

Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model¹

Minsoo Jeong

*Department of Economics
Indiana University*

Joon Y. Park

*Department of Economics
Indiana University and SKKU*

Abstract

We derive the asymptotics of the maximum likelihood estimators for diffusion models. The models considered in the paper are very general, including both stationary and nonstationary diffusions. For such a broad class of diffusion models, we establish the consistency and derive the limit distributions of the exact maximum likelihood estimator, and also the quasi and approximate maximum likelihood estimators based on various versions of approximated transition densities. Our asymptotics are two dimensional, requiring the sampling interval to decrease as well as the time span of sample to increase. The two dimensional asymptotics provide a unifying framework for such a broad class of estimators for both stationary and nonstationary diffusion models. More importantly, they yield the primary asymptotics that are very useful to analyze the exact, quasi and approximate maximum likelihood estimators of the diffusion models, if the samples are collected at high frequency intervals over modest lengths of sampling horizons as in the case of many practical applications.

This version: September 2013

JEL classification: C22, C50, G12

Keywords: diffusion, exact, approximate and quasi maximum likelihood estimations, asymptotics, Euler and Milstein approximations, transition density, positive and null recurrences, consistency, limit distribution.

¹We are grateful to Co-Editor and three anonymous referees for many useful comments. We also would like to thank Yacine Aït-Sahalia, Yoosoon Chang, Xiaohong Chen, Daren Cline, Jean Jacod, and seminar and conference participants at Workshop on Financial Econometrics at Fields Institute, 2010 International Symposium on Financial Engineering and Risk Management, 2011 ISI World Statistics Congress, Yale, Michigan State, Rochester, Michigan and Queens for helpful discussions and suggestions. Park gratefully acknowledges the financial support from the NSF under NSF Grant No. SES-0518619.

1. Introduction

Diffusion models are widely used in the analysis of financial markets. Their popularity may be attributable to many different reasons. However, one of the main reasons appears to be that they are simple and parsimonious, yet flexible enough to generate complicated and realistic dynamics. The reader is referred to Karlin and Taylor (1981) and Karatzas and Shreve (1991) for a good introduction to diffusion models, and to Duffie (2001) and Cochrane (2005) for their applications in economics and finance. Naturally, there is a large literature on the estimation of diffusion models, both parametric and nonparametric. To estimate diffusion models nonparametrically, we may use the standard kernel method, as shown in, e.g., Bandi and Phillips (2003, 2010). For the parametric estimation of diffusion models, available are numerous methods based on a large spectrum of different approaches ranging from the GMM's defined by some orthogonality conditions to the MLE's relying on the exact or approximated transition densities. It seems, however, that the approximate MLE proposed by Aït-Sahalia (2002) is most popular. See Phillips and Yu (2009) for a recent survey on the likelihood-based estimation of diffusion models.

In this paper, we develop a general asymptotic theory of MLE for diffusion models. The limit theories of MLE's available in the literature for diffusion models are highly model-specific and limited to stationary processes. There are two major hurdles in establishing the limit theories for MLE's in diffusion models at any general level. First, except for a few special cases, the transition of a diffusion generally cannot be represented by a closed form density. Therefore, we either have to rely on a complicated numerical method to obtain the exact MLE, or use the quasi and approximate MLE's based on transition densities approximated in a variety of different methods. This makes it difficult to develop a general asymptotic theory that is applicable for all such MLE's. Second, the limit distributions of general nonstationary diffusions are not available except for some simple cases where the underlying diffusion can be transformed into a Brownian motion with drift or an explosive Ornstein-Uhlenbeck process. As a consequence, the asymptotic theories of MLE's in nonstationary diffusions are largely unknown. This is unfortunate, since in many applications we use diffusion models to describe processes that are obviously nonstationary.

The class of diffusion models we consider in the paper is truly broad and includes very general nonstationary, as well as stationary, diffusions. Moreover, our theory is applicable not only for the exact MLE, but also for the quasi and approximate MLE's based on various versions of approximated transition densities, such as among others those implied by the Euler and Milstein approximations of the underlying diffusions and the one obtained by Aït-Sahalia (2002) in his development of the approximate MLE using a closed-form approximation of the transition density by Hermite polynomials. Our asymptotics are two dimensional, having two parameters that are designated respectively to the sampling interval and the time span of sample, as in Bandi and Phillips (2003, 2010) and Bandi and Moloche (2004). More specifically, for the development of our limit theory in the paper we let the sampling interval diminish to zero and the time span of sample increase up to infinity at appropriate rates. In particular, our asymptotics rely on both infill and long-span. This is in contrast with the conventional asymptotics based only on the sample size with the fixed sampling interval.

The two dimensional asymptotics provide a single framework to unify the limit theories of a broad class of the MLE's for both stationary and nonstationary diffusion models. Our main asymptotics do not require stationarity, and the nonstationary diffusions are analyzed exactly in the same manner as the stationary diffusions under very mild regularity conditions. For the stationary diffusions, our approach of course yields the same results as the conventional asymptotics relying only on the sample size. Moreover, the two dimensional asymptotics allow us to consider the exact, quasi and approximate MLE's within a unified framework. In fact, our two dimensional asymptotics provide the distributional results that are much more useful and relevant in practical applications, compared with the conventional one dimensional asymptotics. For instance, as we will explain in more detail below, our asymptotics make it clear that the drift and diffusion term parameters have differing limit behaviors in regards to the sampling frequency and the sample horizon. Furthermore, our theoretical development provides primary asymptotics, which well approximates the finite sample distributions of the MLE's in case of the samples collected at high frequencies for relatively short period of time span. This is usually the case in a majority of practical

applications.

Our asymptotic results reveal many important statistical properties of the MLE's for diffusion models. First, the drift term parameter estimates become consistent only when the sample horizon T increases, whereas the diffusion term parameters can be estimated consistently as long as the sample size increases either by a decrease in sampling interval Δ or by an increase in sample horizon T . The actual convergence rates are determined by the drift and diffusion functions and the recurrence property of the underlying diffusion. For positive recurrent diffusions, they are given respectively by \sqrt{T} and $\sqrt{T/\Delta}$ for the drift and diffusion term parameters. Second, the distributions of the drift and diffusion term parameter estimates become uncorrelated for all large T as Δ shrinks down to zero fast enough. The distributions of the diffusion term parameter estimate become mixed normal for all large T as long as Δ is sufficiently small. On the other hand, the distributions of the drift term parameter estimates are non-Gaussian unless T increases up to infinity. If T reaches to infinity, they become normal in general for stationary diffusions. However, we expect them to be generally non-Gaussian asymptotically and their limit distributions reduce to a generalized version of the Dickey-Fuller distribution appearing in the limit theory of unit root test.

We demonstrate by simulation that our primary asymptotics provide superb approximations for the finite sample distributions of the MLE's even for small sample horizon T , as long as sampling interval Δ is sufficiently small. Our primary asymptotics are particularly useful in approximating the finite sample distributions of the drift term parameters, which are generally quite distant from their limit distributions unless sample horizon T is unrealistically large. In fact, it is shown very clearly in our simulations that our primary asymptotics are very effective in correcting biases and asymptotic critical values of the drift term parameter estimates and their test statistics. Moreover, our simulation results imply that all of the exact, quasi and approximate MLE's considered in the paper should perform comparably in finite samples as long as Δ is small enough. They yield the same primary asymptotics in our asymptotic analysis, from which we may infer that their finite sample distributions are close each other for all T if Δ is sufficiently small relative to T . This,

of course, does not necessarily imply that the quasi and approximate MLE's are always expected to behave as well as the exact MLE in finite samples. At least, however, we may say that the use of the exact MLE is not very compelling when Δ is small, and it is more so if the transition density is not given in a closed form and it is computationally expensive to obtain the exact MLE.

The rest of the paper is organized as follows. In Section 2, we present the background and preliminaries that are necessary to develop our asymptotic theory of the MLE's for diffusion models. A parametric diffusion model is specified and its basic recurrence property is discussed with some examples. Moreover, various MLE's based on the exact and approximated transition densities are introduced. Section 3 develops our framework and some fundamental theories required in the establishment of the asymptotic theory for the MLE's in diffusion models. In particular, continuous approximations of the discrete likelihoods are provided and relevant continuous time asymptotics are presented. Subsequently in Section 4, we obtain our primary asymptotics and derive the limit distributions of the MLE's. Some examples are also given as an illustration of our asymptotic results. In Section 5, we report some simulation results, which demonstrate the relevancy and usefulness of our primary asymptotics in approximating the finite sample distributions of the MLE's. Section 6 concludes the paper. Appendix includes some useful technical lemmas and their proofs, as well as the proofs of the theorems in the paper.

2. Background and Preliminaries

To develop the asymptotics of the MLE's for the diffusion models, it is necessary to introduce some background and preliminary theories on diffusion processes and the MLE's defined from the exact and various other approximated transition densities. Since our theoretical developments are quite extensive and complicated, we need to make some notational conventions to facilitate our exposition. The notation " \sim " is used to denote the asymptotic equivalence, and $P \sim Q$ means that $P/Q \rightarrow 1$ or $P - Q = o(Q)$. On the other hand, " $P \simeq Q$ " just implies that we approximate P by Q , and it does not have any precise mathematical meaning in regards to the proximity between P and Q . Moreover, for a measure

λ on \mathbb{R} that is absolutely continuous with respect to the Lebesgue measure, we also use the same notation λ to denote its density with respect to the Lebesgue measure. This should cause no confusion.

2.1 The Model

We consider the diffusion process X given by the time-homogeneous stochastic differential equation (SDE)

$$dX_t = \mu(X_t, \alpha)dt + \sigma(X_t, \beta)dW_t, \quad (1)$$

where μ and σ are respectively the drift and diffusion functions, and W is the standard Brownian motion. We define $\theta = (\alpha', \beta')'$ to be the parameter in our model, which belongs to the parameter space Θ , with its true value denoted by $\theta_0 = (\alpha_0, \beta_0)$. Moreover, we let $\mathcal{D} = (\underline{x}, \bar{x})$ denote the domain of the diffusion process X , where we allow $\underline{x} = -\infty$ and $\bar{x} = \infty$. Throughout the paper, we assume that a weak solution to the SDE in (1) exists and X is well defined uniquely in probability law. The reader is referred to, e.g., Karlin and Taylor (1981), Karatzas and Shreve (1991), Rogers and Williams (2000) and Revuz and Yor (1999) for more discussions on the solutions to the SDE (1). Finally, we assume that the diffusion X admits a transition density p with respect to the Lebesgue measure, with $p(t, x, \cdot)$ representing the conditional density of X_t given $X_0 = x$. More precise assumptions we need to develop our asymptotic theory for the MLE will be introduced later.

The scale function of the diffusion process X introduced in (1) is defined as

$$s(x, \theta) = \int_w^x \exp\left(-\int_w^y \frac{2\mu(z, \theta)}{\sigma^2(z, \theta)} dz\right) dy \quad (2)$$

for some $w \in \mathcal{D}$. Defined as such, the scale function s is only identified up to an affine transformation, i.e., if s is a scale function, then so is $as + b$ for any constants a and b . A diffusion process $Y_t = s(X_t)$ transformed with its scale function becomes a driftless diffusion and we say that it is in the natural scale. Of course, the scale function of a driftless diffusion is the identity function. We also define the speed density

$$m(x, \theta) = \frac{1}{(\sigma^2 s')(x, \theta)} \quad (3)$$

on \mathcal{D} , where $s^*(x, \theta) = (\partial/\partial x)s(x, \theta)$. The speed measure is defined to be the measure on \mathcal{D} given by the speed density with respect to the Lebesgue measure.²

Our asymptotic theory for the MLE depends crucially on the recurrence property of the underlying diffusion X . To define the recurrence property, we let τ_y be the hitting time of a point y in \mathcal{D} that is given by $\tau_y = \inf\{t \geq 0 | X_t = y\}$. We say that a diffusion is recurrent if $\mathbb{P}\{\tau_y < \infty | X_0 = x\} = 1$ for all x and y in the interior of \mathcal{D} . A recurrent diffusion is said to be null recurrent if $\mathbb{E}(\tau_y | X_0 = x) = \infty$ for all x and y in the interior of \mathcal{D} , and positive recurrent if $\mathbb{E}(\tau_y | X_0 = x) < \infty$. When the drift and diffusion functions satisfy the usual regularity conditions that we will introduce later, the diffusion X in (1) is recurrent if and only if the scale function s in (2) is unbounded at both boundaries \underline{x} and \bar{x} .³ It is positive recurrent if $m(\mathcal{D}, \theta) < \infty$, and null recurrent if $m(\mathcal{D}, \theta) = \infty$. For a positive recurrent diffusion X , we let

$$\pi(x, \theta) = \frac{m(x, \theta)}{m(\mathcal{D}, \theta)}. \quad (4)$$

If the initial value of the process X_0 has density π , then the process X becomes stationary with the time invariant density π . A diffusion which is not recurrent is said to be transient.

Example 2.1 (a) The Brownian motion (BM) with drift is a diffusion generated as

$$dX_t = \alpha dt + \beta dW_t \quad (5)$$

with $\beta > 0$ and $\mathcal{D} = (-\infty, \infty)$. Its transition density can easily be obtained, since the distribution of X_t given $X_0 = x$ is normal with mean $x + \alpha t$ and variance $\beta^2 t$ for $t \geq 0$. The process becomes null recurrent if $\alpha = 0$, in which case the speed measure is given by a scaled Lebesgue measure. It becomes transient, if $\alpha \neq 0$. For the geometric Brownian motion (GBM) given by the SDE $dX_t = \nu X_t dt + \omega X_t dW_t$ with $\omega > 0$, $(\log X_t)$ becomes the BM with drift in (5) with $\alpha = \nu - \omega^2/2$ and $\beta = \omega$.

(b) The Ornstein-Uhlenbeck (OU) process is defined on $\mathcal{D} = (-\infty, \infty)$ as the solution to the SDE

$$dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta dW_t \quad (6)$$

²Following our notational convention discussed earlier, we also use $m(\cdot, \theta)$ to denote the speed measure, as well as the speed density with respect to the Lebesgue measure.

³See Karatzas and Shreve (1991), Chapter 5, Proposition 5.22 for more details.

with $\alpha_2 < 0$ and $\beta > 0$. Vasicek (1977) used the process to model the short-term interest rate. It has the transition given by normal distribution with mean $e^{\alpha_2 t}(x + \alpha_1/\alpha_2)$ and variance $(\beta^2/2\alpha_2)(e^{2\alpha_2 t} - 1)$. It is positive recurrent with time invariant stationary distribution given by normal with mean $-\alpha_1/\alpha_2$ and variance $-\beta^2/2\alpha_2$. The process becomes transient if the mean reversion parameter $\alpha_2 > 0$.

(c) The Feller's square-root (SR) process is given by the SDE

$$dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta\sqrt{X_t}dW_t \quad (7)$$

on $\mathcal{D} = (0, \infty)$, where $\alpha_2 < 0$ and $2\alpha_1/\beta^2 \geq 1$. The process was used by Cox, Ingersol and Ross (1985) to study the term structure of interest rates. The conditional distribution of $\beta^2(e^{\alpha_2 t} - 1)X_t/4\alpha_2$ given $X_0 = x$ follows the noncentral chi-squared distribution with degrees of freedom $4\alpha_1/\beta^2$ and noncentrality parameter $-4\alpha_2 e^{\alpha_2 t} x/\beta^2(e^{\alpha_2 t} - 1)$. It is positive recurrent with the time invariant distribution given by gamma distribution with parameters $2\alpha_1/\beta^2$ and $-2\alpha_2/\beta^2$.

Example 2.2 (a) The constant elasticity of variance (CEV) process is given by the SDE

$$dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta_1 X_t^{\beta_2} dW_t \quad (8)$$

with $\alpha_1 > 0$, $\alpha_2 < 0$, $\beta_1 > 0$, $\beta_2 > 1/2$ and $\mathcal{D} = (0, \infty)$. For this process, we cannot obtain the exact transition density in a closed-form. If $\alpha_1 = 0$ and $\alpha_2 = 0$, then SDE defining CEV process reduces to what is known as the Girsanov SDE. The Girsanov SDE has the trivial solution $X_t \equiv 0$ for all $t \geq 0$. When $\beta_2 < 1/2$, however, it also has a nontrivial weak solution. See, e.g., Rogers and Williams (2000, pp. 175-176).

(b) The nonlinear drift (NLD) diffusion process introduced in Ait-Sahalia (1996) is also used by several authors (with some parameter restrictions) including Ahn and Gao (1999) and Hong and Li (2005) for modeling interest rate processes. It is given by the SDE

$$dX_t = (\alpha_1 + \alpha_2 X_t + \alpha_3 X_t^2 + \alpha_4 X_t^{-1})dt + \sqrt{\beta_1 + \beta_2 X_t + \beta_3 X_t^{\beta_4}} dW_t \quad (9)$$

defined on $\mathcal{D} = (0, \infty)$. The parameter ranges to guarantee the positive recurrent solution for this SDE, i.e., $s(0) = -\infty$, $s(\infty) = \infty$ and $m(\mathcal{D}, \theta) < \infty$, are given by Ait-Sahalia (1996)

as

$$\alpha_3 \leq 0 \text{ and } \alpha_2 < 0 \text{ if } \alpha_3 = 0,$$

$$\alpha_4 > 0 \text{ and } 2\alpha_4 \geq \beta_1 \geq 0, \text{ or } \alpha_4 = 0, \alpha_1 > 0, \beta_1 = 0, \beta_4 > 1 \text{ and } 2\alpha_1 \geq \beta_2 > 0,$$

$$\beta_1 \geq 0 \text{ (and } \beta_3 > 0 \text{ if } \beta_1 = 0 \text{ and } 0 < \beta_4 < 1, \text{ or } \beta_2 > 0 \text{ if } \beta_1 = 0 \text{ and } \beta_4 > 1),$$

$$\beta_3 > 0 \text{ if either } \beta_4 > 1 \text{ or } \beta_2 = 0, \text{ and } \beta_2 > 0 \text{ if either } 0 < \beta_4 < 1 \text{ or } \beta_3 = 0.$$

For a certain set of parameter values, we have $m(\mathcal{D}) = \infty$ and the process becomes null recurrent. For instance, if we set $\alpha_2 = 0$, $\alpha_3 = 0$ and $\beta_3 = 0$ and consider the process given by the SDE

$$dX_t = (\alpha_1 + \alpha_4 X_t^{-1})dt + \sqrt{\beta_1 + \beta_2 X_t}dW_t$$

with $\beta_1 > 0$, $\beta_2 > 0$, $0 < \alpha_1 < \beta_2/2$ and $\alpha_4 > \beta_1/2$, then we have

$$s(x, \theta) \sim c_1 x^{1-2\alpha_1/\beta_2} \text{ as } x \rightarrow \infty \text{ and } c_2 x^{1-2\alpha_4/\beta_1} \text{ as } x \rightarrow 0,$$

$$m(x, \theta) \sim c_3 x^{2\alpha_1/\beta_2-1} \text{ as } x \rightarrow \infty \text{ and } c_4 x^{2\alpha_4/\beta_1} \text{ as } x \rightarrow 0$$

for some constants c_1 , c_2 , c_3 and c_4 , and the process becomes null recurrent.

For the development of our asymptotics, we need to know the divergence rate of the extremal process of X given by $\sup_{t \in [0, T]} X_t$. For several positive recurrent processes that are used widely in economics and finance applications, the exact order of extremal process is well known. The reader is referred to Borkovec and Klüppelberg (1998) for details. For example, the extremal processes of the Ornstein-Uhlenbeck process and the Feller's square root process are respectively of orders $O_p(\sqrt{\log T})$ and $O_p(\log T)$, and the extremal process of the CEV process has order less than or equal to $O_p(T)$ depending upon its parameter values. It is also possible to find appropriate asymptotic bounds of the extremal processes for more general positive recurrent processes, utilizing the result in Davis (1982) which shows that the extremal processes of positive recurrent processes are stochastically bounded by $s^{-1}(T)$ if s^{-1} is regularly varying. In fact, Cline and Jeong (2009) establish that the extremal process is at most of order $O_p(T^r)$ for some $r < \infty$ if μ and σ are regularly varying, provided that $\lim_{x \rightarrow \infty} [x(\mu/\sigma^2)(x)] \neq 1/2$. To obtain the asymptotics

of the extremal process $\inf_{t \in [0, T]} X_t$ for a diffusion having a boundary at the origin, we may use the Ito's lemma to get the drift and diffusion functions of the transformed process $X_t^* = X_t^{-1}$ as

$$dX_t^* = [\sigma^2(X_t^{*-1})X_t^{*3} - \mu(X_t^{*-1})X_t^{*2}]dt - \sigma(X_t^{*-1})X_t^{*2}dW_t,$$

and analyze the extremal process $\sup_{t \in [0, T]} X_t^*$ of X . Note that the drift and diffusion functions are regularly varying for X^* , if they are so for X .

For null recurrent processes, Stone (1963) shows that under suitable regularity conditions on the speed measure of the underlying process, we may find a proper normalization sequence (c_T) , for which the normalized extremal process has a well defined limit distribution. The most well known and useful example of this case is Brownian motion, which has $c_T = \sqrt{T}$. For the general null recurrent processes, if the speed density of the process X^s , $X_t^s = s(X_t)$, is regularly varying with index $r > -1$, then there exists such a normalizing sequence (c_T) , as long as s^{-1} is regularly varying at infinities. The asymptotic behaviors of null recurrent processes will be explored in much more detail in later sections.

2.2 Maximum Likelihood Estimators

Throughout the paper, we assume that the samples of size n collected from the diffusion process (X_t) at interval Δ over time T , i.e.,

$$X_\Delta, X_{2\Delta}, \dots, X_{n\Delta}$$

with $T = n\Delta$, are available, and we denote their observations by $x_\Delta, x_{2\Delta}, \dots, x_{n\Delta}$. Furthermore, we suppose that the exact, approximated or quasi transition density function for the underlying diffusion process (X_t) is available over time interval of length Δ and denoted by $p(\Delta, x, y, \theta)$. The exact, approximate or quasi MLE $\hat{\theta}$ of θ relying on the transition density function $p(\Delta, x, y, \theta)$ is then defined as the maximizer of the log-likelihood

$$\mathcal{L}(\theta) = \sum_{i=1}^n \log p(\Delta, x_{(i-1)\Delta}, x_{i\Delta}, \theta),$$

i.e., $\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}(\theta)$. In the subsequent development of our asymptotic theory, we assume that the parameter space Θ is compact and convex, and the true parameter value θ_0 is an interior point of Θ .

The theoretical results and their derivations that we subsequently develop in the paper are quite complicated and involve various functions with multiple arguments. It will therefore be necessary to make an appropriate convention for the use of notation. First, we will suppress the argument θ in μ , σ , p and all other functions defined from them whenever they are evaluated at θ_0 , to make our presentation simple and more tractable. For instance, we will use $\mu(x)$, $\sigma(x)$ and $p(t, x, y)$, in place of $\mu(x, \alpha_0)$, $\sigma(x, \beta_0)$ and $p(t, x, y, \theta_0)$. Second, for any function f only with a scalar argument x other than θ , i.e., $f(x, \theta)$, we will routinely denote its first and second derivatives with respect to x simply by $f'(x, \theta)$ and $f''(x, \theta)$. As an example, we will write $\sigma'(x)$ or $\sigma'(x, \beta)$, instead of $(\partial/\partial x)\sigma(x, \beta_0)$ or $(\partial/\partial x)\sigma(x, \beta)$. Third, we put the differentiating parameters or variables as subscripts as in $f_\theta(x, \theta)$, $f_y(x, y, \theta)$ or $f_{y\theta}(x, y, \theta)$ to denote the derivatives with respect to the parameters or the derivatives of functions that involve multiple arguments as well as the parameters. Therefore, we use the notation such as $p_{ty\theta}(t, x, y)$ or $p_{ty\theta}(t, x, y, \theta)$. This convention will be made throughout the paper, and should cause no confusion.

For the diffusion models with known and tractable transition densities, we may of course find the exact MLE. As the exact transition density of the diffusion models are generally not available and cannot be given in closed forms, however, we should rely on the approximated transition densities in many cases. The simplest approach to obtain an approximated transition density is to use the Euler scheme. It is based on the first order expansion of SDE in (1), which we write as

$$X_{i\Delta} - X_{(i-1)\Delta} \simeq \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}). \quad (10)$$

The implied transition density for the Euler scheme is given by

$$p_{EU}(\Delta, x, y, \theta) = \frac{1}{\sqrt{2\pi\Delta}\sigma(x, \beta)} \exp \left[-\frac{(y - x - \Delta\mu(x, \alpha))^2}{2\Delta\sigma^2(x, \beta)} \right]. \quad (11)$$

The conditional distribution of $X_{i\Delta}$ given $X_{(i-1)\Delta} = x$ given by the Euler approximation (10) is normal with mean $x + \Delta\mu(x)$ and variance $\Delta\sigma^2(x)$, from which the Euler transition density (11) can easily be derived.

The Milstein scheme introduces an additional term to the expansion of SDE in (1),

which yields

$$\begin{aligned} X_{i\Delta} - X_{(i-1)\Delta} &\simeq \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\ &\quad + \frac{1}{2}(\sigma\sigma^\bullet)(X_{(i-1)\Delta})[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta]. \end{aligned} \quad (12)$$

Unlike the Euler approximation, the Milstein approximation does not yield the normal transition density. The transition density implied by the Milstein approximation is a mixture of normal and chi-square distribution, and given by

$$\begin{aligned} p_{MS}(\Delta, x, y, \theta) & \quad (13) \\ &= \frac{1}{\sqrt{2\pi\Delta} \varpi(x, y, \theta)} \left(\exp \left[-\frac{(\varpi(x, y, \theta) + \sigma(x, \beta))^2}{2\Delta(\sigma\sigma^\bullet)^2(x, \beta)} \right] + \exp \left[-\frac{(\varpi(x, y, \theta) - \sigma(x, \beta))^2}{2\Delta(\sigma\sigma^\bullet)^2(x, \beta)} \right] \right), \end{aligned}$$

where

$$\varpi(x, y, \theta) = [\sigma^2(x, \beta) + \Delta(\sigma\sigma^\bullet)^2(x, \beta) + 2(\sigma\sigma^\bullet)(x, \beta)(y - x - \Delta\mu(x, \alpha))]^{1/2}.$$

The Milstein transition density (13) can easily be obtained by the standard distribution function technique, if we note that the conditional distribution of $X_{i\Delta}$ given $X_{(i-1)\Delta} = x$ is identical to the distribution of $x + \Delta\mu(x, \alpha) + \sqrt{\Delta}\sigma(x, \beta)\mathbb{N}(0, 1) + (\Delta/2)(\sigma\sigma^\bullet)(x, \beta)[\mathbb{N}(0, 1)^2 - 1]$, where $\mathbb{N}(0, 1)$ is the standard normal random variate. The Milstein transition density was also obtained by Elerian (1998).⁴

We may also consider the quasi MLE with the mean and variance obtained from the Milstein approximation, which yields the conditional mean and variance of $X_{i\Delta} - X_{(i-1)\Delta}$ given $X_{(i-1)\Delta} = x$ respectively by

$$\begin{aligned} \mu_M(x, \alpha) &= \Delta\mu(x, \alpha), \\ \sigma_M^2(x, \beta) &= \Delta\sigma^2(x, \beta) + \frac{\Delta^2}{2}(\sigma\sigma^\bullet)^2(x, \beta). \end{aligned}$$

Therefore, we may use the corresponding normal density

$$p_{QM}(\Delta, x, y, \theta) = \frac{1}{\sqrt{2\pi\sigma_M^2(x, \beta)}} \exp \left[-\frac{(y - x - \mu_M(x, \alpha))^2}{2\sigma_M^2(x, \beta)} \right] \quad (14)$$

⁴The final expression of the Milstein transition density in Elerian (1998) is slightly different from ours in (13), though they are identical.

for the quasi MLE based on the Milstein approximation. Compared with the Euler approximation in (11), we have an additional higher order correction term $\Delta^2(\sigma\sigma^\cdot)^2(x, \beta)/2$ for the variance in the approximated normal transition density.

Our subsequent asymptotic theory is also applicable for the closed-form MLE proposed by Ait-Sahalia (2002), which approximates the transition density based on the Lamperti transformation and the Hermite expansion. The method uses the transformation $\tau(x, \beta) = \int_w^x dy/\sigma(y, \beta)$ for some $w \in \mathcal{D}$ to define $X_t^* = \tau(X_t, \beta)$, so that the transformed process X^* satisfies the SDE $dX_t^* = \nu(X_t^*, \theta)dt + dW_t$ with

$$\nu(x, \theta) = \frac{\mu(\tau^{-1}(x, \beta), \alpha)}{\sigma(\tau^{-1}(x, \beta), \beta)} - \frac{1}{2}\sigma^\cdot(\tau^{-1}(x, \beta), \beta). \quad (15)$$

If we denote by p^* and p^{**} the densities of the transitions $X_{(i-1)\Delta}^* \mapsto X_{i\Delta}^*$ and $X_{(i-1)\Delta}^* \mapsto X_{i\Delta}^{**} = \Delta^{-1/2}(X_{i\Delta}^* - X_{(i-1)\Delta}^*)$ respectively, it follows that

$$\begin{aligned} p(\Delta, x, y, \theta) &= \frac{1}{\sigma(y, \beta)} p^*(\Delta, \tau(x, \beta), \tau(y, \beta), \theta) \\ p^*(\Delta, x, y, \theta) &= \Delta^{-1/2} p^{**}(\Delta, x, \Delta^{-1/2}(y - x), \theta). \end{aligned}$$

Note that X is transformed and normalized appropriately for the transition $X_{(i-1)\Delta}^* \mapsto X_{i\Delta}^{**}$, so it has density close to that of standard normal. Therefore, we may approximate p^{**} as

$$p^{**}(\Delta, x, y, \theta) \simeq p_J^{**}(\Delta, x, y, \theta) = \phi(y) \sum_{j=0}^J \eta_j(\Delta, x, \theta) H_j(y), \quad (16)$$

where ϕ is the standard normal density function and (H_j) are the Hermite polynomials, and (η_j) are coefficients obtained from the approximated conditional moments of the process X^* . Once we obtain the transition density p^{**} in a closed-form in this way, we may obtain the approximated transition density of the original process X as

$$p_{AS}(\Delta, x, y, \theta) = \frac{1}{\sqrt{\Delta}\sigma(y, \beta)} p_J^{**}(\Delta, \tau(x, \beta), \Delta^{-1/2}[\tau(y, \beta) - \tau(x, \beta)], \theta), \quad (17)$$

as we have shown above.

Kessler (1997) proposes the quasi MLE based on the normal transition density

$$p_{KS}(\Delta, x, y, \theta) = \frac{1}{\sqrt{2\pi\sigma_K^2(x, \theta)}} \exp \left[-\frac{(y - x - \mu_K(x, \theta))^2}{2\sigma_K^2(x, \theta)} \right] \quad (18)$$

using conditional mean and variance are approximated by

$$\mu_K(x, \theta) = \sum_{j=0}^J \frac{\Delta^j}{j!} L^j x$$

$$\sigma_K^2(x, \theta) = \Delta \sigma^2(x, \beta) \left(1 + \frac{1}{\Delta \sigma^2(x, \beta)} \sum_{j=2}^J \Delta^j \sum_{k=0}^{J-j} \frac{\Delta^k}{k!} L^k \left(\sum_{a,b \geq 1, a+b=j} \frac{L^a x}{a!} \frac{L^b x}{b!} \right) \right)$$

where L is the infinitesimal generator given by $Lf(x) = \mu(x, \alpha)Df(x) + (1/2)\sigma^2(x, \beta)D^2f(x)$ with the usual differential operator D . In practice, σ_K^2 can be negative or zero, which makes it impossible to obtain the log-likelihoods involving $\log(\sigma_K^2)$ and $1/\sigma_K^2$. To avoid this, he suggests to use its Taylor expansion in Δ up to order J .

Our theory also applies to the simulated MLE, which obtains the transition density of the process with simulations. Gihman and Skorohod (1972) show that the transition density of (X_t) can be written by

$$p(\Delta, x, y, \theta) = \sqrt{\frac{\sigma(x, \beta)}{2\pi\Delta\sigma^3(y, \beta)}} \exp \left[-\frac{1}{2\Delta} (\tau(y, \beta) - \tau(x, \beta))^2 + \int_x^y \frac{\mu(z, \alpha)}{\sigma^2(z, \beta)} dz \right]$$

$$\times \mathbb{E} \exp \left[\Delta \int_0^1 \omega \left((1-t)\tau(x, \beta) + t\tau(y, \beta) + \sqrt{\Delta} \tilde{W}_t, \theta \right) dt \right],$$

where \tilde{W} , $\tilde{W}_t = W_t - tW_1$, is Brownian bridge, $\tau(x, \beta)$ is the Lamperti transformation and $\omega(x, \theta) = -(1/2)(\nu^2(x, \theta) + \nu^*(x, \theta))$ with $\nu(x, \theta)$ defined in (15), provided in particular that $|\omega(x, \theta)| = O(x^2)$ as $x \rightarrow \infty$. The expectation part involving Brownian bridge can be obtained from the simulation with arbitrary precision for any given parameter value, so we may obtain the corresponding numerical transition density approximating the true transition density arbitrarily well. Of course, we may use the transition density to obtain the exact MLE even when there is no closed-form solution of the transition density. See Nicolau (2002) for more information on the actual implementation of this approach. On the other hand, utilizing the Chapman-Kolmogorov equation, Pedersen (1995) and Brandt and Santa-Clara (2002) suggest simulating the transition density with

$$p_N(\Delta, x, y, \theta) = \mathbb{E} \left[p^* \left(\frac{\Delta}{N}, X_{\Delta-\Delta/N}^*, y, \theta \right) \middle| X_0^* = x \right],$$

where p^* is an approximated transition density based on, for example, the Euler approximation, and X^* is the corresponding process generated with that approximation. They

show that p_N converges to the true transition density as $N \rightarrow \infty$, and therefore we may use it to obtain the exact ML estimation with arbitrary precision. See Durham and Gallant (2002) for some comparisons among various methods for simulating the unknown transition density.

3. Fundamentals of MLE Asymptotics

In this section, we develop some fundamental theories required to establish the general asymptotics for the MLE's in diffusion models. To more effectively present our asymptotic analysis, we define

$$\ell(\Delta, x, y, \theta) = \Delta \log \left[\sqrt{\Delta} p(\Delta, x, y, \theta) \right],$$

which is the standardized log likelihood function. We will consider various derivatives of the log likelihood function ℓ , as well as the drift and diffusion functions μ and σ . For $f = \ell, \mu$ or σ , we signify its partial derivative $\partial^{i+j+k+\ell} f / \partial a^i \partial b^j \partial c^k \partial d^\ell$ by $f_{a^i b^j c^k d^\ell}$, where (a, b, c, d) are the arguments of f and (i, j, k, ℓ) is any sets of positive integers. Lastly, for any of the derivatives of ℓ that has Δ as one of its argument, say, $f(\Delta, x, y)$, we define $f(0, x, x)$ to be its Δ -limit, i.e., $f(0, x, x) = \lim_{\Delta \rightarrow 0} f(\Delta, x, x)$. Of course, we assume that the Δ -limit of $f(\Delta, x, y)$ exists, whenever we have the expression $f(0, x, x)$ in what follows. Note that the transition density ℓ , and therefore its derivatives too, is meaningfully defined only for $\Delta > 0$. Our standardization of the log likelihood function in Δ ensures that the Δ -limit exists for ℓ and its derivatives. In presenting our asymptotics, we extend our earlier convention and use the notation “ \sim_p ” to denote the asymptotic equivalence in probability. More specifically, $P \sim_p Q$ implies that $P/Q \rightarrow_p 1$, or equivalently, $P - Q = o_p(Q)$.

3.1 Basic Framework and Continuous Approximations

Our asymptotics follow the approach by Wooldridge (1994) and Park and Phillips (2001). If we let $\mathcal{S} = \partial \mathcal{L} / \partial \theta$ and $\mathcal{H} = \partial^2 \mathcal{L} / \partial \theta \partial \theta'$, the asymptotic leading term of $\hat{\theta}$ can be obtained from the first order Taylor expansion of \mathcal{S} , i.e.,

$$\mathcal{S}(\hat{\theta}) = \mathcal{S}(\theta_0) + \mathcal{H}(\tilde{\theta})(\hat{\theta} - \theta_0), \quad (19)$$

where $\tilde{\theta}$ lies in the line segment connecting $\hat{\theta}$ and θ_0 . To derive our asymptotics, we will establish that

$$\text{AD1: } w^{-1}\mathcal{S}(\theta_0) \rightarrow_d N,$$

$$\text{AD2: } w^{-1}\mathcal{H}(\theta_0)w^{-1'} \rightarrow_d M \text{ for some } M \text{ positive definite a.s., and}$$

$$\text{AD3: There is a sequence } v \text{ such that } vw^{-1} \rightarrow 0, \text{ and}$$

$$\sup_{\theta \in \mathcal{N}} |v^{-1}(\mathcal{H}(\theta) - \mathcal{H}(\theta_0))v^{-1'}| \rightarrow_p 0,$$

$$\text{where } \mathcal{N} = \{\theta : |v'(\theta - \theta_0)| \leq 1\},$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$ at appropriate rates for some matrix sequences w and v , and random vector and matrix N and M . Note that w and v are functions of T and Δ , which we suppress for notational simplicity.

As shown in Wooldridge (1994), AD3 together with AD1 and AD2 implies that $\mathcal{S}(\hat{\theta}) = 0$ with probability approaching one and $w^{-1}(\mathcal{H}(\tilde{\theta}) - \mathcal{H}(\theta_0))w^{-1'} = o_p(1)$, as $T \rightarrow \infty$ and $\Delta \rightarrow 0$ at appropriate rates.⁵ We may therefore easily deduce from the first order Taylor expansion (19) that

$$w'(\hat{\theta} - \theta_0) = -[w^{-1}\mathcal{H}(\theta_0)w^{-1'}]^{-1}w^{-1}\mathcal{S}(\theta_0) + o_p(1) \rightarrow_d M^{-1}N \quad (20)$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$ respectively at an appropriate rate. Therefore, once we establish AD3, we only need to find the limit behaviors of the score $\mathcal{S}(\theta_0)$ and Hessian $\mathcal{H}(\theta_0)$. The asymptotics of the MLE would then follow immediately from (20). The subsequent developments of our asymptotic theory will therefore be focused on the analysis of limit behaviors of $\mathcal{S}(\theta_0)$ and $\mathcal{H}(\theta_0)$ and on the establishment of condition in AD3.

To develop our asymptotics more effectively, we introduce functional operators \mathcal{A} and \mathcal{B} that are defined as

$$\mathcal{A}f(t, x, y) = f_t(t, x, y) + \mu(y)f_y(t, x, y) + \frac{1}{2}\sigma^2(y)f_{y^2}(t, x, y)$$

$$\mathcal{B}f(t, x, y) = \sigma(y)f_y(t, x, y)$$

⁵This is shown in Wooldridge (1994) within the usual asymptotic framework relying only on the sample size n . However, it is clear that his argument is also applicable in our context as long as there are proper normalizing sequences w and v .

for f with its derivatives $f_t = \partial f / \partial t$, $f_y = \partial f / \partial y$ and $f_{y^2} = \partial^2 f / \partial y^2$ assumed to exist, and write

$$f(t-s, X_s, X_t) - f(0, X_s, X_s) = \int_s^t \mathcal{A}f(t-s, X_s, X_r) dr + \int_s^t \mathcal{B}f(t-s, X_s, X_r) dW_r. \quad (21)$$

If necessary, we further expand the terms $\mathcal{A}f(t-s, X_s, X_r)$ and $\mathcal{B}f(t-s, X_s, X_r)$ in a similar fashion to obtain

$$\begin{aligned} \mathcal{A}f(t-s, X_s, X_t) - \mathcal{A}f(0, X_s, X_s) &= \int_s^t \mathcal{A}^2 f(r-s, X_s, X_r) dr + \int_s^t \mathcal{B} \mathcal{A}f(r-s, X_s, X_r) dW_r, \\ \mathcal{B}f(t-s, X_s, X_t) - \mathcal{B}f(0, X_s, X_s) &= \int_s^t \mathcal{A} \mathcal{B}f(r-s, X_s, X_r) dr + \int_s^t \mathcal{B}^2 f(r-s, X_s, X_r) dW_r. \end{aligned}$$

Clearly, we may repeatedly apply the procedure to obtain expansions to any arbitrary order.

To obtain the asymptotic leading terms of $\mathcal{S}(\theta_0)$ and $\mathcal{H}(\theta_0)$, we write

$$\mathcal{S}(\theta_0) = \sum_{i=1}^n \ell_\theta(\Delta, X_{(i-1)\Delta}, X_{i\Delta}), \quad \mathcal{H}(\theta_0) = \sum_{i=1}^n \ell_{\theta\theta'}(\Delta, X_{(i-1)\Delta}, X_{i\Delta}), \quad (22)$$

and expand them using (21). If we denote by f any element of the terms in the expansion and assume that it is differentiable, then we have

$$\Delta \sum_{i=1}^n f(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) = \Delta \sum_{i=1}^n f(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) + R_A + R_B \quad (23)$$

with

$$\begin{aligned} R_A &= \Delta \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \mathcal{A}f(r-(i-1)\Delta, X_{(i-1)\Delta}, X_r) dr \\ R_B &= \Delta \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \mathcal{B}f(r-(i-1)\Delta, X_{(i-1)\Delta}, X_r) dW_r, \end{aligned}$$

where R_A and R_B are remainder terms which become negligible asymptotically.

To develop the expansion above formally and rigorously, we need to introduce some technical assumptions. For the convenience of exposition, we momentarily assume that the boundaries \underline{x} or \bar{x} is either $\pm\infty$ or 0. This causes no loss in generality, since we may simply consider $X - \underline{x}$ or $X - \bar{x}$ for more general case.

Assumption 3.1 We assume that (a) $\sigma^2(x, \beta) > 0$, (b) $\mu(x, \alpha)$, $\sigma^2(x, \beta)$ and $\ell(t, x, y, \theta)$ are infinitely differentiable in $t \geq 0$, $x, y \in \mathcal{D}$ and θ in the interior of Θ , and that for any $f(t, x, y, \theta)$ of their derivatives we have $|f(t, x, y, \theta)| \leq g(x)g(y)$ for all $t \geq 0$ small, for all $x, y \in \mathcal{D}$ and for all θ in the interior of Θ , where $g : \mathcal{D} \rightarrow \mathbb{R}$ is locally bounded and $|g(x)| \sim c|x|^p$ at boundaries $\pm\infty$ and $|g(x)| \sim c|x|^{-p}$ at boundary 0 for some constant $c > 0$, (c) $\sup_{0 \leq t \leq T} |X_t| = O_p(T^q)$ if the boundaries are $\pm\infty$ and $(\inf_{0 \leq t \leq T} |X_t|)^{-1} = O_p(T^q)$ if one of the boundaries is 0, and (d) $\Delta T^{4(pq+1)} \rightarrow 0$ as $T \rightarrow \infty$ and $\Delta \rightarrow 0$.

The condition in Assumption 3.1(a) and the differentiability of the drift and diffusion functions in Assumption 3.1(b) are routinely assumed in the study of diffusion models. In particular, they are sufficient for the existence of a weak solution of the SDE (1) up to an explosion time that is unique in probability law. See, e.g., Theorem 5.5.15 of Karatzas and Shreve (1991). In Assumption 3.1(b), we additionally require the existence of an envelop function for all the derivatives of $\mu(x, \alpha)$, $\sigma^2(x, \beta)$ and $\ell(t, x, y, \theta)$ so that we may effectively control them especially near the boundaries. In Assumption 3.1(c), we set the growing and diminishing rates of the underlying diffusion process. We may obtain the rates from the asymptotic behavior of extremal process we discussed earlier. Assumption 3.1(d) makes it explicit that our asymptotics in the paper are derived under the condition $T \rightarrow \infty$ and $\Delta \rightarrow 0$. In particular, the condition requires that Δ decreases fast enough as T increases. Our asymptotic results will therefore be more relevant for the case where Δ is sufficiently small relative to T . Indeed, this is the case in many practical applications of diffusions models, which rely on samples collected at relatively high frequencies over short or moderate lengths of time spans, such as daily observations over a few years.

Now we are ready to deal with the summations in (22), but before that, we introduce the following lemma which is useful to obtain the leading terms in our asymptotics explicitly in terms of μ and σ .

Lemma 3.1 Let ℓ be the normalized log-likelihood for the transition density of (X_t) obtained by using any of the methods introduced in Section 2.2. Then under Assumptions

3.1(a), (b), we have for all $x \in \mathcal{D}$ and θ in the interior of Θ ,

$$\begin{aligned}\ell(0, x, x, \theta) &= 0, & \mathcal{A}\ell(0, x, x, \theta) &= -\frac{\sigma^2(x)}{2\sigma^2(x, \beta)} - \log(\sigma(x, \beta)), \\ \mathcal{B}\ell(0, x, x, \theta) &= 0, & \mathcal{B}^2\ell(0, x, x, \theta) &= -\frac{\sigma^2(x)}{\sigma^2(x, \beta)}\end{aligned}$$

ignoring the terms which do not depend upon θ , and

$$\begin{aligned}\mathcal{A}^2\ell(0, x, x, \theta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + \mu(x)\frac{2\mu(x, \alpha)}{\sigma^2(x, \beta)} + (\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2}(0, x, x, \theta), \\ \mathcal{AB}\ell(0, x, x, \theta) &= \mathcal{BA}\ell(0, x, x, \theta) = \sigma(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)}, \\ \mathcal{B}^3\ell(0, x, x, \theta) &= 0\end{aligned}$$

ignoring the terms which are independent of α .

We may obtain the asymptotics for the score and Hessian functions explicitly using

$$\int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t ds dW_t + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t dW_s dt = \Delta(W_{i\Delta} - W_{(i-1)\Delta})$$

and

$$\int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t dW_s dW_t = \frac{1}{2} [(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta]$$

and Lemma 3.1. For the score of the drift term parameter, we have

$$\begin{aligned}\mathcal{S}_\alpha(\theta_0) &= \frac{1}{\Delta} \sum_{i=1}^n \ell_\alpha(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \\ &\simeq \frac{1}{2} \sum_{i=1}^n (\mathcal{AB} + \mathcal{BA})\ell_\alpha(0, X_{(i-1)\Delta}, X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}),\end{aligned}\tag{24}$$

since $\ell_\alpha(0, x, x) = 0$, $\mathcal{A}\ell_\alpha(0, x, x) = 0$, $\mathcal{B}\ell_\alpha(0, x, x) = 0$, $\mathcal{B}^2\ell_\alpha(0, x, x) = 0$, $\mathcal{B}^3\ell_\alpha(0, x, x) = 0$, and $\mathcal{A}^2\ell_\alpha(0, x, x) = 0$ due to Lemma 3.1. For the score of the diffusion term parameter, we have

$$\begin{aligned}\mathcal{S}_\beta(\theta_0) &= \frac{1}{\Delta} \sum_{i=1}^n \ell_\beta(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \\ &\simeq \frac{1}{2\Delta} \sum_{i=1}^n \mathcal{B}^2\ell_\beta(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) [(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta],\end{aligned}\tag{25}$$

since it follows from Lemma 3.1 that $\ell_\beta(0, x, x) = 0$, $\mathcal{A}\ell_\beta(0, x, x) = 0$ and $\mathcal{B}\ell_\beta(0, x, x) = 0$.

We may similarly analyze the Hessian. For the Hessian of the drift term parameter, we may obtain

$$\begin{aligned}\mathcal{H}_{\alpha\alpha}(\theta_0) &= \frac{1}{\Delta} \sum_{i=1}^n \ell_{\alpha\alpha'}(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \\ &\simeq \frac{\Delta}{2} \sum_{i=1}^n \mathcal{A}^2 \ell_{\alpha\alpha'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) \\ &\quad + \frac{1}{2} \sum_{i=1}^n (\mathcal{AB} + \mathcal{BA}) \ell_{\alpha\alpha'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) (W_{i\Delta} - W_{(i-1)\Delta}),\end{aligned}\quad (26)$$

since we have from Lemma 3.1 that $\ell_{\alpha\alpha'}(0, x, x) = 0$, $\mathcal{A}\ell_{\alpha\alpha'}(0, x, x) = 0$, $\mathcal{B}\ell_{\alpha\alpha'}(0, x, x) = 0$, $\mathcal{B}^2\ell_{\alpha\alpha'}(0, x, x) = 0$, $\mathcal{B}^3\ell_{\alpha\alpha'}(0, x, x) = 0$. Moreover, the Hessian of the diffusion term parameter reduces to

$$\mathcal{H}_{\beta\beta}(\theta_0) = \frac{1}{\Delta} \sum_{i=1}^n \ell_{\beta\beta'}(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) \simeq \sum_{i=1}^n \mathcal{A}\ell_{\beta\beta'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}), \quad (27)$$

since $\ell_{\beta\beta'}(0, x, x) = 0$ and $\mathcal{B}\ell_{\beta\beta'}(0, x, x) = 0$. The leading term of the off-diagonal block $\mathcal{H}_{\alpha\beta}(\theta_0)$ can be also shown to be negligible in the limit.

Lemma 3.2 Under Assumption 3.1, we have

$$\begin{aligned}\mathcal{S}_{\alpha}(\theta_0) &= \int_0^T \frac{\mu_{\alpha}}{\sigma}(X_t) dW_t + O_p(\sqrt{\Delta} T^{4pq+1}) \\ \mathcal{S}_{\beta}(\theta_0) &= \sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_{\beta}}{\sigma}(X_t) dV_t + O_p(\Delta^{-1/4} T^{4pq+7/4})\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}_{\alpha\alpha}(\theta_0) &= - \int_0^T \frac{\mu_{\alpha}\mu'_{\alpha}}{\sigma^2}(X_t) dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t) dW_t + O_p(\sqrt{\Delta} T^{4pq+1}) \\ \mathcal{H}_{\beta\beta}(\theta_0) &= - \frac{2}{\Delta} \int_0^T \frac{\sigma_{\beta}\sigma'_{\beta}}{\sigma^2}(X_t) dt + O_p(\Delta^{-1/2} T^{3pq+1}) \\ \mathcal{H}_{\alpha\beta}(\theta_0) &= O_p(T^{3pq+1})\end{aligned}\quad (28)$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$.

For the asymptotics of the diffusion term parameter β , we only need the first set of results in Lemma 3.1, while for the asymptotics of the drift term parameter α , both the first and second sets of the results in Lemma 3.1 are required.

3.2 Preliminary Continuous Time Asymptotics

Now we establish primary asymptotics for continuous time processes

$$\int_0^T f(X_t)dt \quad \text{and} \quad \int_0^T g(X_t)dW_t \quad (29)$$

as $T \rightarrow \infty$ for some classes of functions $f, g : \mathcal{D} \rightarrow \mathbb{R}^k$. The asymptotics of two continuous time processes in (29) will be referred to as the additive functional asymptotics and the martingale transform asymptotics, respectively, in the paper. For the development of these asymptotics, it will be convenient to introduce

Definition 3.1 We say that f is m -integrable and g is m -square integrable, respectively, if f and $g \otimes g$ are integrable with respect to the speed measure m .

Under our notational convention of using m to denote both the speed measure and its density with respect to the Lebesgue measure, f is m -integrable and g is m -square integrable if and only if mf and $m(g \otimes g)$ are integrable respectively with respect to the Lebesgue measure. We will simply call f integrable and g is square integrable, if f and $g \otimes g$ are integrable with respect to the Lebesgue measure.

For the positive recurrent process X , the continuous asymptotics in (29) is well known, which we give below for the future reference. Recall that we have $m(\mathcal{D}) < \infty$ and the time invariant marginal distribution is given by $\pi = m/m(\mathcal{D})$ for the positive recurrent process. Needless to say, $\pi(f) < \infty$ and $\pi(g \otimes g) < \infty$, if and only if f is m -integrable and g is m -square integrable in this case.

Proposition 3.3 Let Assumption 3.1 hold. If X is positive recurrent and f and g are respectively m -integrable and m -square integrable, then we have

$$\frac{1}{T} \int_0^T f(X_t)dt \rightarrow_{a.s.} \pi(f), \quad \frac{1}{\sqrt{T}} \int_0^T g(X_t)dW_t \rightarrow_d \mathbb{N}(0, \pi(gg'))$$

as $T \rightarrow \infty$.

For positive recurrent processes, both the additive functional and martingale transform asymptotics therefore yield the usual normal limit distributions. Moreover, we need the standard normalizing sequences T and \sqrt{T} , respectively, for their asymptotics.

The additive functional and martingale transform asymptotics for null recurrent processes are little known, and we will fully develop them below. For the null recurrent diffusion X , we consider the transformed process X^s , $X_t^s = s(X_t)$, where s is the scale function. As is well known, X^s becomes a driftless diffusion that is given by

$$dX_t^s = \sigma_s(X_t^s) dW_t, \quad (30)$$

where $\sigma_s = (s^* \sigma) \circ s^{-1}$. Therefore, X^s is in natural scale. The speed measure of X^s is given by the density m_r , $m_r(x) = 1/\sigma_s^2(x)$. For the development of our asymptotics, it is convenient to write

$$f(X_t) = f_s(X_t^s) \quad \text{and} \quad g(X_t) = g_s(X_t^s),$$

where $f_s = f \circ s^{-1}$ and $g_s = g \circ s^{-1}$. Note that f_s and g_s are defined over the entire range of \mathbb{R} for all recurrent processes. The notations f_s and g_s will be used frequently in what follows.

It is well known that f_s and $g_s \otimes g_s$ are integrable with respect to the measure m_r on \mathbb{R} if and only if they are integrable with respect to the measure m on \mathcal{D} , and we have $m_r(f_s) = m(f)$ and $m_r(g_s \otimes g_s) = m(g \otimes g)$. In particular, the speed density m_r of a null recurrent diffusion in natural scale is not integrable on \mathbb{R} , since $m_r(\mathbb{R}) = m(\mathcal{D}) = \infty$. To effectively deal with null recurrent diffusions, we define

Definition 3.2 A null recurrent process is said to be *regular with index $r > -1$* if for its speed density m_r in natural scale, we have

$$m_r(x) = m_r^*(x) + \varepsilon_r(x)$$

where m_r^* is a homogeneous function of degree $r > -1$, and ε_r is a locally integrable function such that $\varepsilon_r(x) = o(|x|^r)$ as $|x| \rightarrow \infty$.

The regularity conditions we introduce in Definition 3.2 are not very stringent and allows for a wide class of non-integrable m_r including all speed densities in natural scale we consider in the examples with appropriate restrictions on their parameter values. For a regular null

recurrent process with index $r > -1$, we have

$$m_r(x)/|x|^r \rightarrow a \text{ or } b \quad (31)$$

respectively as $x \rightarrow \pm\infty$ and $a + b > 0$.

The basic asymptotics for null recurrent processes are given below. We assume that the scale transform has been performed and consider the process X^s in natural scale.

Proposition 3.4 Let X be a regular null recurrent process with index $r > -1$ having speed density m_r in natural scale and driven by Brownian motion W , and define the processes X^{sT} on $[0, 1]$ for each T by $X_t^{sT} = T^{-1/(r+2)} X_{Tt}^s$. Then we have

$$X^{sT} \rightarrow_d X^\circ \quad (32)$$

as $T \rightarrow \infty$ in the space $\mathcal{C}[0, 1]$ of continuous functions defined on $[0, 1]$. Here X° is defined by $X^\circ = B \circ \tau^r$ with

$$\tau_t^r = \inf \left\{ s \left| \int_{\mathbb{R}} m_r^*(x) l(s, x) dx > t \right. \right\}$$

for $0 \leq t \leq 1$, where B is a standard Brownian motion, and l is the local time of B . Moreover, for W^T defined by $W_t^T = T^{-1/2} W_{Tt}$, we have

$$W^T \rightarrow_d W^\circ \quad (33)$$

jointly with (32) in $\mathcal{C}[0, 1]$ as $T \rightarrow \infty$, where W° is a standard Brownian motion given by

$$W_t^\circ = \int_0^t m_r^{*1/2}(X_s^\circ) dX_s^\circ$$

for $0 \leq t \leq 1$.

The limit process X° is defined as a time change of the Brownian motion B with time change τ^r given by the right continuous inverse of $\int_{\mathbb{R}} m_r^*(x) l(\cdot, x) dx$, where l is the local time of B . The stochastic processes defined in this way are called generalized diffusion processes corresponding to the speed density m_r^* . The reader is referred to Kotani and Watanabe (1982) or Itô and McKean (1996) for the details of this class of processes. In particular,

for the speed density m_r^* in Definition 3.2 together with the asymptotes in (31), the limit process X° scaled as

$$\left[\left(\frac{a}{r+1} \right)^{\frac{1}{r+2}} + \left(\frac{b}{r+1} \right)^{\frac{1}{r+2}} \right] X^\circ$$

becomes a skew Bessel process in natural scale of dimension $2(r+1)/(r+2)$ with the skew parameter $a^{1/(r+2)}/(a^{1/(r+2)} + b^{1/(r+2)})$. Note that $0 < 2(r+1)/(r+2) < 2$ if $r > -1$. For the construction and the properties of the skew Bessel process in natural scale, the reader is referred to Watanabe (1995, pp. 160, 164). We call the process a symmetric Bessel process in natural scale if the skew parameter is $1/2$. In case that $b = 0$, the process reduces to a Bessel process in natural scale. Moreover, if $r = 0$, we have $m_r^*(x) = a 1\{x \geq 0\} + b 1\{x < 0\}$, and the limit process X° becomes a skew Brownian motion in natural scale.

Definition 3.3 We say that f is *m-asymptotically homogeneous* if

$$f_s(\lambda x) = \kappa(f_s, \lambda) h(f_s, x) + \delta(f_s, \lambda, x)$$

with

$$|\delta(f_s, \lambda, x)| \leq a(f_s, \lambda) p(f_s, x) + b(f_s, \lambda) q(f_s, \lambda x)$$

as $\lambda \rightarrow \infty$, where (i) $h(f_s, \cdot)$, $p(f_s, \cdot)$ and $q(f_s, \cdot)$ are locally integrable in measures m_r and m_r^* , (ii) $\kappa(f_s, \lambda)$ is nonsingular for all large λ , (iii) $q(f_s, \cdot)$ is locally bounded on $\mathbb{R} \setminus \{0\}$ and vanishing at infinity, and (iv)

$$\limsup_{\lambda \rightarrow \infty} \|\kappa(f_s, \lambda)^{-1} a(f_s, \lambda)\| = 0, \quad \limsup_{\lambda \rightarrow \infty} \|\kappa(f_s, \lambda)^{-1} b(f_s, \lambda)\| < \infty.$$

We call $\kappa(f_s, \cdot)$ and $h(f_s, \cdot)$ respectively the *asymptotic order* and *limit homogeneous function* of f . If we have (i)' $h(f_s, \cdot)$, $p(f_s, \cdot)$ and $q(f_s, \cdot)$ are locally square integrable in measures m_r and m_r^* , in place of (i), then f is said to be *m-square asymptotically homogeneous*.

In particular, we require *m-asymptotically homogeneous* or *m-square asymptotically homogeneous function* f to be given roughly as

$$f_s(\lambda x) \sim \kappa(f_s, \lambda) h(f_s, x)$$

for large λ , where the limit homogeneous function $h(f_s, \cdot)$ of f is integrable or square integrable in both m_r and m_r^* over any compact set containing the origin.

The concept of m -asymptotic homogeneity is closely related to the notion of regular variation. For simplicity, we assume that the underlying diffusion is in natural scale so that the scale function is an identity and that m is a scaled Lebesgue measure, and call a function satisfying the required conditions asymptotically homogeneous instead of m -asymptotically homogeneous. In this case, a function f regularly varying with index $r > -1$ symmetrically at $\pm\infty$ is asymptotically homogeneous with asymptotic order $\kappa(f, \lambda) = f(\lambda)$ and limit homogeneous function $h(f, x) = |x|^r$. Of course, we have regularly varying functions that are not symmetric and have different growth rates at $\pm\infty$, in which case κ and h are determined by the dominating side of $\pm\infty$. The reader is referred to Bingham, Goldie and Teugels (1993) for more details of the regularly varying functions. The main motivation of introducing a new concept here is to extend the notation of regular variation to vector-valued functions. As an example, the vector-valued function $f(x) = (|x|, |x| \log |x|)'$ is asymptotically homogeneous with

$$\kappa(f, \lambda) = \begin{pmatrix} \lambda & 0 \\ \lambda \log \lambda & \lambda \end{pmatrix} \quad \text{and} \quad h(f, x) = \begin{pmatrix} |x| \\ |x| \log |x| \end{pmatrix}.$$

A regularly varying function cannot be asymptotically homogeneous with limit homogeneous function $|x| \log |x|$.

The continuous time asymptotics for the functionals of null recurrent processes may now be readily derived by applying the results in Höpfner and Löcherbach (2003) and Proposition 3.4 to the null recurrent process X^s in natural scale. For null recurrent processes, we consider both classes of integrable and asymptotically homogeneous functions in the sense of Definitions 3.1 and 3.3. Of course, there are functions that are neither integrable nor asymptotically homogeneous in our sense. However, virtually all functions involved in diffusion models that are used in practical applications belong to one of these two function classes.

Theorem 3.5 Let Assumption 3.1 hold and assume that (X_t) is null recurrent and regular with index $r > -1$.

(a) If f is m -integrable and g is m -square integrable, then we have

$$\begin{aligned} \frac{1}{T^{1/(r+2)}} \int_0^T f(X_t) dt &\rightarrow_d Km(f) A^{1/(r+2)} \\ \frac{1}{\sqrt{T^{1/(r+2)}}} \int_0^T g(X_t) dW_t &\rightarrow_d \sqrt{K} m(gg')^{1/2} B \circ A^{1/(r+2)}, \end{aligned}$$

jointly as $T \rightarrow \infty$, where $A^{1/(r+2)}$ is the Mittag-Leffler process with index $1/(r+2)$ at time 1, and B is standard vector Brownian motion independent of $A^{1/(r+2)}$, and

$$K = \frac{\Gamma((r+1)/(r+2))}{\Gamma((r+3)/(r+2))} \frac{(r+2)^{2/(r+2)}}{(a^{1/(r+2)} + b^{1/(r+2)})},$$

where a and b are from (31).

(b) If f is m -asymptotically homogeneous and g is m -square asymptotically homogeneous, then we have

$$\begin{aligned} \frac{1}{T} \kappa(f_s, T^{1/(r+2)})^{-1} \int_0^T f(X_t) dt &\rightarrow_d \int_0^1 h(f_s, X_t^\circ) dt \\ \frac{1}{\sqrt{T}} \kappa(g_s, T^{1/(r+2)})^{-1} \int_0^T g(X_t) dW_t &\rightarrow_d \int_0^1 h(g_s, X_t^\circ) dW_t^\circ \end{aligned}$$

jointly as $T \rightarrow \infty$ in notations defined in Definition 3.3 and Proposition 3.4.

For the details of the Mittag-Leffler process, readers are referred to Bingham (1971) or Höpfner (1990).

The asymptotics for the leading terms of $\mathcal{S}(\theta_0)$ and $\mathcal{H}(\theta_0)$ we present in Lemma 3.2 may be readily derived using our continuous time asymptotics. To obtain the proper asymptotics, we need to assume

Assumption 3.2 We assume that there exist nonsingular sequences $w_\alpha(T)$ and $w_\beta(T)$ such that $w_\alpha(T), w_\beta(T) \rightarrow \infty$,

$$\begin{aligned} w_\alpha^{-1}(T) \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t &\rightarrow_d N_\alpha, \quad w_\beta^{-1}(T) \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \rightarrow_d N_\beta \\ w_\alpha^{-1}(T) \int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2}(X_t) dt w_\alpha^{-1}(T)' &\rightarrow_d M_\alpha, \quad w_\beta^{-1}(T) \int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t) dt w_\beta^{-1}(T)' \rightarrow_d M_\beta \end{aligned}$$

for some $M_\alpha, M_\beta > 0$ a.s. and N_α, N_β , and

$$(w_\alpha \otimes w_\alpha)^{-1}(T) \int_0^T \frac{\mu_{\alpha \otimes \alpha}}{\sigma}(X_t) dW_t \rightarrow_p 0$$

as $T \rightarrow \infty$.

Remark 3.1 The conditions in Assumption 3.2 are not stringent and expected to hold widely, as we discuss below. Denoting I_α and I_β as the identity matrices of the same dimension as α and β respectively,

(a) If (X_t) is positive recurrent, Assumption 3.2 is satisfied with $w_\alpha(T) = \sqrt{T}I_\alpha$ and $w_\beta(T) = \sqrt{T}I_\beta$ if μ_α/σ , σ_β/σ , and $\mu_{\alpha\otimes\alpha}/\sigma$ are m -square integrable, and $m[(\mu_\alpha\mu'_\alpha)/\sigma^2]$, $m[(\sigma_\beta\sigma'_\beta)/\sigma^2] > 0$.

(b) If (X_t) is null recurrent and regular with index $r > -1$, Assumption 3.2 is satisfied with $w_\alpha(T) = \sqrt{T^{1/(r+2)}}I_\alpha$ and $w_\beta(T) = \sqrt{T^{1/(r+2)}}I_\beta$ if μ_α/σ , σ_β/σ , and $\mu_{\alpha\otimes\alpha}/\sigma$ are m -square integrable, and $m[(\mu_\alpha\mu'_\alpha)/\sigma^2]$, $m[(\sigma_\beta\sigma'_\beta)/\sigma^2] > 0$.

(c) Let (X_t) be null recurrent and regular with index $r > -1$, and let $\nu_\alpha = (\mu_\alpha/\sigma) \circ s^{-1}$, $\tau_\beta = (\sigma_\beta/\sigma) \circ s^{-1}$ and $\varpi_\alpha = (\mu_{\alpha\otimes\alpha}/\sigma) \circ s^{-1}$ be m -square asymptotically homogeneous with

$$\int_{|x|\leq\delta} hh'(\nu_\alpha, x)dx, \int_{|x|\leq\delta} hh'(\tau_\beta, x)dx > 0$$

for any $\delta > 0$. Furthermore, let

$$T^{-1/2}(\kappa \otimes \kappa)^{-1}(\nu_\alpha, T^{1/(r+2)})\kappa(\varpi_\alpha, T^{1/(r+2)}) \rightarrow 0$$

as $T \rightarrow \infty$. Then Assumption 3.2 is satisfied with $w_\alpha(T) = \sqrt{T}\kappa(\nu_\alpha, T^{1/(r+2)})$ and $w_\beta(T) = \sqrt{T}\kappa(\tau_\beta, T^{1/(r+2)})$.

If we let $w = \text{diag}(w_\alpha(T), \Delta^{-1/2}w_\beta(T))$, it follows straightforwardly that

Lemma 3.6 Under Assumptions 3.1 and 3.2, we have

$$\begin{aligned} w^{-1}\mathcal{S}(\theta_0) &\sim_p w^{-1} \left(\int_0^T \frac{\mu'_\alpha}{\sigma}(X_t)dW_t, \sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma'_\beta}{\sigma}(X_t)dV_t \right)' \\ w^{-1}\mathcal{H}(\theta_0)w^{-1'} &\sim_p w^{-1} \text{diag} \left(- \int_0^T \frac{\mu_\alpha\mu'_\alpha}{\sigma^2}(X_t)dt, - \frac{2}{\Delta} \int_0^T \frac{\sigma_\beta\sigma'_\beta}{\sigma^2}(X_t)dt \right) w^{-1'} \end{aligned} \quad (34)$$

for small Δ and large T .

Now we have shown in Lemma 3.6 that AD1 and AD2 hold, and it suffices to establish AD3 to derive the asymptotics of the MLE using (20). For AD3, we require

Assumption 3.3 If we let

$$\begin{aligned} f(x, \theta) &= \mu(x) \left(\frac{\mu_{\alpha \otimes \alpha \otimes \alpha}}{\sigma^2} \right) (x, \theta) \\ &\quad - \left(\frac{\mu \mu_{\alpha \otimes \alpha \otimes \alpha} + \mu_{\alpha \otimes \alpha} \otimes \mu_{\alpha} + \mu_{\alpha} \otimes \mu_{\alpha \otimes \alpha} + (I_{\alpha} \otimes C_{\alpha})(\mu_{\alpha \otimes \alpha} \otimes \mu_{\alpha})}{\sigma^2} \right) (x, \theta) \\ g(x, \theta) &= \sigma(x) \left(\frac{\mu_{\alpha \otimes \alpha \otimes \alpha}}{\sigma^2} \right) (x, \theta), \end{aligned} \quad (35)$$

there exists $\varepsilon > 0$ such that

$$\begin{aligned} T^{\varepsilon} (w_{\alpha} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} (T) \sup_{\theta \in \mathcal{N}} \left| \int_0^T f(X_t, \theta) dt \right| &\rightarrow_p 0 \\ T^{\varepsilon} (w_{\alpha} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} (T) \sup_{\theta \in \mathcal{N}} \left| \int_0^T g(X_t, \theta) dW_t \right| &\rightarrow_p 0 \end{aligned}$$

as $T \rightarrow \infty$, where \mathcal{N} is defined as in AD3.

Here we denote I_{α} and C_{α} as the identity matrix and the commutation matrix for square matrices, respectively, of the same dimension as α . Following lemma is useful to check Assumption 3.3.

Lemma 3.7 Let f and g be defined in (35) and denote d as the dimension of θ .

(a) Let X be positive recurrent and denote $\mathcal{N}_{\varepsilon} = \{\theta : \|\theta - \theta_0\| \leq T^{-1/2+\varepsilon}\}$. If there exist p and q such that

$$\|f(x, \theta)\| \leq p(x), \quad \|g(x, \theta_1) - g(x, \theta_2)\| \leq q(x) \|\theta_1 - \theta_2\| \quad (36)$$

for all $x \in \mathcal{D}$ and $\theta, \theta_1, \theta_2 \in \mathcal{N}_{\varepsilon}$ for all large T , and p and $q^{d+\varepsilon}$ are m -integrable for some $\varepsilon > 0$, then Assumption 3.3 is satisfied.

(b) Let X be null recurrent and regular with index $r > -1$, and denote $\mathcal{N}_{\varepsilon} = \{\theta : \|\theta - \theta_0\| \leq T^{-1/[2(r+2)]+\varepsilon}\}$. If for some $\varepsilon > 0$ there exist p and q such that (36) holds, and p and q are m -integrable and m -square integrable respectively, then Assumption 3.3 is satisfied.

(c) Let X be null recurrent and regular with degree $r > -1$, and denote $\mathcal{N}_{\varepsilon} = \{\theta : \|\text{diag}[\kappa'(\nu_{\alpha}, T^{1/(r+2)}), \kappa'(\tau_{\beta}, T^{1/(r+2)})](\theta - \theta_0)\| \leq T^{-1/2+\varepsilon}\}$. Also let $\nu_{\alpha} = (\mu_{\alpha}/\sigma) \circ s^{-1}$ and $\tau_{\beta} = (\sigma_{\beta}/\sigma) \circ s^{-1}$ be m -square asymptotically homogeneous, and suppose that (36) holds

with m -asymptotically homogeneous p and m -square asymptotically homogeneous q , such that

$$\begin{aligned} \|T^{-1/2+\varepsilon}(\kappa \otimes \kappa \otimes \kappa)^{-1}(\nu_\alpha, T^{1/(r+2)})\kappa(p, T^{1/(r+2)})\| &\rightarrow 0 \\ \|T^{-1+\varepsilon}(\kappa \otimes \kappa \otimes \kappa)^{-1}(\nu_\alpha, T^{1/(r+2)})\kappa(q, T^{1/(r+2)})\| &\rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ for some $\varepsilon > 0$. Then Assumption 3.3 is satisfied.

With Assumption 3.3, we can derive

Lemma 3.8 Under Assumptions 3.1-3.3, AD3 holds.

Now we have shown all conditions AD1-AD3 hold under Assumptions 3.1-3.3, and we are ready to establish the asymptotics of the MLE's in general diffusion models.

4. Asymptotic Theory of MLE

4.1 Primary Asymptotics

From the results we obtained in the previous section, it is rather straightforward to have

Theorem 4.1 Under Assumptions 3.1-3.3, we have

$$\begin{aligned} \hat{\alpha} - \alpha_0 &\sim_p \left(\int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2}(X_t) dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t \\ \hat{\beta} - \beta_0 &\sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t) dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \end{aligned}$$

for all small Δ and large T , where V is standard Brownian motion independent of W .

Theorem 4.1 provides the primary asymptotics for the exact and quasi MLE's of diffusion model parameters considered in the paper. They are obtained in particular under Assumption 3.1(d). Therefore, in particular, if T is large and Δ is small sufficiently to satisfy Assumption 3.1(d), we may expect that our primary asymptotics would well approximate the finite sample distributions of the exact, quasi and approximate MLE's in diffusion models. It should be noted that we do not assume $T = \infty$ here. As we will show below, the

standard limit distributions can be obtained straightforwardly by taking T -limits in our primary asymptotics. The standard limit distributions are often little useful, since many practical applications use samples collected at high frequency intervals over only moderately long time spans, such as daily observations spanning a few years of time. We believe that the distributions given by our primary asymptotics are in general much more accurate approximations of the relevant finite sample distributions. This is well demonstrated through simulations in the following section.

Our primary asymptotics reveal many of the important statistical properties of the exact, quasi and approximate MLE in diffusion models. First, the MLE $\hat{\alpha}$ for the drift term parameter and the MLE $\hat{\beta}$ for the diffusion term parameter are uncorrelated for all large T if Δ is sufficiently small relative to T . This, of course, implies that $\hat{\alpha}$ and $\hat{\beta}$ become asymptotically independent if they have jointly normal limit distributions, as will be the case for most nonstationary as well as stationary diffusions. Second, unless $T = \infty$, the distribution of $\hat{\alpha}$ is essentially non-Gaussian in all cases. For many diffusion models, the finite T distribution of $\hat{\alpha}$ is quite different from normal and $\hat{\alpha}$ has a somewhat serious bias problem, as we will show later by simulations. Third, on the other hand, the distribution of $\hat{\beta}$ is mixed normal even in finite T . Upon noticing that V is independent of W , and hence of X , we may indeed easily deduce that

$$\hat{\beta} \simeq_d \text{MN} \left(\beta_0, \frac{\Delta}{2} \left[\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2} (X_t) dt \right]^{-1} \right)$$

for large T and small Δ . Therefore, the finite T distribution of $\hat{\beta}$ is centered at the true value, and we may also expect that $\hat{\beta}$ does not suffer from any serious finite sample bias problem.

Now we discuss the consistency and derive the limit distributions of $\hat{\alpha}$ and $\hat{\beta}$. Under Assumption 3.3, we have in particular that

$$\int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2} (X_t) dt, \quad \frac{1}{\Delta} \int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2} (X_t) dt \rightarrow_p \infty \quad (37)$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$. Therefore, it can be easily deduced from Theorem 4.1 that

Corollary 4.2 Let Assumptions 3.1-3.3 hold. Then $\hat{\alpha}$ and $\hat{\beta}$ are consistent.

The conditions in (37) correspond to the well known minimal excitation condition for the classical regression model. They are expected to hold for a broad class of diffusion models, including virtually all diffusions that are used in practical applications. Note that $\Delta \rightarrow 0$ may be sufficient to satisfy the second condition, whereas for the first condition it is absolutely necessary that we have $T \rightarrow \infty$. This makes it clear that in general we need $T \rightarrow \infty$ for the consistency of the drift term parameter, though $\Delta \rightarrow 0$ is enough to get the consistency of the diffusion term parameter.

4.2 Limit Distributions

For a large class of diffusion models, we may obtain the exact convergence rates for the exact, quasi and approximate MLE's, and find their limit distributions. This will be shown below.

Theorem 4.3 Let Assumptions 3.1-3.3 hold. If X is positive recurrent, and μ_α/σ and σ_β/σ are m -square integrable, then we have

$$\sqrt{T}(\hat{\alpha} - \alpha_0) \rightarrow_d \mathbb{N}\left(0, \pi \left[\frac{\mu_\alpha \mu'_\alpha}{\sigma^2} \right]^{-1}\right), \quad \sqrt{\frac{T}{\Delta}}(\hat{\beta} - \beta_0) \rightarrow_d \mathbb{N}\left(0, \frac{1}{2} \pi \left[\frac{\sigma_\beta \sigma'_\beta}{\sigma^2} \right]^{-1}\right)$$

independently as $T \rightarrow \infty$ and $\Delta \rightarrow 0$.

For a majority of positive recurrent processes, we have the normal asymptotics. This is already well expected. Here we just use a different setting for the asymptotics, i.e., we let $T \rightarrow \infty$ and $\Delta \rightarrow 0$, whereas virtually all the existing literature assumes that $n = T/\Delta \rightarrow \infty$ with either Δ fixed or T fixed.⁶ In general, the convergence rates for the drift term and diffusion term parameters are given by \sqrt{T} and $\sqrt{T/\Delta}$, respectively, for positive recurrent diffusions. Note in particular that the convergence rate for the diffusion term parameter depends only on the sample size $n = T/\Delta$.

Theorem 4.4 Let Assumptions 3.1-3.3 hold. Moreover, let X be null recurrent and regular with index $r > -1$, and assume that μ_α/σ and σ_β/σ are m -square integrable. Then

⁶Kessler (1997) is the only exception. Indeed, he obtains the same asymptotics as ours for the positive recurrent diffusion models with scalar parameters in the drift and diffusion functions.

we have

$$\begin{aligned}\sqrt{T^{1/(r+2)}}(\hat{\alpha} - \alpha_0) &\rightarrow_d \text{MN}\left(0, \left[Km\left(\frac{\mu_\alpha \mu'_\alpha}{\sigma^2}\right) A^{1/(r+2)} \right]^{-1}\right) \\ \sqrt{\frac{T^{1/(r+2)}}{\Delta}}(\hat{\beta} - \beta_0) &\rightarrow_d \text{MN}\left(0, \frac{1}{2} \left[Km\left(\frac{\sigma_\beta \sigma'_\beta}{\sigma^2}\right) A^{1/(r+2)} \right]^{-1}\right)\end{aligned}$$

independently as $T \rightarrow \infty$ and $\Delta \rightarrow 0$, using notation introduced in Theorem 3.5(a).

For the null recurrent processes satisfying the required integrability condition, the limit distributions of both the drift term and diffusion term parameters are mixed normal, with the mixing variate given by a Mittag-Leffler process at time 1. The index of the underlying null recurrent process plays an important role, determining the exact convergence rates of the MLE's and the index of the Mittag-Leffler process in the mixing variate of the limit distributions. Note that the convergence rates of the MLE's here are strictly lower than the case of positive recurrent processes, since $r > -1$. Roughly, this is because the vanishing tails of μ_α/σ and σ_β/σ attenuate the signal from the stochastic trend of X in this case.

Theorem 4.5 Let Assumptions 3.1-3.3 hold. Moreover, let X be null recurrent and regular with index $r > -1$, and assume that μ_α/σ and σ_β/σ are m -square asymptotically homogeneous and define $\nu_\alpha = (\mu_\alpha/\sigma) \circ s^{-1}$ and $\tau_\beta = (\sigma_\beta/\sigma) \circ s^{-1}$. Then we have

$$\begin{aligned}\sqrt{T}\kappa'(\nu_\alpha, T^{1/(r+2)})(\hat{\alpha} - \alpha_0) &\rightarrow_d \left(\int_0^1 hh'(\nu_\alpha, X_t^\circ) dt \right)^{-1} \int_0^1 h(\nu_\alpha, X_t^\circ) dW_t^\circ \\ \sqrt{\frac{T}{\Delta}}\kappa'(\tau_\beta, T^{1/(r+2)})(\hat{\beta} - \beta_0) &\rightarrow_d \frac{1}{\sqrt{2}} \left(\int_0^1 hh'(\tau_\beta, X_t^\circ) dt \right)^{-1} \int_0^1 h(\tau_\beta, X_t^\circ) dV_t^\circ\end{aligned}$$

jointly as $T \rightarrow \infty$ and $\Delta \rightarrow 0$, where V° is a standard Brownian motion independent of W° and X° and other notations are introduced in Theorem 3.5(b).

The limit distributions of the MLE's for null recurrent processes under our asymptotic homogeneity condition have some important common aspects with our previous results. First, the limit distribution of the diffusion term parameter is mixed normal as in the case of null recurrent processes with the integrability condition. This is because V° is independent of X° . The only difference is that the mixing variate here is given by a functional of the

limit process of the underlying diffusion. In contrast, the limit distribution of the drift term parameter is essentially non-Gaussian. Note that W° is independent upon X° as shown in Proposition 3.4. Second, the convergence rates of the MLE's for null recurrent processes under the asymptotic homogeneity condition for μ_α/σ and σ_β/σ are in general faster than those under the integrability condition. In the simple case that $\nu_\alpha(x) = \tau_\beta(x) = |x|^k$ with $r + 2k > -1$, as required to meet our asymptotic homogeneity condition, we have

$$\sqrt{T}\kappa(\nu_\alpha, T^{1/(r+2)}) = \sqrt{T}\kappa(\tau_\beta, T^{1/(r+2)}) = T^{1/2}T^{k/(r+2)} = T^{(r+2k+2)/2(r+2)}$$

and $r + 2k + 2 > 1$.

The asymptotics of the standard test statistics can easily be obtained from our asymptotics for the MLE's, whenever the MLE's have mixed-normal limit distributions. In particular, the standard tests such as Wald, LM and LR tests based on the MLE's have the standard normal or chi-square distribution asymptotically in this case.

Example 4.1 (BM with Drift) For the Brownian motion with drift introduced in (5), it follows directly from Theorem 4.1 that

$$\begin{aligned}\hat{\alpha} - \alpha &\sim_p \frac{\beta W_T}{T} =_d \mathbb{N}\left(0, \frac{\beta^2}{T}\right) \\ \hat{\beta} - \beta &\sim_p \sqrt{\frac{\Delta}{2}} \frac{\beta V_T}{T} =_d \mathbb{N}\left(0, \frac{\beta^2 \Delta}{2T}\right),\end{aligned}$$

and we have $\sqrt{T}(\hat{\alpha} - \alpha) \rightarrow_d \mathbb{N}(0, \beta^2)$ and $\sqrt{T/\Delta}(\hat{\beta} - \beta) \rightarrow_d \mathbb{N}(0, \beta^2/2)$ for the drift and diffusion term parameters.

Example 4.2 (OU Process) For the Ornstein-Uhlenbeck process defined in (6), we have

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta \left(\int_0^T \begin{pmatrix} 1 & X_t \\ X_t & X_t^2 \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} 1 \\ X_t \end{pmatrix} dW_t$$

and

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \beta \frac{V_T}{T},$$

due to Theorem 4.1. Moreover, it follows from Theorem 4.3 that

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \rightarrow_d \mathbb{N} \left(0, \begin{bmatrix} \beta^2 - \frac{2\alpha_1^2}{\alpha_2} & -2\alpha_1 \\ -2\alpha_1 & -2\alpha_2 \end{bmatrix} \right)$$

for the drift term parameters, and

$$\sqrt{T/\Delta}(\hat{\beta} - \beta) \rightarrow_d \mathbb{N}(0, \beta^2/2)$$

for the diffusion term parameter.

We may also consider the case where $\alpha_1 = 0$ and $\alpha_2 = 0$. In this case, we have $X_t = \beta W_t$, from which it follows that

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta \left(\int_0^T \begin{pmatrix} 1 & \beta W_t \\ \beta W_t & \beta^2 W_t^2 \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} 1 \\ \beta W_t \end{pmatrix} dW_t.$$

In particular, X becomes a null recurrent process and it follows from Theorem 4.5 that

$$\begin{pmatrix} \sqrt{T} & 0 \\ 0 & T \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \rightarrow_d \beta \left(\int_0^1 \begin{pmatrix} 1 & \beta W_t \\ \beta W_t & \beta^2 W_t^2 \end{pmatrix} dt \right)^{-1} \int_0^1 \begin{pmatrix} 1 \\ \beta W_t \end{pmatrix} dW_t,$$

and the limit distribution is non-Gaussian and of Dickey-Fuller type. On the other hand, the asymptotics for $\hat{\beta}$ remain the same as above.

Example 4.3 (SR Process) For the Feller's square root process defined in (7), Theorem 4.1 yields

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta \left(\int_0^T \begin{pmatrix} X_t^{-1} & X_t \\ X_t & X_t^2 \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} X_t^{-1/2} \\ X_t^{1/2} \end{pmatrix} dW_t$$

and

$$\hat{\beta} - \beta \sim_p \sqrt{\frac{\Delta}{2}} \beta \frac{V_T}{T}.$$

Under the stationarity condition $2\alpha_1 > \beta^2$, we may also easily deduce

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \rightarrow_d \mathbb{N} \left(0, \beta^2 \begin{bmatrix} 2\alpha_2/(\beta^2 - 2\alpha_1) & -\alpha_1/\alpha_2 \\ -\alpha_1/\alpha_2 & \frac{1}{2}\alpha_1(2\alpha_1 + \beta^2)/\alpha_2^2 \end{bmatrix}^{-1} \right)$$

and

$$\sqrt{T/\Delta}(\hat{\beta} - \beta) \rightarrow_d \mathbb{N}(0, \beta^2/2)$$

from Theorem 4.3.

Example 4.4 (CEV Process) For the CEV process defined in (8), we may deduce from Theorem 4.1 that

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \beta_1 \left(\int_0^T \begin{pmatrix} |X_t|^{-2\beta_2} & X_t |X_t|^{-2\beta_2} \\ X_t |X_t|^{-2\beta_2} & |X_t|^{-2\beta_2+2} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} |X_t|^{-\beta_2} \\ X_t |X_t|^{-\beta_2} \end{pmatrix} dW_t$$

and

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \sim_p \sqrt{\frac{\Delta}{2}} \beta_1 \left(\int_0^T \begin{pmatrix} 1 & \beta_1 \log |X_t| \\ \beta_1 \log |X_t| & \beta_1^2 \log^2 |X_t| \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} 1 \\ \beta_1 \log |X_t| \end{pmatrix} dV_t.$$

If $\alpha_1 > 0$, $\alpha_2 < 0$, $\beta_1 > 0$ and $\beta_2 > 1/2$, the process becomes positive recurrent and we have

$$\sqrt{T} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \rightarrow_d \mathbb{N} \left(0, \beta_1^2 \left[\mathbb{E} \begin{pmatrix} X_t^{-2\beta_2} & X_t^{-2\beta_2+1} \\ X_t^{-2\beta_2+1} & X_t^{-2\beta_2+2} \end{pmatrix} \right]^{-1} \right)$$

and

$$\sqrt{\frac{T}{\Delta}} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \rightarrow_d \mathbb{N} \left(0, \frac{\beta_1^2}{2} \left[\mathbb{E} \begin{pmatrix} 1 & \beta_1 \log(X_t) \\ \beta_1 \log(X_t) & \beta_1^2 \log^2(X_t) \end{pmatrix} \right]^{-1} \right),$$

due to Theorem 4.3.

Example 4.5 (NLD Process) For the positive recurrent nonlinear drift (NLD) diffusion process defined in (9), the asymptotics of the MLE's follow similarly as in the previous examples. Here we derive their asymptotics in the null recurrent case and consider

$$dX_t = (\alpha_1 + \alpha_2 X_t^{-1}) dt + \sqrt{\beta_1 + \beta_2 X_t} dW_t,$$

on $\mathcal{D} = (0, \infty)$ for $\beta_1 > 0$, $\beta_2 > 0$, $0 < \alpha_1 < \beta_2/2$ and $\alpha_2 > \beta_1/2$. The primary asymptotics is given by

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} \sim_p \left(\int_0^T \begin{pmatrix} \frac{1}{\beta_1 + \beta_2 X_t} & \frac{1}{X_t(\beta_1 + \beta_2 X_t)} \\ \frac{1}{X_t(\beta_1 + \beta_2 X_t)} & \frac{1}{X_t^2(\beta_1 + \beta_2 X_t)} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} \frac{1}{\sqrt{\beta_1 + \beta_2 X_t}} \\ \frac{1}{X_t \sqrt{\beta_1 + \beta_2 X_t}} \end{pmatrix} dW_t$$

and

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \sim_p \sqrt{\frac{\Delta}{2}} \left(\int_0^T \begin{pmatrix} \frac{1}{4(\beta_1 + \beta_2 X_t)^2} & \frac{X_t}{4(\beta_1 + \beta_2 X_t)^2} \\ \frac{X_t}{4(\beta_1 + \beta_2 X_t)^2} & \frac{X_t^2}{4(\beta_1 + \beta_2 X_t)^2} \end{pmatrix} dt \right)^{-1} \int_0^T \begin{pmatrix} \frac{1}{2(\beta_1 + \beta_2 X_t)} \\ \frac{X_t}{2(\beta_1 + \beta_2 X_t)} \end{pmatrix} dV_t$$

from Theorem 4.1. For this model, we have

$$s^*(x) = \exp \left(- \int_1^x \frac{2(\alpha_1 + \alpha_2 u^{-1})}{\beta_1 + \beta_2 u} du \right) = \left(\frac{\beta_1 + \beta_2}{\beta_1 + \beta_2 x} \right)^{\frac{2\alpha_1}{\beta_2} - \frac{2\alpha_2}{\beta_1}} x^{-\frac{2\alpha_2}{\beta_1}},$$

from which it follows that

$$\frac{1}{(s \cdot 2\sigma^2) \circ s^{-1}(x)} \sim \begin{cases} \frac{1}{\beta_2 - 2\alpha_1} \left(\frac{\beta_2}{\beta_2 - 2\alpha_1} \right)^{\frac{2\alpha_1}{2\alpha_1 - \beta_2}} \left(\frac{\beta_2}{\beta_1 + \beta_2} \right)^{\frac{2(\alpha_1\beta_1 - \alpha_2\beta_2)}{\beta_1(\beta_2 - 2\alpha_1)}} x^{\frac{\beta_2 - 4\alpha_1}{2\alpha_1 - \beta_2}} & \text{as } x \rightarrow \infty \\ \frac{1}{\beta_1} \left(\frac{\beta_1}{\beta_1 - 2\alpha_2} \right)^{\frac{4\alpha_2}{2\alpha_2 - \beta_1}} \left(\frac{\beta_1}{\beta_1 + \beta_2} \right)^{\frac{4(\alpha_1\beta_1 - \alpha_2\beta_2)}{\beta_2(\beta_1 - 2\alpha_2)}} (-x)^{\frac{4\alpha_2}{\beta_1 - 2\alpha_2}} & \text{as } x \rightarrow -\infty \end{cases}.$$

Therefore, we have $T^{\frac{2\alpha_1}{\beta_2} - 1} X_{Tt}^s \rightarrow_d X_t^r$, where X^r is a generalized diffusion process associated with the speed density

$$\frac{1}{\beta_2 - 2\alpha_1} \left(\frac{\beta_2}{\beta_2 - 2\alpha_1} \right)^{\frac{2\alpha_1}{2\alpha_1 - \beta_2}} \left(\frac{\beta_2}{\beta_1 + \beta_2} \right)^{\frac{2(\alpha_1\beta_1 - \alpha_2\beta_2)}{\beta_1(\beta_2 - 2\alpha_1)}} x^{(\beta_2 - 4\alpha_1)/(2\alpha_1 - \beta_2)} 1\{x \geq 0\}.$$

The limit process X^r becomes $4\alpha_1/\beta_2$ -dimensional Bessel process in natural scale if it is multiplied by

$$\alpha_1^{\frac{2\alpha_1}{\beta_2} - 1} \left(\frac{\beta_2}{\beta_2 - 2\alpha_1} \right)^{-\frac{2\alpha_1}{\beta_2}} \left(\frac{\beta_2}{\beta_1 + \beta_2} \right)^{\frac{2\alpha_1}{\beta_2} - \frac{2\alpha_2}{\beta_1}}.$$

Since $\mu_\alpha \mu'_\alpha / \sigma$ and $\sigma_\beta \sigma'_\beta / \sigma$ are integrable with respect to the speed density of X except for the second diagonal element of $\sigma_\beta \sigma'_\beta / \sigma$, which becomes asymptotically homogenous if composited with s^{-1} , we deduce from Theorem 4.4 that

$$\begin{aligned} T^{\frac{1}{2} - \frac{\alpha_1}{\beta_2}} \begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \end{pmatrix} &\rightarrow_d \text{MN}\left(0, [Km(f_\alpha f'_\alpha) A^{1 - 2\alpha_1/\beta_2}]^{-1}\right) \\ \frac{T^{\frac{1}{2} - \frac{\alpha_1}{\beta_2}}}{\sqrt{\Delta}} (\hat{\beta}_1 - \beta_1) &\rightarrow_d \text{MN}\left(0, [2Km(f_{\beta_1}^2) A^{1 - 2\alpha_1/\beta_2}]^{-1}\right) \\ \sqrt{\frac{T}{\Delta}} (\hat{\beta}_2 - \beta_2) &\rightarrow_d \mathbb{N}(0, 2\beta_2^2), \end{aligned}$$

where

$$\begin{aligned} K &= -\frac{\Gamma(-2\alpha_1/\beta_2)\alpha_1}{\Gamma(2\alpha_1/\beta_2)\beta_2} 4^{\frac{\alpha_1}{\beta_2}} \beta_2^{1 - \frac{2\alpha_1}{\beta_2}} \left(\frac{\beta_2}{\beta_1 + \beta_2} \right)^{\frac{2\alpha_2}{\beta_1} - \frac{2\alpha_1}{\beta_2}} \\ m(x) &= (\beta_1 + \beta_2)^{\frac{2\alpha_2}{\beta_1} - \frac{2\alpha_1}{\beta_2}} (\beta_1 + \beta_2 x)^{\frac{2\alpha_1}{\beta_2} - \frac{2\alpha_2}{\beta_1} - 1} x^{\frac{2\alpha_2}{\beta_1}} \\ f_\alpha(x) &= \left(\frac{\frac{1}{\sqrt{\beta_1 + \beta_2 x}}}{\frac{1}{x\sqrt{\beta_1 + \beta_2 x}}} \right), \quad f_{\beta_1}(x) = \frac{1}{2(\beta_1 + \beta_2 x)}, \end{aligned}$$

and $A^{1 - 2\alpha_1/\beta_2}$ is the Mittag-Leffler process with index $1 - 2\alpha_1/\beta_2$ at time 1.

5. Simulations

We perform the Monte Carlo simulations to examine the relevancy and usefulness of our asymptotic theory in approximating finite sample distributions of the MLE's for diffusion models. For our simulations, we use the CEV model $dX_t = (\alpha_1 + \alpha_2 X_t)dt + \beta_1 X_t^{\beta_2} dW_t$ in Examples 2.2(a) and 4.4 with $\alpha_1 = 0.0072$, $\alpha_2 = -0.09$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$. The parameter values in our simulation model are the estimates obtained by Aït-Sahalia (1999) for the CEV model fitted with the monthly federal funds rate for the period of 1963-1998. We consider the time spans $T = 10$ and $T = 50$ representing 10 and 50 years of data respectively, and the sampling frequencies $\Delta = 0.005$ and $\Delta = 0.1$ respectively for the daily and monthly observations. To obtain the samples used in our simulations, we rely on the Milstein scheme to discretize our model and generate samples at a finer sampling interval $\delta = 0.0005$, and collect the samples at each of the values for Δ considered in our simulation. The simulation iterations are set to be 5000. To save the space, we only present the results for the MLE based on the Milstein approximation. The results for other MLE's are largely identical under our simulation setup.

5.1 Finite Sample Distributions

Figures 1 and 2 show the finite sample distributions of the MLE. We may clearly see that the distribution of the diffusion term parameter gets closer to normal as the sampling frequency increases. This is in contrast with the distribution of the drift term parameter, which remains to be far from being normal even at a relatively high sampling frequency. Our simulation results here are well expected from the asymptotic theory in the paper, especially our primary asymptotics. They show that the asymptotic leading term of the diffusion term parameter is mixed-normal as long as the sampling interval Δ is sufficiently small relative to the time span T of sample. On the other hand, they imply that the asymptotic leading term of the drift term parameter is non-Gaussian at all finite T no matter how small Δ is. It is quite notable to see how well our primary asymptotics approximate the finite sample distributions of the MLE's, particularly when the sampling frequency is relatively high. Indeed, the finite sample distribution of the drift term parameter is remarkably well

approximated by our primary asymptotics at daily frequency for both 10 and 50 years of time span.

Figures 3 and 4 show the finite sample distributions of the t -statistics. In Figure 3, we can see that the actual distribution of the t -statistic is quite distinct from the standard normal distribution even when the sampling frequency is daily and the time span of sample is as large as 50 years. This implies that the t -tests would have serious size distortions in finite samples if we use the critical values from the standard normal distribution. However, it is clearly seen that our primary asymptotics provide quite accurate approximations for the finite sample distribution of the t -statistics for both 10 and 50 years of sample horizon. Therefore, our primary asymptotics can be used to obtain the critical values of the t -test more appropriate in finite samples. In contrast, the distribution for the diffusion term parameter is pretty close to normal at daily frequency, as shown in Figure 4. Again, this is well expected from our asymptotic theory, which shows that the distribution of the t -statistic is normal even for finite T as long as Δ is sufficiently small relative to T .

5.2 Bias and Size Corrections

Clearly, we may use our primary asymptotics to correct for the finite sample bias of the MLE and the size distortion of the t -test in finite samples. This possibility is explored here. Since our primary asymptotics require Δ to be small relative to T , we mainly consider the daily observations for our simulation here. However, our primary asymptotics also work well for the monthly observations in our simulation setup, and yield similar results as the daily observations. For the bias correction of the MLE, we simulate the means of our primary asymptotics, and use the simulated means of the primary asymptotics to adjust for the original estimates. Likewise, to correct the size of the t -test, we use the critical values obtained from the simulated distributions of our primary asymptotics. The sample means and the empirical distributions of the primary asymptotics are computed using 2000 simulated samples. We use the true parameter values in our simulations to obtain the means and distributions of our primary asymptotics. This is because our main purpose is to show how effective our primary asymptotics are in correcting the finite sample bias and size for

the MLE and t -test.⁷

Table 1 shows the biases of the MLE with and without correction. It is notable that the biases of the MLE's for α_1 and α_2 are as big as approximately 600% of their true values for the case of $T = 10$. However, their biases virtually disappear after correction, decreasing down to approximately 1% of their original magnitudes. Even for $T = 50$, the biases in the MLE's for α_1 and α_2 are substantial and the corrections based on their primary asymptotics are well motivated. In contrast, the MLE's for β_1 and β_2 have negligible biases even for the case of $T = 10$, though the magnitudes of the biases are slightly reduced as we increase T from 10 to 50. Naturally, our correction has no effect on their finite sample performance. In Table 2, we compare the actual sizes of the t -tests based on the standard critical values and the critical values obtained from our primary asymptotics. The usual t -tests for α_1 and α_2 have enormous size distortions for the case of $T = 10$, which remain to be significant as T increases up to 50. On the other hand, the t -tests relying on the primary asymptotics have the actual sizes that are virtually identical to their nominal values even when $T = 10$. As expected, the t -tests for β_1 and β_2 show no evidence of finite sample size distortions.

6. Conclusion

In the paper, we develop the asymptotic theory for the MLE's in diffusion models. We consider a wide class of the MLE's, including the exact, quasi and approximate MLE's, which are based on the exact transition density or the transition densities approximated in a variety of methods. In our framework, we accommodate virtually all likelihood-based estimators proposed in the literature. Our assumptions on the underlying diffusion models are also truly general. In particular, we allow for very general nonstationary diffusions as well as stationary diffusions in the development of our asymptotics. Our asymptotic theory provides the exact convergence rates and explicit limit distributions of the MLE's in such a general and flexible context. The convergence rates for the MLE's vary depending upon the drift and diffusion functions and the recurrence property of the underlying diffusion. For the parameters in the drift and diffusion terms, they are given respectively by T^κ and

⁷If based on the estimated parameter values, the bias and size corrections for the MLE and t -test based on our primary asymptotics are less effective, though they still provide substantial improvements.

$\Delta^{-1/2}T^\kappa$ with some constant $0 < \kappa < \infty$. For the positive recurrent diffusions, the MLE's are asymptotically normal. However, for the null recurrent diffusions, they have generally non-Gaussian limit distributions that may be regarded as a generalized version of the limit distribution of the Dickey-Fuller unit root test. The drift and diffusion term parameters are asymptotically uncorrelated, and become independent when their asymptotic distributions are normal or mixed normal.

All the MLE's we consider in the paper have the identical leading terms in our primary asymptotics. Therefore, they are equivalent at least up to the asymptotic order represented by the leading terms in our primary asymptotics. As is well known, however, their relative performances in finite samples vary across different models and parameter values, especially when the sampling frequency Δ is not sufficiently small. It would therefore be interesting to derive their higher order asymptotic expansions and use them to better explain the relative finite sample performances of various MLE's. We believe that the higher order expansions along our approach in the paper will give us some important clues on the finite sample performance of the MLE's. Finally, the continuous time asymptotics developed in the paper can be used in developing asymptotics for many other interesting models. In particular, the continuous time asymptotics we establish in the paper for the general null recurrent diffusions can be used in many other important contexts to analyze the discrete samples collected from null recurrent diffusions. In particular, our asymptotics for null recurrent diffusions make it possible to explore continuous time models involving general nonstationary processes. This will be shown more clearly in our subsequent researches.

Appendix A. Useful Lemmas

A.1. Lemmas

Lemma A1

Let f be twice differentiable and suppose that f and its derivatives satisfy the boundedness condition in Assumption 3.1(b). Also let Assumption 3.1 holds. Then

$$\sum_{i=1}^n f(X_{(i-1)\Delta})\Delta = \int_0^T f(X_t)dt + O_p(\Delta T^{2pq+1})$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$.

Lemma A2

Let $f(t, x, y, \theta)$ be twice differentiable, and f and its derivatives satisfy the boundedness condition in Assumption 3.1(b). Then, under Assumptions 3.1 we have the following as $T \rightarrow \infty$ and $\Delta \rightarrow 0$.

(a) If the following repeated integral only consists of Riemann integrals (dt), then

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^s f(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dt = O_p(\Delta^{k-1} T^{2pq+1})$$

uniformly in $\theta \in \mathcal{N}$, where k is the dimension of the repeated integral.

(b) Otherwise, i.e., if the repeated integral involves Itô integrals (dW_t), then

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^s f(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dW_t = O_p(\Delta^{(2k_1+k_2-1)/2} T^{2pq+1/2}) \quad (\text{A.1})$$

for all $\theta \in \Theta$, where k_1 is the number of dt and k_2 is the number of dW_t . The combination and the order of dt and dW_t can be arbitrary.

Moreover, if we additionally assume that X is either positively recurrent with its time invariant measure π satisfying $\pi(g^{2d}) < \infty$ for g defined in Assumption 3.1(b) and d greater than the dimension of θ , or null recurrent and regular with index $r > -1$, then (A.1) holds uniformly in $\theta \in \mathcal{N}$.

Lemma A3

Define

$$V_t^\Delta = \sqrt{\frac{2}{\Delta}} \left(\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_u dW_s + \int_{(j-1)\Delta}^t \int_{(j-1)\Delta}^s dW_u dW_s \right)$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \dots, n+1$. Then

$$V^\Delta \rightarrow_p V$$

for a standard Brownian motion V independent of W , and $V_T^\Delta - V_T = O_p((\Delta T)^{1/4})$ as $T \rightarrow \infty$ and $\Delta \rightarrow 0$ satisfying $\Delta T \rightarrow 0$.

Lemma A4

Let f be twice differentiable and let f and its derivatives satisfy the boundedness condition in Assumption 3.1(b). Also let Assumption 3.1 holds. Then as $T \rightarrow \infty$ and $\Delta \rightarrow 0$, (a)

$$\sum_{i=1}^n f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) = \int_0^T f(X_t) dW_t + O_p(\sqrt{\Delta} T^{4pq+1/2})$$

and (b)

$$\sqrt{\frac{2}{\Delta}} \sum_{i=1}^n f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_u dW_s = \int_0^T f(X_t) dV_t + O_p(\Delta^{1/4} T^{4pq+7/4}),$$

where V is as defined in Lemma A3.

Lemma A5

Let ℓ be a normalized log-likelihood defined as $\ell(t, x, y, \theta) = t \log(\sqrt{t} p(t, x, y, \theta))$, where $p(t, x, y, \theta)$ is the true transition density of a diffusion process given by $dX_t = \mu(X_t, \alpha)dt + \sigma(X_t, \beta)dW_t$. Then for all $x \in \mathcal{D}$ and θ in the interior of Θ , ℓ satisfies

$$\begin{aligned} \ell(0, x, x, \theta) &= 0, & \ell_y(0, x, x, \theta) &= 0, \\ \ell_t(0, x, x, \theta) &= -\log(\sigma(x, \beta)), & \ell_{y^2}(0, x, x, \theta) &= -\frac{1}{\sigma^2(x, \beta)}, \\ \ell_{ty}(0, x, x, \theta) &= \frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} + \nu(x, \beta), \\ \ell_{t^2}(0, x, x, \theta) + \ell_{ty^2}(0, x, x, \theta)\sigma^2(x, \beta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + \omega(x, \beta) \end{aligned}$$

ignoring terms unrelated with θ , where ν and ω are some functions not depending on α . Also, $\ell_{y^3}(0, x, x, \theta)$ and $\ell_{y^4}(0, x, x, \theta)$ do not depend on α .

A.2. Proofs of Lemmas

Proof of Lemma A1

We have

$$\begin{aligned} \sum_{i=1}^n f(X_{(i-1)\Delta})\Delta &= \int_0^T f(X_t)dt - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (f(X_t) - f(X_{(i-1)\Delta}))dt \\ &= \int_0^T f(X_t)dt - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \left(\mu f' + \frac{\sigma^2 f''}{2} \right) (X_s) ds dt \\ &\quad - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma f'(X_s) dW_s dt \\ &= \int_0^T f(X_t)dt + R_{1T} + R_{2T} \end{aligned}$$

by Itô's lemma. By changing the order of the integrals,

$$R_{1T} = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \left(\mu f' + \frac{\sigma^2 f''}{2} \right) (X_s) ds \leq \Delta \int_0^T \left| \mu f' + \frac{\sigma^2 f''}{2} \right| (X_t) dt = O_p(\Delta T^{2pq+1})$$

from Assumptions 3.1(b) and 3.1(c). Also,

$$R_{2T} = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \sigma f'(X_s) dW_s$$

and this is a martingale whose quadratic variation is

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s)^2 \sigma^2 f'^2(X_s) ds \leq \Delta^2 \int_0^T \sigma^2 f'^2(X_t) dt = O_p(\Delta^2 T^{3pq+1}).$$

Thus both R_{1T} and R_{2T} are of order $O_p(\Delta T^{2pq+1})$ under Assumption 3.1(d).

Proof of Lemma A2

(a) It directly follows from Assumptions 3.1(b) and 3.1(c) that

$$\begin{aligned} & \sup_{\theta \in \mathcal{N}} \left| \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^s f(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dt \right| \\ & \leq \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^s g(X_{(i-1)\Delta}) g(X_r) dr \cdots dt \leq n \Delta^k \left(\sup_{t \in [0, T]} g(X_t) \right)^2 = O_p(\Delta^{k-1} T^{2pq-1}) \end{aligned}$$

for small $\Delta \geq 0$ and some g defined in Assumption 3.1(b), which completes the proof.

(b) We will first show that (A.1) holds for given $\theta \in \Theta$, and the uniform order will be derived in the next part. For fixed $\theta \in \Theta$, we will deal with the two cases separately: one in which the most inner integral is an Itô integral (dW_t), and the other in which the most inner integral is a Riemann integral (dt). For the first case, we revert the order of integrations in (A.1) such that

$$\begin{aligned} & \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_r^{i\Delta} \cdots \int_u^{i\Delta} f(r, X_{(i-1)\Delta}, X_r, \theta) dW_t \cdots ds dW_r \\ & = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f(r, X_{(i-1)\Delta}, X_r, \theta) \int_r^{i\Delta} \cdots \int_u^{i\Delta} dW_t \cdots ds dW_r. \end{aligned} \quad (\text{A.2})$$

We can always write that

$$\int_t^{i\Delta} A_s dB_s = \int_{(i-1)\Delta}^{i\Delta} A_s dB_s - \int_{(i-1)\Delta}^t A_s dB_s = P_i - Q_{i,t},$$

where A and B are semimartingales, and treat P_i as a random variable invariant over $t \in [(i-1)\Delta, i\Delta]$ and $Q_{i,t}$ as a semimartingale adapted to the filtration generated by W . Therefore, if we define a continuous version of (A.2) as

$$\begin{aligned} M_t &= \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} f(r, X_{(i-1)\Delta}, X_r, \theta) \int_r^{i\Delta} \cdots \int_u^{i\Delta} dW_v \cdots ds dW_r \\ &+ \int_{(j-1)\Delta}^t f(r, X_{(j-1)\Delta}, X_r, \theta) \int_r^{j\Delta} \cdots \int_u^{j\Delta} dW_v \cdots ds dW_r \end{aligned}$$

for $t \in [(j-1)\Delta, j\Delta]$, then the quadratic variation of M_T is given by

$$\begin{aligned} [M]_T &= \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f^2(r, X_{(i-1)\Delta}, X_r, \theta) \left(\int_r^{i\Delta} \cdots \int_u^{i\Delta} dW_t \cdots ds \right)^2 dr \\ &\leq \sup_{t \in [0, T]} g^4(X_t) \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_r^{i\Delta} \cdots \int_u^{i\Delta} dW_t \cdots ds \right)^2 dr \\ &= O_p(\Delta^{(2k_1+k_2-1)} T^{4pq+1}), \end{aligned} \quad (\text{A.3})$$

where the order of the summation in the second line can be obtained by taking expectation and changing the order of integrations. From (A.3) we establish that (A.1) holds in the first case.

For the second case, we also revert the order of the integrals, but we will do so only from the most inner Itô integral to the most outer integral. Then we obtain that

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_u^{i\Delta} \cdots \int_w^{i\Delta} \left(\int_{(i-1)\Delta}^u \cdots \int_{(i-1)\Delta}^h f(r, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dv \right) dW_t \cdots ds dW_u. \quad (\text{A.4})$$

Note that in (A.4), the most outer integral becomes Itô integral, and the most inner integrals in the parentheses only consist of Riemann integrals. We can rewrite (A.4) as

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^u \cdots \int_{(i-1)\Delta}^h f(r, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dv \right) \int_u^{i\Delta} \cdots \int_w^{i\Delta} dW_t \cdots ds dW_u \quad (\text{A.5})$$

since the repeated integral in the parentheses is only related with the most outer integral. Similarly as (A.2), the quadratic variation of (A.5) is given by

$$\begin{aligned} &\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^u \cdots \int_{(i-1)\Delta}^h f(r, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dv \right)^2 \left(\int_u^{i\Delta} \cdots \int_w^{i\Delta} dW_t \cdots ds \right)^2 du \\ &\leq \sup_{t \in [0, T]} g^4(X_t) \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^u \cdots \int_{(i-1)\Delta}^h dr \cdots dv \right)^2 \left(\int_u^{i\Delta} \cdots \int_w^{i\Delta} dW_t \cdots ds \right)^2 du \\ &= O_p(\Delta^{(2k_1+k_2-1)} T^{4pq+1}), \end{aligned}$$

which completes the proof.

Uniform Martingale Order In this part we will show that

$$\sup_{\theta \in \mathcal{N}} \left| \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^s f(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r, \theta) dr \cdots dW_t \right| = O_p(\Delta^{k_1+k_2/2-1/2} T^{2pq+1/2}) \quad (\text{A.6})$$

as $T \rightarrow \infty$ and $\Delta \rightarrow 0$. Denote $M_T(\theta) = \Delta^{-k_1-k_2/2+1/2} T^{-2pq-1/2} A_T(\theta)$, where $A_T(\theta)$ is the summation on the left hand side of (A.6). We will show the uniform boundedness of $M_T(\theta)$ by establishing the convergence of finite dimensional distributions and the weak relative compactness, due to Kolmogorov's criterion.

Firstly for the weak relative compactness, it follows from the Hölder inequality and Assumption

3.1(b) that

$$\begin{aligned}
& \mathbb{E}|M_T(\theta_1) - M_T(\theta_2)|^d \tag{A.7} \\
& \leq \frac{1}{\Delta^{(k_1+k_2/2-1/2)dT^{2pqd+d/2}}} \mathbb{E} \left(\sum_{i=1}^n \sup_{v \in [(i-1)\Delta, i\Delta]} \left| \int_{(i-1)\Delta}^v \cdots \int_{(i-1)\Delta}^s (f_{i,r}(\theta_1) - f_{i,r}(\theta_2)) dr \cdots dW_t \right| \right)^d \\
& \leq \frac{n}{\Delta^{(k_1+k_2/2-1/2)dT^{2pqd+d/2}}} \sum_{i=1}^n \mathbb{E} \left(\sup_{v \in [(i-1)\Delta, i\Delta]} \left| \int_{(i-1)\Delta}^v \cdots \int_{(i-1)\Delta}^s (f_{i,r}(\theta_1) - f_{i,r}(\theta_2)) dr \cdots dW_t \right| \right)^d,
\end{aligned}$$

where $f_{i,r}(\theta) = f(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r, \theta)$. To bound the last line of (A.7), we utilize the following two rules:

$$\mathbb{E} \left(\sup_{v \in [s,t]} \left| \int_s^v A_r dW_r \right| \right)^d \leq \mathbb{E} \left(\sup_{v \in [s,t]} \left| \int_s^v A_r^2 dr \right| \right)^{d/2} \leq \Delta^{d/2} \mathbb{E} \left(\sup_{v \in [s,t]} |A_v| \right)^d, \tag{A.8}$$

$$\mathbb{E} \left(\sup_{v \in [s,r]} \left| \int_s^v A_r dr \right| \right)^d \leq \Delta^d \mathbb{E} \left(\sup_{v \in [s,t]} |A_v| \right)^d \tag{A.9}$$

for $d \geq 1$ and $t \geq s$, where A is a semimartingale. Note that we obtain the first inequality of (A.8) due to the Burkholder-Davis-Gundy inequality. We apply (A.8) or (A.9) to the last line of (A.7), depending on whether the most outer integral is an Itô integral or a Riemann integral, respectively. We repeat applying (A.8) and (A.9) for $k-1$ times, then we obtain that

$$\begin{aligned}
\mathbb{E}|M_T(\theta_1) - M_T(\theta_2)|^d & \leq \frac{n}{T^{2pqd+d/2}} \sum_{i=1}^n \mathbb{E} \left(\sup_{v \in [(i-1)\Delta, i\Delta]} \int_{(i-1)\Delta}^v (f_{i,r}(\theta_1) - f_{i,r}(\theta_2))^2 dr \right)^{d/2} \\
& \leq \|\theta_1 - \theta_2\|^d \frac{n}{T^{2pqd+d/2}} \sum_{i=1}^n \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} g^2(X_{(i-1)\Delta}) g^2(X_r) dr \right)^{d/2} \tag{A.10} \\
& \leq \|\theta_1 - \theta_2\|^d \frac{1}{T^{2pqd+d/2-1}} \sqrt{\mathbb{E} \sum_{i=1}^n g^{2d}(X_{(i-1)\Delta}) \Delta} \sqrt{\mathbb{E} \int_0^T g^{2d}(X_t) dt}
\end{aligned}$$

when the most inner integral of the last line of (A.7) is dW_t , and

$$\begin{aligned}
\mathbb{E}|M_T(\theta_1) - M_T(\theta_2)|^d & \leq \frac{n\Delta^{-d/2}}{T^{2pqd+d/2}} \sum_{i=1}^n \mathbb{E} \left(\sup_{v \in [(i-1)\Delta, i\Delta]} \int_{(i-1)\Delta}^v |f_{i,r}(\theta_1) - f_{i,r}(\theta_2)| dr \right)^d \\
& \leq \|\theta_1 - \theta_2\|^d \frac{n\Delta^{-d/2}}{T^{2pqd+d/2}} \sum_{i=1}^n \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} g(X_{(i-1)\Delta}) g(X_r) dr \right)^d \tag{A.11} \\
& \leq \|\theta_1 - \theta_2\|^d \frac{n\Delta^{-d/2}}{T^{2pqd+d/2}} \sum_{i=1}^n \mathbb{E} \left(\Delta \int_{(i-1)\Delta}^{i\Delta} g^2(X_{(i-1)\Delta}) g^2(X_r) dr \right)^{d/2} \\
& \leq \|\theta_1 - \theta_2\|^d \frac{1}{T^{2pqd+d/2-1}} \sqrt{\mathbb{E} \sum_{i=1}^n g^{2d}(X_{(i-1)\Delta}) \Delta} \sqrt{\mathbb{E} \int_0^T g^{2d}(X_t) dt}
\end{aligned}$$

when the most inner integral of the last line of (A.7) is dt , where we obtain the second to fourth inequalities of (A.10) and (A.11) due to Assumption 3.1(b) and the Hölder inequality.

To obtain bounds for the last lines of (A.10) and (A.11), we consider two cases separately: X is positive recurrent, or null recurrent. Firstly, if X is positively recurrent, we have

$$\mathbb{E} \frac{1}{T} \int_0^T g^{2d}(X_t) dt < \infty, \quad \frac{\Delta}{T} \mathbb{E} \sum_{i=1}^n g^{2d}(X_{(i-1)\Delta}) < \infty \quad (\text{A.12})$$

for all large T from the ergodic theorem. Secondly, if X is null recurrent, we let g^{2d} be m -asymptotically homogeneous and regularly varying without loss of generality, then we can deduce from (B.94), (B.95), (B.96), (B.100) and (B.101) that

$$\frac{1}{T^{2pqd+1}} \int_0^T g^{2d}(X_t) dt < \infty \quad (\text{A.13})$$

for large T , since $q = 1/(r+2)$ due to Proposition 3.4. Also, we have

$$\begin{aligned} \frac{\Delta}{T^{2pqd+1}} \sum_{i=1}^n g^{2d}(X_{(i-1)\Delta}) &= \frac{1}{T^{2pqd+1}} \int_0^T g^{2d}(X_t) dt - \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \mathcal{A}g^{2d}(X_s) ds dt \\ &\quad - \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \mathcal{B}g^{2d}(X_s) dW_s dt \\ &= G_{1T} - G_{2T} - G_{3T}, \end{aligned} \quad (\text{A.14})$$

where $\mathcal{A}g(x) = (g \cdot \mu)(x) + (g \cdot \sigma^2)(x)/2$ and $\mathcal{B}g(x) = (g \cdot \sigma)(x)$, due to Itô's lemma.

To bound (A.14), we have $\mathbb{E}G_{1T} < \infty$ for large T from (A.13), and we also deduce that

$$\begin{aligned} \mathbb{E}G_{2T} &= \mathbb{E} \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_s^{i\Delta} \mathcal{A}g^{2d}(X_s) dt ds \\ &= \mathbb{E} \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \mathcal{A}g^{2d}(X_s) ds \leq \mathbb{E} \frac{\Delta}{T^{2pqd+1}} \int_0^T \mathcal{A}g^{2d}(X_t) dt < \infty \end{aligned} \quad (\text{A.15})$$

for large T under Assumption 3.1(d), where the last inequality is obtained similarly as we obtain (A.13). Lastly for G_{3T} , we have

$$G_{3T} = \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_s^{i\Delta} \mathcal{B}g^{2d}(X_s) dt dW_s = \frac{1}{T^{2pqd+1}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \mathcal{B}g^{2d}(X_s) dW_s.$$

Therefore, we deduce from the Burkholder-Davis-Gundy inequality that

$$\begin{aligned} \mathbb{E}G_{3T} &\leq \mathbb{E} \left(\frac{1}{T^{4pqd+2}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s)^2 (\mathcal{B}g^{2d})^2(X_s) ds \right)^{1/2} \\ &\leq \mathbb{E} \left(\frac{\Delta^2}{T^{4pqd+2}} \int_0^T (\mathcal{B}g^{2d})^2(X_t) dt \right)^{1/2} < \infty \end{aligned} \quad (\text{A.16})$$

for large T under Assumption 3.1(d), where the last inequality is obtained similarly as we obtain (A.13). Therefore, we deduce from (A.13), (A.14), (A.15) and (A.16) that

$$\mathbb{E} \frac{\Delta}{T^{2pqd+1}} \sum_{i=1}^n g^{2d}(X_{(i-1)\Delta}) < \infty \quad (\text{A.17})$$

for all large T . We obtain from (A.10), (A.11), (A.13) and (A.17) that

$$\mathbb{E}|M_T(\theta_1) - M_T(\theta_2)|^d \leq C\|\theta_1 - \theta_2\|^d. \quad (\text{A.18})$$

for all large T , $d \geq 4$ and some $C > 0$. Therefore, Kolmogorov's criterion for weak relative compactness is satisfied.

For the convergence of finite dimensional distributions, we focus on the second moment of $M_T(\theta)$ and show its boundedness in any finite dimensional product space of $M_T(\theta)$. If the most outer integral of $M_T(\theta)$ is dW_t , we obtain that

$$\begin{aligned} \mathbb{E}(M_T(\theta))^2 &= \frac{1}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \left(\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_v dW_t \right)^2 \\ &\leq \frac{1}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^t \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_v \right)^2 dt \\ &\leq \frac{1}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \left(\sup_{t \in [(i-1)\Delta, i\Delta]} \left| \int_{(i-1)\Delta}^t \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_v \right| \right)^2. \end{aligned} \quad (\text{A.19})$$

If the most outer integral of $M_T(\theta)$ is dt , we change the order of integrals so that we obtain

$$\begin{aligned} \mathbb{E}(M_T(\theta))^2 &\leq \frac{1}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \left(\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left[\int_{(i-1)\Delta}^v \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_u \right] \left[\int_v^{i\Delta} \cdots \int_v^t dw \cdots dt \right] dW_v \right)^2 \\ &\leq \frac{\Delta^{2a}}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^v \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_u \right)^2 dv \\ &\leq \frac{\Delta^{2a+1}}{\Delta^{2(k_1+k_2/2-1/2)}T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \left(\sup_{v \in [(i-1)\Delta, i\Delta]} \left| \int_{(i-1)\Delta}^v \cdots \int_{(i-1)\Delta}^s f_{i,r}(\theta) dr \cdots dW_u \right| \right)^2, \end{aligned} \quad (\text{A.20})$$

where a is the dimension of the repeated integral $\int_v^{i\Delta} \cdots \int_v^t dw \cdots dt$ in the second line of (A.20).

We repeat applying (A.8) and (A.9) to (A.19) and (A.20) similarly as we did to the last line of (A.7), then we obtain that

$$\begin{aligned} \mathbb{E}(M_T(\theta))^2 &\leq \frac{1}{T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} g^2(X_{(i-1)\Delta}) g^2(X_r) dr \\ &\leq \frac{1}{T^{4pq+1}} \sqrt{\mathbb{E} \sum_{i=1}^n g^4(X_{(i-1)\Delta}) \Delta} \sqrt{\mathbb{E} \int_0^T g^4(X_t) dt} < \infty, \end{aligned} \quad (\text{A.21})$$

when the most inner integral is dW_t , due to the Hölder inequality, (A.13) and (A.17). Similarly we obtain that

$$\begin{aligned} \mathbb{E}(M_T(\theta))^2 &\leq \frac{\Delta^{-1}}{T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \left(\int_{(i-1)\Delta}^{i\Delta} g(X_{(i-1)\Delta}) g(X_r) dr \right)^2 \\ &\leq \frac{1}{T^{4pq+1}} \mathbb{E} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} g^2(X_{(i-1)\Delta}) g^2(X_r) dr < \infty \end{aligned} \quad (\text{A.22})$$

when the most inner integral is dt . Therefore, the boundedness of $M_T(\theta)$ in the finite dimensional product space of $M_T(\theta)$ follows from (A.21) and (A.22).

The finite dimensional result of (A.21) and (A.22), together with the weak relative compactness condition in (A.18), implies that $\mathbb{E}(M_T(\theta))^2 < \infty$ for all large T uniformly in $\theta \in \mathcal{N}$, from which we obtain that $A_T(\theta) = O_p(\Delta^{(2k_1+k_2-1)/2} T^{2pq+1/2})$ uniformly in $\theta \in \mathcal{N}$. The proof is therefore complete.

Proof of Lemma A3

Clearly, V^Δ is a continuous martingale with quadratic variation given by

$$[V^\Delta]_t = \frac{2}{\Delta} \left[\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} (W_s - W_{(i-1)\Delta})^2 ds + \int_{(j-1)\Delta}^t (W_s - W_{(j-1)\Delta})^2 ds \right]$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \dots, n+1$.

We have

$$\begin{aligned} [V^\Delta]_t - t &= \frac{2}{\Delta} \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} [(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta)] ds \\ &\quad + \frac{2}{\Delta} \int_{(j-1)\Delta}^t [(W_s - W_{(j-1)\Delta})^2 - (s - (j-1)\Delta)] ds + O(\Delta) \end{aligned} \quad (\text{A.23})$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \dots, n+1$, uniformly in $t \in [0, T]$. Therefore, ignoring $O(\Delta)$ term in (A.23) that is unimportant, it follows that

$$\begin{aligned} \mathbb{E}([V^\Delta]_t - t)^2 &= \left(\frac{2}{\Delta} \right)^2 \sum_{i=1}^{j-1} \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} [(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta)] ds \right)^2 \\ &\quad + \left(\frac{2}{\Delta} \right)^2 \mathbb{E} \left(\int_{(j-1)\Delta}^t [(W_s - W_{(j-1)\Delta})^2 - (s - (j-1)\Delta)] ds \right)^2 \end{aligned} \quad (\text{A.24})$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \dots, n+1$, due to the independent increment property of Brownian motion. However, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} [(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta)] ds \right)^2 \\ &\leq \Delta \int_{(i-1)\Delta}^{i\Delta} \mathbb{E} [(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta)]^2 ds = \frac{2\Delta^4}{3} \end{aligned} \quad (\text{A.25})$$

for $i = 1, \dots, n$. Moreover, we may deduce from (A.24) and (A.25) that

$$\begin{aligned} \mathbb{E}([V^\Delta]_t - t)^2 &\leq \left(\frac{2}{\Delta} \right)^2 \sum_{i=1}^n \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} [(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta)] ds \right)^2 \\ &= \left(\frac{2}{\Delta} \right)^2 n \frac{2\Delta^4}{3} = \frac{8}{3} \Delta T \rightarrow 0 \end{aligned} \quad (\text{A.26})$$

under our assumption. Consequently, it follows that

$$\sup_{0 \leq t \leq T} \mathbb{E} ([V^\Delta]_t - t)^2 \rightarrow 0 \quad (\text{A.27})$$

in our asymptotic framework. This implies that $V^\Delta \rightarrow_p V$, where V is the standard Brownian motion.

Now we show that V is independent of W . For this, we note that

$$[V^\Delta, W]_t = \sqrt{\frac{2}{\Delta}} \left[\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} (W_s - W_{(i-1)\Delta}) ds + \int_{(j-1)\Delta}^t (W_s - W_{(j-1)\Delta}) ds \right]$$

for $t \in [(j-1)\Delta, j\Delta]$, $j = 1, \dots, n+1$. It follows that

$$\mathbb{E}[V^\Delta, W]_t^2 = \frac{2}{\Delta} \left[\sum_{i=1}^{j-1} \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} (W_s - W_{(i-1)\Delta}) ds \right)^2 + \mathbb{E} \left(\int_{(j-1)\Delta}^t (W_s - W_{(j-1)\Delta}) ds \right)^2 \right] \quad (\text{A.28})$$

for $t \in [(j-1)\Delta, j\Delta]$, $j = 1, \dots, n+1$, due to the independent increment property of Brownian motion. Moreover, by Cauchy-Schwarz we have

$$\begin{aligned} \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} (W_s - W_{(i-1)\Delta}) ds \right)^2 &\leq \Delta \int_{(i-1)\Delta}^{i\Delta} \mathbb{E}(W_s - W_{(i-1)\Delta})^2 ds \\ &= \Delta \int_{(i-1)\Delta}^{i\Delta} (s - (i-1)\Delta) ds = \frac{\Delta^3}{2} \end{aligned} \quad (\text{A.29})$$

for $i = 1, \dots, n$. Therefore, it can be deduced from (A.28) and (A.29) that

$$\mathbb{E}[V^\Delta, W]_t^2 \leq \frac{2}{\Delta} \sum_{i=1}^n \mathbb{E} \left(\int_{(i-1)\Delta}^{i\Delta} (W_s - W_{(i-1)\Delta}) ds \right)^2 \leq \frac{2}{\Delta} n \frac{\Delta^3}{2} = \Delta T,$$

and that

$$\sup_{0 \leq t \leq T} \mathbb{E}[V^\Delta, W]_t^2 \rightarrow 0$$

in our asymptotic framework. This proves that V is independent of W .

For the second statement, note that V_T^Δ is actually a time changed Brownian motion $V_{[V^\Delta]_T}$ from the DDS Brownian motion representation and (A.27). We write

$$\frac{V_{[V^\Delta]_T} - V_T}{(\Delta T)^{1/4}} = U^{T,\Delta} \circ Z^{T,\Delta}$$

denoting

$$U_t^{T,\Delta} = \frac{V_{T+t\sqrt{\Delta T}} - V_T}{(\Delta T)^{1/4}}, \quad Z^{T,\Delta} = \frac{[V^\Delta]_T - T}{\sqrt{\Delta T}}.$$

Note that $(U_t^{T,\Delta})$ is a two-sided Brownian motion for all T and Δ from the scale invariance and time homogeneity, and trivially converges to a two-sided Brownian motion as $T \rightarrow \infty$ and $\Delta \rightarrow 0$. Furthermore we have $Z^{T,\Delta} = O_p(1)$ as $T \rightarrow \infty$ and $\Delta \rightarrow 0$ from (A.26), thus $U^{T,\Delta} \circ Z^{T,\Delta}$ is also $O_p(1)$. This is because for large T and small Δ , there exists M_1 such that

$$\mathbb{P}\{|Z^{T,\Delta}| \geq M_1\} \leq \varepsilon_1$$

for any $\varepsilon_1 > 0$ and also there exists M_2 such that

$$\mathbb{P}\left\{\sup_{t \in [-M_1, M_1]} |U_t^{T, \Delta}| / \sqrt{M_1} \geq M_2\right\} \leq \varepsilon_2$$

for any $\varepsilon_2 > 0$ and all large M_1 , thus there exist M_1 and M_2 such that

$$\mathbb{P}\left\{|U^{T, \Delta} \circ Z^{T, \Delta}| \geq M_2 \sqrt{M_1}\right\} \leq \varepsilon_1 + \varepsilon_2$$

for any $\varepsilon_1, \varepsilon_2 > 0$.

Proof of Lemma A4

For (a), we can deduce from Lemma A2 that

$$\begin{aligned} \sum_{i=1}^n f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) &= \int_0^T f(X_t) dW_t - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (f(X_t) - f(X_{(i-1)\Delta})) dW_t \\ &= \int_0^T f(X_t) dW_t - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right) (X_s) ds dW_t \\ &\quad - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma f^\cdot(X_s) dW_s dW_t \\ &= \int_0^T f(X_t) dW_t + O_p(\Delta T^{4pq+1/2}) + O_p(\sqrt{\Delta} T^{4pq+1/2}). \end{aligned}$$

For (b), we write

$$\begin{aligned} &\sqrt{\frac{2}{\Delta}} \sum_{i=1}^n f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_u dW_s \\ &= \int_0^T f(X_t) dV_t + \int_0^T f(X_t) d(V^\Delta - V)_t - \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (f(X_t) - f(X_{(i-1)\Delta})) dV_t^\Delta \\ &= \int_0^T f(X_t) dV_t + P_T + Q_T, \end{aligned}$$

and will show the order of P_T in Part 1, and the order of Q_T in Part 2.

Part 1 For P_T , we have

$$P_T = f(X_T)(V_T^\Delta - V_T) - \int_0^T (V_t^\Delta - V_t) df(X_t) - [f(X), (V^\Delta - V)]_T$$

from integration by parts. For the first term,

$$f(X_T)(V_T^\Delta - V_T) = O_p(T^{pq}) O_p((\Delta T)^{1/4}) = O_p(\Delta^{1/4} T^{pq+1/4}),$$

and for the second term,

$$\begin{aligned} \int_0^T (V_t^\Delta - V_t) df(X_t) &= \int_0^T (V_t^\Delta - V_t) \left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right) (X_t) dt + \int_0^T (V_t^\Delta - V_t) \sigma f^\cdot(X_t) dW_t \\ &= P_{1T} + P_{2T}. \end{aligned}$$

We can bound P_{1T} by

$$P_{1T} \leq \sqrt{\int_0^T (V_t^\Delta - V_t)^2 dt \int_0^T \left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right)^2 (X_t) dt} = O_p(\Delta^{1/4} T^{3/4}) O_p(T^{4pq+1}) = O_p(\Delta^{1/4} T^{4pq+7/4})$$

and P_{2T} is a martingale whose quadratic variation is given by

$$\int_0^T (V_t^\Delta - V_t)^2 \sigma^2 f^{\cdot\cdot}(X_t) dt \leq \sqrt{\int_0^T (V_t^\Delta - V_t)^4 dt \int_0^T \sigma^4 f^{\cdot\cdot}(X_t) dt} = O_p(\Delta^{1/2} T^{3/2}) O_p(T^{6pq+1/2}),$$

from which $P_{2T} = O_p(\Delta^{1/4} T^{3pq+1})$ follows. For the last term $[f(X), (V^\Delta - V)]_T$, since

$$f(X_t) = f(X_0) + \int_0^t \left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right) (X_s) ds + \int_0^t \sigma f^\cdot(X_s) dW_s$$

and W and V are independent of each other, $[f(X), (V^\Delta - V)]_T$ is the same as the quadratic covariation between

$$\int_0^t \sigma f^\cdot(X_s) dW_s, \quad V_t^\Delta = \sqrt{\frac{2}{\Delta}} \left(\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_u dW_s + \int_{(j-1)\Delta}^t \int_{(j-1)\Delta}^s dW_u dW_s \right).$$

Therefore we deduce

$$[f(X), (V^\Delta - V)]_T = \sqrt{\frac{2}{\Delta}} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \sigma f^\cdot(X_s) \int_{(i-1)\Delta}^s dW_u ds.$$

To obtain its order, note that

$$\begin{aligned} & \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f(X_s) \int_{(i-1)\Delta}^s dW_u ds \\ &= \sum_{i=1}^n f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_u ds + \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (f(X_s) - f(X_{(i-1)\Delta})) \int_{(i-1)\Delta}^s dW_u ds \\ &= P_{3T} + P_{4T}. \end{aligned}$$

We have $P_{3T} = O_p(\Delta T^{pq+1/2})$ from Lemma A2, and

$$P_{4T} \leq T \sup_{t \in [0, T]} \sup_{s \in [0, \Delta]} |f(X_{t+s}) - f(X_t)| \sup_{t \in [0, T]} \sup_{s \in [0, \Delta]} |W_{t+s} - W_t| = O_p(\Delta^{1-\delta} T^{2pq+1-\delta})$$

for any $\delta > 0$, so the order of quadratic covariation becomes

$$[f(X), (V^\Delta - V)]_T = O_p(\Delta^{1/2-\delta} T^{2pq+1-\delta}).$$

Since this is of smaller order than P_{1T} , we have $P_T = O_p(\Delta^{1/4} T^{4pq+7/4})$ as a result.

Part 2 For Q_T ,

$$\begin{aligned} Q_T &= \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right) (X_s) ds dV_t^\Delta + \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \sigma f^\cdot(X_s) dW_s dV_t^\Delta \\ &= Q_{1T} + Q_{2T} \end{aligned}$$

from Itô's lemma. For Q_{1T} , note that

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t f(X_s) ds dV_t^\Delta$$

is a martingale with a quadratic variation

$$\begin{aligned} & \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^t f(X_s) ds \right)^2 d[V^\Delta]_t \\ &= 2 \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t f(X_u) \int_{(i-1)\Delta}^u f(X_s) ds du d[V^\Delta]_t \\ &= 2 \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} ([V^\Delta]_{i\Delta} - [V^\Delta]_u) f(X_u) \int_{(i-1)\Delta}^u f(X_s) ds du \\ &\leq 2 \sqrt{\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} ([V^\Delta]_{i\Delta} - [V^\Delta]_u)^2 du \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f^2(X_u) \left(\int_{(i-1)\Delta}^u f(X_s) ds \right)^2 du} \\ &= Q_{11T} Q_{12T}. \end{aligned}$$

Since the order of $\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} ([V^\Delta]_{i\Delta} - [V^\Delta]_s)^2 ds$ is the same as the order of its expectation being a positive process, we can consider the order of the expectation instead. We have

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} ([V^\Delta]_{i\Delta} - [V^\Delta]_s)^2 ds \right) &= \mathbb{E} \left(\frac{4}{\Delta^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(\int_s^{i\Delta} (W_u - W_s)^2 du \right)^2 ds \right) \quad (\text{A.30}) \\ &= \mathbb{E} \left(\frac{4}{\Delta^2} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \mathbb{E}_{(i-1)\Delta} \left(\int_s^{i\Delta} (W_u - W_s)^2 du \right)^2 ds \right), \end{aligned}$$

where \mathbb{E}_t denotes a conditional expectation with information given up to time t , and since

$$\mathbb{E}_{(i-1)\Delta} \left(\int_s^{i\Delta} (W_u - W_s)^2 du \right)^2 \leq (i\Delta - s) \int_s^{i\Delta} \mathbb{E}_{(i-1)\Delta} (W_u - W_s)^4 du = (i\Delta - s)^4,$$

we have

$$\mathbb{E} \left(\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} ([V^\Delta]_{i\Delta} - [V^\Delta]_s)^2 ds \right) \leq 4\Delta^2 T$$

and $Q_{11T} = O_p(\Delta\sqrt{T})$. For Q_{12T} ,

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f^2(X_u) \left(\int_{(i-1)\Delta}^u f(X_s) ds \right)^2 du \leq \Delta^2 T \sup_{0 \leq t \leq T} |f^2(X_t)| \sup_{0 \leq t \leq T} |f(X_t)|^2 = O_p(\Delta^2 T^{4pq+1}),$$

so $Q_{1T} = O_p(\Delta\sqrt{T}) O_p(\Delta T^{2pq+1/2}) = O_p(\Delta^2 T^{2pq+1})$.

For Q_{2T} , note that

$$\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t f(X_s) dW_s dV_t^\Delta = \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (V_{i\Delta}^\Delta - V_s^\Delta) f(X_s) dW_s$$

changing the order of the integrals, and this is a martingale with a quadratic variation

$$\begin{aligned} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (V_{i\Delta}^\Delta - V_s^\Delta)^2 f^2(X_s) ds &\leq \sqrt{\sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} (V_{i\Delta}^\Delta - V_s^\Delta)^4 ds \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} f^4(X_s) ds} \\ &= O_p(\Delta\sqrt{T})O_p(T^{2pq+1/2}) \end{aligned}$$

since

$$\mathbb{E}_{(i-1)\Delta} \int_{(i-1)\Delta}^{i\Delta} (V_{i\Delta}^\Delta - V_s^\Delta)^4 ds = \frac{36}{5}\Delta^3,$$

thus we can check that Q_T is of smaller order than P_T .

Proof of Lemma A5

For the derivation, we utilize the results in Friedman (1964). Since some theorems in Friedman (1964) deals only with diffusions with bounded supports, we first transform (X_t) with a bounded function, derive asymptotics, and then back-transform them to obtain our desired statement. This is possible because we are only interested in infinitesimal properties of the transition density around $x = y$ and $t = 0$.

The transformation function f can be any bounded monotone function as long as it satisfies proper smoothness and boundary conditions.⁸ In this proof, we will use the logistic function $f(x) = 1/(1 + e^{-x})$ to avoid unnecessary complications in derivation. Denoting $Y_t = f(X_t)$, we have

$$dY_t = \mu^*(Y_t, \theta)dt + \sigma^*(Y_t, \beta)dW_t,$$

where

$$\mu^*(x, \theta) = \left((f \cdot \mu + \frac{1}{2} f \cdot \sigma^2) \circ f^{-1} \right)(x, \theta), \quad \sigma^*(x, \beta) = ((f \cdot \sigma) \circ f^{-1})(x, \beta).$$

With this transformation, (Y_t) is bounded on $(0, 1)$, and μ^* and σ^* are Hölder continuous with exponent $0 < \alpha < 1$, since $\mu \circ f^{-1}$ and $\sigma \circ f^{-1}$ are infinitely differentiable on the support of (Y_t) and slowly varying at both boundaries⁹ from the closure properties of regularly varying functions, together with $(f \cdot \sigma)(x) = x - x^2$ and $(f \cdot \mu)(x) = x - 3x^2 + 2x^3$. Thus the transformed diffusion (Y_t) satisfies the conditions (A_1) and (A_2) on pp. 3 and $(A_3)'$ on pp. 28 of Friedman (1964).

Hereafter we omit superscript $*$ for all the functions related with (Y_t) to simplify the notation. That is, we denote μ, σ, p and ℓ as the drift, diffusion, transition density and normalized likelihood functions of (Y_t) , respectively. Those functions for (X_t) are denoted as μ^o, σ^o, p^o and ℓ^o to avoid confusions. We maintain definitions of α and β as the same. Parameter arguments θ are omitted hereafter.

For the first step, we will derive infinitesimal properties of ℓ , the normalized likelihood of the transformed process (Y_t) . Under given conditions, the transition density as a fundamental solution of the partial differential equation $u_t(t, x) = \sigma^2(x)u_{xx}(t, x)/2 + \mu(x)u_x(t, x)$ is given by

$$p(t, x, y) = \bar{p}(t, x, y) + \int_0^t \int_D \bar{p}(t-s, w, y) q(s, x, w) dw ds \quad (\text{A.31})$$

⁸The conditions are (i) f is bounded, strictly monotone and four-times differentiable, and (ii) f^{-1} is slowly varying at the boundaries and $(\partial/\partial x)f^{-1}(x)$ is regularly varying with index $a > 1/2$ at the boundaries.

⁹We use a natural extension of the definition for regular variation, in the sense that we say $f(x)$ is regularly varying at the boundaries of (a, b) , if $f(a + 1/x)$ and $f(b - 1/x)$ are regularly varying for large x .

from Theorems 8 and 15 on pp. 19 and 28 of Friedman (1964), where $D = (0, 1)$,

$$\bar{p}(t, x, y) = \frac{1}{\sigma(x)\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t\sigma^2(x)}\right)$$

and q is a solution of

$$q(t, x, y) = \bar{q}(t, x, y) + \int_0^t \int_D \bar{q}(t-s, w, y) q(s, x, w) dw ds, \quad (\text{A.32})$$

where

$$\bar{q}(t, x, y) = \frac{1}{2}[\sigma^2(y) - \sigma^2(x)] \frac{\partial^2}{\partial y^2} \bar{p}(t, x, y) + \mu(y) \frac{\partial}{\partial y} \bar{p}(t, x, y).$$

This transition density satisfies the Kolmogorov forward equation,

$$p_t(t, x, y) = (\sigma'^2 + \sigma\sigma'' - \mu')(y)p(t, x, y) + (2\sigma\sigma' - \mu)(y)p_y(t, x, y) + \frac{1}{2}\sigma^2(y)p_{y^2}(t, x, y). \quad (\text{A.33})$$

In terms of the normalized log-likelihood $\ell(t, x, y) = t \log p(t, x, y) + t \log(\sqrt{t})$, this becomes

$$\begin{aligned} t\ell_t(t, x, y) - \ell(t, x, y) - \frac{t}{2} &= -t^2\mu'(y) - t\mu(y)\ell_y(t, x, y) + t^2\sigma'^2(y) + t^2\sigma\sigma''(y) + 2t\sigma\sigma'(y)\ell_y(t, x, y) \\ &\quad + \frac{t}{2}\sigma^2(y)\ell_{y^2}(t, x, y) + \frac{1}{2}\sigma^2(y)\ell_y^2(t, x, y). \end{aligned} \quad (\text{A.34})$$

Now we will derive infinitesimal properties of ℓ . From pp. 16 (4.9) and (4.15) of Friedman (1964),

$$\left| \int_0^t \int_D \bar{p}(t-s, w, x) q(s, x, w) dw ds \right| \leq C \quad (\text{A.35})$$

for some constant C , thus, $\sqrt{t}p(t, x, x) \rightarrow 1/(\sigma(x)\sqrt{2\pi})$ as $t \rightarrow 0$. Then it follows $\ell(t, x, x) = t \log(\sqrt{t}p(t, x, x)) \rightarrow 0$ as $t \rightarrow 0$. From (A.34), letting $y = x$ and taking Taylor expansion as

$$\ell(t, x, x) = t\ell_t(0, x, x) + t^2\ell_{t^2}(\tilde{t}, x, x)/2, \quad \ell_y(t, x, x) = t\ell_{ty}(0, x, x) + t^2\ell_{t^2y}(\tilde{t}, x, x)/2$$

for some $\tilde{t} \in [0, t]$, we can obtain $\ell_{y^2}(t, x, x) \rightarrow -1/\sigma^2(x)$ as $t \rightarrow 0$.

From (A.33), $\sqrt{t}p(t, x, x) = O(1)$, $\sqrt{t}p_y(t, x, x) = O(1)$, $t^{3/2}p_t(t, x, x) = O(1)$ and $t^{3/2}p_{y^2}(t, x, x) = O(1)$ as $t \rightarrow 0$ from (A.31), we should have

$$\lim_{t \rightarrow 0} (t^{3/2}p_t(t, x, x)) = \lim_{t \rightarrow 0} \left(t^{3/2} \frac{1}{2} \sigma^2 p_{y^2}(t, x, x) \right). \quad (\text{A.36})$$

Note that

$$\ell_{y^2}(t, x, x) = -t \frac{p_y^2(t, x, x)}{p^2(t, x, x)} + t \frac{p_{y^2}(t, x, x)}{p(t, x, x)} \rightarrow -\frac{1}{\sigma^2(x)}$$

and $tp_y^2(t, x, x)/p^2(t, x, x) \rightarrow 0$ as $t \rightarrow 0$, thus

$$t^{3/2}p_{y^2}(t, x, x) \rightarrow -\frac{1}{\sqrt{2\pi}\sigma^3(x)}, \quad t^{3/2}p_t(t, x, x) \rightarrow -\frac{1}{2\sqrt{2\pi}\sigma(x)}$$

from (A.36). Also,

$$\ell_t(t, x, x) = \log(\sqrt{t}p(t, x, x)) + \frac{1}{2} + t \frac{p_t(t, x, x)}{p(t, x, x)},$$

and since

$$t \frac{p_t(t, x, x)}{p(t, x, x)} \rightarrow -\frac{1}{2}, \quad \sqrt{t}p(t, x, x) \rightarrow \frac{1}{\sqrt{2\pi}\sigma(x)},$$

we have $\ell_t(0, x, x) = -\log(\sigma(x)) - \log(2\pi)/2$.

From (A.34) let $y = x$ and proceed one step further with the limits $\ell(0, x, x) = 0$, $\ell_y(0, x, x) = 0$, $\ell_{y^2}(0, x, x) = -1/\sigma^2(x)$ and $\ell_t(0, x, x) = -\log(\sigma(x)) - \log\sqrt{2\pi}$, and divide both sides with t^2 , then we have

$$\begin{aligned} \frac{1}{2}\ell_{t^2}(0, x, x) &= (2\sigma\sigma'(x) - \mu(x))\ell_{ty}(0, x, x) - \mu'(x) + \sigma'^2(x) + \sigma\sigma''(x) \\ &\quad + \frac{1}{2}\sigma^2(x)\ell_{ty}^2(0, x, x) + \frac{1}{2}\sigma^2(x)\ell_{ty^2}(0, x, x). \end{aligned} \quad (\text{A.37})$$

We can also take derivatives w.r.t. y on each side of (A.34), and divide them with t , then we have

$$\ell_{ty}(0, x, x) = \frac{\mu(x)}{\sigma^2(x)} + \frac{1}{2}\sigma^2(x)\ell_{y^3}(0, x, x) - \frac{3\sigma'(x)}{\sigma(x)}. \quad (\text{A.38})$$

Finally, take second derivatives w.r.t. y on both sides of (A.34), and divide them with t , then we have

$$\begin{aligned} 2\ell_{ty^2}(0, x, x) &= (4\sigma\sigma'(x) - \mu(x))\ell_{y^3}(0, x, x) - \frac{4\sigma'(x)}{\sigma(x)}\ell_{ty}(0, x, x) + \sigma^2(x)\ell_{ty}\ell_{y^3}(0, x, x) \\ &\quad + \frac{1}{2}\sigma^2(x)\ell_{y^4}(0, x, x) + \frac{2\mu'(x)}{\sigma^2(x)} - 5\left(\frac{\sigma'^2(x)}{\sigma^2(x)} + \frac{\sigma''(x)}{\sigma(x)}\right). \end{aligned} \quad (\text{A.39})$$

Arranging the equations (A.37) and (A.39), we obtain

$$\begin{aligned} \ell_{t^2}(0, x, x) + \sigma^2(x)\ell_{ty^2}(0, x, x) &= -\frac{\mu^2(x)}{\sigma^2(x)} - 2\sigma^3(x)\sigma'(x)\ell_{y^3}(0, x, x) + \frac{3}{4}\sigma^6(x)\ell_{y^3}^2(0, x, x) \\ &\quad + \frac{1}{2}\sigma^4(x)\ell_{y^4}(0, x, x) - 3\sigma\sigma''(x) + 6\sigma'^2(x). \end{aligned} \quad (\text{A.40})$$

Now we will show $\ell_{y^3}(0, x, x)$ does not depend on α . Since

$$\ell_{y^3}(t, x, x) = 2t \frac{p_y^3(t, x, x)}{p^3(t, x, x)} - 3t \frac{p_y p_{y^2}(t, x, x)}{p^2(t, x, x)} + t \frac{p_{y^3}(t, x, x)}{p(t, x, x)}$$

and we have $\sqrt{t}p_y(t, x, x) = O(1)$ and $t^{3/2}p_{y^3}(t, x, x) = O(1)$ from (A.31) together with

$$\sqrt{t}p(t, x, x) \rightarrow \frac{1}{\sigma(x)\sqrt{2\pi}}, \quad t^{3/2}p_{y^2}(t, x, x) \rightarrow -\frac{1}{\sigma^3(x)\sqrt{2\pi}},$$

it is enough to show that $\lim_{t \rightarrow 0} (\sigma^2(x)t^{3/2}p_{y^3}(t, x, x) + 3\sqrt{t}p_y(t, x, x))$ does not depend on α . From (A.31) and pp. 16 (4.14) of Friedman (1964),

$$\begin{aligned} p_y(t, x, x) &= \int_0^t \int_D \bar{p}_y(t-s, w, x) \bar{q}(s, x, w) dw ds + O(1), \\ p_{y^3}(t, x, x) &= \int_0^t \int_D \bar{p}_{y^3}(t-s, w, x) \bar{q}(s, x, w) dw ds + O(t^{-1}). \end{aligned}$$

We will denote $\bar{q}_\mu(t, x, y) = \mu(y)\bar{p}_y(t, x, y)$ excluding the part which does not depend on α from $\bar{q}(t, x, y)$ and will only consider this for the ease of calculation here. If we let

$$\begin{aligned}\bar{p}_{y,1}(t, x, y) &= \frac{x-y}{\sqrt{2\pi}t^{3/2}\sigma^3(y)} \exp\left(-\frac{(x-y)^2}{2t\sigma^2(y)}\right), \\ \bar{p}_{y^3,1}(t, x, y) &= \frac{(x-y)^3 - 3t(x-y)\sigma^2(y)}{\sqrt{2\pi}t^{7/2}\sigma^7(y)} \exp\left(-\frac{(x-y)^2}{2t\sigma^2(y)}\right), \\ \bar{q}_{\mu,2}(t, x, y) &= \frac{(x-y)\mu(x)}{\sqrt{2\pi}t^{3/2}\sigma^3(x)} \exp\left(-\frac{(x-y)^2}{2t\sigma^2(x)}\right),\end{aligned}\tag{A.41}$$

then the remainder terms becomes higher order from pp. 16 (4.14) of Friedman (1964), thus we have

$$\begin{aligned}p_{y,\mu}(t, x, x) &= \int_0^t \int_D \bar{p}_{y,1}(t-s, w, x) \bar{q}_{\mu,2}(s, x, w) dw ds + O(1), \\ p_{y^3,\mu}(t, x, x) &= \int_0^t \int_D \bar{p}_{y^3,1}(t-s, w, x) \bar{q}_{\mu,2}(s, x, w) dw ds + O(t^{-1})\end{aligned}$$

by denoting $p_{y,\mu}$ and $p_{y^3,\mu}$ as the parts related with α . Now from (A.41), we may deduce after some algebra that

$$\begin{aligned}\int_0^t \int_{-\infty}^{\infty} \bar{p}_{y,1}(t-s, w, x) \bar{q}_{\mu,2}(s, x, w) dw ds &= -\frac{\mu(x)}{t^{1/2}2\sqrt{2\pi}\sigma^3(x)}, \\ \int_0^t \int_{-\infty}^{\infty} \bar{p}_{y^3,1}(t-s, w, x) \bar{q}_{\mu,2}(s, x, w) dw ds &= \frac{3\mu(x)}{t^{3/2}2\sqrt{2\pi}\sigma^5(x)},\end{aligned}$$

thus for any $x \in D$,

$$\sqrt{t}p_{y,\mu}(t, x, x) \rightarrow -\frac{\mu(x)}{2\sqrt{2\pi}\sigma^3(x)}, \quad t^{3/2}p_{y^3,\mu}(t, x, x) \rightarrow \frac{3\mu(x)}{2\sqrt{2\pi}\sigma^5(x)}$$

and this leads to our desired result.

Finally, we show that $\ell_{y^4}(0, x, x)$ does not depend on α . Since the score is a martingale, we should have

$$\mathbb{E}(\ell_\alpha(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) | X_{(i-1)\Delta}) = 0.$$

Note that we have

$$\begin{aligned}\mathbb{E}(\ell_\alpha(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) | X_{(i-1)\Delta}) &= \Delta \mathcal{A}\ell_\alpha(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) + \frac{\Delta^2}{2} \mathcal{A}^2\ell_\alpha(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}) \\ &\quad + \frac{\Delta^3}{6} \mathbb{E}(\mathcal{A}^3\ell_\alpha(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) | X_{(i-1)\Delta})\end{aligned}$$

with $\mathcal{A}\ell_\alpha(0, x, x) = 0$, thus denoting

$$c_1 = \mathcal{A}^2\ell_\alpha(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}), \quad c_2(\Delta) = \mathbb{E}(\mathcal{A}^3\ell_\alpha(\Delta, X_{(i-1)\Delta}, X_{i\Delta}) | X_{(i-1)\Delta})$$

given $X_{(i-1)\Delta}$, we have $3c_1 + \Delta c_2(\Delta) = 0$. If $c_1 \neq 0$, $c_2(\Delta) = -3c_1/\Delta$, which is a contradiction from Assumption 3.1(b), thus together with (A.40), we have $\mathcal{A}^2\ell_\alpha(0, x, x) = \frac{3}{4}\sigma^4(x)\ell_{y^4\alpha}(0, x, x) = 0$ for all α .

Summarizing the results, we have

$$\begin{aligned}
\ell(0, x, x) &= 0, & \ell_y(0, x, x) &= 0, \\
\ell_t(0, x, x) &= -\log(\sigma(x)) - \log(\sqrt{2\pi}), & \ell_{y^2}(0, x, x) &= -\frac{1}{\sigma^2(x)}, \\
\ell_{ty}(0, x, x) &= \frac{\mu(x)}{\sigma^2(x)} - \frac{3\sigma'(x)}{\sigma(x)} + \frac{1}{2}\sigma^2(x)\ell_{y^3}(0, x, x), \\
\ell_{t^2}(0, x, x) + \ell_{ty^2}(0, x, x)\sigma^2(x) &= -\frac{\mu^2(x)}{\sigma^2(x)} - 2\sigma^3(x)\sigma'(x)\ell_{y^3}(0, x, x) + \frac{3}{4}\sigma^6(x)\ell_{y^3}^2(0, x, x) \\
&\quad + \frac{1}{2}\sigma^4(x)\ell_{y^4}(0, x, x) - 3\sigma\sigma''(x) + 6\sigma'^2(x)
\end{aligned} \tag{A.42}$$

with $\ell_{y^3}(0, x, x)$ and $\ell_{y^4}(0, x, x)$ not depending on α .

So far we have derived infinitesimal properties for the normalized likelihood of (Y_t) . In the next step, we will back-transform them to obtain the same kind of statements for ℓ^o , the normalized likelihood of (X_t) . Before proceeding, note that we have a relationship

$$\ell^o(t, x, y) = t \log(\sqrt{t}p^o(t, x, y)) = t \log(\sqrt{t}p[t, f(x), f(y)]f'(y)) = \ell(t, f(x), f(y)) + t \log(f'(y))$$

from the formula for functions of random variables, since a distribution of X_t given $X_0 = x$ is the same as the one of $f^{-1}(Y_t)$ given $f^{-1}(Y_0) = x$. From this relationship, we can derive $\ell^o(0, x, x) = 0$ and $\ell_{ty}^o(0, x, x) = 0$ from (A.42) ignoring terms unrelated with θ . We also have $\ell_t^o(0, x, x) = -\log(\sigma(f(x))) = -\log(f'(x)\sigma^o(x)) = -\log(\sigma^o(x))$ ignoring terms unrelated with θ , and $\ell_{y^2}^o(0, x, x) = f'^2(x)\ell_{y^2}(0, f(x), f(x)) = -1/\sigma^{o2}(x)$. For ℓ_{ty}^o ,

$$\begin{aligned}
\ell_{ty}^o(0, x, x) &= f'(x)\ell_{ty}(f(x), f(x), 0) = f'(x) \left(\frac{f'\mu^o + f''\sigma^{o2}/2}{f'^2\sigma^{o2}} \right)(x) + f'(x)v(f(x)) \\
&= \frac{\mu^o}{\sigma^{o2}}(x) + \frac{1}{2} \frac{f''}{f'}(x) + f'(x)v(f(x)),
\end{aligned}$$

where

$$v(x) = -\frac{3\sigma'(x)}{\sigma(x)} + \frac{1}{2}\sigma^2(x)\ell_{y^3}(0, x, x).$$

For $\ell_{t^2}^o + \ell_{ty^2}^o\sigma^{o2}$,

$$\begin{aligned}
&\ell_{t^2}^o(0, x, x) + \ell_{ty^2}^o(0, x, x)\sigma^{o2}(x) \\
&= f''(x)\sigma^{o2}(x)\ell_{ty}(0, f(x), f(x)) + \ell_{t^2}(0, f(x), f(x)) + \ell_{ty^2}(0, f(x), f(x))\sigma^2(f(x)) \\
&= -\frac{\mu^{o2}}{\sigma^{o2}}(x) + \frac{1}{4} \frac{f''^2\sigma^{o2}}{f'^2}(x) + f''(x)\sigma^{o2}(x)v(f(x)) + w(f(x)),
\end{aligned}$$

where v is defined above and

$$w(x) = -2\sigma^3(x)\sigma'(x)\ell_{y^3}(0, x, x) + \frac{3}{4}\sigma^6(x)\ell_{y^3}^2(0, x, x) + \frac{1}{2}\sigma^4(x)\ell_{y^4}(0, x, x) - 3\sigma\sigma''(x) + 6\sigma'^2(x).$$

The third line is by plugging the results of (A.42) in the second line and arranging them. Lastly,

$$\begin{aligned}
\ell_{y^3}^o(0, x, x) &= f'''(x)\ell_y(0, f(x), f(x)) + 3f'(x)f''(x)\ell_{y^2}(0, f(x), f(x)) + f'''(x)\ell_{y^3}(0, f(x), f(x)) \\
&= -\frac{3f'f''}{f'^2\sigma^{o2}}(x) + f'^3(x)\ell_{y^3}(0, f(x), f(x))
\end{aligned}$$

and

$$\begin{aligned}\ell_{y^4}^o(0, x, x) &= f^{\cdots\cdots}(x)\ell_y(0, f(x), f(x)) + (3f^{\cdots 2} + 4f^{\cdot}f^{\cdots\cdots})(x)\ell_{y^2}(0, f(x), f(x)) \\ &\quad + 6f^{\cdot 2}(x)f^{\cdots}(x)\ell_{y^3}(0, f(x), f(x)) + f^{\cdot 4}(x)\ell_{y^4}(0, f(x), f(x)) \\ &= -\left(\frac{3f^{\cdots 2} + 4f^{\cdot}f^{\cdots\cdots}}{f^{\cdot 2}\sigma^o 2}\right)(x) + 6f^{\cdot 2}(x)f^{\cdots}(x)\ell_{y^3}(0, f(x), f(x)) + f^{\cdot 4}(x)\ell_{y^4}(0, f(x), f(x)).\end{aligned}$$

Replacing μ^o , σ^o and ℓ^o with μ , σ and ℓ to recover original notations, we finally obtain the stated result of the lemma.

Appendix B. Proofs of Theorems

Proof of Lemma 3.1

For the exact transition density, we can derive the stated result from Lemma A5, or we can derive from (A.42)

$$\begin{aligned}\ell(0, x, x, \theta) &= 0, \quad \mathcal{A}\ell(0, x, x, \theta) = -\frac{\sigma^2(x)}{2\sigma^2(x, \beta)} - \log(\sigma(x, \beta)) - \log(\sqrt{2\pi}), \\ \mathcal{B}\ell(0, x, x, \theta) &= 0, \quad \mathcal{B}^2\ell(0, x, x, \theta) = -\frac{\sigma^2(x)}{\sigma^2(x, \beta)}, \\ \mathcal{A}^2\ell(0, x, x, \theta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + 2\mu(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} + (\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2}(0, x, x, \theta) \\ &\quad + 3(2\sigma^{\cdot 2}(x, \beta) - \sigma\sigma^{\cdots}(x, \beta)) - \frac{\mu^2(x)}{\sigma^2(x, \beta)} - \frac{\mu(x)}{\sigma^2(x, \beta)}(6\sigma\sigma^{\cdot}(x, \beta) + \sigma\sigma^{\cdot}(x)) \\ &\quad - \frac{\sigma^2(x)}{2\sigma^2(x, \beta)}(2\mu^{\cdot}(x) + \sigma^{\cdot 2}(x)) - \frac{\sigma^3\sigma^{\cdots}(x)}{2\sigma^2(x, \beta)} \\ &\quad + (\sigma^3\sigma^{\cdot}(x) + \mu\sigma^2(x) + \mu(x)\sigma^2(x, \beta) - 2\sigma^3\sigma^{\cdot}(x, \beta))\ell_{y^3}(0, x, x, \beta) \\ &\quad + \frac{3}{4}\sigma^6(x, \beta)\ell_{y^3}^2(0, x, x, \beta) + \frac{1}{4}(2\sigma^4(x, \beta) + \sigma^4(x))\ell_{y^4}(0, x, x, \beta), \\ \mathcal{A}\mathcal{B}\ell(0, x, x, \theta) &= \mathcal{B}\mathcal{A}\ell(0, x, x, \theta) \\ &= \sigma(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{\sigma(x)}{\sigma^2(x, \beta)}(\sigma\sigma^{\cdot}(x) + \mu(x) + 3\sigma\sigma^{\cdot}(x, \beta)) \\ &\quad + \frac{1}{2}\sigma(x)(\sigma^2(x) + \sigma^2(x, \beta))\ell_{y^3}(0, x, x, \beta), \\ \mathcal{B}^3\ell(0, x, x, \theta) &= -\frac{3\sigma^2\sigma^{\cdot}(x)}{\sigma^2(x, \beta)} + \sigma^3(x)\ell_{y^3}(0, x, x, \beta)\end{aligned}\tag{B.43}$$

with $\ell_{y^3}(0, x, x, \beta)$ and $\ell_{y^4}(0, x, x, \beta)$ not depending on α , thus we can check that each term satisfies the stated result. For other approximated ML estimators of diffusion models of which the approximated transition density is given by a function of $\mu(x, \alpha)$ and $\sigma(x, \beta)$, we can utilize symbolic math softwares such as *Mathematica* or *Maple* to show the statements in Lemma A5. For the Gaussian quasi-ML estimators (11), (14) and (18) based on the Euler, Milstein QML and Kessler

approximations respectively, we have

$$\begin{aligned}
\ell(0, x, x, \theta) &= 0, & \ell_y(0, x, x, \theta) &= 0, \\
\ell_t(0, x, x, \theta) &= -\log(\sigma(x, \beta)) - \log(\sqrt{2\pi}), & \ell_{y^2}(0, x, x, \theta) &= -\frac{1}{\sigma^2(x, \beta)}, \\
\ell_{ty}(0, x, x, \theta) &= \frac{\mu(x, \alpha)}{\sigma^2(x, \beta)}, & & \\
\ell_{t^2}(0, x, x, \theta) + \ell_{ty^2}(0, x, x, \theta)\sigma^2(x, \beta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)}, & & \\
\ell_{y^3}(0, x, x, \theta) &= 0, & \ell_{y^4}(0, x, x, \theta) &= 0
\end{aligned} \tag{B.44}$$

as a result. For the Milstein ML estimator (13), we obtain $\ell(0, x, x, \theta)$, $\ell_y(0, x, x, \theta)$, $\ell_t(0, x, x, \theta)$ and $\ell_{y^2}(0, x, x, \theta)$ as the same as (B.44) and

$$\begin{aligned}
\ell_{ty}(0, x, x, \theta) &= \frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{3\sigma^\cdot(x, \beta)}{2\sigma(x, \beta)}, & & \\
\ell_{t^2}(0, x, x, \theta) + \ell_{ty^2}(0, x, x, \theta)\sigma^2(x, \beta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + \frac{9}{4}\sigma^\cdot{}^2(x, \beta), & & \\
\ell_{y^3}(0, x, x, \theta) &= \frac{3\sigma^\cdot(x, \beta)}{\sigma^3(x, \beta)}, & \ell_{y^4}(0, x, x, \theta) &= -\frac{15\sigma^\cdot{}^2(x, \beta)}{\sigma^4(x, \beta)}.
\end{aligned} \tag{B.45}$$

For Ait-Sahalia's estimator in (17), we obtain $\ell(0, x, x, \theta)$, $\ell_y(0, x, x, \theta)$, $\ell_t(0, x, x, \theta)$ and $\ell_{y^2}(0, x, x, \theta)$ as the same as (B.44) and

$$\begin{aligned}
\ell_{ty}(0, x, x, \theta) &= \frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{3\sigma^\cdot(x, \beta)}{2\sigma(x, \beta)}, & & \\
\ell_{t^2}(0, x, x, \theta) + \ell_{ty^2}(0, x, x, \theta)\sigma^2(x, \beta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + \frac{5}{4}\sigma^\cdot{}^2(x, \beta) - \sigma(x, \beta)\sigma^\cdot{}^\cdot(x, \beta), & & \\
\ell_{y^3}(0, x, x, \theta) &= \frac{3\sigma^\cdot(x, \beta)}{\sigma^3(x, \beta)}, & \ell_{y^4}(0, x, x, \theta) &= -\frac{11\sigma^\cdot{}^2(x, \beta)}{\sigma^4(x, \beta)} + \frac{4\sigma^\cdot{}^\cdot(x, \beta)}{\sigma^3(x, \beta)}.
\end{aligned} \tag{B.46}$$

The derivation is basically an algebra involving differentiations and taking limits, and the Mathematica codes showing these steps will be provided separately upon request.¹⁰ With these results, for the Gaussian quasi-ML estimators (11), (14) and (18), we can derive from (B.44) that $\ell(0, x, x, \theta)$, $\mathcal{A}\ell(0, x, x, \theta)$, $\mathcal{B}\ell(0, x, x, \theta)$ and $\mathcal{B}^2\ell(0, x, x, \theta)$ are the same as (B.43) and

$$\begin{aligned}
\mathcal{A}^2\ell(0, x, x, \theta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + 2\mu(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} + (\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2}(0, x, x, \theta) \\
&\quad - \frac{\mu(x)}{\sigma^2(x, \beta)}(\mu(x) + \sigma\sigma^\cdot(x)) - \frac{\sigma^2(x)}{2\sigma^2(x, \beta)}(2\mu^\cdot(x) + \sigma^\cdot{}^2(x) + \sigma\sigma^\cdot{}^\cdot(x)), \\
\mathcal{A}\mathcal{B}\ell(0, x, x, \theta) &= \mathcal{B}\mathcal{A}\ell(0, x, x, \theta) \\
&= \sigma(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{\sigma(x)}{\sigma^2(x, \beta)}(\sigma\sigma^\cdot(x) + \mu(x)), \\
\mathcal{B}^3\ell(0, x, x, \theta) &= -\frac{3\sigma^2\sigma^\cdot(x)}{\sigma^2(x, \beta)}.
\end{aligned}$$

¹⁰One can visit <http://mypage.iu.edu/~jeongm/> for the codes.

For the Milstein ML estimator (13), we can derive from (B.45) that $\ell(0, x, x, \theta)$, $\mathcal{A}\ell(0, x, x, \theta)$, $\mathcal{B}\ell(0, x, x, \theta)$ and $\mathcal{B}^2\ell(0, x, x, \theta)$ are the same as (B.43) and

$$\begin{aligned}
\mathcal{A}^2\ell(0, x, x, \theta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + 2\mu(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} + (\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2}(0, x, x, \theta) \\
&\quad - \frac{\mu(x)}{\sigma^3(x, \beta)}(3\sigma^2\sigma'(x, \beta) - 3\sigma^2(x)\sigma'(x, \beta) + \sigma\sigma'(x)\sigma(x, \beta)) \\
&\quad - \frac{\sigma^2(x)}{2\sigma^2(x, \beta)}(2\mu'(x) + \sigma'^2(x)) - \frac{\sigma^3(x)}{2\sigma^3(x, \beta)}(\sigma''(x)\sigma(x, \beta) - 6\sigma'(x)\sigma'(x, \beta)) \\
&\quad - \frac{1}{4\sigma^4(x, \beta)}(15\sigma^4(x)\sigma'^2(x, \beta) + 4\mu^2(x)\sigma^2(x, \beta) - 9\sigma^4\sigma'^2(x, \beta)), \\
\mathcal{AB}\ell(0, x, x, \theta) &= \mathcal{BA}\ell(0, x, x, \theta) \\
&= \sigma(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{\sigma^2(x)}{2\sigma^3(x, \beta)}(2\sigma'(x)\sigma(x, \beta) - 3\sigma(x)\sigma'(x, \beta)) \\
&\quad - \frac{\sigma(x)}{2\sigma^3(x, \beta)}(2\mu(x)\sigma(x, \beta) + 3\sigma^2\sigma'(x, \beta)), \\
\mathcal{B}^3\ell(0, x, x, \theta) &= -\frac{3\sigma^2\sigma'(x)}{\sigma^2(x, \beta)} + 3\sigma^3(x)\frac{\sigma'(x, \beta)}{\sigma^3(x, \beta)}.
\end{aligned}$$

For Ait-Sahalia's estimator in (17), we can derive from (B.46) that $\ell(0, x, x, \theta)$, $\mathcal{A}\ell(0, x, x, \theta)$, $\mathcal{B}\ell(0, x, x, \theta)$ and $\mathcal{B}^2\ell(0, x, x, \theta)$ are the same as (B.43) and

$$\begin{aligned}
\mathcal{A}^2\ell(0, x, x, \theta) &= -\frac{\mu^2(x, \alpha)}{\sigma^2(x, \beta)} + 2\mu(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} + (\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2}(0, x, x, \theta) \\
&\quad - \frac{\mu(x)}{\sigma^3(x, \beta)}(3\sigma^2\sigma'(x, \beta) - 3\sigma^2(x)\sigma'(x, \beta) + \sigma\sigma'(x)\sigma(x, \beta)) \\
&\quad - \frac{\sigma^2(x)}{2\sigma^2(x, \beta)}(2\mu'(x) + \sigma'^2(x)) - \frac{\sigma^3(x)}{2\sigma^3(x, \beta)}(\sigma''(x)\sigma(x, \beta) - 6\sigma'(x)\sigma'(x, \beta)) \\
&\quad + \frac{1}{4\sigma^4(x, \beta)}(\sigma^4(x)[4\sigma\sigma''(x, \beta) - 11\sigma'^2(x, \beta)] - 4\mu^2(x)\sigma^2(x, \beta)) \\
&\quad - \frac{1}{4}(4\sigma\sigma'' - 5\sigma'^2)(x, \beta), \\
\mathcal{AB}\ell(0, x, x, \theta) &= \mathcal{BA}\ell(0, x, x, \theta) \\
&= \sigma(x)\frac{\mu(x, \alpha)}{\sigma^2(x, \beta)} - \frac{\sigma^2(x)}{2\sigma^3(x, \beta)}(2\sigma'(x)\sigma(x, \beta) - 3\sigma(x)\sigma'(x, \beta)) \\
&\quad - \frac{\sigma(x)}{2\sigma^3(x, \beta)}(2\mu(x)\sigma(x, \beta) + 3\sigma^2\sigma'(x, \beta)), \\
\mathcal{B}^3\ell(0, x, x, \theta) &= -\frac{3\sigma^2\sigma'(x)}{\sigma^2(x, \beta)} + 3\sigma^3(x)\frac{\sigma'(x, \beta)}{\sigma^3(x, \beta)}.
\end{aligned}$$

We can check that each term satisfies the stated result.

Proof of Lemma 3.2

We will only show the derivation for the score of the drift term since other cases can be driven with a similar way. Here, all the functions are evaluated at θ_0 and the arguments are omitted for the

simplicity. For the score term of α , we can apply Itô's lemma subsequently to get

$$\begin{aligned}
\mathcal{S}_\alpha(\theta_0) &= \frac{1}{\Delta} \sum_{i=1}^n \ell_\alpha(\Delta, x_i, y_i) \\
&= \frac{1}{\Delta} \sum_{i=1}^n \ell_\alpha(0, x_i, x_i) + \sum_{i=1}^n \mathcal{A}\ell_\alpha(0, x_i, x_i) + \frac{1}{\Delta} \sum_{i=1}^n \mathcal{B}\ell_\alpha(0, x_i, x_i) \mathcal{W}_{1i} + \frac{\Delta}{2} \sum_{i=1}^n \mathcal{A}^2 \ell_\alpha(0, x_i, x_i) \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \mathcal{B}\mathcal{A}\ell_\alpha(0, x_i, x_i) \mathcal{W}_{2i} + \frac{1}{\Delta} \sum_{i=1}^n \mathcal{A}\mathcal{B}\ell_\alpha(0, x_i, x_i) \mathcal{W}_{3i} + \frac{1}{\Delta} \sum_{i=1}^n \mathcal{B}^2 \ell_\alpha(0, x_i, x_i) \mathcal{W}_{4i} \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \mathcal{B}^3 \ell_\alpha(0, x_i, x_i) \mathcal{W}_{5i} + R,
\end{aligned}$$

where $\mathcal{W}_{1i} = W_{i\Delta} - W_{(i-1)\Delta}$, $\mathcal{W}_{2i} = \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_r ds$, $\mathcal{W}_{3i} = \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dr dW_s$, $\mathcal{W}_{4i} = \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s dW_r dW_s$ and $\mathcal{W}_{5i} = \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^s \int_{(i-1)\Delta}^r dW_u dW_r dW_s$, and

$$\begin{aligned}
R &= \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{A}^3 \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dr ds dt \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{B}\mathcal{A}^2 \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dr ds dW_t \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{A}\mathcal{B}\mathcal{A} \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dr dW_s dt \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{B}^2 \mathcal{A} \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dr dW_s dW_t \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{A}^2 \mathcal{B} \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dW_r ds dt \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{B}\mathcal{A}\mathcal{B} \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dW_r ds dW_t \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \mathcal{A}\mathcal{B}^2 \ell_\alpha(r - (i-1)\Delta, X_{(i-1)\Delta}, X_r) dW_r dW_s dt \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \int_{(i-1)\Delta}^r \mathcal{A}\mathcal{B}^3 \ell_\alpha(u - (i-1)\Delta, X_{(i-1)\Delta}, X_u) dW_u dW_r dW_s dt \\
&\quad + \frac{1}{\Delta} \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^t \int_{(i-1)\Delta}^s \int_{(i-1)\Delta}^r \mathcal{B}^4 \ell_\alpha(u - (i-1)\Delta, X_{(i-1)\Delta}, X_u) dW_u dW_r dW_s dW_t.
\end{aligned}$$

The order of the remainder term can be shown $R = O_p(\sqrt{\Delta} T^{4pq+1})$ from Lemma A2. Note that we have $\ell_\alpha(0, x, x) = 0$, $\mathcal{A}\ell_\alpha(0, x, x) = 0$, $\mathcal{B}\ell_\alpha(0, x, x) = 0$, $\mathcal{A}^2 \ell_\alpha(0, x, x) = 0$, $\mathcal{B}^2 \ell_\alpha(0, x, x) = 0$, $\mathcal{B}^3 \ell_\alpha(0, x, x) = 0$ and $\mathcal{A}\mathcal{B}\ell_\alpha(0, x, x) = \mathcal{B}\mathcal{A}\ell_\alpha(0, x, x) = \mu_\alpha(x)/\sigma(x)$ from Lemma A5. Also note that $\mathcal{W}_{2i} = \frac{\Delta}{2}(W_{i\Delta} - W_{(i-1)\Delta}) + \frac{\Delta}{2\sqrt{3}}(Z_{i\Delta} - Z_{(i-1)\Delta})$ and $\mathcal{W}_{3i} = \frac{\Delta}{2}(W_{i\Delta} - W_{(i-1)\Delta}) - \frac{\Delta}{2\sqrt{3}}(Z_{i\Delta} - Z_{(i-1)\Delta})$,

where Z is a standard Brownian motion independent of W . Thus from Lemma A4(a), we have

$$\begin{aligned}\mathcal{S}_\alpha(\theta_0) &= \frac{1}{2} \sum_{i=1}^n (\mathcal{A}\mathcal{B} + \mathcal{B}\mathcal{A})\ell_\alpha(0, X_{(i-1)\Delta}, X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + O_p(\sqrt{\Delta}T^{4pq+1}) \\ &= \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t + O_p(\sqrt{\Delta}T^{4pq+1}).\end{aligned}$$

Proof of Proposition 3.3

It is well known that

$$\frac{1}{T} \int_0^T f(X_t)dt \rightarrow_{a.s.} \pi(f)$$

as $T \rightarrow \infty$ for positive recurrent (X_t) when $\pi(f) < \infty$. (See, e.g., Theorem V.53.1 and (V.53.5) of Rogers and Williams (2000).) For the second statement, from Theorem 4.1 of van Zanten (2000), we have $K_T M_T \rightarrow_d \mathbb{N}(0, \Sigma)$ as $T \rightarrow \infty$, for non-random invertible matrix sequence K_T such that $\|K_T\| \rightarrow 0$ and $K_T[M]_T K_T' \rightarrow_p \Sigma \geq 0$, where M_T is a vector continuous local martingale and $[M]_T$ is its quadratic variation. Thus it directly follows that as $T \rightarrow \infty$,

$$\frac{1}{\sqrt{T}} \int_0^T g(X_t)dW_t \rightarrow_d \mathbb{N}(0, \pi(gg'))$$

since $T^{-1} \int_0^T (gg')(X_t)dt \rightarrow_{a.s.} \pi(gg')$.

Proof of Proposition 3.4

In the proof, we assume that the required scale transformation has already been done and the process (X_t) is a driftless diffusion with speed density m . We will therefore denote $m = m_r$ and suppress the subscript and superscript “ s ” used in the preliminary scale transformation. We set $T_r = T^{1/(r+2)}$ for $r > -1$ throughout the proof. Note that $T_r/T \rightarrow 0$ as $T \rightarrow \infty$. Define a stopping time τ by

$$\tau_t = \inf \left\{ s \mid \int_{\mathbb{R}} l(s, x)m(x)dx > t \right\},$$

where l is the local time of the Brownian motion B . We have

$$X = B \circ \tau,$$

due to Theorem 47.1 of Rogers and Williams (2000, p. 277).

We define

$$X_t^T = \frac{X_{Tt}}{T_r}, \quad B_t^{rT} = \frac{B_{T_r^2 t}}{T_r}, \quad \tau_t^T = \frac{\tau_{Tt}}{T_r^2} \quad (\text{B.47})$$

for $t \in [0, 1]$, so that

$$X^T = B^{rT} \circ \tau^T. \quad (\text{B.48})$$

It follows that

$$\begin{aligned}\frac{1}{2\varepsilon} \int_0^{T_r^2 t} 1\{|B_s - x| < \varepsilon\}ds &= \frac{T_r^2}{2\varepsilon} \int_0^t 1\{|B_{T_r^2 s} - x| < \varepsilon\}ds \\ &= \frac{T_r^2}{2\varepsilon} \int_0^t 1\left\{T_r \left|B_s^{rT} - \frac{x}{T_r}\right| < \varepsilon\right\}ds \\ &= T_r \frac{T_r}{2\varepsilon} \int_0^t 1\left\{\left|B_s^{rT} - \frac{x}{T_r}\right| < \frac{\varepsilon}{T_r}\right\}ds.\end{aligned}$$

Therefore, if we define l_r to be the local time of B^{rT} , then we have

$$l(T_r^2 t, x) = T_r l_r(t, x/T_r)$$

by taking the limit $\varepsilon \rightarrow 0$. Furthermore, we have

$$\begin{aligned} \tau_{Tt} &= \inf \left\{ s \left| \int_{\mathbb{R}} m(x) l(s, x) dx > Tt \right. \right\} \\ &= T_r^2 \inf \left\{ s \left| \int_{\mathbb{R}} m(x) l(T_r^2 s, x) dx > Tt \right. \right\} \\ &= T_r^2 \inf \left\{ s \left| T_r^2 \int_{\mathbb{R}} m\left(T_r \frac{x}{T_r}\right) l_r\left(s, \frac{x}{T_r}\right) d\left(\frac{x}{T_r}\right) > Tt \right. \right\} \\ &= T_r^2 \inf \left\{ s \left| \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r(s, x) dx > t \right. \right\} \end{aligned}$$

with $m_T(x) = m(T_r x)$, from which it follows that

$$\tau_t^T = \inf \left\{ s \left| \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r(s, x) dx > t \right. \right\}, \quad (\text{B.49})$$

due to our definition in (B.47).

Now we show that

$$\tau^T \rightarrow \tau^r \quad (\text{B.50})$$

almost surely as $T \rightarrow \infty$, in the space $D[0, 1]$ of cadlag functions on $[0, 1]$ endowed with Skorohod topology. Note that we may write

$$|\varepsilon_r(x)| \leq |\varepsilon_r(x)| 1\{|x| \leq M\} + m^*(x) n(x) 1\{|x| > M\}$$

for $M > 0$ such that $M \rightarrow \infty$ and $M/T_r \rightarrow 0$ as $T \rightarrow \infty$, where n is symmetric, bounded and monotonically decreasing to 0 as $|x| \rightarrow \infty$, and we have

$$\left| \frac{T_r^2}{T} m(T_r x) - m^*(x) \right| \leq \frac{T_r^2}{T} |\varepsilon_r(T_r x)| 1\{|T_r x| \leq M\} + m^*(x) n(T_r x) 1\{|T_r x| > M\}. \quad (\text{B.51})$$

Note in particular that $T_r^{r+2} = T$ and $m^*(T_r x) = T_r^r m^*(x)$.

For the first term in (B.51), we have

$$\begin{aligned} & \frac{T_r^2}{T} \int_{|x| \leq M/T_r} |\varepsilon_r(T_r x)| l_r(s, x) dx \\ &= l_r(s, 0) \frac{T_r^2}{T} \int_{|x| \leq M/T_r} |\varepsilon_r(T_r x)| dx + \frac{T_r^2}{T} \int_{|x| \leq M/T_r} |\varepsilon_r(T_r x)| |l_r(s, x) - l_r(s, 0)| dx \\ &\leq l_r(s, 0) \frac{T_r}{T} \int_{|x| \leq M} |\varepsilon_r(x)| dx + \lambda\left(\frac{M}{T_r}\right) \sqrt{l_r(s, 0)} \frac{T_r}{T} \int_{|x| \leq M} |\varepsilon_r(x)| dx \\ &\leq a_T + b_T l_r(s, 0) \end{aligned} \quad (\text{B.52})$$

for all large T , where $\lambda(z) = 2\sqrt{2z \log \log 1/z}$, and a_T and b_T are nonrandom numerical sequences such that $a_T, b_T \rightarrow 0$ as $T \rightarrow \infty$. The second inequality in (B.52) follows from the property of

Brownian local time in, e.g., Borodin (1989, p.20), and for the third inequality we use $\sqrt{l_r(s, 0)} \leq l_r(s, 0) + 1$ and

$$\frac{T_r}{T} \int_{|x| \leq M} |\varepsilon_r(x)| dx = \frac{T_r}{T} O(M^{r+1}) = O((M/T_r)^{r+1}) \rightarrow 0$$

as $T \rightarrow \infty$. For the second term in (B.51), we have

$$\int_{|x| > M/T_r} m^*(x) n(T_r x) l_r(s, x) dx \leq c_T \int_{\mathbb{R}} m^*(x) l_r(s, x) dx, \quad (\text{B.53})$$

where

$$c_T = \sup_{|x| > M/T_r} n(T_r x) \rightarrow 0,$$

since $M \rightarrow \infty$ as $T \rightarrow \infty$. Therefore, it follows from (B.51), (B.52) and (B.53) that

$$\left| \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r(s, x) dx - \int_{\mathbb{R}} m^*(x) l_r(s, x) dx \right| \leq a_T + b_T l_r(s, 0) + c_T \int_{\mathbb{R}} m^*(x) l_r(s, x) dx \quad (\text{B.54})$$

for some nonrandom numerical sequences a_T, b_T and c_T such that $a_T, b_T, c_T \rightarrow 0$ as $T \rightarrow \infty$.

Now we set d_T to be a sequence of numbers such that $d_T \rightarrow 0$ and

$$\frac{1}{b_T} \int_{|x| \leq d_T} m^*(x) dx \rightarrow \infty \quad (\text{B.55})$$

as $T \rightarrow \infty$. We write

$$\int_{|x| \leq d_T} m^*(x) l_r(s, x) dx = l_r(s, 0) \int_{|x| \leq d_T} m^*(x) dx + \int_{|x| \leq d_T} m^*(x) (l_r(s, x) - l_r(s, 0)) dx,$$

and note that we have

$$\begin{aligned} \left| \int_{|x| \leq d_T} m^*(x) (l_r(s, x) - l_r(s, 0)) dx \right| &\leq \lambda_{d_T} \sqrt{l_r(s, 0)} \int_{|x| \leq d_T} m^*(x) dx \\ &\leq \lambda_{d_T} (1 + l_r(s, 0)) \int_{|x| \leq d_T} m^*(x) dx \end{aligned}$$

with $\lambda_{d_T} = 2\sqrt{2}\sqrt{d_T \log \log(1/d_T)}$ for all large T . Therefore, we have

$$\begin{aligned} \int_{|x| \leq d_T} m^*(x) l_r(s, x) dx &\geq l_r(s, 0) \int_{|x| \leq d_T} m^*(x) dx - \lambda_{d_T} (1 + l_r(s, 0)) \int_{|x| \leq d_T} m^*(x) dx \\ &= \left((1 - \lambda_{d_T}) l_r(s, 0) - \lambda_{d_T} \right) \int_{|x| \leq d_T} m^*(x) dx \end{aligned}$$

for all large T , from which it follows that

$$l_r(s, 0) \leq \frac{\int_{\mathbb{R}} m^*(x) l_r(s, x) dx}{(1 - \lambda_{d_T}) \int_{|x| \leq d_T} m^*(x) dx} + \frac{\lambda_{d_T}}{1 - \lambda_{d_T}} \quad (\text{B.56})$$

for large T . Consequently, we may deduce from (B.54) and (B.56) that

$$\begin{aligned}
& \left| \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r(s, x) dx - \int_{\mathbb{R}} m^*(x) l_r(s, x) dx \right| \\
& \leq \left(a_T + \frac{b_T \lambda_{d_T}}{1 - \lambda_{d_T}} \right) + \left(\frac{b_T}{(1 - \lambda_{d_T}) \int_{|x| \leq d_T} m_r(x) dx} + c_T \right) \int_{\mathbb{R}} m^*(x) l_r(s, x) dx \\
& \leq \epsilon + \epsilon \int_{\mathbb{R}} m^*(x) l_r(s, x) dx.
\end{aligned} \tag{B.57}$$

for any $\epsilon > 0$ if T is sufficiently large.

It follows from (B.57) that

$$-\epsilon + (1 - \epsilon) \int_{\mathbb{R}} m^*(x) l_r(s, x) dx \leq \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r(s, x) dx \leq \epsilon + (1 + \epsilon) \int_{\mathbb{R}} m^*(x) l_r(s, x) dx$$

for all $s \geq 0$. Therefore, we have

$$t < (1 - \epsilon) \int_{\mathbb{R}} m^*(x) l_r \left(\tau_{t/(1-\epsilon)}^r + \epsilon, x \right) dx \leq \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r \left(\tau_{t/(1-\epsilon)}^r + \epsilon, x \right) dx + \epsilon,$$

from which and (B.49) it follows that

$$\tau_{t-\epsilon}^T < \tau_{t/(1-\epsilon)}^r + \epsilon \tag{B.58}$$

for all $t > 0$. Moreover, we have

$$t = (1 + \epsilon) \int_{\mathbb{R}} m^*(x) l_r \left(\tau_{t/(1+\epsilon)}^r, x \right) dx \geq \frac{T_r^2}{T} \int_{\mathbb{R}} m_T(x) l_r \left(\tau_{t/(1+\epsilon)}^r, x \right) dx,$$

and we may deduce from (B.49) that

$$\tau_{t/(1+\epsilon)}^r < \tau_{t+\epsilon}^T. \tag{B.59}$$

It is obvious that (B.58) and (B.59) imply (B.50).

On the other hand, we have $B^{rT} =_d B$ for all T , due to the scale invariance property of Brownian motion. Therefore, we have

$$B^{rT} \rightarrow_d B$$

trivially as $T \rightarrow \infty$, which together with (B.48) and (B.50) implies

$$X^T \rightarrow_d X^\circ \tag{B.60}$$

as $T \rightarrow \infty$.

For the convergence of W^T , we note that

$$\mathbb{E} |W_t^T - W_s^T|^4 \leq 3|t - s|^{3/2} \tag{B.61}$$

for all $t, s \geq 0$ and $T > 0$ and the Kolmogorov's criterion for weakly relatively compactness is satisfied, since $W_t^T = T^{-1/2} W_{Tt}$ is a Brownian motion for each T . With this condition satisfied, it

suffices to establish the convergence of the finite dimensional distribution. For each $t \geq 0$, we rewrite W as $dW_t = m^{1/2}(X_t)dX_t$. Then it follows that

$$\begin{aligned} \frac{W_{Tt}}{\sqrt{T}} &= \frac{1}{\sqrt{T}} \int_0^{Tt} m^{1/2}(X_s)dX_s \\ &= \frac{1}{\sqrt{T}} \int_0^t m^{1/2}(X_{Ts})dX_{Ts} \\ &= \frac{1}{\sqrt{T_r}} \int_0^t m^{1/2}(T_r X_s^T)dX_s^T \\ &= \int_0^t m^{*1/2}(X_s^T)dX_s^T + \frac{1}{\sqrt{T_r}} \int_0^t R(T_r X_s^T)dX_s^T \end{aligned} \quad (\text{B.62})$$

denoting $R(x) = \sqrt{m^*(x) + \varepsilon_r(x)} - \sqrt{m^*(x)}$, where the second line follows from change of variables and the last line from Definition 3.2.

The second term of (B.62) is a martingale whose quadratic variation is

$$\frac{1}{T_r} \int_0^t R^2(T_r X_s^T)d[X^T]_s = \int_0^t \frac{R^2(T_r X_s^T)}{m(X_{Ts})}ds = \frac{1}{T_r} \int_0^{\tau_t^T} R_T^2(W_s^{rT})ds \quad (\text{B.63})$$

denoting $R_T(x) = R(T_r x)$, where we obtain the second equality following the same step as in (B.65), from the second line to the last line. Let \bar{n} be defined as in (B.67). Then we can find Q such that

$$\begin{aligned} R_T^2(x) &= 2T_r^r m^*(x) + \varepsilon_r(T_r x) - 2\sqrt{T_r^{2r} m^{*2}(x) + T_r^r m^*(x)\varepsilon_r(T_r x)} \\ &\leq T_r^r (2m^*(x) + |x|^r \bar{n}(x) + 2\sqrt{m^{*2}(x) + m^*(x)|x|^r \bar{n}(x)}) = T_r^r Q(x) \end{aligned}$$

for large T . Since $m^*(x)$ and $|x|^r \bar{n}(x)$ are locally integrable, we can see that Q is also locally integrable from $\sqrt{m^{*2}(x) + m^*(x)|x|^r \bar{n}(x)} \leq m^*(x) + \sqrt{m^*(x)|x|^r \bar{n}(x)}$ and the Hölder inequality. Given $t \geq 0$, we can therefore deduce that

$$\frac{1}{T_r} \int_0^{\tau_t^T} R_T^2(W_s^{rT})ds \leq \int_0^{\tau_t^T} Q(W_s^{rT})ds < \infty$$

for large T with probability arbitrarily close to 1, where the second inequality follows from the same step as in (B.69). Since $R_T^2(x)/T_r^r \rightarrow 0$ for all $x \in \mathbb{R}$ as $T \rightarrow \infty$ from Definition 3.2, we can apply the dominated convergence theorem to show that the quadratic variation of the second term in (B.62) converges to zero in probability as $T \rightarrow \infty$. Therefore, the second term of (B.62) diminishes to zero as $T \rightarrow \infty$.

Now we are left with the first term of (B.62), and we may deduce from (B.60) that

$$\int_0^t m^{*1/2}(X_s^T)dX_s^T \rightarrow_d \int_0^t m^{*1/2}(X_s^\circ)dX_s^\circ \quad (\text{B.64})$$

jointly with (B.60) as $T \rightarrow \infty$ for any $t \geq 0$ given, from which the convergence of the finite dimensional distribution readily follows. Consequently, from (B.61), (B.62) and (B.64) with (B.63) diminishing to zero, we obtain

$$(W_t^T) \rightarrow_d \left(\int_0^t m^{*1/2}(X_s^\circ)dX_s^\circ \right)$$

for $0 \leq t \leq 1$ jointly with (B.60) as $T \rightarrow \infty$. This completes the proof.

Proof of Theorem 3.5

Part (a) follows from Corollary 3.3 of Höpfner and Löcherbach (2003), which is the multivariate version of their Corollary 3.2. The normalizing sequence v and the constant K can be obtained from their Example 3.10, which leads to their Equation (3.6').

For Part (b), we will sequentially establish the additive functional asymptotics and the martingale transform asymptotics below. For the proof of Part (b), we will use the notations we introduce in the proof of Proposition 3.4. Also, following the convention in the proof of Proposition 3.4, we will suppress the subscript and superscript “ s ” and denote $m = m_r$.

Additive Functional Asymptotics We write $f_T(x) = f(T_r x)$ conformably as $m_T(x) = m(T_r x)$ with $T_r = T^{1/(r+2)}$ introduced in the proof of Proposition 3.4. Then we can deduce

$$\begin{aligned}
\frac{\kappa(f, T_r)^{-1}}{T} \int_0^T f(X_t) dt &= \kappa(f, T_r)^{-1} \int_0^1 f_T(X_t^T) dt \\
&= \kappa(f, T_r)^{-1} \int_0^1 f_T((B^{rT} \circ \tau^T)_t) dt \\
&= \frac{T_r^2 \kappa(f, T_r)^{-1}}{T} \int_{\mathbb{R}} m_T(x) \int_0^{\tau_1^T} f_T(B_t^{rT}) l_r(dt, x) dx \\
&= \frac{\kappa(f, T_r)^{-1}}{T_r^r} \int_0^{\tau_1^T} (m_T f_T)(B_t^{rT}) dt,
\end{aligned} \tag{B.65}$$

where l_r is the local time of B^{rT} . In (B.65), the first equality follows from the change of variable in integral from t to t/T and the definition of X^T in (B.47), the second equality is from (B.48), the third equality is due to the change of variable formula in, e.g., Proposition 0.4.10 of Revuz and Yor (1999), and the fourth inequality uses

$$\int_{\mathbb{R}} m_T(x) l_r(dt, x) dx = m_T(B_t^{rT}) dt,$$

which is a generalized version of the so-called occupation times formula in, e.g., Exercise VI.1.15 of Revuz and Yor (1999).

We further deduce

$$\frac{\kappa(f, T_r)^{-1}}{T_r^r} \int_0^{\tau_1^T} (m_T f_T)(B_t^{rT}) dt = \int_0^{\tau_1^T} m^*(B_t^{rT}) h(f, B_t^{rT}) dt + \frac{\kappa(f, T_r)^{-1}}{T_r^r} \int_0^{\tau_1^T} \delta_f(T_r, B_t^{rT}) dt, \tag{B.66}$$

where $\delta_f(\lambda, x) = (mf)(\lambda x) - \lambda^r \kappa(f, \lambda) m^*(x) h(f, x)$. From Definitions 3.2 and 3.3, $\delta_f(\lambda, x)$ can be bounded by

$$|\delta_f(\lambda, x)| \leq \Lambda_1(\lambda) Q_1(x) + \sum_{j=1}^4 \Lambda_{2j}(\lambda) Q_{2j}(\lambda, x),$$

where

$$\begin{aligned}
\Lambda_1(\lambda) &= \lambda^r a(f, \lambda), & Q_1(x) &= m^*(x) p(f, x), \\
\Lambda_{21}(\lambda) &= \lambda^r b(f, \lambda), & Q_{21}(\lambda, x) &= m^*(x) q(f, \lambda x), \\
\Lambda_{22}(\lambda) &= \lambda^r \kappa(f, \lambda), & Q_{22}(\lambda, x) &= |x|^r n(\lambda x) h(f, x), \\
\Lambda_{23}(\lambda) &= \lambda^r a(f, \lambda), & Q_{23}(\lambda, x) &= |x|^r n(\lambda x) p(f, x), \\
\Lambda_{24}(\lambda) &= \lambda^r b(f, \lambda), & Q_{24}(\lambda, x) &= |x|^r n(\lambda x) q(f, \lambda x),
\end{aligned}$$

denoting $n(x) = |x|^{-r} \varepsilon_r(x)$. Note that $Q_1(\cdot)$ is locally integrable and

$$\limsup_{\lambda \rightarrow \infty} \|\lambda^{-r} \kappa(f, \lambda)^{-1} \Lambda_1(\lambda)\| = 0, \quad \limsup_{\lambda \rightarrow \infty} \|\lambda^{-r} \kappa(f, \lambda)^{-1} \Lambda_{2j}(\lambda)\| < \infty$$

for $j = 1, \dots, 4$, from the conditions on $h(f, \cdot)$, $p(f, \cdot)$, $a(f, \cdot)$ and $b(f, \cdot)$ in Definition 3.3.

In the next step, we will show that there exist locally integrable $\bar{Q}_{2j}(\cdot)$ for $j = 1, \dots, 4$ which can bound $Q_{2j}(\lambda, \cdot)$ for large λ , by showing the existence of \bar{n} such that $|n(\lambda x)| \leq \bar{n}(x)$ for all large λ , and that a function f is locally integrable w.r.t. $|x|^r \bar{n}(x)$ as long as it is locally integrable w.r.t. m and m^* . If this holds, $h(f, \cdot)$, $p(f, \cdot)$ and $q(f, \cdot)$ are locally integrable w.r.t. $|x|^r \bar{n}(x)$, thus we can find such $\bar{Q}_{2j}(\cdot)$ replacing $n(\lambda x)$ with $\bar{n}(x)$, since we can choose q satisfying $q(f, \lambda x) \leq q(f, x)$ for all large λ wlog.

To find such \bar{n} , let

$$\bar{n}(x) = (|n|(x) + \delta)1\{|x| \leq M\} + |n|(x)1\{|x| > M\} \quad (\text{B.67})$$

for some $\delta > 0$ and large enough M so that n is monotone on $\pm[M, \infty)$. When $-1 < r < 0$, we have $|n(x)| \leq |x|^{-r} m(x) + c$ for some c from Definition 3.2, thus n is locally bounded in \mathbb{R} from the continuity of m on \mathbb{R} . So $|n(\lambda x)| \leq \bar{n}(x)$ for all large λ since n is diminishing at infinities. Furthermore, when $a, b > 0$ defined in (31), f is locally integrable w.r.t. $|x|^r \bar{n}(x)$ if it is locally integrable w.r.t. m^* since $|x|^r \bar{n}(x) \leq c m^*(x)$ for some c . When either a or b is 0, the integrability on that half line can be dealt with separately. If $a = 0$, we have $|x|^r \bar{n}(x)1\{x \geq 0\} \leq c m(x)1\{x \geq 0\}$ for some c since $m(x) > 0$, thus f is locally integrable w.r.t. $|x|^r \bar{n}(x)$ on \mathbb{R}_+ if it is locally integrable w.r.t. m on \mathbb{R}_+ . The $b = 0$ case is also the same.

If $r > 0$, $|x|^r n(x)$ is locally bounded on \mathbb{R} with $|x|^r n(x) \sim c$ as $x \rightarrow 0$ for some $c > 0$ from $|x|^r n(x) = m(x) - m^*(x)$, $m(x) > 0$, and the continuity of m on \mathbb{R} . From this $n(x) \sim c|x|^{-r}$ as $x \rightarrow 0$ and locally bounded on $\mathbb{R} \setminus \{0\}$, thus $|n(\lambda x)| \leq \bar{n}(x)$ for all large λ since n is diminishing at infinities. Also, note that $|x|^r \bar{n}(x) \leq c m(x)$ for some c since $m(x) > 0$ on \mathbb{R} and \bar{n} is of smaller order than m at infinities. So f is locally integrable w.r.t. $|x|^r \bar{n}(x)$ if it is locally integrable w.r.t. m .

Finally for $r = 0$, $|x|^r n(x) = n(x)$ is locally bounded on \mathbb{R} since $n(x) = m(x) - m^*(x)$ and both m and m^* are locally bounded on \mathbb{R} . Thus $|n(\lambda x)| \leq \bar{n}(x)$ for all large λ since n is diminishing at infinities. It also follows that f is locally integrable w.r.t. $\bar{n}(x)$ if it is locally integrable w.r.t. m since $|\bar{n}(x)| \leq m(x) + c$ for some $c > 0$ and $m(x) > 0$ on \mathbb{R} .

Now we are ready to bound $\delta_f(\lambda, x)$ properly. The second term of (B.66) is bounded by

$$\begin{aligned} & \left| \frac{\kappa(f, T_r)^{-1}}{T_r^r} \int_0^{\tau_1^T} \delta_f(T_r, B_t^{rT}) dt \right| \\ & \leq \frac{\kappa(f, T_r)^{-1}}{T_r^r} \Lambda_1(T_r) \int_0^{\tau_1^T} Q_1(B_t^{rT}) dt + \frac{\kappa(f, T_r)^{-1}}{T_r^r} \sum_{j=1}^4 \Lambda_{2j}(T_r) \int_0^{\tau_1^T} Q_{2j}(T_r, B_t^{rT}) dt \\ & = P_{1T} + P_{2T}, \end{aligned} \quad (\text{B.68})$$

and for P_{1T} ,

$$\int_0^{\tau_1^T} Q_1(B_t^{rT}) dt = \int_{\mathbb{R}} Q_1(x) l_r(\tau_1^T, x) dx \leq \int_{\mathbb{R}} Q_1(x) l_r(\tau_2^r + 1, x) dx \quad (\text{B.69})$$

with probability arbitrarily close to 1 for large T . Thus $\int_0^{\tau_1^T} Q_1(B_t^{rT}) dt = O_p(1)$ as $T \rightarrow \infty$ and it follows that $P_{1T} = o_p(1)$ since $\|T_r^{-r} \kappa(f, T_r)^{-1} \Lambda_1(T_r)\| = o(1)$ as $T \rightarrow \infty$. For P_{2T} also,

$$\int_0^{\tau_1^T} Q_{2j}(T_r, B_t^{rT}) dt = \int_{\mathbb{R}} Q_{2j}(T_r, x) l_r(\tau_1^T, x) dx \leq \int_{\mathbb{R}} Q_{2j}(T_r, x) l_r(\tau_2^r + 1, x) dx$$

for each $j = 1, \dots, 4$ with probability arbitrarily close to 1 for large T . Since $Q_{2j}(T_r, x) \leq \bar{Q}_{2j}(x)$ for all large T and $\int_{\mathbb{R}} \bar{Q}_{2j}(x) l_r(\tau_2^r + 1, x) dx < \infty$ a.s., we can apply the dominated convergence theorem to obtain $\int_0^{\tau_1^T} Q_{2j}(T_r, B_t^{rT}) dt = o_p(1)$ as $T \rightarrow \infty$. Thus we have $P_{2T} = o_p(1)$ since $\|T_r^{-r} \kappa(g, T_r)^{-1} \Lambda_{2j}(T_r)\| = O(1)$ as $T \rightarrow \infty$, and therefore the second term in (B.66) is asymptotically negligible.

For the first term of (B.66),

$$\int_0^{\tau_1^T} m^*(B_t^{rT}) h(f, B_t^{rT}) dt = \int_0^{\tau_1^r} m^*(B_t^{rT}) h(f, B_t^{rT}) dt + \int_{\tau_1^r}^{\tau_1^T} m^*(B_t^{rT}) h(f, B_t^{rT}) dt, \quad (\text{B.70})$$

and the second term is bounded by

$$\begin{aligned} \int_{\tau_1^r}^{\tau_1^T} m^*(B_t^{rT}) h(f, B_t^{rT}) dt &= \int_{\mathbb{R}} m^*(x) h(f, x) [l_r(\tau_1^T, x) - l_r(\tau_1^r, x)] dx \\ &\leq \int_{\mathbb{R}} m^*(x) h(f, x) l_r(\tau_2^r + 1, x) dx \\ &= \int_0^{\tau_2^r + 1} m^*(B_t^{rT}) h(f, B_t^{rT}) dt < \infty \end{aligned}$$

a.s. for large T . Thus we can apply the dominated convergence theorem to show the second term in (B.70) is asymptotically negligible, and the first term converges to

$$\begin{aligned} \int_0^{\tau_1^T} m^*(B_t^{rT}) h(f, B_t^{rT}) dt &= \int_0^{\tau_1^r} m^*(B_t^{rT}) h(f, B_t^{rT}) dt + o_p(1) \\ &\rightarrow_d \int_0^{\tau_1^r} m^*(B_t) h(f, B_t) dt \end{aligned} \quad (\text{B.71})$$

since B^{rT} has the same distribution for all T . Taking the reverse steps of (B.65), we get

$$\int_0^{\tau_1^r} m^*(B_t) h(f, B_t) dt = T_r^r \int_0^1 h(f, X_t^\circ) dt, \quad (\text{B.72})$$

and consequently it follows from (B.66), (B.70), (B.71) and (B.72) that

$$\frac{1}{T} \kappa(f, T_r)^{-1} \int_0^T f(X_t) dt \rightarrow_d \int_0^1 h(f, X_t^\circ) dt \quad (\text{B.73})$$

as $T \rightarrow \infty$.

Martingale Transform Asymptotics To derive the martingale transform asymptotics, we deduce

$$\begin{aligned} \frac{1}{\sqrt{T}} \kappa(g, T_r)^{-1} \int_0^T g(X_t) dW_t &= \frac{1}{\sqrt{T}} \kappa(g, T_r)^{-1} \int_0^1 g_T((B^{rT} \circ \tau^T)_t) dW_{Tt} \\ &= \kappa(g, T_r)^{-1} \int_0^{\tau_1^T} g_T(B_t^{rT}) d(W^T \circ \varsigma^T)_t, \end{aligned} \quad (\text{B.74})$$

where $W_t^T = T^{-1/2} W_{Tt}$ and $\varsigma_t^T = T^{-1} T_r^2 \int_0^t m_T(B_s^{rT}) ds$ from the change of variables. Note that τ_t^T is the right continuous inverse of ς_t^T . Rewriting (B.74), we have

$$\kappa(g, T_r)^{-1} \int_0^{\tau_1^T} g_T(B_t^{rT}) d(W^T \circ \varsigma^T)_t = \int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^T)_t + R_T, \quad (\text{B.75})$$

where the remainder term is given by

$$R_T = \int_0^{\tau_1^T} \left(\kappa(g, T_r)^{-1} g_T(B_t^{rT}) - h(g, B_t^{rT}) \right) d(W^T \circ \varsigma^T)_t.$$

To show $R_T = o_p(1)$, note that R_T is a martingale whose quadratic variation is

$$\begin{aligned} & \int_0^{\tau_1^T} \frac{m_T(B_t^{rT})}{T_r^r} \left(\kappa(g, T_r)^{-1} g_T(B_t^{rT}) - h(g, B_t^{rT}) \right) \left(\kappa(g, T_r)^{-1} g_T(B_t^{rT}) - h(g, B_t^{rT}) \right)' dt \\ &= \kappa(g, T_r)^{-1} \int_0^{\tau_1^T} m^*(B_t^{rT}) \delta \delta'(g, T_r, B_t^{rT}) dt \kappa'(g, T_r)^{-1} \\ &+ \frac{1}{T_r^r} \kappa(g, T_r)^{-1} \int_0^{\tau_1^T} \varepsilon_r(T_r B_t^{rT}) \delta \delta'(g, T_r, B_t^{rT}) dt \kappa'(g, T_r)^{-1} \\ &= R_{1T} + R_{2T}. \end{aligned}$$

Due to Definition 3.3, R_{1T} is bounded by

$$\begin{aligned} & \kappa(g, T_r)^{-1} a(g, T_r) \int_0^{\tau_1^T} m^*(B_t^{rT}) p p'(g, B_t^{rT}) dt a'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \kappa(g, T_r)^{-1} a(g, T_r) \int_0^{\tau_1^T} m^*(B_t^{rT}) p(g, B_t^{rT}) q'(g, T_r B_t^{rT}) dt b'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \kappa(g, T_r)^{-1} b(g, T_r) \int_0^{\tau_1^T} m^*(B_t^{rT}) q(g, T_r B_t^{rT}) p'(g, B_t^{rT}) dt a'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \kappa(g, T_r)^{-1} b(g, T_r) \int_0^{\tau_1^T} m^*(B_t^{rT}) q q'(g, T_r B_t^{rT}) dt b'(g, T_r) \kappa'(g, T_r)^{-1}, \end{aligned}$$

each term of which can be shown asymptotically negligible in the same way as in (B.66)–(B.70). The first term can be dealt with in a similar way as P_{1T} in (B.68), since by denoting $\Lambda_1(\lambda) = \lambda^r(a \otimes a)(g, \lambda)$ and $Q_1(x) = m^*(x)(p \otimes p)(g, x)$, we have

$$\begin{aligned} & (\kappa \otimes \kappa)(g, T_r)^{-1} (a \otimes a)(g, T_r) \int_0^{\tau_1^T} m^*(B_t^{rT}) (p \otimes p)(g, B_t^{rT}) dt \\ &= \frac{1}{T_r^r} (\kappa \otimes \kappa)(g, T_r)^{-1} \Lambda_1(T_r) \int_0^{\tau_1^T} Q_1(B_t^{rT}) dt, \end{aligned}$$

which is of the same form as P_{1T} . Other terms can be shown to be negligible in similar ways as for P_{2T} in (B.68), and omitted here. For R_{2T} , due to Definitions 3.2 and 3.3, R_{2T} is bounded by

$$\begin{aligned} & \frac{1}{T_r^r} \kappa(g, T_r)^{-1} a(g, T_r) \int_0^{\tau_1^T} \varepsilon_r(T_r B_t^{rT}) p p'(g, B_t^{rT}) dt a'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \frac{1}{T_r^r} \kappa(g, T_r)^{-1} a(g, T_r) \int_0^{\tau_1^T} \varepsilon_r(T_r B_t^{rT}) p(g, B_t^{rT}) q'(g, T_r B_t^{rT}) dt b'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \frac{1}{T_r^r} \kappa(g, T_r)^{-1} b(g, T_r) \int_0^{\tau_1^T} \varepsilon_r(T_r B_t^{rT}) q(g, T_r B_t^{rT}) p'(g, B_t^{rT}) dt a'(g, T_r) \kappa'(g, T_r)^{-1} \\ &+ \frac{1}{T_r^r} \kappa(g, T_r)^{-1} b(g, T_r) \int_0^{\tau_1^T} \varepsilon_r(T_r B_t^{rT}) q q'(g, T_r B_t^{rT}) dt b'(g, T_r) \kappa'(g, T_r)^{-1}. \end{aligned} \tag{B.76}$$

We can also apply the same steps for P_{2T} in (B.68) by choosing \bar{n} defined in (B.67) such that $T_r^{-r} \varepsilon_r(T_r x) \leq |x|^r \bar{n}(x)$ for large T , to show that each term of (B.76) become asymptotically negligible. We therefore have $R_T = o_p(1)$.

Now going back to (B.75), we can rewrite the leading term as

$$\begin{aligned} \int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^T)_t &= \int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^r)_t \\ &\quad + \int_0^{\tau_1^T} h(g, B_t^{rT}) d[(W^T \circ \varsigma^T) - (W^T \circ \varsigma^r)]_t, \end{aligned} \quad (\text{B.77})$$

where $\varsigma_t^r = \int_0^t m^*(B_s^{rT}) ds$. To show $\int_0^{\tau_1^T} h(g, B_t^{rT}) d[(W^T \circ \varsigma^T) - (W^T \circ \varsigma^r)]_t = o_p(1)$, we first deduce

$$\begin{aligned} \int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^T)_t &- \int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^r)_t \\ &= \int_0^1 h(g, (B^{rT} \circ \tau^T)_t) dW_t^T - \int_0^{\varsigma^r \circ \tau_1^T} h(g, (B^{rT} \circ \tau^r)_t) dW_t^T \\ &= \int_0^1 [h(g, (B^{rT} \circ \tau^T)_t) - h(g, (B^{rT} \circ \tau^r)_t)] dW_t^T + \int_1^{\varsigma^r \circ \tau_1^T} h(g, (B^{rT} \circ \tau^r)_t) dW_t^T \end{aligned}$$

from the change of variables. The second term is a martingale with quadratic variation

$$\int_1^{\varsigma^r \circ \tau_1^T} hh'(g, (B^{rT} \circ \tau^r)_t) dt = \int_{\tau_1^r}^{\tau_1^T} m^*(B_t^{rT}) hh'(g, B_t^{rT}) dt \rightarrow_p 0$$

from the same step that we used to deal with the second term of (B.70). The first term is also a martingale whose quadratic variation is

$$\begin{aligned} Q_T &= \int_0^1 hh'(g, (B^{rT} \circ \tau^T)_t) dt + \int_0^1 hh'(g, (B^{rT} \circ \tau^r)_t) dt \\ &\quad - \int_0^1 h(g, (B^{rT} \circ \tau^T)_t) h'(g, (B^{rT} \circ \tau^r)_t) dt - \int_0^1 h(g, (B^{rT} \circ \tau^r)_t) h'(g, (B^{rT} \circ \tau^T)_t) dt \\ &= Q_{1T} + Q_{2T} - Q_{3T} - Q_{4T}. \end{aligned} \quad (\text{B.78})$$

To show $Q_T \rightarrow_p 0$, we will show that Q_{1T} , Q_{2T} , Q_{3T} and Q_{4T} all converge to the same limit as $T \rightarrow \infty$, so that they cancel out in the limit. Before we start, we notice that there exists a sequence B^{rT*} such that $B^{rT*} \rightarrow_{a.s.} B^*$ and $\tau^{T*} \rightarrow_{a.s.} \tau^{r*}$ from the Skorohod representation theorem. We will consider this sequence hereafter to show the almost sure convergence of each term in (B.78). We suppress the superscript $*$ hereafter without confusion for notational simplicity.

Firstly we can show

$$Q_{1T} \rightarrow_{a.s.} \int_0^1 hh'(g, X_t^\circ) dt, \quad Q_{2T} \rightarrow_{a.s.} \int_0^1 hh'(g, X_t^\circ) dt \quad (\text{B.79})$$

with the same step as in (B.65) and (B.71). Nextly for the case of Q_{3T} and Q_{4T} , we will utilize the Vitali convergence theorem to show that

$$Q_{3T} \rightarrow_{a.s.} \int_0^1 hh'(g, X_t^\circ) dt, \quad Q_{4T} \rightarrow_{a.s.} \int_0^1 hh'(g, X_t^\circ) dt. \quad (\text{B.80})$$

(See, e.g., Theorem 11.13 of Bartle (2001) for the Vitali convergence theorem.) To apply this theorem, pointwise convergence and uniform integrability are required. Pointwise convergence is trivial since $B^{rT} \rightarrow_{a.s.} B$ and $\tau^T \rightarrow_{a.s.} \tau^r$. For uniform integrability, it is known that a sufficient condition is that there exists $\delta > 0$ such that

$$\int_0^1 h_i^{1+\delta}(g, (B^{rT} \circ \tau^T)_t) h_j^{1+\delta}(g, (B^{rT} \circ \tau^r)_t) dt < \infty \quad (\text{B.81})$$

a.s. uniformly in large T for all i, j , where h_i is the i 'th element of h . (See, e.g., Exercise 11.V of Bartle (2001).) Since there exists $\delta > 0$ that makes $m^*(\cdot)h_i^{2+2\delta}(g, \cdot)$ locally integrable for all i from the local integrability condition on $m^*(\cdot)hh'(g, \cdot)$, we have

$$\begin{aligned} & \int_0^1 h_i^{1+\delta}(g, (B^{rT} \circ \tau^T)_t) h_j^{1+\delta}(g, (B^{rT} \circ \tau^r)_t) dt \\ & \leq \sqrt{\int_0^1 h_i^{2+2\delta}(g, (B^{rT} \circ \tau^T)_t) dt \int_0^1 h_j^{2+2\delta}(g, (B^{rT} \circ \tau^r)_t) dt} \\ & \rightarrow_{a.s.} \sqrt{\int_0^1 h_i^{2+2\delta}(g, X_t^\circ) dt \int_0^1 h_j^{2+2\delta}(g, X_t^\circ) dt} < \infty \end{aligned}$$

a.s. from the same steps as in (B.65) and (B.71). Therefore the uniform integrability condition is satisfied and we can apply the Vitali convergence theorem to obtain (B.80). Thus from (B.79) and (B.80), we have $Q_T \rightarrow_p 0$.

Now we are only left with the leading term of (B.77). Denoting $W^T \rightarrow_d W^\circ$ as $T \rightarrow \infty$, we have

$$\int_0^{\tau_1^T} h(g, B_t^{rT}) d(W^T \circ \varsigma^r)_t \rightarrow_d \int_0^{\tau_1^r} h(g, B_t) d(W^\circ \circ \varsigma^r)_t$$

jointly with (B.71) as $T \rightarrow \infty$, and by the change of variables we have

$$\int_0^{\tau_1^r} h(g, B_t) d(W^\circ \circ \varsigma^r)_t = \int_0^1 h(g, B \circ \tau_t^r) dW_t^\circ = \int_0^1 h(g, X_t^\circ) dW_t^\circ. \quad (\text{B.82})$$

Thus consequently it follows from (B.74), (B.75), (B.77) and (B.82) that

$$\frac{1}{\sqrt{T}} \kappa(g, T_r)^{-1} \int_0^T g(X_t) dW_t \rightarrow_d \int_0^1 h(g, X_t^\circ) dW_t^\circ$$

jointly with (B.73) as $T \rightarrow \infty$.

Proof of Lemma 3.6

It follows from Lemma 3.2, together with the definition of $w = \text{diag}(w_\alpha(T), \Delta^{-1/2} w_\beta(T))$.

Proof of Lemma 3.7

We will prove the statements by showing that there exist positive nondecreasing sequences $\nu_1(T)$ and $\nu_2(T)$ such that

$$\|T^\varepsilon(w_\alpha \otimes w_\alpha \otimes w_\alpha)^{-1}(T) \nu_1(T)\| \rightarrow 0, \quad \frac{1}{\nu_1(T)} \sup_{\theta \in \mathcal{N}} \left| \int_0^T f(X_t, \theta) dt \right| = O_p(1) \quad (\text{B.83})$$

and

$$\|T^\varepsilon(w_\alpha \otimes w_\alpha \otimes w_\alpha)^{-1}(T)\nu_2(T)\| \rightarrow 0, \quad \frac{1}{\nu_2(T)} \sup_{\theta \in \mathcal{N}} \left| \int_0^T g(X_t, \theta) dW_t \right| = O_p(1) \quad (\text{B.84})$$

as $T \rightarrow \infty$. For the second conditions of (B.83) and (B.84) respectively, note that it is enough to show the stochastic boundedness for each element of f and g . So without loss of generality, we will only consider the case when f and g are scalar valued functions hereafter.

Firstly, (B.83) can be shown with the following for each case. For (a), letting $\nu_1(T) = T$,

$$\frac{1}{T} \sup_{\theta \in \mathcal{N}} \left| \int_0^T f(X_t, \theta) dt \right| \leq \frac{1}{T} \int_0^T p(X_t) dt \rightarrow_{a.s.} \pi(p)$$

as $T \rightarrow \infty$ from Proposition 3.3. Thus (B.83) holds since $w_\alpha(T) = \sqrt{T}$. For (b), letting $\nu_1(T) = T^{1/(r+2)}$,

$$\frac{1}{T^{1/(r+2)}} \sup_{\theta \in \mathcal{N}} \left| \int_0^T f(X_t, \theta) dt \right| \leq \frac{1}{T^{1/(r+2)}} \int_0^T p(X_t) dt \rightarrow_d Km(p)A^{1/(r+2)}$$

as $T \rightarrow \infty$ from Theorem 3.5(a). Thus (B.83) holds since $w_\alpha(T) = \sqrt{T^{1/(r+2)}}$. For (c), letting $\nu_1(T) = T\kappa(p, T^{1/(r+2)})$,

$$\frac{1}{T\kappa(p, T^{1/(r+2)})} \sup_{\theta \in \mathcal{N}} \left| \int_0^T f(X_t, \theta) dt \right| \leq \frac{1}{T\kappa(p, T^{1/(r+2)})} \int_0^T p(X_t) dt \rightarrow_d \int_0^1 h(p, X_t^r) dt$$

as $T \rightarrow \infty$ from Theorem 3.5(b). Thus (B.83) holds since $w_\alpha(T) = \sqrt{T}\kappa(p, T^{1/(r+2)})$.

To prove (B.84), we will show that

$$Q_T(\theta) = \frac{1}{\nu_2(T)} \int_0^T g(X_t, \theta) dW_t$$

satisfies the multivariate extension of Kolmogorov's criterion for the weak compactness w.r.t. θ , which is

$$\mathbb{E} \|Q_T(\theta_1) - Q_T(\theta_2)\|^\gamma \leq C \|\theta_1 - \theta_2\|^{d+\epsilon}$$

for some $\gamma, C, \epsilon > 0$ and $\theta_1, \theta_2 \in \mathcal{N}_\varepsilon$ for all large T , where d is the dimension of θ . If $Q_T(\theta)$ satisfies this, it converges to a random variable uniformly in $\theta \in \mathcal{N}_\varepsilon$, thus the second condition of (B.84) is satisfied. (See Theorem XIII.1.8 of Revus and Yor (1999) for Kolmogorov's criterion, and Theorem I.2.1 and Exercise I.2.10 of the same article for its multivariate extension.)

For (a), letting $\nu_2(T) = \sqrt{T}$,

$$\begin{aligned} \mathbb{E} \left(\frac{1}{\sqrt{T}} \int_0^T (g(X_t, \theta_1) - g(X_t, \theta_2)) dW_t \right)^{d+\epsilon} &\leq C_{d+\epsilon} \mathbb{E} \left(\frac{1}{T} \int_0^T (g(X_t, \theta_1) - g(X_t, \theta_2))^2 dt \right)^{\frac{d+\epsilon}{2}} \\ &\leq C_{d+\epsilon} \|\theta_1 - \theta_2\|^{d+\epsilon} \mathbb{E} \left(\frac{1}{T} \int_0^T q^2(X_t) dt \right)^{\frac{d+\epsilon}{2}} \\ &\leq C_{d+\epsilon} \|\theta_1 - \theta_2\|^{d+\epsilon} \frac{1}{T} \int_0^T \mathbb{E} q^{d+\epsilon}(X_t) dt \end{aligned}$$

for some constant $C_{d+\epsilon}$ and all large T , where the first inequality is due to the Burkholder-Davis-Gundy inequality and the last inequality is due to the Hölder inequality. Thus Kolmogorov's criterion is satisfied with

$$\mathbb{E}\|Q_T(\theta_1) - Q_T(\theta_2)\|^{d+\epsilon} \leq C_{d+\epsilon}\pi(q^{d+\epsilon})\|\theta_1 - \theta_2\|^{d+\epsilon}$$

for all large T , which is to be shown. Note that the first condition of (B.84) also holds since $w_\alpha(T) = \sqrt{T}$.

For (b), let $\nu_2(T) = \sqrt{T^{1/(r+2)}}$ and denote $\theta^* = T^{1/[2(r+2)]-\epsilon}(\theta - \theta_0)$. Then

$$Q_T(\theta) = Q_T^*(\theta^*) = \frac{1}{\sqrt{T^{1/(r+2)}}} \int_0^T g_T^*(X_t, \theta^*) dW_t,$$

where $g_T^*(x, \theta^*) = g(x, T^{-1/[2(r+2)]+\epsilon}\theta^* + \theta_0)$, so we can show Kolmogorov's criterion for $Q_T^*(\theta^*)$ instead. We have

$$\begin{aligned} & \mathbb{E}\left(\frac{1}{\sqrt{T^{1/(r+2)}}} \int_0^T (g_T^*(X_t, \theta_1^*) - g_T^*(X_t, \theta_2^*)) dW_t\right)^{d+\epsilon} \\ & \leq C_{d+\epsilon} \mathbb{E}\left(\frac{1}{T^{1/(r+2)}} \int_0^T (g_T^*(X_t, \theta_1^*) - g_T^*(X_t, \theta_2^*))^2 dt\right)^{\frac{d+\epsilon}{2}} \\ & \leq C_{d+\epsilon} T^{(d+\epsilon)(\epsilon-1/[2(r+2)])} \|\theta_1^* - \theta_2^*\|^{d+\epsilon} \mathbb{E}\left(\frac{1}{T^{1/(r+2)}} \int_0^T q^2(X_t) dt\right)^{\frac{d+\epsilon}{2}} \end{aligned}$$

for some constant $C_{d+\epsilon}$ and all large T , from the Burkholder-Davis-Gundy inequality and the condition on g . We also have

$$T^{-\epsilon} \mathbb{E}\left(\frac{1}{T^{1/(r+2)}} \int_0^T q^2(X_t) dt\right)^{\frac{d+\epsilon}{2}} < \infty$$

uniformly for all large T from (B.85), thus Kolmogorov's criterion is satisfied for all large T , which is to be shown. Note that the first condition of (B.84) also holds since $w_\alpha(T) = \sqrt{T^{1/(r+2)}}$.

For (c), let $\nu_2(T) = \sqrt{T}\kappa(q, T^{1/(r+2)})$ and $\theta^* = T^{1/2-\epsilon} \text{diag}[\kappa'(\nu_\alpha, T_r), \kappa'(\tau_\beta, T_r)](\theta - \theta_0)$. We will also show Kolmogorov's criterion for $Q_T^*(\theta^*)$ defined as

$$Q_T^*(\theta^*) = \frac{1}{\sqrt{T}\kappa(q, T^{1/(r+2)})} \int_0^T g_T^*(X_t, \theta^*) dW_t,$$

where $g_T^*(x, \theta^*) = g(x, T^{-1/2+\epsilon} \text{diag}[\kappa'^{-1}(\nu_\alpha, T_r), \kappa'^{-1}(\tau_\beta, T_r)]\theta^* + \theta_0)$. We have

$$\begin{aligned} & \mathbb{E}\left(\frac{1}{\sqrt{T}\kappa(q, T^{1/(r+2)})} \int_0^T (g_T^*(X_t, \theta_1^*) - g_T^*(X_t, \theta_2^*)) dW_t\right)^{d+\epsilon} \\ & \leq C_{d+\epsilon} \mathbb{E}\left(\frac{1}{T\kappa^2(q, T^{1/(r+2)})} \int_0^T (g_T^*(X_t, \theta_1^*) - g_T^*(X_t, \theta_2^*))^2 dt\right)^{\frac{d+\epsilon}{2}} \\ & \leq C_{d+\epsilon} T^{(d+\epsilon)(\epsilon-1/2)} \|\text{diag}[\kappa'^{-1}(\nu_\alpha, T_r), \kappa'^{-1}(\tau_\beta, T_r)](\theta_1^* - \theta_2^*)\|^{d+\epsilon} \times \\ & \quad \mathbb{E}\left(\frac{1}{T\kappa^2(q, T^{1/(r+2)})} \int_0^T q^2(X_t) dt\right)^{\frac{d+\epsilon}{2}} \end{aligned}$$

for some constant $C_{d+\epsilon}$ and all large T , from the Burkholder-Davis-Gundy inequality and the condition on g . We also have

$$T^{-\epsilon} \mathbb{E} \left(\frac{1}{T \kappa^2(q, T^{1/(r+2)})} \int_0^T q^2(X_t) dt \right)^{\frac{d+\epsilon}{2}} < \infty$$

uniformly for all large T from (B.86), thus Kolmogorov's criterion is satisfied for all large T , which is to be shown. Note that the first condition of (B.84) also holds since $w_\alpha(T) = \sqrt{T} \kappa(\nu_\alpha, T^{1/(r+2)})$.

Existence of Moments As before, we assume that the required scale transform has already been done and X is in natural scale. For any $k \geq 1$, we show that

$$T^{-\epsilon} \mathbb{E} \left(\frac{1}{T_r} \int_0^T f(X_t) dt \right)^k < \infty \quad (\text{B.85})$$

uniformly for all large T , if f is integrable in m , and that

$$T^{-\epsilon} \mathbb{E} \left(\frac{1}{T} \kappa(g, T_r)^{-1} \int_0^T g(X_t) dt \right)^k < \infty \quad (\text{B.86})$$

uniformly for all large T , if g is a homogeneous function such that $g(\lambda x) = \kappa(g, \lambda)g(x)$ and g is locally integrable in m .

To show (B.85), we assume without loss of generality that f is nonnegative and has support on a subset of \mathbb{R}_+ . Note that

$$\begin{aligned} \frac{1}{T_r} \int_0^T f(X_t) dt &= T_r \int_0^{\tau_1^T} (m_T f_T)(B_t^{rT}) dt \\ &= T_r \int_{\mathbb{R}} (m_T f_T)(x) l_r(\tau_1^T, x) dx \\ &= \int_{\mathbb{R}} (mf)(x) l_r\left(\tau_1^T, \frac{x}{T_r}\right) dx \\ &\leq \int_{\mathbb{R}} (mf)(x) l_r\left(\tau_2^r + 1, \frac{x}{T_r}\right) dx. \end{aligned} \quad (\text{B.87})$$

The first equality in (B.87) is due to (B.65), the second equality follows directly from the occupation times formula, the third equality is obtained from a simple change of variable in integration, and the last inequality is immediate from $\tau_1^T \leq \tau_2^r + 1$ a.s. for all large T and the nondecreasing property of the additive functional $l_r(\cdot, x)$.

We write

$$l_r(\tau_2^r + 1, \cdot) = l_r(\tau_2^r, \cdot) + l(1, \cdot - X_2^\circ), \quad (\text{B.88})$$

where l is the local time of Brownian motion

$$B_{\tau_2^r + \cdot} - B_{\tau_2^r}$$

which is independent of $X_2^\circ = (B \circ \tau^r)_2$, due to the strong markov property of Brownian motion B . It follows from (B.88) that

$$\begin{aligned} \int_{\mathbb{R}} (mf)(x) l_r\left(\tau_2^r + 1, \frac{x}{T_r}\right) dx &= \int_{\mathbb{R}} (mf)(x) l_r\left(\tau_2^r, \frac{x}{T_r}\right) dx + \int_{\mathbb{R}} (mf)(x) l\left(1, \frac{x}{T_r} - X_2^\circ\right) dx \\ &\leq \int_{\mathbb{R}} (mf)(x) l_r\left(\tau_2^r, \frac{x}{T_r}\right) dx + m(f) \left(\sup_{x \in \mathbb{R}} l(1, x) \right). \end{aligned} \quad (\text{B.89})$$

Moreover, we may readily show that $\sup_{x \in \mathbb{R}} l(1, x)$ has finite moment of any order, using its distribution obtained in, e.g., Borodin (1989, Theorem 4.2, p.13). Consequently, it suffices to show that

$$\mathbb{E} \left(\int_{\mathbb{R}} (mf)(x) l_r \left(\tau_2^r, \frac{x}{T_r} \right) dx \right)^k < \infty \quad (\text{B.90})$$

for all large T , due to (B.87) and (B.89). To show (B.90), we let $M > 0$ be such that $M \rightarrow \infty$ and $M/T_r \rightarrow 0$ and write

$$\int_{\mathbb{R}} (mf)(x) l_r \left(s, \frac{x}{T_r} \right) dx = \int_{|x| \leq M} (mf)(x) l_r \left(s, \frac{x}{T_r} \right) dx + \int_{|x| > M} (mf)(x) l_r \left(s, \frac{x}{T_r} \right) dx \quad (\text{B.91})$$

in what follows.

For the first term in (B.91), we note that

$$\begin{aligned} & \left| \int_{|x| \leq M} (mf)(x) l_r \left(s, \frac{x}{T_r} \right) dx - l_r(s, 0) \int_{|x| \leq M} (mf)(x) dx \right| \\ & \leq \int_{|x| \leq M} (mf)(x) \left| l_r \left(s, \frac{x}{T_r} \right) - l_r(s, 0) \right| dx \\ & \leq \lambda \left(\frac{M}{T_r} \right) \sqrt{l_r(s, 0)} \int_{\mathbb{R}} (mf)(x) dx \\ & \leq \lambda \left(\frac{M}{T_r} \right) (1 + l_r(s, 0)) \int_{\mathbb{R}} (mf)(x) dx \\ & = o(1) (1 + l_r(s, 0)) \int_{\mathbb{R}} (mf)(x) dx \end{aligned}$$

as $T \rightarrow \infty$, where $\lambda(z) = 2\sqrt{2}\sqrt{z \log \log 1/z}$, from which it follows that

$$\int_{|x| \leq M} (mf)(x) l_r \left(s, \frac{x}{T_r} \right) dx \leq a + b l_r(s, 0)$$

for all large T , where $a, b > 0$ are some nonrandom constants. Therefore, we have

$$\int_{|x| \leq M} (mf)(x) l_r \left(\tau_2^r, \frac{x}{T_r} \right) dx \leq a + b l_r(\tau_2^r, 0), \quad (\text{B.92})$$

where $l_r(\tau_2^r, 0)$ is a constant multiple of Mittag-Leffler process whose nonnegative moments exist up to an arbitrary order.

For the second term in (B.91), we write

$$f(x) = \frac{n(x)}{xm(x)},$$

where n is monotonically decreasing and vanishing at infinity. Also, we let $m = m^*$ to simplify the subsequent proof. It is rather clear that the existence of the additional term ε_r in m does not affect our proof. Under the convention, we have

$$m(T_r x) = T_r m(x), \quad f(T_r x) = T_r^{-(r+1)} \frac{n(T_r x)}{xm(x)},$$

and it follows that

$$\begin{aligned}
\int_{|x|>M} (mf)(x) l_r\left(\tau_2^r, \frac{x}{T_r}\right) dx &= T_r \int_{|x|>M/T_r} (mf)(T_r x) l_r(\tau_2^r, x) dx. \\
&\leq n(M) \left(\frac{T_r}{M}\right)^{r+1} \int_{|x|>M/T_r} m(x) l_r(\tau_2^r, x) dx \\
&\leq n(M) \left(\frac{T_r}{M}\right)^{r+1} \int_{\mathbb{R}} m(x) l_r(\tau_2^r, x) dx \\
&= 2n(M) \left(\frac{T_r}{M}\right)^{r+1} \leq 2T^\epsilon
\end{aligned} \tag{B.93}$$

for any $\epsilon > 0$ and for all large T , if we take $M > 0$ appropriately. Note that $r > -1$. Now (B.90) follows directly from (B.91), (B.92) and (B.93), as was to be shown to establish (B.85).

To simplify the proof of (B.86), we assume that $m = m^*$ as before. It is easy to accommodate the existence of the additional term ε_r . We note that

$$\begin{aligned}
\int_0^1 g(X_t^T) dt &= \int_{\mathbb{R}} (mg)(x) l_r(\tau_1^T, x) dx \\
&\leq \int_{\mathbb{R}} (mg)(x) l_r(\tau_2^T + 1, x) dx \\
&= \int_{\mathbb{R}} (mg)(x) l_r(\tau_2^T, x) dx + \int_{\mathbb{R}} (mg)(x) l(1, x - X_2^\circ) dx \\
&= \int_0^2 g(X_t^\circ) dt + \int_{\mathbb{R}} (mg)(x + X_2^\circ) l(1, x) dx.
\end{aligned} \tag{B.94}$$

In what follows, we assume without loss of generality that g is bounded by $p + q$, where p is a power function in modulus with nonnegative power and q is symmetric, locally integrable and monotonically decreasing such that mq is locally integrable.

For the first term, we have

$$\mathbb{E} \int_0^2 p^k(X_t^\circ) dt < \infty \tag{B.95}$$

for any $k \geq 1$, since X° has finite moments up to any order. Moreover, we may readily deduce that

$$T^{-\epsilon} \mathbb{E} \int_0^2 q^k(X_t^\circ) dt < \infty \tag{B.96}$$

with any $\epsilon > 0$, for any $k \geq 1$. To see this, we let $\delta \rightarrow 0$ as $T \rightarrow \infty$, and write

$$\int_{\mathbb{R}} (mq)(x) l_r(s, x) dx = \int_{|x| \leq \delta} (mq)(x) l_r(s, x) dx + \int_{|x| > \delta} (mq)(x) l_r(s, x) dx. \tag{B.97}$$

First, note that

$$\begin{aligned}
\int_{|x| \leq \delta} (mq)(x) l_r(s, x) dx &= l_r(s, 0) \int_{|x| \leq \delta} (mq)(x) dx + \int_{|x| \leq \delta} (mq)(x) (l_r(s, x) - l_r(s, 0)) dx \\
&\leq (l_r(s, 0) + \lambda(\delta) (1 + l_r(s, 0))) \int_{|x| \leq \delta} (mq)(x) dx,
\end{aligned}$$

from which we have

$$\int_{|x| \leq \delta} (mq)(x) l_r(\tau_2^r, x) dx \leq o(\lambda(\delta)) + (1 + o(\lambda(\delta))) l_r(\tau_2^r, 0), \quad (\text{B.98})$$

for all large T . Recall that $l_r(\tau_2^r, 0)$ is a constant multiple of Mittag-Leffler process, which has finite moments up to infinite order. Second, we write

$$\int_{|x| > \delta} (mq)(x) l_r(s, x) dx \leq \delta^{-(r+1)} \int_{|x| > \delta} m(x) l_r(s, x) dx,$$

and therefore,

$$\int_{|x| > \delta} (mq)(x) l_r(\tau_2^r, x) dx \leq \delta^{-(r+1)} \int_{|x| > \delta} m(x) l_r(\tau_2^r, x) dx = 2\delta^{-(r+1)}. \quad (\text{B.99})$$

Now (B.96) follow immediately from (B.97), (B.98) and (B.99), which implies together with (B.95) that the first term in (B.94) has finite moments up to any order that are bounded by $O(T^\epsilon)$ for any $\epsilon > 0$ uniformly for all large T .

For the second term of (B.94), we first note that

$$\int_{\mathbb{R}} p(x + X_2^\circ) l(1, x) dx \leq \int_{\mathbb{R}} (p(x) + p(X_2^\circ)) l(1, x) dx = \int_0^1 p(B_{\tau_2^r+t} - B_{\tau_2^r}) dt + p(X_2^\circ), \quad (\text{B.100})$$

whose expectation is finite. Moreover, we may easily deduce that

$$\mathbb{E} \int_{\mathbb{R}} q(x + y) l(1, x) dx \leq \mathbb{E} \int_{\mathbb{R}} q(x) l(1, x) < \infty$$

for all $y \in \mathbb{R}$, which implies that

$$\mathbb{E} \int_{\mathbb{R}} q(x + X_2^\circ) l(1, x) dx \leq \mathbb{E} \int_{\mathbb{R}} q(x) l(1, x) < \infty. \quad (\text{B.101})$$

Now we may easily deduce from (B.100) and (B.101) that the second term of (B.94) has finite moments to arbitrary order. The proof for (B.86) is therefore complete.

Proof of Lemma 3.8

Here, we will consider each block of the Hessian, $\mathcal{H}_{\alpha\alpha'}(\theta)$, $\mathcal{H}_{\beta\beta'}(\theta)$ and $\mathcal{H}_{\alpha\beta'}(\theta)$ separately. For $\mathcal{H}_{\alpha\alpha'}(\theta)$, from the expansion of the Hessian derived in the same way as in the proof of Lemma 3.2 using Itô's lemma and Lemmas A2 and A5, we have

$$\begin{aligned} \mathcal{H}_{\alpha\alpha'}(\theta) &= \frac{\Delta}{2} \sum_{i=1}^n \mathcal{A}^2 \ell_{\alpha\alpha'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}, \theta) \\ &\quad + \frac{1}{2} \sum_{i=1}^n (\mathcal{A}\mathcal{B} + \mathcal{B}\mathcal{A}) \ell_{\alpha\alpha'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}, \theta) (W_{i\Delta} - W_{(i-1)\Delta}) + O_p(\sqrt{\Delta} T^{4pq+1}) \\ &= \int_0^T f(X_t, \theta) dt + \int_0^T g(X_t, \theta) dW_t + O_p(\sqrt{\Delta} T^{4pq+1}) \\ &= P_T + Q_T + O_p(\sqrt{\Delta} T^{4pq+1}). \end{aligned}$$

where $f(x, \theta) = f_1(x, \theta) + f_2(x, \theta)$ with

$$\begin{aligned} f_1(x, \theta) &= \mu(x) \frac{\mu_{\alpha\alpha'}}{\sigma^2}(x, \theta) - \frac{(\mu\mu_{\alpha\alpha'} + \mu_{\alpha}\mu'_{\alpha})}{\sigma^2}(x, \theta), \\ f_2(x, \theta) &= \frac{1}{2}(\sigma^2(x) - \sigma^2(x, \beta))\ell_{ty^2\alpha\alpha'}(x, \theta), \end{aligned}$$

and $g(x, \theta) = \sigma(x)(\mu_{\alpha\alpha'}/\sigma^2)(x, \theta)$ since $\mathcal{A}^2\ell_{\alpha\alpha'}(0, x, x, \theta) = 2f(x, \theta)$ and $(\mathcal{AB} + \mathcal{BA})\ell_{\alpha\alpha'}(0, x, x, \theta) = 2g(x, \theta)$. For the part involving P_T ,

$$\begin{aligned} &\sup_{\theta \in \mathcal{N}} \left\| v_{\alpha}^{-1} \int_0^T (f(X_t, \theta) - f(X_t, \theta_0)) dt v_{\alpha}^{-1} \right\| \\ &= \sup_{\theta \in \mathcal{N}} \left\| v_{\alpha}^{-1} \int_0^T \left(f_1(X_t, \theta) - f_1(X_t, \theta_0) + f_2(X_t, \theta) \right) dt v_{\alpha}^{-1} \right\|, \end{aligned}$$

and choosing v_{α} and v_{β} such that $\text{diag}(v_{\alpha}, \Delta^{-1/2}v_{\beta}) = T^{-\epsilon}w$,

$$\begin{aligned} &\sup_{\theta \in \mathcal{N}} \left\| v_{\alpha}^{-1} \int_0^T (f_1(X_t, \theta) - f_1(X_t, \theta_0)) dt v_{\alpha}^{-1} \right\| \\ &\leq \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon}(w_{\alpha} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} \int_0^T f_{1\alpha}(X_t, \theta) dt \right\| + \sqrt{\Delta} \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon}(w_{\beta} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} \int_0^T f_{1\beta}(X_t, \theta) dt \right\| \\ &= \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon}(w_{\alpha} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} \int_0^T f_{1\alpha}(X_t, \theta) dt \right\| + O_p(\sqrt{\Delta}T^{4pq+1}), \end{aligned} \quad (\text{B.102})$$

where

$$\begin{aligned} f_{1\alpha}(x, \theta) &= \mu(x) \frac{\mu_{\alpha\alpha\alpha\alpha}}{\sigma^2}(x, \theta) - \frac{\mu\mu_{\alpha\alpha\alpha\alpha} + \mu_{\alpha\alpha\alpha} \otimes \mu_{\alpha} + \mu_{\alpha} \otimes \mu_{\alpha\alpha\alpha} + (I_{\alpha} \otimes C_{\alpha})(\mu_{\alpha\alpha\alpha} \otimes \mu_{\alpha})}{\sigma^2}(x, \theta), \\ f_{1\beta}(x, \theta) &= -2\mu(x) \frac{\mu_{\alpha\alpha\alpha} \otimes \sigma_{\beta}}{\sigma^3}(x, \theta) + 2 \frac{(\mu\mu_{\alpha\alpha\alpha} + \mu_{\alpha} \otimes \mu_{\alpha}) \otimes \sigma_{\beta}}{\sigma^3}(x, \theta). \end{aligned}$$

We also have

$$\begin{aligned} &\sup_{\theta \in \mathcal{N}} \left\| v_{\alpha}^{-1} \int_0^T f_2(X_t, \theta) dt v_{\alpha}^{-1} \right\| \leq \frac{\sqrt{\Delta}}{2} \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon}(w_{\beta} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} \int_0^T f_{2\beta}(X_t, \theta) dt \right\| \\ &= O_p(\sqrt{\Delta}T^{4pq+1}), \end{aligned} \quad (\text{B.103})$$

where $f_{2\beta}(x, \theta) = \ell_{ty^2\alpha\alpha\alpha} \otimes \sigma\sigma_{\beta}(x, \theta)$, and both (B.102) and (B.103) converge to zero in probability from the stated conditions and Assumption 3.1(d). For the part involving Q_T , similarly as in (B.102),

$$\begin{aligned} &\sup_{\theta \in \mathcal{N}} \left\| v_{\alpha}^{-1} \int_0^T (g(X_t, \theta) - g(X_t, \theta_0)) dW_t v_{\alpha}^{-1} \right\| \\ &\leq \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon}(w_{\alpha} \otimes w_{\alpha} \otimes w_{\alpha})^{-1} \int_0^T g_{\alpha}(X_t, \theta) dW_t \right\| + O_p(\sqrt{\Delta}T^{4pq+1}), \end{aligned}$$

where $g_{\alpha}(x, \theta) = \sigma(x)\mu_{\alpha\alpha\alpha\alpha}\sigma^{-2}(x, \theta)$, and this also converges to zero in probability from the assumption.

For $\mathcal{H}_{\beta\beta'}(\theta)$, we have

$$\begin{aligned} \mathcal{H}_{\beta\beta'}(\theta) &= \sum_{i=1}^n \mathcal{A}\ell_{\beta\beta'}(0, X_{(i-1)\Delta}, X_{(i-1)\Delta}, \theta) + O_p(\Delta^{-1/2}T^{4pq+1}) \\ &= \frac{1}{\Delta} \int_0^T h(X_t, \beta) dt + O_p(\Delta^{-1/2}T^{4pq+1}), \end{aligned}$$

where

$$h(x, \beta) = \sigma^2(x) \left(\frac{\sigma_{\beta\beta'}}{\sigma^3} - \frac{3\sigma_{\beta}\sigma'_{\beta}}{\sigma^4} \right) (x, \beta) - \left(\frac{\sigma_{\beta\beta'}}{\sigma} - \frac{\sigma_{\beta}\sigma'_{\beta}}{\sigma} \right) (x, \beta)$$

from Lemma A5, and

$$\begin{aligned} \sup_{\theta \in \mathcal{N}} \left\| \Delta v_{\beta}^{-1} \frac{1}{\Delta} \int_0^T (h(X_t, \beta) - h(X_t, \beta_0)) dt v_{\beta}^{-1} \right\| &\leq \sqrt{\Delta} \sup_{\theta \in \mathcal{N}} \left\| T^{3\epsilon} (w_{\beta} \otimes w_{\beta} \otimes w_{\beta})^{-1} \int_0^T h_{\beta}(X_t, \beta) dt \right\| \\ &= O_p(\sqrt{\Delta} T^{4pq+1}) \rightarrow_p 0, \end{aligned}$$

where

$$\begin{aligned} h_{\beta}(x, \beta) &= \sigma^2(x) \left(\frac{\sigma_{\beta \otimes \beta \otimes \beta}}{\sigma^3} - \frac{3\sigma_{\beta} \otimes \sigma_{\beta \otimes \beta}}{\sigma^4} - \frac{3[\sigma_{\beta \otimes \beta} \otimes \sigma_{\beta} + (I_{\alpha} \otimes C_{\alpha})(\sigma_{\beta \otimes \beta} \otimes \sigma_{\beta})]}{\sigma^4} + \frac{12\sigma_{\beta} \otimes \sigma_{\beta} \otimes \sigma_{\beta}}{\sigma^5} \right) (x, \beta) \\ &\quad - \left(\frac{\sigma_{\beta \otimes \beta \otimes \beta}}{\sigma} - \frac{\sigma_{\beta} \otimes \sigma_{\beta \otimes \beta}}{\sigma^2} - \frac{\sigma_{\beta \otimes \beta} \otimes \sigma_{\beta} + (I_{\alpha} \otimes C_{\alpha})(\sigma_{\beta \otimes \beta} \otimes \sigma_{\beta})}{\sigma} + \frac{\sigma_{\beta} \otimes \sigma_{\beta} \otimes \sigma_{\beta}}{\sigma^2} \right) (x, \beta). \end{aligned}$$

The case for the off-diagonal blocks of the Hessian $\mathcal{H}_{\alpha\beta'}(\theta)$ is similar to the one for $\mathcal{H}_{\alpha\alpha'}(\theta)$, and we can find $\varepsilon > 0$ such that

$$T^{\varepsilon} \sup_{\theta \in \mathcal{N}} \left| \sqrt{\Delta} w_{\alpha}^{-1}(T) (\mathcal{H}_{1\alpha\beta'}(\theta) - \mathcal{H}_{1\alpha\beta'}(\theta_1^*)) w_{\beta}^{-1'}(T) \right| = O_p(\sqrt{\Delta} T^{4pq+1+\varepsilon}) \rightarrow_p 0$$

from Assumption 3.1(d).

Proof of Theorem 4.1

AD1, AD2 and AD3 hold with Assumptions 3.1-3.3 and Lemmas 3.6, 3.8, thus from (20) we can obtain the stated result.

Proof of Corollary 4.2

We have

$$\begin{aligned} w_{\alpha}(T)(\hat{\alpha} - \alpha_0) &\sim_p \left(w_{\alpha}^{-1}(T) \int_0^T \frac{\mu_{\alpha}\mu'_{\alpha}}{\sigma^2}(X_t) dt w_{\alpha}^{-1}(T) \right)^{-1} w_{\alpha}^{-1}(T) \int_0^T \frac{\mu_{\alpha}}{\sigma}(X_t) dW_t = O_p(1), \quad (\text{B.104}) \\ \sqrt{\frac{2}{\Delta}} w_{\beta}(T)(\hat{\beta} - \beta_0) &\sim_p \left(w_{\beta}^{-1}(T) \int_0^T \frac{\sigma_{\beta}\sigma'_{\beta}}{\sigma^2}(X_t) dt w_{\beta}^{-1}(T) \right)^{-1} w_{\beta}^{-1}(T) \int_0^T \frac{\sigma_{\beta}}{\sigma}(X_t) dV_t = O_p(1) \end{aligned}$$

from Theorem 4.1 and Assumption 3.2. Since $w_{\alpha}(T) \rightarrow \infty$ and $\Delta^{-1/2} w_{\beta}(T) \rightarrow \infty$, $\hat{\alpha}$ and $\hat{\beta}$ are consistent.

Proof of Theorem 4.3 and 4.4

We can obtain the stated result by applying Proposition 3.3 (or 3.5(a) for Theorem 4.4) to each term of (B.104) with $w_{\alpha}(T) = w_{\beta}(T) = \sqrt{T}$ (or $\sqrt{T^{1/(r+2)}}$ for Theorem 4.4).

Proof of Theorem 4.5

Let V° be defined as a limiting process such that $T^{-1/2}V_{Tt} \rightarrow_d V_t^\circ$ as $T \rightarrow \infty$ for $t \geq 0$. Then the stated result follows from Theorem 3.5(b) and Theorem 4.1, together with the independency of V° with B and W° . For those independencies, it suffices to show that

$$\mathbb{E}(V_t^T B_t^{rT}) = 0, \quad \mathbb{E}(V_t^T W_t^T) = 0 \quad (\text{B.105})$$

for all $t \in [0, 1]$ and $T > 0$ from Exercise IV.2.22 and Exercise V.4.25 of Revuz and Yor (1999), where $V_t^T = T^{-1/2}V_{Tt}$. V is independent of W , and therefore of X as well, consequently V and B^{rT} are independent with each other since B^{rT} is given by $B_t^{rT} = T_r^{-1}(X \circ \varsigma)_{T_r^2 t}$, where $\varsigma_t = \inf\{s \mid \int_0^s \sigma^2(X_r)dr > t\}$ from the DDS Brownian motion representation. We can deduce (B.105) from the independency of V with B^{rT} and W , therefore V° is independent of B and W° , which completes the proof.

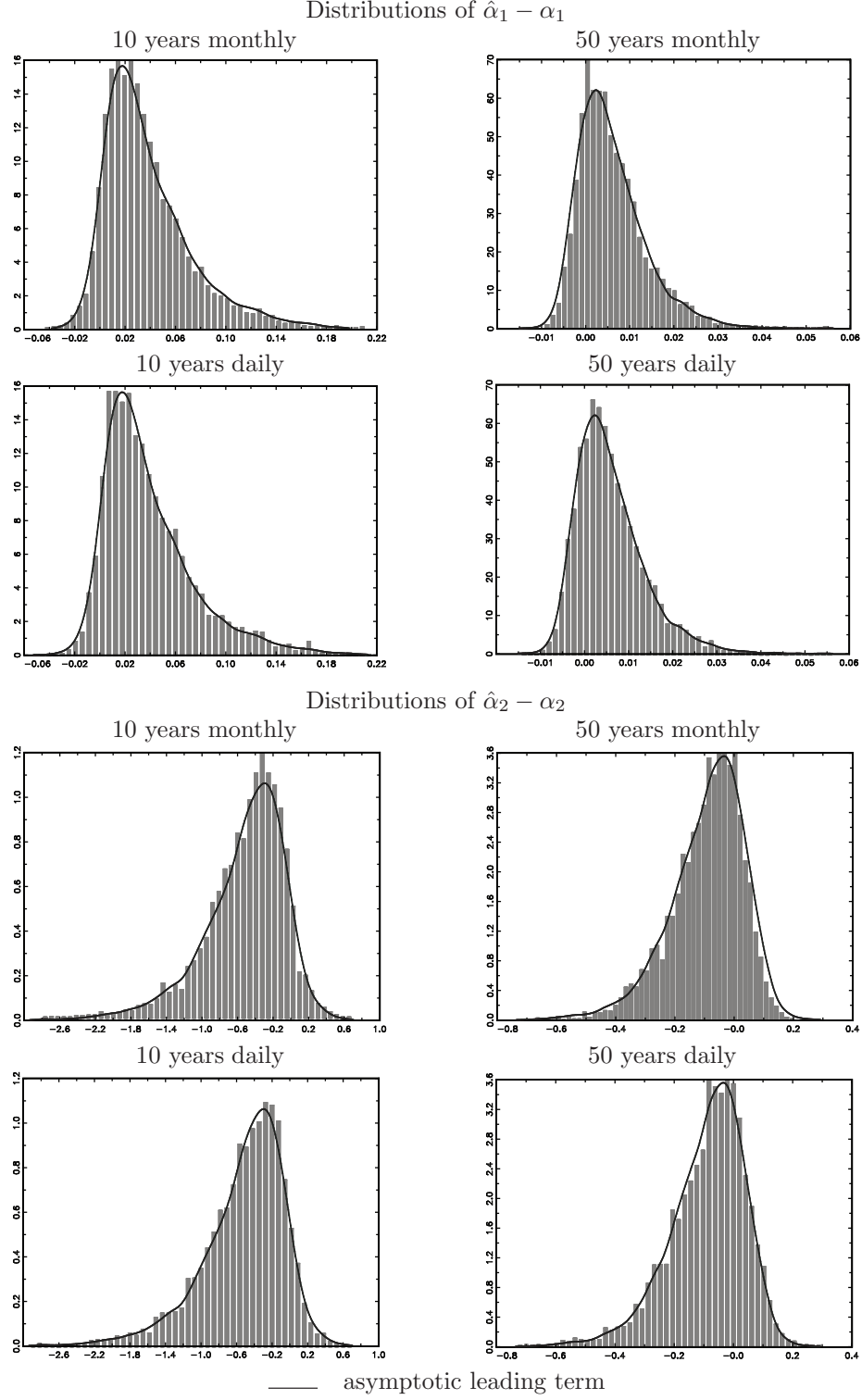
References

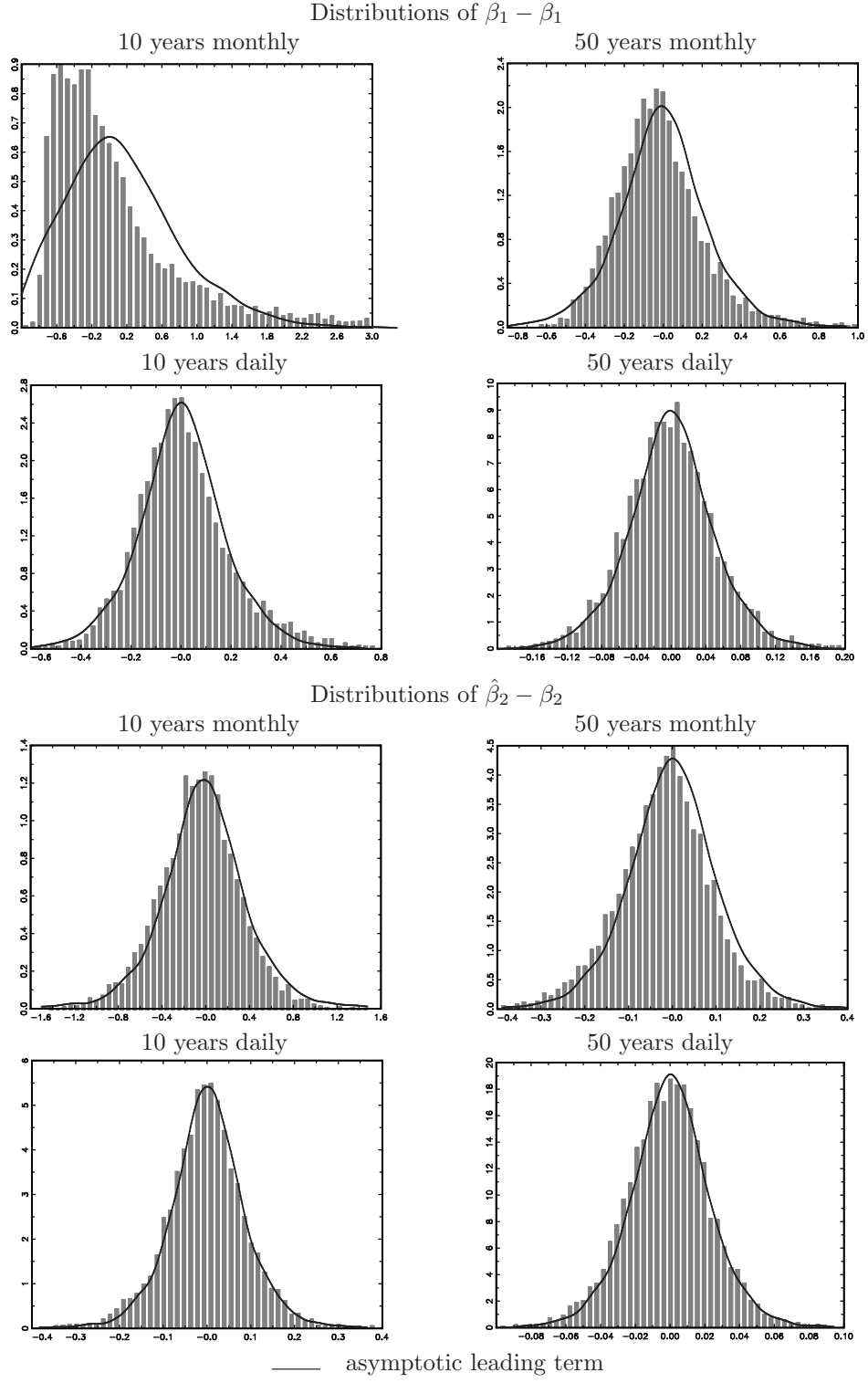
- AHN, D.-H., AND B. GAO (1999): "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, 12, 721-762.
- AÏT-SAHALIA, Y. (1996): "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385-426.
- (1999): "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361-1395.
- (2002): "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223-262.
- BANDI, F. M., AND G. MOLOCHE (2004): "On the Functional Estimation of Multivariate Diffusion Processes," Working paper.
- BANDI, F. M., AND P. C. B. PHILLIPS (2003): "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241-283.
- (2010): "Nonstationary Continuous-Time Processes," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. P. Hansen. Amsterdam, North Holland, pp. 139-201.
- BARTLE, R. G. (2001): *A Modern Theory of Integration*, American Mathematical Society.
- BINGHAM, N. H. (1971): "Limit Theorems for Occupation Times of Markov Processes," *Probability Theory and Related Fields*, 17, 1-22.
- BINGHAM, N. H., C. M. GOLDIE, AND J. L. TEUGELS (1993): *Regular Variation*, Cambridge University Press, Cambridge.
- BORKOVEC, M., AND C. KLÜPPELBERG (1998): "Extremal Behaviour of Diffusion Models in Finance," *Extremes*, 1, 47-80.
- BRANDT M. W. AND P. SANTA-CLARA (2002): "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, 63, 161-210.
- CLINE, D. B.H., AND M. JEONG (2009): "Limit Theorems for the Integrals of Diffusion Processes," Working paper.
- COCHRANE, J. H. (2005): *Asset Pricing*, Princeton University Press, NJ.

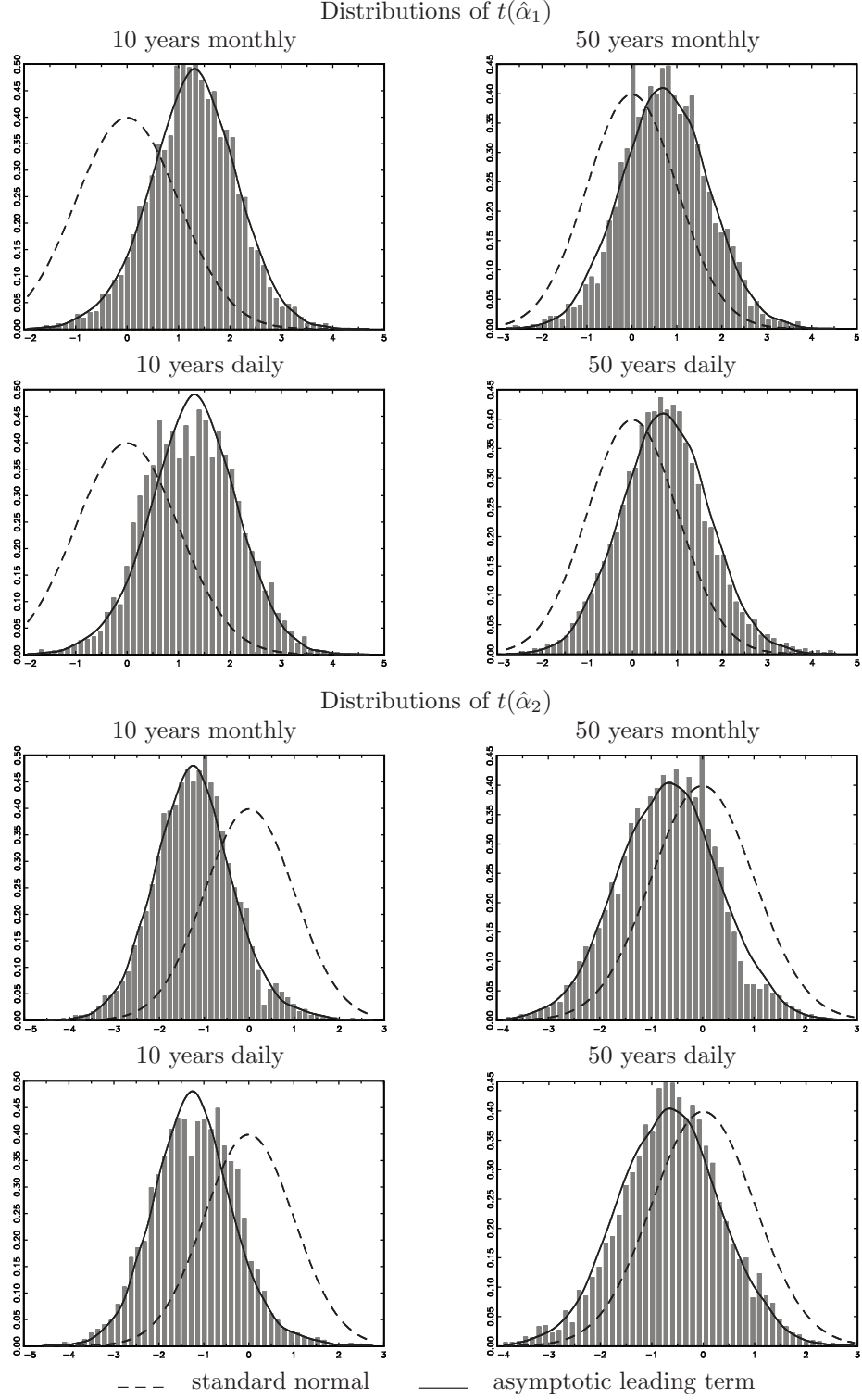
- COX, J. C., J. E. INGERSOLL, AND S. A. ROSS (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.
- DAVIS, R. A. (1982): "Maximum and Minimum of One-Dimensional Diffusions," *Stochastic Processes and their Applications*, 13, 1-9.
- DUFFIE, D. (2001): *Dynamic Asset Pricing Theory*, Princeton University Press, NJ.
- DURHAM, G. B., AND A. R. GALLANT (2002): "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," *Journal of Business and Economic Statistics*, 20, 297-316.
- ELERIAN, O. (1998): "A Note on the Existence of a Closed Form Conditional Transition Density for the Milstein Scheme," Economics Discussion Paper 1998-W18, Nuffield College, Oxford.
- FRIEDMAN, A. (1964): *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, N. J.
- GIHMAN, I. I. AND A. V. SKOROHOD (1972): *Stochastic Differential Equations*, Springer-Verlag, New York.
- HONG, Y., AND H. LI (2005): "Nonparametric Specification Testing for Continuous-Time Models with Applications to Interest Rate Term Structures," *Review of Financial Studies*, 18, 37-84.
- HÖPFNER, R. (1990): "Null Recurrent Birth-and-Death Processes, Limits of Certain Martingales, and Local Asymptotic Mixed Normality," *Scandinavian Journal of Statistics*, 17, 201-215.
- HÖPFNER, R., AND E. LÖCHERBACH (2003): "Limit Theorems for Null Recurrent Markov Processes," *Memoirs of the AMS*, 768, Providence, Rhode Island.
- ITÔ, K., AND H. P. MCKEAN (1996): *Diffusion Processes and their Sample Paths*, Springer-Verlag, New York.
- KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*, New York, Springer-Verlag.
- KARLIN, S., AND H. M. TAYLOR (1981): *A Second Course in Stochastic Processes*, New York, Academic Press.
- KESSLER, M. (1997): "Estimation of an Ergodic Diffusion from Discrete Observations," *Scandinavian Journal of Statistics*, 24, 211-229.
- KOTANI, S., AND S. WATANABE (1982): "Krein's Spectral Theory of Strings and Generalized Diffusion Processes," *Functional Analysis in Markov Processes*, Lecture Notes in Mathematics, 923, 235-259.
- NICOLAU, J. A. O. (2002): "A New Technique for Simulating the Likelihood of Stochastic Differential Equations," *Econometrics Journal*, 5, 91-103.
- PARK, J. Y., AND P. C. B. PHILLIPS (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117-161.
- PEDERSEN, A. R. (1995): "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 257-279.
- PHILLIPS, P. C. B., AND J. YU (2009): "Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance," *Handbook of Financial Time Series*, New York, Springer.
- REVUZ, D. AND M. YOR (1999): *Continuous Martingale and Brownian Motion*, New York, Springer-Verlag.
- ROGERS, L. C. G. AND D. WILLIAMS (2000): *Diffusions, Markov Processes, and Martingales*,

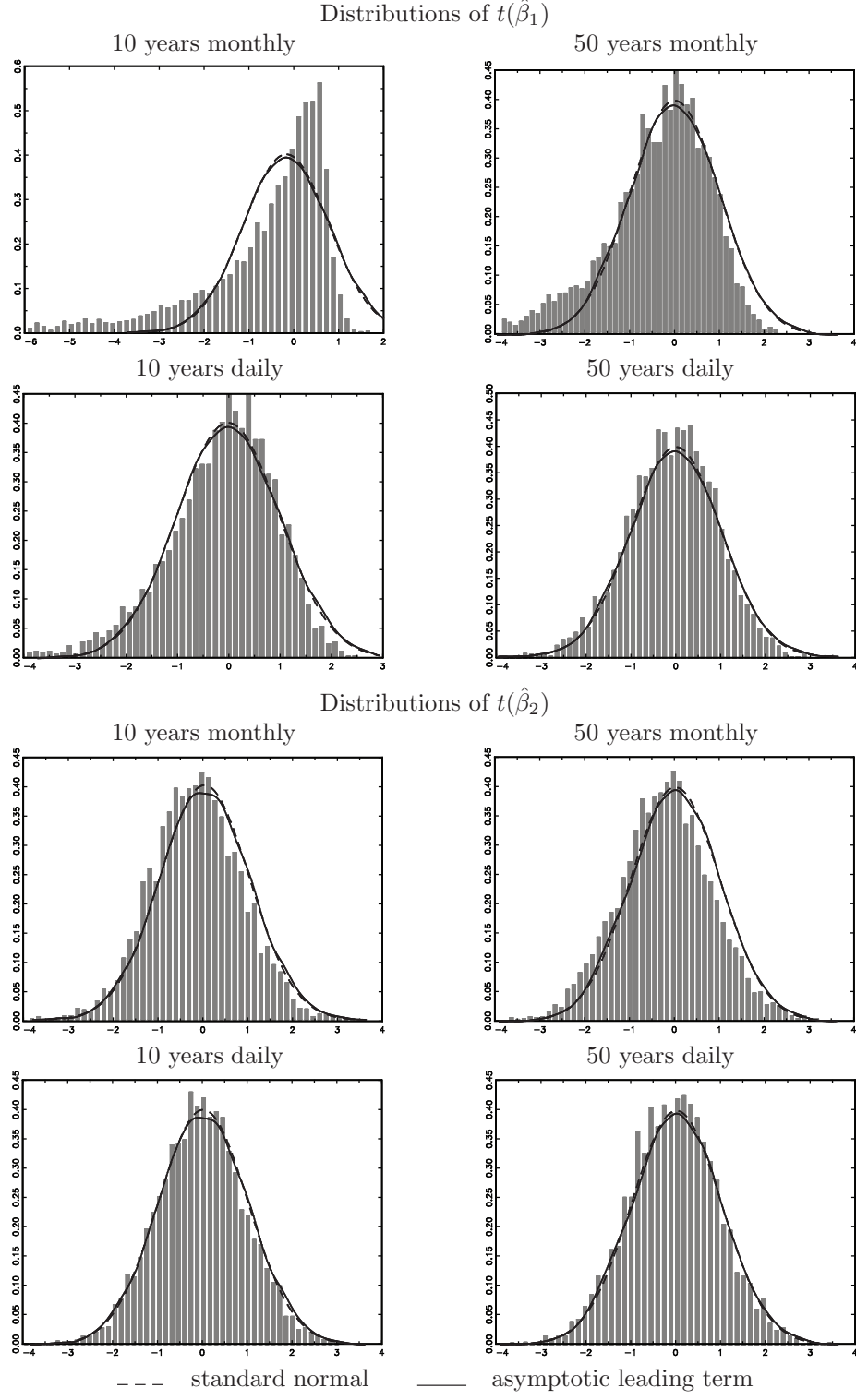
Cambridge University Press.

- STONE, C. (1963): "Limit Theorems for Random Walks, Birth and Death Processes, and Diffusion Processes," *Illinois Journal of Mathematics*, 7, 638-660.
- VAN ZANTEN, H. (2000): "A Multivariate Central Limit Theorem for Continuous Local Martingales," *Statistics and Probability Letters*, 50, 229-235.
- VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.
- WATANABE, S. (1995): "Generalized Arc-Sine Laws for One-Dimensional Diffusion Processes and Random Walks," *Proceedings of Symposia in Pure Mathematics*, 57, 157-172.
- WOOLDRIDGE, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. Amsterdam, North Holland, pp. 2639-2738.

Figure 1: Finite Sample Distributions of $\hat{\alpha} - \alpha$

Figure 2: Finite Sample Distributions of $\hat{\beta} - \beta$

Figure 3: Finite Sample Distributions of $t(\hat{\alpha})$

Figure 4: Finite Sample Distributions of $t(\hat{\beta})$

		Uncorrected				Corrected			
		α_1	α_2	β_1	β_2	α_1	α_2	β_1	β_2
10 years	Bias	0.04082	-0.55809	0.00885	-0.00496	-0.00061	0.00264	0.00936	-0.00477
	(%)	(567.0%)	(620.1%)	(1.1%)	(0.3%)	(8.4%)	(2.9%)	(1.2%)	(0.3%)
	SD	0.03590	0.45166	0.18887	0.08480	0.03590	0.45166	0.18887	0.08480
	RMSE	0.05436	0.71796	0.18908	0.08495	0.03591	0.45167	0.18911	0.08494
50 years	Bias	0.00639	-0.10022	-0.00021	-0.00063	0.00004	-0.00080	0.00018	-0.00044
	(%)	(88.7%)	(111.4%)	(0.03%)	(0.04%)	(0.6%)	(0.9%)	(0.02%)	(0.03%)
	SD	0.00762	0.12177	0.05151	0.02341	0.00762	0.12177	0.05151	0.02341
	RMSE	0.00994	0.15771	0.05151	0.02342	0.00762	0.12177	0.05151	0.02342

Table 1: Bias Correction for MLE

		Nominal size		Actual size					
		Uncorrected				Corrected			
		α_1	α_2	β_1	β_2	α_1	α_2	β_1	β_2
10 years	1%	0.070	0.066	0.025	0.009	0.008	0.007	0.029	0.010
	5%	0.202	0.189	0.062	0.049	0.052	0.052	0.062	0.048
	10%	0.312	0.296	0.102	0.095	0.103	0.105	0.100	0.093
50 years	1%	0.029	0.028	0.011	0.011	0.012	0.013	0.009	0.011
	5%	0.101	0.092	0.049	0.048	0.056	0.051	0.045	0.046
	10%	0.177	0.166	0.096	0.099	0.105	0.099	0.095	0.096

Table 2: Size Correction for t -Test