Journal of Econometrics 209 (2019) 256–288

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A new delta expansion for multivariate diffusions via the Itô-Taylor expansion

Nian Yang [a], Nan Chen [b], Xiangwei Wan [c,*]

[a] *Department of Finance and Insurance, School of Economics, Nanjing University, Nanjing, China*
[b] *Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong*
[c] *Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China*

ARTICLE INFO

Article history:
Received 16 October 2017
Received in revised form 1 January 2019
Accepted 8 January 2019
Available online 21 January 2019

JEL classification:
C13
C32
C63

Keywords:
Closed-form density expansion
Delta expansion
Itô-Taylor expansion
Multivariate diffusions
Maximum likelihood estimation

ABSTRACT

In this paper we develop a new delta expansion approach to deriving analytical approximation to the transition densities of multivariate diffusions using the Itô-Taylor expansion of the conditional expectation of the Dirac delta function. Our approach yields an explicit recursive formulas for the expansion coefficients and is universally applicable for a wide spectrum of models, particularly the time-inhomogeneous non-affine irreducible multivariate diffusions. We show that this new approach can be viewed as an extension of Aït-Sahalia (2002) and Lee et al. (2014) to the case of multivariate models. The derived expansions are proved to converge to the true probability density as the observational time interval shrinks. The obtained approximations can thereby be used to carry out the maximum likelihood estimation for the diffusions with discretely observed data. Extensive numerical experiments demonstrate the accuracy and effectiveness of our approach.

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 (see, e.g., Lo, 1988). However, except a handful of simple instances, explicit expressions of the transition density for 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 of diffusion processes.

Following this rich literature, we propose a new delta expansion approach in the current paper to deriving density approximations for multivariate time-inhomogeneous diffusion processes. The new approach represents the transition density as a conditional expectation of the Dirac delta function on the diffusion process, and expands it into a sum of Hermite polynomial based terms by applying the celebrated Itô-Taylor expansion on it. Due to the analytical form of the delta expansion, we can easily obtain the corresponding approximate maximum likelihood estimator for statistical inference when substituting it into the log-likelihood function of observations from the diffusion process sampled at finite time intervals.

The proposed approach contributes to the literature in two aspects. First, compared with the existent expansion methods, the most salient advantage of the new approach is that we can establish an explicit recursive formula among the expansion coefficients for the density of a general time-inhomogeneous multivariate diffusion process. As far as we know, Lee et al.

* Corresponding author.
E-mail addresses: yangnian@nju.edu.cn (N. Yang), nchen@se.cuhk.edu.hk (N. Chen), xwwan@sjtu.edu.cn (X. Wan).

https://doi.org/10.1016/j.jeconom.2019.01.003
0304-4076/

(2014) derive a similar result in their investigation on univariate diffusions. Our approach greatly generalizes their findings. This new structural discovery brings forth to the delta expansion significant computational efficiency: we can easily compute higher order expansion coefficients from the ones of lower orders through this explicit formula. In addition, since only function differentiation is involved in our expansion, we may even utilize some symbolic computing softwares to speed up calculation of the expansion coefficients.

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. This result is new for the Hermite polynomial based approximations. Aït-Sahalia (2008) notes that the classical Hermite expansion in general may not converge if applied to irreducible processes. 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.

Diffusion density approximation and the related MLE have been well investigated in the literature. Aït-Sahalia (2002) makes a fruitful breakthrough in which he uses Hermite polynomials as the orthogonal basis to approximate the transition density of a univariate time-homogeneous diffusion. Later on, Lee et al. (2014) obtain their delta expansion after some term rearrangements of the Hermite expansion developed in Aït-Sahalia (2002). It turns out that our expansion approach is closely related to their methods. As discussed in Section 3.1, the delta expansion proposed in the current paper allows one degree of freedom in a parameter μ_0. We prove that our expansion will yield the same approximation as what Lee et al. (2014) construct in the case of univariate time-homogeneous diffusions when we take $\mu_0 = 0$ in our expansion formula. In this sense, our method can be viewed as one natural extension to the multivariate time-inhomogeneous setting of the Hermite approach employed in Aït-Sahalia (2002). However, we need to emphasize that the flexibility in μ_0 does grant our method some efficiency in computing the expansion coefficients because we find that other choice of μ_0 than 0 will typically lead to different but much simpler expansion formulas.

One crucial step in the method proposed in Aït-Sahalia (2002) is to apply the Lamperti transform to unitize the process volatility. Unfortunately, not every multivariate process is amenable to such a transformation. To circumvent the difficulty in dealing with the irreducible processes, Aït-Sahalia (1999, 2008) suggest using the Kolmogorov backward and forward equations to determine a small-time expansion of the diffusion probability densities. Choi (2013, 2015) extend this approach to the time-inhomogeneous diffusions. By nature, our delta expansion is different from theirs. We expand the density of irreducible processes in an ascending power order of $\sqrt{\Delta}$, the square root of the length of observational time interval, while their method leads to an approximation series in the integer order of Δ.

For time-homogeneous irreducible diffusions, Li (2013) proposes an alternative approximation to their density functions based on the theory of Malliavin calculus. His method involves heavily computation of the conditional expectation of the multiplication of iterated Itô integrals. As a sideline contribution of this paper, we explicitly express such expectations as a linear combination of the Hermite polynomials and thereby develop a recursive formula for his expansion. This finding casts new insights into Li's method; that is, the small-time expansion of Li (2013), although being derived from more advanced tools such as the Malliavin calculus, is essentially an expansion consisting of the Hermite polynomials. Moreover, we can show through the symbolic computation function in *Mathematica* that his expansion leads to the same result as our delta expansion with $\mu_0 = 0$ up to any given order for one- and two-dimensional diffusions with general drift and volatility coefficients.

Although it still remains as an interesting open problem to theoretically justify whether Li (2013)'s expansion is identical to our delta expansion with $\mu_0 = 0$ for general time-homogeneous diffusions in higher dimension, the differences between the two methods are obvious. First, our method is extended to the time-inhomogeneous cases. Second, the flexibility in the choice of μ_0, which his expansion lacks of, enables us to simplify our delta expansion a lot.

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 noted previously, 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. 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 Aït-Sahalia and Yu (2006), Yu (2007), Filipović et al. (2013), and Li and Chen (2016), just to name a few.

The rest of the paper is organized as follows. Section 2 defines the requirements on the diffusion model considered in the paper. In Section 3, the major part of the paper, we develop our delta expansion to approximate the transition density of a general multivariate diffusion process via the Itô-Taylor expansion and establish its convergence to the true transition density. Section 4 presents some convergence results about the resulting approximate maximum likelihood estimators. We discuss the relationship of our approach to the other existing methods in Section 5. Section 6 contains numerical evidence of the performance of the approximate transition densities and the approximate maximum likelihood estimators under various diffusions. Technical 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 $h = (h_1, h_2, \ldots, h_m)$ 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 ith 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 $\partial_h := \partial^{|h|}/(\partial x_1^{h_1} \cdots \partial x_m^{h_m})$ to be the partial derivatives with respect to the state variable $x := (x_1, x_2, \ldots, x_m)^\top \in D$. 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 model

Consider a multivariate diffusion process

$$dX(t) = \mu(t, X(t); \theta)dt + \sigma(t, X(t); \theta)dW(t), \tag{1}$$

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 functions μ and σ 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 θ belongs to a compact set $\Theta \subset \mathbb{R}^L$. Define the variance–covariance (or diffusion) matrix of the diffusion X by

$$\nu(t, x; \theta) := \sigma(t, x; \theta)\sigma(t, x; \theta)^\top. \tag{2}$$

Later we will use ν_{ij} to denote its (i, j)-element, $1 \leq i, j \leq m$.

Given any 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,

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

In order to build up an MLE for θ, we assume that a sequence of values of the state variables is observed over a discrete time grid $\{t = t_i : i = 0, 1, \ldots, n\}$. By the Markovian property of (1), the log-likelihood function[1] is then given by

$$\ell_n(\theta) := \sum_{i=1}^n \ln p(t_i, X(t_i)|t_{i-1}, X(t_{i-1}); \theta). \tag{3}$$

We have the maximum likelihood estimator of θ defined by the maximizer of the following optimization problem

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

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 $p(t', x'|t, x; \theta)$ are unavailable in most cases. As noted in the introduction, the focus of this paper is to establish an analytical approximation to the function p via an Itô-Taylor expansion based approach. Through it, we obtain an approximation to the MLE $\hat{\theta}_n$ and analyze its convergence properties in Section 4. Besides the MLE, our density approximation can be potentially used in some other contexts such as option pricing, constructing test statistics for diffusions, and so on.

Below are some technical conditions that we need for developing our approximation. Most of them are standard in the literature of MLE for diffusion processes. The first assumption is

Assumption 2.1. Let $D = \Pi_{i=1}^m(\underline{x}_i, \bar{x}_i)$ be the domain of diffusion X defined in (1). It is possible that $\underline{x}_i = -\infty$ and/or $\bar{x}_i = +\infty$. Moreover, the boundary of D is unattainable for the process X.

Note that in the applications relevant in finance, when the SDEs are used to model asset prices or nominal interest rates, their domains are often taken as $D = \Pi_{i=1}^m(0, +\infty)$. 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 the boundary behaviors of multivariate diffusions in Chapter 11 of his book.

The proposed delta expansion requires us to repeatedly differentiate the drift $\mu(t, x; \theta)$ and the volatility matrix $\sigma(t, x; \theta)$. Hence additionally, we need

Assumption 2.2. All the components of $\mu(t, x; \theta)$ and $\sigma(t, x; \theta)$ are infinitely differentiable in (t, x) at any $(t, x, \theta) \in (0, +\infty) \times D \times \Theta$.

Furthermore, the differentiability of (μ, σ) implies that they are locally Lipschitz continuous, thus ensuring that the solution to SDE (1) is strongly unique in the sense of Definition 5.2.3 in Karatzas and Shreve (1991).

However, the above two assumptions are still not sufficient to guarantee the existence of a transition probability density for X defined in (1). Therefore, we need to enhance our requirements by imposing two classical conditions as below:

[1] As in Aït-Sahalia (2002, 2008), we ignore the unconditional density of the first observation $(t_0, X(t_0))$.

Assumption 2.3. 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 $(t, x, \theta) \in [0, +\infty) \times D \times \Theta$.

and

Assumption 2.4. $\mu(t, x; \theta)$ and $\nu(t, x; \theta)$ are bounded and their derivatives exhibit at most polynomial growth in x for $(t, x, \theta) \in [0, +\infty) \times D \times \Theta$.

It is worth stressing that these two assumptions are conventionally proposed to provide sufficient (but not necessary) conditions for the existence of the transition density (see, e.g., Friedman, 1964, Chapter 1, Theorem 10; Friedman, 1975, Chapter 6, Theorem 4.5). For convenience, the theoretical proof of the convergence results of the paper is based on these conditions. However, the practical usefulness of the proposed delta expansion covers a wide range of commonly used models, rather than confined to those strictly satisfying Assumptions 2.3 and 2.4. For instance, these two assumptions are not necessary for the algorithm to compute the expansion coefficients when we derive the density approximation in Theorems 3.1 and 3.3. In addition, the numerical experiments in the paper also suggest that the approximations developed from our expansion method performs very well for many processes that may not satisfy them. It points to the possibility of relaxing these two conditions. We leave the related theoretical justification for future works.

3. The delta expansion of the transition density

This section is devoted to the development of the main results of the paper, how to obtain our delta expansion to approximate the transition probability density of a diffusion process. As the first step to this end, we need to use the Itô-Taylor expansion to expand the density. We motivate the Itô-Taylor expansion method by first presenting the heuristic idea behind it in Section 3.1.

3.1. Heuristic idea behind the Itô -Taylor expansion

Consider any sufficiently smooth function $G(s, y)$. By the Itô formula, we have

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

where ν_{ij} is the (i, j)-element of the diffusion matrix ν (cf. Eq. (2) for its definition). Let $\mathbb{E}^{t,x}[\cdot]$ denote expectation conditional on $X(t) = x$. Taking expectations on both sides of the above equality, we have

$$\mathbb{E}^{t,x}[G(s, X(s))] = G(t, x) + \int_t^s \mathbb{E}^{t,x}[(\partial_u + \mathcal{L})G(u, X(u))]du, \tag{4}$$

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

$$(\mathcal{L}G)(u, y) = \sum_{i=1}^m \mu_i(u, y; \theta)\partial_{e_i} G(u, y) + \frac{1}{2}\sum_{i,j=1}^m \nu_{ij}(u, y; \theta)\partial_{e_i+e_j} G(u, y), \tag{5}$$

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

$$\mathbb{E}^{t,x}[(\partial_{u_1} + \mathcal{L})G(u_1, X(u_1))] = (\partial_t + \mathcal{L})G(t, x) + \int_t^{u_1} \mathbb{E}^{t,x}[(\partial_{u_2} + \mathcal{L})^2 G(u_2, X(u_2))]du_2. \tag{6}$$

Substituting (6) back into (4), we have

$$\mathbb{E}^{t,x}[G(s, X(s))] = G(t, x) + (\partial_t + \mathcal{L})G(t, x) \cdot (s - t) + \mathbb{E}^{t,x}\left[\int_t^s du_1 \int_t^{u_1} (\partial_{u_2} + \mathcal{L})^2 G(u_2, X(u_2))du_2\right].$$

In this way, repeatedly applying the expansion for J times yields

$$\mathbb{E}^{t,x}[G(s, X(s))] = \sum_{N=0}^J \frac{(s - t)^N}{N!}(\partial_t + \mathcal{L})^N G(t, x) + \mathcal{R}_J, \tag{7}$$

where the remainder term $\mathcal{R}_J$ is given by

$$\mathcal{R}_J = \mathbb{E}^{t,x}\left[\int_t^s du_1 \int_t^{u_1} du_2 \cdots \int_t^{u_J} (\partial_{u_{J+1}} + \mathcal{L})^{J+1} G(u_{J+1}, X(u_{J+1}))du_{J+1}\right].$$

Now, we turn to apply the Itô-Taylor expansion (7) to approximate the transition density p. The density function admits the following expression (see, e.g., Watanabe, 1987; Li, 2013):

$$p(t', x'|t, x; \theta) = \mathbb{E}^{t,x}[\delta_{x'}(X(t'))],$$

where $\delta_{x'}(\cdot)$ is the Dirac delta function centered at x'. However, the function $\delta_{x'}(\cdot)$ does not have any derivatives in the classical sense. To circumvent this obstacle from the irregularity of $\delta_{x'}(\cdot)$, we introduce a sequence of smooth functions to approximate it. More precisely, for any given constant vector $\mu_0 \in \mathbb{R}^m$, we may define a function G such that, for $t \le s < t'$ and $y \in \mathbb{R}^m$,

$$G_{t',x'}(s, y) := \frac{1}{(2\pi(t'-s))^{\frac{m}{2}} \det(\nu_0)^{\frac{1}{2}}} \exp\left(-\frac{(x'-y-(t'-s)\mu_0)^\top \nu_0^{-1}(x'-y-(t'-s)\mu_0)}{2(t'-s)}\right), \tag{8}$$

where $\nu_0 := \nu(t, x; \theta) \in \mathbb{R}^{m\times m}$, the diffusion matrix of the process X when $X_t = x$. In other words, if we fix t', s and x', $G_{t',x'}(s, \cdot)$ is simply the corresponding 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:

$$\lim_{s\uparrow t'} \mathbb{E}^{t,x}[G_{t',x'}(s, X(s))] = \mathbb{E}^{t,x}[\delta_{x'}(X(t'))] = p(t', x'|t, x; \theta). \tag{9}$$

Since the function $G_{t',x'}$ is infinitely differentiable for any fixed t' and x', we now can invoke the Itô-Taylor expansion (7) to expand the conditional expectation $\mathbb{E}^{t,x}[G_{t',x'}(s, X(s))]$ on the left-hand side of (9). That is, if we omit the residual term $\mathcal{R}_J$, then

$$\mathbb{E}^{t,x}[G_{t',x'}(s, X(s))] \approx \sum_{N=0}^{J} \frac{(s-t)^N}{N!} \left[(\partial_s + \mathcal{L})^N G_{t',x'}(s, y)\Big|_{s=t,y=x}\right],$$

where the infinitesimal generator $\mathcal{L}$ is acting on the state variable y. From (9), letting s tend to t' in the above expression results in our density approximation, we have

$$p(t', x'|t, x; \theta) \approx \sum_{N=0}^{J} \frac{(t'-t)^N}{N!} \left[(\partial_s + \mathcal{L})^N G_{t',x'}(s, y)\Big|_{s=t,y=x}\right] =: p^{(J)}(t', x'|t, x; \theta) \tag{10}$$

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

Note that in principle other forms of mollifiers are possible for approximating the Dirac delta function. Compared with the other alternatives, the major advantage of the choice of the normal kernel (8) is that repeatedly differentiating it, as what we do in (10), will lead to the Hermite polynomials, from which we can construct recursive formulae (cf. (18)–(19)) to facilitate the computation of the coefficients of our expansion. In addition, it turns out that, under the aforementioned μ_0 and ν_0, the leading term of the resulting expansion (cf. (17)) is identical to the marginal distribution of the following process $\tilde{X}$ at $s = t'$:

$$d\tilde{X}(s) = \mu_0 ds + \sigma(t; x, \theta)dW(s), \ \ s \ge t,$$

and $\tilde{X}(t) = x$. Obviously the process $\tilde{X}$ is a multivariate Brownian motion with constant drift μ_0 and constant volatility matrix $\sigma(t; x, \theta)$. In this sense, our method essentially expands the transition probability density of the original process X around that of $\tilde{X}$. Recall that the process X after time t, if $X(t) = x$, satisfies

$$dX(s) = \mu(s, X(s); \theta)ds + \sigma(s, X(s); \theta)dW(s), \ \ s \ge t. \tag{11}$$

The choice of ν_0 thus entails that these two processes have the same volatilities at the initial time point $s = t$. From this observation, we anticipate that the expansion based on such ν_0 should provide a good approximation over a small time scale.[2] Theorem 3.2 in the next section corroborates this intuition.

[2] This anticipation can be understood easily if we also choose $\mu_0 = \mu(t; x, \theta)$ in the expansion. In such case, the two processes have the same drift and volatility coefficients at the initial point $s = t$. Intuitively, the drifted Brownian motion $\tilde{X}$ under these choices of μ_0 and ν_0 should evolve closely to the process of X within a small time horizon after t because $\mu(s, X(s); \theta)$ and $\sigma(s, X(s); \theta)$ will not change too much away from $\mu(t, x; \theta)$ and $\sigma(t, x; \theta)$. This intuition is exactly the origin of the Euler method for numerically simulating the solution of an SDE. Kloeden and Platen (1992) discuss the convergence properties of the Euler method when small time steps are taken. Furthermore, we need to point out that, in contrast to the choice of ν_0, the value of μ_0 is not crucial for the convergence of our expansion, although a proper μ_0 will significantly simplify its expression. On one hand, the drift term is of order "dt", while the volatility term is of order "$\sqrt{dt}$". Thus, for a small time expansion, ν_0 would play a more important role for the convergence analysis. On the other hand, thanks to the Girsanov theorem, the probability distributions of $\tilde{X}$ under different choices of μ_0, as long as they share the same volatility matrix $\sigma(t, x; \theta)$, are absolutely continuous with each other. The differences in the expansion expressions led by different μ_0 are mainly caused by the corresponding Radon–Nikodym derivative. We show in Theorem 3.2 that such differences will not affect the rate of the convergence.

3.2. The delta expansion from the Itô -Taylor expansion approach

Noting that the derivation in Section 3.1 is heuristic, in this section we shall establish the convergence property of our Itô-Taylor density approximation rigorously. 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 $G_{t',x'}$. Define some more notations to facilitate the presentation of the analytical form of $p^{(J)}$. Let $\phi(z;\Sigma)$ be the density of the m-dimensional normal distribution with mean 0 and covariance matrix Σ, i.e.,

$$\phi(z;\Sigma)=\frac{1}{(2\pi)^{m/2}\det(\Sigma)^{1/2}}\exp\left(-\frac{z^{\top}\Sigma^{-1}z}{2}\right). \tag{12}$$

For an m-dimensional nonnegative integer vector $h=(h_1,\ldots,h_m)$, define $H_h(z;\Sigma)$ to be the corresponding multivariate Hermite polynomial associated with this normal density; that is,

$$H_h(z;\Sigma):=(-1)^{|h|}\phi^{-1}(z;\Sigma)\partial_h\phi(z;\Sigma). \tag{13}$$

Willink (2005) provides a recursive approach to computing these multivariate Hermite polynomials. Namely, for any vector h with $h_j\geq 0, j=1,\ldots,m$, we have

$$H_{h+e_k}(z;\Sigma)=\left(\sum_{j=1}^{m}\Sigma^{(kj)}z_j\right)H_h(z;\Sigma)-\sum_{j=1}^{m}\Sigma^{(kj)}h_jH_{h-e_j}(z;\Sigma), \tag{14}$$

where $\Sigma^{(kj)}$ is the (k,j)-element of the matrix Σ^{-1}. In addition, $H_0(z;\Sigma)=1$ and $H_h(z;\Sigma)=0$ for any h with $\min\{h_1,\ldots,h_m\}<0$.

With the help of the Hermite polynomials, we can show

Theorem 3.1 (*Itô-Taylor Expansion*). *Suppose that Assumptions 2.1–2.2 hold and $\nu(t,x;\theta)$ is non-degenerate. Let $\nu_0=\nu(t,x;\theta)$ and μ_0 be a constant vector. For $J\geq 1$, there exists a sequence of functions $\{w_{N,h}\}$ such that*

$$p^{(J)}(t',x'|t,x;\theta)=q(t',x'|t,x)\left(1+\sum_{N=1}^{J}\sum_{|h|=1}^{\lfloor 3N/2\rfloor}\frac{w_{N,h}(t,x)H_h(z;\nu_0)}{N!}\Delta^{N-\frac{|h|}{2}}\right), \tag{15}$$

where $\lfloor 3N/2\rfloor$ is the largest integer less than or equal to $3N/2$, and $H_h(z;\nu_0)$ are the Hermite polynomials defined through (13) with

$$\Delta=t'-t,\quad z=\frac{x'-x-\mu_0\Delta}{\sqrt{\Delta}}, \tag{16}$$

and

$$q(t',x'|t,x):=\frac{1}{(2\pi\Delta)^{m/2}\det(\nu_0)^{1/2}}\exp\left(-\frac{z^{\top}\nu_0^{-1}z}{2}\right). \tag{17}$$

Theorem 3.1 clearly characterizes the structure of $p^{(J)}$: it can be expressed in terms of a linear combination of Hermite polynomials. More importantly, we can analytically determine the coefficient functions w in a recursive fashion specified as follows. This feature makes our expansion computationally more appealing than other Hermite polynomial based methods.

The series of functions $\{w_{N,h}(s,y)\}$ is defined on $[0,+\infty)\times D$. Each of them is indexed by a positive integer N and an m-dimensional integer valued vector $h=(h_1,\ldots,h_m)$ with $h_j\geq -2$ for all $j=1,\ldots,m$. Furthermore, they satisfy the following three relations.

(i) For any $N\geq 1$, $w_{N,h}(s,y)\equiv 0$ if either $\min\{h_1,\ldots,h_m\}<0$, either $h=0$, or $|h|>2N$.

(ii) When $N=1$, we have

$$\begin{cases} w_{1,e_i}(s,y)=\mu_i(s,y;\theta)-\mu_{0i}, & i=1,\ldots,m;\\ w_{1,2e_i}(s,y)=\frac{1}{2}(\nu_{ii}(s,y;\theta)-\nu_{ii}(t,x;\theta)), & i=1,\ldots,m;\\ w_{1,e_i+e_j}(s,y)=\nu_{ij}(s,y;\theta)-\nu_{ij}(t,x;\theta), & i\neq j,\ i,j=1,\ldots,m, \end{cases} \tag{18}$$

where $\mu_i(\cdot,\cdot;\theta)$ is the i-component of the drift vector μ and $\nu_{ij}(\cdot,\cdot;\theta)$ is the (i,j)-element of the diffusion matrix ν.

(iii) When $N>1$ and all the components in h are nonnegative, and $0<|h|\leq 2N$, we have

$$w_{N,h}(s,y)=(\partial_s+\mathcal{L})w_{N-1,h}(s,y)+\sum_{i=1}^{m}\mathcal{A}_iw_{N-1,h-e_i}(s,y)+\frac{1}{2}\sum_{i,j=1}^{m}(\nu_{ij}(s,y;\theta)-\nu_{ij}(t,x;\theta))w_{N-1,h-e_i-e_j}(s,y), \tag{19}$$

where the infinitesimal generator $\mathcal{L}$ is given in (5), and the operators $\{\mathcal{A}_i, i = 1, \ldots, m\}$ are defined as following: for a smooth function f,

$$(\mathcal{A}_i)f(s,y) = (\mu_i(s,y;\theta) - \mu_{0i})f(s,y) + \sum_{j=1}^{m} \nu_{ij}(s,y;\theta)\partial_{e_j} f(s,y). \tag{20}$$

The operators $\mathcal{L}$ and $\mathcal{A}_i$ are all acting on the state variable y.

Eq. (19) relates a function with higher indices to those with lower ones. We can repeatedly use it to compute the explicit form of $w_{N,h}$ for any N and h. Meanwhile, only differentiation operations are involved in the computation. That makes the implementation of our expansion convenient: we can even simply use some symbolic computation programs such as Mathematica to accomplish the derivation. In this way, the method suggested in this paper avoids some complex preprocessing for the symbolic calculation in solving Kolmogorov PDEs that some other expansion methods require.

The next theorem, as the key step to establish the convergence of the delta expansion in Theorem 3.3, characterizes the uniform convergence rate of the expansion $p^{(J)}$ to the true transition density p.

Theorem 3.2. *Suppose that Assumptions 2.1–2.4 hold. Define $p^{(J)}(t',x'|t,x;\theta)$ through (15). Then, given any positive integer $J > 2m-1$ and compact subset $D^c \subset D$, as $\Delta = t'-t \to 0$, we have*

$$\sup_{(t,x,x',\theta)\in[0,T]\times D^c\times D\times\Theta} \left| p^{(J)}(t',x'|t,x;\theta) - p(t',x'|t,x;\theta)\right| = \mathcal{O}\left(\Delta^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil - \frac{m}{2}}\right), \tag{21}$$

where $\lceil (J+1)/2\rceil$ is the smallest integer larger than or equal to $(J+1)/2$.

This theorem provides a theoretical guarantee of the accuracy of our Itô-Taylor expansion based approximation. It shows that the absolute error between $p^{(J)}$ and the true density p is uniformly bounded by a quantity of order $\Delta^{(\lceil (J+1)/2\rceil - m)/2}$. The conditions in the theorem statement (cf. Assumptions 2.1–2.4) further indicate another advantage of the Itô-Taylor expansion; that is, its convergence property is insensitive to the special structure of the underlying diffusion. Under this expansion, the above approximation order universally applies for a wide range of models, whether they are univariate or multivariate, reducible or irreducible, affine or non-affine, or time-homogeneous or not. In a sharp contrast, the previous literature, including Aït-Sahalia (2002, 2008), Egorov et al. (2003), Choi (2013, 2015), Li (2013), and Filipović et al. (2013), develop model-specific expansion methods to obtain approximations to the probability densities of processes of different types.

The proof of Theorem 3.2 is deferred to Appendix A.2. The key observation in it is that the coefficients $w_{N,h}(t,x)$ are all zeros for $|h| > 3N/2$ under the choice of $\nu_0 = \nu(t,x;\theta)$ (cf. Lemma A.2). We thereby build up a tight upper bound estimation on the approximation error in order to establish the convergence of the expansion. We find that such observation may not be true for some other choices of ν_0, which implies that the resulting expansion will be divergent, consistent with the intuition discussed in footnote 2.

The index of J in Theorem 3.2 is used to stress how many times we operate the differentiation $\partial_t + \mathcal{L}$ for expanding the density. Hence the previous Itô-Taylor expansion is not according to the power order of Δ, different from how a majority of the existing methods present their results. For the convenience of making a comparison between our expansion and the others, we arrive at a *Delta expansion* in the following theorem by rearranging the expansion terms in an ascending order of $\sqrt{\Delta}$.

Theorem 3.3 (*Delta Expansion*)**.** *For any integer $K \geq 1$, define the delta expansion of the transitional probability density $p(t',x'|t,x;\theta)$ by*

$$p^{(K,\Delta)}(t',x'|t,x;\theta) = q(t',x'|t,x)\left(1 + \sum_{k=1}^{K} \Delta^{\frac{k}{2}} C_k\right), \tag{22}$$

where the coefficient C_k is given by

$$C_k = \sum_{N=\lceil (k+1)/2\rceil}^{2k} \frac{1}{N!} \sum_{|h|=2N-k} w_{N,h}(t,x) H_h(z;\nu_0). \tag{23}$$

Suppose the assumptions in Theorem 3.2 hold. Then, given any positive integer $K > m-1$ and compact subset $D^c \subset D$, as $\Delta = t'-t \to 0$,

$$\sup_{(t,x,x',\theta)\in[0,T]\times D^c\times D\times\Theta} \left| p^{(K,\Delta)}(t',x'|t,x;\theta) - p(t',x'|t,x;\theta)\right| = \mathcal{O}\left(\Delta^{\frac{K+1}{2} - \frac{m}{2}}\right). \tag{24}$$

We need to emphasize that this new delta expansion does not change the order of the convergence. More accurately, when we take $J = 2K$, both expansions are of the same order. However, as shown in the proof of Theorem 3.3, the delta expansion is obtained essentially by omitting those terms of order higher than $\Delta^{(\lceil J/2\rceil - m)/2}$ in $p^{(J)}$. That is why the new expansion usually leads to much simpler expressions than the Itô-Taylor expansion, especially when we choose $\mu_0 = \mu(t,x;\theta)$ in the expansion. In this way, the delta expansion yields computational efficiency in the numerical experiments; see more discussion in Example A.2 in Appendix A.1. In addition to the numerical advantage, another appealing feature of the delta expansion in the theoretical aspect is that it can be shown to be equivalent to the Hermite expansion developed in Aït-Sahalia (2002) and the small-time expansion of Li (2013) in the case of time-homogeneous univariate diffusions. In this sense, the delta expansion can be regarded as a generalization of these existing methods. For these two reasons, we focus on the discussion of the delta expansion from now on.

3.3. The expansions for reducible diffusions

Aït-Sahalia (2008) introduces an important concept of reducibility to identify a class of multivariate diffusions that are amenable to his multivariate Hermite expansion approach. A diffusion is said to be reducible if and only if there exists a one-to-one transformation of the diffusion into a new one whose diffusion matrix is the identity matrix. Every univariate diffusion is reducible in this sense by means of the Lamperti transform. However, that is not true for multivariate diffusions. Aït-Sahalia (2008) provides a necessary and sufficient conditions for the reducibility (cf. Section 3 in his paper). Choi (2013) further extends the discussion of reducibility to the time-inhomogeneous diffusions.

Thanks to the identity diffusion matrix of the transformed process, approximating its probability density is computationally more tractable because it is "closer" to a standard normal than that of the original process (see, e.g., Aït-Sahalia, 2002). Once we have an expansion for the transition density of the transformed process, we can obtain the corresponding approximation to the original object of interest, by applying the Jacobian formula. Hence, without loss of generality, we only focus on a reducible diffusion process after the transformation in this subsection. Abuse the notation a little bit by using the same X to refer to the post-transformation process; that is, it satisfies the following SDE

$$dX(t) = \mu(t, X(t); \theta)dt + dW(t). \tag{25}$$

Note that the above diffusion (25) is just a special case of (1) when $\sigma = Id_m$ with Id_m being an $m \times m$ identity matrix. We can then easily invoke the computation in Theorems 3.1 and 3.2 to establish its Jth order Itô-Taylor expansion. In particular,

$$p^{(J)}(t', x'|t, x; \theta) = \Delta^{-\frac{m}{2}}\phi(z; Id_m)\left(1 + \sum_{N=1}^{J}\sum_{|h|=1}^{N}\frac{w_{N,h}(t, x)H_h(z; Id_m)}{N!}\Delta^{N-\frac{|h|}{2}}\right), \tag{26}$$

where $\phi(z; Id_m)$ and $H_h(z; Id_m)$ are the normal density function and the Hermite polynomials defined through (12) and (14), respectively, with Σ replaced by Id_m. Here z and $\{w_{N,h}\}$ are still defined by (16) and the recursive relations (i)–(iii), respectively.

We need to stress that the reducibility does bring forth to us significant computational advantages. We show in Lemma A.2 that $w_{N,h} \equiv 0$ if $|h| > N$. Due to this structural feature, the inner sum in the Itô-Taylor expansion (26) only needs to add up to N terms. In contrast, we need up to $3N/2$ summands in the expansion of an irreducible process (cf. (15)). We include Examples A.1 and A.2 in the appendix for an illustration of this point. The reducibility allows us to enhance the error order of the Itô-Taylor expansion from $\mathcal{O}(\Delta^{(\lceil (J+1)/2\rceil - m)/2})$ to $\mathcal{O}(\Delta^{(J+1)/2-m/2})$. Therefore, we have Kth order the delta expansion below by taking $J = K$ in the Itô-Taylor expansion (26).

Corollary 3.1. *For any integer $K \geq 1$, define the Kth order delta expansion $p^{(K,\Delta)}$ for the diffusion (25) by*

$$p^{(K,\Delta)}(t', x'|t, x; \theta) = \Delta^{-\frac{m}{2}}\phi(z; Id_m)\left(1 + \sum_{k=1}^{K}\Delta^{\frac{k}{2}}C_k\right), \tag{27}$$

where

$$C_k = \sum_{N=\lceil (k+1)/2\rceil}^{k}\frac{1}{N!}\sum_{|h|=2N-k} w_{N,h}(t, x)H_h(z; Id_m). \tag{28}$$

Given any positive integer $K > m - 1$ and compact subset $D^c \subset D$, as $\Delta = t' - t \to 0$,

$$\sup_{(t,x,x',\theta)\in[0,T]\times D^c\times D\times\Theta}\left|p^{(K,\Delta)}(t', x'|t, x; \theta) - p(t', x'|t, x; \theta)\right| = \mathcal{O}\left(\Delta^{\frac{K+1}{2}-\frac{m}{2}}\right). \tag{29}$$

Note that the outer sum in (28) only contains k terms rather than $2k$ terms, leading to a much simpler expression (cf. (23) for the irreducible case).

4. Approximate maximum likelihood estimators

In this section, we shall use the delta expansion $p^{(K,\Delta)}$ in (22) (or equivalently,the Itô-Taylor expansion $p^{(J)}$ in (15)) as an approximate to the true but unknown transition density p to compute approximately MLE. To analyze the convergence properties of such approximations, assume that the parameter space Θ is a compact subset of $\mathbb{R}^L$ and denote the true value of the parameter vector to be $\theta_0 \in \Theta$. In addition, suppose that the drift vector $\mu(t, x; \theta)$ and the diffusion matrix $\nu(t, x; \theta)$ in the model are infinitely continuous differentiable with respect to $\theta \in \Theta$. We have observed a set of data $\{X(t_0), X(t_1), \ldots, X(t_n)\}$ over a discrete time grid. For simplicity, assume that $t_i - t_{i-1} = \Delta$ for all $i = 1, \ldots, n$. Suppose the log-likelihood function $\ell_n(\theta)$ (cf. (3)) has a unique maximizer $\hat{\theta}_n \in \Theta$. Then, it should be the true MLE of parameter θ. But $\hat{\theta}_n$ is not computable because we do not know the exact form of $\ell_n(\theta)$.

Replacing the true probability density p in (3) with its approximation $p^{(K,\Delta)}$,[3] we have an approximate log-likelihood function

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

Maximizing $\ell_n^{(K,\Delta)}(\theta)$ over Θ leads to the approximate MLE $\hat{\theta}_n^{(K,\Delta)}$. We can show that

Theorem 4.1. *Fix the sample size n and $K > m - 1$. Under Assumptions 2.1–2.4,*

$$\hat{\theta}_n^{(K,\Delta)} - \hat{\theta}_n \to 0$$

in $\mathbb{P}_{\theta_0}$-probability as $\Delta \to 0$.[4]

Theorem 4.1 constitutes a very useful step towards establishing the asymptotic consistency of our approximate MLE $\hat{\theta}_n^{(K,\Delta)}$. For instance, if we happen to know that the true MLE $\hat{\theta}_n$ converges to θ_0 as $\Delta \to 0$, then we can choose a subsequence $\Delta_n \to 0$ such that $\hat{\theta}_n^{(K,\Delta_n)} - \theta_0 \to 0$ as $n \to +\infty$, following the proof in Aït-Sahalia (2008). Aït-Sahalia (2002) and Chang and Chen (2011) investigated the asymptotic properties of the true MLEs for univariate diffusions and use them to show the consistency of their MLE approximations. To our best knowledge, the corresponding results about the convergence of the true MLEs in multivariate diffusions are challenging and still open in the literature. We provide some supportive evidences in numerical experiments to show that the approximate MLE from our delta expansion converge to the true parameter values. A thorough theoretic investigation is beyond the scope of the current paper. We leave it for future research.

5. Relations to the existing density approximations

In this section, we discuss the relation of our delta expansion to some other existing approximations, including the Hermite expansion approach proposed in Aït-Sahalia (2002) and Lee et al. (2014), the Malliavin calculus theoretic approach suggested by Li (2013), and the Kolmogorov equation based method in Aït-Sahalia (1999, 2008) and Choi (2013, 2015).

5.1. Relation to the Hermite expansions

Aït-Sahalia (2002) pioneers the investigation on how to use Hermite polynomial series to approximate diffusion transition densities. To make our comparison more concrete, let us briefly review the main results in that paper. It mainly focuses on univariate processes. Note that any univariate process must be reducible in the sense of Section 3.3. We may start from the following setup; that is, let X be a univariate time-homogeneous diffusion defined by

$$dX(t) = \mu(X(t))dt + dW(t), \tag{30}$$

where $W(t)$ is a one-dimensional standard Brownian motion. And denote the infinitesimal generator corresponding to (30) by

$$\mathcal{L} = \mu(x)\frac{\partial}{\partial x} + \frac{1}{2}\frac{\partial^2}{\partial x^2}. \tag{31}$$

For any given states x and x', use $p(\Delta, x'|x)$ to indicate the transition probability of $X(t+\Delta) = x'$ starting from $X(t) = x$.

The Hermite expansion proposed in Aït-Sahalia (2002) consists of two steps; see Eqs. (2.7) and (4.1)–(4.3) therein. First, expand the true density $p(\Delta, x'|x)$ in an orthogonal basis comprising Hermite polynomials $\{H_h, h \geq 0\}$ to obtain

$$p(\Delta, x'|x) = \frac{1}{\sqrt{\Delta}}\phi(z)\sum_{h=0}^{\infty} \eta^{(h)}(\Delta, x)H_h(z), \tag{32}$$

where $\phi(\cdot)$ is the standard normal density and $z = (x' - x)/\sqrt{\Delta}$. Obviously each coefficient of the above expansion admits the following expectation based representation

$$\eta^{(h)}(\Delta, x) = \frac{1}{h!}\mathbb{E}^{t,x}\left[H_h\left(\frac{X(t+\Delta) - x}{\sqrt{\Delta}}\right)\right].$$

[3] $p^{(K,\Delta)}$ 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 $\max\{p^{(K,\Delta)}, \varepsilon/K\}$ for some fixed small $\varepsilon > 0$. Similar procedures are also used in Aït-Sahalia (2002) and Egorov et al. (2003).

[4] A similar convergence conclusion can be shown to be true for the approximate MLE $\hat{\theta}_n^{(J)}$ based on the Itô-Taylor expansion $p^{(J)}$.

The second step applies Taylor expansion, in conjunction with the process infinitesimal operator $\mathcal{L}$ (cf. (31)), to compute the conditional expectation on the right-hand side. We then have

$$\eta^{(h)}(\Delta, x) = \frac{1}{h!}\sum_{k=0}^{\infty}\frac{\Delta^k}{k!}\left(\mathcal{L}^k H_h\left(\frac{y-x}{\sqrt{\Delta}}\right)\right)\Bigg|_{y=x}. \tag{33}$$

Substituting (33) back into (32) yields a double infinite sum to expand $p(\Delta, x'|x)$.

A variety of ways of gathering the terms are possible in order to obtain an approximation to the transition density. For instance, Aït-Sahalia (2002) suggests expanding (32) and (33) up to prefixed orders of J and K respectively. That leads to,

$$\hat{p}^{(J,K)}(\Delta, x'|x) = \frac{1}{\sqrt{\Delta}}\phi(z)\sum_{h=0}^{J}\left(\sum_{k=0}^{K}\frac{\Delta^k}{k!}\frac{1}{h!}\left(\mathcal{L}^k\cdot H_h\left(\frac{y-x}{\sqrt{\Delta}}\right)\right)\Bigg|_{y=x}\right)H_h(z). \tag{34}$$

He shows that a relatively small value for K, say $K = 3$, will deliver satisfactory numerical accuracy in approximating the densities of many univariate diffusion models. Lee et al. (2014) consider an alternative on the basis of the explicit form of $\mathcal{L}^k H_h$ they find.

It turns out that our delta expansion produces essentially the same approximation as what Lee et al. (2014) construct if we choose $\mu_0 = 0$ in our expansion formula. Indeed, interchanging the order of summation in (34) and taking J as a function of K, Lee et al. (2014) reach the following approximated Hermite expansion

$$\hat{p}^{(K,\Delta)} := \frac{\phi(z)}{\sqrt{\Delta}}\sum_{k=0}^{2K}\Delta^{\frac{k}{2}}\left(\sum_{N=\lceil k/2\rceil}^{k}\frac{1}{N!}\zeta_{k-N,2N-k}(x)H_{2N-k}(z)\right), \tag{35}$$

where the function ζ is defined in (75) (see also Definition 1 and Eq. (18) in their paper). We prove that

Proposition 5.1. *Let $p^{(2K,\Delta)}$ be the delta expansion obtained from our approach (cf. (27)) by taking $\mu_0 = 0$. Then,*

$$\hat{p}^{(K,\Delta)} = p^{(2K,\Delta)}. \tag{36}$$

In the sense of the proposition, the delta expansion proposed in this paper can be viewed as a rearranged Hermite expansion in the case of univariate diffusions.

As noted by Aït-Sahalia (2008), the classical Hermite expansion will not in general converge if applied to irreducible processes. Contributing to this literature, our new method builds up expansions that can converge to the true probability densities. In addition, the new expansion admits one degree of freedom in the choice of μ_0. When we take $\mu_0 = \mu(t, x; \theta)$, the resulting approximation will be significantly different from what we can obtain from any other existing methods. Example A.2 shows that such choice of μ_0 can greatly simplify the expressions of our expansion. We defer the related discussion on the numerical performance to the next section.

5.2. *Relation to the expansion of Li (2013)*

Li (2013) presents a small-time asymptotic expansion of transition densities for multivariate diffusions based on the theory of Malliavin calculus developed in Watanabe (1987) and Yoshida (1992). To facilitate the comparison, we manage to establish in Appendix B a new recursive algorithm for his method. Note that Li's expansion heavily relies on computation of the conditional expectation of the multiplication of iterated Itô integrals. As a sideline contribution of this paper, we explicitly reduce such expectations to a linear combination of the Hermite polynomials; see Propositions B.2 and B.3. This new result enables us to compute the expansion coefficients of a given multivariate time-homogeneous diffusion up to any arbitrary order in a more efficient way under Li's expansion. More importantly, it also casts a new insight into the relation between Li's expansion and the Hermite polynomial based expansions: the density approximation led by his method is essentially an expansion consisting of the Hermite polynomials.

With the help of the new recursive relation developed in Appendix B, we can show through the symbolic computation function in *Mathematica* that the small-time asymptotic method in Li (2013) leads to the same expansion result as our delta expansion with $\mu_0 = 0$ up to any given order for one- and two-dimensional diffusions with general drift and volatility coefficients.[5]

Although it remains as an interesting open problem to theoretically justify whether these two methods are the same for general multivariate time-homogeneous diffusions, we still need to emphasize several appealing features of the proposed delta expansion. First, the techniques that we use are more elementary, completely avoiding the heavy machinery of Malliavin calculus. Second, our approach enjoys larger flexibility in the expansion. For example, taking $\mu_0 = \mu(t, x; \theta)$ leads to different but simpler expansions than what Li (2013) obtains.[6] Finally, as Li (2013) focuses on time-homogeneous processes, it is not obvious how his results would be extended to the time-inhomogeneous cases that we consider in this paper.

[5] We have implemented the two algorithms presented in Appendix B.1 and Theorem 3.3 in *Mathematica*. The code is available upon request.

[6] In Example A.2, the formulas in cases (a) and (b) are obtained by taking $\mu_0 = 0$ and $\mu_0 = \mu(t, x; \theta)$, respectively. Obviously, the formulas in case (b) is much simpler than that in case (a), whereas the latter is the same as that provided by Li (2013).

5.3. Relations to other methods

The multivariate diffusions are rarely reducible. An entirely different research venue is therefore pursued in Aït-Sahalia (1999, 2008). It starts from a key observation that the transition probability density of a diffusion process should be a solution to the Kolmogorov forward and backward equations. By postulating an appropriate form for the solution as an expansion of both time and space, one can compute it approximately up to the relevant order. Choi (2013, 2015) extends this Kolmogorov method to multivariate time-inhomogeneous cases. Our Itô-Taylor expansion approach is essentially different from the above Kolmogorov equation method for irreducible diffusions, as we expand the density in an ascending power order of $\sqrt{\Delta}$ in (15) while their method expands it along the integer order of Δ. Such distinction can be easily seen from the unrelated expressions of both expansions.

6. Numerical experiments

Below we undertake some numerical experiments to examine the performance of our delta expansion based approximate densities and the associated approximate MLE. Nine different types of models are considered: the Ornstein–Uhlenbeck (OU) model, the CIR model, the SEV-ND model (Aït-Sahalia, 1996), the bivariate OU (BOU) model, the time-inhomogeneous bivariate OU model (BOUI), the Heston model, the non-affine GARCH, the stochastic volatility CEV (SVCEV) model, and the time-inhomogeneous trivariate BDFS model (EBDFS, cf. Choi (2013) and Balduzzi et al. (1996)). The first three are univariate diffusions, in which the SEV-ND is one model that explicit Lamperti transform is not available. The fourth and sixth models (i.e., BOU and Heston) are multivariate reducible and irreducible diffusions, respectively. All the models of OU, CIR, BOU, and Heston are affine and time-homogeneous. In contrast, the BOUI and EBDFS models are time-inhomogeneous and the GARCH and SVCEV models are non-affine. The purpose of taking so many processes is to investigate how the approach will perform under a wide spectrum of diffusions.

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 6.1, we provide more detailed modeling information about these nine processes. In Section 6.2, we illustrate through numerical experiments that the approximate density stemmed from our expansion converges to the true density in a very fast manner. In Section 6.3, we provide Monte Carlo evidence to show the accuracy and efficiency of our approximate MLE.

6.1. 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 α. 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

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

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.

Model 3 (*The SEV-ND Model*)**.**

$$dX(t) = (\alpha_0 + \alpha_1 X(t) + \alpha_2 X^2(t) + \alpha_3 X^{-1}(t))dt + \sqrt{\beta_0 + \beta_1 X(t) + \beta_2 X^{\beta_3}(t)}dW(t).$$

Aït-Sahalia (1996) proposes the above one-dimensional diffusion process to model the short term rate. Bakshi et al. (2006) use this process to model equity volatility dynamics. A prominent feature of this model is that it allows a nonlinear drift and a stochastic elasticity of variance. Under different choices of model parameters, it nests several theoretically appealing models that admit constant elasticity of variance with nonlinear, linear, and constant drift, respectively. As pointed by Bakshi et al. (2006), one may encounter a problem when using the method of Aït-Sahalia (2002) because the SEV-ND model does not allow an explicit formula for the Lamperti transform. Bakshi et al. (2006) overcome this difficulty by reducing it to a set of one-dimensional integrals.

Model 4. *Bivariate Ornstein–Uhlenbeck (BOU) Model.*

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

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.* κ *is a* 2×2 *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 κ 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

$$\lambda = \frac{1}{2\mathrm{tr}(\kappa)\det(\kappa)}\left(\det(\kappa)Id_2 + (\kappa - \mathrm{tr}(\kappa)Id_2)(\kappa - \mathrm{tr}(\kappa)Id_2)^\top\right)$$

with Id_2 being a 2×2 identity matrix.

Model 5. *Bivariate Time-Inhomogeneous Ornstein–Uhlenbeck (BOUI) Model.*

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

is obtained if we add a deterministic term βt *on the drift coefficient of Model 4, where* $\beta = (\beta_1, \beta_2)^\top$.

Similarly 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

$$\alpha + \beta t + e^{-\kappa(t'-t)}(x - \alpha - \beta t) + e^{-\kappa(t'-t)}\int_0^{t'-t} e^{\kappa u}\kappa\beta u du.$$

Model 6. *Consider the following model*

$$d\begin{pmatrix} S(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \mu S(t) \\ \kappa(\alpha - Y(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{(1-\rho^2)Y(t)}S(t) & \rho\sqrt{Y(t)}S(t) \\ 0 & \sigma Y^\beta(t) \end{pmatrix} dW(t),$$

where $W(t) = (W_1(t), W_2(t))^\top$ *is a 2-dimensional standard Brownian motion,* μ, κ, α, ρ, *and* β *are all constants,* $\beta \geq 1/2$. *Express the dynamic of* $S(t)$ *in terms of* $X(t) = \ln(S(t))$. *We have*

$$d\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \mu - Y(t)/2 \\ \kappa(\alpha - Y(t)) \end{pmatrix} dt + \begin{pmatrix} \sqrt{(1-\rho^2)Y(t)} & \rho\sqrt{Y(t)} \\ 0 & \sigma Y^\beta(t) \end{pmatrix} dW(t). \tag{37}$$

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 (37), 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); when β is unspecified, it is the stochastic volatility CEV (SVCEV) model. The Heston model is affine and irreducible, while the GARCH and SVCEV models are examples of non-affine processes with the latter being more nonlinear.

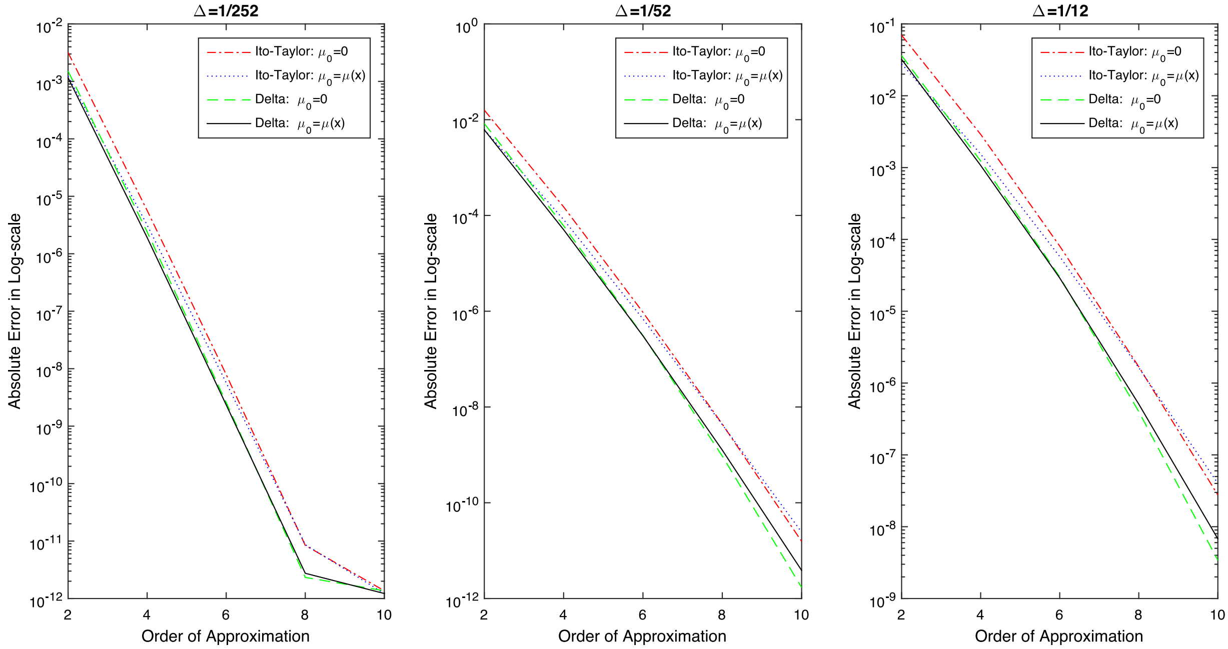

Fig. 1. The Itô-Taylor and delta expansions with different μ_0 under the CIR model. *Notes*: The maximum absolute error between the true densities p and the delta expansion $p^{(K,\Delta)}$ in (27) (the Itô-Taylor expansion $p^{(J)}$ in (26)) is defined as $\max_{x'\in\mathcal{D}}|p(t',x'|t,x;\theta)-p^{(K,\Delta)}(t',x'|t,x;\theta)|$ $(\max_{x'\in\mathcal{D}}|p(t',x'|t,x;\theta)-p^{(J)}(t',x'|t,x;\theta)|)$, 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 CIR model: $(\kappa,\alpha,\sigma)=(0.5,0.06,0.15)$.

Model 7. *Extended BDFS (EBDFS) Model.*

$$d\begin{pmatrix}X_1(t)\\X_2(t)\\X_3(t)\end{pmatrix}=\begin{pmatrix}k_1\left(X_2(t)-X_1(t)\right)\\k_2\left(\alpha_2+\beta_2 t-X_2(t)\right)\\k_3\left(\alpha_3-X_3(t)\right)\end{pmatrix}dt+\begin{pmatrix}\sqrt{(1-\rho^2)X_3(t)} & 0 & \rho\sqrt{X_3(t)}\\0 & \sigma_{21}e^{\sigma_{22}t} & 0\\0 & 0 & \sigma_{31}\sqrt{X_3(t)}\end{pmatrix}dW(t). \tag{38}$$

Balduzzi et al. (1996) extended the univariate diffusion model of the short rate to a three-factor model by introducing a stochastic long-run mean and a stochastic volatility. Choi (2013) added the time-inhomogeneity by using the Hull–White model for the second factor. Thus, the model is time-inhomogeneous and irreducible.

6.2. Density approximation

This subsection examines the accuracy of our expansion $p^{(J)}$ defined in (15) or $p^{(K,\Delta)}$ in (22) as an approximation to the true transition probability density of the underlying process. Furthermore, for reducible diffusions, we first perform the Lamperti transform, and then use (26) or (27) to compute the expansion formulas. For the convenience of comparison, we consider the following four models: OU, CIR, BOU, and BOUI, because the transition densities of all of them are explicitly known. We use the maximum absolute error between the Jth order Itô-Taylor or the Kth order delta expansion and the true density as a measure of approximation error. As noted in Section 3, we provide recursive relations between the expansion coefficients. In which only differentiation is involved. We use Mathematica in the numerical experiments to compute the expansion coefficients $w_{N,h}$.

We take the value of κ 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 κ 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.

Fig. 1 plots the approximation errors of four kinds of our expansions for the CIR model: the Itô-Taylor expansion by taking $\mu_0=\mu(t,x;\theta)$ and $\mu_0=0$, and the delta expansion by taking $\mu_0=\mu(t,x;\theta)$ and $\mu_0=0$. The numerical performances for the expansion under the choices of $\mu_0=\mu(t,x;\theta)$ and $\mu_0=0$ are comparable. Since the former leads to much simpler expansion formulas (see Example A.2), from now on we take $\mu_0=\mu(t,x;\theta)$ for all other models. Moreover, the delta expansion is more accurate than the Itô-Taylor expansion. Therefore, we use the delta expansion to demonstrate the accuracy of our density approximation.

Fig. 2 displays the approximation errors of the delta expansion under the four models whose true transition densities are known. Two general patterns arise in the experiment outcomes. First, for a fixed number of terms K, the error of our density

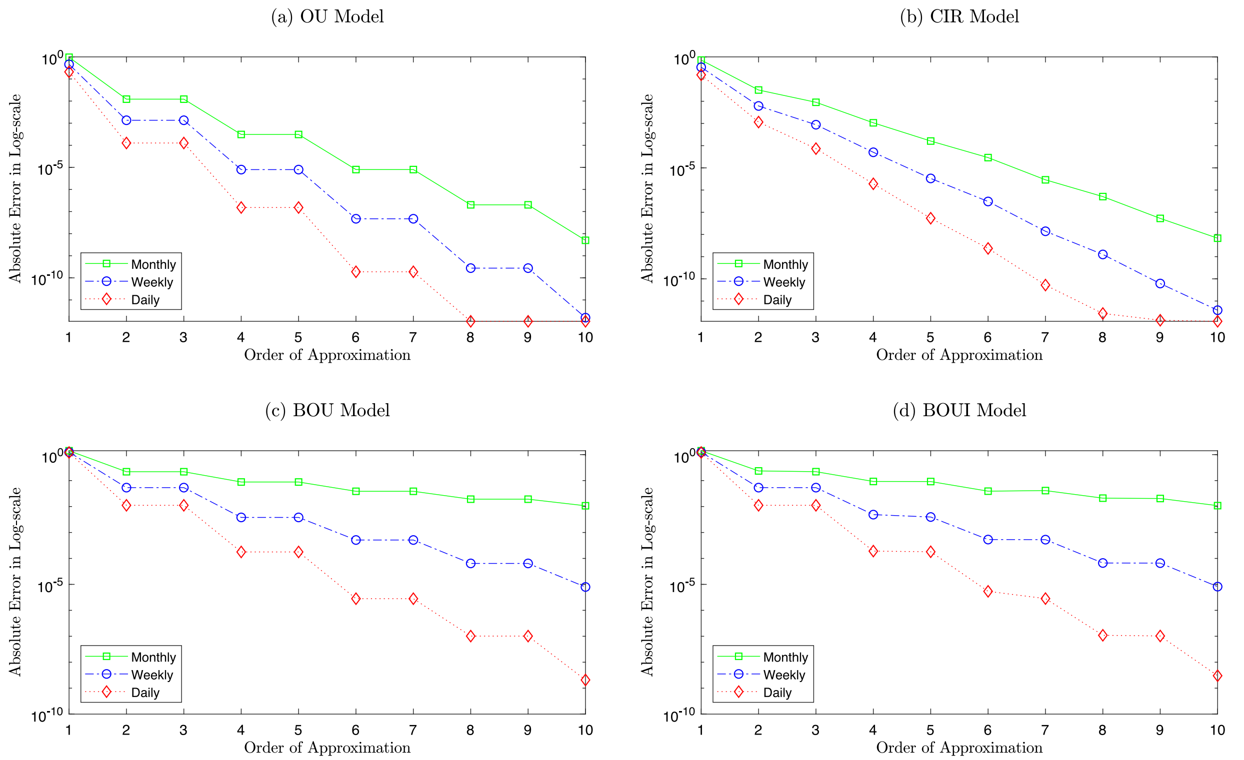

Fig. 2. Maximum absolute error between true and approximate densities. *Notes*: The maximum absolute error between the true densities p and the delta expansion $p^{(K,\Delta)}$ in (27) is similar to that defined in Fig. 1. All formulas take the advantage of the reducibility. We use the following values 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: $(\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, \beta_1, \beta_2, t) = (0, 0, 5, 1, 10, 0.1, 0.1, 0)$.

approximation decreases as the observational time interval Δ shrinks. When we change Δ from 1/12 to 1/252, i.e., the observation frequency from monthly to daily, the maximum absolute error of the delta expansions reduced very significantly. Second, when we fix the observation frequency Δ, the expansion with a larger K will lead to a smaller approximation error. Both patterns corroborate the theoretical statements in Eq. (29).

Through Fig. 3, we intend to compare the maximum absolute error of the delta expansion proposed in this paper with those ones given by Aït-Sahalia (2008) and Choi (2013). Based on our theoretic results in Eq. (29), we should use a 6th order delta expansion to compare with the 3rd order density expansion of Aït-Sahalia (2008),[7] which are both of order $\mathcal{O}(\Delta^3)$. In this sense, the performance of Aït-Sahalia (2008) is better than our delta expansion. To get a similar accuracy to the 3rd order density expansion of Aït-Sahalia (2008), the delta expansion needs about 8 terms. It is worth to mention that, Aït-Sahalia (2008) applies the reducible Kolmogorov method to the OU, CIR and BOU models, which are time-homogeneous reducible models. The reducible Kolmogorov method only involves a single series expansion in the time variable, which is more accurate than that of the irreducible Kolmogorov method involving a double series expansion in the time and state variables. For the time-inhomogeneous BOU model, our 4th order delta expansion is better that the 2nd order density expansion of the reducible Kolmogorov method of Choi (2013),[8] which theoretically are both of order $\mathcal{O}(\Delta^2)$.

6.3. Monte Carlo evidences for the approximate MLE

In this subsection, we shall provide Monte Carlo evidences for the BOU, BOUI, SEV-ND, Heston, GARCH, SVCEV, and EBDFS models to investigate the performance of the approximate MLE resulting from the delta expansion.

Recall that n is the number of observations in each path. In light of the following decomposition

$$\hat{\theta}_n^{(K,\Delta)} - \theta_0 = (\hat{\theta}_n^{(K,\delta)} - \hat{\theta}_n) + (\hat{\theta}_n - \theta_0),$$

[7] 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/.

[8] We thank Seungmoon Choi for sharing with us his 2nd order density approximate formulas for the BOUI model.

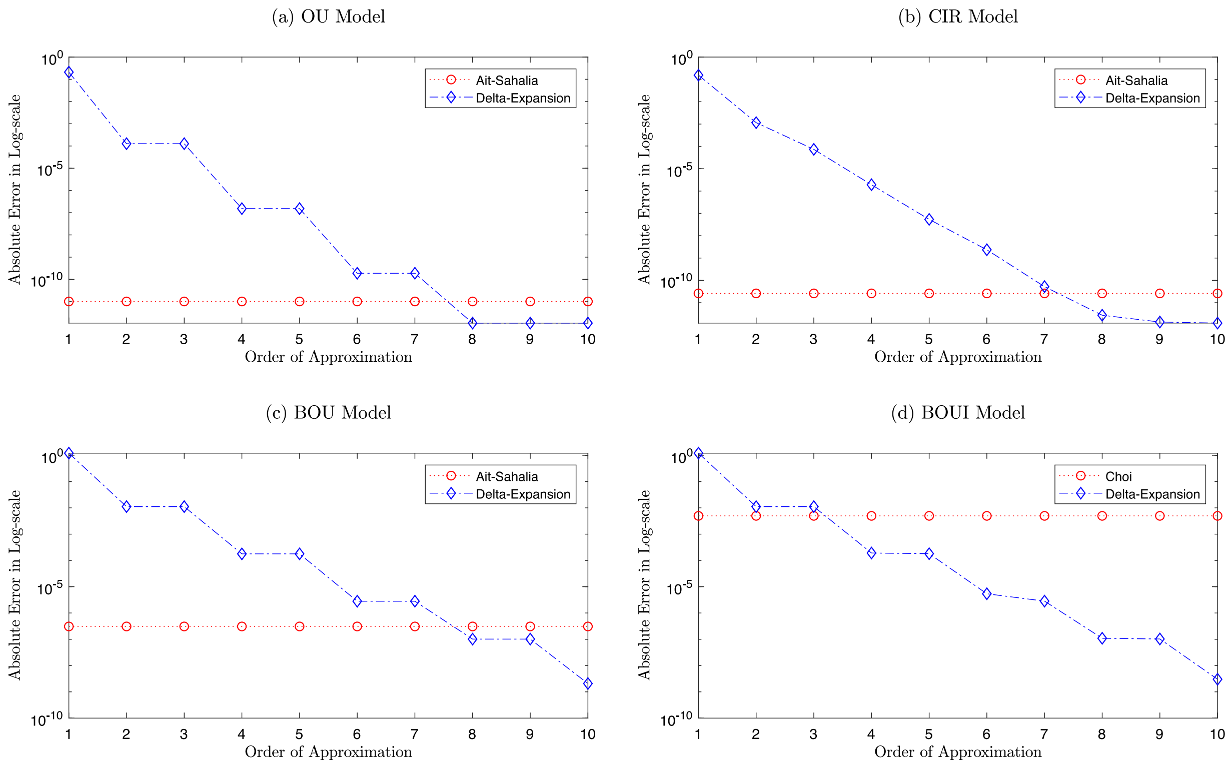

Fig. 3. Comparison of maximum absolute error between different expansions. *Notes*: This picture compares the maximum absolute error of our delta expansion and that of the 3rd order approximate formula of Aït-Sahalia (2008) for the OU, CIR and BOU models and the 2nd order approximate formulas of Choi (2013) for the BOUI model as the benchmarks. The formulas of Aït-Sahalia (2008) and Choi (2013) are derived through their reducible methods. The definition of maximum absolute error is similar to that defined in Fig. 1. The values of parameters in these four models are given in the following: OU: $(\kappa, \alpha, \sigma, \Delta) = (0.5, 0.06, 0.03, 1/252)$; CIR: $(\kappa, \alpha, \sigma, \Delta) = (0.5, 0.06, 0.15, 1/252)$; BOU: $(\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, \Delta) = (0, 0, 5, 1, 10, 1/252)$; BOUI: $(\alpha_1, \alpha_2, \kappa_{11}, \kappa_{21}, \kappa_{22}, \beta_1, \beta_2, \Delta, t) = (0, 0, 5, 1, 10, 0.1, 0.1, 1/252, 0)$.

we identify two sources of errors contributing to the estimation error of our delta expansion (i.e., (22) with $\mu_0 = \mu(t, x; \theta)$) based MLE since the true densities are explicitly known for the BOU and BOUI models. 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 $\hat{\theta}_n^{(K,\Delta)} - \hat{\theta}_n$, the difference between our approximate MLE based on the delta expansion and the true MLE, decreases rapidly as K increases, whereas we take $\mu_0 = \mu(t, x; \theta)$ in the delta expansion since it simplifies the expansion formulas. 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. Similar patterns arise for the estimator $\hat{\theta}_n^{(J)}$ based on the Itô-Taylor expansion. Moreover, with the same order of accuracy (i.e. $K = J$), the performance of $\hat{\theta}_n^{(K,\Delta)}$ is better than that of $\hat{\theta}_n^{(J)}$. Therefore, for the purpose of estimating θ_0, our delta expansion based estimator $\hat{\theta}_n^{(K,\Delta)}$ with a relatively small order (i.e. $K = 4$) can be used as a meaningful substitute for the (generally incomputable) MLE $\hat{\theta}_n$. From now on, we focus on the performance of $\hat{\theta}_n^{(K,\Delta)}$ for different models. We also report the estimation results from the 3rd order expansion of Aït-Sahalia (2008) in Table 1. Note that the BOU process is an example of the reducible multivariate diffusions, amenable to the Lamperti transform. Similarly to the density approximation, to be comparable with his 3rd order estimator, we need to construct the estimator using 8 terms in the delta expansion.

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 delta expansion. As we can see, the accuracy of the MLE yielded by the 4th order delta expansion is comparable to that of the approximate MLE using the 2nd order Choi's expansion, while both expansions are theoretically of order $\mathcal{O}(\Delta^2)$. Like what Table 1 reveals, the approximation error caused by the delta 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 3 compares the performance of a variety of MLEs under the SEV-ND model, including the approximate MLE developed in Aït-Sahalia (2008), and the approximate MLE based on our delta expansion. We can see that the accuracy of the MLE yielded by the 2nd order delta expansion is comparable to that of the approximate MLE using the 1st order Aït-Sahalia's expansion, while both expansions are theoretically of order $\mathcal{O}(\Delta)$. The biases from the approximate MLEs obtained from two methods are both higher order of the parameter values.

Table 1
Monte Carlo evidence for the BOU model with different order of expansion.

θ_0	$\hat{\theta}_n - \theta_0$	$\hat{\theta}_n^{(AS,3)} - \hat{\theta}_n$	$\hat{\theta}_n^{(4,\Delta)} - \hat{\theta}_n$	$\hat{\theta}_n^{(6,\Delta)} - \hat{\theta}_n$	$\hat{\theta}_n^{(8,\Delta)} - \hat{\theta}_n$	$\hat{\theta}_n^{(4)} - \hat{\theta}_n$	$\hat{\theta}_n^{(6)} - \hat{\theta}_n$	$\hat{\theta}_n^{(8)} - \hat{\theta}_n$
$\kappa_{11} = 5$	0.43	−0.0019	0.0032	−0.00032	0.00044	−0.047	−0.0022	−0.00016
	(1.13)	(0.045)	(0.083)	(0.077)	(0.065)	(0.24)	(0.059)	(0.046)
$\kappa_{21} = 1$	0.020	−0.0010	0.0051	−0.0031	−0.0029	−0.021	−0.0015	−0.00069
	(1.19)	(0.029)	(0.10)	(0.065)	(0.044)	(0.24)	(0.035)	(0.035)
$\kappa_{22} = 10$	0.62	−0.0033	−0.0038	0.039	0.000039	−0.19	−0.0076	−0.0014
	(1.55)	(0.030)	(0.26)	(0.13)	(0.061)	(0.50)	(0.077)	(0.074)
$\alpha_1 = 0$	0.0049	−0.00094	−0.0012	−0.0012	−0.00065	−0.0016	−0.00075	−0.0001
	(0.063)	(0.015)	(0.020)	(0.018)	(0.017)	(0.023)	(0.019)	(0.016)
$\alpha_2 = 0$	−0.00033	0.000080	0.000043	−0.000029	−0.000031	−0.000027	−0.000027	−0.000083
	(0.034)	(0.0044)	(0.0064)	(0.0058)	(0.0053)	(0.0077)	(0.0056)	(0.0060)

Notes: We use θ_0 to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., $\Delta = 1/52$). The first column reports true parameter values θ_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 $\hat{\theta}_n^{(AS,3)}$ developed by Aït-Sahalia (2008). The remaining columns report the differences between true maximum likelihood estimator $\hat{\theta}_n$ and estimators $\hat{\theta}_n^{(K,\Delta)}$, $\hat{\theta}_n^{(J)}$ developed in this paper, with the standard derivation in parentheses. The estimators $\hat{\theta}_n^{(K,\Delta)}$ and $\hat{\theta}_n^{(J)}$ are based on the delta expansion in (27) and the Itô-Taylor expansion in (26), respectively, by taking $\mu_0 = \mu(t, x; \theta)$. The order of our delta and Itô-Taylor expansions takes values from $K, J = 4, 6, 8$.

Table 2
Monte Carlo evidence for the BOUI model.

θ_0	$\hat{\theta}_n - \theta_0$	$\hat{\theta}_n^{(Choi,2)} - \hat{\theta}_n$	$\hat{\theta}_n^{(4,\Delta)} - \hat{\theta}_n$
$\kappa_{11} = 5$	0.56	0.088	0.011
	(1.17)	(0.29)	(0.23)
$\kappa_{21} = 1$	0.077	−1.04	0.000043
	(1.21)	(1.93)	(0.17)
$\kappa_{22} = 10$	0.66	−0.36	0.016
	(1.55)	(0.40)	(0.32)
$\alpha_1 = 0$	−0.00038	0.0022	0.0067
	(0.091)	(0.075)	(0.076)
$\alpha_2 = 0$	0.0029	0.0029	0.00046
	(0.061)	(0.035)	(0.034)
$\beta_1 = 0.1$	0.00040	−0.000068	−0.00073
	(0.019)	(0.012)	(0.012)
$\beta_2 = 0.1$	−0.00080	0.00021	−0.000064
	(0.011)	(0.0055)	(0.0053)

Notes: We use θ_0 to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., we take $\Delta = 1/52$). The first column reports true parameter values θ_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 $\hat{\theta}_n^{(Choi,2)}$ developed by Choi (2013) in the third column. The fourth column shows the differences between $\hat{\theta}_n$ and the 4th order approximate estimator $\hat{\theta}_n^{(4,\Delta)}$ using our delta expansion in (27) with $\mu_0 = \mu(t, x; \theta)$. All standard deviations are reported in the parentheses.

Table 3
Monte Carlo evidence for the SEV-ND model.

θ_0	$\hat{\theta}_n^{(2,\Delta)} - \hat{\theta}_n$		$\hat{\theta}_n^{(AS,1)} - \hat{\theta}_n$	
	Mean	Stdev	Mean	Stdev
$\alpha_0 = -4.643 \times 10^{-3}$	7.2×10^{-8}	7.0×10^{-5}	-3.4×10^{-5}	9.8×10^{-5}
$\alpha_1 = 4.333 \times 10^{-2}$	1.5×10^{-5}	5.5×10^{-4}	3.5×10^{-4}	6.6×10^{-4}
$\alpha_2 = -1.143 \times 10^{-1}$	-1.1×10^{-5}	4.6×10^{-4}	-6.6×10^{-5}	4.8×10^{-4}
$\alpha_3 = -1.304 \times 10^{-4}$	2.8×10^{-7}	1.0×10^{-6}	5.3×10^{-8}	2.6×10^{-7}
$\beta_0 = 1.108 \times 10^{-4}$	2.2×10^{-7}	1.1×10^{-6}	4.8×10^{-8}	5.3×10^{-7}
$\beta_1 = -1.883 \times 10^{-3}$	-3.3×10^{-5}	4.7×10^{-5}	1.6×10^{-6}	8.1×10^{-6}
$\beta_2 = 9.681 \times 10^{-3}$	2.1×10^{-5}	1.6×10^{-4}	-1.1×10^{-6}	4.2×10^{-5}
$\beta_3 = 2.073$	0.052	0.051	0.069	0.043

Notes: We use θ_0 to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., we take $\Delta = 1/52$). We display the differences between the true value and the 2nd order approximate estimator $\hat{\theta}_n^{(\Delta,2)}$ obtained from the delta expansion in (27) with $\mu_0 = \mu(t, x; \theta)$, and the differences between the true value and the 1st order approximate estimator $\hat{\theta}_n^{(AS,1)}$.

We consider the Heston, GARCH, and SVCEV models in Table 4, demonstrating the performance of our method for irreducible and nonaffine diffusion processes. None of these models admits closed-form expressions for their transition densities. Hence, we compute the difference between the true parameter values and the delta expansion based MLE to assess the method's accuracy. Aït-Sahalia (2008) expands the log-likelihood functions of such processes from their accompanying

Table 4
Monte Carlo evidence for Heston, GARCH, and SVCEV models.

θ_0	Heston model ($\beta = 1/2$)		GARCH model ($\beta = 1$)		SVCEV model (β)	
	$\hat{\theta}_n^{(AS,2)} - \theta_0$	$\hat{\theta}^{(4,\Delta)} - \theta_0$	$\hat{\theta}_n^{(AS,2)} - \theta_0$	$\hat{\theta}^{(4,\Delta)} - \theta_0$	$\hat{\theta}_n^{(AS,2)} - \theta_0$	$\hat{\theta}^{(4,\Delta)} - \theta_0$
$\sigma = 0.25$	0.00064	0.00061	0.00060	0.00060	0.015	0.017
	(0.0064)	(0.0064)	(0.0064)	(0.0064)	(0.099)	(0.10)
$\rho = -0.8$	−0.00075	−0.00070	−0.00069	−0.00070	−0.0010	−0.00099
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
$\alpha = 0.1$	0.00024	0.00024	−0.000033	0.0000045	0.000072	0.00010
	(0.0074)	(0.0074)	(0.0023)	(0.0023)	(0.0031)	(0.0031)
$\mu = 0.03$	−0.0028	−0.0028	0.0022	0.00051	−0.0017	−0.0031
	(0.082)	(0.081)	(0.079)	(0.081)	(0.053)	(0.051)
$\kappa = 3$	0.15	0.16	0.15	0.16	0.14	0.14
	(0.54)	(0.54)	(0.54)	(0.54)	(0.54)	(0.54)
$\beta = 0.8$	–	–	–	–	−0.0025	0.00077
	–	–	–	–	(0.15)	(0.15)

Notes: We use θ_0 to generate 1000 sample paths. Each of them contains 500 weekly observations (i.e., $\Delta = 1/52$). The three subtables correspond to the results from the Heston, GARCH, and SVCEV models, respectively. The first column of each subtable displays the bias of the 2nd order approximate estimator $\hat{\theta}_n^{(AS,2)}$ developed by Aït-Sahalia (2008); the second column illustrates the bias of the 4th order approximate estimator $\hat{\theta}_n^{(4,\Delta)}$ using our delta expansion in (22) with $\mu_0 = \mu(t, x; \theta)$. All standard deviations are reported in the parentheses.

Table 5
Monte Carlo evidence for the EBDFS model.

θ_0	500 weekly observations		500 daily observations		5000 daily observations	
	$\hat{\theta}_n^{(2,\Delta)} - \theta_0$	$\hat{\theta}_n^{(Choi,1)} - \theta_0$	$\hat{\theta}_n^{(2,\Delta)} - \theta_0$	$\hat{\theta}_n^{(Choi,1)} - \theta_0$	$\hat{\theta}_n^{(2,\Delta)} - \theta_0$	$\hat{\theta}_n^{(Choi,1)} - \theta_0$
$k_1 = 10$	−0.69	1.38	0.85	5.18	−0.25	1.54
	(1.07)	(2.19)	(2.92)	(12.52)	(0.99)	(1.47)
$k_2 = 7$	0.062	0.64	3.33	3.56	0.70	0.85
	(1.12)	(1.43)	(3.70)	(3.89)	(0.50)	(0.53)
$k_3 = 3$	0.21	0.20	1.76	2.38	0.071	0.013
	(0.71)	(0.85)	(2.38)	(1.95)	(0.35)	(0.41)
$\alpha_2 = 0.06$	0.0081	0.0096	0.0020	0.0019	0.0010	0.0010
	(0.0044)	(0.0029)	(0.0061)	(0.0074)	(0.0023)	(0.0023)
$\beta_2 = 0.001$	−0.00084	−0.0010	−0.00096	−0.0010	−0.00095	−0.00095
	(0.00060)	(0.00052)	(0.0053)	(0.0050)	(0.00020)	(0.00020)
$\alpha_3 = 0.1$	0.0000087	0.0029	−0.000041	0.0045	0.00027	0.0041
	(0.0014)	(0.0019)	(0.0099)	(0.020)	(0.0.0010)	(0.0016)
$\sigma_{21} = 0.03$	0.00032	0.00032	0.00019	0.00020	−0.00028	−0.00014
	(0.0012)	(0.0013)	(0.0011)	(0.0011)	(0.00029)	(0.00030)
$\sigma_{22} = 0.001$	0.000097	0.000039	−0.000057	−0.00016	0.00047	−0.000028
	(0.0049)	(0.0050)	(0.015)	(0.015)	(0.00031)	(0.0012)
$\sigma_{31} = 0.05$	0.000064	0.00012	0.000013	0.000037	0.00036	0.00039
	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.00029)	(0.00029)
$\rho = 0.5$	−0.0023	0.0049	−0.0010	0.0031	0.0016	0.0039
	(0.031)	(0.034)	(0.032)	(0.036)	(0.0084)	(0.0088)

Notes: We use θ_0 to generate 1000 sample paths with weekly observations (i.e., $\Delta = 1/52$). The three subtables correspond to the results from 500 weekly, as well as 500 and 5000 daily observations, respectively. The first column of each subtable illustrates the bias of the 2nd order approximate estimator $\hat{\theta}_n^{(2,\Delta)}$ using our delta expansion in (22) with $\mu_0 = \mu(t, x; \theta)$. The second column displays the difference between the true values and the 1st order approximate estimator $\hat{\theta}_n^{(Choi,1)}$ developed by Choi (2013). All standard deviations are reported in the parentheses.

Kolmogorov equations to develop approximate MLEs. We also include the estimation results from his method in Table 4 for the purpose of comparison.[9] The biases of our estimators from the 4th order delta expansion, which is of order $\mathcal{O}(\Delta^2)$, 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), which is of order $\mathcal{O}(\Delta^2)$.

Table 5 illustrates the performance of our estimator for the EBDFS, which is a three-dimensional irreducible, time-inhomogeneous diffusion. Since an explicit-form transition density for the EBDFS model does not exist, we compute the difference between the true parameter values and the 2nd order delta expansion based MLE. We also report the estimation results from the 1st order expansion of Choi (2013).[10] Both expansions are theoretically as accurate as $\mathcal{O}(\Delta)$. Overall, the performances of two estimators are comparable. As noted by Choi (2013), both the biases and standard errors are small relative to the true values, and have declined significantly as the monitored dates increase from 500 to 5000.

9 The 2nd order density approximation formulas for the Heston, GARCH, and SVCEV models are also available on Yacine Aït-Sahalia's website.

10 We thank Seungmoon Choi for sharing with us his 1st order approximation formulas for the EBDFS.

Table 6
Comparison between asymptotic and finite-sample standard deviation for the BOU model.

θ_0	500 observations		2000 observations		5000 observations	
	ASD	FSSD	ASD	FSSD	ASD	FSSD
$\kappa_{11} = 5$	1.02	1.11	0.51	0.52	0.32	0.34
$\kappa_{21} = 1$	1.08	1.19	0.54	0.55	0.34	0.35
$\kappa_{22} = 10$	1.45	1.51	0.73	0.73	0.46	0.44
$\alpha_1 = 0$	0.065	0.061	0.032	0.030	0.020	0.019
$\alpha_2 = 0$	0.033	0.033	0.016	0.017	0.010	0.011

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 $\Delta = 1/52$. All the results for FSSD are based on the 4th order approximate estimator $\hat{\theta}_n^{(4,\Delta)}$ using our delta expansion in (27) with $\mu_0 = \mu(t, x; \theta)$.

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,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, i(\theta_0)^{-1}),$$

as $n \to \infty$ with Δ 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 6, 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.

We carry out the same experiments for some other non-stationary diffusions. The results in Table 7 show that the standard deviations of our estimators for the BOUI, Heston, GARCH, and SVCEV models all decrease at a rate of $\sqrt{n}$, similar to that in the stationary case.

7. Conclusion

This paper constructs a closed-form delta expansion of the transition densities for time-inhomogeneous irreducible multivariate diffusions. The explicit recursive formulas for the expansion coefficients in the method enables us to easily compute the approximation through some symbolic computing softwares. We manage to prove that our expansions will converge to the true density under a set of very mild technical conditions. Numerical experiments for wide-ranging models illustrate the efficiency and accuracy of the expansions of the density and the resulted approximate MLE. In addition, we build explicit connections in the case of time-homogeneous diffusions between our delta expansion under a specially chosen parameter and some existing methods such as Aït-Sahalia (2002) and Lee et al. (2014).

Acknowledgments

We thank the co-editor Yacine Aït-Sahalia and two anonymous referees for their substantial comments. We are grateful to Dacheng Xiu and George J. Jiang for their valuable suggestions. The first and third authors thank the National Natural Science Foundation of China (NSFC) for the financial support under Grant No. 71271110 and Grant Nos. 71401103 and 71421002, respectively. Part of the research of the second author is supported by Hong Kong Research Grant Council under GRF Grant No. CUHK14201114. The authors are also grateful to the financial support from Center for Financial Engineering, The Chinese University of Hong Kong.

Appendix A. The proofs

A.1. Technical lemmas and examples

Lemma A.1. *Fix (t', x') and (t, x).*[11] *For $N \geq 1$,*

$$(\partial_s + \mathcal{L})^N q(t', x'|s, y) = \sum_{|h|=1}^{2N} w_{N,h}(s, y)\partial_h q(t', x'|s, y), \tag{39}$$

[11] Thus, $\nu_0 = \nu(t, x; \theta)$ is also fixed.

Table 7
Finite-sample standard deviation for the BOUI, Heston, GARCH, SVCEV models with different number of observations.

θ_0	500 observations	2000 observations	5000 observations
	BOUI		
$\kappa_{11} = 5$	1.18	0.52	0.34
$\kappa_{21} = 1$	1.22	0.56	0.33
$\kappa_{22} = 10$	1.55	0.73	0.44
$\alpha_1 = 0$	0.093	0.039	0.025
$\alpha_2 = 0$	0.063	0.029	0.017
$\beta_1 = 0$	0.018	0.0021	0.00052
$\beta_2 = 0$	0.012	0.0013	0.00033
	Heston		
$\sigma = 0.25$	0.0065	0.0033	0.0028
$\rho = -0.8$	0.013	0.0066	0.0042
$\alpha = 0.1$	0.0076	0.0033	0.0020
$\mu = 0.03$	0.087	0.032	0.020
$\kappa = 3$	0.50	0.24	0.19
	GARCH		
$\sigma = 0.25$	0.0065	0.0031	0.0022
$\rho = -0.8$	0.013	0.0064	0.0040
$\alpha = 0.1$	0.0022	0.0011	0.00067
$\mu = 0.03$	0.078	0.036	0.020
$\kappa = 3$	0.51	0.24	0.16
	SVCEV		
$\sigma = 0.25$	0.088	0.043	0.029
$\rho = -0.8$	0.013	0.0065	0.0041
$\alpha = 0.1$	0.0032	0.0017	0.0010
$\mu = 0.03$	0.054	0.032	0.018
$\kappa = 3$	0.52	0.24	0.18
$\beta = 0.8$	0.14	0.074	0.053

Notes: This table presents the finite-sample standard deviation for the BOUI, Heston, GARCH, and SVCEV 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 $\Delta = 1/52$. The results for the BOUI and the rest are based on $\hat{\theta}_n^{(4,\Delta)}$ and $\hat{\theta}_n^{(2,\Delta)}$ using our delta expansion in (22) with $\mu_0 = \mu(t, x; \theta)$, respectively.

where the operator $\mathcal{L}$ is defined by (5) acting on the state variable y, the multivariate normal density $q(t', x'|s, y)$ is defined by (17), and the coefficient function $w_{N,h}(s, y)$ is defined by (18) and *(19).*

Proof of Lemma A.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

$$\partial_s q = -\mathcal{L}_0 q, \tag{40}$$

where $\mathcal{L}_0$ is a differential operator acting on y defined by

$$\mathcal{L}_0 = \sum_{i=1}^{m} \mu_{0i} \partial_{e_i} + \frac{1}{2} \sum_{i,j=1}^{m} \nu_{0ij} \partial_{e_i+e_j}.$$

Then we have

$$(\partial_s + \mathcal{L})q = (-\mathcal{L}_0 + \mathcal{L})q = \sum_{i=1}^{m} (\mu_i - \mu_{0i}) \partial_{e_i} q + \frac{1}{2} \sum_{i,j=1}^{m} (\nu_{ij} - \nu_{0ij}) \partial_{e_i+e_j} q.$$

Recalling the definition of $w_{N,h}$ in (18), (39) holds for $N = 1$.

Next, assume that (39) holds for N. Then, for $N + 1$, we have

$$(\partial_s + \mathcal{L})^{N+1} q = (\partial_s + \mathcal{L}) \Big(\sum_{|h|=1}^{2N} w_{N,h} \partial_h q \Big) = \sum_{|h|=1}^{2N} \Big(\partial_s (w_{N,h} \partial_h q) + \mathcal{L}(w_{N,h} \partial_h q) \Big). \tag{41}$$

Applying (40) to the first term gives that

$$\begin{aligned} \partial_s (w_{N,h} \partial_h q) &= (\partial_s w_{N,h}) \partial_h q + w_{N,h} (\partial_s \partial_h q) = (\partial_s w_{N,h}) \partial_h q + w_{N,h} (\partial_h \partial_s q) \\ &= (\partial_s w_{N,h}) \partial_h q - w_{N,h} (\partial_h \mathcal{L}_0 q) = (\partial_s w_{N,h}) \partial_h q - w_{N,h} \mathcal{L}_0 (\partial_h q). \end{aligned}$$

The second term on the right hand side of (41) follows

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

Putting them together, we have

$$(\partial_s + \mathcal{L})^{N+1} q = \sum_{|h|=1}^{2N}((\partial_s + \mathcal{L})w_{N,h})\partial_h q + \sum_{|h|=1}^{2N}\sum_{i=1}^m \mathcal{A}_i w_{N-1,h-e_i}\partial_{h+e_i} q + \sum_{|h|=1}^{2N}\frac{1}{2}\sum_{i,j=1}^m w_{N,h}\cdot(\nu_{ij} - \nu_{0ij})\partial_{h+e_i+e_j} q.$$

Rewriting the index in the summation gives that

$$(\partial_s + \mathcal{L})^{N+1} q = \sum_{|h|=1}^{2N}\Big((\partial_s + \mathcal{L})w_{N,h}\Big)\partial_h q + \sum_{|h|=2}^{2N+1}\sum_{i=1}^m \mathcal{A}_i w_{N-1,h-e_i}\partial_h q + \sum_{|h|=3}^{2N+2}\frac{1}{2}\sum_{i,j=1}^m (\nu_{ij} - \nu_{0ij}) w_{N,h-e_i-e_j}\partial_h q$$
$$= \sum_{|h|=1}^{2N+2} w_{N+1,h}\partial_h q.$$

The last equality holds by the definition of $w_{N+1,h}$ in (19) 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, (39) holds for $N+1$. □

Lemma A.2. *Fix t and x. Recall that the series of functions $\{w_{N,h}(s,y) : s \geq 0, y \in D\}$ is defined by the recursive relations (i)–(iii). The following statement holds: for each $N \geq 1$,*

$$w_{N,h}(s,y)|_{s=t,y=x} = 0, \quad \text{if} \quad 3N/2 < |h| \leq 2N. \tag{42}$$

Moreover, if $\nu(t,x;\theta) \equiv Id_m$, then for each $N \geq 1$,

$$w_{N,h}(s,y) = 0, \quad \text{if} \quad N < |h| \leq 2N. \tag{43}$$

Proof of Lemma A.2. The statement (42) obviously holds for $N = 1$. Indeed, for $N = 1$, $3N/2 < |h| \leq 2N$ implies that $|h| = 2$. Thus by (18), we have

$$\begin{cases} w_{1,2e_i}(t,x) = \frac{1}{2}\Big(\nu_{ii}(s,y;\theta)|_{s=t,y=x} - \nu_{ii}(t,x;\theta)\Big) = 0, & i = 1,\ldots,m; \\ w_{1,e_i+e_j}(t,x) = \nu_{ij}(s,y;\theta)|_{s=t,y=x} - \nu_{ij}(t,x;\theta) = 0, & i \neq j,\ i,j = 1,\ldots,m. \end{cases} \tag{44}$$

Inspired by (44), by Taylor's Theorem and Assumption 2.2, we can expand the function $w_{N,h}(s,y)$ at (t,x) up to the Kth order as follows:

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

where a_0 is a nonnegative integer number, $a = (a_1,\ldots,a_m)$ 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 $\xi^{N,h}_{a_0,a} := \partial_s^{a_0}\partial_a w_{N,h}(s,y)|_{s=t,y=x}$ is a constant, and the function $\Omega^{N,h}_{a_0,a}(s,y)$ satisfies $\lim_{s\to t, y\to x}\Omega^{N,h}_{a_0,a}(s,y) = 0$. Thus, we can define the *Order* of the function $w_{N,h}(s,y)$ at (t,x) as follows

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

Obviously, $w_{N,h}(s,y)|_{s=t,y=x} = 0$ if $Or(w_{N,h}) \geq 1$. Then, we can prove the statement (42) by showing a stronger statement below: for each $N \geq 1$

$$Or(w_{N,h}) \geq 2|h| - 3N, \quad \text{if } 3N/2 < |h| \leq 2N. \tag{45}$$

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

$$w_{N,h}(s,y) = \sum_{a_0+|a|=2|h|-3N} \big(\xi^{N,h}_{a_0,a} + \Omega^{N,h}_{a_0,a}(s,y)\big)\cdot(s-t)^{a_0}(y-x)^a. \tag{46}$$

We use mathematical induction to verify that (45) holds. For $N = 1$, (45) holds by (44). Assume that (45) 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 Eq. (19), 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 (45) holds for $N+1$. Thus, we have proved that (42) holds.

If $\nu(t,x;\theta)\equiv Id_m$, we directly prove that (43) holds using the mathematical induction. For $N=1$, (18) holds by (44). Assume that (43) holds for $N\leq 1$, that is, $w_{N,h}(s,y)=0$ for $N<|h|\leq 2N$. Thus the derivatives of $w_{N,h}(s,y)$ are all zeros $N<|h|\leq 2N$. Then by (19), for $|h|>N+1$

$$w_{N+1,h}(s,y)=(\partial_s+\mathcal{L})w_{N-1,h}(s,y)+\sum_{i=1}^{m}\mathcal{A}_i w_{N,h-e_i}(s,y)=0. \tag{47}$$

Therefore, (43) holds for $N+1$. □

Example A.1 (*The CIR Model Without Lamperti Transform*)**.** Consider the CIR Model 2 without Lamperti transform. We first verify the result (42) in *Lemma* A.2 for $N=1,2,3,4$. By a direct computation (cf. Eqs. (18) and (19)), the coefficients $\{w_{N,h}(s,y),3N/2<h\leq 2N\}$ are given by

$$w_{1,2}(s,y)=\frac{1}{2}\sigma^2(y-x);$$
$$w_{2,4}(s,y)=\frac{1}{4}\sigma^4(y-x)^2;$$
$$w_{3,5}(s,y)=\frac{3}{4}\sigma^4(y-x)\Big(\big(\kappa(\alpha-x)-\mu_0\big)(y-x)-\sigma^2x\Big);$$
$$w_{3,6}(s,y)=\frac{1}{8}\sigma^6(y-x)^3;$$
$$w_{4,7}(s,y)=\frac{1}{4}\sigma^6(y-x)^2\Big(2\big(\kappa(\alpha-x)-\mu_0\big)(y-x)+3\sigma^2x\Big);$$
$$w_{4,8}(s,y)=\frac{1}{16}\sigma^8(y-x)^4.$$

It is obvious that all above are zeros when $y=x$.

Moreover, if we take $\mu_0=\mu(t,x)$, the first several terms in the delta expansion (22) are given below:

$$C_1=\frac{1}{4\sigma^2x^2}\left(z^3-3\sigma^2xz\right);$$
$$C_2=\frac{1}{32\sigma^4x^4}\Big(\sigma^4x^3\left(8\alpha\kappa-3\sigma^2-24\kappa x\right)+\sigma^2x^2z^2\left(-8\alpha\kappa+21\sigma^2+24\kappa x\right)-11\sigma^2xz^4+z^6\Big);$$
$$C_3=\frac{1}{384\sigma^6x^6}\Big(z\left(-16\sigma^4x^3z^2\left(-13\alpha\kappa+15\sigma^2+33\kappa x\right)+6\sigma^2x^2z^4\left(-4\alpha\kappa+25\sigma^2+12\kappa x\right)\right.$$
$$\left.+3\sigma^4x^4\left(-88\alpha\kappa\sigma^2+15\sigma^4+64\kappa^2x^2-8\kappa x\left(8\alpha\kappa-21\sigma^2\right)\right)-24\sigma^2xz^6+z^8\right)\Big),$$

where $z=(x'-x-\mu_0\Delta)/\sqrt{\Delta}$ and $\mu_0=\kappa(\alpha-x)$.

Example A.2 (*The CIR Model after Lamperti Transform*)**.** The CIR Model (2) is reducible. Thus, we first perform a Lamperti transform to obtain a simplified process $Y(t)=2\sqrt{X(t)}/\sigma$ satisfying

$$dY(t)=\left(\frac{\lambda}{Y(t)}-\frac{\kappa}{2}Y(t)\right)dt+dW(t),\qquad \lambda=\frac{4\kappa\alpha-\sigma^2}{2\sigma^2}. \tag{48}$$

Then, we find the approximations formulas for (48) using the delta expansion (27) with different μ_0.

(a) The delta expansion when $\mu_0=0$.

$$C_1=z\left(\frac{\lambda}{y}-\frac{\kappa y}{2}\right);$$
$$C_2=\frac{1}{2}\left(z^2-1\right)\left(-\frac{\kappa}{2}-\frac{\lambda}{y^2}+\left(\frac{\lambda}{y}-\frac{\kappa y}{2}\right)^2\right);$$
$$C_3=\frac{1}{48y^3}z\Big(\kappa^3y^6\left(-\left(z^2-3\right)\right)+6\kappa^2y^4\left(-3\lambda+(\lambda+1)z^2-2\right)-12\kappa\lambda^2y^2\left(z^2-3\right)$$
$$+8(\lambda-1)\lambda\left(-3\lambda+(\lambda-2)z^2+3\right)\Big);$$

where $z=(y'-y)/\sqrt{\Delta}$.

(b) The delta expansion when $\mu_0 = \mu(t, y)$.

$$
\begin{aligned}
C_1 &= 0;\\
C_2 &= -\frac{\left(z^2-1\right)\left(2\lambda+\kappa y^2\right)}{4y^2};\\
C_3 &= \frac{z\left(3\kappa^2 y^4-4\lambda\left(3\lambda-2z^2+3\right)\right)}{24y^3};
\end{aligned}
$$

where $z = (y' - y - \mu_0\Delta)/\sqrt{\Delta}$ and $\mu_0 = \lambda/y - \kappa y/2$.

Comparing the formulas in (a) with that in (b), we can find that the expansion formulas are significantly simplified when we take $\mu_0 = \mu(t, y)$. In addition, comparing the terms in (b) with that presented in Example A.1, we can see that the expansion formulas are much simpler after performing the Lamperti transform.

A.2. The proofs

Proof of Theorem 3.1. Fix (t, x) and (t', x'). For any $s \in [t, t')$ and $y \in D$, by the definition in (10), we have

$$
p^{(J)}(t', x'|s, y; \theta) = \sum_{N=0}^{J} \frac{(t'-s)^N}{N!}(\partial_s + \mathcal{L})^N q(t', x'|s, y), \tag{49}
$$

where the operator $\mathcal{L}$ is defined by (5) acting on state variable y. By (39) in Lemma A.1, we have

$$
p^{(J)}(t', x'|s, y; \theta) = q(t', x'|s, y) + \sum_{N=1}^{J}\sum_{|h|=1}^{2N} \frac{(t'-s)^N}{N!} w_{N,h}(s, y)\partial_h q(t', x'|s, y). \tag{50}
$$

Note that (cf. (13))

$$
\partial_h q(t', x'|s, y) = (t'-s)^{-\frac{|h|}{2}} H_h(z; \nu_0) q(t', x'|s, y) \tag{51}
$$

with $z = (x' - y - \mu_0(t'-s))/\sqrt{t'-s}$. Plugging it into (50), we have

$$
p^{(J)}(t', x'|s, y; \theta) = q(t', x'|s, y)\Bigg(1 + \sum_{N=1}^{J}\sum_{|h|=1}^{2N} \frac{(t'-s)^{N-\frac{|h|}{2}}}{N!} w_{N,h}(s, y) H_h(z; \nu_0)\Bigg). \tag{52}
$$

Taking $s = t$ and $y = x$, by (42) in Lemma A.2, we see that Eq. (15) holds. This completes the proof. □

Proof of Theorem 3.2. Firstly, fix (t, x) and μ_0, then 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)

$$
(\partial_s + \mathcal{L}) p(t', x'|s, y; \theta) = 0, \quad \lim_{t'-s\to 0} p(t', x'|s, y; \theta) = \delta(x' - y), \tag{53}
$$

where $\mathcal{L}$ is defined in (5). Applying $(\partial_s + \mathcal{L})$ to $p^{(J)}(t', x'|s, y; \theta)$ (cf. (49)), we have

$$
\begin{aligned}
(\partial_s + \mathcal{L}) p^{(J)}(t', x'|s, y; \theta) &= \sum_{N=0}^{J} (\partial_s + \mathcal{L}) \left(\frac{(t'-s)^N}{N!}(\partial_s + \mathcal{L})^N q(t', x'|s, y)\right)\\
&= \frac{(t'-s)^J}{J!}(\partial_s + \mathcal{L})^{J+1} q(t', x'|s, y) := \psi_J(s, y; t, x, t', x').
\end{aligned} \tag{54}
$$

Note that

$$
\lim_{t'-s\to 0} q(t', x'|s, y) = \delta(x' - y). \tag{55}
$$

To establish a similar initial condition for $p^{(J)}$, consider any test function $\varphi(\cdot)$, which is continuous with compact support on D. By (52),

$$\int_D \left(p^{(J)}(t', x'|s, y; \theta) - q(t', x'|s, y)\right)\varphi(x')dx'$$
$$= \sum_{N=1}^{J}\sum_{|h|=1}^{2N} \frac{(t'-s)^{N-|h|/2}}{N!} w_{N,h}(s, y)\int_D q(t', x'|s, y)H_h(z; \nu_0)\varphi(x')dx', \tag{56}$$

where $z = (x' - y - \mu_0(t'-s))/\sqrt{t'-s}$. Note that (cf. (12))

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

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

$$\int_D q(t', x'|s, y)H_h(z; \nu_0)\varphi(x')dx' = \int_{D_Z} \phi(z; \nu_0)H_h(z; \nu_0)\varphi(y + \mu_0(t'-s) + z\sqrt{t'-s})dz.$$

Then, taking $t' - s \to 0$ on both sides of (56), we have

$$\lim_{t'-s\to 0}\int_D \left(p^{(J)}(t', x'|s, y; \theta) - q(t', x'|s, y)\right)\varphi(x')dx' = \varphi(y)\sum_{N=1}^{J}\sum_{|h|=2N}\frac{1}{N!}w_{N,h}(s, y)\int_{D_Z}\phi(z; \nu_0)H_h(z; \nu_0)dz, \tag{57}$$

which is zero when $s = t$ and $y = x$, because $w_{N,h}(t, x) = 0$ for all $|h| = 2N$ (cf. (42)).

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

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

Then, by (53) and (54), it satisfies the following Kolmogorov PDE:

$$(\partial_s + \mathcal{L})r^{(J)}(t', x'|s, y; \theta) = \psi_J(s, y; t, x, t', x'). \tag{58}$$

Besides, by (53), (55) and (57), when $s = t$ and $y = x$, the initial condition becomes

$$\lim_{t'-t\to 0} r^{(J)}(t', x'|t, x; \theta) = 0. \tag{59}$$

Thus,

$$r^{(J)}(t', x'|t, x; \theta) = \int_t^{t'}\int_D \psi_J(s, y; t, x, t', x')\cdot p(s, y|t, x; \theta)dyds. \tag{60}$$

Recalling (54) and (39), we decompose ψ_J into three terms:

$$\psi_J := \psi_J^{(1)} + \psi_J^{(2)} + \psi_J^{(3)}, \tag{61}$$

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

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

and $I_1 = \left\{h\middle|1 \le |h| \le 3(J+1)/2\right\}$, $I_2 = \left\{h\middle|3(J+1)/2 < |h| < 2(J+1)\right\}$, $I_3 = \left\{h\middle||h| = 2(J+1)\right\}$. Then, by (60) and (61), we can rewrite the error term into a summation $r^{(J)} = r_1^{(J)} + r_2^{(J)} + r_3^{(J)}$, where

$$r_i^{(J)}(t', x'|t, x; \theta) = \int_t^{t'}\int_D \psi_J^{(i)}(s, y; t, x, t', x')\cdot p(s, y|t, x; \theta)dyds, \quad i = 1, 2, 3.$$

Recalling $z = (x' - y)/\sqrt{t'-s}$ and (51), we further have

$$r_i^{(J)} = \int_t^{t'}\int_D\sum_{h\in I_i}\frac{(t'-s)^{J-\frac{|h|}{2}}}{J!}H_h(z; \nu_0)q(t', x|s, y)w_{J+1,h}(s, y)p(s, y|t, x; \theta)dyds. \tag{62}$$

Next, consider the bounds for $r_1^{(J)}$, $r_2^{(J)}$, and $r_3^{(J)}$, respectively. To bound each remainder term in (62), the basic idea is to decompose the integrand in (62) 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 them respectively.

Consider the first term r_1^J with $h \in I_1$ (i.e., $1 \le |h| \le 3(J+1)/2$). Recall that $\partial_h q(t', x'|s, y) = (t'-s)^{-|h|/2}H_h(z; \nu_0)q(t', x'|s, y)$ in (51), together with Theorem 1 in Chapter 9 of Friedman (1964), we have

$$|H_h(z; \nu_0)q(t', x|s, y)| \le C(t'-s)^{-\frac{m}{2}}e^{-\frac{\lambda_0\|x'-y\|^2}{2(t'-s)}}, \tag{63}$$

where λ_0, C are positive constants depending only on $\nu(t,x;\theta), t, x, J$. Moreover, by Assumptions 2.3 and 2.4 and (6.12) in Chapter 1 of Friedman (1964), there exists $\lambda_1 > 0$ such that

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

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

$$|w_{J+1,h}(s,y)p(s,y|t,x;\theta)| \le C(s-t)^{-\frac{m}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}}. \tag{64}$$

Combining (63) and (64), we have

$$\begin{aligned}&\int_D |H_h(z;\nu_0)q(t',x|s,y)w_{J+1,h}(s,y)p(s,y|t,x;\theta)|dy\\ \le& C\int_D (t'-s)^{-\frac{m}{2}} e^{-\frac{\lambda\|x'-y\|^2}{2(t'-s)}}(s-t)^{-\frac{m}{2}} e^{-\frac{\lambda\|y-x\|^2}{2(s-t)}}dy \le C(t'-t)^{-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}},\end{aligned}$$

where $\lambda = \min\{\lambda_0,\lambda_2\}$ is a positive constant depending only on t, x, and J. Recalling (62) and $h \in I_1$ (i.e., $1 \le |h| \le 3(J+1)/2$), thus we have

$$\begin{aligned}|r_1^{(J)}| \le& \sum_{h\in I_1} C(t'-t)^{-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-s)}} \int_t^{t'} \frac{(t'-s)^{J-\frac{|h|}{2}}}{J!}ds\\ \le& C(t'-t)^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}}\\ =&\mathcal{O}((t'-t)^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil-\frac{m}{2}}),\ \text{for small}\ (t'-t).\end{aligned} \tag{65}$$

For the term $r_2^{(J)}$ with $h \in I_2$ (i.e., $3(J+1)/2 < |h| < 2(J+1)$), the inequality in (63) still holds. Using the expansion of $w_{J+1,h}$ in (46), we have an alternative bound as follows

$$\begin{aligned}|w_{J+1,h}(s,y)p(s,y|t,x;\theta)| \le& C(s-t)^{a_0+\frac{|a|}{2}-\frac{m}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}}\\ =& C(s-t)^{\frac{a_0}{2}+\frac{2|h|-3(J+1)}{2}-\frac{m}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}},\end{aligned} \tag{66}$$

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

$$\left|\left(\frac{y-x}{\sqrt{s-t}}\right)^a p(s,y|t,x;\theta)\right| \le C\left|\left(\frac{y-x}{\sqrt{s-t}}\right)^a\right| \frac{e^{-\frac{\lambda_1\|y-x\|^2}{2(s-t)}}}{(s-t)^{\frac{m}{2}}} \le C(s-t)^{-\frac{m}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}}.$$

Then, combining (62), (63) with (66) and $h \in I_2$ (i.e., $3(J+1)/2 < |h| < 2(J+1)$), we have

$$|r_2^{(J)}| \le C\sum_{h\in I_2}(t'-t)^{-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}} \cdot \int_t^{t'} \frac{(t'-s)^{J-\frac{|h|}{2}}}{J!}(s-t)^{\frac{a_0}{2}+\frac{2|h|-3(J+1)}{2}}ds \tag{67}$$

$$\begin{aligned}=&C\sum_{h\in I_2}(t'-t)^{\frac{a_0}{2}+\frac{|h|-(J+1)}{2}-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}} \cdot B\Big(1+J-\tfrac{|h|}{2}, 1+\tfrac{a_0}{2}+\tfrac{2|h|-3(J+1)}{2}\Big)\\ =&\mathcal{O}((t'-t)^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil-\frac{m}{2}}),\ \text{for small}\ (t'-t),\end{aligned} \tag{68}$$

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 (67) is not integrable around $s = t'$. To overcome the problem, we use integration by parts to reduce the order of differentiation by one. By $\partial_h q(t',x'|s,y) = \partial_{e_i}\partial_{h-e_i}q(t',x'|s,y)$ and integration by parts,[12] then we have,

$$\begin{aligned}&\int_D \partial_{e_i}\partial_{h-e_i}q(t',x'|s,y)\cdot w_{J+1,h}(s,y)p(s,y|t,x;\theta)dy\\ =&\int_D \partial_{h-e_i}q(t',x'|s,y)\cdot\Big(\partial_{e_i}w_{J+1,h}(s,y)\cdot p(s,y|t,x;\theta) + w_{J+1,h}(s,y)\cdot\partial_{e_i}p(s,y|t,x;\theta)\Big)dy.\end{aligned}$$

Recalling (51), similarly to (63), we have

$$|\partial_{h-e_i}q(t',x'|s,y)| = |(t'-s)^{-\frac{|h|-1}{2}}H_{h-e_i}(z;\nu_0)q(t',x'|s,y)| \le C(t'-s)^{-\frac{|h|-1+m}{2}} e^{-\frac{\lambda_0\|x'-y\|^2}{2(t'-s)}}. \tag{69}$$

12 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.

Mimicking the proof of (66), by (46), we have

$$|\partial_{e_i} w_{J+1,h}(s,y)\cdot p(s,y|t,x;\theta)| \le C(s-t)^{\frac{a_0}{2}+\frac{2|h|-3(J+1)-1}{2}-\frac{m}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}}. \tag{70}$$

Similarly to (64), by (46) and (6.13) in Chapter 1 of Friedman (1964), we have

$$|w_{J+1,h}(s,y)\cdot \partial_{e_i} p(s,y|t,x;\theta)| \le C(s-t)^{\frac{a_0}{2}+\frac{2|h|-3(J+1)}{2}-\frac{m+1}{2}} e^{-\frac{\lambda_2\|y-x\|^2}{2(s-t)}}. \tag{71}$$

Recall $r_3^{(J)}$ defined in (60) and (61) with $h \in I_3$ (i.e., $|h| = 2(J+1)$). Then, using similar arguments for $r_2^{(J)}$ and combining (69)–(71), we have

$$\begin{aligned}
|r_3^{(J)}| \le & C\sum_{h\in I_3}(t'-t)^{-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}} \cdot \int_t^{t'} \frac{(t'-s)^{J-(|h|-1)/2}}{J!}(s-t)^{\frac{a_0}{2}+\frac{2|h|-3(J+1)}{2}-\frac{1}{2}} ds \\
= & C\sum_{h\in I_3}(t'-t)^{\frac{a_0}{2}+\frac{|h|-(J+1)}{2}-\frac{m}{2}} e^{-\frac{\lambda\|x'-x\|^2}{2(t'-t)}} \cdot B\left(1+J-\tfrac{|h|-1}{2}, 1+\tfrac{a_0}{2}+\tfrac{2|h|-3(J+1)}{2}-\tfrac{1}{2}\right) \\
= & \mathcal{O}((t'-t)^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil-\frac{m}{2}}), \text{ for small } (t'-t).
\end{aligned} \tag{72}$$

Finally, putting (65), (68), and (72) into together, the error $|r^{(J)}(t',x'|t,x;\theta)| = \mathcal{O}((t'-t)^{\frac{1}{2}\lceil\frac{J+1}{2}\rceil-\frac{m}{2}})$ as $t'-t \to 0$, uniformly for $(t,x,x',\theta) \in [0,T]\times D^c \times D \times \Theta$. □

Proof of Theorem 3.3. Rewrite (15) of Theorem 3.1 as

$$p^{(J)}(t',x'|t,x;\theta) = q(t',x'|t,x)\left(1+A_J\right), \tag{73}$$

where A_J is defined as

$$A_J = \sum_{N=1}^{J}\sum_{i=1}^{\lfloor 3N/2\rfloor}\sum_{|h|=i}\frac{w_{N,h}(t,x)H_h(z;\nu_0)}{N!}\Delta^{N-i/2}.$$

Letting $k/2 = N - i/2$ i.e. $k = 2N - i$ or $i = 2N - k$, we have

$$A_J = \sum_{N=1}^{J}\sum_{k=2N-\lfloor 3N/2\rfloor}^{2N-1}\sum_{|h|=2N-k}\frac{w_{N,h}(t,x)H_h(z;\nu_0)}{N!}\Delta^{k/2},$$

We exchange the order of summation with respect to N and k:

$$A_J = \sum_{k=1}^{2J-1}\Delta^{k/2}\sum_{N=\lceil (k+1)/2\rceil}^{(2k)\wedge J}\frac{1}{N!}\sum_{|h|=2N-k} w_{N,h}(t,x)H_h(z;\nu_0),$$

Choose $J = 2K$ for some positive integer K. Then

$$\begin{aligned}
A_J &= \sum_{k=1}^{K}\Delta^{k/2}C_k + \sum_{k=K+1}^{4K-1}\Delta^{k/2}\sum_{N=\lceil (k+1)/2\rceil}^{2K}\frac{1}{N!}\sum_{|h|=2N-k} w_{N,h}(t,x)H_h(z;\nu_0) \\
&= \sum_{k=1}^{K}\Delta^{k/2}C_k + \mathcal{O}\left(\Delta^{\frac{K+1}{2}}\right),
\end{aligned} \tag{74}$$

where C_k is defined in (23). Dropping the high order terms, we get the delta expansion $p^{(K,\Delta)}$ as defined in (22).

Note that $w_{N,h}(t,x)H_h(z;\nu_0)$ is uniformly bounded for any N, h and $(t,x,x',\theta) \in [0,T]\times D^c\times D\times\Theta$. Then by (73)–(74), (22) and $J = 2K$, we have

$$\sup_{(t,x,x',\theta)\in[0,T]\times D^c\times D\times\Theta}\left|p^{(K,\Delta)}(t',x'|t,x;\theta) - p^{(J)}(t',x'|t,x;\theta)\right| = \mathcal{O}\left(\Delta^{\frac{K+1}{2}-\frac{m}{2}}\right).$$

Together with (21), we get (24). □

Proof of Theorem 4.1. For fixed $x \in D^c$, $T > 0$ and $0 \le t < t' \le T$, let

$$R^{(K,\Delta)}(t',x'|t,x;\Theta) := \sup_{\theta\in\Theta}|p(t',x'|t,x;\theta) - p^{(K,\Delta)}(t',x'|t,x;\theta)|.$$

According to Theorem 3.3, for fixed x and t, $p^{(K,\Delta)}(t',x'|t,x)$ 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 Δ_ϵ independent of x' and θ, such that for all $\Delta \le \Delta_\epsilon$, we have

$R^{(K,\Delta)}(t', x'|t, x; \Theta) < \epsilon$, and

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

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

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

Then,

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

where $\pi_t(x; \theta_0)$ is the marginal density of X_t at the true parameter value θ_0. Since the probability is bounded and the density is integrable, by the dominated convergence theorem, we have

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

that is, $R^{(K,\Delta)}(t', X_{t'}|t, X_t; \Theta)$ converges to zero in probability, as $\Delta \to 0$. And the convergence of $\sup_{\theta\in\Theta} |\ell_n^{(K,\Delta)}(\theta) - \ell_n(\theta)|$ 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 $\hat{\theta}_n^{(K,\Delta)}$ and its convergence in probability to the true MLE $\hat{\theta}_n$ are guaranteed by the continuous differentiability in θ of the log-likelihood function $\ell_n(\theta)$ and the approximate log-likelihood functions $\ell_n^{(K,\Delta)}(\theta)$ for all K. □

Proof of Proposition 5.1. Recall that $\zeta_{l,j}(x)$ is given by (cf. Definition 1 of Lee et al., 2014): for $l, j \geq 1$,

$$\zeta_{l,j}(x) = \mathcal{L}\zeta_{l-1,j}(x) + \left(\mu(x) + \frac{\partial}{\partial x}\right)\zeta_{l,j-1}(x). \tag{75}$$

Furthermore, $\zeta_{0,0}(x) = 1$ and $\zeta_{l,j}(x) = 0$ if $l < 0$ or $j < 0$.

By the above definition of ζ, we know that $\zeta_{0,0}(x) = 1$; $\zeta_{k-\lceil k/2\rceil,2\lceil k/2\rceil-k}(x) = \zeta_{k/2,0} = 0$ if k is an even number. Furthermore, $\lceil k/2\rceil = \lceil (k+1)/2\rceil$ if k is an odd number. Using these facts, we can rewrite (35) as follows:

$$\hat{p}^{(K,\Delta)} = \frac{\phi(z)}{\sqrt{\Delta}}\left(1 + \sum_{k=1}^{2K}\Delta^{\frac{k}{2}}\left(\sum_{N=\lceil (k+1)/2\rceil}^{k}\frac{1}{N!}\zeta_{k-N,2N-k}(x)H_{2N-k}(z)\right)\right). \tag{76}$$

A key observation is that the following statement holds:

$$\zeta_{N-h,h}(x) = w_{N,h}(x), \text{ for } 0 \leq h \leq N, \tag{77}$$

where $w(x)$ is defined by the recursive relations (i)–(iii), i.e., (18)–(19). Noting that the model (30) is time-homogeneous, thus we can omit the time variable in the coefficient $w(t, x)$. Moreover, the corresponding Eq. (19) does not contain the time derivatives. Then, combining (77) with (76), we have

$$\hat{p}^{(K,\Delta)} = \frac{\phi(z)}{\sqrt{\Delta}}\left(1 + \sum_{k=1}^{2K}\Delta^{\frac{k}{2}}\left(\sum_{N=\lceil (k+1)/2\rceil}^{k}\frac{1}{N!}w_{N,2N-k}(x)H_{2N-k}(z)\right)\right).$$

Recalling the definition of $p^{(2K,\Delta)}$ in (27), we have proved that the equality (36) holds.

Therefore, we only need to verify (77), which is proved by mathematical induction over N. For $N = 0$, both sides of (77) are equal to one by definition. Assume (77) holds for N. For $N + 1$,

$$\begin{aligned}\zeta_{N+1-h,h}(x) &= \mathcal{L}\zeta_{N-h,h}(x) + \left(\mu(x) + \frac{\partial}{\partial x}\right)\zeta_{N-(h-1),h-1}(x)\\ &= \mathcal{L}w_{N,h}(x) + \left(\mu(x) + \frac{\partial}{\partial x}\right)w_{N,h-1}(x)\\ &= w_{N+1,h}(x),\end{aligned} \tag{78}$$

where we have used (75) and (19) to get the first and third equalities, respectively. □

Appendix B. A new explicit recursive algorithm for the expansion of Li (2013)

B.1. The explicit recursive algorithm

In this section we present a new explicit recursive algorithm to compute the expansion coefficients in the density approximation of Li (2013). To derive the explicit algorithm and facilitate the comparison with the delta expansion, we use the Itô integral, instead of the Stratonovich integral, to rewrite the expansion algorithm.

Consider a rescaled m-dimensional diffusion process $X(t) := (X_1(t), \ldots, X_m(t))^\top$ satisfying

$$dX^\epsilon(t) = \epsilon^2 \mu(X^\epsilon(t))dt + \epsilon\sigma(X^\epsilon(t))dW(t), \tag{79}$$

where $\{W(t) := (W_1(t), \ldots, W_m(t))^\top, t \geq 0\}$ is a m-dimensional standard Brownian motion. The corresponding infinitesimal generator and differential operators are

$$\mathcal{L}_0 = \sum_{i=1}^{m} \mu_i(x)\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \nu_{ij}(x)\partial^2_{x_i x_j}, \tag{80}$$

$$\mathcal{L}_j = \sum_{i=1}^{m} \sigma_{ij}(x)\partial_{x_i}, \quad j = 1, \ldots, m. \tag{81}$$

For convenience, let

$$\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \{0, 1, \ldots, m\}^n$$

associated with the "norm" $\|\mathbf{i}\| = \sum_{l=1}^{n}(2 \cdot 1_{\{i_l=0\}} + 1_{\{i_l\neq 0\}})$, and define the set

$$\mathcal{M}_k = \{\mathbf{i} = (i_1, i_2, \ldots, i_n) \big| \|\mathbf{i}\| = k\}.$$

Note that we can recursively define $(\mathcal{M}_k)_{k\geq 0}$ below:

$$\mathcal{M}_0 = \emptyset; \quad \mathcal{M}_1 = \{(1), \ldots, (m)\}; \tag{82}$$

$$\mathcal{M}_{k+1} = \{\mathbf{i}\big| i_1 = 0,\ (i_2, \ldots, i_n) \in \mathcal{M}_{k-1}\} \bigcup_{a=1}^{m} \{\mathbf{i}\big| i_1 = a,\ (i_2, \ldots, i_n) \in \mathcal{M}_k\}. \tag{83}$$

Similarly to the proof of Theorem 3.3 in Watanabe (1987) and Lemma 1 in Li (2013), we have the following pathwise expansion of $X^\epsilon(1)$. The difference is that we present the expansion via the Itô integral.

Lemma B.1. *The expansion of the random variable $X^\epsilon(1)$ now reads as*

$$X^\epsilon = \sum_{k=0}^{J} \epsilon^k F_k + \mathcal{O}\left(\epsilon^{J+1}\right), \tag{84}$$

where $F_0 = x$, and the expansion coefficients have the following general form

$$F_k = \sum_{\mathbf{i}\in\mathcal{M}_k} C_{\mathbf{i}}(x) \cdot \mathbb{I}_{\mathbf{i}}(1), \tag{85}$$

where $\mathbf{i} = (i_1, i_2, \ldots, i_n)$ and

$$C_{\mathbf{i}}(x) = (\mathcal{L}_{i_n} \circ \cdots \circ \mathcal{L}_{i_2})\sigma_{\cdot i_1}(x). \tag{86}$$

Here $\sigma_{\cdot 0} = \mu$ and $\sigma_{\cdot i_1} = (\sigma_{1i_1}, \ldots, \sigma_{mi_1})^\top$. The iterated Itô integral $\mathbb{I}_{\mathbf{i}}(t)$ is defined through

$$\mathbb{I}_{\mathbf{i}}(t) = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{k-1}} dW_{i_k}(t_k) \cdots dW_{i_2}(t_2)dW_{i_1}(t_1), \tag{87}$$

where $W_0(t) = t$.

Consider the following changing of variable

$$Y^\epsilon := D(x)\frac{X^\epsilon(1) - x}{\epsilon} \to B(1), \text{ as } \epsilon \to 0, \tag{88}$$

where $B(1)$ is a correlated Brownian motion with covariance matrix

$$\Sigma(x) = D(x)\sigma(x)\sigma(x)^\top D(x),$$

and $D(x)$ is the diagonal matrix defined by

$$D(x) = diag\left(\left(\sum_{j=1}^{m}\sigma_{1j}^2(x)\right)^{-1/2}, \ldots, \left(\sum_{j=1}^{m}\sigma_{mj}^2(x)\right)^{-1/2}\right). \tag{89}$$

By Lemma B.1, we have the expansion of $Y^\epsilon(1)$:

$$Y^\epsilon = \sum_{i=0}^{J} D(x)F_{i+1}\epsilon^i + \mathcal{O}(\epsilon^{J+1}). \tag{90}$$

Therefore, we have the following expansion of the composition of the Dirac delta function and the random variable Y^ϵ:

$$\delta(Y^\epsilon - y) := \sum_{k=0}^{J} \Phi_k(y)\epsilon^k + \mathcal{O}(\epsilon^{J+1}). \tag{91}$$

On the other hand, the transition density $p_X(\Delta, x'|x)$ of X^ϵ (from (t, x) to (t', x')) is given by

$$p_X(\Delta, x'|x) = \Delta^{-\frac{m}{2}} \det(D(x))\mathbb{E}[\delta(Y^\epsilon - y)|X(0) = x], \tag{92}$$

where $y := D(x)\frac{x'-x}{\sqrt{\Delta}}$ and $\Delta = t' - t$. Thus the density $p_X(\Delta, x'|x)$ admits the expansion below:

$$p_X(\Delta, x'|x) = \Delta^{-\frac{m}{2}} \det(D(x)) \sum_{k=0}^{J} \epsilon^k \Omega_k(y) + \mathcal{O}(\epsilon^{J+1}), \tag{93}$$

where $\Omega_k(y) := \mathbb{E}[\Phi_k(y)]$ is the expansion coefficient of the transition density. Then using integration by parts formula, we can derive the explicit formula for the expansion coefficient in Proposition B.1. Before presenting it, we define a set $\mathcal{S}_k$ with $\mathbf{j} := (j_1, j_2, \ldots, j_l)$ as follows:

$$\mathcal{S}_k = \{\mathbf{j}\ |j_\omega \geq 1, \omega = 1, \ldots, l;\ j_1 + j_2 + \cdots + j_l = k, ;\ l = 1, 2, \ldots\}. \tag{94}$$

Note that we can also define $(\mathcal{S}_k)_{k\geq 1}$ recursively by $\mathcal{S}_1 = \{(1)\}$ and

$$\mathcal{S}_{k+1} = \{\mathbf{j}\ |j_1 = 1,\ (j_2, \ldots, j_l) \in \mathcal{S}_k\} \cup \{\mathbf{j}\ |(j_1 - 1, j_2, \ldots, j_l) \in \mathcal{S}_k\}. \tag{95}$$

Proposition B.1. *The explicit formula for the expansion coefficient $\Omega_k(y)$ is given by*

$$\Omega_k(y) = \sum_{(j_1,j_2,\ldots,j_l)\in\mathcal{S}_k} \frac{(-1)^l}{l!} \sum_{(r_1,\ldots,r_l)\in\{1,2,\ldots,m\}^l} \prod_{\omega=1}^{l} D_{r_\omega r_\omega}(x)\partial^l_{y_{r_1}\cdots y_{r_l}}\left(\mathbb{E}\left[\prod_{\omega=1}^{l} F_{j_\omega+1,r_\omega}\middle|W(1) = z\right]\phi(y;\ \Sigma(x))\right), \tag{96}$$

where $z = \sigma(x)^{-1}D(x)^{-1}y$, $\mathcal{S}_k$, $F_{j_\omega+1}$ and $D_{r_\omega r_\omega}(x)$ are defined by (95), (85) and (89), respectively. $\phi(y;\ \Sigma(x))$ is the m-dimensional normal density function with mean 0 and covariance matrix $\Sigma(x) = D(x)\sigma(x)\sigma(x)^\top D(x)$. The conditional expectation is given by

$$\mathbb{E}\left[\prod_{\omega=1}^{l} F_{j_\omega+1,r_\omega}\middle|W(1) = z\right] = \sum_{\substack{\mathbf{i}_\omega\in\mathcal{M}_{j_\omega+1}\\ \omega=1,\ldots,l}} \left(\prod_{\omega=1}^{l} C_{\mathbf{i}_\omega,r_\omega}(x)\cdot\mathbb{E}\left[\prod_{\omega=1}^{l}\mathbb{I}_{\mathbf{i}_\omega}(1)\middle|W(1) = z\right]\right), \tag{97}$$

where $C_{\mathbf{i}_\omega,r_\omega}(x)$ is for (86), and $\mathcal{M}_j$ is recursively defined by (82) and (83).

To obtain the explicit formula for (96), a major obstacle is to compute the conditional expectation of the product of iterated Itô integrals in (97). Li (2013) introduces an algorithm to compute this conditional expectation, which is simplified by Li et al. (2016).

We contribute an explicit recursive algorithm to compute the expansion coefficients $\Omega_k(y)$. First, in contrast to the algorithms in Li (2013); Li et al. (2016), the newly derived algorithm does not require the conversion from iterated Stratonovich integrals to Itô integrals because we state the expansion using the Itô integral directly. Second and most importantly, we establish an explicit recursive formula for the conditional expectation of the product of iterated Itô integrals in the following two propositions. Specifically, in Proposition B.2, we prove that the conditional expectation of an iterated Itô integral in fact corresponds to a Hermite polynomial. Proposition B.3 states that the conditional expectation of the product of iterated Itô integrals is a linear combination of the Hermite polynomials. The results in these propositions are new to the literature.

Proposition B.2. *The conditional expectation of the iterated Itô integral defined in (87) can be expressed as a standard Hermite polynomial below:*

$$\mathbb{E}[\mathbb{I}_{\mathbf{i}}(t)|W(t) = x] = \frac{\sqrt{t}^{\|\mathbf{i}\|}}{n!} H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{x}{\sqrt{t}}\right), \tag{98}$$

where $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \{0, 1, \ldots, m\}^n$, $\mathbf{n_i} = (n_\mathbf{i}(1), \ldots, n_\mathbf{i}(m))$ *with* $n_\mathbf{i}(a)$ *being the number of a (for* $a = 0, 1, \ldots, m$*) in* $\mathbf{i}$*, and* $\|\mathbf{i}\| = 2n_\mathbf{i}(0) + \sum_{a=1}^m n_\mathbf{i}(a)$*. The multivariate Hermite polynomial in (98) is defined through the univariate standard Hermite polynomial as follows:*

$$H_{\mathbf{n_i}}\left(\frac{x}{\sqrt{t}}\right) = \prod_{a=1}^{m} H_{n_\mathbf{i}(a)}\left(\frac{x_a}{\sqrt{t}}\right). \tag{99}$$

Proposition B.3. *The conditional expectation of the product of iterated Itô integrals is a linear combination of the Hermite polynomials as below*

$$\mathbb{E}\left[\prod_{\omega=1}^{l} \mathbb{I}_{\mathbf{i}_\omega}(t)\Big| W(t) = x\right] = \sum_{0\le\alpha\le\lfloor \mathbf{n}(\vec{\mathbf{i}})/2\rfloor} \tilde{w}_{\alpha,\vec{\mathbf{i}}} \cdot \left(\frac{\sqrt{t}^{\|\vec{\mathbf{i}}\|}}{(\ell(\vec{\mathbf{i}}) - |\alpha|)!} H_{\mathbf{n}(\vec{\mathbf{i}})-2\alpha}\left(\frac{x}{\sqrt{t}}\right)\right), \tag{100}$$

where $\vec{\mathbf{i}} := \{\mathbf{i}_1, \ldots, \mathbf{i}_l\}$, $\mathbf{n}(\vec{\mathbf{i}}) = (n_{\vec{\mathbf{i}}}(1), \ldots, n_{\vec{\mathbf{i}}}(m))$ *with* $n_{\vec{\mathbf{i}}}(a) = \sum_{\omega=1}^{l} n_{\mathbf{i}_\omega}(a)$, $\|\vec{\mathbf{i}}\| = \sum_{\omega=1}^{l} \|\mathbf{i}_\omega\|$*, and* $\ell(\vec{\mathbf{i}})$ *is the total length of all vectors* $\mathbf{i}_\omega$ *(for* $\omega = 1, \ldots, l$*) in* $\vec{\mathbf{i}}$, $\alpha = (\alpha_1, \ldots, \alpha_m)$ *with* $|\alpha| = \alpha_1 + \cdots + \alpha_m$ *and* $\min(\alpha) = \min(\alpha_1, \ldots, \alpha_m)$*. The coefficient* $\tilde{w}_{\alpha,\vec{\mathbf{i}}}$ *is recursively determined as follows:* $\tilde{w}_{\alpha,\vec{\mathbf{i}}} = 1_{\{\alpha=0\}}$ *if* $\vec{\mathbf{i}} = \emptyset$ *or* $\{\mathbf{i}_1\}$; $\tilde{w}_{\alpha,\vec{\mathbf{i}}} = 0$ *if* $\min(\alpha) < 0$ *or* $\min(\lfloor \mathbf{n}(\vec{\mathbf{i}})/2\rfloor - \alpha) < 0$*; and*

$$\tilde{w}_{\alpha,\vec{\mathbf{i}}} = \sum_{\omega_1=1}^{l} \tilde{w}_{\alpha,\vec{\mathbf{i}}_{-\omega_1}} + \sum_{1\le\omega_1<\omega_2\le l} 1_{\{i_{\omega_1,1}=i_{\omega_2,1}\neq 0\}} \tilde{w}_{\alpha-e_{i_{\omega_1,1}},\vec{\mathbf{i}}_{-\omega_1-\omega_2}}, \tag{101}$$

where e_a *(for* $a = 1, \ldots, m$*) is the m-dimensional index vector, in which the ath component is 1, and the others are 0;* $\vec{\mathbf{i}}_{-\omega_1}$ *means replacing* $\mathbf{i}_{\omega_1} = (i_{\omega_1,1}, i_{\omega_1,2}, \ldots, i_{\omega_1,n_{\omega_1}})$ *with* $-\mathbf{i}_{\omega_1} = (i_{\omega_1,2}, \ldots, i_{\omega_1,n_{\omega_1}})$ *in the set* $\vec{\mathbf{i}}$*.*

B.2. Related proofs for Appendix B.1

Lemma B.2. *Let* $(\mathcal{F}_t)_{t\ge0}$ *be the natural filtration generated by a standard m-dim Brownian motion* $\{W(t), t \ge 0\}$*. Given* $s \le t$*, for a* $\mathcal{F}_s$*-measurable random variable Z, the following equation holds:*

$$\mathbb{E}\left[Z|W(t) = x\right] = \mathbb{E}\left[\mathbb{E}[Z|W(s)]|W(t) = x\right]. \tag{102}$$

Moreover, for a $\mathcal{F}_t$ *adapted process* $\{f(t), t \ge 0\}$*, we have*

$$\mathbb{E}\left[\int_0^t f(s)dW(s)|W(t) = x\right] = \mathbb{E}\left[\int_0^t \mathbb{E}[f(s)|W(s)]dW(s)|W(t) = x\right]. \tag{103}$$

Proof of Lemma B.2. Assume $m = 1$ without loss of generality because each component of the m-dimensional Brownian motion is mutually independent. Then by the iterated conditional expectation, we have

$$\begin{aligned}\mathbb{E}\left[Z|W(t) = x\right] =& \mathbb{E}\left[\mathbb{E}[Z|W(s), W(t)]|W(t) = x\right]\\ =&\mathbb{E}\left[\int Zg(Z|W(s), W(t))dZ|W(t) = x\right],\end{aligned} \tag{104}$$

where $g(Z|W(s), W(t))$ is the conditional density function of Z conditional on $W(s)$ and $W(t)$. By the definition of conditional density and the Markov property of the Brownian motion, we have

$$\begin{aligned}g(Z|W(s), W(t)) =& \frac{g(Z, W(s), W(t))}{g(W(s), W(t))}\\ =& \frac{g(W(t)|Z, W(s))g(Z, W(s))}{g(W(t)|W(s))g(W(s))} = \frac{g(Z, W(s))}{g(W(s))} = g(Z|W(s)).\end{aligned}$$

Substituting the above formula into (104), we have

$$\mathbb{E}\left[Z|W(t) = x\right] = \mathbb{E}\left[\int Zg(Z|W(s))dZ|W(t) = x\right] = \mathbb{E}\left[\mathbb{E}[Z|W(s)]|W(t) = x\right].$$

Next, we consider (103). Let $t_i = i \cdot t/n$ for $i = 0, 1, \ldots, n$. By the definition of the Itô integral, we have

$$\int_0^t f(s)dW(s) = \lim_{n\to\infty} \sum_{i=1}^{n} f(t_{i-1})(W(t_i) - W(t_{i-1})), \tag{105}$$

and

$$\int_0^t \mathbb{E}[f(s)|W(s)]dW(s) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[f(t_{i-1})|W(t_{i-1})](W(t_i) - W(t_{i-1})). \tag{106}$$

By the formula (102), we have

$$
\begin{aligned}
&\mathbb{E}\left[\Big(\mathbb{E}[f(t_{i-1})|W(t_{i-1})]\Big)(W(t_i)-W(t_{i-1}))|W(t)=x\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\Big(\mathbb{E}[f(t_{i-1})|W(t_{i-1})]\Big)W(t_i)|W(t_i)\right]|W(t)=x\right]\\
&-\mathbb{E}\left[\mathbb{E}\left[\Big(\mathbb{E}[f(t_{i-1})|W(t_{i-1})]\Big)W(t_{i-1})|W(t_{i-1})\right]|W(t)=x\right]\\
=&\mathbb{E}\left[\Big(\mathbb{E}\left[f(t_{i-1})W(t_i)|W(t_i)\right]\Big)|W(t)=x\right]-\mathbb{E}\left[\Big(\mathbb{E}\left[f(t_{i-1})W(t_{i-1})|W(t_{i-1})\right]\Big)|W(t)=x\right]\\
=&\mathbb{E}\left[f(t_{i-1})(W(t_i)-W(t_{i-1}))|W(t)=x\right].
\end{aligned}
\tag{107}
$$

By (105), (106) and (107), we know that (103) holds. $\square$

Proof of Proposition B.2. We prove (98) by mathematical induction on the length of $\mathbf{i}$ (i.e., n). If $n=1$, (98) obviously holds. Suppose (98) holds for $n>1$. For $\mathbf{i}=(i_1,\ldots,i_n,i_{n+1})$, we consider two cases $i_1=0$ and $i_1\neq 0$.

If $i_1=0$, then

$$
\mathbf{I}_{\mathbf{i}}(t)=\int_0^t \mathbf{I}_{-\mathbf{i}}(s)ds,
$$

where $-\mathbf{i}:=(i_2,\ldots,i_{n+1})$ means deleting the first element of $\mathbf{i}$. By (102) of Lemma B.2, we have

$$
\begin{aligned}
\mathbb{E}[\mathbf{I}_{\mathbf{i}}(t)|W(t)=x]&=\int_0^t \mathbb{E}\left[\mathbf{I}_{-\mathbf{i}}(s)|W(t)=x\right]ds\\
&=\int_0^t \mathbb{E}\left[\mathbb{E}[\mathbf{I}_{-\mathbf{i}}(s)|W(s)]|W(t)=x\right]ds\\
&=\int_0^t \mathbb{E}\left[\frac{\sqrt{s}^{\|\mathbf{i}\|-2}}{n!}H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{W(s)}{\sqrt{s}}\right)|W(t)=x\right]ds,
\end{aligned}
$$

where the last holds since (98) is true for n.

Let $|\mathbf{n}_{\mathbf{i}}|=n_{\mathbf{i}}(1)+\cdots n_{\mathbf{i}}(m)$ and note that

$$
\begin{aligned}
&\mathbb{E}\left[\sqrt{s}^{-|\mathbf{n}_{\mathbf{i}}|}H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{W(s)}{\sqrt{s}}\right)|W(t)=x\right]\\
=&\prod_{a=1}^m \sqrt{s}^{-n_{\mathbf{i}}(a)}\int_{\mathbb{R}}\frac{\sqrt{2\pi t}}{e^{-\frac{x_a^2}{2t}}}\frac{e^{-\frac{(x_a-\xi_a)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}\frac{e^{-\frac{\xi_a^2}{2s}}}{\sqrt{2\pi s}}H_{n_{\mathbf{i}}(a)}\left(\frac{\xi_a}{\sqrt{s}}\right)d\xi_a\\
=&\prod_{a=1}^m \sqrt{2\pi t}\cdot e^{\frac{x_a^2}{2t}}\int_{\mathbb{R}}(-1)^{n_{\mathbf{i}}(a)}\partial_{\xi_a}^{n_{\mathbf{i}}(a)}\left(\frac{e^{-\frac{\xi_a^2}{2s}}}{\sqrt{2\pi s}}\right)\frac{e^{-\frac{(x_a-\xi_a)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}d\xi_a\\
=&\prod_{a=1}^m \sqrt{2\pi t}\cdot e^{\frac{x_a^2}{2t}}\int_{\mathbb{R}}(-1)^{n_{\mathbf{i}}(a)}\frac{e^{-\frac{\xi_a^2}{2s}}}{\sqrt{2\pi s}}\partial_{x_a}^{n_{\mathbf{i}}(a)}\left(\frac{e^{-\frac{(x_a-\xi_a)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}\right)d\xi_a\\
=&\prod_{a=1}^m \sqrt{2\pi t}\cdot e^{\frac{x_a^2}{2t}}(-1)^{n_{\mathbf{i}}(a)}\partial_{x_a}^{n_{\mathbf{i}}(a)}\left(\int_{\mathbb{R}}\frac{e^{-\frac{\xi_a^2}{2s}}}{\sqrt{2\pi s}}\frac{e^{-\frac{(x_a-\xi_a)^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}}d\xi_a\right)\\
=&\prod_{a=1}^m \sqrt{t}^{-|\mathbf{n}_{\mathbf{i}}|}H_{n_{\mathbf{i}}(a)}\left(\frac{x_a}{\sqrt{t}}\right)=\sqrt{t}^{-|\mathbf{n}_{\mathbf{i}}|}H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{x}{\sqrt{t}}\right).
\end{aligned}
\tag{108}
$$

Then, noting that $\|\mathbf{i}\|=|\mathbf{n}_{\mathbf{i}}|+2n_{\mathbf{i}}(0)$ and $|\mathbf{n}_{\mathbf{i}}|+n_{\mathbf{i}}(0)=n+1$, we have

$$
\int_0^t \mathbb{E}\left[\frac{\sqrt{s}^{\|\mathbf{i}\|-2}}{n!}H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{W(s)}{\sqrt{s}}\right)|W(t)=x\right]ds=\frac{\sqrt{t}^{\|\mathbf{i}\|}}{(n+1)!}H_{\mathbf{n}_{\mathbf{i}}}\left(\frac{x}{\sqrt{t}}\right),
\tag{109}
$$

then, (98) is true for $n+1$ when $i_1=0$.

For the case $i_1\neq 0$, assume $i_1=1$ without loss of generality. Then,

$$
\mathbf{I}_{\mathbf{i}}(t)=\int_0^t \mathbf{I}_{-\mathbf{i}}(s)dW_1(s).
$$

Applying (103) in Lemma B.2 repeatedly, we have

$$\begin{aligned}\mathbb{E}[\mathbf{I_i}(t)|W(t)=x] =&\mathbb{E}\left[\int_0^t \mathbf{I_{-i}}(s)dW_{i_1}(s)|W(t)=x\right]\\ =&\mathbb{E}\left[\int_0^t \mathbb{E}[\mathbf{I_{-i}}(s)|W(s)]dW_1(s)|W(t)=x\right]\\ =&\mathbb{E}\left[\int_0^t \frac{\sqrt{s}^{\|\mathbf{i}\|-1}}{n!}H_{\mathbf{n_{-i}}}\left(\frac{W(s)}{\sqrt{s}}\right)dW_1(s)|W(t)=x\right]\\ =&\mathbb{E}\left[\int_0^t \frac{s^n}{n!}Q_{1'}\left(\prod_{a=2}^m Q_a\right)dW_1(s)\Big|W(t)=x\right],\end{aligned} \tag{110}$$

where

$$Q_{1'} =\sqrt{s}^{-(n_{p\mathbf{i}}(1)-1)}H_{n_{\mathbf{i}}(1)-1}\left(\frac{W_1(s)}{\sqrt{s}}\right); \tag{111}$$

$$Q_a =\mathbb{E}\left[\sqrt{s}^{-n_{\mathbf{i}}(a)}H_{n_{\mathbf{i}}(a)}\left(\frac{W_a(s)}{\sqrt{s}}\right)|W_a(t)=x_a\right], \quad a=2,\ldots,m.$$

Similarly to the derivation of (108), we can obtain the formula for Q_a (for $a \geq 2$)

$$Q_a = \sqrt{t}^{-n_{\mathbf{i}}(a)}H_{n_{\mathbf{i}}(a)}\left(\frac{x_a}{\sqrt{t}}\right).$$

Substituting the above formula back to (110), then we have

$$\mathbb{E}[\mathbf{I_i}(t)|W(t)=x] = Q_1\prod_{a=2}^m\left(\sqrt{t}^{-n_{\mathbf{i}}(a)}H_{n_{\mathbf{i}}(a)}\left(\frac{x_a}{\sqrt{t}}\right)\right). \tag{112}$$

where

$$\begin{aligned}Q_1 :=&\mathbb{E}\left[\int_0^t \frac{s^n}{n!}Q_{1'}dW_1(s)\Big|W(t)=x\right]\\ =&\mathbb{E}\left[\int_0^t \frac{s^n}{n!}\sqrt{s}^{-(n_{\mathbf{i}}(1)-1)}H_{n_{\mathbf{i}}(1)-1}\left(\frac{W_1(s)}{\sqrt{s}}\right)dW_1(s)\Big|W_1(t)=x_1\right].\end{aligned} \tag{113}$$

We only need to derive the formula for Q_1. By the Itô formula, we have

$$\begin{aligned}&\frac{t^{n+1}\sqrt{t}^{-n_{\mathbf{i}}(1)}}{(n+1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{W_1(t)}{\sqrt{t}}\right)\\ =&\frac{n_{\mathbf{i}}(1)!t^{n+1-n_{\mathbf{i}}(1)}}{(n+1)!}\cdot\frac{\sqrt{t}^{n_{\mathbf{i}}(1)}}{n_{\mathbf{i}}(1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{W_1(t)}{\sqrt{t}}\right)\\ =&\int_0^t (n+1-n_{\mathbf{i}}(1))\frac{n_{\mathbf{i}}(1)!s^{n-n_{\mathbf{i}}(1)}}{(n+1)!}\frac{\sqrt{s}^{n_{\mathbf{i}}(1)}}{n_{\mathbf{i}}(1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{W_1(s)}{\sqrt{s}}\right)ds\\ &+\int_0^t \frac{n_{\mathbf{i}}(1)!s^{n+1-n_{\mathbf{i}}(1)}}{(n+1)!}\cdot\frac{\sqrt{s}^{n_{\mathbf{i}}(1)-1}}{(n_{\mathbf{i}}(1)-1)!}H_{n_{\mathbf{i}}(1)-1}\left(\frac{W_1(s)}{\sqrt{s}}\right)dW_1(s)\\ =&\frac{n+1-n_{\mathbf{i}}(1)}{n+1}\int_0^t \frac{s^n}{n!}\cdot\sqrt{s}^{-n_{\mathbf{i}}(1)}H_{n_{\mathbf{i}}(1)}\left(\frac{W_1(s)}{\sqrt{s}}\right)ds\\ &+\frac{n_{\mathbf{i}}(1)}{n+1}\int_0^t \frac{s^n}{n!}\cdot\sqrt{s}^{-(n_{\mathbf{i}}(1)-1)}H_{n_{\mathbf{i}}(1)-1}\left(\frac{W_1(s)}{\sqrt{s}}\right)dW_1(s).\end{aligned}$$

Taking expectation on both sides conditional on $W_1(t)=x_1$, we have

$$\frac{t^{n+1}\sqrt{t}^{-n_{\mathbf{i}}(1)}}{(n+1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{x_1}{\sqrt{t}}\right) = \frac{n+1-n_{\mathbf{i}}(1)}{n+1}\frac{t^{n+1}\sqrt{t}^{-n_{\mathbf{i}}(1)}}{(n+1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{x_1}{\sqrt{t}}\right)+\frac{n_{\mathbf{i}}(1)}{n+1}Q_1,$$

where the first term on the right hand side can be derived similarly to (109), and the second term is from the definition of Q_1 in (113). Thus, the formula for Q_1 is given by

$$Q_1 = \frac{t^{n+1}\sqrt{t}^{-n_{\mathbf{i}}(1)}}{(n+1)!}H_{n_{\mathbf{i}}(1)}\left(\frac{x_1}{\sqrt{t}}\right). \tag{114}$$

Recalling that $\|\mathbf{i}\| = |\mathbf{n_i}| + 2n_{\mathbf{i}}(0)$ and $|\mathbf{n_i}| + n_{\mathbf{i}}(0) = n + 1$, combining with (112) and (114), we know that (98) is true for $n + 1$ when $i_1 \neq 0$. $\square$

Proof of Proposition B.3. We prove it by mathematical induction on the total length of all vectors $\mathbf{i}_\omega$ (for $\omega = 1, \ldots, l$) in $\vec{\mathbf{i}}$, i.e., $\ell(\vec{\mathbf{i}})$. For $\ell(\vec{\mathbf{i}}) = 0$, that is, $\vec{\mathbf{i}} = \emptyset$. The product of iterated Itô integrals on the left of (100) degenerates to 1, then $\tilde{w}_{\alpha,\emptyset} = 1_{\{\alpha=0\}}$. For $\ell(\vec{\mathbf{i}}) = 1$, that is, $\vec{\mathbf{i}} = \{(i_1)\}$, then (100) holds from (98) under the setting $\tilde{w}_{\alpha,\{(i_1)\}} = 1_{\{\alpha=0\}}$. Assume (100) is true for $\ell(\vec{\mathbf{i}}) \leq N$. Consider the case $\ell(\vec{\mathbf{i}}) = N + 1$. By the Itô's formula, we have

$$\begin{aligned}\prod_{\omega=1}^{l} \mathbb{I}_{\mathbf{i}_\omega}(t) = &\sum_{\omega_1=1}^{l} \int_0^t \prod_{\omega\neq\omega_1} \mathbb{I}_{\mathbf{i}_\omega}(s)\mathbb{I}_{-\mathbf{i}_{\omega_1}}(s)dW_{i_{\omega_1,1}}(s)\\ &+ \sum_{1\leq\omega_1<\omega_2\leq l} 1_{\{i_{\omega_1,1}=i_{\omega_2,1}\neq 0\}} \int_0^t \prod_{\omega\neq\omega_1,\omega_2} \mathbb{I}_{\mathbf{i}_\omega}(s)\mathbb{I}_{-\mathbf{i}_{\omega_1}}(s)\mathbb{I}_{-\mathbf{i}_{\omega_2}}(s)ds. \end{aligned} \tag{115}$$

Using (103), we have

$$\underbrace{\mathbb{E}\Big[\int_0^t \prod_{\omega\neq\omega_1} \mathbb{I}_{\mathbf{i}_\omega}(s)\mathbb{I}_{-\mathbf{i}_{\omega_1}}(s)dW_{i_{\omega_1,1}}(s)\Big|W(t)=y\Big]}_{T_{\omega_1}} = \mathbb{E}\Big[\int_0^t \underbrace{\mathbb{E}\Big[\prod_{\omega\neq\omega_1} \mathbb{I}_{\mathbf{i}_\omega}(s)\mathbb{I}_{-\mathbf{i}_{\omega_1}}(s)\Big|W(s)\Big]}_{T'_{\omega_1}} dW_{i_{\omega_1,1}}(s)\Big|W(t)=y\Big].$$

Since (100) is true for $\ell(\vec{\mathbf{i}}) \leq N$, we have

$$T'_{\omega_1} = \sum_{0\leq\alpha\leq\lfloor\mathbf{n}(\vec{\mathbf{i}}_{-\mathbf{i}_{\omega_1}})/2\rfloor} \tilde{w}_{\alpha,\vec{\mathbf{i}}_{-\omega_1}} \cdot \left(\frac{\sqrt{s}^{\|\vec{\mathbf{i}}_{-\omega_1}\|}}{(\ell(\vec{\mathbf{i}}_{-\omega_1}) - |\alpha|)!} H_{\mathbf{n}(\vec{\mathbf{i}}_{-\omega_1})-2\alpha}\left(\frac{W(s)}{\sqrt{s}}\right)\right).$$

Mimicking the proof of (109) and (114) leads to that

$$T_{\omega_1} = \sum_{0\leq\alpha\leq\lfloor\mathbf{n}(\vec{\mathbf{i}}_{-\mathbf{i}_{\omega_1}})/2\rfloor} \tilde{w}_{\alpha,\vec{\mathbf{i}}_{-\omega_1}} \cdot \left(\frac{\sqrt{t}^{\|\vec{\mathbf{i}}\|}}{(\ell(\vec{\mathbf{i}}) - |\alpha|)!} H_{\mathbf{n}(\vec{\mathbf{i}})-2\alpha}\left(\frac{x}{\sqrt{t}}\right)\right). \tag{116}$$

Under the condition $i_{\omega_1,1} = i_{\omega_2,1} \neq 0$, define

$$T_{\omega_1,\omega_2} := \mathbb{E}\left[\int_0^t \prod_{\omega\neq\omega_1,\omega_2} \mathbb{I}_{\mathbf{i}_\omega}(s)\mathbb{I}_{-\mathbf{i}_{\omega_1}}(s)\mathbb{I}_{-\mathbf{i}_{\omega_2}}(s)ds\Big|W(t)=y\right].$$

Using (102) and (109), similarly to the derivation of (116), we have

$$\begin{aligned}T_{\omega_1,\omega_2} &= \sum_{0\leq\alpha\leq\lfloor\mathbf{n}(\vec{\mathbf{i}})/2\rfloor - e_{i_{\omega_1,1}}} \tilde{w}_{\alpha,\vec{\mathbf{i}}_{-\omega_1-\omega_2}} \left(\frac{\sqrt{t}^{\|\vec{\mathbf{i}}\|}}{(\ell(\vec{\mathbf{i}}) - 1 - |\alpha|)!} H_{\mathbf{n}(\vec{\mathbf{i}})-2e_{i_{\omega_1,1}}-2\alpha}\left(\frac{x}{\sqrt{t}}\right)\right)\\ &= \sum_{e_{i_{\omega_1,1}}\leq\alpha\leq\lfloor\mathbf{n}(\vec{\mathbf{i}})/2\rfloor} \tilde{w}_{\alpha-e_{i_{\omega_1,1}},\vec{\mathbf{i}}_{-\omega_1-\omega_2}} \left(\frac{\sqrt{t}^{\|\vec{\mathbf{i}}\|}}{(\ell(\vec{\mathbf{i}}) - |\alpha|)!} H_{\mathbf{n}(\vec{\mathbf{i}})-2\alpha}\left(\frac{x}{\sqrt{t}}\right)\right). \end{aligned} \tag{117}$$

Taking conditional expectation on both sides of (115), and by (116) and (117), we get (100) for $\ell(\vec{\mathbf{i}}) = N + 1$, where the coefficients are recursively defined by (101). $\square$

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Rev. Financ. Stud. 9, 385–426.
Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. J. Finance 54, 1361–1395.
Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.
Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Ann. Statist. 36, 906–937.
Aït-Sahalia, Y., Mykland, .P.A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.
Aït-Sahalia, Y., Yu, J., 2006. Saddlepoint approximations for continuous-time markov processes. J. Econometrics 134, 507–551.
Bakshi, G., Ju, N., Ou-Yang, H., 2006. Estimation of continuous-time models with an application to equity volatility dynamics. J. Financ. Econ. 82, 227–249.
Balduzzi, P., Das, S.R., Foresi, S., Sundaram, R.K., 1996. A simple approach to three factor affine term structure models. J. Fixed Incom 6, 43–53.
Chang, J., Chen, S.X., 2011. On the approximate maximum likelihood estimation for diffusions processes. Ann. Statist. 39, 2820–2851.
Choi, S., 2013. Closed-form likelihood expansions for multivariate time-inhomogeneous diffusions. J. Econometrics 174, 45–65.
Choi, S., 2015. Explicit form of approximate transition probability density functions of diffusion processes. J. Econometrics 187, 57–73.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.
Duan, J.C., 1995. The GARCH option pricing model. Math. Finance 5, 13–32.
Egorov, A.V., Li, H., Xu, Y., 2003. Maximum likelihood estimation of time-inhomogeneous diffusions. J. Econometrics 114, 107–139.
Fan, J., Zhang, C., 2003. A reexamination of diffusion estimators with applications to financial model validation. J. Amer. Statist. Assoc. 98, 118–134.
Filipović, D., Mayerhofer, E., Schneider, P., 2013. Density approximations for multivariate affine jump-diffusion processes. J. Econometrics 176, 93–111.
Friedman, A., 1964. Partial Differential Equations of Parabolic Type. Dover Publications, New York.
Friedman, A., 1975. Stochastic Differential Equations and Applications: Volume 1. Academic Press, New York.
Friedman, A., 1976. Stochastic Differential Equations and Applications: Volume 2. Academic Press, New York.
Hansen, L.P., Sargent, T.J., 1983. The dimensionality of the aliasing problem in models with rational spectral densities. Econometrica 51, 377–387.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6, 327–343.
Kallenberg, O., 2002. Foundations of Modern Probability. Springer, New York.
Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, New York.
Kessler, M., 1997. Estimation of an ergodic diffusion from discrete observations. Scand. J. Statist. 24, 211–229.
Kessler, M., Rahbek, A., 2004. Identification and inference for multivariate cointegrated and ergodic gaussian diffusions. Stat. Inference Stoch. Process. 7, 137–151.
Kloeden, P.E., Platen, E., 1992. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.
Kristensen, D., Mele, A., 2011. Adding and subtracting Black–Scholes: a new approach to approximating derivative prices in continuous-time models. J. Financ. Econ. 102, 390–415.
Lee, Y.D., Song, S., Lee, E.K., 2014. The delta expansion for the transition density of diffusion models. J. Econometrics 178, 694–705.
Li, C., 2013. Maximum-likelihood estimation for diffusion processes via closed-form density expansions. Ann. Statist. 41, 1350–1380.
Li, C., An, Y., Chen, D., Li, Q., Si, N., 2016. Efficicient computation of the likelihood expansions for diffusion models. IIE Trans. 48, 1156–1171.
Li, C., Chen, D., 2016. Estimating jump-diffusions using closed-form likelihood expansions. J. Econometrics 195, 51–70.
Li, C., Li, X., 2015. A closed-form expansion approach for pricing discretely monitored variance swaps. Oper. Res. Lett. 43, 450–455.
Lo, A.W., 1988. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econ. Theory 4, 231–247.
Milstein, G.N., 1975. Approximate integration of stochastic differential equations. Theory Probab. Appl. 19, 557–562.
Nelson, D.B., 1990. ARCH models as diffusion approximations. J. Econometrics 45, 7–38.
Pedersen, A.R., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Statist. 22, 55–71.
Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. J. Finance 52, 1973–2002.
Uchida, M., Yoshida, N., 2012. Adaptive estimation of an ergodic diffusion process based on sampled data. Stochastic Process. Appl. 122, 2885–2924.
Vasicek, O., 1977. An equilibrium characterisation of the term structure. J. Financ. Econ. 5, 177–188.
Watanabe, S., 1987. Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. Ann. Probab. 15, 1–39.
Willink, R., 2005. Normal moments and Hermite polynomials. Statist. Probab. Lett. 73, 271–275.
Xiu, D., 2014. Hermite polynomial based expansion of european option prices. J. Econometrics 179, 158–177.
Yoshida, N., 1992. Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. Probab. Theory Related Fields 92, 275–311.
Yu, J., 2007. Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese yuan. J. Econometrics 141, 1245–1280.