w_{N,h} \partial_h q) + L(w_{N,h} \partial_h q) \right).$$  \hspace{1cm} (22)

Applying (21) to the first term gives that

$$\partial_s (w_{N,h} \partial_h q) = (\partial_s w_{N,h}) \partial_h q + w_{N,h} (\partial_s \partial_h q) = (\partial_s w_{N,h}) \partial_h q + w_{N,h} \partial_{\partial_h q}$$

$$= (\partial_s w_{N,h}) \partial_h q - w_{N,h} (\partial_h L_0 q) = (\partial_s w_{N,h}) \partial_h q - w_{N,h} L_0 (\partial_h q).$$

The second term on the right hand side of (22) follows

$$L(w_{N,h} \partial_h q) = \sum_{i=1}^{m} \mu_i \partial e_i (w_{N,h} \partial_h q) + \frac{1}{2} \sum_{i,j=1}^{m} \nu_{ij} \partial e_i + \nu_{ij} (w_{N,h} \partial_h q)$$

$$= (L w_{N,h}) \partial_h q + w_{N,h} L(\partial_h q) + \sum_{i,j=1}^{m} \nu_{ij} (\partial e_j w_{N,h}) \partial_h e_j q.$$
Putting them together, we have
\[
(\partial_s + \mathcal{L})^{N+1} q = \sum_{|h|=1}^{2N} ((\partial_s + \mathcal{L}) w_{N,h} \partial_h q
\]
\[
+ \sum_{|h|=1}^{2N} \sum_{i=1}^{m} \nu_{ij} (\partial_{e_j} w_{N,h}) \partial_{h+e_i} q + \sum_{|h|=1}^{2N} \sum_{i=1}^{m} w_{N,h} \cdot \mu_i \partial_{h+e_i} q
\]
\[
+ \sum_{|h|=1}^{2N} \frac{1}{2} \sum_{i,j=1}^{m} w_{N,h} \cdot (\nu_{ij} - \nu_{0ij}) \partial_{h+e_i+e_j} q.
\]
Rewriting the index in the summation gives that
\[
(\partial_s + \mathcal{L})^{N+1} q = \sum_{|h|=1}^{2N} ((\partial_s + \mathcal{L}) w_{N,h} \partial_h q
\]
\[
+ \sum_{|h|=2}^{2N+1} \sum_{i,j=1}^{m} \nu_{ij} (\partial_{e_j} w_{N,h-e_i}) \partial_{h} q + \sum_{|h|=2}^{2N+1} \sum_{i=1}^{m} \mu_i w_{N,h-e_i} \partial_{h} q
\]
\[
+ \sum_{|h|=3}^{2N+2} \frac{1}{2} \sum_{i,j=1}^{m} (\nu_{ij} - \nu_{0ij}) w_{N,h-e_i-e_j} \partial_{h} q
\]
\[
= \sum_{|h|=1}^{2N+2} w_{N+1,h} \partial_{h} q.
\]
The last equality holds by the definition of \( w_{N+1,h} \) in (17) and the assignment that \( w_{N,h}(s,y) \equiv 0 \) if \( \min\{h_1, \ldots, h_m\} < 0 \), or \( h = 0 \), or \( |h| > 2N \). Hence, (20) holds for \( N + 1 \).

Lemma 2. Fix \( t \) and \( x \). Recalling the functions \( \{w_{N,h}(s,y) \equiv w_{N,h}(s,y; t, x, \theta) : s \geq 0, y \in D\} \) are defined by (16) and (17), the following statement holds: for each \( N \geq 1 \),
\[
w_{N,h}(s,y)|_{s=t,y=x} = 0, \quad \text{if} \quad 3N/2 < |h| \leq 2N.
\]
Proof of Lemma 2. The statement (23) obviously holds for \( N = 1 \). Indeed, for \( N = 1 \), \( 3N/2 < |h| \leq 2N \) implies that \( |h| = 2 \). Thus by (16), we have
\[
\left\{
\begin{array}{l}
w_{1,2e_i}(t,x) = \frac{1}{2} \left( \nu_{ii}(s,y; \theta)|_{s=t,y=x} - \nu_{ii}(t,x; \theta) \right) = 0, \quad i = 1, \ldots, m; \\
w_{1,e_i+e_j}(t,x) = \nu_{ij}(s,y; \theta)|_{s=t,y=x} - \nu_{ij}(t,x; \theta) = 0, \quad i \neq j, \quad i, j = 1, \ldots, m.
\end{array}
\right.
\]
Inspired by (24), by Taylor’s Theorem and Assumption A.2, we can expand the function $w_{N,h}(s, y)$ at $(t, x)$ up to the $K$-th order as follows:

$$w_{N,h}(s, y) = \sum_{0 \leq a_0 + |a| \leq K} \xi_{a_0, a}^{N,h} \cdot (s - t)^{a_0} (y - x)^a + \sum_{a_0 + |a| = K} \Omega_{a_0, a}^{N,h}(s, y) \cdot (s - t)^{a_0} (y - x)^a,$$

where $a_0$ is a nonnegative integer number, $a = (a_1, \cdots, a_m)$ is a vector index with nonnegative integer components, and $x^a = x_1^{a_1} \cdots x_m^{a_m}$. For each $a_0, a$, the coefficient $\xi_{a_0, a}^{N,h} := \partial_{s}^{a_0} \partial_{a} w_{N,h}(s, y)|_{s=t, y=x}$ is a constant, and the function $\Omega_{a_0, a}^{N,h}(s, y)$ satisfies $\lim_{s \to t, y \to x} \Omega_{a_0, a}^{N,h}(s, y) = 0$. Thus, we can define the Order of the function $w_{N,h}(s, y)$ at $(t, x)$ as follows

$$Or(w_{N,h}) := \min\{k \mid \xi_{a_0, a}^{N,h} \neq 0, a_0 + |a| = k\}.$$

Obviously, $w_{N,h}(s, y)|_{s=t, y=x} = 0$ if $Or(w_{N,h}) \geq 1$. Then, we can prove the statement (23) by showing a stronger statement below: for each $N \geq 1$

$$Or(w_{N,h}) \geq 2|h| - 3N, \quad \text{if} \quad 3N/2 < |h| \leq 2N. \quad (25)$$

It is equivalent to say, for each $N \geq 1$ and $3N/2 < |h| \leq 2N,$

$$w_{N,h}(s, y) = \sum_{a_0 + |a| = 2|h| - 3N} \left(\xi_{a_0, a}^{N,h} + \Omega_{a_0, a}^{N,h}(s, y)\right) \cdot (s - t)^{a_0} (y - x)^a. \quad (26)$$

We use mathematical induction to verify that (25) holds. For $N = 1$, (25) holds by (24). Assume that (25) holds for $N$. Note that (i) the Order of a summation of functions is at least the minimum Order of each function; (ii) a first order differential operator acting on a function will decrease its Order at most by 1; (iii) the Order of a multiplication of functions is the summation of the Order of each function; (iv) $Or(\nu_{ij}(s, y; \theta) - \nu_{ij}(t, x; \theta)) \geq 1$. Therefore for $N + 1$, by equation (17), we have

$$Or(w_{N+1,h}) \geq \min\{Or(w_{N,h}) - 2, Or(w_{N,h-e_i}), Or(w_{N,h-e_i}) - 1, 1 + Or(w_{N,h-e_i-e_j})\}$$

$$\geq \min\{(2|h| - 3N) - 2, (2(|h| - 1) - 3N) - 1, 1 + (2(|h| - 2) - 3N)\}$$

$$\geq 2|h| - 3(N + 1).$$

Hence we have verified that (25) holds for $N + 1$. This completes the proof. \qed
Example 1. Regarding the CIR model (cf. Model 2), we verify the result (23) in Lemma 2 for \( N = 1, 2, 3, 4 \). By a direct computation (cf. Equations (16) and (17)), the coefficients \( \{ w_{N,h}(s,y), 3N/2 < h \leq 2N \} \) are given by

\[
\begin{align*}
    w_{1,2}(s,y) &= \frac{1}{2} \sigma^2(y - x); \\
    w_{2,4}(s,y) &= \frac{1}{4} \sigma^4(y - x)^2; \\
    w_{3,5}(s,y) &= \frac{3}{4} \sigma^6(y - x) + \frac{3}{4} \kappa(\alpha - y) \sigma^4(y - x)^2; \\
    w_{3,6}(s,y) &= \frac{1}{8} \sigma^6(y - x)^3; \\
    w_{4,7}(s,y) &= \frac{3}{4} \sigma^8(y - x)^2 + \frac{1}{2} \kappa(\alpha - y) \sigma^6(y - x)^3; \\
    w_{4,8}(s,y) &= \frac{1}{16} \sigma^8(y - x)^4.
\end{align*}
\]

It is obviously that all above are zeros when \( y = x \).

A.3 The Proofs

Proof of Theorem 1. Fix \((t, x)\) and \((t', x')\). For any \( s \in [t, t') \) and \( y \in D \), by the definition in (10), we have

\[
p^{(J)}(t', x'|s, y; \theta) = \sum_{N=0}^{J} \frac{(t' - s)^N}{N!} (\partial_s + \mathcal{L})^N q(t', x'|s, y),
\]

where the operator \( \mathcal{L} \) is defined by (4) acting on state variable \( y \). By (20) in Lemma 1, we have

\[
p^{(J)}(t', x'|s, y; \theta) = q(t', x'|s, y) + \sum_{N=1}^{J} \sum_{|h|=1}^{2N} \frac{(t' - s)^N}{N!} w_{N,h}(s,y) \partial_h q(t', x'|s, y).
\]

Note that (cf. (12))

\[
\partial_h q(t', x'|s, y) = (t' - s)^{-|h|} H_h(z; \nu_0) q(t', x'|s, y)
\]

with \( z = (x' - y)/\sqrt{t' - s} \). Plugging it into (28), we have

\[
p^{(J)}(t', x'|s, y; \theta) = q(t', x'|s, y) \left( 1 + \sum_{N=1}^{J} \sum_{|h|=1}^{2N} \frac{(t' - s)^{N-|h|}}{N!} w_{N,h}(s,y) H_h(z; \nu_0) \right).
\]

27
Taking \( s = t \) and \( y = x \), by (23) in Lemma 2, we see that equation (14) holds. This completes the proof.

**Proof of Theorem 2.** Firstly, fix \((t, x)\), then \(\mu = 0 \) and \(\nu_0 = \nu(t, x; \theta)\) are also fixed. For \( s \in [t, t') \) and \( y \in D \), consider \( p(t', x'|s, y; \theta) \) and \( p^{(j)}(t', x'|s, y; \theta) \). Note that, \( p(t', x'|s, y; \theta) \) satisfies the backward Kolmogorov PDE associated with SDE (1) (see, e.g., Section 5.1 in Karatzas and Shreve, 1991)

\[
(\partial_s + \mathcal{L})p(t', x'|s, y; \theta) = 0, \quad \lim_{t' - s \to 0} p(t', x'|s, y; \theta) = \delta(x' - y),
\]

where \( \mathcal{L} \) is defined in (4). Applying \( (\partial_s + \mathcal{L}) \) to \( p^{(j)}(t', x'|s, y; \theta) \) (cf. (27)), we have

\[
(\partial_s + \mathcal{L})p^{(j)}(t', x'|s, y; \theta) = \sum_{N=1}^{J} \left( \frac{(t' - s)^N}{N!} (\partial_s + \mathcal{L})^N q(t', x'|s, y) \right) = \frac{(t' - s)^J}{J!} (\partial_s + \mathcal{L})^{J+1} q(t', x'|s, y) := \psi_J(s, y; t, x, t', x').
\]

Note that

\[
\lim_{t' - s \to 0} q(t', x'|s, y) = \delta(x' - y).
\]

To establish a similar initial condition for \( p^{(j)} \), consider any test function \( \varphi(\cdot) \), which is continuous with compact support on \( D \). By (30),

\[
\int_D \left( p^{(j)}(t', x'|s, y; \theta) - q(t', x'|s, y) \right) \varphi(x') dx' = \sum_{N=1}^{J} \sum_{|h|=1}^{2N} \frac{(t' - s)^{N-|h|/2}}{N!} w_{N,h}(s, y) \int_D q(t', x'|s, y) H_h(z; \nu_0) \varphi(x') dx',
\]

where \( z = (x' - y)/\sqrt{t' - s} \). Note that (cf. (11))

\[
q(t', x'|s, y) = (t' - s)^{-m/2} \phi(z; \nu_0).
\]

Changing the variable from \( x' \) to \( z \), and denoting \( D_Z \) as the domain of \( z \), we have

\[
\int_D q(t', x'|s, y) H_h(z; \nu_0) \varphi(x') dx' = \int_{D_Z} \phi(z; \nu_0) H_h(z; \nu_0) \varphi(y + z\sqrt{t' - s}) dz.
\]
Then, taking $t' - s \to 0$ on both sides of (34), we have

$$
\lim_{t' - s \to 0} \int_D \left( p^{(J)}(t', x'|s, y; \theta) - q(t', x'|s, y) \right) \varphi(x') dx' = \varphi(y) \sum_{N=1}^{J} \sum_{|h|=2N} \frac{1}{N!} w_{N,h}(s, y) \int_{D_2} \phi(z; \nu_0) H_h(z; \nu_0) dz,
$$

(35)

which is zero when $s = t$ and $y = x$, because $w_{N,h}(t, x) = 0$ for all $|h| = 2N$ (cf. (23)).

For any $J$, such that $J > 2m - 1$, define the error between the partial sum $p^{(J)}$ and the true density $p$ as

$$
r^{(J)}(t', x'|t, x; \theta) = p^{(J)}(t', x'|t, x; \theta) - p(t', x'|t, x; \theta).
$$

Then, by (31) and (32), it satisfies the following Kolmogorov PDE:

$$
(\partial_s + \mathcal{L})r^{(J)}(t', x'|t, x; \theta) = \psi_J(s, y; t, x, t', x').
$$

(36)

Besides, by (31), (33) and (35), when $s = t$ and $y = x$, the initial condition becomes

$$
\lim_{t' - t \to 0} r^{(J)}(t', x'|t, x; \theta) = 0.
$$

(37)

Thus,

$$
r^{(J)}(t', x'|t, x; \theta) = \int_t^{t'} \int_D \psi_J(s, y; t, x, t', x') \cdot p(s, y|t, x; \theta) dy ds.
$$

(38)

Recalling (32) and (20), we decompose $\psi_J$ into three terms:

$$
\psi_J := \psi^{(1)}_J + \psi^{(2)}_J + \psi^{(3)}_J,
$$

(39)

where for $i = 1, 2, 3$,

$$
\psi^{(i)}_J(s, y; t, x, t', x') = \sum_{h \in I_i} \frac{(t' - s)^J}{J!} w_{J+1,h}(s, y) \partial_h q(t', x'|s, y),
$$

and $I_1 = \{ h | 1 \leq |h| \leq 3(J + 1)/2 \}$, $I_2 = \{ h | 3(J + 1)/2 < |h| < 2(J + 1) \}$, $I_3 = \{ h | |h| = 2(J + 1) \}$. Then, by (38) and (39), we can rewrite the error term into a summation $r^{(J)} = r^{(J)}_1 + r^{(J)}_2 + r^{(J)}_3$, where

$$
r^{(J)}_i(t', x'|t, x; \theta) = \int_t^{t'} \int_D \psi^{(i)}_J(s, y; t, x, t', x') \cdot p(s, y|t, x; \theta) dy ds, \quad i = 1, 2, 3.
$$
Recalling \( z = (x' - y) / \sqrt{t' - s} \) and (29), we further have

\[
\begin{align*}
    r_i^{(J)} &= \int_t^{t'} \int_{D_{h\in I_i}} \left( \frac{(t' - s)^J - m}{J!} \right) H_h(z; \nu_0)q(t', x|s, y)w_{J+1,h}(s, y)p(s, y|t, x; \theta)dyds.
\end{align*}
\]

(40)

Next, consider the bounds for \( r_1^{(J)}, r_2^{(J)}, \) and \( r_3^{(J)} \), respectively. To bound each remainder term in (40), the basic idea is to decompose the integrand in (40) into two parts: \( H_h(z; \nu_0)q(t', x|s, y) \) and \( w_{J+1,h}(s, y)p(s, y|t, x; \theta) \), and bound each of them.

Consider the first term \( r_1^J \) with \( h \in I_1 \) (i.e., \( 1 \leq |h| \leq 3(J + 1)/2 \)). Recall that 
\[
    \partial_h q(t', x'|s, y) = (t' - s)^{-|h|/2}H_h(z; \nu_0)q(t', x'|s, y)
\]
in (29), together with Theorem 1 in Chapter 9 of Friedman (1964), we have

\[
|H_h(z; \nu_0)q(t', x|s, y)| \leq C(t' - s)^{-\frac{m}{2} e^{-\frac{\lambda_0 \|x' - y\|^2}{2(t' - s)}}},
\]

(41)

where \( \lambda_0, C \) are positive constants depending only on \( \nu(t, x; \theta) \), \( t, x, J \). Moreover, by Assumptions A.3 and A.4 and (6.12) in Chapter 1 of Friedman (1964), there exists \( \lambda_1 > 0 \) such that

\[
    |p(s, y|t, x; \theta)| \leq C(s - t)^{-\frac{m}{2} e^{-\frac{\lambda_1 \|y - x\|^2}{2(s - t)}}}.
\]

Since \( w_{J+1,h} \) is a polynomial and \( x \) belongs to a compact set \( D^c \), there exists \( \lambda_2 \in (0, \lambda_1) \) such that

\[
|w_{J+1,h}(s, y)p(s, y|t, x; \theta)| \leq C(s - t)^{-\frac{m}{2} e^{-\frac{\lambda_2 \|y - x\|^2}{2(s - t)}}}.
\]

(42)

Combining (41) and (42), we have

\[
\begin{align*}
    \int_D |H_h(z; \nu_0)q(t', x|s, y)w_{J+1,h}(s, y)p(s, y|t, x; \theta)|dy \\
    \leq C \int_D (t' - s)^{-\frac{m}{2} e^{-\frac{\lambda_1 \|x' - y\|^2}{2(t' - s)}}} (s - t)^{-\frac{m}{2} e^{-\frac{\lambda_1 \|y - x\|^2}{2(s - t)}}} dy \\
    \leq C(t' - t)^{-\frac{m}{2} e^{-\frac{\lambda_1 \|x' - x\|^2}{2(t' - t)}}},
\end{align*}
\]

where \( \lambda = \min \{\lambda_0, \lambda_2\} \) is a positive constant depending only on \( t, x, \) and \( J \).
(40) and \(h \in I_1\) (i.e., \(1 \leq |h| \leq 3(J + 1)/2\), thus we have

\[
|r_1^{(J)}| \leq \sum_{h \in I_1} C(t' - t)^{-\frac{m}{2}} e^{-\frac{\lambda \|y' - x\|^2}{2(t' - s)}} \int_t^{t'} (t' - s)^{\frac{J - |h|}{2}} \frac{ds}{J!}
\]

\[
\leq C(t' - t)^{\frac{J+1}{4} - \frac{m}{2}} e^{-\frac{\lambda \|y' - x\|^2}{2(t' - s)}}
\]

\[
= \mathcal{O}((t' - t)^{\frac{J+1}{4} - \frac{m}{2}}), \text{ for small } (t' - t). \tag{43}
\]

For the term \(r_2^{(J)}\) with \(h \in I_2\) (i.e., \(3(J + 1)/2 < |h| < 2(J + 1)\)), the inequality in (41) still holds. Using the expansion of \(w_{J+1,h}\) in (26), we have an alternative bound as follows

\[
|w_{J+1,h}(s,y)p(s,y|t,x;\theta)| \leq C(s - t)^{a_0 + \frac{|a|}{2} - \frac{m}{2}} e^{-\frac{\lambda_2 \|y - x\|^2}{2(s-t)}}
\]

\[
= C(s - t)^{a_0 + \frac{2|h| - 3(J + 1)}{2} - \frac{m}{2}} e^{-\frac{\lambda_2 \|y - x\|^2}{2(s-t)}}, \tag{44}
\]

where \(a_0\) is a nonnegative integer satisfying \(a_0 + |a| = 2|h| - 3(J + 1)\), and the first inequality holds due to the following fact

\[
\left| \left( \frac{y - x}{\sqrt{s - t}} \right)^a p(s,y|t,x;\theta) \right| \leq C \left| \left( \frac{y - x}{\sqrt{s - t}} \right)^a \right| e^{-\frac{\lambda_2 \|y - x\|^2}{2(s-t)}} \leq C(s - t)^{-\frac{m}{2}} e^{-\frac{\lambda_2 \|y - x\|^2}{2(s-t)}}.
\]

Then, combining (40), (41) with (44) and \(h \in I_2\) (i.e., \(3(J + 1)/2 < |h| < 2(J + 1)\)), we have

\[
|r_2^{(J)}| \leq C \sum_{h \in I_2} (t' - t)^{-\frac{m}{2}} e^{-\frac{\lambda \|y' - x\|^2}{2(t' - s)}} \int_t^{t'} (t' - s)^{\frac{J - |h|}{2}} \frac{ds}{J!} (s - t)^{a_0 + \frac{2|h| - 3(J + 1)}{2}}
\]

\[
= C \sum_{h \in I_2} (t' - t)^{a_0 + \frac{|h| - (J + 1)}{2} - \frac{m}{2}} e^{-\frac{\lambda \|y' - x\|^2}{2(t' - s)}} \cdot B(1 + J - |h|, 1 + a_0 + \frac{2|h| - 3(J + 1)}{2})
\]

\[
= \mathcal{O}((t' - t)^{\frac{J+1}{4} - \frac{m}{2}}), \text{ for small } (t' - t), \tag{46}
\]

where the beta function \(B(x, y)\) is finite for \(x > 0\) and \(y > 0\).

For the term \(r_3^{(J)}\) with \(h \in I_3\) (i.e., \(|h| = 2(J + 1)\)), the integrand \((t' - s)^{J-|h|/2} = (t' - s)^{-1}\) in (45) is not integrable around \(s = t'\). To overcome the problem, we use integration by parts to reduce the order of differentiation by one. By \(\partial_h q(t', x'|s, y) =\)
Recalling (29), similar to (41), we have
\[ \partial_{e_i} \partial_{h-e_i} q(t', x'|s, y) \] and integration by parts,\(^6\) then we have,
\[ \int_D \partial_{e_i} \partial_{h-e_i} q(t', x'|s, y) \cdot w_{J+1,h}(s, y)p(s, y|t, x; \theta) dy = \int_D \partial_{h-e_i} q(t', x'|s, y) \cdot \left( \partial_{e_i} w_{J+1,h}(s, y) \cdot p(s, y|t, x; \theta) + w_{J+1,h}(s, y) \cdot \partial_{e_i} p(s, y|t, x; \theta) \right) dy. \]

Recalling (29), similar to (41), we have
\[ |\partial_{h-e_i} q(t', x'|s, y)| = |(t' - s)^{-\frac{|h| - 1}{2}} H_{h-e_i}(z; \nu_0) q(t', x'|s, y)| \leq C(t' - s)^{-\frac{|h| - 1 + m}{2}} e^{-\frac{\lambda_0\|x'-y\|^2}{2(s-t)}}. \tag{47} \]

Similar to (44), by (26), we have
\[ |\partial_{e_i} w_{J+1,h}(s, y) \cdot p(s, y|t, x; \theta)| \leq C(s - t)^{\frac{a_0}{2} + \frac{2|h| - 3(J+1) - 1}{2} - \frac{m}{2}} e^{-\frac{\lambda_0\|y-x\|^2}{2(s-t)}}. \tag{48} \]

Similar to (42), by (26) and (6.13) in Chapter 1 of Friedman (1964), we have
\[ |w_{J+1,h}(s, y) \cdot \partial_{e_i} p(s, y|t, x; \theta)| \leq C(s - t)^{\frac{a_0}{2} + \frac{2|h| - 3(J+1) - 1}{2} - \frac{m+1}{2}} e^{-\frac{\lambda_0\|y-x\|^2}{2(s-t)}}. \tag{49} \]

Recall \( r_3^{(J)} \) defined in (38) and (39) with \( h \in I_3 \) (i.e., \( |h| = 2(J+1) \)). Then, similar to the argument for \( r_2^{(J)} \), combining (47-49), we have
\[ |r_3^{(J)}| \leq C \sum_{h \in I_3} (t' - t)^{-\frac{m}{2}} e^{-\frac{\lambda_0\|x'-x\|^2}{2(s-t)}} \cdot \int_t^{t'} \frac{(t' - s)^{-J-\frac{|h| - 1}{2}}}{J!} (s - t)^{\frac{a_0}{2} + \frac{2|h| - 3(J+1) - 1}{2} - \frac{\frac{m+1}{2}}{2}} ds \]
\[ = C \sum_{h \in I_3} (t' - t)^{-\frac{m}{2}} e^{-\frac{\lambda_0\|x'-x\|^2}{2(s-t)}} \cdot B(1, \frac{|h| - 1}{2} - \frac{1}{2}) (s - t)^{-\frac{a_0}{2} + \frac{2|h| - 3(J+1) - 1}{2} - \frac{m+1}{2}} \]
\[ = O((t' - t)^{-\frac{m}{2}}), \text{ for small } (t' - t). \tag{50} \]

Finally, putting (43), (46), and (50) into together, the error \(|p^{(J)}(t', x'|t, x; \theta)| = O((t' - t)^{-\frac{a_0}{2} - \frac{m}{2}}) \) as \( t' - t \to 0 \), uniformly for \((t, x, x', \theta) \in [0, T] \times D^e \times D \times \Theta. \]

**Proof of Theorem 3.** For fixed \( x \in D^e, \ T > 0 \) and \( 0 < t < t' < T \), let
\[ R^{(J)}(t', x'|t, x; \Theta):= \sup_{\Theta \in \Theta} |p(t', x'|t, x; \theta) - p^{(J)}(t', x'|t, x; \theta)|. \]

\(^6\)Since the values at the boundaries are of order \( \exp(-c_0\|x\|^2/(s-t)) \) (if it is not zero), which decays faster than any polynomials, we do not consider the values at the boundaries when using integration by parts.
According to Theorem 2, for fixed $x$ and $t$, $p^{(J)}(t', x'|t, x)$ converges to $p(t', x'|t, x; \theta)$ uniformly in $x' \in D$ and $\theta \in \Theta$ as $\Delta \to 0$. That is, for any $\epsilon > 0$, there exists a positive $\Delta_\epsilon$ independent of $x'$ and $\theta$, such that for all $\Delta \leq \Delta_\epsilon$, we have $R^{(J)}(t', x'|t, x; \Theta) < \epsilon$, and

$$\mathbb{E}_{\theta_0}[R^{(J)}(t', X(t')|t, X(t); \Theta)|X(t) = x] = \int_D R^{(J)}(t', x'|t, x; \Theta)p(t', x'|t, x; \theta_0)dx'$$
$$\leq \epsilon \int_D p(t', x'|t, x; \theta_0)dx' = \epsilon.$$

Thus, by Chebyshev’s inequality, the sequence $R^{(J)}(t', X_{t'}|t, X_t; \Theta)$ converges to zero in probability, given $X_t = x$, that is, given any $\zeta > 0,$

$$\lim_{\Delta \to 0} \text{Prob}[[R^{(J)}(t', X_{t'}|t, X_t; \Theta)] > \zeta |X_t = x; \theta_0] = 0.$$

Then,

$$\text{Prob}[[R^{(J)}(t', X_{t'}|t, X_t; \Theta)] > \zeta; \theta_0]$$
$$= \int_D \text{Prob}[[R^{(J)}(t', X_{t'}|t, X_t; \Theta)] > \zeta |X_t = x; \theta_0]\pi_t(x; \theta_0)dx,$$

where $\pi_t(x; \theta_0)$ is the marginal density of $X_t$ at the true parameter value $\theta_0$. Since the probability is bounded and the density is integrable, by the dominated convergence theorem, we have

$$\lim_{\Delta \to 0} \text{Prob}[[R^{(J)}(t', X_{t'}|t, X_t; \Theta)] > \zeta; \theta_0] = 0,$$

that is, $R^{(J)}(t', X_{t'}|t, X_t; \Theta)$ converges to zero in probability, as $\Delta \to 0$. And the convergence of $\sup_{\theta \in \Theta} |\ell^{(J)}_n(\theta) - \ell_n(\theta)|$ to zero in probability follows from the continuity of the logarithm functions (e.g. Lemma 4.3 in Kallenberg, 2002). The existence of the approximate MLE $\hat{\theta}^{(J)}_n$ and its convergence in probability to the true MLE $\hat{\theta}_n$ are guaranteed by the continuous differentiability in $\theta$ of the log-likelihood function $\ell_n(\theta)$ and the approximate log-likelihood functions $\ell^{(J)}_n(\theta)$ for all $J$. \qed
References


Notes: The maximum absolute error between the true densities $p$ and the Itô-Taylor expansion $p^{(J)}$ (14) is defined as $\max_{x' \in \mathcal{D}} |p(t', x'| t, x; \theta) - p^{(J)}(t', x'| t, x; \theta)|$, where $x$, $t$, and $t'$ are fixed, the region $\mathcal{D}$ of forward state $x'$ is large enough to include several standard deviations from the mean. We use the following values for the parameters in these four models: OU: $(\kappa, \alpha, \sigma) = (0.5, 0.06, 0.03)$; CIR: $(\kappa, \alpha, \sigma) = (0.5, 0.06, 0.15)$; BOU: $(\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}) = (0, 0, 5, 1, 10)$; BOUI: $(\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, \beta_1, \beta_2) = (0, 0, 5, 1, 10, 0.1, 0.1)$. 

38
Figure 2: Comparison of Maximum Absolute Error between Different Expansions

Notes: This picture compares the maximum absolute error of our Itô-Taylor expansion and that of Aït-Sahalia (2008) and Choi(2013). The definition of maximum absolute error is similar to that defined in Figure 1. The values of parameters in these four models are given in the following: OU: \((\kappa, \alpha, \sigma, t' - t) = (0.5, 0.06, 0.03, 1/52)\); CIR: \((\kappa, \alpha, \sigma, t' - t) = (0.5, 0.06, 0.15, 1/52)\); BOU: \((\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, t' - t) = (0, 0, 5, 1, 10, 1/252)\); BOUI: \((\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, \beta_1, \beta_2, t' - t, t) = (0, 0, 5, 1, 10, 0.1, 0.1, 1/252, 0)\).
Table 1: Monte Carlo Evidence for the BOU Model with Different Order of Expansion

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$\hat{\theta}_n - \theta_0$</th>
<th>$\hat{\theta}_n^{(AS)} - \theta_n$</th>
<th>$\hat{\theta}_n^{(4)} - \theta_n$</th>
<th>$\hat{\theta}_n^{(5)} - \theta_n$</th>
<th>$\hat{\theta}_n^{(6)} - \theta_n$</th>
<th>$\hat{\theta}_n^{(7)} - \theta_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{11} = 5$</td>
<td>0.41888</td>
<td>0.00186</td>
<td>0.00774</td>
<td>-0.00368</td>
<td>0.00222</td>
<td>0.00080</td>
</tr>
<tr>
<td></td>
<td>(1.11974)</td>
<td>(0.04472)</td>
<td>(0.14139)</td>
<td>(0.10338)</td>
<td>(0.06851)</td>
<td>(0.06916)</td>
</tr>
<tr>
<td>$\kappa_{21} = 1$</td>
<td>0.03538</td>
<td>0.00105</td>
<td>0.01423</td>
<td>-0.00017</td>
<td>0.00276</td>
<td>-0.00123</td>
</tr>
<tr>
<td></td>
<td>(1.19813)</td>
<td>(0.02933)</td>
<td>(0.15667)</td>
<td>(0.11368)</td>
<td>(0.08693)</td>
<td>(0.05589)</td>
</tr>
<tr>
<td>$\kappa_{22} = 10$</td>
<td>0.61443</td>
<td>0.00324</td>
<td>0.04357</td>
<td>-0.000490</td>
<td>0.00239</td>
<td>-0.00097</td>
</tr>
<tr>
<td></td>
<td>(1.54834)</td>
<td>(0.02971)</td>
<td>(0.27662)</td>
<td>(0.22413)</td>
<td>(0.10551)</td>
<td>(0.10502)</td>
</tr>
<tr>
<td>$\alpha_1 = 0$</td>
<td>0.00486</td>
<td>0.00091</td>
<td>0.00021</td>
<td>0.00105</td>
<td>0.00132</td>
<td>0.00052</td>
</tr>
<tr>
<td></td>
<td>(0.06284)</td>
<td>(0.01477)</td>
<td>(0.02000)</td>
<td>(0.01959)</td>
<td>(0.01962)</td>
<td>(0.01983)</td>
</tr>
<tr>
<td>$\alpha_2 = 0$</td>
<td>-0.00033</td>
<td>0.00009</td>
<td>0.00003</td>
<td>-0.00008</td>
<td>0.00003</td>
<td>-0.00003</td>
</tr>
<tr>
<td></td>
<td>(0.03351)</td>
<td>(0.00443)</td>
<td>(0.00675)</td>
<td>(0.00653)</td>
<td>(0.00637)</td>
<td>(0.00563)</td>
</tr>
</tbody>
</table>

Notes: We use $\theta_0$ to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., we take $t' - t = 1/52$). The first column reports true parameter values $\theta_0$. The second column reports the bias and the standard derivation (values in parentheses) of the true maximum likelihood estimator $\hat{\theta}_n$. The third column shows the difference between true maximum likelihood estimator $\hat{\theta}_n$ and the 3rd-order approximate estimator $\hat{\theta}_n^{(AS)}$ developed by Aït-Sahalia (2008). The remaining columns report the differences between true maximum likelihood estimator $\hat{\theta}_n$ and Itô-Taylor expansion estimator $\hat{\theta}_n^{(J)}$ developed in this paper, with the standard derivation in parentheses. The order of our Itô-Taylor expansion takes values from $J = 4$ to $J = 7$.  


Table 2: Monte Carlo Evidence for the BOUI Model

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$\hat{\theta}_n - \theta_0$</th>
<th>$\hat{\theta}^{(Choi)}_n - \hat{\theta}_n$</th>
<th>$\hat{\theta}^{(4)}_n - \hat{\theta}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{11} = 5$</td>
<td>0.59446</td>
<td>0.08046</td>
<td>-0.01367</td>
</tr>
<tr>
<td></td>
<td>(1.17004)</td>
<td>(0.26004)</td>
<td>(0.33544)</td>
</tr>
<tr>
<td>$\kappa_{21} = 1$</td>
<td>0.10497</td>
<td>-1.11154</td>
<td>0.00273</td>
</tr>
<tr>
<td></td>
<td>(1.24774)</td>
<td>(1.94661)</td>
<td>(0.34327)</td>
</tr>
<tr>
<td>$\kappa_{22} = 10$</td>
<td>0.62750</td>
<td>-0.37080</td>
<td>-0.11460</td>
</tr>
<tr>
<td></td>
<td>(1.57026)</td>
<td>(0.38605)</td>
<td>(0.52059)</td>
</tr>
<tr>
<td>$\alpha_1 = 0$</td>
<td>-0.00163</td>
<td>0.00546</td>
<td>0.00263</td>
</tr>
<tr>
<td></td>
<td>(0.08942)</td>
<td>(0.07228)</td>
<td>(0.07931)</td>
</tr>
<tr>
<td>$\alpha_2 = 0$</td>
<td>0.00205</td>
<td>0.00137</td>
<td>-0.00024</td>
</tr>
<tr>
<td></td>
<td>(0.06047)</td>
<td>(0.03442)</td>
<td>(0.03417)</td>
</tr>
<tr>
<td>$\beta_1 = 0.1$</td>
<td>0.00112</td>
<td>-0.00057</td>
<td>-0.00047</td>
</tr>
<tr>
<td></td>
<td>(0.01839)</td>
<td>(0.01124)</td>
<td>(0.01241)</td>
</tr>
<tr>
<td>$\beta_2 = 0.1$</td>
<td>-0.00085</td>
<td>0.00046</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td>(0.01122)</td>
<td>(0.00540)</td>
<td>(0.00560)</td>
</tr>
</tbody>
</table>

Notes: We use $\theta_0$ to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., we take $t' - t = 1/52$). The first column reports true parameter values $\theta_0$. The second column contains the bias of the true maximum likelihood estimator $\hat{\theta}_n$. We display the differences between the true maximum likelihood estimator $\hat{\theta}_n$ and the 2nd order approximate estimator $\hat{\theta}^{(Choi)}_n$ developed by Choi (2013) in the third column. The fourth column shows the differences between $\hat{\theta}_n$ and the 4th-order approximate estimator $\hat{\theta}^{(4)}_n$ using our Itô-Taylor expansion in (14). All standard deviations are reported in the parentheses.
Table 3: Monte Carlo Evidence for Heston and GARCH Models

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>Heston Model ($\beta = 1/2$)</th>
<th>GARCH Model ($\beta = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\theta}_n^{(AS)} - \theta_0$</td>
<td>$\hat{\theta}_n^{(4)} - \theta_0$</td>
</tr>
<tr>
<td>$\sigma = 0.25$</td>
<td>0.00064</td>
<td>0.00060</td>
</tr>
<tr>
<td></td>
<td>(0.00644)</td>
<td>(0.00643)</td>
</tr>
<tr>
<td>$\rho = -0.8$</td>
<td>-0.0075</td>
<td>-0.0059</td>
</tr>
<tr>
<td></td>
<td>(0.01339)</td>
<td>(0.01335)</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>0.00024</td>
<td>0.00026</td>
</tr>
<tr>
<td></td>
<td>(0.00735)</td>
<td>(0.00747)</td>
</tr>
<tr>
<td>$\mu = 0.03$</td>
<td>-0.00284</td>
<td>-0.00342</td>
</tr>
<tr>
<td></td>
<td>(0.08155)</td>
<td>(0.08164)</td>
</tr>
<tr>
<td>$\kappa = 3$</td>
<td>0.14815</td>
<td>0.14498</td>
</tr>
<tr>
<td></td>
<td>(0.53684)</td>
<td>(0.54298)</td>
</tr>
</tbody>
</table>

Notes: We use $\theta_0$ to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., we take $t' - t = 1/52$). The two subtables correspond to the results from the Heston and GARCH models, respectively. The first column of each subtable displays the bias of the 2nd order approximate estimator $\hat{\theta}_n^{(AS)}$ developed by Ait-Sahalia (2008); the second column illustrates the bias of the 4th-order approximate estimator $\hat{\theta}_n^{(4)}$ using our Itô-Taylor expansion in (14). All standard deviations are reported in the parentheses.
Table 4: Comparison between Asymptotic and Finite-Sample Standard Deviation for the BOU Model

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>500 observations</th>
<th>2000 observations</th>
<th>5000 observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASD</td>
<td>FSSD</td>
<td>ASD</td>
</tr>
<tr>
<td>$\kappa_{11} = 5$</td>
<td>1.02 1.12</td>
<td>0.51 0.51</td>
<td>0.32 0.33</td>
</tr>
<tr>
<td>$\kappa_{21} = 1$</td>
<td>1.08 1.18</td>
<td>0.54 0.54</td>
<td>0.34 0.34</td>
</tr>
<tr>
<td>$\kappa_{22} = 10$</td>
<td>1.45 1.53</td>
<td>0.73 0.72</td>
<td>0.46 0.43</td>
</tr>
<tr>
<td>$\alpha_1 = 0$</td>
<td>0.065 0.063</td>
<td>0.032 0.030</td>
<td>0.020 0.019</td>
</tr>
<tr>
<td>$\alpha_2 = 0$</td>
<td>0.033 0.034</td>
<td>0.016 0.017</td>
<td>0.010 0.011</td>
</tr>
</tbody>
</table>

Notes: This table presents the asymptotic standard deviation (ASD) and the finite-sample standard deviation (FSSD) as the number of observations are 500, 2000, and 5000, respectively. The number of simulation trials is 1000 for all three cases. The length of the time interval is fixed at $t' - t = 1/52$. All the results for FSSD are based on the 4th-order approximate estimator $\hat{\theta}_n^{(4)}$ using our Itô-Taylor expansion in (14).
Table 5: Finite-Sample Standard Deviation for the BOUI, Heston, GARCH Models with Different Number of Observations

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>BOUI</th>
<th>Heston</th>
<th>GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 observations</td>
<td>2000 observations</td>
<td>5000 observations</td>
</tr>
<tr>
<td>$\kappa_{11} = 5$</td>
<td>1.18</td>
<td>0.50</td>
<td>0.34</td>
</tr>
<tr>
<td>$\kappa_{21} = 1$</td>
<td>1.20</td>
<td>0.54</td>
<td>0.33</td>
</tr>
<tr>
<td>$\kappa_{22} = 10$</td>
<td>1.52</td>
<td>0.71</td>
<td>0.48</td>
</tr>
<tr>
<td>$\alpha_1 = 0$</td>
<td>0.094</td>
<td>0.037</td>
<td>0.021</td>
</tr>
<tr>
<td>$\alpha_2 = 0$</td>
<td>0.062</td>
<td>0.030</td>
<td>0.015</td>
</tr>
<tr>
<td>$\beta_1 = 0$</td>
<td>0.019</td>
<td>0.0020</td>
<td>0.00050</td>
</tr>
<tr>
<td>$\beta_2 = 0$</td>
<td>0.012</td>
<td>0.0014</td>
<td>0.00038</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Heston</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 0.25$</td>
<td>0.0064</td>
<td>0.0031</td>
<td>0.0021</td>
</tr>
<tr>
<td>$\rho = -0.8$</td>
<td>0.013</td>
<td>0.0070</td>
<td>0.0042</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>0.0075</td>
<td>0.0036</td>
<td>0.0021</td>
</tr>
<tr>
<td>$\mu = 0.03$</td>
<td>0.082</td>
<td>0.037</td>
<td>0.022</td>
</tr>
<tr>
<td>$\kappa = 3$</td>
<td>0.54</td>
<td>0.25</td>
<td>0.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma = 0.25$</td>
<td>0.0064</td>
<td>0.0030</td>
<td>0.0020</td>
</tr>
<tr>
<td>$\rho = -0.8$</td>
<td>0.013</td>
<td>0.0065</td>
<td>0.0042</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>0.0023</td>
<td>0.0013</td>
<td>0.0008</td>
</tr>
<tr>
<td>$\mu = 0.03$</td>
<td>0.080</td>
<td>0.046</td>
<td>0.030</td>
</tr>
<tr>
<td>$\kappa = 3$</td>
<td>0.55</td>
<td>0.25</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Notes: This table presents the finite-sample standard deviation for the BOUI, Heston, GARCH Models. The number of observations are 500, 2000, and 5000, respectively. The number of simulation trials is 1000 for all three cases. The length of the time interval is fixed at $t' - t = 1/52$. All the results are based on the 4th-order approximate estimator $\hat{\theta}_n^{(4)}$ using our Itô-Taylor expansion in (14).