On the Approximate Maximum Likelihood Estimation for Diffusion Processes

Jinyuan Chang[1] and Song Xi Chen[1,2]

[1] Department of Business Statistics and Econometrics
Guanghua School of Management
Peking University
[2] Department of Statistics
Iowa State University

Emails: changjinyuan1986@yahoo.com.cn; songchen@iastate.edu

Abstract

The transitional density of a diffusion process is generally unknown, which prevents the full maximum likelihood estimation (MLE) based on discretely observed sample paths. Aït-Sahalia (1999, 2002) proposed Edgeworth type series approximations to the transitional densities of diffusion processes, which lead to the approximate maximum likelihood estimation (AMLE) for parameters. The consistency and the rate of convergence of the AMLE are established, which reveal the roles played by the number of terms used in the density approximation and the sampling length between successive observations. We find conditions under which the AMLE have the same asymptotic distribution as that of the full MLE. A first order approximation to the Fisher information matrix is proposed.

AMS 2000 subject classifications: Primary 62M05; Secondary 62F12.

Keywords and phrases: Asymptotic normality; Consistency; Diffusion process; Discrete observation; Edgeworth expabnsion; Maximum likelihood estimation.

1 Introduction

Continuous-time diffusion processes defined by stochastic differential equations (Øksendal, 2000) are the basic stochastic modeling tools in the modern financial theory and applications. Diffusion models are commonly employed to describe the price dynamics of a financial asset or a portfolio of assets. An eminent application is in deriving the price of a derivative contract on an asset or a group of assets. The celebrated Black-Scholes option pricing formula (Black and Scholes, 1973; Merton, 1973) is obtained by assuming that the log price process of the underlying asset follows an Ornstein-Uhlenbeck diffusion process. The widely used Vasicek (Vasicek, 1977) and Cox-Ingersol-Ross (Cox, Ingersoll and Ross, 1985) pricing formulae for the zero coupon bond were developed based on two specific mean-reverting diffusion processes with a constant or the square root (Feller, 1952) diffusion functions respectively. Other pricing formulae have also been developed for assets defined by other processes; see Bakshi, Cao and Chen (1997) and Dumas, Fleming and Whaley (1998). In the implementations of the aforementioned pricing formulae, the parameters of the diffusion processes which describe the underlying assets dynamics have to be estimated based on empirical observations. Sundaresan (2001) gave a comprehensive survey on the financial applications of continuous-time stochastic models which was largely the diffusion processes. Fan (2005) provided an overview on nonparametric estimation for diffusion processes. Other related works include Bibby and Sørensen (1995), Wang (2002), Fan and Zhang (2003), Fan and Wang (2007).

Estimating parameters of diffusion processes face several challenges. One is that despite being continuous-time models, the processes are only observed at discrete time points rather than observed continuously over time. The discrete observations prevent the use of the relatively straight forward likelihood expressions (Prakasa Rao, 1999) available for continuously observed diffusion processes. Another challenge is that despite the diffusion processes are Markovian, their transitional densities from one time point to the next are unknown except for only a few specific processes. This means that the efficient maximum likelihood estimation (MLE) can not be readily implemented for most of these processes.

In path breaking works, Aït-Sahalia (1999, 2002) established Edgeworth type series expansions to approximate the transitional densities of univariate diffusion processes. Similar expansions have been proposed for multivariate processes in Aït-Sahalia (2008). These density approximations, as advocated by Aït-Sahalia, are then employed to form approximated likelihood functions, which are maximized to obtain the approximate maximum likelihood estimators (AMLEs). Aït-Sahalia (2002, 2008) demonstrated that the approximate likelihood converges to the true likelihood as the number of terms in the series expansions goes to infinity. He also provided some results on the consistency of the AMLEs. Numerical evaluations of the transitional density approximations as conducted in Aït-Sahalia (1999), Stramer and Yan (2007a, 2007b) and others have shown good performance in the numerical approximation of the underlying transitional densities. The approach has opened a very accessible route for obtaining parameter estimators for diffusion processes, and for estimating other quantities which are functional of the transitional density, as commonly encountered in finance. Indeed, Aït-Sahalia and Kimmel (2005, 2010) demonstrated two such applications in stochastic volatility models and the affine term structure models, respectively. Tang and Chen (2009) provided some results on the approximated MLE based on the one-term expansion for the mean-reverting processes. They revealed that there is an extra leading order bias term in the AMLE due to the density approximation.

Despite the above mentioned results on the transitional density approximation and the AMLE, there are some key questions to be addressed. One is on the consistency of the AMLE. While Aït-Sahalia (2002, 2008) contained some results on consistency, there is much to be explored. There are two key ingredients in Aït-Sahalia's density approximation. One is J, the number of terms used in the Edgeworth type approximation, and the other is δ, the length of the sampling interval between successive observations. In this paper, we study explicitly the roles played by J and δ on the consistency of the AMLE, and quantify their roles on the rate of convergence. Another question is under what conditions on J and δ, the AMLE has the same asymptotic distribution as the full MLE. Here, we consider two regimes: (i) δ is fixed and $J \to \infty$; (ii) J is fixed but $\delta \to 0$, representing two views of asymptotics. In the case of $\delta \to 0$, it is found that $J \geqslant 2$ is necessary to ensure the AMLE having the same asymptotic normality as the MLE. Like the transitional density, the Fisher information matrix, the quantity that defines the efficiency of the full MLE, is unknown analytically, even the underlying transitional density is known. We show in this paper an approximation to the Fisher information matrix can be obtained based on the one-term density approximation.

The paper is organized as follows. In Section 2, we outline the transitional density approximations of Aït-Sahalia (1999, 2002). Some preliminary analysis needed for studying the AMLE is presented in Section 3. Section 4 establishes the consistency and the rates of convergence of the AMLE. Asymptotic normality of the AMLE and its equivalence to the full MLE are addressed in Section 5. Section 6 discuss the approximation for the Fisher information matrix. Simulation results are reported in Section 7. Technical conditions and details of proofs are relegated to Appendix.

2 Transitional Density Approximation

Consider a univariate diffusion process $(X_t)_{t\geq 0}$ defined by a stochastic differential equation

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dB_t, \tag{2.1}$$

where μ and σ are respectively the drift and diffusion functions, B_t is the standard Brownian motion. Both the drift and diffusion functions are known except for an unknown parameter vector θ taking values in a set $\Theta \subseteq \mathbb{R}^d$.

Given a sampling interval $\delta > 0$, let $f_X(x|x_0,\delta;\theta)$ be the transitional density of $X_{t+\delta}$ given $X_t = x_0$ for $(x_0, x) \in \mathcal{X} \times \mathcal{X}$, where $\mathcal{X}$ is the domain of X_t. Despite the parametric forms of the drift and the diffusion functions are available in (2.1), a closed-form expression for $f_X(x|x_0,\delta;\theta)$ is not generally available for most of the processes. In most cases, the density is only known to satisfy the Kolmogorov backward and forward partial differential equations. In path-breaking works, Aït-Sahalia (1999, 2002) proposed Edgeworth type expansions to approximate the transitional density.

The approach of Aït-Sahalia is the following. He first transformed X_t to a diffusion process with unit diffusion function by

$$Y_t = \gamma(X_t;\theta) := \int^{X_t} \frac{du}{\sigma(u;\theta)}, \tag{2.2}$$

which satisfies

$$dY_t = \mu_Y(Y_t;\theta)dt + dB_t,$$

where

$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2}\frac{\partial\sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta).$$

Let $f_Y(y|y_0,\delta;\theta)$ be the transitional density of $Y_{t+\delta}$ given $Y_t = y_0$. The two density functions are related according to

$$f_X(x_t|x_{t-1},\delta;\theta) = \frac{f_Y(\gamma(x_t;\theta)|\gamma(x_{t-1};\theta),\delta;\theta)}{\sigma(x_t;\theta)}. \tag{2.3}$$

To ensure convergence of the expansions, Aït-Sahalia standardized $Y_{t+\delta}$ by $Z_t = \delta^{-1/2}(Y_{t+\delta} - y_0)$. Let $f_Z(z|y_0,\delta;\theta)$ denote the conditional density of $Z_{t+\delta}$ given $Z_t = 0$, which is related to f_Y by

$$f_Z(z|y_0,\delta;\theta) = \delta^{1/2} f_Y(\delta^{1/2}z + y_0|y_0,\delta;\theta).$$

Let $\{H_j(z)\}_{j=1}^{\infty}$ be the Hermite polynomials which are defined by

$$H_j(z) = \phi^{-1}(z)(-1)^j \frac{d^j\phi(z)}{dz^j},$$

which are orthogonal with respect to the standard normal density ϕ, namely $\int H_j(z)H_k(z)\phi(x)dx = 0$ if $j \neq k$. A formal Hermite orthogonal series expansion to the density $f_Z(z|y_0, \delta; \theta)$ is

$$f_Z^H(z|y_0, \delta; \theta) = \phi(z) \sum_{j=0}^{\infty} \eta_j(y_0, \delta; \theta) H_j(z), \tag{2.4}$$

where the coefficients

$$\begin{aligned}\eta_j(y_0, \delta; \theta) &= (j!)^{-1} \int H_j(z) f_Z(z|y_0, \delta; \theta) dz = (j!)^{-1} \int H_j\{\delta^{-1/2}(y - y_0)\} f_Y(y|y_0, \delta; \theta) dy \\ &= (j!)^{-1} \mathbb{E}\left[H_j(\delta^{-1/2}(Y_{t+\delta} - y_0))|Y_t = y_0\right].\end{aligned}$$

The last conditional expectation is unknown. Aït-Sahalia proposed Taylor expansions for this conditional expectation with respect to the sampling interval δ based on the infinitesimal generator of Y_t. For twice continuously differentiable function g, the infinitesimal generator of Y_t is

$$\mathcal{A}_\theta g(y) = \mu_Y(y; \theta) \frac{\partial g}{\partial y} + \frac{1}{2} \frac{\partial^2 g}{\partial y^2}. \tag{2.5}$$

A K-term Taylor series expansion to $\mathbb{E}[H_j(\delta^{-1/2}(Y_{t+\delta} - y_0))|Y_t = y_0]$ is

$$\begin{aligned}\mathbb{E}\left[H_j(\delta^{-1/2}(Y_{t+\delta} - y_0))|Y_t = y_0\right] = &\sum_{k=0}^{K} {\mathcal{A}_\theta}^k H_j(\delta^{-1/2}(y - y_0))|_{y=y_0} \delta^k / k! \\ &+ \mathbb{E}\left[{\mathcal{A}_\theta}^{k+1} H_j(\delta^{-1/2}(Y_{t+\delta^*} - y_0))|Y_t = y_0\right] \delta^{k+1}/(k+1)!.\end{aligned} \tag{2.6}$$

Substituting the approximation (2.6) to the orthogonal expansion (2.4) followed by gathering terms according to the powers of δ, a J-term approximation to the transitional density $f_Y(y, \delta|y_0; \theta)$ is

$$f_Y^{(J)}(y|y_0, \delta; \theta) = \delta^{-1/2} \phi\left(\frac{y - y_0}{\delta^{1/2}}\right) \exp\left(\int_{y_0}^{y} \mu_Y(u; \theta) du\right) \sum_{j=0}^{J} c_j(y|y_0; \theta) \delta^j / j!,$$

where $c_0(y|y_0; \theta) \equiv 1$ and for $j \geqslant 1$,

$$c_j(y|y_0; \theta) = j(y - y_0)^{-j} \int_{y_0}^{y} (w - y_0)^{j-1} \cdot \left\{\lambda_Y(w; \theta) c_{j-1}(w|y_0; \theta) + \frac{1}{2} \frac{\partial^2 c_{j-1}(w|y_0; \theta)}{\partial w^2}\right\} dw.$$

Here $\lambda_Y(y; \theta) = -\{\mu_Y^2(y; \theta) + \partial \mu_Y(y; \theta)/\partial y\}/2$.

Transforming back from y to x via (2.2) and (2.3), the J-term expansion to $f_X(x|x_0, \delta; \theta)$ is

$$\begin{aligned}f_X^{(J)}(x|x_0, \delta; \theta) = &\sigma^{-1}(x; \theta) \delta^{-1/2} \phi\left(\frac{\gamma(x; \theta) - \gamma(x_0; \theta)}{\delta^{1/2}}\right) \\ &\cdot \exp\left\{\int_{x_0}^{x} \frac{\mu_Y(\gamma(u; \theta); \theta)}{\sigma(u; \theta)} du\right\} \sum_{j=0}^{J} c_j(\gamma(x; \theta)|\gamma(x_0; \theta); \theta) \delta^j / j!.\end{aligned} \tag{2.7}$$

Aït-Sahalia (2002) demonstrated that as $J \to \infty$,

$$f_X^{(J)}(x|x_0, \delta; \theta) \to f_X(x|x_0, \delta; \theta) \tag{2.8}$$

uniformly with respect to $\theta \in \Theta$ and x_0 over compact subsets of $\mathcal{X}$. The convergence is also uniformly with respect to x over subsets of $\mathcal{X}$ depending on the property of $\sigma(x; \theta)$.

If $\sum_{j=0}^{\infty} |c_j(y|y_0, \delta; \theta)|\delta^j/j! < \infty$ on $\mathcal{Y} \times \mathcal{Y}$ with probability one, where $\mathcal{Y}$ is the domain of Y_t, we can define $\widetilde{A}_3(x|x_0, \delta; \theta) = \log\{\sum_{j=0}^{\infty} c_j(y|y_0; \theta)\delta^j/j!\}$. Then, the result in (2.8) implies that

$$\log f_X(x|x_0, \delta; \theta) = -\log\sqrt{2\pi\delta} + A_1(x|x_0, \delta; \theta) + A_2(x|x_0, \delta; \theta) + \widetilde{A}_3(x|x_0, \delta; \theta), \tag{2.9}$$

where

$$\begin{aligned} A_1(x|x_0, \delta; \theta) &= -\log\{\sigma(x; \theta)\} - \frac{1}{2\delta}\{\gamma(x; \theta) - \gamma(x_0; \theta)\}^2, \\ A_2(x|x_0, \delta; \theta) &= \int_{x_0}^{x} \frac{\mu_Y(\gamma(u; \theta); \theta)}{\sigma(u; \theta)} du \quad \text{and} \\ A_3(x|x_0, \delta; \theta) &= \log\left\{\sum_{j=0}^{J} c_j(\gamma(x; \theta)|\gamma(x_0; \theta); \theta)\delta^j/j!\right\}. \end{aligned}$$

Expression (2.9) is the starting point for our analysis.

Given a set of discrete observations $\{X_{t\delta}\}_{t=1}^{n}$ with equal sampling length δ of the diffusion process $(X_t)_{t\geq 0}$, to simplify notations, we write X_t for $X_{t\delta}$, and hide δ in the expressions for the transitional density f_X and its approximations. At the same time, we use f and $f^{(J)}$ to express f_X and $f_X^{(J)}$ respectively. Based on the J-term expansion (2.7), the J-term approximate log-likelihood function given in Aït-Sahalia (2002) is

$$\begin{aligned} \ell_n^{(J)}(\theta) &= \sum_{t=1}^{n} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta) \\ &= -n\log\sqrt{2\pi\delta} + \sum_{t=1}^{n} A_1(X_t|X_{t-1}, \delta; \theta) + \sum_{t=1}^{n} A_2(X_t|X_{t-1}, \delta; \theta) + \sum_{t=1}^{n} A_3(X_t|X_{t-1}, \delta; \theta). \end{aligned}$$

Let $\hat{\theta}_n^{(J)} = \arg\max_{\theta\in\Theta} \ell_n^{(J)}(\theta)$ be the approximate MLE (AMLE) and $\hat{\theta}_n$ be the true MLE that maximizes the full likelihood $\ell_n(\theta) = \sum_{t=1}^{n} \log f(X_t|X_{t-1}, \delta; \theta)$.

We consider two asymptotic regimes in our analysis. The first regime is that

$$\delta \text{ is fixed but } J \to \infty.$$

which is the situation considered in Aït-Sahalia (2002). The second regime allows that

$$J \text{ is fixed, } \delta \to 0 \text{ but } n\delta \to \infty,$$

which is more tuned with an implementation of the density approximation with a fixed number of terms. In this case, to ensure the consistency and asymptotic normality of the AMLE, δ has to converge to zero at an appropriate speed while the total amount of observation time $T =: n\delta \to \infty$. More specifications on the speed of $\delta \to 0$ will be provided when we study the consistency and asymptotic distribution of the AMLE.

3 Preliminaries

Under regular circumstances as assumed by Condition (A.2) (ii) in Appendix, the full MLE $\hat{\theta}_n$ and the J-term approximate MLE $\hat{\theta}_n^{(J)}$ satisfy their respective likelihood score equations so that

$$\sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1}, \delta; \hat{\theta}_n) = \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f^{(J)}(X_t|X_{t-1}, \delta; \hat{\theta}_n^{(J)}) = 0. \tag{3.1}$$

Subtracting $\sum_{t=1}^{n} \partial \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0)/\partial\theta$ from both sides of (3.1),

$$\begin{aligned} & \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f^{(J)}(X_t|X_{t-1}, \delta; \hat{\theta}_n^{(J)}) - \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0) \\ = & \sum_{t=1}^{n} \frac{\partial}{\partial\theta} [\widetilde{A}_3(X_t|X_{t-1}, \delta; \theta_0) - A_3(X_t|X_{t-1}, \delta; \theta_0)] \\ & + \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1}, \delta; \hat{\theta}_n) - \sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1}; \theta_0). \end{aligned} \tag{3.2}$$

Carrying out Taylor expansions on both sides of (3.2),

$$\begin{aligned} & n^{-1} \sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0)(\hat{\theta}_n^{(J)} - \theta_0) \\ & + \frac{1}{2} \left[E_d \otimes (\hat{\theta}_n^{(J)} - \theta_0)^T \right] \cdot n^{-1} \sum_{t=1}^{n} \frac{\partial^3}{\partial\theta\partial\theta^T\partial\theta} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta^*) \cdot (\hat{\theta}_n^{(J)} - \theta_0) \\ = \quad & n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial\theta} [\widetilde{A}_3(X_t|X_{t-1}, \delta; \theta_0) - A_3(X_t|X_{t-1}, \delta; \theta_0)] \\ & + n^{-1} \sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f(X_t|X_{t-1}, \delta; \theta_0)(\hat{\theta}_n - \theta_0) \\ & + \frac{1}{2} \left[E_d \otimes (\hat{\theta}_n - \theta_0)^T \right] \cdot n^{-1} \sum_{t=1}^{n} \frac{\partial^3}{\partial\theta\partial\theta^T\partial\theta} \log f(X_t|X_{t-1}, \delta; \theta^{**}) \cdot (\hat{\theta}_n - \theta_0) \end{aligned} \tag{3.3}$$

where E_d is the $d \times d$ identity matrix, θ^* is on the joint line between $\hat{\theta}_n^{(J)}$ and θ_0, and θ^{**} is on the joint line between $\hat{\theta}_n$ and θ_0. Here we define

$$\frac{\partial^3}{\partial\theta\partial\theta^T\partial\theta} \log f(X_t|X_{t-1}, \delta; \theta) := \begin{pmatrix} \partial^3 \log f(X_t|X_{t-1}, \delta; \theta)/\partial\theta\partial\theta^T\partial\theta_1 \\ \vdots \\ \partial^3 \log f(X_t|X_{t-1}, \delta; \theta)/\partial\theta\partial\theta^T\partial\theta_d \end{pmatrix},$$

which is a $d^2 \times d$ matrix, and $\partial^3 \log f^{(J)}(X_t|X_{t-1}, \delta; \theta^*)/\partial\theta\partial\theta^T\partial\theta$ is similarly defined.

Let $\Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0)$ and $\Delta_{n2}(\hat{\theta}_n, \theta_0)$ denote the last terms on the left and right hand sides of

(3.3), respectively. Furthermore, let

$$F_n(\theta_0, J, \delta) = n^{-1} \sum_{t=1}^{n} \frac{\partial^2 [\widetilde{A}_3(X_t|X_{t-1}, \delta; \theta_0) - A_3(X_t|X_{t-1}, \delta; \theta_0)]}{\partial\theta\partial\theta^T},$$

$$U_n(\theta_0, J, \delta) = n^{-1} \sum_{t=1}^{n} \frac{\partial [\widetilde{A}_3(X_t|X_{t-1}, \delta; \theta_0) - A_3(X_t|X_{t-1}, \delta; \theta_0)]}{\partial\theta} \quad \text{and}$$

$$N_n(\theta_0, J, \delta) = n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0)}{\partial\theta\partial\theta^T}.$$

Then, (3.3) can be written as

$$\begin{aligned} &N_n(\theta_0, J, \delta)(\hat{\theta}_n^{(J)} - \theta_0) + \Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0) \\ &= U_n(\theta_0, J, \delta) + [N_n(\theta_0, J, \delta) + F_n(\theta_0, J, \delta)]\,(\hat{\theta}_n - \theta_0) + \Delta_{n2}(\hat{\theta}_n, \theta_0). \end{aligned} \tag{3.4}$$

This expression is useful in our studies for the consistency and asymptotic distribution for the AMLE $\hat{\theta}_n^{(J)}$ for (i) fixed δ but $J \to \infty$ and (ii) $J \geqslant 2$ being fixed and $\delta \to 0$. A more elaborate expansion than (3.4) with quadratic terms will be provided for the case of $J = 1$ while $\delta \to 0$.

Let $||A||_2 = \{\rho(A^T A)\}^{1/2}$ be the spectral norm of a matrix A, where $\rho(A^T A)$ denotes the largest eigen-value of $A^T A$. The following proposition describes properties for the quantities appeared in (3.4).

Proposition 1 *Under Conditions (A.1), (A.3)-(A.4), (A.6)-(A.7) given in Appendix, there exists a positive constant Δ such that for any positive integer J and $\delta \in (0, \Delta)$,*

(a) $\mathbb{E}\{F_n(\theta_0, J, \delta)\}$, $\mathbb{E}\{U_n(\theta_0, J, \delta)\}$ and $\mathbb{E}\{N_n(\theta_0, J, \delta)\}$ exist;

(b) $\Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0) = O_p\{||\hat{\theta}_n^{(J)} - \theta_0||_2^2\}$ and $\Delta_{n2}(\hat{\theta}_n, \theta_0) = O_p\{||\hat{\theta}_n - \theta_0||_2^2\}$.

Let $I(\delta) = I(\delta; \theta_0)$ be the Fisher information matrix, which we assume is invertible in Condition (A.5). It is expected that the expected value of $N_n(\theta_0, J, \delta)$, denoted by $N(\theta_0, J, \delta)$, will converge to $-I(\delta)$, as $J \to \infty$ for each fixed δ or J being fixed but $\delta \to 0$. The following proposition confirms this and provides the speed of the convergence.

Proposition 2 *Under Conditions (A.1), (A.4), (A.6)-(A.7) given in Appendix, there exists a constant C, that is not dependent on J and δ, and a $\bar{\Delta} > 0$ such that for any positive integer J and $\delta \in (0, \bar{\Delta})$,*

$$||N(\theta_0, J, \delta) + I(\delta)||_2 \leqslant C\delta^{J+1}.$$

As $I(\delta)$ is invertible for each fixed $\delta > 0$, $N_n(\theta_0, J, \delta)$ will be invertible with probability approaching one as $J \to \infty$ with δ being fixed. However, under the second asymptotic regime where $\delta \to 0$, the limit of the Fisher information $I(0) := \lim_{\delta\to 0} I(\delta)$, as well as $N(\theta_0, J, 0)$, may be singular. This is the case for some Ornstein-Uhlenbeck processes (Vasicek model) as shown in Section 6.

The following proposition provides another account on $N(\theta_0, J, \delta)$ and its deviation from $-I(\delta)$, as well as the convergence of $N^{-1}(\theta_0, J, \delta)U(\theta_0, J, \delta)$, where $U(\theta_0, J, \delta)$ denote the expected value of $U_n(\theta_0, J, \delta)$.

Proposition 3 *Under Conditions (A.1), (A.3)-(A.7) given in Appendix,*

$$||N^{-1}(\theta_0, J, \delta)I(\delta) + E_d||_2 = O(\delta^J) \quad \textit{and} \quad ||N^{-1}(\theta_0, J, \delta)U(\theta_0, J, \delta)||_2 = O(\delta^J)$$

under either (i) for any fixed $\delta \in (0, \bar{\Delta})$, *where* $\bar{\Delta}$ *is the quantity in Proposition 2, and* $J \to \infty$ *or (ii) for any fixed* J, $\delta \to 0$.

4 Consistency

We consider in this section the consistency of the AMLE $\hat{\theta}_n^{(J)}$ and establish its convergence rate. For a fixed sampling interval δ, Aït-Sahalia (2002) proved that there existed a sequence $J_n \to \infty$ such that $\hat{\theta}_n^{(J_n)} - \hat{\theta}_n \xrightarrow{p} 0$ under P_{θ_0} as $n \to \infty$, where P_{θ_0} is the underlying probability measure. Based on the consistency of $\hat{\theta}_n$, we know that the consistency of $\hat{\theta}_n^{(J_n)}$ is hold. For a fixed J, Aït-Sahalia (2008) proved that there existed a sequence $\{\delta_n\}$ vanishing to zero such that $\sqrt{n}I^{1/2}(\delta_n)(\hat{\theta}_n^{(J)} - \theta_0) = O_p(1)$.

In this paper, we will give more explicit guidelines on how to select the afore-mentioned sequences J_n and δ_n so that the AMLE is consistent. We will also establish the rates of convergence of $\hat{\theta}_n^{(J)}$. Our study begins with (3.1). The following two propositions are needed, whose proofs are given in Appendix.

Proposition 4 *Under Conditions (A.1), (A.4), (A.6)-(A.7) given in Appendix, there exist two finite positive constants* $\widetilde{\Delta}$ *and* C, *not dependent on* J *and* δ, *such that for any* J *and* $\delta \in (0, \widetilde{\Delta}]$,

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\theta) - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f^{(J)}(X_t|X_{t-1},\delta;\theta)\right|\right|_2\right\} \leqslant C\delta^{J+1}.$$

Proposition 5 *Under Conditions (A.1), (A.3)-(A.4), (A.6)-(A.7) given in Appendix, there exists a constant* $\dot{\Delta} > 0$ *such that*

$$\sup_{\theta\in\Theta}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\theta) - \mathbb{E}\left\{\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\theta)\right\}\right|\right|_2 \xrightarrow{p} 0$$

for (i) $\delta \in (0, \dot{\Delta}]$ *being fixed,* $n \to \infty$, *or (ii)* $n \to \infty$, $\delta \to 0$ *but* $n\delta \to \infty$.

From Proposition 4,

$$\sup_{\theta\in\Theta}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\theta) - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f^{(J)}(X_t|X_{t-1},\delta;\theta)\right|\right|_2 \xrightarrow{p} 0$$

for either (i) $\delta \in (0, \widetilde{\Delta}]$ being fixed, $J \to \infty$ and $n \to \infty$, or (ii) J being fixed, $n \to \infty$ and $\delta \to 0$.

Based on these two propositions,

$$
\begin{aligned}
&\left\| \mathbb{E}\left\{ \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1},\delta;\hat{\theta}_n^{(J)}) \right\} \right\|_2 \\
\leqslant\ & \sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1},\delta;\theta) - \mathbb{E}\left\{ \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1},\delta;\theta) \right\} \right\|_2 \\
&+ \sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f(X_t|X_{t-1},\delta;\theta) - \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta} \log f^{(J)}(X_t|X_{t-1},\delta;\theta) \right\|_2 \\
\xrightarrow{p}\ & 0,
\end{aligned}
$$

for either (i) $\delta \in (0, \widetilde{\Delta} \wedge \dot{\Delta}]$ being fixed, $J \to \infty$ and $n \to \infty$, or (ii) J being fixed, $n \to \infty$, $\delta \to 0$ but $n\delta \to \infty$. Hence, noting Condition (A.2) (i), we have the consistency of the AMLE $\hat{\theta}_n^{(J)}$.

Theorem 1 *Under Conditions (A.1)-(A.4), (A.6)-(A.7) given in Appendix,* $\hat{\theta}_n^{(J)} - \theta_0 \xrightarrow{p} 0$ *under either (i)* $\delta \in (0, \widetilde{\Delta} \wedge \dot{\Delta}]$ *being fixed;* $J \to \infty$ *and* $n \to \infty$*, or (ii)* J *being fixed,* $n \to \infty$*,* $\delta \to 0$ *but* $n\delta \to \infty$*.*

From Proposition 2 and Condition (A.5), multiply $N^{-1}(\theta_0, J, \delta)$ on both side of (3.4), we have

$$
\begin{aligned}
\hat{\theta}_n^{(J)} - \theta_0 = \quad & N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) - N^{-1}(N_n - N)(\hat{\theta}_n^{(J)} - \theta_0) \\
& - N^{-1}\Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0) + N^{-1}\Delta_{n2}(\hat{\theta}_n, \theta_0).
\end{aligned} \tag{4.1}
$$

From this together with Proposition 4 and Theorem 1, we can establish the rate of convergence for the AMLE.

Theorem 2 *Under Conditions (A.1)-(A.7) given in Appendix,*

$$
\hat{\theta}_n^{(J)} - \theta_0 = \begin{cases} O_p(\delta^{J+1} + n^{-1/2}\delta^{-1/2}), & \text{if } \delta \in (0, \widetilde{\Delta} \wedge \dot{\Delta}] \text{ is fixed and } J \to \infty; \\ O_p(\delta^{J} + n^{-1/2}\delta^{-1/2}), & \text{if } J \text{ is fixed, } \delta \to 0 \text{ but } n\delta^2 \to \infty. \end{cases}
$$

The above theorem reveals the impacts of the sampling interval δ and the number of terms J used in the density approximation on the convergence rate. In particular, the rate for the AMLE has an extra δ^{J+1} or δ^J term in addition to the standard rate $(n\delta)^{-1/2}$ of the full MLE. This extra term is the result of the density approximation. And its particular form suggests that the sampling interval δ has to be less than 1 in order to make the AMLE $\hat{\theta}_n^{(J)}$ converge to θ_0. It is apparent that the higher the J is, the less impact the extra term has on the AMLE $\hat{\theta}_n^{(J)}$.

5 Asymptotic Distribution

In this section, we consider the asymptotic distribution of the AMLE $\hat{\theta}_n^{(J)}$. Throughout this section, as in Condition (A.2) (iii), we assume the full MLE $\hat{\theta}_n$ is asymptotically normally distributed such that for any $\delta > 0$,

$$\sqrt{n} I^{1/2}(\delta)(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, E_d) \quad \text{as } n \to \infty. \tag{5.2}$$

The question of this section is under what conditions the AMLE has the same asymptotic normality as the full MLE. Our investigations are organized according to the two asymptotic regimes mentioned at the end of Section 2.

5.1 Fixed δ and $J \to \infty$

This is a simple case to treat. Under this setting, we note from Proposition 2 and Condition (A.5) that, $N^{-1}(\theta_0, J, \delta) = O(1)$ uniformly for any J. Utilizing the result in Theorem 2, the expansion (3.4) becomes

$$\begin{aligned}
&\hat{\theta}_n^{(J)} - \theta_0 \\
&= N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J+1} + n^{-1}\delta^{-1} + \delta^{2J+2}) \\
&= N^{-1}U_n - N^{-1}I(\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J+1} + n^{-1}\delta^{-1/2} + \delta^{2J+2}) \\
&= N^{-1}U_n + (\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J-1/2} + n^{-1}\delta^{-1/2} + \delta^{2J+2}).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sqrt{n} I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \\
&= \sqrt{n} I^{1/2}(\delta) N^{-1} U_n + \sqrt{n} I^{1/2}(\delta)(\hat{\theta}_n - \theta_0) + O_p(\delta^{J-1/2} + n^{-1/2}\delta^{-1/2} + n^{1/2}\delta^{2J+2}).
\end{aligned}$$

Since $U_n = O_p(\delta^{J+1})$, if $n\delta^{2J+2} \to 0$, then $\sqrt{n} I^{1/2}(\delta) N^{-1} U_n = o_p(1)$ and

$$\sqrt{n} I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \xrightarrow{d} N(0, E_d).$$

Therefore, the AMLE has the same asymptotic distribution as the full MLE $\hat{\theta}_n$. This is attained by requesting $n\delta^{2J+2} \to 0$ in addition to $J \to \infty$. The latter condition prescribes a rule on the selection of the $J = J_n(\delta)$. By choosing an $\epsilon > 0$ so that $\delta^{2J+2} = n^{-1-\epsilon}$ for each pair of n and δ, then

$$J = J_n(\delta) = \frac{-1-\epsilon}{2\log\delta}\log n - 1 > \frac{-1}{2\log\delta}\log n - 1.$$

The integer truncation of the above lower bound can be used as a reference value for the number of term used in the density approximation for each given pair of (n, δ).

Table 1 reports such reference values of J assigned by the above formula for a set of (n, δ) combinations commonly encountered in empirical studies. It shows that for monthly frequency or less ($\delta \leqslant 1/12$), one term approximation is adequate, and for $\delta = 1/4$, $J = 2$ is needed. However, there is a dramatic increase in J as the sampling length is larger than 1/4: demanding at least four terms for $\delta = 1/2$ (half yearly) or at least ten terms for $\delta = 3/4$. The number of

Table 1: The least approximation term selection to guarantee the AMLE has the same asymptotic distribution as the full MLE for special sampling interval δ and sample size n

δ	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
1/252	1	1	1	1
1/52	1	1	1	1
1/12	1	1	1	1
1/4	2	2	2	2
1/2	4	4	5	5
3/4	10	12	13	14

terms also increases for these higher δ values as n increases, although the rate of this increase is much slower than that as δ is increased. The latter may be understood that for a given δ, as n increases, the chance of having extreme values in the tails of the transitional distribution increases. As the density approximation is less accurate in the tails than in the main body of the distribution, there is a need for having more terms in the density approximation.

5.2 J fixed, $\delta \to 0$ and $n\delta \to \infty$

Our starting point is the expansion (4.1). As $N^{-1}(N_n - N) = o_p(1)$ if $n\delta^2 \to \infty$ which we will assume in the rest of this section when we are dealing with diminishing δ. It is readily understood that in order for $\hat{\theta}_n^{(J)}$ having the same asymptotic distribution as $\hat{\theta}_n$, it is required that

$$N^{-1}U_n,\ N^{-1}\Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0) \text{ and } N^{-1}\Delta_{n2}(\hat{\theta}_n, \theta_0) \text{ are all the smaller order of } \hat{\theta}_n^{(J)} - \theta_0. \tag{5.3}$$

We will demonstrate in the following that (5.3) can be attained by strengthening the condition that $n\delta^2 \to \infty$ to $n\delta^3 \to \infty$ while at the same time restricting $J \geqslant 2$. Hence, under these circumstances, $\hat{\theta}_n^{(J)}$ has the same asymptotic distribution as $\hat{\theta}_n$. Later we will demonstrate that this equivalence in the asymptotic distribution is quite unlikely for $J = 1$.

Our analysis needs to expand (3.3) to the quadratic terms. To this end, let us define

$$M_n(\theta_0, J, \delta) = n^{-1} \sum_{t=1}^{n} \frac{\partial^3}{\partial\theta\partial\theta^T\partial\theta} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0) \quad \text{and}$$

$$T_n(\theta_0, J, \delta) = n^{-1} \sum_{t=1}^{n} \frac{\partial^3}{\partial\theta\partial\theta^T\partial\theta} \log f(X_t|X_{t-1}, \delta; \theta_0).$$

By further expanding to quadratic terms, (4.1) can be written as

$$\begin{aligned}
& \hat{\theta}_n^{(J)} - \theta_0 \\
= \ & N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) - N^{-1}(N_n - N)(\hat{\theta}_n^{(J)} - \theta_0) \\
& - \frac{1}{2}N^{-1}\left[E_d \otimes (\hat{\theta}_n^{(J)} - \theta_0)^T\right] M_n(\hat{\theta}_n^{(J)} - \theta_0) + \frac{1}{2}N^{-1}\left[E_d \otimes (\hat{\theta}_n - \theta_0)^T\right] T_n(\hat{\theta}_n - \theta_0) \\
& - N^{-1}\tilde{\Delta}_{n1}(\hat{\theta}_n^{(J)}, \theta_0) + N^{-1}\tilde{\Delta}_{n2}(\hat{\theta}_n, \theta_0),
\end{aligned} \tag{5.4}$$

where $\tilde{\Delta}_{n1}(\hat{\theta}_n^{(J)}, \theta_0)$ and $\tilde{\Delta}_{n2}(\hat{\theta}_n, \theta_0)$ are remainder terms. By applying the same method in the proof of Proposition 1, it can be shown that $\tilde{\Delta}_{n1}(\hat{\theta}_n^{(J)}, \theta_0) = O_p\{||\hat{\theta}_n^{(J)} - \theta_0||_2^3\}$ and $\tilde{\Delta}_{n2}(\hat{\theta}_n, \theta_0) = O_p\{||\hat{\theta}_n - \theta_0||_2^3\}$.

In order to make $\hat{\theta}_n^{(J)}$ have the same asymptotic distribution as $\hat{\theta}_n$, the two quadratic terms on the right of (5.4) have to be smaller order of $\hat{\theta}_n^{(J)} - \theta_0$ and $\hat{\theta}_n - \theta_0$ respectively, namely

$$N^{-1}\left[E_d \otimes (\hat{\theta}_n^{(J)} - \theta_0)^T\right] \cdot M_n \cdot (\hat{\theta}_n^{(J)} - \theta_0) = o_p(||\hat{\theta}_n^{(J)} - \theta_0||_2)$$

or equivalently

$$N^{-1}\left[E_d \otimes (\hat{\theta}_n^{(J)} - \theta_0)^T\right] = o_p(1); \tag{5.5}$$

and

$$N^{-1}\left[E_d \otimes (\hat{\theta}_n - \theta_0)^T\right] \cdot T_n \cdot (\hat{\theta}_n - \theta_0) = o_p(||\hat{\theta}_n - \theta_0||_2)$$

or equivalently

$$n\delta^3 \to \infty, \tag{5.6}$$

since $\hat{\theta}_n - \theta_0 = O_p\{(n\delta)^{-1/2}\}$ and $N^{-1} = O(\delta^{-1})$.

As $\hat{\theta}_n^{(J)} - \theta_0 = O_p\{\delta^J + (n\delta)^{-1/2}\}$, (5.5) requires that $\delta^{J-1} + n^{-1/2}\delta^{-3/2} \to 0$. Hence, in order to make $\hat{\theta}_n^{(J)}$ have the same asymptotic distribution as $\hat{\theta}_n$, it is necessary to have

$$J \geqslant 2 \text{ and } n\delta^3 \to \infty. \tag{5.7}$$

It can be readily checked from (5.2) and (5.4) that under (5.7),

$$\sqrt{n} I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \xrightarrow{d} N(0, E_d).$$

Now we consider the case of $J = 1$. To ensure the remainder terms $N^{-1}\Delta_{n1}(\hat{\theta}_n^{(J)}, \theta_0)$ and $N^{-1}\Delta_{n2}(\hat{\theta}_n, \theta_0)$ are negligible, by a similar argument applied above for the case of $J \geqslant 2$, it is also necessary to assume $n\delta^3 \to \infty$. From Theorem 2, $\hat{\theta}_n^{(1)} - \theta_0 = O_p(\delta + (n\delta)^{-1/2})$. To gain insight on the situation, we need to find out the order of magnitude of the quadratic term in (5.4), namely the order of magnitude of

$$S_n = N^{-1}\left[E_d \otimes (\hat{\theta}_n^{(1)} - \theta_0)^T\right] M_n(\hat{\theta}_n^{(1)} - \theta_0) - N^{-1}\left[E_d \otimes (\hat{\theta}_n - \theta_0)^T\right] T_n(\hat{\theta}_n - \theta_0).$$

With this notation, (5.4) can be written as

$$\hat{\theta}_n^{(J)} - \theta_0 = N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) - \frac{1}{2}S_n + o_p\{(n\delta)^{-1/2}\} + O_p(\delta^2). \tag{5.8}$$

Define an operator between two vectors A and B:

$$A * B = \left[E_d \otimes A^T\right] \cdot M_n \cdot B + \left[E_d \otimes B^T\right] \cdot M_n \cdot A.$$

By repeated substitutions, it can be shown that

$$\begin{aligned} S_n = \; & \frac{1}{2}N^{-1}\left[(N^{-1}U_n) * (N^{-1}U_n)\right] + \frac{1}{2}N^{-1}\left[\left(\frac{1}{2}S_n\right) * \left(\frac{1}{2}S_n\right)\right] \\ & - N^{-1}\left[(N^{-1}U_n) * \left(\frac{1}{2}S_n\right)\right] + o_p(\delta). \end{aligned}$$

As $U_n = O_p(\delta^2)$ for $J = 1$ and $N^{-1} = O(\delta^{-1})$, it can be deduced from the above equation that $S_n = O_p(\delta)$. Hence, for $J = 1$ if we require $n\delta^3 \to \infty$, the quadratic term S_n will contribute to the leading order of $\hat{\theta}_n^{(1)} - \theta_0$. If we do not require $n\delta^3 \to \infty$, then the sum of remainder terms, $N^{-1}\tilde{\Delta}_{n1}(\hat{\theta}_n^{(J)}, \theta_0) + N^{-1}\tilde{\Delta}_{n2}(\hat{\theta}_n, \theta_0)$ will not be controlled. Hence, if $J = 1$, it is very likely that the asymptotic distribution of $\hat{\theta}_n^{(J)}$ will differ from that of $\hat{\theta}_n$ unless $U_n = 0$ with probability one. In the rare case of $U_n = 0$, it is possible for $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n$ to share the same limiting distribution.

Therefore, in order to guarantee that $\hat{\theta}_n^{(J)}$ has the same asymptotic distribution as $\hat{\theta}_n$ under $\delta \to 0$, we need to use the AMLE based on at least two-term expansions, while satisfying $n\delta^3 \to \infty$, which we will assume in the rest of this section.

Note that $\hat{\theta}_n^{(J)} - \theta_0 = O_p(\delta^J + n^{-1/2}\delta^{-1/2})$. Then,

$$\begin{aligned}
&\hat{\theta}_n^{(J)} - \theta_0 \\
&= N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J-1} + n^{-1}\delta^{-3/2}) + N^{-1} \cdot O_p(\delta^{2J} + n^{-1}\delta^{-1}) \\
&= N^{-1}U_n - N^{-1}I(\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J-1} + n^{-1}\delta^{-3/2}) + N^{-1} \cdot O_p(\delta^{2J} + n^{-1}\delta^{-1}) \\
&= N^{-1}U_n + (\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J-1} + n^{-1}\delta^{-3/2}) + N^{-1} \cdot O_p(\delta^{2J} + n^{-1}\delta^{-1}).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
& \sqrt{n}I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \\
= \; & \sqrt{n}I^{-1/2}IN^{-1}U_n + \sqrt{n}I^{1/2}(\delta)(\hat{\theta}_n - \theta_0) + O_p(\delta^{J-1} + n^{-1/2}\delta^{-3/2}) \\
& + \sqrt{n}I^{-1/2}IN^{-1} \cdot O_p(\delta^{2J} + n^{-1}\delta^{-1}) \\
= \; & \sqrt{n}I^{1/2}(\hat{\theta}_n - \theta_0) + O_p(n^{1/2}\delta^{J+1/2}) + O_p(\delta^{J-1} + n^{-1/2}\delta^{-3/2} + n^{1/2}\delta^{2J-1/2}).
\end{aligned}$$

Hence, for any $J \geqslant 2$ such that $n\delta^3 \to \infty$ and $n\delta^{2J+1} \to 0$,

$$\sqrt{n}I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \xrightarrow{d} N(0, E_d).$$

This result shows that, when δ vanishes to zero, in order to guarantee the AMLE has the same asymptotic distribution as full MLE, we need to pick the approximation order $J \geqslant 2$, while maintaining $n\delta^3 \to \infty$ and $n\delta^{2J+1} \to 0$.

The following theorem summarizes the asymptotic normality under both asymptotic regimes.

Theorem 3 *Under Conditions (A.1)-(A.7) given in Appendix,*

$$\sqrt{n}I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0) \xrightarrow{d} N(0, E_d),$$

for (i) $\delta \in (0, \widetilde{\Delta} \wedge \dot{\Delta}]$ being fixed, $n \to \infty$, $J \to \infty$ but $n\delta^{2J+2} \to 0$ or (ii) $J \geqslant 2$ being fixed, $n \to \infty$, $\delta \to 0$ but $n\delta^3 \to \infty$ and $n\delta^{2J+1} \to 0$.

5.3 Asymptotic bias and variance

The remainder of this section is devoted to the consideration of the asymptotic bias and variance of the AMLE under the two asymptotic regimes. Given our analysis in the early part of this section, our consideration will be focused on the situations where the asymptotic normality of the AMLE can be assumed, namely under (i) δ being fixed, $J \to \infty$, $n \to \infty$ but $n\delta^{2J+2} \to 0$ or (ii) $J \geqslant 2$ being fixed, $\delta \to 0$, $n\delta^3 \to \infty$ but $n\delta^{2J+1} \to 0$.

In the case of δ being fixed and $J \to \infty$, from (5.4) and provided $n\delta^{2J+2} \to 0$, we have

$$\begin{aligned}
& \hat{\theta}_n^{(J)} - \theta_0 \\
= \;& N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) \\
& - N^{-1}(N_n - N)N^{-1}U_n - N^{-1}(N_n - N)N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) \\
& - \frac{1}{2}N^{-1}\left\{E_d \otimes \left[N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0)\right]^T\right\} M_n \left[N^{-1}U_n + N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0)\right] \\
& + \frac{1}{2}N^{-1}\left[E_d \otimes (\hat{\theta}_n - \theta_0)^T\right] T_n(\hat{\theta}_n - \theta_0) + O_p(n^{-3/2}) \\
= \;& N^{-1}U_n + \left[E_d - N^{-1}(N_n - N)\right] N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0) + O_p(n^{-1/2}\delta^{J+1}) + O_p(n^{-3/2}).
\end{aligned}$$

Then, the leading order bias of $\hat{\theta}_n^{(J)}$ is

$$B(\theta_0, J, \delta) = N^{-1}U + \mathbb{E}\left\{\left[E_d - N^{-1}(N_n - N)\right] N^{-1}(N_n + F_n)(\hat{\theta}_n - \theta_0)\right\}, \tag{5.9}$$

and the leading order variance is

$$V(\theta_0, J, \delta) = N^{-1}I(\delta)Var(\hat{\theta}_n)I(\delta)N^{-1}. \tag{5.10}$$

In the case of $J \geqslant 2$ being fixed, $\delta \to 0$ and $n\delta^3 \to \infty$ but $n\delta^{2J+1} \to 0$, it can be shown by a similar argument to that for the fixed δ case above, the asymptotic bias and variance have the same forms as (5.9) and (5.10), respectively. Both (5.9) and (5.10) will be used to calibrate with the simulated bias and variance in the simulation study in Section 7.

6 Approximating Fisher Information Matrix

We demonstrate in this section that the approximation of the transitional density provides a way to approximate the Fisher information matrix. Fisher information matrix $I(\delta)$ is a key quantity associated with inference based on the full MLE. It defines the asymptotic efficiency and the rate of convergence. From Proposition 2, a natural candidate to approximate $I(\delta)$ is $-N(\theta_0, J, \delta)$ based on the J-term expansion. To simplify our expedition, our consideration here is focused under the following diffusion process

$$dX_t = \mu(X_t; \eta)dt + \sigma(X_t; \xi)dB_t, \tag{6.1}$$

where $\eta = (\eta_1, \cdots, \eta_{d_1})^T$ and $\xi = (\xi_1, \cdots, \xi_{d_2})^T$ are distinct drift and diffusion parameters respectively. The whole parameter $\theta = (\eta^T, \xi^T)^T$. Here, we provide an explicit expression $N(\theta_0, 1, \delta)$ based on the one-term density expansion. Expressions for higher J values may be made via more extensive derivations.

Recall that the one-term ($J = 1$) transitional density approximation is

$$
\begin{aligned}
&\log f^{(1)}(x|x_0, \delta; \theta) \\
= &-\frac{1}{2}\log 2\pi\delta - \log\sigma(x;\xi) - \frac{1}{2\delta}\left(\gamma(x;\xi) - \gamma(x_0;\xi)\right)^2 + \int_{x_0}^{x}\left\{\frac{\mu(u;\eta)}{\sigma^2(u;\xi)} - \frac{1}{2\sigma(u;\xi)}\frac{\partial\sigma(u;\xi)}{\partial u}\right\}du \\
&+ \log\left\{1 + c_1(\gamma(x;\xi)|\gamma(x_0;\xi);\theta)\cdot\delta\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
c_1(\gamma(x;\xi)|\gamma(x_0;\xi);\theta) = \frac{1}{2}\Bigg\{ &-\left[\frac{\mu(x;\eta)}{\sigma(x;\xi)} - \frac{\mu(x_0;\eta)}{\sigma(x_0;\xi)}\right] + \frac{1}{2}\left[\frac{\partial\sigma(x;\xi)}{\partial x} - \frac{\partial\sigma(x_0;\xi)}{\partial x_0}\right] \\
&- \int_{x_0}^{x}\left[\frac{\mu(u;\eta)}{\sigma(u;\xi)} - \frac{1}{2}\frac{\partial\sigma(u;\xi)}{\partial u}\right]^2\frac{du}{\sigma(u;\xi)}\Bigg\}\Bigg/\int_{x_0}^{x}\frac{du}{\sigma(u;\xi)}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{\partial^2\log f^{(1)}}{\partial\eta_i\partial\eta_j} &= \int_{x_0}^{x}\frac{\partial^2\mu(u;\eta)}{\partial\eta_i\partial\eta_j}\frac{du}{\sigma^2(u;\xi)} + \delta\cdot\frac{\partial^2 c_1}{\partial\eta_i\partial\eta_j}\frac{1}{1+c_1\delta} - \delta^2\cdot\frac{\partial c_1}{\partial\eta_i}\frac{\partial c_1}{\partial\eta_j}\frac{1}{(1+c_1\delta)^2}, \\
\frac{\partial^2\log f^{(1)}}{\partial\eta_i\partial\xi_j} &= -2\int_{x_0}^{x}\frac{\partial\mu(u;\eta)}{\partial\eta_i}\frac{\partial\sigma(u;\xi)}{\partial\xi_j}\frac{du}{\sigma^3(u;\xi)} + \delta\cdot\frac{\partial^2 c_1}{\partial\eta_i\partial\xi_j}\frac{1}{1+c_1\delta} - \delta^2\cdot\frac{\partial c_1}{\partial\eta_i}\frac{\partial c_1}{\partial\xi_j}\frac{1}{(1+c_1\delta)^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\log f^{(1)}}{\partial\xi_i\partial\xi_j} = &-\frac{\partial^2\sigma(x;\xi)}{\partial\xi_i\partial\xi_j}\frac{1}{\sigma(x;\xi)} + \frac{\partial\sigma(x;\xi)}{\partial\xi_i}\frac{\partial\sigma(x;\xi)}{\partial\xi_j}\frac{1}{\sigma^2(x;\xi)} \\
&- \frac{1}{\delta}\int_{x_0}^{x}\frac{\partial\sigma(u;\xi)}{\partial\xi_i}\frac{du}{\sigma^2(u;\xi)}\int_{x_0}^{x}\frac{\partial\sigma(u;\xi)}{\partial\xi_j}\frac{du}{\sigma^2(u;\xi)} \\
&+ \frac{1}{\delta}\int_{x_0}^{x}\frac{du}{\sigma(u;\xi)}\int_{x_0}^{x}\left[\frac{\partial^2\sigma(u;\xi)}{\partial\xi_i\partial\xi_j}\frac{1}{\sigma^2(u;\xi)} - \frac{\partial\sigma(u;\xi)}{\partial\xi_i}\frac{\partial\sigma(u;\xi)}{\partial\xi_j}\frac{2}{\sigma^3(u;\xi)}\right]du \\
&+ \int_{x_0}^{x}\Bigg\{\left[\frac{6\mu(u;\xi)}{\sigma^4(u;\xi)} - \frac{\partial\sigma(u;\xi)}{\partial u}\frac{1}{\sigma^3(u;\xi)}\right]\frac{\partial\sigma(u;\xi)}{\partial\xi_i}\frac{\partial\sigma(u;\xi)}{\partial\xi_j} \\
&\qquad - \left[\frac{2\mu(u;\xi)}{\sigma^3(u;\xi)} - \frac{\partial\sigma(u;\xi)}{\partial u}\frac{1}{2\sigma^2(u;\xi)}\right]\frac{\partial^2\sigma(u;\xi)}{\partial\xi_i\partial\xi_j} \\
&\qquad + \left[\frac{\partial^2\sigma(u;\xi)}{\partial u\partial\xi_i}\frac{\partial\sigma(u;\xi)}{\partial\xi_j} + \frac{\partial^2\sigma(u;\xi)}{\partial u\partial\xi_j}\frac{\partial\sigma(u;\xi)}{\partial\xi_i}\right]\frac{1}{2\sigma^2(u;\xi)} \\
&\qquad - \frac{\partial^3\sigma(u;\xi)}{\partial u\partial\xi_i\partial\xi_j}\frac{1}{2\sigma(u;\xi)}\Bigg\}du \\
&+ \delta\cdot\frac{\partial^2 c_1}{\partial\xi_i\partial\xi_j}\frac{1}{1+c_1\delta} - \delta^2\cdot\frac{\partial c_1}{\partial\xi_i}\frac{\partial c_1}{\partial\xi_j}\frac{1}{(1+c_1\delta)^2}.
\end{aligned}
$$

Let μ_i, μ_{ij} and so on denote partial derivatives with respect to η_i, η_i and η_j, respectively; and σ_i and $\sigma_{x,j}$ and so on denote partial derivatives with respect to ξ_i, and x and ξ_j, respectively.

Then, it can be shown that

$$
\begin{aligned}
\left.\frac{\partial^2 c_1}{\partial\eta_i\partial\eta_j}\right|_{x=x_0} &= -\sigma^{-2}\mu_i\mu_j - \mu\sigma^{-2}\mu_{ij} + \sigma^{-1}\mu_{ij}\sigma_x - \frac{1}{2}\mu_{xij},\\
\left.\frac{\partial^2 c_1}{\partial\eta_i\partial\xi_j}\right|_{x=x_0} &= 2\mu\sigma^{-3}\mu_i\sigma_j - \sigma^{-2}\mu_i\sigma_x\sigma_j + \sigma^{-1}\mu_i\sigma_{xj}\\
\left.\frac{\partial^2 c_1}{\partial\xi_i\partial\xi_j}\right|_{x=x_0} &= -3\mu^2\sigma^{-4}\sigma_i\sigma_j + 2\mu\sigma^{-3}\sigma_x\sigma_i\sigma_j + \mu^2\sigma^{-3}\sigma_{ij} - \mu\sigma^{-2}\sigma_x\sigma_{ij}\\
&\quad - \mu\sigma^{-2}\sigma_{xi}\sigma_j - \mu\sigma^{-2}\sigma_{xj}\sigma_i + \mu\sigma^{-1}\sigma_{xij} + \frac{1}{4}\sigma_{xx}\sigma_{ij} - \frac{1}{4}\sigma_{xi}\sigma_{xj}\\
&\quad - \frac{1}{4}\sigma_x\sigma_{xij} + \frac{1}{4}\sigma_{xxi}\sigma_j + \frac{1}{4}\sigma_{xxj}\sigma_i + \frac{1}{4}\sigma\sigma_{xxij}.
\end{aligned}
$$

Let $\mathcal{A}$ denote the infinitesimal generator of the diffusion process (6.1), which is similar to (2.5). Define

$$
g_1(x,x_0) = \int_{x_0}^{x}\sigma_i\sigma^{-2}du\int_{x_0}^{x}\sigma_j\sigma^{-2}du
$$

and

$$
g_2(x,x_0) = \int_{x_0}^{x}\sigma^{-1}du\int_{x_0}^{x}\left[\sigma^{-2}\sigma_{ij} - 2\sigma^{-3}\sigma_i\sigma_j\right]du.
$$

Then,

$$
\begin{aligned}
\mathcal{A}g_1|_{x=x_0} &= (\sigma^{-2}\sigma_i\sigma_j)|_{x=x_0},\\
\mathcal{A}^2g_1|_{x=x_0} &= (2\mu^2\sigma^{-4}\sigma_i\sigma_j - 8\mu\sigma^{-3}\sigma_x\sigma_i\sigma_j + 4\sigma^{-2}\sigma_x^2\sigma_i\sigma_j + 2\sigma^{-2}\mu_x\sigma_i\sigma_j\\
&\quad + 2\mu\sigma^{-2}\sigma_{xi}\sigma_j + 2\mu\sigma^{-2}\sigma_{xj}\sigma_i - 2\sigma^{-1}\sigma_x\sigma_{xi}\sigma_j - 2\sigma^{-1}\sigma_x\sigma_{xj}\sigma_i\\
&\quad - 2\sigma^{-1}\sigma_{xx}\sigma_i\sigma_j + \frac{1}{2}\sigma_{xi}\sigma_{xj} + \frac{1}{2}\sigma_{xxi}\sigma_j + \frac{1}{2}\sigma_{xxj}\sigma_i)|_{x=x_0},\\
\mathcal{A}g_2|_{x=x_0} &= \sigma^{-1}\sigma_{ij} - 2\sigma^{-2}\sigma_i\sigma_j,\\
\mathcal{A}^2g_2|_{x=x_0} &= (-4\mu^2\sigma^{-4}\sigma_i\sigma_j + 20\mu\sigma^{-3}\sigma_x\sigma_i\sigma_j + 2\mu^2\sigma^{-3}\sigma_{ij} - 4\sigma^{-2}\mu_x\sigma_i\sigma_j\\
&\quad - 15\sigma^{-2}\sigma_x^2\sigma_i\sigma_j - 7\mu\sigma^{-2}\sigma_x\sigma_{ij} - 6\mu\sigma^{-2}\sigma_{xi}\sigma_j - 6\mu\sigma^{-2}\sigma_{xj}\sigma_i\\
&\quad + 2\sigma^{-1}\mu_x\sigma_{ij} + 6\sigma^{-1}\sigma_{xx}\sigma_i\sigma_j + 9\sigma^{-1}\sigma_x\sigma_{xi}\sigma_j + 9\sigma^{-1}\sigma_x\sigma_{xj}\sigma_i\\
&\quad + 3\sigma^{-1}\sigma_x^2\sigma_{ij} + 3\mu\sigma^{-1}\sigma_{xij} - 2\sigma_{xx}\sigma_{ij} - \frac{5}{2}\sigma_x\sigma_{xij}\\
&\quad - 4\sigma_{xi}\sigma_{xj} - 2\sigma_{xxi}\sigma_j - 2\sigma_{xxj}\sigma_i + \sigma\sigma_{xxij})|_{x=x_0}.
\end{aligned}
$$

Hence, from the above expressions,

$$
\mathbb{E}\left(\frac{\partial^2\log f^{(1)}}{\partial\eta_i\partial\eta_j}\right) = \delta\cdot\mathbb{E}\left(\frac{\partial^2 c_1}{\partial\eta_i\partial\eta_j}\right) + O(\delta^2) =: \delta\cdot N_{11}^{(1)} + O(\delta^2),
$$

$$
\mathbb{E}\left(\frac{\partial^2\log f^{(1)}}{\partial\eta_i\partial\xi_j}\right) = \delta\cdot\mathbb{E}\left(\frac{\partial^2 c_1}{\partial\eta_i\partial\xi_j}\right) + O(\delta^2) =: \delta\cdot N_{12}^{(1)} + O(\delta^2)
$$

and

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2 \log f^{(1)}}{\partial\xi_i\partial\xi_j}\right) =& -\mathbb{E}\{\sigma^{-1}\sigma_{ij}+\sigma^{-2}\sigma_i\sigma_j\} - \mathbb{E}[\mathcal{A}g_1|_{x=x_0}] + \mathbb{E}[\mathcal{A}g_2|_{x=x_0}] \\
& - \frac{\delta}{2}\cdot\mathbb{E}[\mathcal{A}^2 g_1|_{x=x_0}] + \frac{\delta}{2}\cdot\mathbb{E}[\mathcal{A}^2 g_2|_{x=x_0}] + \delta\cdot\mathbb{E}\left(\frac{\partial^2 c_1}{\partial\eta_i\partial\eta_j}\right) \\
& + O(\delta^2) \\
=:& -2\mathbb{E}(\sigma^{-2}\sigma_i\sigma_j) + \delta\cdot N_{22}^{(1)} + O(\delta^2),
\end{aligned}$$

where

$$\begin{aligned}
N_{11}^{(1)} =& \mathbb{E}\left(-\sigma^{-2}\mu_i\mu_j - \mu\sigma^{-2}\mu_{ij} + \sigma^{-1}\mu_{ij}\sigma_x - \frac{1}{2}\mu_{xij}\right), \\
N_{12}^{(1)} =& \mathbb{E}\left(2\mu\sigma^{-3}\mu_i\sigma_j - \sigma^{-2}\mu_i\sigma_x\sigma_j + \sigma^{-1}\mu_i\sigma_{xj}\right), \\
N_{22}^{(1)} =& \mathbb{E}\Bigg(-6\mu^2\sigma^{-4}\sigma_i\sigma_j + 16\mu\sigma^{-3}\sigma_x\sigma_i\sigma_j + 2\mu^2\sigma^{-3}\sigma_{ij} - 3\sigma^{-2}\mu_x\sigma_i\sigma_j - \frac{19}{2}\sigma^{-2}\sigma_x^2\sigma_i\sigma_j \\
& - \frac{9}{2}\mu\sigma^{-2}\sigma_x\sigma_{ij} - 5\mu\sigma^{-2}\sigma_{xi}\sigma_j - 5\mu\sigma^{-2}\sigma_{xj}\sigma_i + \sigma^{-1}\mu_x\sigma_{ij} + 4\sigma^{-1}\sigma_{xx}\sigma_i\sigma_j \\
& + \frac{11}{2}\sigma^{-1}\sigma_x\sigma_{xi}\sigma_j + \frac{11}{2}\sigma^{-1}\sigma_x\sigma_{xj}\sigma_i + \frac{3}{2}\sigma^{-1}\sigma_x^2\sigma_{ij} + \frac{5}{2}\mu\sigma^{-1}\sigma_{xij} - \frac{3}{4}\sigma_{xx}\sigma_{ij} \\
& - \frac{5}{2}\sigma_{xi}\sigma_{xj} - \frac{3}{2}\sigma_x\sigma_{xij} - \sigma_{xxi}\sigma_j - \sigma_{xxj}\sigma_i + \frac{3}{4}\sigma\sigma_{xxij}\Bigg).
\end{aligned}$$

Thus,

$$N(\theta_0, 1, \delta) = \begin{pmatrix} \delta\cdot N_{11}^{(1)} & \delta\cdot N_{12}^{(1)} \\ \delta\cdot N_{12}^{(1)T} & -2\cdot\mathbb{E}(\sigma^{-2}\sigma_i\sigma_j) + \delta\cdot N_{22}^{(1)} \end{pmatrix} + O(\delta^2). \tag{6.2}$$

We learn from Proposition 2 that $-N(\theta_0, 1, \delta)$ provides a leading order approximation to $I(\delta)$ with a reminder term at the order of δ^2. Equation (6.2) confirms that as $\delta \to 0$, given the asymptotic normality of the full MLE $\hat{\theta}_n$ as conveyed by (5.2), that the convergence rate of the full MLE for the drift parameters η is $(n\delta)^{1/2}$ whereas that for the diffusion parameters ξ is $n^{1/2}$, faster than the drift parameter estimator. This differential rate of convergence has been observed in various empirical studies (Ball and Torous, 1996), and was revealed theoretically by Tang and Chen (2009) for linear drift mean reverting diffusion processes. The result in (6.2) shows the differential convergence rates between the drift and the diffusion parameters for much general diffusion processes.

In the rest of the section, we will derive the approximate Fisher information matrix approximation for two specific diffusion processes. Both are widely employed in the modeling of the interest rate dynamics.

6.1 Vasicek's Model

Consider Vasicek's Model (Vaiscek, 1976),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{6.3}$$

which is also the Ornstein-Uhlenbeck process. The conditional distribution of X_t given X_{t-1} is

$$X_t|X_{t-1} \sim N\left\{X_{t-1}e^{-\kappa\delta} + \alpha(1-e^{-\kappa\delta}), \frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta})\right\}$$

and the stationary distribution of $\{X_t\}$ is

$$X_t \sim N\left(\alpha, \frac{\sigma^2}{2\kappa}\right). \tag{6.4}$$

The log of the transitional density is

$$\begin{aligned}&\log f(X_t|X_{t-1},\delta;\theta)\\ =& -\frac{1}{2}\log\pi - \frac{1}{2}\log\left(\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta})\right) - \frac{(X_t - X_{t-1}e^{-\kappa\delta} - \alpha(1-e^{-\kappa\delta}))^2}{\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta})}.\end{aligned}$$

Let $\theta = (\kappa,\alpha,\sigma)^T$ and $P(X_t, X_{t-1}, \theta) = X_t - X_{t-1}e^{-\kappa\delta} - \alpha(1-e^{-\kappa\delta})$, then

$$P(X_t, X_{t-1},\theta)|X_{t-1} \sim N\left\{0, \frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta})\right\}. \tag{6.5}$$

The second derivatives of $\log f(X_t|X_{t-1},\delta;\theta)$ are, respectively,

$$\begin{aligned}\frac{\partial^2 \log f}{\partial\kappa^2} =& -\frac{1}{2\kappa^2} + \frac{2\delta^2 e^{2\kappa\delta}}{(e^{2\kappa\delta}-1)^2} - \frac{2\kappa\delta^2(X_{t-1}-\alpha)^2}{\sigma^2(e^{2\kappa\delta}-1)}\\ &+ \frac{4\delta e^{2\kappa\delta}[(1-\kappa\delta)e^{2\kappa\delta} - (1+\kappa\delta)]P^2(X_t,X_{t-1},\theta)}{\sigma^2(e^{2\kappa\delta}-1)^3}\\ &+ P(X_t,X_{t-1},\theta)L_1(X_{t-1},\theta),\end{aligned}$$

$$\frac{\partial^2 \log f}{\partial\alpha^2} = -\frac{2\kappa(e^{\kappa\delta}-1)^2}{\sigma^2(e^{2\kappa\delta}-1)}, \quad \frac{\partial^2 \log f}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{6\kappa e^{2\kappa\delta}P^2}{\sigma^4(e^{2\kappa\delta}-1)},$$

$$\frac{\partial^2 \log f}{\partial\kappa\partial\alpha} = \frac{2\kappa\delta(X_{t-1}-\alpha)(e^{\kappa\delta}-1)}{\sigma^2(e^{2\kappa\delta}-1)} + P(X_t,X_{t-1},\theta)L_2(X_{t-1},\theta),$$

$$\frac{\partial^2 \log f}{\partial\kappa\partial\sigma} = \frac{2e^{2\kappa\delta}[e^{2\kappa\delta}-(1+2\kappa\delta)]P^2}{\sigma^3(e^{2\kappa\delta}-1)^2} + P(X_t,X_{t-1},\theta)L_3(X_{t-1},\theta)$$

$$\text{and} \quad \frac{\partial^2 \log f}{\partial\alpha\partial\sigma} = P(X_t,X_{t-1},\theta)L_4(X_{t-1},\theta),$$

where $L_i(X_{t-1},\theta)$, for $i = 1,\cdots,4$ are measurable functions of X_{t-1} for given θ.

From (6.4) and (6.5), it yields that the information matrix of $\theta = (\kappa,\alpha,\sigma)^T$ is $I(\delta) = (I_{ij})_{3\times3}$ where

$$I_{11} = \frac{1}{2\kappa^2} + \frac{\delta[\kappa\delta + \kappa\delta e^{2\kappa\delta} - 2e^{2\kappa\delta} + 2]}{\kappa(e^{2\kappa\delta}-1)^2} = \frac{\delta}{2\kappa} + O(\delta^2), \quad I_{12} = I_{21} = 0,$$

$$I_{13} = I_{31} = \frac{(1+2\kappa\delta) - e^{2\kappa\delta}}{\sigma\kappa(e^{2\kappa\delta}-1)} = -\frac{\delta}{\sigma} + O(\delta^2), \quad I_{22} = \frac{2\kappa(e^{\kappa\delta}-1)^2}{\sigma^2(e^{2\kappa\delta}-1)} = \frac{\kappa^2\delta}{\sigma^2} + O(\delta^2),$$

$$I_{23} = I_{32} = 0, \quad \text{and} \quad I_{33} = \frac{2}{\sigma^2}.$$

These mean that

$$I(\delta) = \begin{pmatrix} \delta \cdot (2\kappa)^{-1} & 0 & -\delta \cdot \sigma^{-1} \\ 0 & \delta \cdot \kappa^2 \sigma^{-2} & 0 \\ -\delta \cdot \sigma^{-1} & 0 & 2\sigma^{-2} \end{pmatrix} + O(\delta^2). \tag{6.6}$$

Hence, $I(0) = \lim_{\delta \to 0} I(\delta)$ is singular, an issue we have raised earlier and led us to assume $\delta I^{-1}(\delta)$'s largest eigen-value being bounded in Condition (A.5).

Using the approximation formula in (6.2), we have

$$N(\theta, 1, \delta) = \begin{pmatrix} -\delta \cdot (2\kappa)^{-1} & 0 & \delta \cdot \sigma^{-1} \\ 0 & -\delta \cdot \kappa^2 \sigma^{-2} & 0 \\ \delta \cdot \sigma^{-1} & 0 & -2\sigma^{-2} \end{pmatrix} + O(\delta^2).$$

It means the leading order term of $-N(\theta, 1, \delta)$ is identical with that of the true Fisher information matrix in (6.6).

6.2 Cox-Ingersoll-Ross Model

Consider Cox-Ingersoll-Ross (CIR) Model (Cox, Ingersoll and Ross, 1985)

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t. \tag{6.7}$$

which is also Feller (1952)'s square root processes.

Let $\theta = (\kappa, \alpha, \sigma)^T$ and $c = 4\kappa\sigma^{-2}(1 - e^{-\kappa\delta})^{-1}$, the conditional distribution of cX_t given X_{t-1} is

$$cX_t | X_{t-1} \sim \chi^2_\nu(\lambda),$$

where the distribution is a non-central χ^2 distribution with degree of freedom $\nu = 4\kappa\alpha\sigma^{-2}$ and non-central parameter $\lambda = cX_{t-1}e^{-\kappa\delta}$. The transitional density of $X_{t+\delta}$ given X_t is

$$f(X_t | X_{t-1}, \delta; \theta) = \frac{c}{2} e^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2\sqrt{uv}),$$

where $u = cX_{t-1}e^{-\kappa\delta}/2$, $v = cX_t/2$, $q = 2\kappa\alpha/\sigma^2 - 1 \geqslant 0$ and I_q is the modified Bessel function of the first kind of order q. If $2\kappa\alpha > \sigma^2$, then the stationary distribution of $\{X_t\}$ is

$$X_t \sim \Gamma\left(\frac{2\kappa\alpha}{\sigma^2}, \frac{\sigma^2}{2\kappa}\right). \tag{6.8}$$

The log transitional density function is

$$\log f(X_t | X_{t-1}, \delta; \theta) = \log c - (u + v) + \frac{q}{2}(\log v - \log u) + \log I_q(2\sqrt{uv}) - \log 2.$$

Although the second partial derivations of the log transitional density function can be derived after some labor that involved with differentiating the modified Bessel function of first kind, acquiring an expression for the Fisher information matrix is a rather hard task, largely due to the difficulty in deriving the expectations. In contrast, using the approximation formula (6.2), we can obtain the approximation for opposite Fisher information matrix

$$N(\theta_0, 1, \delta) = \begin{pmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ N_{31} & N_{32} & N_{33} \end{pmatrix} + O(\delta^2),$$

where

$$N_{11} = \delta \cdot \sigma^{-2} \cdot \mathbb{E}\{X_t^{-1}(\alpha - X_t)^2\}, \qquad N_{12} = N_{21} = \delta \cdot \mathbb{E}\left\{2\kappa\sigma^{-2}X_t^{-1}(\alpha - X_t) - \frac{1}{2}X_t^{-1}\right\},$$

$$N_{13} = N_{31} = -\delta \cdot 2\kappa\sigma^{-3} \cdot \mathbb{E}\{X_t^{-1}(\alpha - X_t)^2\}, \qquad N_{22} = \delta \cdot \kappa^2\sigma^{-2} \cdot \mathbb{E}X_t^{-1},$$

$$N_{23} = N_{32} = -\delta \cdot 2\kappa^2\sigma^{-3} \cdot \mathbb{E}\{X_t^{-1}(\alpha - X_t)\} \qquad \text{and}$$

$$N_{33} = 2\sigma^{-2} - \delta \cdot 3\kappa\sigma^{-2} + \delta \cdot \mathbb{E}\left\{6\kappa^2\sigma^{-4}X_t^{-1}(\alpha - X_t)^2 - 6\kappa\sigma^{-2}X_t^{-1}(\alpha - X_t) + \frac{9}{4}X_t^{-1} + \sigma^{-1}X_t^{-1}\right\}.$$

More explicit form of the approximation may be obtained by cultivating the marginal distribution of X_t. Under (6.8), we can get

$$\mathbb{E}X_t^{-1} = \frac{\sigma^2}{2\kappa\alpha - \sigma^2} \quad \text{and} \quad \mathbb{E}X_t = \alpha.$$

Then,

$$N_{11} = \delta \cdot \frac{\alpha^2\sigma^2 - 2\kappa\alpha^2 + \alpha\sigma^2}{2\kappa\alpha\sigma^2 - \sigma^4}, \quad N_{12} = N_{21} = \delta \cdot \frac{4\kappa\alpha\sigma^2 - \sigma^4 - 8\kappa^2\alpha + 4\kappa\sigma^2}{4\kappa\alpha\sigma^2 - 2\sigma^4},$$

$$N_{13} = N_{31} = -\delta \cdot \frac{2\kappa\alpha^2\sigma^2 - 4\kappa^2\alpha^2 + 2\kappa\alpha\sigma^2}{2\kappa\alpha\sigma^3 - \sigma^5}, \quad N_{22} = \delta \cdot \frac{\kappa^2}{2\kappa\alpha - \sigma^2},$$

$$N_{23} = -\delta \cdot \frac{2\kappa^2\alpha\sigma^2 - 4\kappa^3\alpha + 2\kappa^2\sigma^2}{2\kappa\alpha\sigma^3 - \sigma^5}, \quad \text{and}$$

$$N_{33} = \frac{2}{\sigma^2} + \delta \cdot \frac{24\kappa^2\alpha^2\sigma^2 - 48\kappa^3\alpha^2 + 48\kappa^2\alpha\sigma^2 - 24\kappa\alpha\sigma^4 + 36\kappa\sigma^4 + 4\sigma^5 + 9\sigma^6}{8\kappa\alpha\sigma^4 - 4\sigma^6}.$$

Using $-N(\theta_0, 1, \delta)$, we can get the approximation of Fisher information matrix. This approximation may be used in carrying out statistical inference on the CIR processes.

7 Simulation

We report results from simulation studies which are designed to confirm the theoretical findings on the AMLE as reported in the earlier sections. To allow verification with the full MLE, we considered the Vasicek and CIR diffusion models reported in the previous section as both models permit the full MLE. The two regimes of the asymptotic were experimented: the fixed δ and the diminishing δ with $n\delta^3 \to \infty$.

The first part of the simulation is about the case which δ is fixed. The parameters used in the simulated Vasicek and CIR models were $\theta = (\kappa, \alpha, \sigma)^T = (0.858, 0.0891, 0.0468)$ and

$\theta = (\kappa, \alpha, \sigma)^T = (0.892, 0.09, 0.1817)$, respectively. The sampling interval δ was 1/12 and 1/4, and the order of the density approximation J was 1 and 2, respectively. For each δ and J, the sample size n was set at 500, 1000 and 2000 respectively. In addition to bias and standard deviation, we consider

$$\text{RMSD}(J) = \sqrt{\mathbb{E}||\hat{\theta}_n^{(J)} - \hat{\theta}_n||_2^2},$$

the squared root of the expected square of modulated deviations between $\hat{\theta}_n^{(J)}$ and $\hat{\theta}_n$, as an overall performance measure.

Table 2 and 3 summarize the simulation for the fixed δ case. They report the average bias and standard deviation (SD) for the full MLE and the AMLEs with $J = 1$ and $J = 2$, as well as the RMSD between the AMLEs and the full MLE, for both the Vasicek and the CIR models. To give the simulation results more perspectives and to confirm the derived approximated bias and variance formulae in Section 6, we also computed the asymptotic bias and standard deviation based on the formulae (5.9) and (5.10). We observe from Tables 2 and 3 that at each δ (1/12 and 1/4) experimented, the bias and the standard deviation of all three estimators for all three parameters became smaller as n was increased. These confirmed the consistency of the estimators. The tables also showed that there was a good agreement among the three estimators in terms of the performance measures. It appeared that the bias and the variance of the AMLE with $J = 1$ and $J = 2$ were quite comparible to each other. However, by comparing the rows of RMSD, it was clear that in most of the cases (except for $n = 500$ of CIR model), the RMSD for $J = 2$ was smaller than $J = 1$, signaling the AMLE with $J = 2$ was closer to the full MLE than that of the AMLE with $J = 1$. This indicates that the AMLEs with $J = 2$ were indeed closer to those with $J = 1$, as confirmed by our early analysis. The asymptotic bias and standard deviation predicted for the AMLE with $J = 1$ and 2 offer more insights, and showed good agreement between the simulated results and the predicted values by the theory, which is very assuring. We also observed that for $\delta = 1/4$, the AMLE with $J = 2$ performs better than AMLE with $J = 1$, which somehow reflects Table 1 which shows that $J = 2$ is preferred than $J = 1$ at this frequency. When δ was fixed at 1/12, we see the performance between $J = 1$ and $J = 2$ was largely similar.

The second part of the simulation was devoted to diminishing δ case. Here we wanted to confirm the differential behavior of the AMLEs in the limiting distribution between $J = 1$ and $J \geqslant 2$, as revealed in Section 5. The Vasicek model with $\theta = (\kappa, \alpha, \sigma)^T = (0.892, 0.09, 0.1817)$ was considered. We tried to create two scenarios: (i) $n\delta^3 \to \infty$ and (ii) $n\delta^3 \to 0$, while $\delta \to 0$. They were created by choosing $\delta = n^{-1/6}$ and $\delta = n^{-1/2}$ respectively, whiling selecting $n = 500, 1000, 2000, 4000$ and 8000 respectively, to create two streams of asymptotic sequences. For each n and δ, we generated repeatedly the Vasicek sample paths 1000 times. For each simulated sample path, we obtained the AMLEs $\hat{\theta}_n^{(J)}$ for $J = 1$ and 2 respectively, and compute the Wald statistics

$$W_n(J) = n(\hat{\theta}_n^{(J)} - \theta_0)^T I(\delta)(\hat{\theta}_n^{(J)} - \theta_0).$$

If $\sqrt{n}I^{1/2}(\delta)(\hat{\theta}_n^{(J)} - \theta_0)$ is asymptotically standard normally distributed in $\mathbb{R}^d$, then the Wald statistic $W_n(J) \xrightarrow{d} \chi_3^2$. Based on the 1000 Wald statistics from the simulations, we then performed the Kolmogorov-Smirnov (K-S) test to test $H_0 : W_n(J) \sim \chi_3^2$ or not for each of the designed sequeneces of (n, δ) generated under the two scenarios. Table 4 reports the p-values of the test, which show that for $J = 1$, under both scenarios, the p-values of the K-S test became smaller and hence the above null hypothesis was rejected as n was increased. For $J = 2$, the p-values of the K-S test were sharply difference between the two scenarios. In particular, the p-values was mostly quite large under the scenario of $n\delta^3 \to \infty$, and they were largely significant (small) when δ was diminishing at the faster rate of $n^{-1/2}$ such that $n\delta^3 \to 0$. These were consistent with our theoretical findings in Section 5.

Appendix

We need the following technical assumptions in our analysis.

(A.1) (i) Θ is a compact set in $\mathbb{R}^d$, and the true parameter θ_0 is an interior point of Θ; (ii) for all values of the parameters θ, Assumption 1-3 in Aït-Sahalia (2002) hold; (iii) the drift function $\mu(x;\theta)$ is a bona fide function of θ for each x.

(A.2) (i) For every $\delta > 0$,

$$\mathbb{E}\left\{\frac{\partial \log f(X_t|X_{t-1},\delta;\theta_0)}{\partial\theta}\right\} = 0,$$

and θ_0 is the only root of $\mathbb{E}\left\{\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\theta)\right\} = 0$. (ii) the MLE $\hat{\theta}_n$ and the J-term approximate MLE $\hat{\theta}_n^{(J)}$ satisfy, respectively,

$$\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\hat{\theta}_n) = 0 \quad \text{and}$$

$$\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f^{(J)}(X_t|X_{t-1},\delta;\hat{\theta}_n^{(J)}) = 0.$$

And (iii) $\hat{\theta}_n$ is consistent to θ_0 and asymptotically normal such that (5.2) is satisfied.

(A.3) There exist finite positive constants Δ and K_1 such that, for $l = 1, 2, 3$, any $\delta \in (0, \Delta]$, $i_1, i_2, i_3 \in \{1, \cdots, d\}$ and $j = 1$ and 2,

$$\mathbb{E}\sup_{\theta\in\Theta}\left\{\left|\frac{\partial^l A_j(X_t|X_{t-1},\delta;\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_l}}\right|^2\right\} \leqslant K_1.$$

(A.4) There exist finite positive constants ν_l for $q = 0, 1, 2$ and 3, $\Delta > 0$ and K_2 such that $\nu_0 > 3$, $\nu_2 > \nu_1 > 3$, $\nu_3 > 1$ and for any $i_1, \cdots, i_3 \in \{1, \cdots, d\}$ and $\delta \in (0, \Delta]$,

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left[\sum_{l=0}^{\infty}\left|\frac{\partial^q c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta);\theta\right)}{\partial\theta_{i_1}\cdots\partial\theta_{i_q}}\right|\frac{\Delta^l}{l!}\right]^{\nu_l}\right\} \leqslant K_2.$$

(A.5) For any $\delta > 0$, the Fisher information matrix

$$I(\delta) := \mathbb{E}\left\{\frac{\partial^2 \log f(X_t|X_{t-1},\delta;\theta_0)}{\partial\theta\partial\theta^T}\right\}$$

is invertible and as $\delta \to 0$ the largest eigen-values of $\delta I^{-1}(\delta)$ is bounded away from infinity.

(A.6) For each positive integer K, which may be infinite, and any $\delta \in (0, \Delta]$,

$$\mathbb{P}\left\{\inf_{\theta\in\Theta}\left|\sum_{l=0}^{K} c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta);\theta\right)\frac{\delta^l}{l!}\right| = 0\right\} = 0,$$

(A.7) For any $\beta > 1$ and $\eta > 0$, there exist $\Delta(\beta,\eta) > 0$, then for any $\delta \in (0, \Delta(\beta,\eta)]$ and K, where K may be infinite,

$$\mathbb{P}\left\{\inf_{\theta\in\Theta}\left|\sum_{l=0}^{K} c_l(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta);\theta)\frac{\delta^l}{l!}\right| < \eta^{1/\beta}\right\} < \eta.$$

(A.1) and (A.2) are standard requirements for maximum likelihood estimators. (A.1) (ii) contains conditions on the smoothness of the drift and the diffusion which ensures the existence of a unique solution to (2.1) as well as the infinite differentiability of the transitional density $f(x|x_0,\delta;\theta)$ with respect to x, x_0 and δ, and three time differentiable with respect to θ (Friedman, 1964). The second part of (A.2) is the simplified approach of Cramër (1946) assuming the MLEs are the solutions of the likelihood score equations. (A.3) is needed to guarantee the third derivative of $\log f(X_t|X_{t-1},\delta;\theta)$ with respect to θ can be controlled by an integrable function, whileas Condition (A.4) ensures the absolutely convergence of the infinite series $\sum_{l=0}^{\infty}|c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)|\delta^l/l! = \exp\{\widetilde{A}_3(x|x_0,\delta;\theta)\}$ as Aït-Sahalia (2002) has provided conditions on the non-degeneracy of the diffusion function and the boundary condition, which together with the late part of Condition (A.1) leads to the convergence of the above infinite series $\exp\{\widetilde{A}_3(x|x_0,\delta;\theta)\}$. (A.4) is also needed to allow exchange of differentiation and summation for the infinite series. The first part of the (A.5) is of standard in likelihood inference. Its second part reflects the fact that for some processes $\lim_{\delta\to 0} I(\delta)$ may be singular, as conveyed in our discussion in Section 6 for the Vasicek process. Condition (A.6) is needed to guarantee the derivatives of log transitional density and log approximated transitional density exist with probability one. Condition (A.7) is needed to manage the denominators in the derivatives of the log of the approximated transitional density, ensuring the probability of their taking small values can be controlled uniformly.

We shall give the proofs for the propositions and theorems mentioned in Sections 3-4. We first present some lemmas about the true transitional density and its approximations, which we will use in later proofs.

Lemma 1 *Under (A.1) and (A.4), for any $\delta \in (0, \Delta)$, the infinite series*

$$\sum_{l=0}^{\infty} c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}$$

absolutely converges with probability 1, and for $k = 1, 2$ and 3, and $i_1, i_2, i_3 \in \{1, \cdots, d\}$,

$$\frac{\partial^k}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}}\sum_{l=0}^{\infty} c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!} = \sum_{l=0}^{\infty}\frac{\partial^k}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}.$$

PROOF: Firstly, we consider the absolutely convergence of the infinite series. Let $S_n(\delta) = \sum_{l=0}^{n} c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\delta^l/l!$.

For a fixed $\delta \in (0,\Delta)$ and $\theta \in \Theta$,

$$\mathbb{P}\left\{\max_{M\leqslant m\leqslant N}|S_m(\delta)-S_M(\delta)|>\epsilon\right\}\leqslant \mathbb{P}\left\{\sum_{l=M+1}^{N}|c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)|\frac{\delta^l}{l!}>\epsilon\right\}.$$

Applying Markov inequality,

$$\mathbb{P}\left\{\max_{M\leqslant m\leqslant N}|S_m(\delta)-S_M(\delta)|>\epsilon\right\}\leqslant \epsilon^{-2}\cdot\mathbb{E}\left\{\sum_{l=M+1}^{N}|c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)|\frac{\delta^l}{l!}\right\}^2.$$

Letting $N\to\infty$, we get from (A.4),

$$\mathbb{P}\left\{\sup_{m\geqslant M}|S_m-S_M|>\epsilon\right\}\leqslant \epsilon^{-2}\cdot\mathbb{E}\left\{\sum_{l=M+1}^{\infty}|c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)|\frac{\delta^l}{l!}\right\}^2\to 0 \quad \text{as} \quad M\to\infty.$$

If we let $\omega_M=\sup_{m,n\geqslant M}|S_m-S_n|$, then $\omega_M\downarrow$ as $M\uparrow$ and

$$\mathbb{P}(\omega_M>2\epsilon)\leqslant \mathbb{P}\left\{\sup_{m\geqslant M}|S_m-S_M|>\epsilon\right\}\to 0$$

as $M\to\infty$. Hence, $\omega_M\downarrow 0$ almost surely. Then, we attain the absolutely convergence of the infinite series. Actually, this absolute convergence is uniform on Θ.

Next, we consider the exchange between the differentiation and the summation. The key is to prove that

$$\sum_{l=0}^{\infty}\frac{\partial}{\partial\theta_i}c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}$$

is uniformly convergent on Θ with probability 1. Using the same method above and from (A.4), the result is correct. Then,

$$\frac{\partial}{\partial\theta_i}\sum_{l=0}^{\infty}c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}=\sum_{l=0}^{\infty}\frac{\partial}{\partial\theta_i}c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}$$

for any $i\in\{1,\cdots,d\}$ with probability 1. Using the same approach, we can show the exchange between differentiation and the summation is also valid for $k=2$ and 3, respectively. □

Lemma 2 *Under (A.6) and (A.7), for any positive $\beta>1$, there exists two constants $m(\beta)<\infty$ and $\Delta_1(\beta)>0$ such that for any $\delta\in(0,\Delta_1(\beta)]$ and J, where J can be infinity, then*

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J}c_l\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta)\right)\frac{\delta^l}{l!}\right|^{-\beta}\right\}<m(\beta).$$

PROOF: Let

$$K(J,\delta)=\inf_{\theta\in\Theta}\left|\sum_{j=0}^{J}c_j\left(\gamma(X_t;\theta)|\gamma(X_{t-1};\theta);\theta\right)\frac{\delta^j}{j!}\right|-1.$$

Then

$$-1\leqslant K(J,\delta)\leqslant\left|\sum_{j=1}^{J}c_j\left(\gamma(X_t;\theta_0)|\gamma(X_{t-1};\theta_0);\theta_0\right)\frac{\delta^j}{j!}\right|.$$

Note that (A.6), it implies that $\mathbb{P}(K(J,\delta)=-1)=0$ for any $\delta\in(0,\Delta]$. Define $\tilde{K}(J,\delta)$ such that $1+\tilde{K}(J,\delta)=(1+K(J,\delta))^{\beta}$, then $\mathbb{P}(\tilde{K}(J,\delta)=-1)=0$ for any $\delta\in(0,\Delta]$. For any $\varepsilon\in(0,1)$,

$$\begin{aligned}
\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J}c_l\frac{\delta^l}{l!}\right|^{-\beta}\right\} &= \mathbb{E}\left[\frac{1}{1+\tilde{K}(J,\delta)}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]+\mathbb{E}\left[\frac{1}{1+\tilde{K}(J,\delta)}1_{\{\tilde{K}(J,\delta)\geqslant-1+\varepsilon\}}\right]\\
&\leqslant \mathbb{E}\left[\frac{1}{1+\tilde{K}(J,\delta)}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]+\frac{1}{\varepsilon}\\
&\leqslant \mathbb{E}\left\{\sum_{i=0}^{\infty}|\tilde{K}(J,\delta)|^i1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right\}+\frac{1}{\varepsilon}\\
&\leqslant \sum_{i=0}^{\infty}\mathbb{E}\left[|\tilde{K}(J,\delta)|^i1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]+\frac{1}{\varepsilon},
\end{aligned}$$

where $1_{\{\}}$ is the indicator function. The last inequality is based on Fatou's lemma. By Hölder inequality, for $i\geqslant 1$,

$$\begin{aligned}
&\mathbb{E}\left[|\tilde{K}(J,\delta)|^{i+1}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\\
&=\mathbb{E}\left\{|\tilde{K}(J,\delta)|^{i-1}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\cdot|\tilde{K}(J,\delta)|^{2}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right\}\\
&\leqslant \left\{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\right\}^{(i-1)/i}\left\{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{2i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\right\}^{1/i} \qquad (7.9)\\
&=\left\{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\right\}\left\{\frac{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{2i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]}{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]}\right\}^{1/i}.
\end{aligned}$$

Denote the second factor on the right hand side by $T_i(J,\delta)$. We claim that each element of the sequence $\{T_i(J,\delta)\}_{i\geqslant 1}$ can be controlled by a constant which is strictly less than 1. To appreciate this, let $\alpha=(1-\varepsilon)^{-2}$, then for any $i\geqslant 1$,

$$\mathbb{E}\left[|\tilde{K}(J,\delta)|^{2i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\leqslant 1=\alpha(1-\varepsilon)^2\leqslant\alpha^i\cdot\left\{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{2}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\right\}^i.$$

On the other hand, applying Jensen inequality, for any $i\geqslant 1$,

$$\left\{\mathbb{E}\left[|\tilde{K}(J,\delta)|^{i}1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]\right\}^{1/i}\geqslant\mathbb{E}\left[|\tilde{K}(J,\delta)|1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right].$$

Hence, for any $i \geqslant 1$,

$$T_i(J,\delta) \leqslant \alpha \cdot \frac{\mathbb{E}\left[|\tilde{K}(J,\delta)|^2 1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]}{\mathbb{E}\left[|\tilde{K}(J,\delta)| 1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right]}.$$

Choosing $\eta \in (0,1)$, by (A.7), we know that for any J and $\delta \in (0, \Delta(\eta,\beta)]$,

$$\mathbb{P}\left\{-1 < \tilde{K}(J,\delta) < -1+\eta\right\} < \eta.$$

Hence, for any J, $\delta \in (0, \Delta \wedge \Delta(\eta,\beta)]$ and $i \geqslant 1$,

$$T_i(J,\delta) \leqslant \alpha \cdot \frac{\eta + (1-\eta)^2}{1-\varepsilon} = \alpha^{3/2} \cdot \left[\eta + (1-\eta)^2\right].$$

If the right hand of the above inequality can be controlled by a constant which is strictly less than 1, we prove our claim. In the following, we will prove that we can find (η,α) such that $\eta \in (0,1)$, $\alpha > 1$ and $\alpha^{3/2} \cdot [\eta + (1-\eta)^2] < 1$.

For a fixed $\alpha > 1$, we consider the solution for

$$\eta < \alpha^{-3/2}/2 \quad \text{and} \quad (1-\eta)^2 < \alpha^{-3/2}/2.$$

These equations are equivalent to

$$1 - \frac{1}{\sqrt{2}\alpha^{3/4}} < \eta < \frac{1}{2\alpha^{3/2}}.$$

As $1 - 1/\sqrt{2} < 1/2$, we can pick a $\alpha > 1$ but sufficiently near 1 to guarantee that

$$1 - \frac{1}{\sqrt{2}\alpha^{3/4}} < \frac{1}{2\alpha^{3/2}}.$$

Hence, we can pick $\alpha > 1$ and $\eta > 0$ such that, for any J and $\delta \in (0, \Delta \wedge \Delta(\eta,\beta)]$,

$$T_i(J,\delta) \leqslant \alpha^{3/2} \cdot \left[\eta + (1-\eta)^2\right] < 1.$$

At the same time, we know

$$\mathbb{E}\left[|\tilde{K}(J,\delta)| 1_{\{-1<\tilde{K}(J,\delta)<-1+\varepsilon\}}\right] \leqslant 1.$$

Then, for any J and $\delta \in (0, \Delta \wedge \Delta(\eta,\beta)]$,

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l \frac{\delta^l}{l!}\right|^{-\beta}\right\} \leqslant \frac{1}{1-\alpha^{3/2}\cdot[\eta+(1-\eta)^2]} + \frac{1}{\varepsilon} =: m(\beta) < \infty.$$

Hence, we complete the proof of Lemma 2. □

Lemma 3 *Under (A.1), (A.3)-(A.4), (A.6)-(A.7), there exist two constants $M_1 < \infty$ and $\Delta_2 > 0$ such that, for any J, where J can be infinity, $\delta \in (0, \Delta_2)$ and $i, j, k \in \{1, \cdots, d\}$,*

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial^3 \log f^{(J)}(X_t|X_{t-1}, \delta; \theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right\} < M_1.$$

PROOF: From the definition of A_3, if $J = \infty$, then $A_3 = \widetilde{A}_3$. Note that

$$\begin{aligned}\left|\frac{\partial^3 A_3}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leqslant\ & \left|\sum_{l=0}^{J} c_l \frac{\delta^l}{l!}\right|^{-1} \sum_{l=0}^{J}\left|\frac{\partial^3 c_l}{\partial\theta_i\partial\theta_j\partial\theta_k}\frac{\delta^l}{l!}\right| \\ & + \left|\sum_{l=0}^{J} c_l \frac{\delta^l}{l!}\right|^{-2}\left\{\sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\frac{\delta^l}{l!}\right| \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_k}\frac{\delta^l}{l!}\right|\right. \\ & \left. + \sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_k}\frac{\delta^l}{l!}\right| \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_j}\frac{\delta^l}{l!}\right| + \sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_j\partial\theta_k}\frac{\delta^l}{l!}\right| \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\frac{\delta^l}{l!}\right|\right\} \\ & + 2\cdot\left|\sum_{l=0}^{J} c_l \frac{\delta^l}{l!}\right|^{-3} \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\frac{\delta^l}{l!}\right| \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_j}\frac{\delta^l}{l!}\right| \cdot \sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_k}\frac{\delta^l}{l!}\right|.\end{aligned}$$

Then, applying Hölder inequality,

$$\begin{aligned} & \mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial^3 A_3}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right\} \\ \leqslant\ & \left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{\nu_3}{\nu_3-1}}\right\}^{\frac{\nu_3-1}{\nu_3}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial^3 c_l}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\frac{\delta^l}{l!}\right]^{\nu_3}\right\}^{1/\nu_3} \\ & + \left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_1}{\nu_1-2}}\right\}^{\frac{\nu_1-2}{\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_k}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1} \\ & + \left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_1}{\nu_1-2}}\right\}^{\frac{\nu_1-2}{\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_k}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1} \\ & + \left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_1}{\nu_1-2}}\right\}^{\frac{\nu_1-2}{\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_j\partial\theta_k}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1} \\ & + 2\cdot\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{3\nu_1}{\nu_1-3}}\right\}^{(\nu_1-3)/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1} \\ & \qquad \cdot\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{J}\left|\frac{\partial c_l}{\partial\theta_k}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}.\end{aligned}$$

Choose $\Delta_2 = \Delta \wedge \Delta_1(\nu_3/(\nu_3-1)) \wedge \Delta_1(2\nu_1/(\nu_1-2)) \wedge \Delta_1(3\nu_1/(\nu_1-3))$. Note Lemma 2 and (A.4), then for any J and $\delta \in (0, \Delta_2]$,

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial^3 A_3}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right\} < C$$

where C is a finite constant which is not dependent on J and δ. On the other hand, with (A.3), we can say there exists a constant $M_1 < \infty$ such that for any J and $\delta \in (0, \Delta_2]$,

$$\mathbb{E}\left\{\sup_{\theta\in\Theta}\left|\frac{\partial^3 \log f^{(J)}(X_t|X_{t-1},\delta;\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right\} < M_1.$$

Hence, we complete the proof of Lemma 3. □

Proof of Proposition 1: Using the same method in the proof of Lemma 3, we know (a) is hold. On the other hand, Lemma 3 implies (b). □

Proof of Proposition 2: By the definition of $\widetilde{A}_3(X_t|X_{t-1},\delta;\theta)$ and $A_3(X_t|X_{t-1},\delta;\theta)$, the $(i,j)-$th element of the matrix $\partial^2[\widetilde{A}_3(X_t|X_{t-1},\delta;\theta) - A_3(X_t|X_{t-1},\delta;\theta)]/\partial\theta\partial\theta^T$ is

$$\begin{aligned}
&\left(\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right)^{-1}\sum_{l=J+1}^{\infty}\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\frac{\delta^l}{l!} - \left(\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right)^{-1}\left(\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right)^{-1}\sum_{l=0}^{J}\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\frac{\delta^l}{l!}\sum_{l=J+1}^{\infty} c_l\frac{\delta^l}{l!}\\
&-\left(\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right)^{-2}\sum_{l=J+1}^{\infty}\frac{\partial c_l}{\partial\theta_i}\frac{\delta^l}{l!}\sum_{l=0}^{\infty}\frac{\partial c_l}{\partial\theta_j}\frac{\delta^l}{l!} + \left(\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right)^{-2}\left(\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right)^{-1}\sum_{l=J+1}^{\infty} c_l\frac{\delta^l}{l!}\sum_{l=0}^{J}\frac{\partial c_l}{\partial\theta_i}\frac{\delta^l}{l!}\sum_{l=0}^{J}\frac{\partial c_l}{\partial\theta_j}\frac{\delta^l}{l!}\\
&+\left(\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right)^{-1}\left(\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right)^{-2}\sum_{l=J+1}^{\infty} c_l\frac{\delta^l}{l!}\sum_{l=0}^{J}\frac{\partial c_l}{\partial\theta_i}\frac{\delta^l}{l!}\sum_{l=0}^{\infty}\frac{\partial c_l}{\partial\theta_j}\frac{\delta^l}{l!}.
\end{aligned}$$

Then, for any $i, j \in \{1, \cdots, d\}$, applying Hölder inequality,

$$\begin{aligned}
&|I_{ij}(\delta) + N_{ij}(\theta_0, J, \delta)|\\
\leqslant\ &\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right|^{-\frac{\nu_2}{\nu_2-1}}\right\}^{\frac{\nu_2-1}{\nu_2}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_2}\right\}^{1/\nu_2}\\
&+\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{2(\nu_0\wedge\nu_2)}{\nu_0\wedge\nu_2-2}}\right\}^{\frac{\nu_0\wedge\nu_2-2}{2(\nu_0\wedge\nu_2)}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right|^{-\frac{2(\nu_0\wedge\nu_2)}{\nu_0\wedge\nu_2-2}}\right\}^{\frac{\nu_0\wedge\nu_2-2}{2(\nu_0\wedge\nu_2)}}\\
&\qquad\cdot\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{J}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_0\wedge\nu_2}\right\}^{\frac{1}{\nu_0\wedge\nu_2}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}|c_l|\frac{\delta^l}{l!}\right]^{\nu_0\wedge\nu_2}\right\}^{\frac{1}{\nu_0\wedge\nu_2}}\\
&+\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_1}{\nu_1-2}}\right\}^{\frac{\nu_1-2}{\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{\infty}\left|\frac{\partial c_l}{\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\\
&+\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J} c_l\frac{\delta^l}{l!}\right|^{-\frac{4\nu_0\nu_1}{(\nu_1-2)\nu_0-\nu_1}}\right\}^{\frac{(\nu_1-2)\nu_0-\nu_1}{2\nu_0\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{\infty} c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_0\nu_1}{(\nu_1-2)\nu_0-\nu_1}}\right\}^{\frac{(\nu_1-2)\nu_0-\nu_1}{2\nu_0\nu_1}}\\
&\qquad\cdot\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}|c_l|\frac{\delta^l}{l!}\right]^{\nu_0}\right\}^{1/\nu_0}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{J}\left|\frac{\partial c_l}{\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}
\end{aligned}$$

$$+\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{J}c_l\frac{\delta^l}{l!}\right|^{-\frac{2\nu_0\nu_1}{(\nu_1-2)\nu_0-\nu_1}}\right\}^{\frac{(\nu_1-2)\nu_0-\nu_1}{2\nu_0\nu_1}}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left|\sum_{l=0}^{\infty}c_l\frac{\delta^l}{l!}\right|^{-\frac{4\nu_0\nu_1}{(\nu_1-2)\nu_0-\nu_1}}\right\}^{\frac{(\nu_1-2)\nu_0-\nu_1}{2\nu_0\nu_1}}$$

$$\cdot\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}|c_l|\frac{\delta^l}{l!}\right]^{\nu_0}\right\}^{1/\nu_0}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{J}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}\left\{\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=1}^{\infty}\left|\frac{\partial c_l}{\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\nu_1}\right\}^{1/\nu_1}.$$

On the other hand, for any $\alpha>0$ and $\delta\in(0,\Delta]$, we have the following inequalities

$$\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}|c_l|\frac{\delta^l}{l!}\right]^{\alpha}\leqslant\left(\frac{\delta}{\Delta}\right)^{\alpha(J+1)}\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{\infty}|c_l|\frac{\Delta^l}{l!}\right]^{\alpha},$$

$$\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\delta^l}{l!}\right]^{\alpha}\leqslant\left(\frac{\delta}{\Delta}\right)^{\alpha}\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{\infty}\left|\frac{\partial c_l}{\partial\theta_i}\right|\frac{\Delta^l}{l!}\right]^{\alpha}$$

and

$$\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=J+1}^{\infty}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\right|\frac{\delta^l}{l!}\right]^{\alpha}\leqslant\left(\frac{\delta}{\Delta}\right)^{\alpha(J+1)}\mathbb{E}\sup_{\theta\in\Theta}\left[\sum_{l=0}^{\infty}\left|\frac{\partial^2 c_l}{\partial\theta_i\partial\theta_j}\right|\frac{\Delta^l}{l!}\right]^{\alpha}.$$

Then, noting (A.4) and Lemma 2, we know there exists a constant $\bar{\Delta}>0$ such that for any J and $\delta\in(0,\bar{\Delta}]$,

$$|N_{ij}(\theta_0,J,\delta)+I_{ij}(\delta)|\leqslant C\delta^{J+1},$$

where C is not dependent on J and δ. Hence, we complete the proof of Proposition 2. □

Proof of Proposition 3: Recall Proposition 2, then

$$||I^{-1}(\delta)N(\theta_0,J,\delta)+E||_2\leqslant||I^{-1}(\delta)||_2\cdot||N(\theta_0,J,\delta)+I(\delta)||_2\leqslant C\delta^J.$$

If $C\delta^J<1$, then

$$||N^{-1}(\theta_0,J,\delta)I(\delta)+E||_2\leqslant\frac{||I^{-1}(\delta)N(\theta_0,J,\delta)+E||_2}{1-||I^{-1}(\delta)N(\theta_0,J,\delta)+E||_2}.$$

From Proposition 2, if $C\delta^{J+1}<1$, then

$$||N^{-1}(\theta_0,J,\delta)+I^{-1}(\delta)||_2\leqslant\frac{||I^{-1}(\delta)||_2^2||N(\theta_0,J,\delta)+I(\delta)||_2}{1-||I^{-1}(\delta)||_2||N(\theta_0,J,\delta)+I(\delta)||_2}.$$

On the other hand, using the same method in the proof of Proposition 2, we have

$$||U(\theta_0,J,\delta)||_2=O(\delta^{J+1}),$$

for any positive J and $\delta\in(0,\bar{\Delta})$. Hence,

$$||N^{-1}(\theta_0,J,\delta)I(\delta)+E||_2=O(\delta^J)\ \text{ and }\ ||N^{-1}(\theta_0,J,\delta)U(\theta_0,J,\delta)||_2=O(\delta^J),$$

under either (i) for fixed $\delta\in(0,\bar{\Delta})$ and $J\to\infty$, or (ii) for fixed J, $\delta\to0$. □

Proof of Proposition 4: Use the same method in the proof of Proposition 2. $\square$

Proof of Proposition 5: We'll use Corollary 2.1 in Newey (1989) to prove this proposition. We only need to verify three conditions under two situations mentioned in Proposition 5.

(i) For any $i \in \{1, \cdots, d\}$,

$$\mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta)\right\} \quad \text{is equicontinuous;}$$

(ii) For any $i \in \{1, \cdots, d\}$,

$$\sup_{\theta \in \Theta} \left|\left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta^T} \log f(X_t|X_{t-1}, \delta; \theta) \right|\right|_2 = O_p(1);$$

(iii) For any $i \in \{1, \cdots, d\}$ and $\theta \in \Theta$,

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta) - \mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta)\right\} \xrightarrow{p} 0.$$

For any $\theta^*, \theta^{**} \in \Theta$, note that

$$\begin{aligned}
&\mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta^*)\right\} - \mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta^{**})\right\} \\
&= \mathbb{E}\left\{\frac{\partial^2}{\partial \theta_i \partial \theta^T} \log f(X_t|X_{t-1}, \delta; \bar{\theta})\right\} \cdot (\theta^* - \theta^{**}),
\end{aligned}$$

where $\bar{\theta}$ is on the joint line between θ^* and θ^{**}. Then

$$\begin{aligned}
&\left|\mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta^*)\right\} - \mathbb{E}\left\{\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta^{**})\right\}\right| \\
&\leqslant \left|\left|\mathbb{E}\left\{\frac{\partial^2}{\partial \theta_i \partial \theta^T} \log f(X_t|X_{t-1}, \delta; \bar{\theta})\right\}\right|\right|_2 \cdot ||\theta^* - \theta^{**}||_2.
\end{aligned}$$

For any $j \in \{1, \cdots, d\}$, use the same method in the proof of Lemma 3, we know that there exists a constant C, which is not dependent on J and δ, and $\hat{\Delta} > 0$ such that, for any J and $\delta \in (0, \hat{\Delta}]$,

$$\mathbb{E}\left\{\sup_{\theta \in \Theta} \left|\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X_t|X_{t-1}, \delta; \theta)\right|\right\} < C.$$

Hence, (i) and (ii) can be established.

To verify (iii), we note that

$$\frac{\partial}{\partial \theta_i} \log f(X_t|X_{t-1}, \delta; \theta) = \frac{\partial}{\partial \theta_i} A_1(X_t|X_{t-1}, \delta; \theta) + \frac{\partial}{\partial \theta_i} A_2(X_t|X_{t-1}, \delta; \theta) + \frac{\partial}{\partial \theta_i} \widetilde{A}_3(X_t|X_{t-1}, \delta; \theta).$$

From (A.3), Lemma 3 and Lemma 4 in Aït-Sahalia and Mykland (2004), we know that there exists a positive constant κ such that for any $t_1 < t_2$,

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg\{ \left[\frac{\partial}{\partial\theta_i}\log f(X_{t_1}|X_{t_1-1},\delta;\theta) - \mathbb{E}\left\{\frac{\partial}{\partial\theta_i}\log f(X_{t_1}|X_{t_1-1},\delta;\theta)\right\}\right] \\
&\qquad\qquad \cdot \left[\frac{\partial}{\partial\theta_i}\log f(X_{t_2}|X_{t_2-1},\delta;\theta) - \mathbb{E}\left\{\frac{\partial}{\partial\theta_i}\log f(X_{t_2}|X_{t_2-1},\delta;\theta)\right\}\right] \Bigg\}\Bigg| \\
&\leqslant C\cdot \exp\{-\kappa(t_2-t_1)\delta\},
\end{aligned}$$

where

$$C = \mathbb{E}\Bigg\{ \left[\frac{\partial}{\partial\theta_i}\log f(X_t|X_{t-1},\delta;\theta) - \mathbb{E}\left\{\frac{\partial}{\partial\theta_i}\log f(X_t|X_{t-1},\delta;\theta)\right\}\right]^2 \Bigg\}.$$

Then,

$$\begin{aligned}
&\mathbb{E}\Bigg\{\frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta_i}\log f(X_t|X_{t-1},\delta;\theta) - \mathbb{E}\left\{\frac{\partial}{\partial\theta_i}\log f(X_t|X_{t-1},\delta;\theta)\right\}\right] \Bigg\}^2 \\
&\leqslant \frac{C}{n} + \frac{C}{n}\cdot\frac{\exp\{-\kappa\delta\}}{1-\exp\{-\kappa\delta\}} \leqslant\ 3\left[2K_1 + K_2\cdot m\left(\frac{2\nu_1}{\nu_1-2}\right)\right]\cdot\left\{\frac{1}{n} + \frac{1}{n[\exp(\kappa\delta)-1]}\right\} \\
&\to 0,
\end{aligned}$$

under the two situations mentioned in the statement of Proposition 5. Hence, we complete the proof. □

Proof of Theorem 2: For fixed δ, from Theorem 1 and (4.1), we know that the leading order term of $\hat\theta_n^{(J)} - \theta_0$ contains two parts, one is $N^{-1}U_n$, and the other is $N^{-1}(N_n + F_n)(\hat\theta_n - \theta_0)$. Hence, $\hat\theta_n^{(J)} - \theta_0 = O_p(\delta^{J+1} + n^{-1/2}\delta^{-1/2})$.

For J fixed and $\delta \to 0$, Proposition 4 implies

$$\mathbb{E}\left\{\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\hat\theta_n^{(J)})\right|\right|_2\right\} \leqslant C\delta^{J+1}.$$

Then,

$$\mathbb{E}\left\{\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\hat\theta_n^{(J)}) - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(X_t|X_{t-1},\delta;\hat\theta_n)\right|\right|_2\right\} \leqslant C\delta^{J+1}.$$

This means that

$$\mathbb{E}\left\{\left|\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(X_t|X_{t-1},\delta;\tilde\theta)\cdot(\hat\theta_n^{(J)} - \hat\theta_n)\right|\right|_2\right\} \leqslant C\delta^{J+1},$$

where $\tilde\theta$ is on the joining line between $\hat\theta_n^{(J)}$ and $\hat\theta_n$. Hence,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(X_t|X_{t-1},\delta;\tilde\theta)\cdot(\hat\theta_n^{(J)} - \hat\theta_n) = O_p(\delta^{J+1}).$$

Since $\tilde{\theta} \xrightarrow{p} \theta_0$ and $\hat{\theta}_n^{(J)} - \hat{\theta}_n = o_p(1)$,

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f(X_t|X_{t-1}, \delta; \theta_0) \cdot (\hat{\theta}_n^{(J)} - \hat{\theta}_n) = O_p(\delta^{J+1}).$$

On the other hand, from Proposition 2, we know

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f(X_t|X_{t-1}, \delta; \theta_0) - \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \log f^{(J)}(X_t|X_{t-1}, \delta; \theta_0) = O_p(\delta^{J+1}).$$

Then $N_n(\hat{\theta}_n^{(J)} - \hat{\theta}_n) = O_p(\delta^{J+1})$. Using the same way in verifying (iii) in the proof of Proposition 5, we know $N_n - N = O_p(n^{-1/2})$. As $n\delta^2 \to \infty$, then $N(\hat{\theta}_n^{(J)} - \hat{\theta}_n) = O_p(\delta^{J+1})$. Hence, $\hat{\theta}_n^{(J)} - \hat{\theta}_n = O_p(\delta^J)$. At the same time, we know $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2}\delta^{-1/2})$. Then,

$$\hat{\theta}_n^{(J)} - \theta_0 = O_p(\delta^J + n^{-1/2}\delta^{-1/2}).$$

This completes the proof of Theorem 2. □

References

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions, *Journal of 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, *Annals of Statistics*, **36**, 906-937.

Aït-Sahalia, Y., Fan, J. and Peng, H. (2009). Nonparametric transition-Based tests for jump-diffusions, *Journal of the American Statistical Association*, **104**, 1102-1116.

Aït-Sahalia, Y. and Kimmel, R. (2005). Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics*, **83**, 413-452.

Aït-Sahalia, Y. and Kimmel, R. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions, **98**, 113-144.

Aït-Sahalia, Y. and Mykland, P. (2004). Estimators of diffusions with randomly spaced discrete observations: a general theory, *Annals of Statistics*, **32**, 2186-2222.

Bakshi, G. Cao, C. and Chen, Z. (1997). Empirical performance of alternative option pricing models, *Journal of Finance*, **52**, 2003-2049.

Ball, C. and Torous, W. (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, **3**, 215-238.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637-654.

Bibby, B. and Sørensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17-39.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. (1985). A theory of term structure of interest rates, *Econometrica*, **53**, 385-407.

Cramër, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ.

Dumas, B., Fleming, J. and Whaley, R. E. (1998). Implied volatility functions: empirical tests, *Journal of Finance*, **53**, 2059-2106.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.*, **20**, 317-357.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, **102**, 1349-1362.

Fan, J. and Zhang, C.M. (2003). A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, **98**, 118-134.

Feller, W. (1952). The parabolic differential equations and the associated semi-groups of transformations, *Annals of Mathematics*, **55**, 468-519.

Friedman, A. (1964). *Partial Differential Equations of Parabolic Type.* Prentice-Hall, Englewood Cliffs, NJ.

Merton, R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4**, 141-183.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity, *Econometrica*, **59**, 1161-1167.

Prakasa Rao, B. L. S. (1999). *Semimartingales and Statistical Inference*, Chapman and Hall/CRC, London.

Øksendal, B. (2000). *Stochastic Differential Equations: An Introduction with Applications*, Fifth Edition, Berlin: Springer.

Sundaresan, S. M. (2000). Continuous time finance: a review and assessment, *Journal of Finance*, **55**, 1569-1622.

Stramer, O. and Yan, J. (2007). On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation, *Journal of Computational and Graphical Statistics*, **16**, 672-691.

Stramer, O. and Yan, J. (2007). Asymptotics of an efficient Monte Carlo estimation for the transition density of diffusion processes. *Methodology and Computing in Applied Probability*, **9**, 483-496.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes, *Journal of Econometrics*, **149**, 65-81.

Vasicek, O. (1977): An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177-188.

Wang, Y. (2002) Asymptotic nonequivalence of GARCH models and diffusions. *The Annals of Statistics*, **30**, 754-783.

Table 2: Simulated average bias (Bias) and standard deviations (SD) for the full MLE and the two AMLE $\hat{\theta}_n^{(J)}$ with $J = 1$ and 2 respectively for Vasicek Model; The rows headed with ABias and ASD are asymptotic bias and SD based on formulae (5.9) and (5.10); and the rows headed with RMSD represent the root of mean squared deviation between $\hat{\theta}_n$ and $\hat{\theta}_n^{(J)}$.

$\kappa = 0.858, \alpha = 0.0891, \sigma = 0.0468, \delta = 1/12$										
Method	Statistics	$n = 500$			$n = 1000$			$n = 2000$		
		κ	α	σ	κ	α	σ	κ	α	σ
MLE	Bias	0.0992	0.0002	4.39e-5	0.0518	-0.0002	7.05e-5	0.0245	-3.97e-5	2.69e-5
	SD	0.2307	0.0085	0.0016	0.1624	0.0058	0.0011	0.1114	0.0042	0.0008
$J = 1$	Bias	0.0896	0.0002	4.14e-5	0.0419	-0.0002	6.68e-5	0.0149	-3.34e-5	2.30e-5
	ABias	0.0908	0.0003	4.55e-5	0.0446	-0.0001	0.0001	0.0179	-2.63e-5	4.55e-5
	SD	0.2255	0.0085	0.0016	0.1586	0.0058	0.0011	0.1091	0.0041	0.0008
	ASD	0.2251	0.0084	0.0016	0.1585	0.0057	0.0011	0.1088	0.0041	0.0008
	RMBD	0.0173	0.0002	1.36e-5	0.0100	0.0001	7.39e-6	0.0100	0.0001	6.27e-6
$J = 2$	Bias	0.0992	0.0002	4.39e-5	0.0520	-0.0002	7.06e-5	0.0246	-4.01e-5	2.70e-5
	ABias	0.1016	0.0002	4.55e-5	0.0529	-0.0002	0.0001	0.0249	-2.98e-5	4.55e-5
	SD	0.2309	0.0085	0.0016	0.1625	0.0058	0.0011	0.1115	0.0042	0.0008
	ASD	0.2366	0.0085	0.0016	0.1666	0.0058	0.0011	0.1143	0.0042	0.0008
	RMBD	0.0062	1.28e-5	1.05e-5	0.0008	9.14e-6	7.80e-7	0.0006	7.37e-6	7.80e-7
$\kappa = 0.858, \alpha = 0.0891, \sigma = 0.0468, \delta = 1/4$										
Method	Statistics	$n = 500$			$n = 1000$			$n = 2000$		
		κ	α	σ	κ	α	σ			
MLE	Bias	0.0380	4.09e-5	9.12e-5	0.0170	1.83e-5	3.66e-5	0.0084	-5.72e-5	4.00e-5
	SD	0.1366	0.0050	0.0016	0.0957	0.0034	0.0012	0.0647	0.0024	0.0008
$J = 1$	Bias	0.0127	5.63e-5	7.13e-5	-0.0095	2.81e-5	6.83e-6	-0.0191	-4.90e-5	9.21e-6
	ABias	0.0174	0.0002	0.0001	-0.0097	1.69e-5	3.29e-5	-0.0085	0.0001	4.55e-5
	SD	0.1290	0.0050	0.0016	0.0905	0.0034	0.0012	0.0611	0.0024	0.0008
	ASD	0.1215	0.0047	0.0016	0.0849	0.0032	0.0012	0.0576	0.0023	0.0008
	RMBD	0.0332	0.0005	0.0001	0.0316	0.0004	0.0001	0.0300	0.0003	0.0001
$J = 2$	Bias	0.0396	4.17e-5	9.43e-5	0.0186	1.58e-5	3.96e-5	0.0100	-5.80e-5	4.34e-5
	ABias	0.0376	0.0001	0.0001	0.0161	1.45e-5	4.55e-5	0.0071	-0.0001	4.55e-5
	SD	0.1386	0.0050	0.0016	0.0966	0.0034	0.0012	0.0652	0.0024	0.0008
	ASD	0.1403	0.0050	0.0016	0.0982	0.0034	0.0012	0.0665	0.0024	0.0008
	RMBD	0.0316	0.0002	0.0001	0.0063	0.0001	1.59e-5	0.0042	4.68e-5	1.02e-5

Table 3: Simulated average bias (Bias) and standard deviations (SD) for the full MLE and the two AMLE $\hat{\theta}_n^{(J)}$ with $J = 1$ and 2 respectively for CIR Model; The rows headed with ABias and ASD are asymptotic bias and SD based on formulae (5.9) and (5.10); and the rows headed with RMSD represent the root of mean squared deviation between $\hat{\theta}_n$ and $\hat{\theta}_n^{(J)}$.

$\kappa = 0.892, \alpha = 0.09, \sigma = 0.1817, \delta = 1/12$

Method	Statistics	$n = 500$			$n = 1000$			$n = 2000$		
		κ	α	σ	κ	α	σ	κ	α	σ
MLE	Bias	0.0980	0.0001	0.0003	0.0521	-1.54e-5	3.86e-5	0.0295	-0.0002	0.0002
	SD	0.2389	0.0093	0.0060	0.1596	0.0067	0.0043	0.1082	0.0048	0.0030
$J = 1$	Bias	0.0910	0.0004	0.0003	0.0435	0.0002	4.35e-5	0.0199	0.0001	0.0002
	A.Bias	0.0818	0.0005	0.0003	0.0411	0.0004	3.17e-5	0.0213	0.0002	0.0002
	SD	0.2340	0.0093	0.0060	0.1558	0.0067	0.0043	0.1053	0.0048	0.0031
	A.SD	0.2169	0.0091	0.0060	0.1452	0.0066	0.0040	0.1181	0.0047	0.0030
	RMBD	0.0200	0.0009	0.0004	0.0173	0.0003	0.0002	0.0173	0.0004	0.0005
$J = 2$	Bias	0.0978	0.0001	0.0003	0.0521	-2.22e-5	3.81e-5	0.0294	-0.0002	0.0002
	ABias	0.0984	0.0001	0.0003	0.0525	-3.43e-5	2.69e-5	0.0299	-0.0002	0.0002
	SD	0.2405	0.0093	0.0060	0.1603	0.0067	0.0043	0.1088	0.0048	0.0030
	ASD	0.2389	0.0093	0.0060	0.1596	0.0067	0.0043	0.1105	0.0048	0.0030
	RMBD	0.0224	0.0004	0.0004	0.0141	2.66e-5	3.91e-5	0.0068	0.0001	0.0003

$\kappa = 0.892, \alpha = 0.09, \sigma = 0.1817, \delta = 1/4$

Method	Statistics	$n = 500$			$n = 1000$			$n = 2000$		
		κ	α	σ	κ	α	σ	κ	α	σ
MLE	Bias	0.0371	-6.38e-5	0.0004	0.0218	-0.0002	0.0003	0.0103	-3.06e-5	3.05e-5
	SD	0.1437	0.0055	0.0065	0.0968	0.0039	0.0045	0.0696	0.0028	0.0033
$J = 1$	Bias	0.0234	0.0008	0.0005	0.0070	0.0007	0.0006	-0.0057	0.0010	0.0006
	ABias	0.0207	0.0008	0.0004	0.0095	0.0007	0.0003	-0.0011	0.0006	0.0005
	SD	0.1338	0.0054	0.0065	0.0861	0.0037	0.0045	0.0607	0.0027	0.0037
	ASD	0.1159	0.0064	0.0067	0.0823	0.0044	0.0047	0.0592	0.0027	0.0034
	RMSD	0.0447	0.0018	0.0017	0.0447	0.0020	0.0021	0.0424	0.0020	0.0027
$J = 2$	Bias	0.0388	-0.0001	0.0003	0.0186	-0.0003	0.0003	0.0069	-9.87e-5	1.33e-5
	ABias	0.0513	-0.0001	0.0002	0.0262	-0.0003	0.0001	0.0147	-0.0001	1.06e-5
	SD	0.2256	0.0055	0.0069	0.0980	0.0039	0.0045	0.0698	0.0028	0.0033
	ASD	0.1938	0.0055	0.0065	0.0969	0.0039	0.0045	0.0697	0.0028	0.0033
	RMSD	0.1622	0.0004	0.0021	0.0200	0.0001	0.0002	0.0100	0.0001	0.0001

Table 4: P-values of Kolmogorov-Smirnov test for $W_n(J) \sim \chi^2_3$.

Situation	n	δ	$J = 1$	$J = 2$
$\delta = n^{-1/6}$	500	0.3550	0.3524	0.0587
	1000	0.3162	0.4595	0.5830
	2000	0.2817	0.1149	0.2710
	4000	0.2510	0.0019	0.8309
	8000	0.2236	5.74e-8	0.6002
$\delta = n^{-1/2}$	500	0.0447	5.04e-7	2.45e-8
	1000	0.0316	0.0003	9.72e-5
	2000	0.0224	0.0006	0.0003
	4000	0.0158	0.1109	0.0851
	8000	0.0112	0.0470	0.0367