The Annals of Statistics
2013, Vol. 41, No. 3, 1350–1380
DOI: 10.1214/13-AOS1118

MAXIMUM-LIKELIHOOD ESTIMATION FOR DIFFUSION PROCESSES VIA CLOSED-FORM DENSITY EXPANSIONS

BY CHENXU LI[1]

Guanghua School of Management, Peking University

This paper proposes a widely applicable method of approximate maximum-likelihood estimation for multivariate diffusion process from discretely sampled data. A closed-form asymptotic expansion for transition density is proposed and accompanied by an algorithm containing only basic and explicit calculations for delivering any arbitrary order of the expansion. The likelihood function is thus approximated explicitly and employed in statistical estimation. The performance of our method is demonstrated by Monte Carlo simulations from implementing several examples, which represent a wide range of commonly used diffusion models. The convergence related to the expansion and the estimation method are theoretically justified using the theory of Watanabe [*Ann. Probab.* **15** (1987) 1–39] and Yoshida [*J. Japan Statist. Soc.* **22** (1992) 139–159] on analysis of the generalized random variables under some standard sufficient conditions.

1. Introduction. Diffusion processes governed by stochastic differential equations (hereafter SDE) are widely used in describing the phenomenon of random fluctuations over time, and even become indispensable for analyzing high-frequency data; see, for example, Mykland and Zhang [52]. Practical application of diffusion models calls for statistical inference based on discretely monitored data. The literature has seen a wide spectrum of asymptotically efficient estimation methods, for example, those based on various contrast functions proposed in Yoshida [69], Kessler [39], Kessler and Sørensen [40] and the references given in Sørensen [59]. Taking the efficiency, feasibility and generality into account, maximum-likelihood estimation (hereafter MLE) can be a choice among others. However, for the increasingly complex real-world dynamics, likelihood functions (transition densities) are generally not known in closed-form and thus involve significant challenges in valuation. This leads to various methods of approximation and the resulting approximate MLE. The focus of this paper is to propose a widely applicable closed-form asymptotic expansion for transition density and thus to apply it in approximate MLE for multivariate diffusion process.

Received November 2011; revised April 2013.

[1]This research was supported by the Guanghua School of Management, the Center for Statistical Sciences, and the Key Laboratory of Mathematical Economics and Quantitative Finance (Ministry of Education) at Peking University, as well as the National Natural Science Foundation of China (Project 11201009).

MSC2010 subject classifications. Primary 62F12, 62M05; secondary 60H10, 60J60, 60H30.

Key words and phrases. Asymptotic expansion, diffusion, discrete observation, maximum-likelihood estimation, transition density.

1.1. *Background*. To approximate likelihood functions, Yoshida [69] proposed to discretize continuous likelihood functions (see, e.g., Basawa and Prakasa Rao [13]); many others focused on direct approximation of likelihood functions (transition densities) for discretely monitored data, see surveys in, for example, Phillips and Yu [56], Jensen and Poulsen [36], Hurn, Jeisman and Lindsay [34] and the references therein. In particular, among various numerical methods, Lo [46] proposed to employ a numerical solution of Kolmogorov equation for transition density; Pedersen [55], Brandt and Santa-Clara [21], Durham and Gallant [25], Stramer and Yan [60], Beskos and Roberts [17], Beskos et al. [16], Beskos, Papaspiliopoulos and Roberts [15] and Elerian, Chib and Shephard [28] advocated the application of various Monte Carlo simulation methods; Yu and Phillips [73] developed an exact Gaussian method for models with a linear drift function; Jensen and Poulsen [36] resorted to the techniques of binomial trees. Since all these numerical methods are computationally demanding, real-world implementation has necessitated the development of analytical methods for efficiently approximating transition density. An adhoc approach is to approximate the model by discretization, for example, the Euler scheme, and then use the transition density of the discretized model. Elerian [27] refined such an approximation via the second order Milstein scheme. Kessler [39] and Uchida and Yoshida [63] employed a more sophisticated normal-distribution-based approximation via higher order expansions of the mean and variance.

For approximate MLE of diffusions, Dacunha-Castelle and Florens-Zmirou [23] is one of the earliest attempts to apply the idea of small-time expansion of transition densities, which in principle can be made arbitrarily accurate. However, their method relies on implicit representation of moments of Brownian bridge functionals, and thus requires Monte Carlo simulation in implementation. A milestone is the ground-breaking work of Aït-Sahalia [1–3], which established the theory of Hermite-polynomial-based analytical expansion for transition density of diffusion models and the corresponding approximate MLE. Along the line of Aït-Sahalia [1–3], a number of substantial refinements and applications emerged in the literature of likelihood-based statistical inference (see surveys in Aït-Sahalia [4]); see, for example, Bakshi and Ju [11], Bakshi, Ju and Ou-Yang [12], Aït-Sahalia and Mykland [7, 8], Aït-Sahalia and Kimmel [5, 6], Li [45], Egorov, Li and Xu [26], Schaumburg [58], Aït-Sahalia and Yu [9], Yu [72], Filipović, Mayerhofer and Schneider [31], Tang and Chen [61], Xiu [66] and Chang and Chen [22].

1.2. *Expansion for likelihood functions and approximate MLE*. Starting from the celebrated Edgeworth expansion for distribution of standardized summation of independently identically distributed random variables (see, e.g., Chapter XVI in Feller [30], Chapter 2 in Hall [33] and Chapter 5 in McCullagh [47]), asymptotic expansions have become powerful tools for statistics, econometrics and many other disciplines in science and technology. Taking dependence of random variables into

account, Mykland [48–51] established the theory, calculation and various statistical applications of martingale expansion, which is further developed in Yoshida [70, 71].

Having an analogy with these Edgeworth-type expansions and motivated by MLE for diffusion processes, I propose a new small-time asymptotic expansion of transition density for multivariate diffusions based on the theory of Watanabe [65] and Yoshida [67, 68]. However, in contrast to the traditional Edgeworth expansions, our expansion does not require the knowledge of generally implicit moments, cumulants or characteristic function of the underlying variable, and thus it is applicable to a wide range of diffusion processes. Moreover, in analogy to the verification of validity given in, for example, Bhattacharya and Ghosh [18], Mykland [48–51] and Yoshida [70, 71] for Edgeworth type expansions, the uniform convergence rate (with respect to various parameters) of our density expansion is proved under some sufficient conditions on the drift and diffusion coefficients of the underlying diffusion using the theory of Watanabe [65] and Yoshida [67, 68]. Consequently, the approximate MLE converges to the true one, and thus inherits its asymptotic properties. Such results are further demonstrated through numerical tests and Monte Carlo simulations for some representative examples.

In comparison to the expansion proposed by Aït-Sahalia [1–3], our method is able to bypass the challenge resulting from the discussion of reducibility, the explicity of the *Lamperti* transform (see, e.g., Section 5.2 in Karatzas and Shreve [38]) and its inversion, as well as the iterated equations for expressing correction terms, which in general lead to multidimensional integrals; see Bakshi, Ju and Ou-Yang [12]. Thus it renders an algorithm for practically obtaining a closed-form expansion (without integrals and implicit transforms) for transition density up to any arbitrary order, which serves as a widely applicable tool for approximate MLE. Even after the *Lamperti* transform, our expansion employs a completely different nature comparing with those proposed in Aït-Sahalia [1–3], which hinge on expansions in an orthogonal basis consisting of Hermite polynomials and expansions of each coefficient expressed by an expectation of a smooth functional of the transformed variable via an iterated Dynkin formula; see Section 4 in Aït-Sahalia [2].

Moreover, our method is different from the existing theory of large-deviations-based expansions, which were discussed in, for example, Azencott [10], Bismut [20], Ben Arous [14] and Léandre [43], and given probabilistic representation in Watanabe [65] for the purpose of investigating the analytical structure of heat kernel in differential geometry. Large-deviations-based asymptotic expansions involve Riemannian distance (implied by the true but generally unknown transition density) and higher order correction terms. Except for some special cases, they rarely admit closed-form expressions by solving the corresponding variational problems. However, for practical implementation of statistical estimation, relatively simple closed-form approximations are usually favorable.

The rest of this paper is organized as follows. In Section 2, the model is introduced with some technical assumptions and the maximum-likelihood estimation problem is formulated. In Section 3, the transition density expansion is proposed with closed-form correction terms of any arbitrary order for general multivariate diffusion processes, and the uniform convergence of the expansion is established. In Section 4, numerical performance of the density expansion is demonstrated through examples. In Section 5, the asymptotic properties of the consequent approximate MLE are established. In Section 6, Monte Carlo evidence for the approximate MLE is provided. In Section 7, the paper is concluded and some opportunities for future research are outlined. Appendix A provides an algorithm for explicitly calculating a type of conditional expectation, which plays an important role in the closed-form expansion. Appendix B contains all proofs. The supplementary material [44] collects some concrete formulas for illustration, figures for exhibiting detailed numerical performance, additional and alternative output of simulation results, more examples, a brief introduction to the theory of Watanabe–Yoshida and the proof of a technical lemma.

2. The model and maximum-likelihood estimation. Assuming known parametric form of the drift vector function $\mu = (\mu_1, \ldots, \mu_m) : \mathbb{R}^m \to \mathbb{R}^m$ and the dispersion matrix $\sigma = (\sigma_{ij})_{m\times d}: \mathbb{R}^m \to \mathbb{R}^{m\times d}$ with unknown parameter θ belonging to a compact set $\Theta \subset \mathbb{R}^k$, an m-dimensional time-homogenous diffusion X is modeled by an SDE,

$$dX(t) = \mu\big(X(t);\theta\big)\,dt + \sigma\big(X(t);\theta\big)\,dW(t), \qquad X(0) = x_0, \tag{2.1}$$

where $\{W(t)\}$ is a d-dimensional standard Brownian motion. Let $E \subset \mathbb{R}^m$ denote the state space of X. Without loss of generality, we assume $m = d$ throughout the paper.

By the time-homogeneity nature of diffusion X, let $p_X(\Delta, x|x_0;\theta)$ denote the conditional density of $X(t+\Delta)$ given $X(t) = x_0$, that is,

$$\mathbb{P}\big(X(t+\Delta) \in dx|X(t) = x_0\big) = p_X(\Delta, x|x_0;\theta)\,dx.$$

Based on the discrete observations of X at time grids $\{\Delta, 2\Delta, \ldots, n\Delta\}$, which correspond to the daily, weekly or monthly monitoring, etc., the likelihood function is constructed as

$$l_n(\theta) = \prod_{i=1}^{n} p_X\big(\Delta t, X(i\Delta)|X\big((i-1)\Delta\big);\theta\big); \tag{2.2}$$

the corresponding log-likelihood function admits the following form:

$$\ell_n(\theta) = \sum_{i=1}^{n} L_i(\theta), \tag{2.3}$$

where the log transition density is

$$L_i(\theta) = \log\big[p_X\big(\Delta, X(i\Delta)|X\big((i-1)\Delta\big);\theta\big)\big]. \tag{2.4}$$

Maximum-likelihood estimation is to identify the optimizer in $\theta \in \Theta$ for (2.2) or equivalently (2.3). However, except for some simple models, (2.2) and (2.3) rarely admit closed-form expressions.

For ease of exposition, we introduce some technical assumptions. Let $A(x;\theta) = \sigma(x;\theta)\sigma(x;\theta)^T$ denote the diffusion matrix.

ASSUMPTION 1. The diffusion matrix $A(x;\theta)$ is positive definite, that is, $\det A(x;\theta) > 0$, for any $(x,\theta) \in E \times \Theta$.

ASSUMPTION 2. For each integer $k \geq 1$, the kth order derivatives in x of the functions $\mu(x;\theta)$ and $\sigma(x;\theta)$ exist, and they are uniformly bounded for any $(x,\theta) \in E \times \Theta$.

ASSUMPTION 3. The transition density $p_X(\Delta, x|x_0;\theta)$ is continuous in $\theta \in \Theta$, and the log-likelihood function (2.3) admits a unique maximizer in the parameter set Θ.

Assumptions 1 and 2 are conventionally proposed in the study of stochastic differential equations; see, for example, Ikeda and Watanabe [35]. They are sufficient (but not necessary) to guarantee the existence and uniqueness of the solution and other desirable technical properties. For convenience, the theoretical proofs given in Appendix B are based on these conditions. However, as is shown in Sections 4 and 6, numerical examples suggest that the method proposed in this paper is applicable to a wide range of commonly used models, rather than confined to those strictly satisfying these sufficient (but not necessary) conditions. Assumption 3 collects two standard conditions for maximum likelihood estimation. In particular, for the continuity (and higher differentiability) of the transition density in the parameter, sufficient conditions based on the smoothness of the drift and dispersion functions can be found in, for example, Azencott [10] and Aït-Sahalia [2]. Theoretical relaxation of these conditions may involve case-by-case treatment and standard approximation argument, which is beyond the scope of this paper and can be regarded as a future research topic.

3. A closed-form expansion for transition density. The method of approximate maximum-likelihood estimation proposed in this paper relies on a closed-form expansion for transition density of any arbitrary diffusion process. Bypassing the discussion of the *Lamperti* transform and the reducibility issue as in Aït-Sahalia [1–3], our starting point stands on the fact that the transition density can be expressed as

$$p_X(\Delta, x|x_0;\theta) = \mathbb{E}\big[\delta\big(X(\Delta) - x\big)|X(0) = x_0\big], \tag{3.1}$$

where $\delta(z)$ is the Dirac Delta function centered at 0 for some variable z. More precisely, $\delta(z)$ is defined as a generalized function (distribution) such that it is zero for all values of z except when it is zero, and its integral from $-\infty$ to ∞ is equal to one; see, for example, Kanwal [37] for more details. Watanabe [65] established the validity of (3.1) through the theory of generalized random variables and expressed correction terms of large-deviations-based density expansion as implicit expectation forms by separately treating the cases of diagonal ($x = x_0$) and off-diagonal ($x \neq x_0$). In particular, the off-diagonal ($x \neq x_0$) expansion depends on a generally implicit variational formulation for Riemanian distance. From the viewpoint of statistical applications where $X(\Delta) \neq X(0)$ (corresponding to $x \neq x_0$) happens almost surely, the expansion proposed in Watanabe [65] is impractical due to high computational costs. In the literature of statistical inference, (3.1) has been employed in Pedersen [55] for simulation-based approximate MLE. In this section, we propose a new expansion of the transition density which universally treats the diagonal ($x = x_0$) and off-diagonal ($x \neq x_0$) cases. Heuristically speaking, our method hinges on a Taylor-like expansion of a standardized version of $\delta(X(\Delta) - x)$, which results in closed-form formulas for any arbitrary correction term.

3.1. *Basic setup and notation.* Let $\epsilon = \sqrt{\Delta}$ be a small parameter based on which an asymptotic expansion is carried out. By rescaling the model (2.1) to bring forth finer local behavior of the diffusion process, we let $X^\epsilon(t) := X(\epsilon^2 t)$. Integral substitution and the Brownian scaling property yield that

$$dX^\epsilon(t) = \epsilon^2 \mu\big(X^\epsilon(t); \theta\big)\, dt + \epsilon \sigma\big(X^\epsilon(t); \theta\big)\, dW^\epsilon(t), \qquad X^\epsilon(0) = x_0, \tag{3.2}$$

where $\{W^\epsilon(t)\}$ is a m-dimensional standard Brownian motion. For notation simplicity, we write the scaled Brownian motion $W^\epsilon(t)$ as $W(t)$ and drop the parameter θ in what follows.

Let us introduce a vector function $b(x) = (b_1(x), b_2(x), \ldots, b_m(x))^T$ defined by

$$b_i(x) = \mu_i(x) - \frac{1}{2} \sum_{k=1}^{m} \sum_{j=1}^{m} \sigma_{kj}(x) \frac{\partial}{\partial x_k} \sigma_{ij}(x) \tag{3.3}$$

and construct the following differential operators:

$$\mathcal{A}_0 := \sum_{i=1}^{m} b_i(x) \frac{\partial}{\partial x_i} \quad \text{and} \quad \mathcal{A}_j := \sum_{i=1}^{m} \sigma_{ij}(x) \frac{\partial}{\partial x_i} \qquad \text{for } j = 1, \ldots, m, \tag{3.4}$$

which map vector-valued functions to vector-valued functions of the same dimension, respectively. More precisely, for any $\nu \in \mathbb{N}$ and a ν-dimensional vector-valued function $\varphi(x) = (\varphi_1(x), \varphi_2(x), \ldots, \varphi_\nu(x))^T$,

$$\big(\mathcal{A}_0(\varphi)\big)(x) = \left(\sum_{i=1}^{m} b_i(x) \frac{\partial \varphi_1(x)}{\partial x_i}, \sum_{i=1}^{m} b_i(x) \frac{\partial \varphi_2(x)}{\partial x_i}, \ldots, \sum_{i=1}^{m} b_i(x) \frac{\partial \varphi_\nu(x)}{\partial x_i} \right)^T$$

and

$$(\mathcal{A}_j(\varphi))(x) = \left(\sum_{i=1}^{m} \sigma_{ij}(x) \frac{\partial \varphi_1(x)}{\partial x_i}, \sum_{i=1}^{m} \sigma_{ij}(x) \frac{\partial \varphi_2(x)}{\partial x_i}, \ldots, \sum_{i=1}^{m} \sigma_{ij}(x) \frac{\partial \varphi_v(x)}{\partial x_i}\right)^T$$

for $j = 1, 2, \ldots, m$.

For an index $\mathbf{i} = (i_1, \ldots, i_n) \in \{0, 1, 2, \ldots, m\}^n$ and a right-continuous stochastic process $\{f(t)\}$, define an iterated Stratonovich integral with integrand f as

$$J_{\mathbf{i}}[f](t) := \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} f(t_n) \circ dW_{i_n}(t_n) \cdots \circ dW_{i_2}(t_2) \circ dW_{i_1}(t_1), \tag{3.5}$$

where $\circ$ denotes stochastic integral in the Stratonovich sense. Note that $J_{\mathbf{i}}[f](t)$ is recursively defined from inside to outside; see page 174 of Kloeden and Platen [41]. For ease of exposition, the order of iterated integrations defined in this paper is the reverse of that in Kloeden and Platen [41] for any arbitrary index. To lighten the notation, for $f \equiv 1$, the integral $J_{\mathbf{i}}[1](t)$ is abbreviated to $J_{\mathbf{i}}(t)$. By convention, let $W_0(t) := t$ and define

$$\|\mathbf{i}\| := \sum_{k=1}^{n} [2 \cdot 1_{\{i_k = 0\}} + 1_{\{i_k \neq 0\}}] \tag{3.6}$$

as a "norm" of index $\mathbf{i}$, which counts an index k with $i_k = 0$ twice.

By viewing $X^\epsilon(1)$ as a function of ϵ, it is natural to obtain a pathwise expansion in ϵ with random coefficients, which serves as a foundation for our transition density expansion. According to Watanabe [65], I introduce the following coefficient function $C_{\mathbf{i}}(x_0)$ defined by iterative application of the differential operators (3.4):

$$C_{\mathbf{i}}(x_0) := \mathcal{A}_{i_n}(\cdots(\mathcal{A}_{i_3}(\mathcal{A}_{i_2}(\sigma_{\cdot i_1}))) \cdots)(x_0) \tag{3.7}$$

for an index $\mathbf{i} = (i_1, \ldots, i_n)$. Here, for $i_1 \in \{1, 2, \ldots, m\}$, the vector $\sigma_{\cdot i_1}(x) = (\sigma_{1 i_1}(x), \ldots, \sigma_{m i_1}(x))^T$ denotes the i_1th column vector of the dispersion matrix $\sigma(x)$, for $i_1 = 0$, $\sigma_{\cdot 0}(x)$ refers to the vector $b(x)$ defined in (3.3).

Using vector function (3.3), the scaled diffusion (3.2) can be equivalently expressed as the following stochastic differential equation in the Stratonovich sense (see, e.g., Section 3.3 in Karatzas and Shreve [38]), that is,

$$dX^\epsilon(t) = \epsilon^2 b\big(X^\epsilon(t)\big)\, dt + \epsilon \sigma\big(X^\epsilon(t)\big) \circ dW(t).$$

Thus, similarly to Theorem 3.3 in Watanabe [65], it is easy to obtain a closed-form pathwise expansion of $X^\epsilon(1)$ from successive applications of the Itô formula.

LEMMA 1. *$X^\epsilon(1)$ admits the following pathwise asymptotic expansion*:

$$X^\epsilon(1) = \sum_{k=0}^{J} F_k \epsilon^k + \mathcal{O}\big(\epsilon^{J+1}\big) \tag{3.8}$$

for any $J \in N$. Here, $F_0 = x_0$ and F_k can be written as a closed-form linear combination of iterated Stratonovich integrals, that is,

$$F_k = \sum_{\|\mathbf{i}\|=k} C_{\mathbf{i}}(x_0) J_{\mathbf{i}}(1) \tag{3.9}$$

for $k = 1, 2, \ldots$, where the integral $J_{\mathbf{i}}(1)$, the norm $\|\mathbf{i}\|$ and coefficient $C_{\mathbf{i}}(x_0)$ are defined in (3.5), (3.6) *and* (3.7), *respectively.*

For any arbitrary dimension $r = 1, 2, \ldots, m$, one has the element-wise form of the expansion (3.8) as $X_r^\epsilon(1) = \sum_{k=0}^J F_{k,r}\epsilon^k + \mathcal{O}(\epsilon^{J+1})$ where

$$F_{k,r} = \sum_{\|\mathbf{i}\|=k} C_{\mathbf{i},r}(x_0) J_{\mathbf{i}}(1) \tag{3.10}$$

with

$$C_{\mathbf{i},r}(x_0) := \mathcal{A}_{i_n}\big(\cdots\big(\mathcal{A}_{i_3}\big(\mathcal{A}_{i_2}(\sigma_{ri_1})\big)\big)\cdots\big)(x_0)$$

for $\mathbf{i} = (i_1, \ldots, i_n)$. Note that (3.8) is different from the Wiener chaos decomposition (see, e.g., Nualart [53]), which employs an alternative way of representing random variables. The validity of the pathwise expansion (3.8) and other expansions introduced in the next subsection can be rigorously guaranteed by the theory of Watanabe [65] and Yoshida [67, 68]. For ease of exposition, we focus on the derivation of density expansion in this and the following subsection and articulate the validity issue in Section 3.3.

We introduce an m-dimensional correlated Brownian motion

$$B(t) = \big(B_1(t), B_2(t), \ldots, B_m(t)\big) \qquad \text{with } B_k(t) = \frac{\sum_{j=1}^m \sigma_{kj}(x_0) W_j(t)}{\sqrt{\sum_{j=1}^m \sigma_{kj}^2(x_0)}} \tag{3.11}$$

for $k = 1, 2, \ldots, m$. Thus, the leading term F_1 can be expressed as

$$F_1 = \left(\sqrt{\sum_{j=1}^m \sigma_{1j}^2(x_0)}B_1(1), \sqrt{\sum_{j=1}^m \sigma_{2j}^2(x_0)}B_2(1), \ldots, \sqrt{\sum_{j=1}^m \sigma_{mj}^2(x_0)}B_m(1)\right).$$

Let $D(x)$ be a diagonal matrix defined by

$$D(x) := \operatorname{diag}\left(\frac{1}{\sqrt{\sum_{j=1}^m \sigma_{1j}^2(x)}}, \frac{1}{\sqrt{\sum_{j=1}^m \sigma_{2j}^2(x)}}, \ldots, \frac{1}{\sqrt{\sum_{j=1}^m \sigma_{mj}^2(x)}}\right). \tag{3.12}$$

It follows that $B(t) = D(x_0)\sigma(x_0)W(t)$ and $D(x_0)F_1 = B(1)$. Furthermore, the correlation of $B_k(t)$ and $B_l(t)$ for $k \neq l$ is given by

$$\rho_{kl}(x_0) := \operatorname{Corr}\big(B_k(t), B_l(t)\big) = \frac{\sum_{j=1}^m \sigma_{kj}(x_0)\sigma_{lj}(x_0)}{\sqrt{\sum_{j=1}^m \sigma_{kj}^2(x_0)}\sqrt{\sum_{j=1}^m \sigma_{lj}^2(x_0)}}.$$

So, the covariance matrix of $B(1)$ is

$$\Sigma(x_0) = \big(\rho_{ij}(x_0)\big)_{m\times m} = D(x_0)\sigma(x_0)\sigma(x_0)^T D(x_0). \tag{3.13}$$

It follows that Assumption 1 is equivalent to the positive definite property of the correlation matrix $\Sigma(x_0)$ and the nonsingularity of the dispersion matrix $\sigma(x_0)$, that is, $\det A(x_0) > 0 \Longleftrightarrow \det \Sigma(x_0) > 0 \Longleftrightarrow \det \sigma(x_0) > 0$. Finally, for any index $i \in \{1, 2, \ldots, m\}$ and differentiable function $u(y)$ with $y \in \mathbb{R}^m$, we introduce the following differential operator:

$$\mathcal{D}_i u(y) := \frac{\partial u(y)}{\partial y_i} - u(y)\big(\Sigma(x_0)^{-1} y\big)_i, \tag{3.14}$$

where $(\Sigma(x_0)^{-1}y)_i$ denotes the ith element of the vector $\Sigma(x_0)^{-1}y$.

3.2. *Asymptotic expansion for transition densities: A general framework.* Employing the scaled diffusion $X^\epsilon(t) = X(\epsilon^2 t)$ with $\epsilon = \sqrt{\Delta}$, the expectation representation (3.1) for transition density can be expressed as

$$p_X(\Delta, x|x_0; \theta) = \mathbb{E}\big[\delta\big(X^\epsilon(1) - x\big)|X^\epsilon(0) = x_0\big]. \tag{3.15}$$

To guarantee the convergence, our expansion procedure begins with standardizing $X^\epsilon(1)$ to

$$Y^\epsilon := D(x_0)\frac{X^\epsilon(1) - x_0}{\epsilon} = D(x_0)\frac{X^\epsilon(1) - x_0}{\sqrt{\Delta}}, \tag{3.16}$$

which converges to a nonconstant random variable (a multivariate normal in our case), see Watanabe [65] and Yoshida [67, 68] for a similar setting. Indeed, based on the Brownian motion defined in (3.11) and the fact $D(x_0)F_1 = B(1)$, the jth component of $Y^\epsilon(1)$ satisfies that

$$Y_j^\epsilon := \frac{X_j^\epsilon(1) - x_{0j}}{\epsilon\sqrt{\sum_{i=1}^d \sigma_{ji}^2(x_0)}} \to B_j(1) \qquad \text{as } \epsilon \to 0 \tag{3.17}$$

for $j = 1, 2, \ldots, m$. It is worth noting that Watanabe [65] employed an alternative standardization method (see Theorem 3.7 in Watanabe [65]) in constructing the implicit expectation representation for the correction terms of large-deviations-based density expansion for the case of $x \neq x_0$; see Theorem 3.8 in Watanabe [65].

Owing to (3.16), the pathwise expansion (3.8) implies that

$$Y^\epsilon = \sum_{i=0}^{J} Y_i \epsilon^i + \mathcal{O}\big(\epsilon^{J+1}\big) \qquad \text{with } Y_i = D(x_0)F_{i+1} \tag{3.18}$$

for any $J \in \mathbb{N}$. Thus, based on (3.15), a Jacobian transform resulting from the change of variable in (3.16) yields the following representation of the density of $X^\epsilon(1)$ based on that of Y^ϵ, that is,

$$p_X(\Delta, x|x_0; \theta) = \left(\frac{1}{\sqrt{\Delta}}\right)^m \det D(x_0)\mathbb{E}\big[\delta\big(Y^\epsilon - y\big)|X(0) = x_0\big],$$

where $y = D(x_0)(x - x_0)/\sqrt{\Delta}$. For ease of exposition, the initial condition $X(0) = x_0$ is omitted in what follows. So, the key task is to develop an asymptotic expansion for $\mathbb{E}[\delta(Y^\epsilon - y)]$ around $\epsilon = 0$.

Based on the theory of Watanabe [65] and Yoshida [67, 68], the Dirac Delta function can be manipulated as a function for many purposes, though it can be formally defined as a distribution. Based on the expansion of Y^ϵ and heuristic application of classical rule for differentiating composite functions [the Dirac Delta function $\delta(\cdot - y)$ acting on Y^ϵ as a function of ϵ], one is able to obtain a Taylor expansion of $\delta(Y^\epsilon - y)$ as

$$\delta\big(Y^\epsilon - y\big) = \sum_{k=0}^{J} \Phi_k(y)\epsilon^k + \mathcal{O}\big(\epsilon^{J+1}\big) \tag{3.19}$$

for any $J \in \mathbb{N}$, where $\Phi_k(y)$ represents the coefficient of the kth expansion term. Thus, the following expansion is immediately implied:

$$\mathbb{E}\big[\delta\big(Y^\epsilon - y\big)\big] := \sum_{k=0}^{J} \Omega_k(y)\epsilon^k + \mathcal{O}\big(\epsilon^{J+1}\big), \tag{3.20}$$

where $\Omega_k(y) := \mathbb{E}\Phi_k(y)$ will be explicitly derived and the remainder term is interpreted in the sense of classical calculus. Thus, the approximate transition density for X up to the Jth order is proposed as

$$\begin{aligned} p_X^{(J)}(\Delta, x|x_0;\theta) &:= \left(\frac{1}{\epsilon}\right)^m \det D(x_0) \sum_{k=0}^{J} \Omega_k\left(D(x_0)\frac{x - x_0}{\epsilon}\right)\epsilon^k \\ &= \left(\frac{1}{\sqrt{\Delta}}\right)^m \det D(x_0) \sum_{k=0}^{J} \Omega_k\left(D(x_0)\frac{x - x_0}{\sqrt{\Delta}}\right)\Delta^{k/2}. \end{aligned} \tag{3.21}$$

The convergence of this expansion (guaranteed by the theory of Watanabe [65] and Yoshida [67, 68]) will be discussed in Section 3.3.

As outlined in the whole framework, our idea naturally originates from pathwise expansion of a standardized random variable. However, explicit calculation of the correction terms Ω_k is still a challenging issue. In what follows, we will give a general closed-form formula. Based on (3.17), (3.18), (3.19) and (3.20), it is straightforward to find the leading term as

$$\begin{aligned} \Omega_0(y) &= \mathbb{E}\big[\delta(Y_0 - y)\big] = \mathbb{E}\big[\delta\big(B(1) - y\big)\big] \\ &= \phi_{\Sigma(x_0)}(y) := \frac{\exp(-y^T \Sigma(x_0)^{-1} y/2)}{(2\pi)^{m/2}(\det \Sigma(x_0))^{1/2}}, \end{aligned} \tag{3.22}$$

where $\Sigma(x_0)$ is defined in (3.13).

To express $\Omega_k(y)$ for arbitrary $k \in \mathbb{N}$, we introduce an index set

$$\begin{aligned} S_k = \big\{&(l, \mathbf{r}(l), \mathbf{j}(l))|l = 1, 2, \ldots, \mathbf{r}(l) = (r_1, r_2, \ldots, r_l) \in \{1, 2, \ldots, m\}^l, \\ &\mathbf{j}(l) = (j_1, j_2, \ldots, j_l) \text{ with } j_i \geq 1 \text{ and } j_1 + j_2 + \cdots + j_l = k\big\}. \end{aligned} \tag{3.23}$$

As building blocks, let $P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(z)$ denote a multivariate function in $z = (z_1, z_2, \ldots, z_m) \in \mathbb{R}^m$ defined by the conditional expectation of multiplication of iterated Stratonovich integrals with arbitrary indices $\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_l$, that is,

$$P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(z) := \mathbb{E}\Bigg(\prod_{\omega=1}^{l} J_{\mathbf{i}_\omega}(1) | W(1) = z\Bigg), \tag{3.24}$$

which can be explicitly calculated as a multivariate polynomial according to an effective algorithm proposed in Appendix A.

Now, we will give an explicit formula for obtaining any arbitrary correction term $\Omega_k(y)$ under any arbitrary multivariate diffusion process in the following proposition, which can be implemented using only basic and explicit calculations in any symbolic software package, for example, Mathematica.

THEOREM 1. *For any $k \in \mathbb{N}$, the correction term $\Omega_k(y)$ in (3.21) admits the following explicit expression*:

$$\Omega_k(y) = \Bigg(\sum_{(l,\mathbf{r}(l),\mathbf{j}(l))\in S_k} Q_{(l,\mathbf{r}(l),\mathbf{j}(l))}(y)\Bigg)\phi_{\Sigma(x_0)}(y), \tag{3.25}$$

where $Q_{(l,\mathbf{r}(l),\mathbf{j}(l))}(y)$ is a polynomial explicitly calculated from

$$\begin{aligned} &Q_{(l,\mathbf{r}(l),\mathbf{j}(l))}(y) \\ &\quad = \frac{(-1)^l}{l!} \sum_{\{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)\,|\,\|\mathbf{i}_\omega\|=j_\omega+1,\omega=1,2,\ldots,l\}} \prod_{\omega=1}^{l} \big[C_{\mathbf{i}_\omega,r_\omega}(x_0) D_{r_\omega r_\omega}(x_0)\big] \\ &\qquad \times \mathcal{D}_{r_1}\big(\mathcal{D}_{r_2}\big(\cdots \mathcal{D}_{r_l}\big(P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}\big(\sigma(x_0)^{-1} D(x_0)^{-1} y\big)\big)\cdots\big)\big) \end{aligned} \tag{3.26}$$

for the index $(l, \mathbf{r}(l), \mathbf{j}(l)) = (l, (r_1, r_2, \ldots, r_l), (j_1, j_2, \ldots, j_l)) \in S_k$. Here, S_k, $\phi_{\Sigma(x_0)}(y)$, $\|\cdot\|$, $C_{\mathbf{i}_\omega,r_\omega}(x_0)$, $D_{r_\omega r_\omega}(x_0)$, $\mathcal{D}_{r_i}$ and $P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(\cdot)$ are defined in (3.23), (3.22), (3.6), (3.7), (3.12), (3.14) and (3.24), respectively.

PROOF. See Appendix B. □

An algorithm for explicitly calculating conditional expectation (3.24), which plays an important role in completing the closed-form correction terms as proposed in Theorem 1, is given in Appendix A. Regardless of the dimension of diffusion processes, I concretely exemplify the closed-form expression (3.25) by the first three correction terms in the supplementary material [44]. With Ω_k given by (3.25), a closed-form expansion for transition density can be constructed via (3.21).

3.3. *Convergence of the expansion.* In this subsection, we establish the uniform convergence of the asymptotic expansion (3.21), which will serve as an important building block for the asymptotic properties of approximate maximum-likelihood estimation discussed in Section 5. Theoretically speaking, unlike the Hermite-polynomial-based method in Aït-Sahalia [1–3], which allows justification of convergence as more correction terms are added, our new method is a Taylor-like asymptotic expansion, which is established in the neighborhood of $\Delta = 0$. However, as demonstrated in the numerical experiments and Monte Carlo evidence in Sections 4 and 6, respectively, accuracy of the expansion is enhanced as J increases while holding Δ fixed. Based on the theory of Watanabe [65] and Yoshida [67, 68], the following result implies uniform convergence of our asymptotic expansion of transition density jointly in the whole state space E for the forward variable x, the whole set Θ for the parameter θ, and an arbitrary compact subset $K \subset E$ for the backward variable x_0.

THEOREM 2. *Under the Assumptions 1 and 2, the transition density expansion* (3.21) *satisfies*

$$\sup_{(x,x_0,\theta)\in E\times K\times\Theta} \left|p_X^{(J)}(\Delta, x|x_0;\theta) - p_X(\Delta, x|x_0;\theta)\right| = \mathcal{O}\left(\Delta^{(J+1-m)/2}\right) \tag{3.27}$$

as $\Delta \to 0$ *for* $J \geq m$.

PROOF. See Appendix B. □

It deserves to note that (3.27) gives a theoretical (not necessarily tight) upper bound estimate of the uniform approximation error of $p_X^{(J)}(\Delta, x|x_0;\theta) - p_X(\Delta, x|x_0;\theta)$. The effects of dimensionality can be seen as resulting from the multiplier $\Delta^{-m/2}$ in the expansion (3.21), which leads to the error magnitude $\Delta^{(J+1-m)/2}$. When J is taken sufficiently large as $J \geq m$, the uniform error is controlled by taking $\Delta \to 0$.

4. Numerical performance of density approximation. In this section, we employ three representative and analytically tractable examples (the mean-reverting Ornstein–Uhlenbeck process, the Feller square root process and the double mean-reverting Ornstein–Uhlenbeck process) with explicitly known transition densities to demonstrate the numerical performance of the transition density asymptotic expansion proposed in Section 3. For all of the examples investigated in this and the subsequent sections, we provide the first several expansion terms calculated from the general formula (3.22) and (3.25) in the supplementary material [44]. Higher order correction terms involved in the numerical implementation are documented in the form of Mathematica notebook, which will be provided upon request. The density expansions will be used in Monte Carlo analysis for approximate maximum likelihood estimation in Section 6.

The mean-reverting Ornstein–Uhlenbeck process (also known as the Vasicek model in financial applications) labeled as MROU is specified as:

MODEL 1. The MROU (mean-reverting Ornstein–Uhlenbeck) model,

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

The Gaussian nature of the MROU model renders a closed-form transition density, which serves as a benchmark for explicit comparison with our asymptotic expansion approximations. In the numerical experiments, we choose a parameter set $\kappa = 0.5$, $\alpha = 0.06$ and $\sigma = 0.03$ similar to those employed in Aït-Sahalia [2].

The Feller square root process (also known as the Cox–Ingersoll–Ross model in financial applications) labeled as SQR is specified as:

MODEL 2. The SQR (Feller's square root) model,

$$dX(t) = \kappa\big(\alpha - X(t)\big)\,dt + \sigma\sqrt{X(t)}\,dW(t).$$

The combination of the mean-reverting feature and the Bessel nature (see, e.g., Chapter XI in Revuz and Yor [57]) renders closed-form transition densities. In particular, we concentrate on the case where zero is unattainable, that is, the Feller condition $2\kappa\alpha - \sigma^2 > 0$ holds; see Feller [29]. In the numerical experiments, we choose a parameter set $\kappa = 0.5$, $\alpha = 0.06$ and $\sigma = 0.15$ similar to those employed in Aït-Sahalia [2].

We recall that, for the one-dimensional diffusions investigated in Aït-Sahalia [1, 2] and the so-called reducible multivariate diffusions discussed in Aït-Sahalia [3], the density expansions proposed in Aït-Sahalia [1–3] begin with a so-called *Lamperti* transform, which transforms the marginal distribution to locally normal. Whenever applied, let $\gamma(\cdot;\theta)$ denote such a transform, and let $Z(t) = \gamma(X(t);\theta)$ denote the process after the transform. Thus, taking one-dimensional cases as an example, the expansion for the transition density of X can be constructed from

$$p_X^{(J)}(\Delta, x|x_0;\theta) := \sigma(x;\theta)^{-1} p_Z^{(J)}\big(\Delta, \gamma(x;\theta)|\gamma(x_0;\theta);\theta\big). \tag{4.1}$$

As momentarily demonstrated in the numerical results, a combination of the *Lamperti* transform and our expansion leads to faster convergence, compared with the direct expansion. A heuristic reason for this phenomenon is as follows. As seen from Section 3, our expansion is carried out around a normal distribution. After a *Lamperti* transform, the diffusion behaves locally as a Brownian motion, which facilitates the convergence. Therefore, the *Lamperti* transform may accelerate the convergence of expansion, and thus it is recommended to apply it whenever it exists and is explicit.

For multivariate cases, we employ a popular double mean-reverting Ornstein–Uhlenbeck model (see, e.g., Aït-Sahalia [3]) labeled as DMROU, whose transition density is bivariate correlated normal:

MODEL 3. The DMROU (double mean-reverting Ornstein–Uhlenbeck) model,

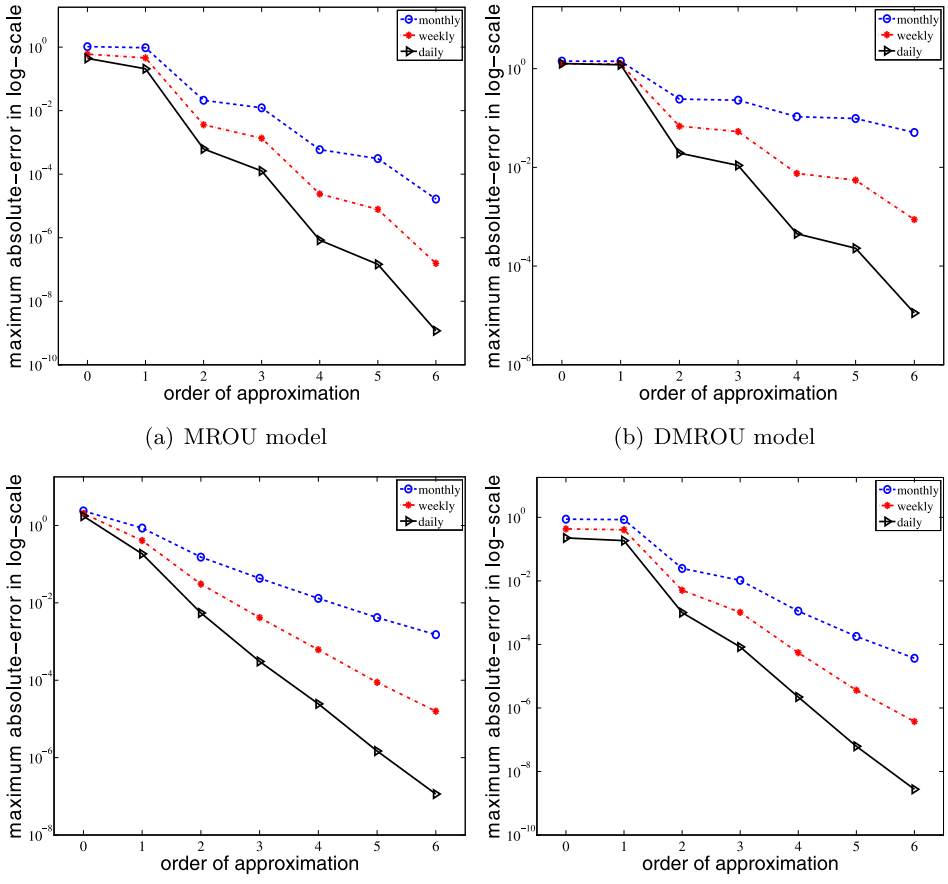
$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \left(\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} - \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \right) dt + d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix},$$

where $\{(W_1(t), W_2(t))\}$ is a standard two-dimensional Brownian motion.

According to the classification in Dai and Singleton [24], the DMROU model is a multivariate affine diffusion process of the $A_0(2)$ type. In the numerical experiments, we choose the parameters as $\kappa_{11} = 5$, $\kappa_{21} = 1$, $\kappa_{22} = 10$ and $\alpha_1 = \alpha_2 = 0$ similar to those employed in Aït-Sahalia [3].

Based on the explicit expressions of the true transition densities of the three models (see, e.g., Aït-Sahalia [1–3]), we exhibit the error of Jth order approximation $e_X^{(J)}(\Delta, x|x_0; \theta) = p_X^{(J)}(\Delta, x|x_0; \theta) - p_X(\Delta, x|x_0; \theta)$ for the time increment Δ. The numerical investigation is performed at a region $\mathcal{D}$, which is several standard deviations around the mean of the forward position (i.e., $\mathbb{E}(X(\Delta)|X(0) = x_0)$), and as an indicator of the overall performance, the uniform error $\max_{x\in\mathcal{D}} |e_X^{(J)}(\Delta, x|x_0; \theta)|$ is considered. In Figure 1(a), (b), (c) and (d), the uniform errors for the above three benchmark models (MROU, SQR and DMROU) are plotted for monthly, weekly and daily monitoring frequencies ($\Delta = 1/12, 1/52, 1/252$) and different orders of approximation ($J = 1, 2, 3, \ldots, 6$). Especially for the SQR model, the plots are provided for both a direct expansion in Figure 1(c) and an expansion with *Lamperti* transform acceleration [see (4.1)] in Figure 1(d). Such numerical evidence demonstrates that the approximation error tends to decrease as the monitoring increment shrinks (Δ decreases) or more correction terms are included (J increases), and that the combination with *Lamperti* transform may accelerate the convergence. As seen from the dynamics of the SQR model, the volatility function $\sigma(x) = \sigma\sqrt{x}$ violates Assumption 2 at the point $x = 0$. However, the numerical performance exhibited in Figure 1(c) and (d) suggests that the technical assumptions given in Section 2 are sufficient but not necessary in order to guarantee numerical convergence of the density expansion and the resulting properties of the approximate MLE. From theoretical perspectives, the singularity at $x = 0$ may lead to a significant challenge in mathematically verifying the convergence of transition density expansion, which can be regarded as a future research topic.

In the supplementary material [44], we document detailed performance of the density approximation for the MROU, SQR (for both the direct expansion and the accelerated approach via *Lamperti* transform) and DMROU models, respectively. For the former two one-dimensional cases, that is, the MROU and SQR models, we plot the errors of approximation corresponding to weekly monitoring frequency and orders ranging from $J = 1, 2, \ldots, 6$. For the latter set of graphs, we plot the contours of the approximation errors for the DMROU model corresponding to weekly monitoring frequency and orders ranging from $J = 1, 2, \ldots, 6$.



(a) MROU model

(b) DMROU model

(c) SQR model (Direct Expansion)

(d) SQR model (with acceleration by *Lamperti* transform)

FIG. 1. *Maximum absolute errors of density approximation for Models* 1, 2 *and* 3.

The asymptotic expansion proposed in this paper is essentially different from that in Aït-Sahalia [1–3] and other existing large-deviations-based results. First, the expansion proposed here includes correction terms corresponding to any order of $\epsilon = \sqrt{\Delta}$; however, in Aït-Sahalia [1–3] and other methods, expansions include only integer orders of Δ (even orders of $\epsilon = \sqrt{\Delta}$). Second, the expansion terms in Aït-Sahalia [1–3] appear to be longer than the corresponding orders in the expansion proposed in this paper. Taking the MROU and the SQR model, for example, the mean-reverting correction starts from the leading order in Aït-Sahalia's expansion; however, in our expansion, the leading order term is the density of a normal distribution, and the first appearance of mean-reverting drift parameters is deferred to the correction term corresponding to first order of ϵ.

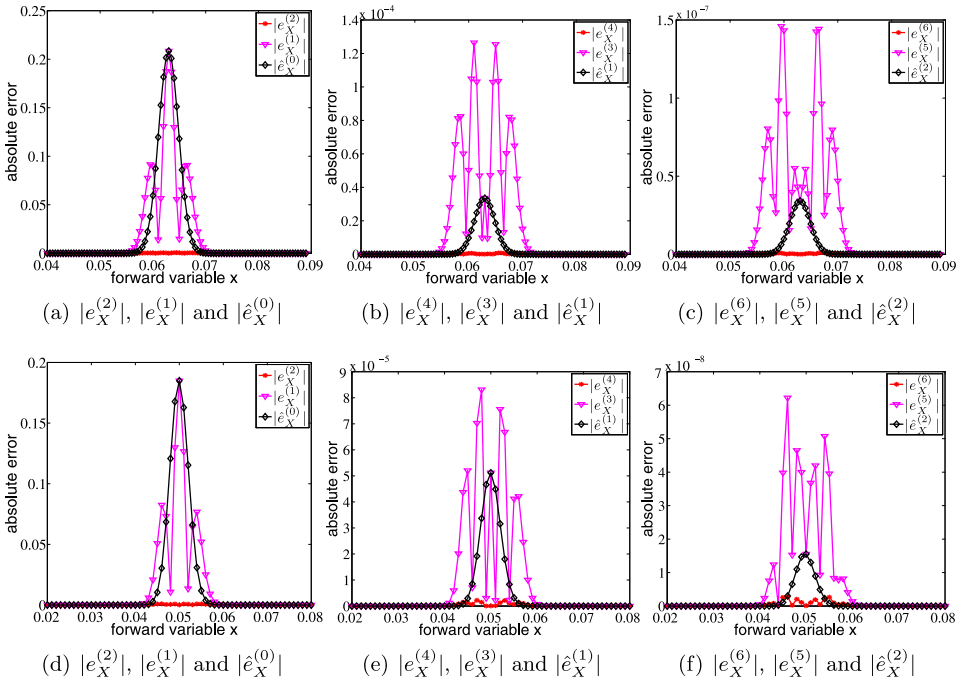


FIG. 2. *Cross comparisons of absolute approximation errors* (*corresponding to different orders of expansion*) *of the proposed method* ($|e_X^{(J)}|$) *and that of Aït-Sahalia* [2] ($|\widehat{e}_X^{(J)}|$).

Let $\widehat{e}_X^{(J)}(\Delta, x|x_0;\theta) = \widehat{p}_X^{(J)}(\Delta, x|x_0;\theta) - p_X(\Delta, x|x_0;\theta)$ denote the approximation error of Aït-Sahalia's Jth order expansion, where $\widehat{p}_X^{(J)}$ is defined in equation (2.14) as in Aït-Sahalia [2]. For the method proposed in this paper, approximation errors are denoted by $e_X^{(J)}(\Delta, x|x_0;\theta) = p_X^{(J)}(\Delta, x|x_0;\theta) - p_X(\Delta, x|x_0;\theta)$. Without loss of generality, I employ the MROU and the SQR models to numerically illustrate the comparison of errors resulting from the method of Aït-Sahalia [1–3] and those from this paper. Considering different expressions and arrangements of correction terms, I make a cross comparison of absolute errors for different orders from the two methods as exhibited in Figure 2(a), (b) and (c) for the MROU model as well as Figure 2(d), (e) and (f) for the SQR model. In particular, we consider the *Lamperti* transform acceleration for the SQR model in order to parallel the method in Aït-Sahalia [2]. In the comparison, the orders range from $J = 1, 2, 3, \ldots, 6$ for our method, while $J = 0, 1, 2$ for that of Aït-Sahalia [2]. Without loss of generality, the monitoring frequency is chosen as $\Delta = 1/52$. As we will see, the absolute errors resulting from each two consecutive orders $J = 2K-1$ and $J = 2K$ of the expansion proposed in this paper sandwich that resulting from the order $K-1$ of the expansion proposed in Aït-Sahalia [2], for $K = 0, 1, 2$. The two methods both admit small magnitude of errors resulting from low order

approximations and are comparable to each other as more correction terms are included.

5. Approximate maximum-likelihood estimation. This section is devoted to a method of approximate MLE based on the asymptotic expansion for transition density proposed in Section 3. Similar to Aït-Sahalia [1–3], the Jth order expansion of the log-density can be given by

$$l_X^{(J)}(\Delta, x|x_0;\theta) := -\frac{m}{2}\log\Delta + \log[\det D(x_0)] + \sum_{k=0}^{J}\Lambda_k\left(D(x_0)\frac{x-x_0}{\sqrt{\Delta}}\right)\epsilon^k$$

for any $J = 0, 1, 2, \ldots$, where the correction terms Λ_k can be explicitly calculated from straightforward differentiation of the density expansion (3.21).

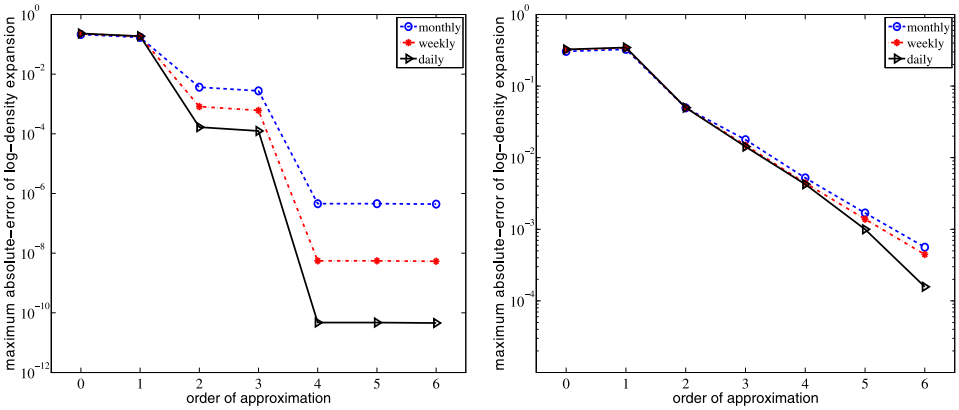
Without loss of generality, we employ Model 1 (MROU) and Model 2 (SQR) to illustrate the convergence of uniform errors of the log-density expansions $(\max_{x\in\mathcal{D}}|l_X^{(J)}(\Delta, x|x_0;\theta) - \log p_X(\Delta, x|x_0;\theta)|)$ in Figure 3(a) and (b) in a similar way as Figure 1(a)–(d) do for the uniform errors of density expansions. For the MROU model, Figure 3(a) shows the uniform errors of its log-density expansions. For the SQR model, Figure 3(b) plots the uniform errors of its *Lamperti*-transformed log-density expansions which are naturally calculated from

$$l_X^{(J)}(\Delta, x|x_0;\theta) := -\log\sigma(x;\theta) + l_Z^{(J)}\big(\Delta, \gamma(x;\theta)|\gamma(x_0;\theta);\theta\big), \tag{5.1}$$

where γ is the *Lamperti* transform and $Z(t) = \gamma(X(t);\theta)$.

By analogy to the log-likelihood function $\ell_n(\theta)$ in (2.3), we introduce the Jth order approximate log-likelihood function

$$\ell_n^{(J)}(\theta) = \sum_{i=1}^{n} L_i^{(J)}(\theta), \tag{5.2}$$



(a) MROU model (b) SQR model (with *Lamperti* transform)

FIG. 3. *Maximum absolute errors of log-density approximation for Models 1 and 2.*

where $L_i^{(J)}(\theta) = l_X^{(J)}(\Delta, X(i\Delta)|X((i-1)\Delta); \theta)$. According to Assumption 3, we assume, for simplicity, that the true log-likelihood function $\ell_n(\theta)$ admits a unique maximizer $\widehat{\theta}_n$, which serves as the true maximum-likelihood estimator. Similarly, let $\widehat{\theta}_n^{(J)}$ be the approximate maximum-likelihood estimator of order J obtained from maximizing $\ell_n^{(J)}(\theta)$. Setting up or refining technical conditions for ensuring the identification and the usual asymptotic properties of the true but generally incomputable MLE is beyond the scope of this paper, and can be investigated as a future research topic; see, for example, Aït-Sahalia [2] for the discussion of one-dimensional cases. As a consequence of Theorem 2, we set up the convergence of the approximate MLE $\widehat{\theta}_n^{(J)}$ to the true MLE $\widehat{\theta}_n$ in what follows.

PROPOSITION 1. *Under Assumptions 1 and 2, the approximate maximum-likelihood estimator obtained from optimizing* (5.2) *satisfies that, for the fixed sample size* n,

$$\widehat{\theta}_n^{(J)} - \widehat{\theta}_n \xrightarrow{P} 0 \tag{5.3}$$

as $\Delta \to 0$ *for* $J \geq m$.

PROOF. See Appendix B. □

Though the convergence in (5.3) is theoretically justified as the monitoring increment Δ shrinks to 0 for any fixed order J, the convergence of (5.3) may also hold as $J \to \infty$ for a range of fixed values of Δ. This is analogous to the Taylor expansion in classical calculus. The respective effects on the discrepancy between the approximate MLE and the true MLE resulting from shrinking Δ and increasing expansion orders J are illustrated via Monte Carlo evidence in Section 6. In particular, we will demonstrate numerically that for an arbitrary Δ, a larger order J results in a better approximation of the MLE.

6. Monte Carlo evidence of approximate maximum-likelihood estimation. To further demonstrate the convergence issues discussed in the previous sections, we provide Monte Carlo evidence of approximate maximum-likelihood estimation for the three models discussed in Section 4. Let N denote the number of sample paths generated from the transition distributions; let n denote the number of observations on each path. For finite-sample results, we report the mean and standard deviation of the discrepancy between the MLE and the true parameter value (i.e., $\widehat{\theta}_n - \theta^{\text{True}}$), and the discrepancy between the approximate MLE and the MLE (i.e., $\widehat{\theta}_n^{(J)} - \widehat{\theta}_n$).

For the two Gaussian models, that is, the MROU model and the DMROU model, the situation considered here is restricted to the stationary case. Therefore, the asymptotic variance of the maximum-likelihood estimator is given by the inverse

of Fisher's information matrix, which is the lowest possible variance of all estimators. So, as in Aït-Sahalia [1–3], we assume that $\kappa > 0$ for the MROU model, and $\kappa_{11} > 0$ and $\kappa_{22} > 0$ for the DMROU model. By the nature of stationarity, one has the local asymptotic normal structure for the maximum-likelihood estimator $\widehat{\theta}_n$, that is,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, i(\theta)^{-1}\big) \tag{6.1}$$

as $n \to \infty$ with Δ fixed. Here, the Fisher information matrix is calculated as

$$i(\theta) = -\mathbb{E}\bigg(\frac{\partial^2 L_1(\theta)}{\partial\theta\,\partial\theta^T}\bigg), \tag{6.2}$$

where ^{T} denotes matrix transposition.

Without loss of generality, we analyze the results of the MROU model in what follows. As seen from Table 1, the asymptotic distribution of $\widehat{\theta}_n - \theta^{\text{True}}$ is calculated from (6.1) and (6.2). The small discrepancy between the finite-sample and asymptotic standard deviations of $\widehat{\theta}_n - \theta^{\text{True}}$ indicates that the choice of sample size $n = 1000$ is approaching an optimality. When the monitoring frequency Δ shrinks, or when the order of approximation J increases, the approximate MLEs obtained from maximizing the approximate log-likelihood function (5.2) get closer to the exact (but usually incomputable) MLEs, and thus get closer to the true parameter, if the sample size n is large enough. This can be seen by a comparison of some outputs with relatively larger bias and standard deviations resulting from relatively lower order expansions or larger monitoring increments with those improved outputs resulting from relatively higher order expansions and smaller monitoring increments. This phenomenon reconciles our discussions in Section 5.

TABLE 1
Monte Carlo evidence for the MROU model

Parameters $\boldsymbol{\theta}^{\text{True}}$	**Asymptotic** $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}^{\text{True}}$		**Finite sample** $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}^{\text{True}}$		**Finite sample** $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}^{(3)} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$		**Finite sample** $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}^{(6)} - \widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$	
	Mean	**Stddev**	**Mean**	**Stddev**	**Mean**	**Stddev**	**Mean**	**Stddev**
$\Delta = 1/52$								
$\kappa = 0.5$	0	0.229136	0.245175	0.329396	0.013477	0.014645	0.000002	0.000102
$\alpha = 0.06$	0	0.013682	0.000329	0.015202	0.000002	0.000318	0.000000	0.000003
$\sigma = 0.03$	0	0.000674	0.000021	0.000675	0.000003	0.000015	−0.000000	0.000000
$\Delta = 1/12$								
$\kappa = 0.5$	0	0.111867	0.054162	0.124773	0.028923	0.014382	−0.000003	0.000297
$\alpha = 0.06$	0	0.006573	0.000097	0.006440	0.000002	0.000174	0.000000	0.000014
$\sigma = 0.03$	0	0.000685	0.000022	0.000687	0.000025	0.000022	0.000000	0.000001

Notes. The number of simulation trials is $N = 5000$ and the number of observations on each path is $n = 1000$.

While holding the length of sampling interval Δ fixed, the approximation error $\widehat{\theta}_n^{(J)} - \widehat{\theta}_n$ decreases and is dominated by the intrinsic sampling error $\widehat{\theta}_n - \theta^{\mathrm{True}}$ as J increases. Therefore, according to Aït-Sahalia [1–3], a small-order approximation (e.g., the $\widehat{\theta}_n^{(6)}$ for the MROU model) is adequate enough for replacing the true MLE $\widehat{\theta}_n$ for the purpose of estimating unknown parameter θ. According to Aït-Sahalia [1–3], once the approximation error resulting from replacing the true density p_X by its approximation, say $p_X^{(J)}$, is dominated by the sampling error [usually estimated from asymptotic variance computed from (6.1)] due to the true maximum-likelihood estimation, such $p_X^{(J)}$ is appropriate in practice. Such a proper replacement has an effect that is statistically indiscernible from the sampling variation of the true yet incomputable MLE $\widehat{\theta}_n$ around θ. As a result of the fast development of modern computation and optimization technology, calculation of high-order likelihood approximations will become increasingly feasible; thus errors between approximate MLE and MLE can be improved to become arbitrarily small, at least in principle.

Owing to the limited space in this paper and the similarity in the pattern of results to those of the MROU model, we collect the simulation results for the DMROU and the SQR models in the supplementary material [44]. In particular, for the SQR model, the simulation results will demonstrate that a combination of the *Lamperti* transform and our expansion may enhance the efficiency of the estimation. Moreover, in the supplementary material [44], we will investigate two more sophisticated data-generating processes (arising from financial modeling) with rich drift and diffusion specifications, in which the Lamperti transform either requires computationally demanding implicit integration and inversion or does not exist due to a multivariate irreducible specification; see Aït-Sahalia [3].

7. Concluding remarks. This paper contributes a method for approximate maximum-likelihood estimation (MLE) of multivariate diffusion processes from discretely sampled data, based on a closed-form asymptotic expansion for transition density, for which any arbitrary order of corrections can be systematically obtained through a generally implementable algorithm. Numerical examples and Monte Carlo evidence for illustrating the performance of density asymptotic expansion and the resulting approximate MLE are provided in order to demonstrate the wide applicability of the method. Based on some sufficient (but not necessary) technical conditions, the convergence and asymptotic properties are theoretically justified. Owing to the limited space of this paper which focuses on introducing a method of estimation, investigations on more asymptotic properties related to the approximate MLE can be regarded as a future research topic, for example, a tighter upper bound for the discrepancy (5.3) based on the error estimate of the transition density expansion (3.27), as well as the consistency and asymptotic distribution of the approximate MLE under various sampling schemes in terms of monitoring frequency and observational horizon; see, for example, Yoshida [69], Kessler [39],

Bibby and Sørensen [19] and Genon-Catalot and Jacod [32]. In this regard, we note Chang and Chen [22] for analyzing the asymptotic properties of the approximate MLE proposed in Aït-Sahalia [2] of one-dimensional diffusion processes. One may also apply the idea for explicitly approximating transition density in various other aspects of statistical inference, for which explicit asymptotic expansions of certain quantities are helpful.

APPENDIX A: EXPLICIT CALCULATION OF CONDITIONAL EXPECTATION (3.24)

In this section, we expatiate on a general algorithm for explicitly calculating the conditional expectation (3.24) of multiplication of iterated Stratonovich integrals as a multivariate polynomial in $z = (z_1, z_2, \ldots, z_m) \in \mathbb{R}^m$. In addition to theoretical interests, iterated stochastic integral plays important roles in many applications arising from stochastic modeling, for example, the analysis of convergence rate of various methods for approximating solutions to stochastic differential equations; see Kloeden and Platen [41]. Special cases for conditional expectations of iterated Itô stochastic integrals (without integral with respect to the time variable) can be found in, for example, Nualart, Üstünel and Zakai [54], Yoshida [67, 68] and Kunitomo and Takahashi [42].

To present our algorithm, similar to the definition of iterated Stratonovich integral, we define

$$I_{\mathbf{i}}[f](t) := \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} f(t_n)\, dW_{i_n}(t_n) \cdots dW_{i_2}(t_2)\, dW_{i_1}(t_1) \tag{A.1}$$

as an iterated Itô integral for an arbitrary index $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \{0, 1, 2, \ldots, m\}^n$ with a right-continuous integrand f. To lighten the notation, the integral $I_{\mathbf{i}}[1](t)$ is abbreviated to $I_{\mathbf{i}}(t)$.

Before discussing details in the following subsections, we briefly outline a general algorithm, which can be implemented using any symbolic packages, for example, Mathematica. Throughout our discussion, the iterated (Stratonovich or Itô) stochastic integrals may involve integrations with respect to not only Brownian motions but also time variables.

ALGORITHM.

- Convert each iterated Stratonovich integral in (3.24) to a linear combination of iterated Itô integrals;
- Convert each multiplication of iterated Itô integrals resulting from the previous step to a linear combination of iterated Itô integrals;
- Compute the conditional expectation of iterated Itô integral via an explicit construction of Brownian bridge.

A.1. Conversion from iterated Stratonovich integrals to Itô integrals. Denote by $l(\mathbf{i}) := l((i_1, \ldots, i_n)) = n$ the length of the index $\mathbf{i}$. Denote by $-\mathbf{i}$ an index

obtained from deleting the first element of $\mathbf{i}$. In particular, if $l(\mathbf{i}) = 0$, we define $J_{\mathbf{i}}[f](t) = f(t)$ by slightly extending the definition (3.5). According to page 172 of Kloeden and Platen [41], we have the following conversion algorithm: for the case of $l(\mathbf{i}) = 0$ or 1, we have $J_{\mathbf{i}}(t) = I_{\mathbf{i}}(t)$; for the case of $l(\mathbf{i}) \geq 2$, we have

$$J_{\mathbf{i}}(t) = I_{(i_1)}\big[J_{-\mathbf{i}}(\cdot)\big](t) + 1_{\{i_1=i_2\neq 0\}} I_{(0)}\big[\tfrac{1}{2} J_{-(-\mathbf{i})}(\cdot)\big](t). \tag{A.2}$$

For example, if $l(\mathbf{i}) = 2$, one has

$$J_{\mathbf{i}}(t) = I_{\mathbf{i}}(t) + \tfrac{1}{2} 1_{\{i_1=i_2\neq 0\}} I_{(0)}(t).$$

Thus, with the conversion algorithm (A.2), we convert each iterated Stratonovich integral in (3.24) to a linear combination of iterated Itô integrals. Thus, the product $\prod_{\omega=1}^{l} J_{\mathbf{i}_\omega}(1)$ can be expanded as a linear combination of multiplication of Itô integrals.

A.2. Conversion from multiplication of Itô integrals to a linear combination. We provide a simple recursion algorithm for converting a multiplication of iterated Itô integrals to a linear combination. According to Lemma 2 in Tocino [62], a product of two Itô integrals as defined in (A.1) satisfies that

$$\begin{aligned} I_{\boldsymbol{\alpha}}(t) I_{\boldsymbol{\beta}}(t) &= \int_0^t I_{\boldsymbol{\alpha}}(s) I_{-\boldsymbol{\beta}}(s)\, dW_{\beta_1}(s) + \int_0^t I_{-\boldsymbol{\alpha}}(s) I_{\boldsymbol{\beta}}(s)\, dW_{\alpha_1}(s) \\ &\quad + \int_0^t I_{-\boldsymbol{\alpha}}(s) I_{-\boldsymbol{\beta}}(s) 1_{\{\alpha_1=\beta_1\neq 0\}}\, ds \end{aligned} \tag{A.3}$$

for any arbitrary indices $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)$. Iterative applications of this relation render a linear combination form of $I_{\boldsymbol{\alpha}}(t) I_{\boldsymbol{\beta}}(t)$. Inductive applications of such an algorithm convert a product of any number of iterated Itô integrals to a linear combination. Therefore, our immediate task is reduced to the calculation of conditional expectations of iterated Itô integrals.

A.3. Conditional expectation of iterated Itô integral. We focus on the explicit calculation of conditional expectations of the following type:

$$\begin{aligned} &\mathbb{E}\big(I_{\mathbf{i}}(1) | W(1) = z\big) \\ &\quad = \mathbb{E}\bigg(\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} dW_{i_n}(t_n) \cdots dW_{i_2}(t_2)\, dW_{i_1}(t_1) | W(1) = z\bigg). \end{aligned} \tag{A.4}$$

By an explicit construction of Brownian bridge (see page 358 in Karatzas and Shreve [38]), we obtain the following distributional identity, for any $k = 1, 2, \ldots, m$:

$$\big(W_k(t) | W(1) = z\big) \overset{\mathcal{D}}{=} \big(W_k(t) | W_k(1) = z_k\big) \overset{\mathcal{D}}{=} BB_k^z(t) := \mathcal{B}_k(t) - t\mathcal{B}_k(1) + tz_k,$$

where $\mathcal{B}_k$'s are independent Brownian motions and $BB_k^z(t) := \mathcal{B}_k(t) - t\mathcal{B}_k(1) + tz_k$ is distributed as a Brownian bridge starting from 0 and ending at z_k at time 1.

For ease of exposition, we also introduce $\mathcal{B}_0(t) \equiv 0$ and $z_0 = 1$. Therefore, the condition $W(1) = z$ in (A.4) can be eliminated since

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} dW_{i_n}(t_n) \cdots dW_{i_2}(t_2)\, dW_{i_1}(t_1) | W(1) = z\Big) \\
&\quad = \mathbb{E}\Big(\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} d\big(\mathcal{B}_{i_n}(t_n) - t_n \mathcal{B}_{i_n}(1) + t_n z_{i_n}\big) \cdots \\
&\qquad\qquad d\big(\mathcal{B}_{i_2}(t_2) - t_2 \mathcal{B}_{i_2}(1) + t_2 z_{i_2}\big) \\
&\qquad\qquad d\big(\mathcal{B}_{i_1}(t_1) - t_1 \mathcal{B}_{i_1}(1) + t_1 z_{i_1}\big)\Big).
\end{aligned} \tag{A.5}$$

An early attempt using the idea of Brownian bridge to deal with conditional expectation (A.4) can be found in Uemura [64], which investigated the calculation of heat kernel expansion in the diagonal case. It is worth mentioning that, instead of giving a method for explicitly calculating (A.4), Uemura [64] employed discretization of stochastic integrals to show that (A.4) has the structure of a multivariate polynomial in z with unknown coefficients. Therefore, the validity of the above derivation can be seen from the definition of stochastic integral as a limit of discretized summation. In particular, the random variables $\mathcal{B}_{i_1}(1), \mathcal{B}_{i_2}(1), \ldots, \mathcal{B}_{i_n}(1)$ are not involved in the integral in (A.5). The integrals with respect to $d\mathcal{B}_{i_k}(t_k)$ are in the sense of usual stochastic integrals; the integrals with respect to dt_k are in the sense of Lebesgue integrals.

By expanding the right-hand side of (A.5) and collecting terms according to monomials of z_i's, we express (A.4) as a multivariate polynomial in z:

$$\mathbb{E}\big(I_{\mathbf{i}}(1) | W(1) = z\big) = \sum_{k=0}^{n} \sum_{\{l_1, l_2, \ldots, l_k\} \subset \{1,2,\ldots,n\}} c(l_1, l_2, \ldots, l_k) z_{i_{l_1}} z_{i_{l_2}} \cdots z_{i_{l_k}},$$

where the coefficients are determined by

$$\begin{aligned}
&c(l_1, l_2, \ldots, l_k) \\
&\quad := \mathbb{E} \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} d\big(\mathcal{B}_{i_n}(t_n) - t_n \mathcal{B}_{i_n}(1)\big) \cdots \\
&\qquad\qquad d\big(\mathcal{B}_{i_{l_k+1}}(t_{l_k+1}) - t_{l_k+1} \mathcal{B}_{i_{l_k+1}}(1)\big) \\
&\qquad\qquad dt_{l_k}\, d\big(\mathcal{B}_{i_{l_k-1}}(t_{l_k-1}) - t_{l_k-1} \mathcal{B}_{i_{l_k-1}}(1)\big) \cdots \\
&\qquad\qquad d\big(\mathcal{B}_{i_{l_2+1}}(t_{l_2+1}) - t_{l_2+1} \mathcal{B}_{i_{l_2+1}}(1)\big) \\
&\qquad\qquad dt_{l_2}\, d\big(\mathcal{B}_{i_{l_2-1}}(t_{l_2-1}) - t_{l_2-1} \mathcal{B}_{i_{l_2-1}}(1)\big) \cdots \\
&\qquad\qquad d\big(\mathcal{B}_{i_{l_1+1}}(t_{l_1+1}) - t_{l_1+1} \mathcal{B}_{i_{l_1+1}}(1)\big) \\
&\qquad\qquad dt_{l_1}\, d\big(\mathcal{B}_{i_{l_1-1}}(t_{l_1-1}) - t_{l_1-1} \mathcal{B}_{i_{l_1-1}}(1)\big) \cdots \\
&\qquad\qquad d\big(\mathcal{B}_{i_1}(t_1) - t_1 \mathcal{B}_{i_1}(1)\big).
\end{aligned} \tag{A.6}$$

Algebraic calculation from expanding the terms like $d(\mathcal{B}_{i_n}(t_n) - t_n\mathcal{B}_{i_n}(1))$ simplifies (A.6) as a linear combination of expectations of the following form: $\mathbb{E}(\mathcal{B}_{m_1}(1)\mathcal{B}_{m_2}(1)\cdots\mathcal{B}_{m_r}(1)I_{\mathbf{j}}(1))$ where $I_{\mathbf{j}}(1)$ is an iterated Itô integral.

By viewing $\mathcal{B}_{m_i}(1)$ as $\int_0^1 d\mathcal{B}_{m_i}(t_1)$, we have

$$\mathbb{E}\big(\mathcal{B}_{m_1}(1)\mathcal{B}_{m_2}(1)\cdots\mathcal{B}_{m_r}(1)I_{\mathbf{j}}(1)\big) = \mathbb{E}\left(\prod_{i=1}^{r}\int_0^1 d\mathcal{B}_{m_i}(t_1)I_{\mathbf{j}}(1)\right). \tag{A.7}$$

To calculate this expectation, we use the algorithm proposed in Section A.2 to convert $\prod_{i=1}^{r}\int_0^1 d\mathcal{B}_{m_i}(t_1)I_{\mathbf{j}}(1)$ to a linear combination of iterated Itô integrals. Finally, we need to calculate expectation of iterated Itô integrals without conditioning. For any arbitrary index $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \{0, 1, 2, \ldots, m\}^n$, we have

$$\begin{aligned}\mathbb{E}I_{\mathbf{i}}(1) &= \mathbb{E}\left(\int_0^1\int_0^{t_1}\cdots\int_0^{t_{n-1}} dW_{i_n}(t_n)\cdots dW_{i_2}(t_2)\,dW_{i_1}(t_1)\right)\\ &= \int_0^1\int_0^{t_1}\cdots\int_0^{t_{n-1}} dt_n\cdots dt_2\,dt_1 \equiv \frac{1}{n!},\end{aligned}$$

if $\mathbf{i} = (i_1, i_2, \ldots, i_n) = (0, 0, \ldots, 0)$ and $\mathbb{E}I_{\mathbf{i}}(1) = 0$, otherwise (by the martingale property of stochastic integrals).

APPENDIX B: PROOFS

B.1. Proof of Theorem 1. Using the chain rule and the Taylor theorem, the kth ($k \geq 1$) order correction term for $\delta(Y^\epsilon - y)$ admits the following form:

$$\Phi_k(y) = \sum_{(l,\mathbf{r}(l),\mathbf{j}(l))\in S_k}\frac{1}{l!}\partial^{\mathbf{r}}\delta\big(B(1)-y\big)Y_{j_1,r_1}Y_{j_2,r_2}\cdots Y_{j_l,r_l}, \tag{B.1}$$

where $\partial^{\mathbf{r}}$ denotes $\frac{\partial}{\partial x_{r_1}}\frac{\partial}{\partial x_{r_2}}\cdots\frac{\partial}{\partial x_{r_l}}$ for simplicity. Thus, taking expectation of (B.1) and applying (3.18), we obtain that

$$\begin{aligned}\Omega_k(y) &= \mathbb{E}\Phi_k(y)\\ &= \sum_{(l,\mathbf{r}(l),\mathbf{j}(l))\in S_k}\frac{1}{l!}D_{r_1r_1}(x_0)D_{r_2r_2}(x_0)\cdots D_{r_lr_l}(x_0)\\ &\qquad\times\mathbb{E}\big(\partial^{\mathbf{r}}\delta\big(B(1)-y\big)F_{j_1+1,r_1}F_{j_2+1,r_2}\cdots F_{j_l+1,r_l}\big).\end{aligned}$$

Employing the integration-by-parts property of the Dirac delta function (see, e.g., Section 2.6 in Kanwal [37]), the conditional expectation can be computed as

$$\begin{aligned}&\mathbb{E}\big[\partial^{\mathbf{r}}\delta\big(B(1)-y\big)F_{j_1+1,r_1}F_{j_2+1,r_2}\cdots F_{j_l+1,r_l}\big]\\ &\quad=\int_{b\in\mathbb{R}^d}\mathbb{E}\big[\partial^{\mathbf{r}}\delta\big(B(1)-y\big)F_{j_1+1,r_1}F_{j_2+1,r_2}\cdots F_{j_l+1,r_l}|B(1)=b\big]\\ &\qquad\times\phi_{\Sigma(x_0)}(b)\,db\end{aligned}$$

$$= (-1)^l \partial^{\mathbf{r}}\big[\mathbb{E}\big[F_{j_1+1,r_1}F_{j_2+1,r_2}\cdots F_{j_l+1,r_l}|W(1)=\sigma(x_0)^{-1}D(x_0)^{-1}y\big] \times \phi_{\Sigma(x_0)}(y)\big],$$

where $\phi_{\Sigma(x_0)}(y)$ is given in (3.22). By plugging in (3.10), we have that

$$\mathbb{E}\big[F_{j_1+1,r_1}F_{j_2+1,r_2}\cdots F_{j_l+1,r_l}|W(1)=z\big] = \sum_{\{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)|\|\mathbf{i}_\omega\|=j_\omega+1,\omega=1,2,\ldots,l\}} \prod_{\omega=1}^{l} C_{\mathbf{i}_\omega,r_\omega}(x_0)P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(z),$$

where $P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(z)$ is defined in (3.24). Formula (3.25) follows from the fact that

$$\frac{\partial}{\partial y_i}\big(P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(\sigma(x_0)^{-1}D(x_0)^{-1}y)\phi_{\Sigma(x_0)}(y)\big) = \bigg(\frac{\partial}{\partial y_i}P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(\sigma(x_0)^{-1}D(x_0)^{-1}y) - P_{(\mathbf{i}_1,\mathbf{i}_2,\ldots,\mathbf{i}_l)}(\sigma(x_0)^{-1}D(x_0)^{-1}y)\big(\Sigma(x_0)^{-1}y\big)_i\bigg)\phi_{\Sigma(x_0)}(y)$$

as well as the definition of the differential operators in (3.14).

REMARK 1. The above conditioning argument can be justified, when $\partial^{\mathbf{r}}\delta(B(1)-y)$ is regarded as a generalized Wiener functional (random variable) and the expectation is interpreted in the corresponding generalized sense as in Watanabe [65].

B.2. Proof of Theorem 2. Now, based on Assumption 2, we introduce the following uniform upper bounds. For $k \geq 1$, let μ_k and σ_k be the uniform upper bounds of the kth order derivative of μ and σ, respectively, that is,

$$\bigg|\frac{\partial^{(k)}\mu(x;\theta)}{\partial x^k}\bigg| \leq \mu_k \quad \text{and} \quad \bigg|\frac{\partial^{(k)}\sigma(x;\theta)}{\partial x^k}\bigg| \leq \sigma_k \tag{B.2}$$

for $(x,\theta) \in \mathbb{R}^m \times \Theta$. Also, let μ_0 and σ_0 denote the uniform upper bounds of $|\mu(x_0;\theta)|$ and $|\sigma(x_0;\theta)|$ on $(x_0,\theta) \in K \times \Theta$, respectively, that is,

$$\big|\mu(x_0;\theta)\big| \leq \mu_0 \quad \text{and} \quad \big|\sigma(x_0;\theta)\big| \leq \sigma_0 \tag{B.3}$$

for $(x_0,\theta) \in K \times \Theta$. In order to establish the uniform convergence in Theorem 2, we introduce the following lemma. When the dependence of parameters is emphasized, we express $X^\epsilon(1)$ as $X^\epsilon(1;\theta,x_0)$ and express the standardized random variable Y^ϵ defined in (3.16) as

$$Y^\epsilon(\theta,x_0) = D(x_0)\big(X^\epsilon(1;\theta,x_0)-x_0\big)/\sqrt{\Delta}.$$

In this Appendix, we employ standard notation of Malliavin calculus (see, e.g., Nualart [53] and Ikeda and Watanabe [35]) and the theory of Watanabe [65] and Yoshida [67, 68]. For the readers' convenience, a brief survey of some relative theory is provided in the supplementary material [44].

LEMMA 2. *Under Assumption 2, the pathwise expansion (3.18) holds in the sense of D^∞ uniformly in $(x_0, \theta) \in K \times \Theta$, that is,*

$$\left\| Y^\epsilon(\theta, x_0) - \sum_{k=0}^{J} \frac{1}{k!} \frac{\partial^{(k)} Y^\epsilon(\theta, x_0)}{\partial \epsilon^k}\bigg|_{\epsilon=0} \epsilon^k \right\|_{D_p^s} = \mathcal{O}\big(\epsilon^{J+1}\big)$$

for any $J \in \mathbb{N}$, $p \geq 1$ and $s \in \mathbb{N}$.

PROOF. See the supplementary material [44]. □

Because of Assumption 1, Theorem 3.4 in Watanabe [65] guarantees the uniform nondegenerate condition, that is,

$$\limsup_{\epsilon \to 0} \mathbb{E}\big[\det\big(\Sigma\big(Y^\epsilon(\theta, x_0)\big)\big)^{-p}\big] < \infty \qquad \text{for any } p \in (0, +\infty).$$

Let $\Lambda = \mathbb{R}^m$ denote a set of indices. For any $y = (y_1, \ldots, y_m) \in \Lambda$, let us consider a generalized function defined as $T_y(z) := \delta(z - y)$, which is a Schwartz distribution, that is, $T_y \in \mathcal{S}'(\mathbb{R}^m)$. Applying Theorem 2.3 in Watanabe [65] and Theorem 2.2 in Yoshida [68], we obtain that $T_y(Y^\epsilon(\theta, x_0))$ admits the following asymptotic expansion: for any arbitrary $J \in \mathbb{N}$,

$$\delta\big(Y^\epsilon(\theta, x_0) - y\big) := \sum_{k=0}^{J} \Phi_{k,(\theta,x_0)}(y)\epsilon^k + \mathcal{O}\big(\epsilon^{J+1}\big) \qquad \text{in } D^{-\infty},$$

uniform in $y \in \Lambda$, $x_0 \in K$ and $\theta \in \Theta$. Here, the correction term $\Phi_{k,(\theta,x_0)}(y)$ is given in (B.1). Therefore, we obtain that

$$\sup_{y \in \Lambda, x_0 \in K, \theta \in \Theta} \left| \mathbb{E}\left(\delta\big(Y^\epsilon(\theta, x_0) - y\big) - \sum_{k=0}^{J} \Phi_{k,(\theta,x_0)}(y)\epsilon^k \right) \right| = \mathcal{O}\big(\epsilon^{J+1}\big).$$

Hence, by taking into account the transform (3.16), we obtain that

$$\sup_{\substack{(x,x_0,\theta) \\ \in E \times K \times \Theta}} \left| \mathbb{E}\delta\big(X^\epsilon(1) - x\big) - \frac{\det D(x_0)}{\sqrt{\Delta^m}} \sum_{k=0}^{J} \Omega_k\left(\frac{D(x_0)(x - x_0)}{\sqrt{\Delta}} \right)\epsilon^k \right| = \mathcal{O}\big(\Delta^{(J+1-m)/2}\big),$$

which yields (3.27).

B.3. Proof of Proposition 1. For $J \geq m$, let

$$R^{(J)}(\Delta, x | x_0; \Theta) := \sup_{\theta \in \Theta} \big| p_X(\Delta, x | x_0; \theta) - p_X^{(J)}(\Delta, x | x_0; \theta) \big|.$$

By Theorem 2, there exists a constant $C > 0$ such that $R^{(J)}(\Delta, x|x_0; \Theta) \leq C\epsilon^{J+1-m}$ for any $x \in E$ and ϵ sufficiently small. Thus, for any positive integer k, it follows that

$$\mathbb{E}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big|^k | X(t) = x_0\big] \leq C^k \epsilon^{k(J+1-m)} \to 0 \qquad \text{as } \epsilon \to 0.$$

By the Chebyshev inequality, $R^{(J)}(\Delta, X(t+\Delta)|X(t); \Theta)$ converges to zero in probability given $X(t) = x_0$, that is, for any $\varepsilon > 0$,

$$\mathbb{P}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big| > \varepsilon | X(t) = x_0\big] \to 0 \qquad \text{as } \epsilon \to 0.$$

By conditioning, it follows that

$$\begin{aligned}&\mathbb{P}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big| > \varepsilon\big] \\ &\quad = \int_R \mathbb{P}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big| > \varepsilon | X(t) = x_0\big]\mathbb{P}\big(X(t) \in dx_0\big).\end{aligned}$$

Because of the fact that

$$0 \leq \mathbb{P}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big| > \varepsilon | X(t) = x_0\big] \leq 1$$

and $\int_R \mathbb{P}(X(t) \in dx_0) = 1$, it follows from the Lebesgue dominated convergence theorem that

$$\mathbb{P}\big[\big|R^{(J)}\big(\Delta, X(t+\Delta)|X(t); \Theta\big)\big| > \varepsilon\big] \to 0 \qquad \text{as } \epsilon \to 0,$$

that is,

$$\mathbb{P}\Big[\sup_{\theta \in \Theta} \big|p_X\big(\Delta, X(t+\Delta)|X(t); \theta\big) - p_X^{(J)}\big(\Delta, X(t+\Delta)|X(t); \theta\big)\big| > \varepsilon\Big] \to 0$$

as $\epsilon \to 0$. Now, we obtain that

$$p_X^{(J)}\big(\Delta, X(t+\Delta)|X(t); \theta\big) - p_X\big(\Delta, X(t+\Delta)|X(t); \theta\big) \xrightarrow{\mathbb{P}} 0 \tag{B.4}$$

as $\epsilon \to 0$ uniformly in $\theta \in \Theta$. Following similar lines of argument as those in the proof of Theorem 2 in Aït-Sahalia [2] and Theorem 3 in Aït-Sahalia [3], we arrive at

$$L_i^{(J)}(\theta) \xrightarrow{\mathbb{P}} L_i(\theta) \qquad \text{as } \epsilon \to 0 \text{ uniformly in } \theta \in \Theta$$

by the convergence in (B.4) and continuity of logarithm. Hence, for any arbitrary $n > 0$, one obtains the convergence of log-likelihood $\ell_n^{(J)}(\theta) \xrightarrow{\mathbb{P}} \ell_n(\theta)$ uniformly in θ. Finally, the convergence of $\widehat{\theta}_n^{(J)} - \widehat{\theta}_n \xrightarrow{\mathbb{P}} 0$ as $\epsilon \to 0$ follows directly from Assumption 3 and the standard method employed in Aït-Sahalia [2, 3].

Acknowledgements. I am very grateful to Professor Peter Bühlmann (Co-Editor), the Associate Editor and three anonymous referees for the constructive suggestions. I also thank Professors Yacine Aït-Sahalia, Mark Broadie, Song Xi Chen, Ioannis Karatzas, Per Mykland, Nakahiro Yoshida and Lan Zhang for helpful comments.

SUPPLEMENTARY MATERIAL

Maximum-likelihood estimation for diffusion processes via closed-form density expansions—Supplementary material (DOI: 10.1214/13-AOS1118SUPP; .pdf). This supplementary material contains (1) closed-form formulas for $\Omega_1(y)$, $\Omega_2(y)$ and $\Omega_3(y)$, (2) closed-form expansion formulas for the examples, (3) detailed plots of errors for the examples, (4) simulation results for the DMROU and SQR models, (5) an alternative exhibition of the simulation results, (6) two more examples for simulation study, (7) a brief survey of the Malliavin Calculus and Watanabe–Yoshida Theory and (8) proof of Lemma 2.

REFERENCES

[1] AÏT-SAHALIA, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *J. Finance* **54** 1361–1395.

[2] AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262. MR1926260

[3] AÏT-SAHALIA, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36** 906–937. MR2396819

[4] AÏT-SAHALIA, Y. (2009). Estimating and testing continuous-time models in finance: The role of transition densities. *Annual Review of Financial Economics* **1** 341–359.

[5] AÏT-SAHALIA, Y. and KIMMEL, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* **83** 413–452.

[6] AÏT-SAHALIA, Y. and KIMMEL, R. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics* **98** 113–144.

[7] AÏT-SAHALIA, Y. and MYKLAND, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71** 483–549. MR1958137

[8] AÏT-SAHALIA, Y. and MYKLAND, P. A. (2004). Estimators of diffusions with randomly spaced discrete observations: A general theory. *Ann. Statist.* **32** 2186–2222. MR2102508

[9] AÏT-SAHALIA, Y. and YU, J. (2006). Saddlepoint approximations for continuous-time Markov processes. *J. Econometrics* **134** 507–551. MR2327401

[10] AZENCOTT, R. (1984). Densité des diffusions en temps petit: Développements asymptotiques. In *Lecture Notes in Math. Seminar on Probability XVIII* **1059** 402–498. Springer, Berlin.

[11] BAKSHI, G. and JU, N. (2005). A refinement to Aït-Sahalia's (2003) "Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach." *J. Bus.* **78** 2037–2052.

[12] BAKSHI, G., JU, N. and OU-YANG, H. (2006). Estimation of continuous-time models with an application to equity volatility dynamics. *Journal of Financial Economics* **82** 227–249.

[13] BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980). *Statistical Inference for Stochastic Processes*: *Probability and Mathematical Statistics*. Academic Press, London. MR0586053

[14] BEN AROUS, G. (1988). Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. *Stochastics* **25** 125–153. MR0999365

[15] BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *Ann. Statist.* **37** 223–245. MR2488350

[16] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 333–382. MR2278331

[17] BESKOS, A. and ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15** 2422–2444. MR2187299

[18] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. MR0471142

[19] BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39. MR1354454

[20] BISMUT, J.-M. (1984). *Large Deviations and the Malliavin Calculus*. *Progress in Mathematics* **45**. Birkhäuser, Boston, MA. MR0755001

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

[22] CHANG, J. and CHEN, S. X. (2011). On the approximate maximum likelihood estimation for diffusion processes. *Ann. Statist.* **39** 2820–2851. MR3012393

[23] DACUNHA-CASTELLE, D. and FLORENS-ZMIROU, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19** 263–284. MR0872464

[24] DAI, Q. and SINGLETON, K. J. (2000). Specification analysis of affine term structure models. *J. Finance* **55** 1943–1978.

[25] DURHAM, G. B. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. With comments and a reply by the authors. MR1939904

[26] EGOROV, A. V., LI, H. and XU, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *J. Econometrics* **114** 107–139. MR1962374

[27] ELERIAN, O. (1998). A note on the existence of a closed form conditional transition density for the Milstein scheme. Working papers, Univ. Oxford Economics.

[28] ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375

[29] FELLER, W. (1951). Two singular diffusion problems. *Ann. of Math.* (2) **54** 173–182. MR0054814

[30] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, *Vol.* 2. Wiley, New York.

[31] FILIPOVIĆ, D., MAYERHOFER, E. and SCHNEIDER, P. (2011). Density approximations for multivariate affine jump-diffusion processes. Preprint.

[32] GENON-CATALOT, V. and JACOD, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **29** 119–151. MR1204521

[33] HALL, P. (1995). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

[34] HURN, A., JEISMAN, J. and LINDSAY, K. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics* **5** 390–455.

[35] IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. *North-Holland Mathematical Library* **24**. North-Holland, Amsterdam. MR1011252

[36] JENSEN, B. and POULSEN, R. (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives* **9** 1–15.

[37] KANWAL, R. P. (2004). *Generalized Functions*: *Theory and Applications*, 3rd ed. Birkhäuser, Boston, MA. MR2075881

[38] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940

[39] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **24** 211–229. MR1455868

[40] KESSLER, M. and SØRENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314. MR1681700

[41] KLOEDEN, P. E. and PLATEN, E. (1999). *The Numerical Solution of Stochastic Differential Equations*. Springer, New York.

[42] KUNITOMO, N. and TAKAHASHI, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims. *Math. Finance* **11** 117–151. MR1807851

[43] LÉANDRE, R. (1988). Applications quantitatives et géométriques du calcul de Malliavin. In *Stochastic Analysis* (*Paris*, 1987) (M. Metivier and S. Watanabe, eds.). *Lecture Notes in Math.* **1322** 109–133. Springer, Berlin. MR0962867

[44] LI, C. (2013). Supplement to "Maximum-likelihood estimation for diffusion processes via closed-form density expansions." DOI:10.1214/13-AOS1118SUPP.

[45] LI, M. (2010). A damped diffusion framework for financial modeling and closed-form maximum likelihood estimation. *J. Econom. Dynam. Control* **34** 132–157. MR2576110

[46] LO, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* **4** 231–247. MR0959611

[47] MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman & Hall, London. MR0907286

[48] MYKLAND, P. A. (1992). Asymptotic expansions and bootstrapping distributions for dependent variables: A martingale approach. *Ann. Statist.* **20** 623–654. MR1165585

[49] MYKLAND, P. A. (1993). Asymptotic expansions for martingales. *Ann. Probab.* **21** 800–818. MR1217566

[50] MYKLAND, P. A. (1994). Bartlett type identities for martingales. *Ann. Statist.* **22** 21–38. MR1272073

[51] MYKLAND, P. A. (1995). Martingale expansions and second order inference. *Ann. Statist.* **23** 707–731. MR1345195

[52] MYKLAND, P. A. and ZHANG, L. (2010). The econometrics of high frequency data. In *Statistical Methods for Stochastic Differential Equations* (M. Kessler, A. Lindner and M. Sørensen, eds.). Chapman & Hall/CRC, Boca Raton, FL.

[53] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Springer, Berlin. MR2200233

[54] NUALART, D., ÜSTÜNEL, A. S. and ZAKAI, M. (1988). On the moments of a multiple Wiener–Itô integral and the space induced by the polynomials of the integral. *Stochastics* **25** 233–240. MR0999370

[55] PEDERSEN, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Stat.* **22** 55–71. MR1334067

[56] PHILLIPS, P. C. and YU, J. (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In *Handbook of Financial Time Series* (T. Andersen, R. Davis, J. P. Kreib and T. Mikosch, eds.). Springer, New York.

[57] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357

[58] SCHAUMBURG, E. (2001). Maximum likelihood estimation of jump processes with applications to finance. Ph.D. thesis, Princeton Univ.

[59] SØRENSEN, M. (2012). Estimating functions for diffusion-type processes. In *Statistical Methods for Stochastic Differential Equations* (M. Kessler, A. Lindner and M. Sørensen, eds.). *Monographs on Statistics and Applied Probability* **124** 1–107. Chapman and Hall/CRC, Boca Raton, FL. MR2976982

[60] STRAMER, O. and YAN, J. (2007). On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. *J. Comput. Graph. Statist.* **16** 672–691. MR2351085

[61] TANG, C. Y. and CHEN, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *J. Econometrics* **149** 65–81. MR2515045

[62] TOCINO, A. (2009). Multiple stochastic integrals with Mathematica. *Math. Comput. Simulation* **79** 1658–1667. MR2488113

[63] UCHIDA, M. and YOSHIDA, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.* **122** 2885–2924. MR2931346

[64] UEMURA, H. (1987). On a short time expansion of the fundamental solution of heat equations by the method of Wiener functionals. *J. Math. Kyoto Univ.* **27** 417–431. MR0910227

[65] WATANABE, S. (1987). Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. *Ann. Probab.* **15** 1–39. MR0877589

[66] XIU, D. (2011). Dissecting and deciphering European option prices using closed-form series expansion. Technical report, Univ. Chicago Booth School of Business.

[67] YOSHIDA, N. (1992). Asymptotic expansion for statistics related to small diffusions. *J. Japan Statist. Soc.* **22** 139–159. MR1212246

[68] 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. MR1165514

[69] YOSHIDA, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41** 220–242. MR1172898

[70] YOSHIDA, N. (1997). Malliavin calculus and asymptotic expansion for martingales. *Probab. Theory Related Fields* **109** 301–342. MR1481124

[71] YOSHIDA, N. (2001). Malliavin calculus and martingale expansion. *Bull. Sci. Math.* **125** 431–456. MR1869987

[72] 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. MR2413501

[73] YU, J. and PHILLIPS, P. C. (2001). A Gaussian approach for estimating continuouse time models of short term interest rates. *Econom. J.* **4** 211–225.

DEPARTMENT OF BUSINESS STATISTICS
AND ECONOMETRICS
GUANGHUA SCHOOL OF MANAGEMENT
PEKING UNIVERSITY
BEIJING, 100871
P.R. CHINA
E-MAIL: cxli@gsm.pku.edu.cn