ASYMPTOTICS FOR THE MAXIMUM LIKELIHOOD ESTIMATORS OF DIFFUSION MODELS

A Dissertation

by

MINSOO JEONG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2008

Major Subject: Economics

ASYMPTOTICS FOR THE MAXIMUM LIKELIHOOD ESTIMATORS OF DIFFUSION MODELS

A Dissertation

by

MINSOO JEONG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Joon Y. Park
Committee Members,	Yoosoon Chang
	Qi Li
	Daren B.H. Cline
Head of Department,	Larry Oliver

December 2008

Major Subject: Economics

ABSTRACT

Asymptotics for the Maximum Likelihood Estimators
of Diffusion Models. (December 2008)

Minsoo Jeong, B.A., M.A., Seoul National University
Chair of Advisory Committee: Joon Y. Park

In this paper I derive the asymptotics of the exact, Euler, and Milstein ML estimators for diffusion models, including general nonstationary diffusions. Though there have been many estimators for the diffusion model, their asymptotic properties were generally unknown. This is especially true for the nonstationary processes, even though they are usually far from the standard ones. Using a new asymptotics with respect to both the time span T and the sampling interval Δ, I find the asymptotics of the estimators and also derive the conditions for the consistency. With this new asymptotic result, I could show that this result can explain the properties of the estimators more correctly than the existing asymptotics with respect only to the sample size n. I also show that there are many possibilities to get a better estimator utilizing this asymptotic result with a couple of examples, and in the second part of the paper, I derive the higher order asymptotics which can be used in the bootstrap analysis.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Park, and my committee members, Dr. Cline, Dr. Chang, and Dr. Li, for their patience and guidance throughout the course of this research. Thanks also to Yongok Choi and Daehee Jeong, and to other friends and colleagues at Texas A&M University and Rice University for the helpful discussions and suggestions. Finally, thanks to my parents for their endless love and encouragement, without which I could have not managed to finish my dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The diffusion model was originally designed and has long been used to model the stochastic dynamics arising in physics and biology. In recent decades, however, it also has gotten much attention from the financial and economics fields, and they applied the diffusion to the various financial and economics problems. Merton (1971) and Black and Scholes (1973) are the most popular and significant works which established the foundation of option pricing theory in finance. Vasicek (1977) and Cox, Ingersoll and Ross (1985) are also well known works which have considered the diffusion processes to model the interest rate term structure. Nowadays most of the financial theories are written in terms of the continuous time framework, so the importance of the diffusion model cannot be emphasized more.

As representing the importance and the popularity of the model, numerous estimation methods have been proposed, among which the main consideration in this paper is the maximum likelihood estimation. Unlike the discrete time model estimation, the main difficulties in the estimation of the diffusion model arises from the fact that we cannot obtain the transition density in a closed form solution in most of the cases, so we need to approximate it to do the estimation. The Euler scheme is the most easiest and simplest way of the approximation, while the Milstein scheme gives us a finer result with a higher order approximation of the data generating process. There are also other various approximation methods proposed by many literatures, and among them, Aït-Sahalia (2002)'s method is one of the most popular methods in practice.

The journal model is *Econometrica.*

For each of those estimation methods, the corresponding asymptotic theories were also provided, but mostly they could only deal with the stationary cases with a few exceptions. We were mostly interested in the stationary processes in the past, but in recent years people are getting more and more doubtful about the stationary assumption even for the basic financial processes such as the interest rate or the exchange rate processes. Moreover, the existing asymptotics with respect to the number of samples is not enough to deal with the continuous time processes such as the diffusion model. For example, it has been long been noted that there is a huge magnitude of bias in the drift term parameter estimation of the diffusion models, but it was just a well known phenomenon without reasonable asymptotic theory that can explain it. In this paper, I propose a new asymptotic theory that can address this problem, also without a restrictive stationary assumption. The basic concept for this new asymptotics has mostly come from the ideas in Park and Phillips (2001), Aït-Sahalia and Park (2008a) and Aït-Sahalia and Park (2008b). For the introduction and the background theories of the diffusion processes, readers are recommended to refer to Karlin and Taylor (1981), Revuz and Yor (1999) and Karatzas and Shreve (1991).

In Chapter II, I derive the first order asymptotics of the exact, Euler and Milstein maximum likelihood estimator of the diffusion models, and in Chapter III, I derive the higher order asymptotics for the estimators. Various examples for the popular diffusion models in finance and economics are also illustrated. In the Appendix, the proofs for the theorems in the paper and other useful lemmas to derive them are introduced.

CHAPTER II

ASYMPTOTICS FOR THE MAXIMUM LIKELIHOOD ESTIMATORS OF DIFFUSION MODELS

In the first chapter, I deal with the first order asymptotics of the Maximum Likelihood estimators of diffusion models.

A. Background

Consider the time-homogeneous stochastic differential equation

$$dX_t = \mu(X_t, \alpha)dt + \sigma(X_t, \beta)dW_t \tag{2.1}$$

where μ and σ are the drift and diffusion functions, respectively. I will denote $\theta = (\alpha', \beta')'$ hereafter. I let $\mathcal{D} = (\underline{x}, \bar{x})$ denotes the domain of the diffusion process X_t. The Euler approximation of this SDE is

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

and the closed-form solution of this approximated transition density from x to y with an interval Δ is given by

$$p_{_E}(x, y) = \frac{1}{\sqrt{2\pi\Delta}\,\sigma(x)} \exp\left[-\frac{\big(y - x - \Delta\mu(x)\big)^2}{2\Delta\sigma^2(x)} \right]$$

suppressing the parameter arguments for each function. Milstein approximation of this SDE is

$$\begin{aligned} X_{i\Delta} - X_{(i-1)\Delta} \simeq\ & \mu(X_{(i-1)\Delta})\Delta + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\ & + \frac{1}{2}\sigma\sigma^{\cdot}(X_{(i-1)\Delta})\big[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta\big], \end{aligned}$$

where $f^{\cdot}(x,\theta)$ denotes a derivative $\partial/\partial x\, f(x,\theta)$. We denote $f_\theta(x,\theta)$ as a derivative with respect to the parameter, $\partial/\partial\theta\, f(x,\theta)$. In the case of the Euler approximation, the approximated transition density is the normal distribution, but in the case of the Milstein approximation, the approximation error is reduced more with a mixture of a normal and a chi-squared distribution, and the approximated transition density from x to y with an interval Δ is given by,

$$p_{_M}(x,y) = \frac{1}{\sqrt{2\pi\Delta}\,\tau(x,y)}\left(\exp\left[\frac{-\big(\tau(x,y)+\sigma(x)\big)^2}{2\Delta\sigma^2\sigma^{\cdot 2}(x)}\right] + \exp\left[\frac{-\big(\tau(x,y)-\sigma(x)\big)^2}{2\Delta\sigma^2\sigma^{\cdot 2}(x)}\right]\right),$$

where

$$\tau(x,y) = \big[\sigma^2(x) + \Delta\,\sigma^2\sigma^{\cdot 2}(x) + 2\,\sigma\sigma^{\cdot}(x)\big(y - x - \Delta\,\mu(x)\big)\big]^{1/2}$$

suppressing the parameter arguments for each function.

With a sample of time span T and the sampling interval Δ, the Euler and Milstein ML estimator $\hat\theta$ is defined as an estimator which minimizes the log-likelihood function

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

over $\theta \in \Theta$, where $n = T/\Delta$, i.e.,

$$\hat\theta = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \mathcal{L}(\theta).$$

Here $\hat p$ represents either $p_{_E}$ or $p_{_M}$. We assume that Θ is compact and convex, and θ_0 is an interior point of Θ. The Milstein ML estimation method was first proposed in Elerian (1998). Replacing $\hat p$ with the true transition density p, we can perform the exact ML estimation, but it is only restricted to the cases when we know the true transition density in a closed-form, such as Ornstein-Uhlenbeck, Feller's square root, and Brownian motion with drift.

Letting $\mathcal{S} = \partial\mathcal{L}/\partial\theta$ and $\mathcal{H} = \partial^2\mathcal{L}/\partial\theta\partial\theta'$, the asymptotic distribution of $\hat\theta$ can be obtained from the first order Taylor expansion of $\mathcal{S}$, which is written as

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

where $\tilde{\theta}$ lies in the line segment connecting $\hat{\theta}$ and θ_0. If the following conditions hold as $T \to \infty$ and $\Delta \to 0$ for some appropriate matrix sequence w, (w is a function of both T and Δ but I will suppress the subscript for the simplicity.)

AD1: $w^{-1}\mathcal{S}(\theta_0) = O_p(1)$.

AD2: $w^{-1}\mathcal{H}(\theta_0)w^{-1\prime} = O_p(1)$ and $w'\mathcal{H}^{-1}(\theta_0)w = O_p(1)$.

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

$$\sup_{\theta \in \mathcal{N}} \left| v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1\prime} \right| \to_p 0,$$

where $\mathcal{N} = \{\theta : |v'(\theta - \theta_0)| \leq 1\}$. ($v$ is also a function of both T and Δ.)

we can derive the asymptotic leading term of the estimator. Wooldridge (1994) shows that AD3 together with AD1 and AD2 implies[1]

AD4: $\mathcal{S}(\hat{\theta}) = 0$ with probability approaching to one as $T \to \infty$ and $\Delta \to 0$.

AD5: $w^{-1}\big(\mathcal{H}(\tilde{\theta}) - \mathcal{H}(\theta_0)\big)w^{-1\prime} = o_p(1)$ and $w'(\hat{\theta} - \theta_0) = O_p(1)$.

Thus, with these conditions, we have

$$\begin{aligned} w^{-1}\mathcal{S}(\hat{\theta}) &= w^{-1}\mathcal{S}(\theta_0) + w^{-1}\mathcal{H}(\theta_0)w^{-1\prime}w'(\hat{\theta} - \theta_0) + w^{-1}\big(\mathcal{H}(\tilde{\theta}) - \mathcal{H}(\theta_0)\big)w^{-1\prime}w'(\hat{\theta} - \theta_0) \\ &= w^{-1}\mathcal{S}(\theta_0) + w^{-1}\mathcal{H}(\theta_0)w^{-1\prime}w'(\hat{\theta} - \theta_0) + o_p(1) \end{aligned}$$

[1] Weak dependency is originally assumed to show the asymptotic normality of the estimator, but it turns out that without the weak dependency condition, we can still show AD4 and AD5 as long as we can find a proper normalizing sequence w.

so with probability approaching to one, $w^{-1}\mathcal{S}(\hat{\theta}) = 0$ and

$$w'(\hat{\theta} - \theta_0) = -\, w'\mathcal{H}(\theta_0)^{-1}\mathcal{S}(\theta_0) + o_p(1).$$

So the rest of the steps are just to find the leading terms of $\mathcal{H}(\theta_0)$ and $\mathcal{S}(\theta_0)$.

B. Assumptions

1. Assumption Set 1

This set of assumptions to show the asymptotics of the Euler and Milstein ML Estimators. We assume the following assumptions to make AD1 - AD3 hold.

Assumption 1. *$\mu(x,\alpha)$ has its derivatives up to 6th order, and $\sigma(x,\beta)$ has its derivatives up to the 7th order, w.r.t. x on $\mathcal{D}$. $\mu(x,\alpha)$ and $\sigma(x,\beta)$ and their derivatives w.r.t. x have their derivatives up to the 6th order, w.r.t. θ on the interior of Θ.*

Assumption 2. *Letting $f(x)$ be each of those functions in Assumption 13 or $\sigma^{-1}(x)$, $f(x)$ is locally bounded on the domain $\mathcal{D}$, and there exists a positive nondecreasing function κ_f such that*

$$\frac{1}{\kappa_f(T)} \sup_{t\in[0,T]} \left|f(X_t)\right| \to_p 0$$

$$T^{-p}\kappa_f(T) \to 0$$

as $T \to \infty$ for some $p < \infty$. We call κ_f as the asymptotic function of f.

This assumption is to get proper bounds for the remainder terms which appears in the derivation of the asymptotic terms. This can be guaranteed by the limit theorems of the extremal process of diffusions, together with the appropriate boundary conditions of the function f. For the properties of the extremal process of diffusion models, one can refer to Berman (1964), Davis (1982), and Stone (1963), and for the properties for the function f, if f is regularly varying at both boundaries of $\mathcal{D}$ then it often is possible to verify Assumption 2, as we will discuss further below.

To see more about this, note that the asymptotic property of $\sup_{t\in[0,T]}\left|f(X_t)\right|$ is determined by the asymptotic properties of $\sup_{t\in[0,T]}\left|X_t\right|$ and the supremum of the properly centered reciprocal of X_t, together with the boundary properties of the function f, so firstly we can use the following result in Davis (1982),

$$\lim_{T\to\infty}\left|\mathbf{P}\left(\sup_{t\in[0,T]}\left|X_t\right|\le u_T\right)-\exp\left(-\frac{T}{S(u_T)M(\mathcal{D})}\right)\right|=0$$

for any $u_T\to\bar{x}$, for the positive recurrent processes. S is the scale function and M is the speed measure of the process X_t. If we assume that $\mu(x)$ and $\sigma(x)$ are regularly varying at both boundaries, taking $u_T=T^{1+\varepsilon}$ for $\varepsilon\ge 0$, we always have $T/S(u_T)=O(1)$ so the extremal process normalized with such u_T always degenerates to zero, or has a non-degenerating distribution. For the properties of the reciprocal of X_t, we can apply Itô's lemma to get the drift and diffusion function of the transformed process first, and then we can apply the above result with the same manner. We will explain more about this in the examples later. For null recurrent processes, the derivation mostly depends on each case, but one can refer to Stone (1963) and Cline, Jeong and Park (2008) for the most general cases. Once we know the asymptotics of the suprema, rest steps are easy with the regular variation property of function f. It will be more explained in the examples later.

Assumption 3. *There exist positive nondecreasing functions w_α and w_β such that*

$$w_\alpha^{-2}(T)\int_0^T\frac{\mu_\alpha^2}{\sigma^2}(X_t)dt\quad\textit{and}\quad w_\beta^{-2}(T)\int_0^T\frac{\sigma_\beta^2}{\sigma^2}(X_t)dt$$

converge in distribution to some almost surely positive definite random variables as $T\to\infty$.

This can be easily shown for the positive recurrent processes with $w_\alpha(T)$ and $w_\beta(T)$ being $\sqrt{T}$, and for other cases, we can get reasonable conditions for it to hold as in Cline, Jeong and Park (2008), which utilizes the result in Stone (1963), Kasahara (1975) and Höpfner and Löcherbach (2003). It will be more dealt with in the examples and in Cline, Jeong and Park (2008). We let $w = \mathrm{Diag}\big(w_\alpha(T), \Delta^{-1/2} w_\beta(T)\big)$ hereafter.

Assumption 4. $\sigma^2(x) > 0$ *for any* $x \in \mathcal{D}$.

This is to guarantee the existence of the integrals of the function of the process, for example,

$$\int_0^T \frac{\mu\sigma_\beta}{\sigma^5}(X_t)dt < \infty,$$

which appears in the asymptotic expansions. The key point here is that what is in the denominator is always $\sigma(X)$, so the existence of the integral is guaranteed by the continuity of the process X_t, together with the local boundedness of μ and σ and their derivatives.

For Ornstein-Uhlenbeck process and Brownian motion, we can easily check this since the diffusion function is constant as $\sigma(x) = \beta$, so the above integral becomes

$$\int_0^T \frac{\mu(X_t)}{\beta^5} dt < \infty,$$

which is guaranteed from the continuity of X_t, local boundedness of $\mu(x)$, and $\sigma^2(x) > 0$. As for CEV or Feller's square root process, the above integral becomes

$$\int_0^T \frac{\mu(X_t)\log(X_t)}{X_t^{4\beta_2}} dt < \infty,$$

which is again guaranteed by the continuity of X_t, local boundedness of $\mu(x)$, and $\sigma^2(x) > 0$.

Assumption 5. *The asymptotic functions satisfy,*

$$\Delta T \to 0$$
$$\Delta^{1/4}\kappa_1(T\kappa_2(T)) \to 0$$

as $T \to \infty$ and $\Delta \to 0$, where κ_1 and κ_2 represent any combinations of the asymptotic functions in Assumption 2.

This Assumption requires that Δ should decrease fast enough as T increases. This is a technical condition for the proofs and it does not restrict the model. Though it seems to require a bunch of complicated conditions for the all possible combinations of κ_1 and κ_2, it turns out that we only need to check this condition for the fastest increasing function κ_f among others and it is not difficult to check.

Assumption 6. *As $T \to \infty$, we have*

$$T^{-\varepsilon}\frac{\dot{\kappa}(T)}{\kappa(T)} \to 0$$

for any $\varepsilon > 0$, where κ represents one of the asymptotic function κ_f in Assumption 2, and $\dot{\kappa}$ represents corresponding asymptotic function of the derivative of f with respect to the parameter.

This requires that the order difference between the derivatives is not too big, and it is of course satisfied by many functional classes, such as the power functions and the logarithmic function. It is also not difficult to check this condition since we only need to check for one or two functional classes which are related with the model. Any diffusion processes having polynomial drift and diffusion functions, such as Ornstein-Uhlenbeck, Feller's square root, Brownian motion with drift, CEV and AS-CEV of course satisfy this condition. More will be shown in the examples later.

Assumption 7. *Defining* $\mathcal{N}_{T,\Delta} = \{\theta : |v'(\theta - \theta_0)| \le 1\}$ *with* v *satisfying* $vw^{-1} \to 0$, *we have*

$$\sup_{\theta \in \mathcal{N}_{T,\Delta}} \left| \frac{\kappa(T, \theta_0)}{\kappa(T, \theta)} \right| \to 1$$

as $T \to \infty$ *subject to Assumption 5, where* κ *represents one of the asymptotic functions in Assumption 2. Hereafter I suppress the subscript such as* $\mathcal{N}$ *for the simplicity.*

This is also satisfied by many functional classes, including the power functions and the logarithmic function. Examples for these Assumption 6 and 7 will be dealt with more in Example 2. For Assumption 5-7, it looks as if at the first glance that it will be very complicated and troublesome to check all the conditions, but as in the Example 2, it turns out that we only need to check a few extremal cases for most of the diffusion models used in practice, and we only need to check the conditions for a functional class.

Example 1. *(Ornstein-Uhlenbeck):* Consider a process

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

with $\alpha_2 > 0$, $\beta > 0$ and $\mathcal{D} = (-\infty, \infty)$. It is easy to see that both the drift function $\mu(x) = \alpha_2(\alpha_1 - x)$ and the diffusion function $\sigma(x) = \beta$ satisfy the differentiability condition in the domain of the process $\mathcal{D}$. For Assumption 5, they are conditions for the decreasing rate of Δ, and it is satisfied if

$$\Delta T^4 \to 0$$

as $T \to \infty$ and $\Delta \to 0$. For Assumption 6, it is easy to check that

$$T^{-\varepsilon} \frac{\dot{\kappa}(T)}{\kappa(T)} = \frac{1}{T^{\varepsilon}} \to 0.$$

Here, Assumption 7 is also obvious since in this Ornstein-Uhlenbeck case, all the asymptotic order functions do not depend on the parameter value.

Example 2. *(CEV):* Consider a process

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

with $\alpha_1 > 0$, $\alpha_2 > 0$, $\beta_1 > 0$, $\beta_2 > 1/2$ also satisfying Assumption 4, and $\mathcal{D} = (0, \infty)$. It is also easy to see that both the drift function $\mu(x) = \alpha_2(\alpha_1 - x)$ and the diffusion function $\sigma(x) = \beta_1 x^{\beta_2}$ satisfy the differentiability condition in the domain of the process $\mathcal{D}$, and they all satisfy Assumption 2. Borkovec and Klüppelberg (1998) shows some examples of the properties of the extremal processes of the commonly used diffusion models, and we can check that the supremum of the CEV process can be bounded with a sequence $\nu(T) = T$. (The actual rate of $\nu(T)$ is different for each parameter setting, but here I only consider the biggest order for the simplicity.) Applying Itô's lemma, we can easily check that this also holds with the reciprocal of the process, that is, $\sup_{t\in[0,T]} \left|X_t^{-1}\right| = O_p(T)$ for $\beta_2 > 1/2$, since

$$dY_t = \left(\alpha_2 Y_t - \alpha_1\alpha_2 Y_t^2 + \beta_1^2 Y_t^{3-2\beta_2}\right)dt - \beta_1 Y_t^{2-\beta_2} dW_t,$$

denoting $Y_t = X_t^{-1}$. So Assumption 2 is satisfied for each μ and σ and their derivatives since they are all regularly varying at both boundaries. For example, if $f(x) = x^2$,

$$\sup_{t\in[0,T]} \left|X_t^2\right| \leq \left(\sup_{t\in[0,T]} \left|X_t\right|\right)^2 = O_p(T^2)$$

so $\kappa_f(x) = x^2$, and if $f(x) = 1/x^3$,

$$\sup_{t\in[0,T]} \left|X_t^{-3}\right| \leq \left(\sup_{t\in[0,T]} \left|X_t^{-1}\right|\right)^3 = O_p(T^3)$$

so $\kappa_f(x) = x^3$. For Assumption 3, refer to Cline, Jeong and Park (2008).

For Assumption 5, it is enough to check with the biggest order κ_f. When $1/2 < \beta_2 < 7/2$, the biggest order becomes $\log(T)^6 T^{7-\beta_2}$, thus the condition is satisfied if

$$\Delta^{1/4} \log^6 \big(\log^6(T) T^{8-\beta_2} \big) (\log^6(T) T^{8-\beta_2})^{7-\beta_2} \to 0$$

as $T \to \infty$ and $\Delta \to 0$. When $\beta_2 \geq 7/2$, the biggest order is $\log(T)^6 T^{\beta_2}$ so the condition becomes

$$\Delta^{1/4} \log^6 \big(\log^6(T) T^{\beta_2+1} \big) (\log^6(T) T^{\beta_2+1})^{\beta_2} \to 0$$

as $T \to \infty$ and $\Delta \to 0$. Note again that these are just the technical conditions for the proof, to deal with the remainder terms. For Assumption 6, we have, for example,

$$T^{-\varepsilon} \frac{\kappa_{\sigma_{\beta_2}}(T)}{\kappa_\sigma(T)} = T^{-\varepsilon} \frac{\log(T) T^{\beta_2}}{T^{\beta_2}} = \frac{\log(T)}{T^\varepsilon} \to 0$$

as $T \to \infty$ for any $\varepsilon > 0$. For Assumption 7, it suffices to show it holds for a power function, for example, $\kappa(T, \beta) = T^\beta$. To check this, note first that $w = T^{1/2}$ for this CEV model, and for large enough $T > 1$,

$$\sup_{\beta \in \mathcal{N}} \left| \frac{\kappa(T, \beta)}{\kappa(T, \beta_0)} \right| = \sup_{\beta \in \mathcal{N}} \left| T^{\beta - \beta_0} \right| \leq \sup_{\beta \in \mathcal{N}} \left| T^{|\beta - \beta_0|} \right| = \sup_{\beta \in \mathcal{N}} T^{|\beta - \beta_0|} \leq T^{T^{-\varepsilon}}$$

by choosing $v = T^{1/2 - \varepsilon}$ for some $\varepsilon > 0$, and also,

$$\sup_{\beta \in \mathcal{N}} \left| T^{\beta - \beta_0} \right| \geq \sup_{\beta \in \mathcal{N}} \left| T^{-|\beta - \beta_0|} \right| = \sup_{\beta \in \mathcal{N}} T^{-|\beta - \beta_0|} \geq \inf_{\beta \in \mathcal{N}} T^{-|\beta - \beta_0|} \geq T^{-T^{-\epsilon}}$$

for some $\epsilon > 0$. We have both $T^{T^{-\varepsilon}} \to 1$ and $T^{-T^{-\epsilon}} \to 1$ so the assumption is satisfied for this case.

Example 3. *(AS-CEV):* Consider a process

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

living on $\mathcal{D} = (0, \infty)$. With a condition $\alpha_3 < 0$, $\beta_4 > 0$ and $\alpha_4 > \beta_1^2$ together with $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 > 0$, we can show that this process satisfies Assumption 2 with $\sup_{t\in[0,T]} |X_t| = O_p(T)$ and $\sup_{t\in[0,T]} \left|X_t^{-1}\right| = O_p(T)$, since

$$
\begin{aligned}
dY_t = \Big[&\beta_3^2 Y_t^{3-2\beta_4} + 2\beta_1\beta_3 Y_t^{3-\beta_4} + 2\beta_2\beta_3 Y_t^{2-\beta_4} - \alpha_3 - \alpha_2 Y_t \\
&+ \big(\beta_2(\beta_2 - 2\beta_1) - \alpha_1\big) Y_t^2 + (\beta_1^2 - \alpha_4) Y_t^3\Big] dt + \big(\beta_2 Y_t + \beta_1 Y_t^2 + \beta_3 Y_t^{2-\beta_4}\big) dW_t
\end{aligned}
$$

where $Y_t = X_t^{-1}$. For $2\alpha_4 < -3\beta_1^2$ case, it can be also dealt with with the result in Cline, Jeong and Park (2008), and it is also not difficult to show that it satisfies the rest of the assumptions.

2. Assumption Set 2

This set of assumptions is to show the asymptotics of the exact ML estimator. We denote $\ell(x, y, \Delta) = \log p(x, y, \Delta)$, where p is the true transition density of the diffusion model. Parameter arguments are suppressed here.

Assumption 8. *$\ell(x, y, \Delta)$ and its derivatives w.r.t. the parameters, y, and Δ up to the third order satisfy Assumption 2 and 5-7.*

Assumption 9. *The following derivatives of the log-likelihood function ℓ satisfy*

$$
\begin{aligned}
&\ell_\alpha(x, x, 0) = 0 && \ell_{\alpha\alpha'}(x, x, 0) = 0 \\
&\ell_{\alpha\alpha' y}(x, x, 0) = 0 && \ell_{\alpha\beta'}(x, x, 0) = 0 \\
&\ell_{\alpha y}(x, x, 0) = \frac{\mu_\alpha}{\sigma^2}(x) && \ell_\beta(x, x, 0) = -\frac{\sigma_\beta}{\sigma}(x) \\
&\lim_{\Delta\to 0} \Delta \ell_{\beta yy}(x, x, \Delta) = \frac{2\sigma_\beta}{\sigma^3}(x) && \lim_{\Delta\to 0} \sqrt{\Delta}\,\ell_{\alpha\beta' y}(x, x, \Delta) = 0
\end{aligned}
$$

and

$$
\begin{aligned}
&\ell_{\alpha\Delta}(x,x,0)+\frac{1}{2}\ell_{\alpha yy}(x,x,0)\sigma^2(x)=\frac{\mu\mu_\alpha}{\sigma^2}(x)\\
&\ell_{\alpha\alpha'\Delta}(x,x,0)+\frac{1}{2}\ell_{\alpha\alpha' yy}(x,x,0)\sigma^2(x)=-\frac{\mu_\alpha\mu'_\alpha}{\sigma^2}(x)\\
&\lim_{\Delta\to 0}\left[\ell_{\beta\beta'}(x,x,\Delta)+\frac{\Delta}{2}\ell_{\beta\beta' yy}(x,x,\Delta)\sigma^2(x)\right]=-\frac{2\sigma_\beta\sigma'_\beta}{\sigma^2}(x)\\
&\lim_{\Delta\to 0}\left[\sqrt{\Delta}\ell_{\alpha\beta'\Delta}(x,x,\Delta)+\frac{\sqrt{\Delta}}{2}\ell_{\alpha\beta' yy}(x,x,\Delta)\sigma^2(x)\right]=0.
\end{aligned}
$$

Assumption 8 and Assumption 9 are the crucial conditions so that the estimators have the proper limit distributions. The following assumptions are technical conditions to deal with the remainder terms deriving the asymptotic first order terms.

ED1: There exists $K_{T,\Delta}$ such that $\sum_{i=1}^n f(X_{(i-1)\Delta},0)=O_p(K_{T,\Delta})$ and

$$
\sup_{0<\tilde{\Delta}<\Delta}\sum_{i=1}^n\big(f(X_{(i-1)\Delta},\tilde{\Delta})-f(X_{(i-1)\Delta},0)\big)=o_p(K_{T,\Delta})
$$

as $T\to\infty$ and $\Delta\to 0$ satisfying Assumption 5.

ED2: There exists $M_{T,\Delta}$ such that $\sum_{i=1}^n f(X_{(i-1)\Delta},X_{(i-1)\Delta},\Delta)(X_{i\Delta}-X_{(i-1)\Delta})^2=O_p(M_{T,\Delta})$ and

$$
\begin{aligned}
\sum_{i=1}^n\sup_{\tilde{y}_i\in[X_{(i-1)\Delta},X_{i\Delta}]}&\big(f(X_{(i-1)\Delta},\tilde{y}_i,\Delta)-f(X_{(i-1)\Delta},X_{(i-1)\Delta},\Delta)\big)(X_{i\Delta}-X_{(i-1)\Delta})^2\\
&=o_p(M_{T,\Delta})
\end{aligned}
$$

as $T\to\infty$ and $\Delta\to 0$ satisfying Assumption 5.

Assumption 10. *Denoting* $f(x,\Delta)$ *as each of the following functions,*

$$\ell_{\alpha\Delta\Delta}(x,x,\Delta),\ \ell_{\alpha y\Delta}(x,x,\Delta)\sigma(x),\ \ell_{\alpha y\Delta}(x,x,\Delta)\mu(x),\ \ell_{\alpha yy\Delta}(x,x,\Delta)\sigma^2(x),$$
$$\ell_{\beta\Delta}(x,x,\Delta),\ \ell_{\beta y\Delta}(x,x,\Delta)\sigma(x),\ \ell_{\beta y\Delta}(x,x,\Delta)\mu(x),\ \ell_{\beta yy\Delta}(x,x,\Delta)\sigma^2(x),$$
$$\ell_{\alpha\alpha\Delta\Delta}(x,x,\Delta),\ \ell_{\alpha\alpha y\Delta}(x,x,\Delta)\sigma(x),\ \ell_{\alpha\alpha y\Delta}(x,x,\Delta)\mu(x),\ \ell_{\alpha\alpha yy\Delta}(x,x,\Delta)\sigma^2(x),$$
$$\ell_{\beta\beta\Delta}(x,x,\Delta),\ \ell_{\beta\beta y\Delta}(x,x,\Delta)\sigma(x),\ \ell_{\beta\beta y\Delta}(x,x,\Delta)\mu(x),\ \ell_{\beta\beta yy\Delta}(x,x,\Delta)\sigma^2(x),$$
$$\ell_{\alpha\beta\Delta\Delta}(x,x,\Delta),\ \ell_{\alpha\beta y\Delta}(x,x,\Delta)\sigma(x),\ \ell_{\alpha\beta y\Delta}(x,x,\Delta)\mu(x),\ \ell_{\alpha\beta yy\Delta}(x,x,\Delta)\sigma^2(x),$$

it satisfies ED1.

Assumption 11. *Denoting* $f(x,y,\Delta)$ *as each of the following functions,*

$$\ell_{\alpha yy}(x,y,\Delta),\ \ell_{\beta yy}(x,y,\Delta),\ \ell_{\alpha\alpha yy}(x,y,\Delta),\ \ell_{\beta\beta yy}(x,y,\Delta),\ \ell_{\alpha\beta yy}(x,y,\Delta),$$

it satisfies ED2.

Assumption 12. *There exists a sequence* v *such that* $vw^{-1} \to 0$*, and such that*

$$\sup_{\theta\in\mathcal{N}} \left| v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1\prime} \right| \to_p 0$$

where $\mathcal{N} = \{\theta : |v'(\theta-\theta_0)| \le 1\}$*.*

It is only a matter of time to check these conditions and one can easily check them for the models with known transition densities.

Example 4. *(Ornstein-Uhlenbeck):* Consider a process

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

with $\alpha_2 > 0$, $\beta > 0$ and $D = (-\infty, \infty)$. Checking Assumption 10-12 is not difficult but can be tedious since we should apply almost same steps to the various given functions. Here I will only check a couple of functions among the whole conditions as an example. Application to other functions is straightforward. For Assumption 10, consider the following example,

$$\ell_{\alpha_1 y\Delta}(x,\Delta) = \frac{2\alpha_2^2 e^{\alpha_2\Delta}}{\beta^2(e^{\alpha_2\Delta}+1)^2}.$$

Here, $K_{T,\Delta} = T/\Delta$ since

$$\sum_{i=1}^{n} \ell_{\alpha_1 y\Delta}(x,0) = \sum_{i=1}^{n} \frac{\alpha_2^2}{2\beta^2} = \frac{T\alpha_2^2}{2\Delta\beta^2} = O_p(T/\Delta)$$

We have

$$\begin{aligned}\sum_{i=1}^{n}(\ell_{\alpha_1 y\Delta}(x,\tilde{\Delta}) - \ell_{\alpha_1 y\Delta}(x,0)) &= \sum_{i=1}^{n}\left(\frac{2\alpha_2^2 e^{\alpha_2\tilde{\Delta}}}{\beta^2(e^{\alpha_2\tilde{\Delta}}+1)^2} - \frac{\alpha_2^2}{2\beta^2}\right)\\ &= \sum_{i=1}^{n}\frac{\alpha_2^2\beta^2(4e^{\alpha_2\tilde{\Delta}} - (e^{\alpha_2\tilde{\Delta}}+1)^2)}{2\beta^4(e^{\alpha_2\tilde{\Delta}}+1)^2}\\ &= -\sum_{i=1}^{n}\frac{\alpha_2^4\tilde{\Delta}^2}{8\beta^2} + O_p(T\tilde{\Delta}^2) = O_p(T\Delta) = o_p(K_{T,\Delta})\end{aligned}$$

since $\tilde{\Delta} \leq \Delta$, satisfying Assumption 10. For this Ornstein-Uhlenbeck process, Assumption 11 becomes obvious since there is no $\tilde{y}$ in any of the functions in the condition. For Assumption 9, let us take a look at the second derivative with respect to α_1. Note that

$$\ell_{\alpha_1\alpha_1}(x,y,\Delta) = -\frac{2\alpha_2(e^{\alpha_2\Delta}-1)}{\beta^2(e^{\alpha_2\Delta}+1)}.$$

Thus, taking $v = T^{-1/2+\varepsilon}$ for some $\varepsilon > 0$,

$$\begin{aligned}
&\sup_{\theta\in\mathcal{N}}\left|T^{-1+2\varepsilon}\sum_{i=1}^{n}\left(\frac{2\alpha_2(e^{\alpha_2\Delta}-1)}{\beta^2(e^{\alpha_2\Delta}+1)}-\frac{2\alpha_{2,0}(e^{\alpha_{2,0}\Delta}-1)}{\beta_0^2(e^{\alpha_{2,0}\Delta}+1)}\right)\right| \\
&\quad=\sup_{\theta\in\mathcal{N}}\left|T^{2\varepsilon}\left(\frac{\alpha_2^2}{\beta^2}-\frac{\alpha_{2,0}^2}{\beta_0^2}\right)+O_p(T^{2\varepsilon}\Delta)\right| \\
&\quad=\sup_{\theta\in\mathcal{N}}\left|T^{2\varepsilon}\frac{\alpha_{2,0}^2(\beta^2-\beta_0^2)-\beta_0^2(\alpha_2^2-\alpha_{2,0}^2)}{\beta^2\beta_0^2}+O_p(T^{2\varepsilon}\Delta)\right| \\
&\quad=T^{2\varepsilon}O_p(T^{-1/2})+O_p(T^{2\varepsilon}\Delta)
\end{aligned}$$

so we can choose any $0 < \varepsilon < 1/4$ to make it converge to zero. For Assumption 12,

$$\ell_{\alpha_1 y}(x,x,\Delta)=\frac{2\alpha_2 e^{\alpha_2\Delta}}{\beta(e^{\alpha_2\Delta}+1)}$$

for example, so we have

$$\ell_{\alpha_1 y}(x,x,0)=\frac{\alpha_2}{\beta^2}=\frac{\mu_{\alpha_1}(x)}{\sigma(x)}$$

satisfying Assumption 12.

C. First Order Asymptotics

If the conditions AD1 - AD3 hold, we can easily derive the following result from the steps described in Section 2. Hereafter, $A \approx B$ denote that $A - B$ is of smaller order than B.

Theorem 1. *With Assumptions 1 to 7, the asymptotic first order terms of Euler, and Milstein ML estimators are obtained as the following, and with Assumptions 8 to*

12, the asymptotic first order distribution of the exact ML estimator is obtained as

$$\hat{\alpha} - \alpha \approx \left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t$$
$$\hat{\beta} - \beta \approx \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t$$

as $T \to \infty$ *and* $\Delta \to 0$ *under Assumption 5, where* V *is a standard Brownian motion independent of* W*.*

Proof of this theorem is omitted here since it easily follows from the following propositions with the same steps already described at the end of the previous section.

Proposition 1. *For Euler, and Milstein ML estimators, with Assumptions 1 to 7, AD1 and AD2 hold with* $\mathcal{S}(\theta_0)$ *having its leading term as the following, and for the exact ML estimator, the same holds with Assumptions 8 to 9,*

$$\int_0^T \frac{\mu_\alpha}{\sigma}(X_t, \theta_0)dW_t \quad \text{and} \quad \sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t, \theta_0)dV_t$$

for the drift term parameters and the diffusion term parameters, respectively, and also, $\mathcal{H}(\theta_0)$ *having its leading term as*

$$\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t, \theta_0)dt \quad \text{and} \quad \frac{2}{\Delta} \int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t, \theta_0)dt$$

for the drift term parameters and the diffusion term parameters, respectively. Also, the leading term of $\mathcal{H}(\theta_0)$ *becomes a block diagonal matrix in probability as* $T \to \infty$ *and* $\Delta \to 0$ *under Assumption 5.*

Proposition 2. *For the Euler and Milstein ML estimators, with Assumptions 1 to 7, AD3 holds.*

Example 1. *(Ornstein-Uhlenbeck):* For the Ornstein-Uhlenbeck process

$$dX_t = \alpha_2(\alpha_1 - X_t)dt + \beta dW_t,$$

with $\alpha_2 > 0$, note that the drift function is $\mu(x, \alpha_1, \alpha_2) = \alpha_2(\alpha_1 - x)$ and the diffusion function is $\sigma(x, \beta) = \beta$. Applying these functions to the asymptotic distribution in Theorem 1, we have

$$\begin{aligned}
\hat{\alpha}_1 - \alpha_1 &\approx \frac{\beta}{\alpha_2}\frac{W_T}{T} \\
\hat{\alpha}_2 - \alpha_2 &\approx \beta \left(\int_0^T (\alpha_1 - X_t)^2 dt \right)^{-1} \int_0^T (\alpha_1 - X_t) dW_t \\
\hat{\beta} - \beta &\approx \sqrt{\frac{\Delta}{2}} \beta \frac{V_T}{T},
\end{aligned}$$

thus

$$\begin{aligned}
\sqrt{T}(\hat{\alpha}_1 - \alpha_1) &\to_d \mathbf{N}(0, \beta^2/\alpha_2^2) \\
\sqrt{T}(\hat{\alpha}_2 - \alpha_2) &\to_d \mathbf{N}(0, 2\alpha_2)
\end{aligned}$$

for the drift term parameters, and

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

for the diffusion term parameter as $T \to \infty$ and $\Delta \to 0$, since Ornstein-Uhlenbeck process is stationary. Note here that the leading terms of $\hat{\alpha}_1$ and $\hat{\beta}$ is normal even in finite T, while the leading term of $\hat{\alpha}_2$ is non-normal in finite T. Figure 1 shows the difference between the normal distribution and the first order term obtained from Theorem 1. Even for this simplest stationary Ornstein-Uhlenbeck process, we can see that the distributions are quite different.

Example 2. *(Geometric Brownian Motion):* For the geometric Brownian motion

$$dX_t = \alpha X_t dt + \beta X_t dW_t,$$

$T = 5$ ($\alpha_2 = 0.25$, $\alpha_1 = 0$ and $\beta = 0.02$)

(Dotted line is the density function of $\mathbf{N}\big(0, 2\alpha_2/\sqrt{T}\big)$.)

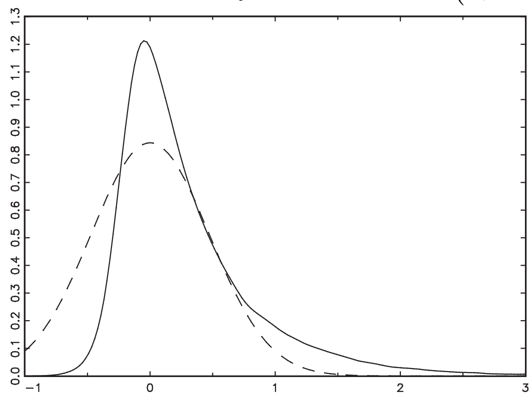

Fig. 1.— First Order Distribution of $\hat{\alpha}_2 - \alpha_2$

we can log transform the process to have

$$d \log X_t = \left(\alpha - \frac{\beta^2}{2}\right) dt + \beta dW_t.$$

For this transformed process, we have $\mu(x, \alpha^*) = \alpha^* x$ for the drift function denoting $\alpha^* = \alpha - \beta^2/2$, and $\sigma(x, \beta) = \beta x$ for the diffusion function. Applying these functions to Theorem 1, we have

$$\sqrt{T}(\hat{\alpha}^* - \alpha^*) \to_d \mathbf{N}(0, \beta^2)$$

for the drift term parameter, and

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

for the diffusion term parameter.

Example 3. *(Feller's Square Root):* For Feller's square root process

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

with $2\alpha_1\alpha_2 \geq \beta^2$, we have

$$\begin{aligned}
\hat{\alpha}_1 - \alpha_1 &\approx \frac{h_{22}s_1 - h_{12}s_2}{h_{11}h_{22} - h_{12}^2} \\
\hat{\alpha}_2 - \alpha_2 &\approx \frac{h_{11}s_2 - h_{12}s_1}{h_{11}h_{22} - h_{12}^2} \\
\hat{\beta} - \beta &\approx \sqrt{\frac{\Delta}{2}}\beta\frac{V_T}{T},
\end{aligned}$$

where

$$\begin{aligned}
h_{11} &= \int_0^T \frac{\alpha_2^2}{\beta_1^2 X_t}dt, \quad h_{22} = \int_0^T \frac{(\alpha_1 - X_t)^2}{\beta_1^2 X_t}dt, \quad h_{12} = \int_0^T \frac{\alpha_2(\alpha_1 - X_t)}{\beta_1^2 X_t}dt \\
s_1 &= \int_0^T \frac{\alpha_2}{\beta_1 X_t^{1/2}}dW_t, \quad s_2 = \int_0^T \frac{\alpha_1 - X_t}{\beta_1 X_t^{1/2}}dW_t.
\end{aligned}$$

Note that we have

$$\begin{aligned}
\frac{1}{T}\int_0^T \frac{(\alpha_1 - X_t)^2}{X_t}dt &\to_p \frac{\alpha_1\beta^2}{2\alpha_2\alpha_1 - \beta^2} \\
\frac{1}{T}\int_0^T \frac{(\alpha_1 - X_t)}{X_t}dt &\to_p \frac{\beta^2}{2\alpha_2\alpha_1 - \beta^2} \\
\frac{1}{T}\int_0^T \frac{1}{X_t}dt &\to_p \frac{2\alpha_2}{2\alpha_2\alpha_1 - \beta^2}
\end{aligned}$$

since X_t is stationary for $2\alpha_2\alpha_1 > \beta^2$, so

$$\begin{aligned}
\sqrt{T}(\hat{\alpha}_1 - \alpha_1) &\to_d \mathbf{N}\left(0,\ \alpha_1\beta^2/\alpha_2^2\right) \\
\sqrt{T}(\hat{\alpha}_2 - \alpha_2) &\to_d \mathbf{N}\left(0,\ 2\alpha_2\right)
\end{aligned}$$

as $T \to \infty$.

Example 4. *(CEV - Constant Elasticity of Variance):* For a positive recurrent CEV process

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

we have

$$\begin{aligned}
\hat{\alpha}_2 - \alpha_2 &\approx \frac{h_{11}s_2 - h_{12}s_1}{h_{11}h_{22} - h_{12}^2} \\
\hat{\alpha}_1 - \alpha_1 &\approx \frac{h_{22}s_1 - h_{12}s_2}{h_{11}h_{22} - h_{12}^2} \\
\hat{\beta}_2 - \beta_2 &\approx \sqrt{\frac{\Delta}{2}} \frac{h_{33}s_4 - h_{34}s_3}{h_{33}h_{44} - h_{34}^2} \\
\hat{\beta}_1 - \beta_1 &\approx \sqrt{\frac{\Delta}{2}} \frac{h_{44}s_3 - h_{34}s_4}{h_{33}h_{44} - h_{34}^2},
\end{aligned}$$

where

$$\begin{aligned}
h_{11} &= \int_0^T \frac{\alpha_2^2}{\beta_1^2 X_t^{2\beta_2}} dt, \quad h_{22} = \int_0^T \frac{(\alpha_1 - X_t)^2}{\beta_1^2 X_t^{2\beta_2}} dt, \quad h_{12} = \int_0^T \frac{\alpha_2(\alpha_1 - X_t)}{\beta_1^2 X_t^{2\beta_2}} dt \\
s_1 &= \int_0^T \frac{\alpha_2}{\beta_1 X_t^{\beta_2}} dW_t, \quad s_2 = \int_0^T \frac{\alpha_1 - X_t}{\beta_1 X_t^{\beta_2}} dW_t
\end{aligned}$$

and

$$\begin{aligned}
h_{33} &= \frac{T}{\beta_1^2}, \quad h_{44} = \int_0^T \log^2(X_t)dt, \quad h_{34} = \frac{1}{\beta_1}\int_0^T \log(X_t)dt \\
s_3 &= \frac{V_T}{\beta_1}, \quad s_4 = \int_0^T \log(X_t)dV_t.
\end{aligned}$$

Corollary 1. *With Assumptions 1 to 7, the asymptotic first order terms of the t-statistics of the Euler, and Milstein ML estimators are obtained as the following, and with Assumptions 8 to 12, the asymptotic first order distribution of the t-statistics of* *the exact ML estimator is obtained as*

$$t(\hat{\alpha}_k) \approx \left[\left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t \right]_k \Bigg/ \left[\left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \right)^{-1} \right]_{kk}^{1/2}$$
$$t(\hat{\beta}_k) \approx \left[\left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t \right]_k \Bigg/ \left[\left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \right)^{-1} \right]_{kk}^{1/2}$$

as $T \to \infty$ and $\Delta \to 0$ under Assumption 5, where V is a standard Brownian motion independent of W. a_k is the k'th element of a vector a, and A_{kk} is the (k,k) element of a matrix A.

Example 5. *(Ornstein-Uhlenbeck):* For a process

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

with $\alpha_2 > 0$, we have

$$t(\hat{\alpha}_2) \approx \left(\int_0^T (\alpha_1 - X_t)^2 \right)^{-1/2} \int_0^T (\alpha_1 - X_t)dW_t$$
$$t(\hat{\alpha}_1) \approx \frac{W_T}{\sqrt{T}} \sim \mathbf{N}(0,1)$$
$$t(\hat{\beta}) \approx \frac{V_T}{\sqrt{T}} \sim \mathbf{N}(0,1).$$

Note that we have also $t(\hat{\alpha}_2) \to_d \mathbf{N}(0,1)$ as $T \to \infty$. Figure 2 shows the standard normal density function, actual histogram of $t(\hat{\alpha}_2)$ obtained from the simulation, and the distribution of the leading term obtained from Corollary 1. We can see that the actual histogram of the t-statistic is closer to the limit distribution than to the standard normal density function.

Example 6. *(CEV - Constant Elasticity of Variance):* For a positive recurrent CEV process

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

$T = 5$ ($\alpha_2 = 0.25$, $\alpha_1 = 0$ and $\beta = 0.02$)

(Dotted line is the standard normal density function.)

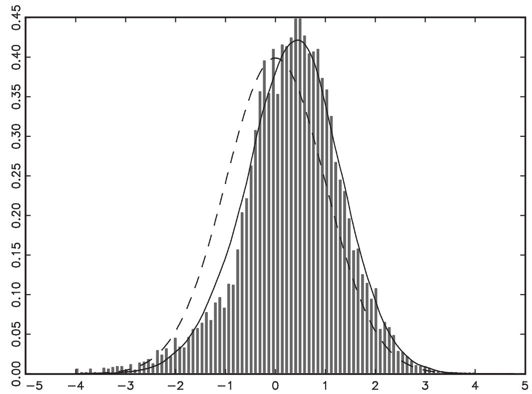

Fig. 2.— First Order Distribution and the Histogram of $t(\hat{\alpha}_2)$ – OU

we have

$$
\begin{aligned}
t(\hat{\alpha}_2) &\approx \frac{h_{11}s_2 - h_{12}s_1}{\left[h_{11}(h_{11}h_{22} - h_{12}^2)\right]^{1/2}} \\
t(\hat{\alpha}_1) &\approx \frac{h_{22}s_1 - h_{12}s_2}{\left[h_{22}(h_{11}h_{22} - h_{12}^2)\right]^{1/2}} \\
t(\hat{\beta}_2) &\approx \frac{h_{33}s_4 - h_{34}s_3}{\left[h_{33}(h_{33}h_{44} - h_{34}^2)\right]^{1/2}} \\
t(\hat{\beta}_1) &\approx \frac{h_{44}s_3 - h_{34}s_4}{\left[h_{44}(h_{33}h_{44} - h_{34}^2)\right]^{1/2}},
\end{aligned}
$$

where each term is defined as same as above. Figure 3 shows the standard normal density function, actual histogram of $t(\hat{\alpha}_2)$ obtained from the simulation, and the distribution of the leading term obtained from Corollary 1. As in the Ornstein-Uhlenbeck case, the distribution of the leading term explains the actual histogram quite well.

$T = 5$ ($\alpha_2 = 0.09$, $\alpha_1 = 0.08$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$)

(Dotted line is the standard normal density function.)

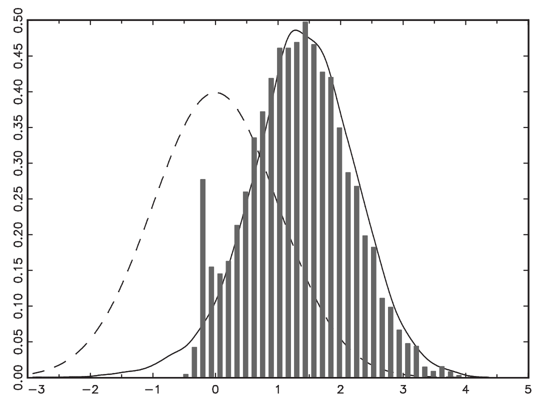

Fig. 3.— First Order Distribution and the Histogram of $t(\hat{\alpha}_2)$ – CEV

1. Consistency and the Convergence Rate of the Estimator

From Theorem 1, we can check that the Milstein ML estimator is consistent as long as

$$\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t)dt \to \infty$$

for the drift term parameters and

$$\frac{1}{\Delta}\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt \to \infty$$

for the diffusion term parameters, and also these determine the convergence rate. To understand more about this in a specific case, let us consider the CEV model first,

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

For the CEV case, note that these conditions are

$$\int_0^T \frac{(\kappa_1 - X_t)^2}{X_t^{2\beta_2}} dt \to \infty, \qquad \int_0^T X_t^{-2\beta_2} dt \to \infty,$$
$$\frac{1}{\Delta}\int_0^T \beta_1^{-2} dt \to \infty, \qquad \frac{1}{\Delta}\int_0^T \log(X_t)^2 dt \to \infty.$$

With suitable parameter restrictions as in the example above, these convergence rates become $\sqrt{T}$, $\sqrt{T}$, $\sqrt{T/\Delta}$ and $\sqrt{T/\Delta}$, and we can easily see that the drift term parameters will not be consistent unless $T \to \infty$, while for diffusion term parameter estimators, they will be still consistent if $\Delta \to 0$. This is an interesting property of the diffusion process estimation. This property of the diffusion estimator is well known among those who study the diffusion process, but here, I present this theoretical result in an explicit expression of the asymptotic distribution. For the Brownian motion with drift

$$dX_t = \alpha dt + \beta dW_t,$$

the above conditions become

$$\int_0^T \beta^{-2} dt \to \infty \quad \text{and} \quad \frac{1}{\Delta}\int_0^T \beta^{-2} dt \to \infty$$

for α and β, respectively, so the convergence rates for each parameters are $\sqrt{T}$ and $\sqrt{T/\Delta}$. In this case also, the convergence rate of the drift term parameter does not depend on the sampling interval Δ, while the convergence rate of the diffusion term parameter depends on both T and Δ.

2. Mixed Normal Property of the Estimator

Since X and W are not independent of each other, the distribution of the drift term estimator

$$\hat{\alpha} - \alpha \approx \left(\int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t) dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t$$

is very non-standard and far from normal distribution in general. On the other hand, for the diffusion term estimator, X and V are independent of each other, so we can show that the leading term of the diffusion term estimator is mixed normal as,

$$\begin{aligned} \hat{\beta} - \beta &\approx \sqrt{\frac{\Delta}{2}} \left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t) dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \\ &\sim \mathbf{N}\left(0, \frac{\Delta}{2} \left(\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t) dt \right)^{-1} \right). \end{aligned}$$

From this, we can expect that the diffusion term parameter estimator will behave in more standard way than the drift term parameter estimator, and moreover, since this is the mean-zero mixed normal distribution, we can expect that it will suffer less from the bias problem.

For a single diffusion term parameter model, the leading term of the t-statistic of the diffusion term parameter estimator is

$$\begin{aligned} t(\hat{\beta}) &\approx \left(\int_0^T \frac{\sigma_\beta^2}{\sigma^2}(X_t) dt \right)^{-1} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \Big/ \left(\int_0^T \frac{\sigma_\beta^2}{\sigma^2}(X_t) dt \right)^{-1/2} \\ &= \left(\int_0^T \frac{\sigma_\beta^2}{\sigma^2}(X_t) dt \right)^{-1/2} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \\ &\sim \mathbf{N}(0,1) \end{aligned}$$

so we can check that it follows the standard normal distribution even if the process is nonstationary.

D. Monte Carlo Study

1. Performance Comparison

In this section, I perform Monte Carlo simulations to assess the performance of the Milstein ML estimator. The simulations are designed for two goals.

Firstly I consider the performance of the estimator in different time span T and different sampling interval Δ. From the asymptotic result illustrated in Section 4, we expect that the estimator will perform better as the time span increases and the sampling interval decreases, but if we only focus on the drift term parameters, decreasing the sampling interval will not help much to estimate them more accurately. Thus, with this theoretical background, we may be able to say that obtaining intra-day high frequency data will only give a marginal help on estimating the drift term. So if we are only interested in the drift term estimation, and if we suspect that the high frequency data is contaminated with the microstructure errors, then we can just use the daily or monthly data for the estimation without worrying about the loss of the information. This property of the diffusion estimator is shown in the following MSE comparison. For this simulation, I generated process with the CEV model

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

To increase the accuracy of the data generation, I generated the process with the Milstein approximation, with finer sampling interval $\tilde{\Delta} = \Delta/1000$, and resampled it to make a data of the sampling interval Δ. The simulation iterations are set to be 1000. As expected from the asymptotic result, while the MSEs decrease drastically as the time span T increases in the first part of Table I, in the second part of Table I, the MSEs for the drift term parameters stay almost still at a fixed level even though the sampling interval is getting smaller and smaller.

Table I.

MSE Comparison for Various Time Span T and Sampling Interval Δ

$\alpha_2 = 1, \ \ \alpha_1 = 1, \ \ \beta_1 = 0.1, \ \ \beta_2 = 1.1$

$\Delta = 0.01$

	α_2	α_1	β_1	β_2
$T = 1$	50.197	7.071×10^{-3}	1.027×10^{-4}	3.313
$T = 2$	11.794	4.549×10^{-3}	4.397×10^{-5}	1.104
$T = 4$	2.627	2.425×10^{-3}	1.877×10^{-5}	0.480

$\alpha_2 = 1, \ \ \alpha_1 = 1, \ \ \beta_1 = 0.1, \ \ \beta_2 = 1.1$

$T = 10$

	α_2	α_1	β_1	β_2
$\Delta = 0.2$	0.412	9.427×10^{-4}	2.268×10^{-4}	1.726
$\Delta = 0.05$	0.348	9.113×10^{-4}	4.181×10^{-5}	0.470
$\Delta = 0.02$	0.393	8.785×10^{-4}	1.399×10^{-5}	0.261

Our next Monte Carlo simulation is for the performance comparison with the estimation method introduced in Aït-Sahalia (2002). This is one of the most widely used among other estimation methods, so I picked this for the comparison. While this is a good estimator, I show that the Milstein ML estimator is as good as this in the estimation performance. Moreover, the ease of application is a lot less complicated than that, and also the computation time is a lot less than that. The computation time for each estimator also depends on the parameter settings, but in the following simulation, the calculation time was almost 10 times longer than the Milstein estimator.

The simulation settings for this is $T = 5, \ 20$, and $\Delta = 0.005, \ 0.025, \ 0.1$,

Table II.

Performance Comparison ($T = 5$)

IQR_{50} ($\alpha_2 = 0.09$, $\alpha_1 = 0.08$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$)

$T = 5$		α_1	α_2	β_1	β_2
Daily	Euler	0.03379	1.1496	0.3648	0.1787
	Milstein	0.03382	1.1495	0.3642	0.1775
	Aït-Sahalia	0.03459	1.1865	0.3522	0.1697
Weekly	Euler	0.03452	1.1677	0.8353	0.4217
	Milstein	0.03442	1.1637	0.8249	0.4118
	Aït-Sahalia	0.03460	1.1891	0.8632	0.4052
Monthly	Euler	0.03667	1.1924	1.6001	0.8290
	Milstein	0.03646	1.1901	1.6098	0.7940
	Aït-Sahalia	0.03689	1.3037	1.9914	0.7649

representing 5 and 20 years of data observed in daily, weekly, and monthly basis. The parameter settings are based on the estimation result in Aït-Sahalia (1999). The comparison criteria is IQR_{50}. IQR_{50} is defined as $\text{IQR}_{50}=|q_{75} - q_{25}|$ where q_i is the i-th quantile of the empirical distribution, and it helps to assess the performance of estimators when the estimators suffers from possible outliers. As shown in Table II and Table III, between Milstein ML and Aït-Sahalia's estimators, neither one dominates the other and it is hard to tell which one performs better. As for the Euler and Milstein ML estimators, we can also check that Milstein ML estimator generally performs better than Euler ML estimator, especially when the sampling interval is relatively large. Table IV is the outlier counts for each estimators. We can see that the method in Aït-Sahalia (2002) suffers from outliers of big magnitude.

Table III.

Performance Comparison ($T = 20$)

IQR_{50} ($\alpha_2 = 0.09$, $\alpha_1 = 0.08$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$)

$T = 20$		α_1	α_2	β_1	β_2
	Euler	0.03018	0.3198	0.1197	0.0567
Daily	Milstein	0.03036	0.3201	0.1182	0.0567
	Aït-Sahalia	0.03033	0.3167	0.1172	0.0554
	Euler	0.03057	0.3164	0.2633	0.1280
Weekly	Milstein	0.03057	0.3170	0.2617	0.1260
	Aït-Sahalia	0.03034	0.3183	0.2602	0.1242
	Euler	0.03174	0.3182	0.5031	0.2644
Monthly	Milstein	0.03167	0.3168	0.4990	0.2601
	Aït-Sahalia	0.03178	0.3239	0.5289	0.2567

2. Hypothesis Testing

From the form of the asymptotic distribution of the parameter estimates, one question easily arises about the hypothesis testing. If the limiting distribution is not normal, and still we use the critical values obtained under the normality, then it is obvious that the size of the test will be very different from the actual size. For example, the t-statistics for α_2 and α_1 of the CEV model have the following limiting distributions,

$$t(\hat{\alpha}_1) \approx \frac{h_{22}s_1 - h_{12}s_2}{\left[h_{22}(h_{11}h_{22} - h_{12}^2)\right]^{1/2}} \tag{2.2}$$

$$t(\hat{\alpha}_2) \approx \frac{h_{11}s_2 - h_{12}s_1}{\left[h_{11}(h_{11}h_{22} - h_{12}^2)\right]^{1/2}}, \tag{2.3}$$

Table IV.

Outlier Comparison

Outliers greater than $10^6 \times \mathrm{IQR}_{50}$ (out of 10000)

		$T=5$				$T=20$			
		α_1	α_2	β_1	β_2	α_1	α_2	β_1	β_2
Daily	Euler	0	0	0	0	0	0	0	0
	Milstein	0	0	0	0	0	0	0	0
	Aït-Sahalia	9	0	70	0	2	0	63	0
Weekly	Euler	0	0	0	0	0	0	0	0
	Milstein	0	1	0	0	0	0	0	0
	Aït-Sahalia	4	0	32	1	2	1	12	0
Monthly	Euler	0	0	0	0	0	0	0	0
	Milstein	1	2	0	0	0	0	1	0
	Aït-Sahalia	6	5	40	0	0	0	4	0

where

$$h_{11} = \int_0^T \frac{\alpha_2^2}{\beta_1^2 X_t^{2\beta_2}} dt, \quad h_{22} = \int_0^T \frac{(\alpha_1 - X_t)^2}{\beta_1^2 X_t^{2\beta_2}} dt, \quad h_{12} = \int_0^T \frac{\alpha_2(\alpha_1 - X_t)}{\beta_1^2 X_t^{2\beta_2}} dt$$
$$s_1 = \int_0^T \frac{\alpha_2}{\beta_1 X_t^{\beta_2}} dW_t, \quad s_2 = \int_0^T \frac{\alpha_1 - X_t}{\beta_1 X_t^{\beta_2}} dW_t,$$

so we can hardly expect that it will follow the standard normal distribution. We can check this from the simulation and Figure 4 shows the simulated distributions for each random variable (2.2) and (2.3).

So unless we know the exact limiting distribution, we can only use the critical values for the normal distribution so this problem can be applied to any cases when

$T = 5$ ($\alpha_2 = 0.09$, $\alpha_1 = 0.08$, $\beta_1 = 0.8$ and $\beta_2 = 1.5$)

(Dotted lines are for the standard normal density function.)

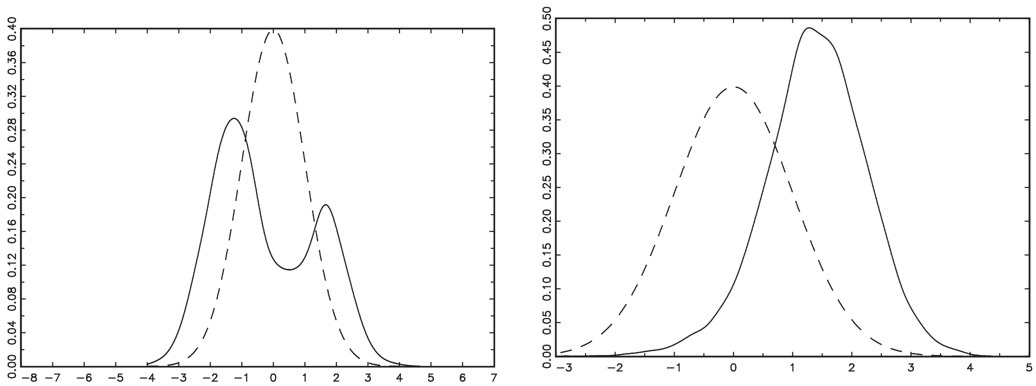

Fig. 4.— First Order Distributions of $t(\hat{\alpha}_1)$ and $t(\hat{\alpha}_2)$

we are estimating diffusion processes. In Table V and Table VI, we present the simulation results showing the discrepancies between the actual and the simulated size of the tests, and also show that this property of the estimator is not only for the Milstein ML estimator, but also same for other diffusion estimators such as Aït-Sahalia (2002)'s closed-form ML estimator. Table VII shows the comparison result between the standard normal, bootstrap and the limit distribution obtained in (2) and (3). For the limit distributions, I used estimated parameter values. As we can see here, both bootstrap and first order limit distribution performed better than standard normal critical values.

E. Application to the Estimation

This limit theorem for the diffusion estimators can be used to enhance the performance of the estimators. Followings are a couple of examples.

Table V.

Size of t-Statistics – Milstein ML estimation

$T = 5,\ \Delta = 0.005$

		α_1	α_2	β_1	β_2
	1%	0.07	0.129	0.000	0.016
One-sided	5%	0.107	0.384	0.010	0.055
	10%	0.129	0.554	0.052	0.101
	1%	0.405	0.109	0.041	0.010
Two-sided	5%	0.498	0.306	0.083	0.061
	10%	0.541	0.452	0.121	0.112

1. Time Change Bias Correction Method

Assume that we have the following process

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

As illustrated in the previous examples, the estimator for α usually produces a big bias even for the simple stationary processes such as the Ornstein-Uhlenbeck process. Choi and Park (2008) shows that, from the idea that $\hat{\alpha}$ has the following leading term,

$$\hat{\alpha} - \alpha \approx \left(\int_0^T \frac{\mu_\alpha^2(X_t, \alpha)}{\sigma^2(X_t, \beta)} dt \right)^{-1} \int_0^T \frac{\mu_\alpha(X_t, \alpha)}{\sigma(X_t, \beta)} dW_t,$$

we can think of a time change to make the denominator a constant c, so that,

$$\hat{\alpha}^{\tau_c} - \alpha \approx \left(\int_0^{\tau_c} \frac{\mu_\alpha^2(X_t, \alpha)}{\sigma^2(X_t, \beta)} dt \right) \int_0^{\tau_c} \frac{\mu_\alpha(X_t, \alpha)}{\sigma(X_t, \beta)} dW_t = \frac{1}{c} \int_0^{\tau_c} \frac{\mu_\alpha(X_t, \alpha)}{\sigma(X_t, \beta)} dW_t.$$

Table VI.

Size of t-Statistics – Aït-Sahalia's method

$T = 5,\ \Delta = 0.005$

		α_1	α_2	β_1	β_2
	1%	0.082	0.084	0.000	0.006
One-sided	5%	0.134	0.314	0.008	0.060
	10%	0.156	0.502	0.056	0.098
	1%	0.304	0.052	0.004	0.002
Two-sided	5%	0.392	0.192	0.018	0.032
	10%	0.440	0.330	0.036	0.078

Since this is a martingale which is mean-zero, we can expect that this estimator will have no bias, and we can construct an estimator utilizing this fact. One can refer to Choi and Park (2008) for more on this.

2. Bias Correction Using the Rate of Convergence

Note that for a positive recurrent process, we have

$$\frac{1}{T}\int_0^T f(X_t)dt \to_{a.s.} \int_{\mathcal{D}} f(x)p(x)dx \tag{2.4}$$

$$\frac{1}{\sqrt{T}}\left(\int_0^T f(X_t)dt - T\int_{\mathcal{D}} f(x)p(x)dx\right) \to_d \mathbf{N}(0,c)$$

Table VII.

Size Adjustment

Size of t-statistics – Milstein ML estimation

$T = 5,\ \Delta = 0.005$

		$t(\hat{\alpha}_1)$			$t(\hat{\alpha}_2)$		
		Std. Nor.	Bootst.	Lim. Dist.	Std. Nor.	Bootst.	Lim. Dist.
One-side	1%	0.065	0.049	0.053	0.136	0.062	0.047
	5%	0.106	0.104	0.087	0.389	0.169	0.158
	10%	0.133	0.133	0.116	0.569	0.267	0.256
Two-side	1%	0.408	0.191	0.361	0.121	0.122	0.098
	5%	0.505	0.286	0.423	0.313	0.186	0.178
	10%	0.551	0.382	0.467	0.461	0.257	0.241

Critical values based on:

Std. Nor. – standard normal distribution

Bootst. – parametric bootstrap method

Lim. Dist. – limit distribution simulated with the estimated parameter values

for some constant c, where $p(x) = m(x)/M(\mathcal{D})$, with proper conditions. (See Khasminskii (2001).) From this, we can check the order of the bias of the estimator,

$$\begin{aligned}
\mathbb{E}(\hat{\alpha} - \alpha) &\approx \mathbb{E}\left(\int_0^T \frac{\mu_\alpha^2}{\sigma^2}(X_t)dt \right)^{-1} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t \\
&= \mathbb{E}\,\frac{C}{T} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t + \mathbb{E}\,\frac{\mathbf{N}(0,c)}{T^{3/2}C^2} \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t + o_p(T^{-1}) \\
&= 0 + O_p(T^{-1}),
\end{aligned}$$

where $C = \left(\int_{\mathcal{D}} \frac{\mu_\alpha^2}{\sigma^2}(x)p(x)dx\right)^{-1}$. Now using this information, we can think of a method to correct the bias by setting up the following simple regression relationship,

$$\mathbb{E}\,\hat{\alpha}_i - \alpha = \frac{c}{T_i} + \varepsilon_i$$

for each different T_i. We can estimate $\hat{c}$ by subsampling with different time span T_i, and the bias corrected estimator $\tilde{\alpha}$ becomes

$$\tilde{\alpha} = \hat{\alpha} - \frac{\hat{c}}{T}.$$

Table VIII is the simulation table with this correction method. If we have null-

Table VIII.

Performance Comparison (α_2)

CEV ($\alpha_1 = 0.08$, $\alpha_2 = 0.09$, $\beta_1 = 0.8$, $\beta_2 = 1.5$)

$T = 5$, $\Delta = 0.005$

	Median bias	IQR_{50}
Original	0.949	1.133
Bias corrected	0.209	0.981

recurrent diffusion processes (with suitable conditions), the convergence rate of the bias will become $T^{-1/2}$, not T^{-1}, since the integral in (2.4) will converge to a random variable, not to a constant, so in this case, we can also apply this fact to the above correction method.

CHAPTER III

ASYMPTOTIC EXPANSIONS FOR THE MAXIMUM LIKELIHOOD ESTIMATORS OF DIFFUSION MODELS

In this chapter, I deal with the second and the higher order asymptotics of the maximum likelihood estimators of diffusion models

A. Background

Consider a time-homogeneous stochastic differential equation

$$dX_t = \mu(X_t, \alpha)dt + \sigma(X_t, \beta)dW_t \tag{3.1}$$

where μ and σ are the drift and diffusion functions, respectively. We will denote $\theta = (\alpha', \beta')'$ hereafter. We let $\mathcal{D} = (\underline{x}, \bar{x})$ denotes the domain of the diffusion process X_t. Euler approximation of this SDE is

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

and the closed-form solution of this approximated transition density is given by

$$p_{_E}(x, y) = \frac{1}{\sqrt{2\pi\Delta}\,\sigma(x)} \exp\left[-\frac{\big(y - x - \Delta\mu(x)\big)^2}{2\Delta\sigma^2(x)} \right],$$

denoting $x = X_{(i-1)\Delta}$ and $y = X_{i\Delta}$, and suppressing the parameter arguments for each function. Milstein approximation of this SDE is

$$\begin{aligned} X_{i\Delta} - X_{(i-1)\Delta} \simeq\ & \mu(X_{(i-1)\Delta})\Delta + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\ & + \frac{1}{2}\sigma\sigma^{\cdot}(X_{(i-1)\Delta})\big[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta\big], \end{aligned}$$

where $a^{\cdot}(x, \theta)$ denotes a derivative $\partial/\partial x \, a(x, \theta)$ (I define $a_{\cdot}(x, \theta)$ as a derivative $\partial/\partial\theta \, a(x, \theta)$). In the case of the Euler approximation, the approximated transition density is a normal distribution, but in the case of the Milstein approximation, the approximation error is reduced more with a mixture of a normal and a chi-squared distribution, and the approximated transition density is given by,

$$p_{_M}(x, y) = \frac{1}{\sqrt{2\pi\Delta}\, \tau(x, y)} \left(\exp\left[\frac{-\big(\tau(x, y) + \sigma(x)\big)^2}{2\Delta\sigma^2\sigma^{\cdot 2}(x)} \right] + \exp\left[\frac{-\big(\tau(x, y) - \sigma(x)\big)^2}{2\Delta\sigma^2\sigma^{\cdot 2}(x)} \right] \right),$$

where

$$\tau(x, y) = \big(\sigma^2(x) + \Delta\, \sigma^2\sigma^{\cdot 2}(x) + 2\, \sigma\sigma^{\cdot}(x)\big(y - x - \Delta\, \mu(x)\big)\big)^{1/2}$$

denoting $x = X_{(i-1)\Delta}$ and $y = X_{i\Delta}$, and suppressing the parameter arguments for each function.

The Euler and Milstein ML estimator $\hat{\theta}$ is defined as an estimator which minimizes the log-likelihood function

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

over $\theta \in \Theta$, i.e.,

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \mathcal{L}(\theta).$$

Here $\hat{p}$ represents either $p_{_E}$ or $p_{_M}$. We assume that Θ is compact and convex, and θ_0 is an interior point of Θ. The Milstein ML estimation method was first proposed in Elerian (1998). Replacing $\hat{p}$ with the true transition density p, we can perform the exact ML estimation, but it is only restricted to the cases when we know the true transition density in a closed-form, such as O-U, Feller, and BM with drift.

B. Assumptions

Here I adopt Assumptions 2-4 and 6-7 from Part 1. For the following assumptions, they are basically same as Assumption 1 and 5 in Part 1, but only requires higher order conditions.

Assumption 13. *$\mu(x,\alpha)$ has its derivatives up to 7th order, and $\sigma(x,\beta)$ has its derivatives up to the 8th order, w.r.t. x on $\mathcal{D}$. $\mu(x,\alpha)$ and $\sigma(x,\beta)$ have their derivatives up to the 7th order, w.r.t. θ on the interior of Θ. (We assume only piecewise differentiability.) These functions satisfy the conditions in Assumption 2, 6 and 7.*

Assumption 14. *The asymptotic order functions satisfy,*

$$\Delta T^3 \to 0$$

$$\Delta \kappa_1^8 (T\kappa_2(\nu(T))) \to 0$$

as $T \to \infty$ and $\Delta \to 0$, where κ_1 and κ_2 represent any combinations of the order functions in Assumption 13.

C. Asymptotic Higher Order Expansions

Let us denote $\mathcal{S} = \partial\mathcal{L}/\partial\theta$, $\mathcal{H} = \partial^2\mathcal{L}/\partial\theta\theta'$ and $\mathcal{J} = \partial^3\mathcal{L}/\partial\theta\otimes\theta\theta'$. Then by the Taylor expansion of the score function around θ_0, we have

$$\mathcal{S}(\hat{\theta}) = \mathcal{S}(\theta_0) + \mathcal{H}(\theta_0)(\hat{\theta}-\theta_0) + \frac{1}{2}\big(I_k\otimes(\hat{\theta}-\theta_0)'\big)\mathcal{J}(\tilde{\theta})(\hat{\theta}-\theta_0), \tag{3.2}$$

where $\tilde{\theta}$ is a value in the line segment connecting θ_0 and $\hat{\theta}$. Here, $\mathcal{J}$ is the derivative of $\mathcal{H}$ represented by a $k^2 \times k$ matrix (where k is the number of parameters), i.e.,

$$\mathcal{J}(\theta) = \begin{pmatrix} \mathcal{J}_1(\theta) \\ \vdots \\ \mathcal{J}_k(\theta) \end{pmatrix}, \quad \text{where} \quad \mathcal{J}_j(\theta) = \frac{\partial \mathcal{H}(\theta)}{\partial \theta_j}.$$

Rewriting the second term of the above expansion as the following,

$$\begin{aligned} \begin{pmatrix} (\hat{\theta}-\theta_0)'\mathcal{J}_1(\tilde{\theta})(\hat{\theta}-\theta_0) \\ \vdots \\ (\hat{\theta}-\theta_0)'\mathcal{J}_k(\tilde{\theta})(\hat{\theta}-\theta_0) \end{pmatrix} &= \begin{pmatrix} (\hat{\theta}-\theta_0)'\mathcal{J}_1(\theta_0)(\hat{\theta}-\theta_0) \\ \vdots \\ (\hat{\theta}-\theta_0)'\mathcal{J}_k(\theta_0)(\hat{\theta}-\theta_0) \end{pmatrix} \\ &\quad + \begin{pmatrix} (\hat{\theta}-\theta_0)'\big(\mathcal{J}_1(\tilde{\theta})-\mathcal{J}_1(\theta_0)\big)(\hat{\theta}-\theta_0) \\ \vdots \\ (\hat{\theta}-\theta_0)'\big(\mathcal{J}_k(\tilde{\theta})-\mathcal{J}_k(\theta_0)\big)(\hat{\theta}-\theta_0) \end{pmatrix} \\ &= A_T + B_T. \end{aligned}$$

If B_T is of smaller order than A_T, we can get the following approximation

$$\mathcal{S}(\hat{\theta}) \simeq \mathcal{S}(\theta_0) + \mathcal{H}(\theta_0)(\hat{\theta}-\theta_0) + \frac{1}{2}\big(I_k \otimes (\hat{\theta}-\theta_0)'\big)\mathcal{J}(\theta_0)(\hat{\theta}-\theta_0)$$

replacing $\mathcal{J}(\tilde{\theta})$ with $\mathcal{J}(\theta_0)$. This can be shown from the following conditions,

SD1: $\rho_i^{-1}\mathcal{J}_i(\theta_0)\rho_i^{-1\prime} = O_p(1)$ for each $i = 1, \ldots, k$

SD2: There is a sequence ϱ_i such that $\varrho_i \rho_i^{-1} \to 0$, and such that

$$\sup_{\theta \in \mathcal{N}} \left| \varrho_i^{-1}(\mathcal{J}_i(\theta) - \mathcal{J}_i(\theta_0))\varrho_i^{-1\prime} \right| \to_p 0$$

for each $i = 1, \ldots, k$, where $\mathcal{N} = \{\theta : |\varrho_i'(\theta - \theta_0)| \leq 1\}$.

From Wooldridge (1994), SD1 and SD2 together with AD1 and AD2 in Part 1 implies

SD3: $\rho_i^{-1}\big(\mathcal{J}_i(\tilde{\theta}) - \mathcal{J}_i(\theta_0)\big)\rho_i^{-1\prime} \to_p 0$.

Thus, with SD1 and SD2, the above approximation becomes valid.

Now going back to the Taylor approximation above, with the first order condition $\mathcal{S}(\hat{\theta}) = 0$ for the maximum likelihood estimation, we have

$$\begin{aligned}\hat{\theta} - \theta_0 &\simeq -\mathcal{H}(\theta_0)^{-1}\mathcal{S}(\theta_0) - \frac{1}{2}\mathcal{H}(\theta_0)^{-1}\big(I_k \otimes (\hat{\theta} - \theta_0)'\big)\mathcal{J}(\theta_0)(\hat{\theta} - \theta_0) \\ &= C_T + D_T.\end{aligned} \tag{3.3}$$

To get the second order expansion of the estimator, it is enough to get the first order term from D_T, while we need to obtain both the first and the second order term from C_T.

Proposition 3. *For Euler, and Milstein ML estimators, the first and the second order terms of $\mathcal{S}(\theta_0)$ and $\mathcal{H}(\theta_0)$, and the leading terms of $\mathcal{J}(\theta_0)$ are as shown in the Appendix 1 and Appendix 2, respectively.*

Note that this proposition also accounts to SD1.

Proposition 4. *For Euler, and Milstein ML estimators defined above, SD2 holds.*

The proof of Proposition 2 is omitted here since the same steps can be applied as in the proof of Proposition 1 in Part 1, replacing H with $\mathcal{J}$.
Now combining the above results together, we have the following result,

Theorem 2. *The asymptotic expansions of Euler, and Milstein ML estimators are obtained as*

$$\begin{aligned}\hat{\alpha} - \alpha_0 &\approx -H_{\alpha\alpha,1}^{-1}S_{\alpha,1} - \frac{1}{2}H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} \\ &\qquad - \sqrt{\Delta}H_{\alpha\alpha,1}^{-1}\Big(H_{\alpha\alpha,2}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} + S_{\alpha,2} - H_{\alpha\beta,1}H_{\beta\beta,1}^{-1}S_{\beta,1}\Big) \\ \hat{\beta} - \beta_0 &\approx -\sqrt{\Delta}H_{\beta\beta,1}^{-1}S_{\beta,1} - \Delta^{3/4}H_{\beta\beta,1}^{-1}S_{\beta,2}\end{aligned}$$

where each term is defined in Appendix 1 and Appendix 2, respectively.

We can also only consider the case when Δ is small enough to make the Δ-order terms negligible. By the Taylor expansion of the score function around θ_0, we have

$$\begin{aligned}\mathcal{S}(\hat{\theta}) = \mathcal{S}(\theta_0) + \mathcal{H}(\theta_0)(\hat{\theta}-\theta_0) + \frac{1}{2}\big(I_k \otimes (\hat{\theta}-\theta_0)'\big)\mathcal{J}(\theta_0)(\hat{\theta}-\theta_0) \\ + \frac{1}{6}\big(I_k \otimes (\hat{\theta}-\theta_0)'\big)\mathcal{K}(\tilde{\theta})\big((\hat{\theta}-\theta_0)\otimes(\hat{\theta}-\theta_0)\big),\end{aligned}$$

where $\tilde{\theta}$ is a value in the line segment connecting θ_0 and $\hat{\theta}$. Here, $\mathcal{J}$ is as defined in the above, and $\mathcal{K}$ is the derivative of $\mathcal{J}$ represented by a $k^2 \times k^2$ matrix (where k is the number of parameters), i.e., $\mathcal{K} = \partial^4\mathcal{L}/\partial\theta\theta'\otimes\theta\theta'$. We can represent this as

$$\mathcal{K}(\theta) = \begin{pmatrix} \mathcal{K}_{11}(\theta) & \cdots & \mathcal{K}_{1k}(\theta) \\ \vdots & \ddots & \vdots \\ \mathcal{K}_{k1}(\theta) & \cdots & \mathcal{K}_{kk}(\theta) \end{pmatrix}, \quad \text{where} \quad \mathcal{K}_{ij}(\theta) = \frac{\partial^2\mathcal{H}(\theta)}{\partial\theta_i\theta_j},$$

With the same type of the conditions for $\mathcal{K}_{ij}$,

SD1′: $\rho_{ij}^{-1}\mathcal{K}_{ij}(\theta_0)\rho_{ij}^{-1\prime} = O_p(1)$ for each $i,j = 1,\ldots,k$

SD2′: There is a sequence ϱ_{ij} such that $\varrho_{ij}\rho_{ij}^{-1} \to 0$, and such that

$$\sup_{\theta\in\mathcal{N}} \big|\varrho_{ij}^{-1}(\mathcal{K}_{ij}(\theta) - \mathcal{K}_{ij}(\theta_0))\varrho_{ij}^{-1\prime}\big| \to_p 0$$

for each $i,j = 1,\ldots,k$, where $\mathcal{N} = \{\theta : |\varrho_{ij}'(\theta-\theta_0)| \leq 1\}$.

we have

SD3′: $\rho_{ij}^{-1}\big(\mathcal{K}_{ij}(\tilde{\theta}) - \mathcal{K}_{ij}(\theta_0)\big)\rho_{ij}^{-1\prime} \to_p 0$,

which makes the following approximation valid,

$$\begin{aligned}\hat{\theta}-\theta_0 \simeq &- \mathcal{H}(\theta_0)^{-1}\mathcal{S}(\theta_0) - \frac{1}{2}\mathcal{H}(\theta_0)^{-1}\big(I_k\otimes(\hat{\theta}-\theta_0)'\big)\mathcal{J}(\theta_0)(\hat{\theta}-\theta_0) \\ &- \frac{1}{6}\mathcal{H}(\theta_0)^{-1}\big(I_k\otimes(\hat{\theta}-\theta_0)'\big)\mathcal{K}(\theta_0)\big((\hat{\theta}-\theta_0)\otimes(\hat{\theta}-\theta_0)\big) \\ =&\ A_T + B_T + C_T.\end{aligned} \tag{3.4}$$

Now the rest of the steps are to get the third order asymptotic expansion of A_T, the second order asymptotic expansion of B_T, and the first order asymptotic expansion of C_T. Note that the higher order terms containing Δ order becomes negligible in this setup, so we only need to consider the terms without Δ. For this, we need the following assumptions instead of Assumption 13-14.

Assumption 15. *$\mu(x,\alpha)$ has its derivatives up to 8th order, and $\sigma(x,\beta)$ has its derivatives up to the 9th order, w.r.t. x on $\mathcal{D}$. $\mu(x,\alpha)$ and $\sigma(x,\beta)$ have their derivatives up to the 8th order, w.r.t. θ on the interior of Θ. (We assume only piecewise differentiability.) These functions satisfy the conditions in Assumption 2.*

Under these additional assumptions, we have the following result,

Proposition 5. *For Euler, and Milstein ML estimators, the first and the second order terms of $\mathcal{J}(\theta_0)$, and the leading terms of $\mathcal{K}(\theta_0)$ are as shown in the Appendix 1 and Appendix 2, respectively.*

Theorem 3. *The asymptotic expansions of Euler, and Milstein ML estimators are obtained as*

$$
\begin{aligned}
\hat{\alpha} - \alpha_0 \approx& -H_{\alpha\alpha,1}^{-1} S_{\alpha,1} - \frac{1}{2} H_{\alpha\alpha,1}^{-1} \left(I_k \otimes S'_{\alpha,1} H_{\alpha\alpha,1}^{-1}\right) J_{\alpha\alpha\alpha,1} H_{\alpha\alpha,1}^{-1} S_{\alpha,1} \\
&- \frac{1}{6} H_{\alpha\alpha,1}^{-1} \left(I_k \otimes S'_{\alpha,1} H_{\alpha\alpha,1}^{-1}\right) K_{\alpha\alpha\alpha\alpha,1} \left(H_{\alpha\alpha,1}^{-1} S_{\alpha,1} \otimes H_{\alpha\alpha,1}^{-1} S_{\alpha,1}\right) \\
\hat{\beta} - \beta_0 \approx& -\sqrt{\Delta} H_{\beta\beta,1}^{-1} S_{\beta,1}
\end{aligned}
$$

where each term is defined in Appendix 1 and Appendix 2, respectively.

Followings are examples.

Example 1. *(Ornstein-Uhlenbeck):* For the Ornstein-Uhlenbeck process

$$
dX_t = \alpha_2(\alpha_1 - X_t)dt + \beta dW_t,
$$

note that the drift function is $\mu(x, \alpha_1, \alpha_2) = \alpha_2(\alpha_1 - x)$ and the diffusion function is $\sigma(x, \beta) = \beta$. Applying these functions to the asymptotic distribution in Theorem 2, we have

$$
\begin{aligned}
\hat{\alpha}_1 - \alpha_1 \approx & -\frac{s_1}{h_{11}} - \left[\frac{(h_{12,1} + h_{12,2})s_2}{h_{11}h_{22}}\right] \\
& - \left[\frac{\alpha_2 h_{22} h_{12,1}^2 s_1 - h_{12,1}(s_2^2 h_{11} + h_{22} s_1^2) - h_{12,2}(3 s_2^2 h_{11} + 2 h_{22} s_1^2)}{\alpha_2 h_{22}^2 h_{11}^2}\right] \\
\hat{\alpha}_2 - \alpha_2 \approx & -\frac{s_2}{h_{22}} + \left[\frac{h_{12,1} s_1}{h_{11} h_{22}}\right] - \left[\frac{s_2(h_{12,1}^2 - 4h_{12,2}^2)}{h_{11} h_{22}^2}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
& s_1 = \frac{\alpha_2}{\beta} W_T,\ s_2 = \frac{1}{\beta}\int_0^T (\alpha_1 - X_t) dW_t,\ h_{11} = -\frac{\alpha_2^2}{\beta^2} T, \\
& h_{22} = -\frac{1}{\beta^2}\int_0^T (\alpha_1 - X_t)^2 dt,\ h_{12,1} = -\frac{\alpha_2}{\beta^2}\int_0^T (\alpha_1 - X_t) dt,\ h_{12,2} = \frac{W_t}{\beta}.
\end{aligned}
$$

Note that the order of these terms are $T^{-1/2}$, T^{-1} and $T^{-3/2}$, respectively. If we consider the case when the decreasing rate of Δ is fairly slow to make all the higher T-order terms negligible, we have

$$
\hat{\alpha}_2 - \alpha_2 \approx -\frac{s_2}{h_{22}} - \sqrt{\Delta}\frac{s_{2,d}}{h_{22}},
$$

where $s_{2,d} = -V_T/\sqrt{2}$.

Example 2. *(Feller's Square Root):* For the process

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

we have

$$\hat{\alpha}_1 - \alpha_1 \approx -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} - \left[\frac{(h_{12,1}s_2 - h_{22}s_1)(h_{12,1}s_1 - h_{11}s_2)}{\alpha_2(h_{11}h_{22} - h_{12,1}^2)^2}\right]$$
$$+ \left[\frac{\alpha_2 h_{11}^2 h_{22}^2 h_{12,2}(3h_{11}s_2^2 + 2h_{22}s_1^2) + h_{11}h_{12,1}^3 s_2(5h_{22}s_1^2 - h_{11}s_2^2)}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4}\right.$$
$$+ \frac{\alpha_2 h_{11}h_{22}h_{12,1}^2 h_{12,2}(7h_{11}s_2^2 + 9h_{22}s_1^2) - h_{11}^2 h_{22}h_{12,1}s_2(h_{11}s_2^2 + 16h_{22}s_1^2)}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4}$$
$$\left. - \frac{\alpha_2 h_{12,2}^4(4h_{11}h_{12,2}s_2^2 + 9h_{22}h_{12,2}s_1^2) - 5\alpha_2^2 h_{12,1}^5 h_{12,2}^2 s_2}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4}\right]$$
$$\hat{\alpha}_2 - \alpha_2 \approx -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2}$$
$$+ \left[\frac{\alpha_2^2(4h_{11}^3 h_{22}^2 h_{12,2}^2 s_2 - 6h_{11}h_{12,1}^4 h_{12,2}^2 s_2 + 2h_{12,1}^5 h_{12,2}^2 s_1 + 2h_{11}h_{22}h_{12,1}^3 h_{12,2}^2 s_1)}{\alpha_2^2(h_{11}h_{22} - h_{12,1}^2)^4}\right.$$
$$\left. - \frac{6\alpha_2 h_{11}^2 h_{22}h_{12,1}h_{12,2}(h_{11}s_2^2 + h_{22}s_1^2) + 2h_{11}^2 h_{12,1}^2 s_2(h_{11}s^2 + 4h_{22}s_1^2)}{\alpha_2^2(h_{11}h_{22} - h_{12,1}^2)^4}\right],$$

where

$$s_1 = \frac{\alpha_2}{\beta_1}\int_0^T \frac{1}{\sqrt{X_t}}dW_t,\ s_2 = \frac{1}{\beta_1}\int_0^T \frac{\alpha_1 - X_t}{\sqrt{X_t}}dW_t,\ h_{11} = -\frac{\alpha_2^2}{\beta_1^2}\int_0^T \frac{1}{X_t}dt,$$
$$h_{22} = -\frac{1}{\beta_1^2}\int_0^T \frac{(\alpha_1 - X_t)^2}{X_t}dt,\ h_{12,1} = -\frac{\alpha_2}{\beta_1^2}\int_0^T \frac{\alpha_1 - X_t}{X_t}dt,\ h_{12,2} = \frac{1}{\beta_1}\int_0^T \frac{1}{\sqrt{X_t}}dW_t.$$

Note that the order of these terms are $T^{-1/2}$, T^{-1} and $T^{-3/2}$ for α_1, while those are $T^{-1/2}$ and $T^{-3/2}$ for α_2 since the T^{-1} order term vanishes. If we consider the case when the decreasing rate of Δ is fairly slow to make all the higher T-order terms negligible, we have

$$
\begin{aligned}
\hat{\alpha}_1 - \alpha_1 \approx & -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} \\
& + \sqrt{\Delta}\Bigg[\frac{h_{12,1}\big[(h_{33}h_{44} - h_{34}^2)s_{2,d} + (h_{24}h_{34} - h_{23}h_{44})s_3 - (h_{24}h_{33} - h_{23}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)} \\
& \qquad - \frac{h_{22}\big[(h_{33}h_{44} - h_{34}^2)s_{1,d} + (h_{14}h_{34} - h_{13}h_{44})s_3 - (h_{14}h_{33} - h_{13}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)}\Bigg] \\
\hat{\alpha}_2 - \alpha_2 \approx & -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2} \\
& + \sqrt{\Delta}\Bigg[\frac{h_{12,1}\big[(h_{33}h_{44} - h_{34}^2)s_{1,d} + (h_{14}h_{34} - h_{13}h_{44})s_3 - (h_{14}h_{33} - h_{13}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)} \\
& \qquad - \frac{h_{11}\big[(h_{33}h_{44} - h_{34}^2)s_{2,d} + (h_{24}h_{34} - h_{23}h_{44})s_3 - (h_{24}h_{33} - h_{23}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)}\Bigg],
\end{aligned}
$$

for the Milstein ML estimation case, and

$$
\begin{aligned}
\hat{\beta}_1 - \beta_1 &\approx -\sqrt{\Delta}\frac{h_{44}s_3 - h_{34,1}s_4}{h_{33}h_{44} - h_{34,1}^2} + \Delta^{-3/2}\frac{h_{34,1}s_{4,d}}{h_{33}h_{44} - h_{34,1}^2} \\
\hat{\beta}_2 - \beta_2 &\approx -\sqrt{\Delta}\frac{h_{33}s_4 - h_{34,1}s_3}{h_{33}h_{44} - h_{34,1}^2} - \Delta^{-3/2}\frac{h_{33}s_{4,d}}{h_{33}h_{44} - h_{34,1}^2},
\end{aligned}
$$

where

$$
\begin{aligned}
h_{33} &= -\frac{2T}{\beta_1^2},\ h_{44} = -2\int_0^T \log(X_t)^2 dt,\ h_{34} = -\frac{2}{\beta_1}\int_0^T \log(X_t)dt, \\
s_{1,d} &= \frac{\sqrt{2}\alpha_2}{2}\int_0^T \frac{1}{X_t}dV_t,\ s_{2,d} = \frac{\sqrt{2}\alpha_1}{2}\int_0^T \frac{1}{X_t}dV_t - \sqrt{2}V_T, \\
s_3 &= \frac{\sqrt{2}}{\beta_1}\int_0^T X_t^{\alpha_2 - 1/2}dV_t,\ s_4 = \frac{\sqrt{2}\alpha_1}{\beta_1}\int_0^T \log(X_t)X_t^{\alpha_2 - 1/2}dV_t, \\
h_{14} &= \frac{3\alpha_2}{2}\int_0^T \frac{\log(X_t)}{X_t}dt,\ h_{13} = \frac{3\alpha_2}{2\beta_1}\int_0^T \frac{1}{X_t}dt,\ h_{23} = \frac{3}{2\beta_1}\int_0^T \frac{\alpha_1 - X_t}{X_t}dt, \\
h_{24} &= \frac{3}{2}\int_0^T \frac{(\alpha_1 - X_t)\log(X_t)}{X_t}dt,\ s_{4,d} = \frac{2\beta_1}{3^{1/4}}\int_0^T X_t^{-1/2}\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}dU_t.
\end{aligned}
$$

For the Euler ML estimation case, we have

$$\hat{\alpha}_1 - \alpha_1 \approx -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} - \sqrt{\Delta}\left[\frac{h_{22}s_{1,d} - h_{12,1}s_{2,d}}{h_{11}h_{22} - h_{12,1}^2}\right]$$
$$\hat{\alpha}_2 - \alpha_2 \approx -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2} - \sqrt{\Delta}\left[\frac{h_{11}s_{2,d} - h_{12,1}s_{1,d}}{h_{11}h_{22} - h_{12,1}^2}\right]$$

with the followings replaced as

$$s_{1,d} = -\frac{\alpha_2}{2\sqrt{2}}\int_0^T \frac{1}{X_t}dV_t,\ \ s_{2,d} = -\frac{\alpha_1}{2\sqrt{2}}\int_0^T \frac{1}{X_t}dV_t - \frac{1}{2\sqrt{2}}V_T.$$

Example 3. *(CEV - Constant Elasticity of Variance):* For a positive recurrent CEV process

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

we have

$$\begin{aligned}
\hat{\alpha}_1 - \alpha_1 \approx& -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} - \left[\frac{(h_{12,1}s_2 - h_{22}s_1)(h_{12,1}s_1 - h_{11}s_2)}{\alpha_2(h_{11}h_{22} - h_{12,1}^2)^2}\right] \\
&+ \left[\frac{\alpha_2 h_{11}^2 h_{22}^2 h_{12,2}(3h_{11}s_2^2 + 2h_{22}s_1^2) + h_{11}h_{12,1}^3 s_2(5h_{22}s_1^2 - h_{11}s_2^2)}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4}\right. \\
&\quad + \frac{\alpha_2 h_{11}h_{22}h_{12,1}^2 h_{12,2}(7h_{11}s_2^2 + 9h_{22}s_1^2) - h_{11}^2 h_{22}h_{12,1}s_2(h_{11}s_2^2 + 16h_{22}s_1^2)}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4} \\
&\quad \left. - \frac{\alpha_2 h_{12,2}^4(4h_{11}h_{12,2}s_2^2 + 9h_{22}h_{12,2}s_1^2) - 5\alpha_2^2 h_{12,1}^5 h_{12,2}^2 s_2}{\alpha_2^2(h_{11}h_{22} - h_{12}^2)^4}\right] \\
\hat{\alpha}_2 - \alpha_2 \approx& -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2} \\
&+ \left[\frac{\alpha_2^2(4h_{11}^3 h_{22}^2 h_{12,2}^2 s_2 - 6h_{11}h_{12,1}^4 h_{12,2}^2 s_2 + 2h_{12,1}^5 h_{12,2}^2 s_1 + 2h_{11}h_{22}h_{12,1}^3 h_{12,2}^2 s_1)}{\alpha_2^2(h_{11}h_{22} - h_{12,1}^2)^4}\right. \\
&\quad \left. - \frac{6\alpha_2 h_{11}^2 h_{22}h_{12,1}h_{12,2}(h_{11}s_2^2 + h_{22}s_1^2) + 2h_{11}^2 h_{12,1}^2 s_2(h_{11}s^2 + 4h_{22}s_1^2)}{\alpha_2^2(h_{11}h_{22} - h_{12,1}^2)^4}\right],
\end{aligned}$$

where

$$s_1 = \frac{\alpha_2}{\beta_1}\int_0^T \frac{1}{X_t^{\beta_2}}dW_t,\ s_2 = \frac{1}{\beta_1}\int_0^T \frac{\alpha_1 - X_t}{X_t^{\beta_2}}dW_t,\ h_{11} = -\frac{\alpha_2^2}{\beta_1^2}\int_0^T \frac{1}{X_t^{2\beta_2}}dt,$$
$$h_{22} = -\frac{1}{\beta_1^2}\int_0^T \frac{(\alpha_1 - X_t)^2}{X_t^{2\beta_2}}dt,\ h_{12,1} = -\frac{\alpha_2}{\beta_1^2}\int_0^T \frac{\alpha_1 - X_t}{X_t^{2\beta_2}}dt,\ h_{12,2} = \frac{1}{\beta_1}\int_0^T \frac{1}{X_t^{\beta_2}}dW_t.$$

Note that the order of these terms are $T^{-1/2}$, T^{-1} and $T^{-3/2}$ for α_1, while those are $T^{-1/2}$ and $T^{-3/2}$ for α_2 since the T^{-1} order term vanishes. If we consider the case when the decreasing rate of Δ is fairly slow to make all the higher T-order terms negligible, we have

$$\begin{aligned}
\hat{\alpha}_1 - \alpha_1 \approx & -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} \\
& + \sqrt{\Delta}\Bigg[\frac{h_{12,1}\big[(h_{33}h_{44} - h_{34}^2)s_{2,d} + (h_{24}h_{34} - h_{23}h_{44})s_3 - (h_{24}h_{33} - h_{23}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)} \\
& \qquad - \frac{h_{22}\big[(h_{33}h_{44} - h_{34}^2)s_{1,d} + (h_{14}h_{34} - h_{13}h_{44})s_3 - (h_{14}h_{33} - h_{13}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)}\Bigg] \\
\hat{\alpha}_2 - \alpha_2 \approx & -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2} \\
& + \sqrt{\Delta}\Bigg[\frac{h_{12,1}\big[(h_{33}h_{44} - h_{34}^2)s_{1,d} + (h_{14}h_{34} - h_{13}h_{44})s_3 - (h_{14}h_{33} - h_{13}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)} \\
& \qquad - \frac{h_{11}\big[(h_{33}h_{44} - h_{34}^2)s_{2,d} + (h_{24}h_{34} - h_{23}h_{44})s_3 - (h_{24}h_{33} - h_{23}h_{34})s_4\big]}{(h_{11}h_{22} - h_{12,1}^2)(h_{33}h_{44} - h_{34}^2)}\Bigg],
\end{aligned}$$

for the Milstein ML estimation case, and

$$\begin{aligned}
\hat{\beta}_1 - \beta_1 &\approx -\sqrt{\Delta}\frac{h_{44}s_3 - h_{34}s_4}{h_{33}h_{44} - h_{34}^2} + \Delta^{-3/2}\frac{h_{34}s_{4,d}}{h_{33}h_{44} - h_{34}^2} \\
\hat{\beta}_2 - \beta_2 &\approx -\sqrt{\Delta}\frac{h_{33}s_4 - h_{34}s_3}{h_{33}h_{44} - h_{34}^2} - \Delta^{-3/2}\frac{h_{33}s_{4,d}}{h_{33}h_{44} - h_{34}^2},
\end{aligned}$$

where

$$h_{33} = -\frac{2T}{\beta_1^2},\ h_{44} = -2\int_0^T \log(X_t)^2 dt,\ h_{34} = -\frac{2}{\beta_1}\int_0^T \log(X_t)dt,$$
$$s_{1,d} = \sqrt{2}\alpha_2\beta_2\int_0^T \frac{1}{X_t}dV_t,\ s_{2,d} = \sqrt{2}\alpha_1\beta_2\int_0^T \frac{1}{X_t}dV_t - \frac{1+2\beta_2}{\sqrt{2}}V_T,$$
$$s_3 = \frac{\sqrt{2}}{\beta_1}\int_0^T X_t^{\alpha_2-\beta_2}dV_t,\ s_4 = \frac{\sqrt{2}\alpha_1}{\beta_1}\int_0^T \log(X_t)X_t^{\alpha_2-\beta_2}dV_t,$$
$$h_{14} = 3\alpha_2\beta_2\int_0^T \frac{\log(X_t)}{X_t}dt,\ h_{13} = \frac{3\alpha_2\beta_2}{\beta_1}\int_0^T \frac{1}{X_t}dt,\ h_{23} = \frac{3\beta_2}{\beta_1}\int_0^T \frac{\alpha_1 - X_t}{X_t}dt,$$
$$h_{24} = 3\beta_2\int_0^T \frac{(\alpha_1 - X_t)\log(X_t)}{X_t}dt,\ s_{4,d} = \frac{2\beta_1}{3^{1/4}}\int_0^T X_t^{\beta_2-1}\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}dU_t.$$

For the Euler ML estimation case, we have

$$\hat{\alpha}_1 - \alpha_1 \approx -\frac{h_{22}s_1 - h_{12,1}s_2}{h_{11}h_{22} - h_{12,1}^2} - \sqrt{\Delta}\left[\frac{h_{22}s_{1,d} - h_{12,1}s_{2,d}}{h_{11}h_{22} - h_{12,1}^2}\right]$$
$$\hat{\alpha}_2 - \alpha_2 \approx -\frac{h_{11}s_2 - h_{12,1}s_1}{h_{11}h_{22} - h_{12,1}^2} - \sqrt{\Delta}\left[\frac{h_{11}s_{2,d} - h_{12,1}s_{1,d}}{h_{11}h_{22} - h_{12,1}^2}\right]$$

with the followings replaced as

$$s_{1,d} = -\frac{\alpha_2\beta_2}{\sqrt{2}}\int_0^T \frac{1}{X_t}dV_t,\ s_{2,d} = -\frac{\alpha_1\beta_2}{\sqrt{2}}\int_0^T \frac{1}{X_t}dV_t - \frac{1-\beta_2}{\sqrt{2}}V_T.$$

CHAPTER IV

CONCLUSION

In this paper, I introduced a new asymptotics for the diffusion model estimation, and derived the asymptotic first and the higher order terms according to this asymptotics. As mentioned in the introduction, I could show where the big bias for the drift term parameter estimator comes from using this asymptotics, and could also show that we have very different characteristics for the drift and diffusion parameters. As we know the source of the bias and the distortion of the distribution, we can also think of many ways to correct them. In this paper I suggested a couple of correction methods which could successfully reduce the bias of the estimator and could get a more correct size for the hypothesis testing. Though the correction methods are in the baby steps now, I expect that there are many possibilities to utilize this new asymptotic result to get more efficient estimators and better test statistics with a correct size.

REFERENCES

Aït-Sahalia, Y. (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.

Aït-Sahalia, Y., and J. Y. Park (2008a): "Nonparametric Estimation of Diffusion Models," working paper, Texas A&M University.

——— (2008b): "Specification Testing for Nonstationary Diffusions," working paper, Texas A&M University.

Berman, S. M. (1964): "Limiting Distribution of the Maximum of a Diffusion Process," *Annals of Mathematical Statistics*, 35, 319-329.

Black, F., and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 673-654.

Borkovec, M., and C. Klüppelberg (1998): "Extremal Behaviour of Diffusion Models in Finance," *Extremes*, 1, 47-80.

Cline, D. B.H., M. Jeong, and J. Y. Park (2008): "Limit Theorems for the Integral of Diffusion Processes," working paper, Texas A&M University.

Cox, J., J. Ingersoll, and S. 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.

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.

Höpfner, R., and E. Löcherbach (2003): "Limit Theorems for Null Recurrent Markov Processes," Mem. Amer. Math. Soc., 161, no. 768.

Karatzas, I., and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, New York, NY: Springer-Verlag.

Karlin, S., and H. M. Taylor (1981): *A Second Course in Stochastic Processes,* New York, NY: Academic Press.

Kasahara, Y. (1975): "Spectral Theory of Generalized Second Order Differential Operators and its Applications to Markov Processes," *Japanese Journal of Mathematics*, 1, 67-84.

Khasminskii, R. (2001): "Limit Distributions of Some Integral Functionals for Null-Recurrent Diffusions," *Stochastic Processes and their Applications*, 92, 1-9.

Merton, R. C. (1971): "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3, 373-413.

Park, J. Y., and P. C. B. Phillips (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117-161.

Revuz, D. and M. Yor (1999): *Continuous Martingale and Brownian Motion,* New York, NY: Springer-Verlag.

Stone, C. (1963): "Limit Theorems for Random Walks, Birth and Death Processes, and Diffusion Processes," *Illinois Journal of Mathematics*, 7, 638-660.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-186.

Wooldridge, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics,* Vol. IV, ed. by R. F. Eagle and D. L. McFadden. Amsterdam, Elsvier, pp. 2639-2738.

APPENDIX A

PROOFS, LEMMAS, AND THE ASYMPTOTIC EXPANSIONS

A. Proofs and Useful Lemmas for Chapter II

I assume Assumptions 1-7 for the following lemmas. Here, $\mathbb{E}_t$ denotes a conditional expectation with information given up to time t. Hereafter, I define

$$e^+(x_i, y_i) = \exp\left(\frac{\sqrt{\sigma^2(x_i) + \Delta\sigma^2\sigma^{\cdot 2}(x_i) + 2\sigma\sigma^{\cdot}(x_i)(y_i - x_i - \Delta\mu(x_i))}}{\Delta\sigma\sigma^{\cdot 2}(x_i)}\right)$$
$$e^-(x_i, y_i) = 1/e^+(x_i, y_i),$$

for the simplicity.

1. Proof of Proposition 1

Part 1: Euler ML Case

Denote $x_i = X_{(i-1)\Delta}$ and $y_i = X_{i\Delta}$. Note that we have the scores of the likelihood $\mathcal{L}$ as $\mathcal{S}(\theta_0) = \sum_{i=1}^n \left(\ell_\alpha(x_i, y_i), \ell_\beta(x_i, y_i)\right)'$, where

$$\ell_\alpha(x_i, y_i) = \frac{\mu_\alpha(x_i)}{\sigma^2(x_i)}\big(y_i - x_i - \Delta\mu(x_i)\big)$$
$$\ell_\beta(x_i, y_i) = \frac{\sigma_\beta(x_i)}{\Delta\sigma^3(x_i)}\left[\big(y_i - x_i - \Delta\mu(x_i)\big)^2 - \Delta\sigma^2(x_i)\right]$$

and for the Hessians, we have

$$\mathcal{H}(\theta_0) = \sum_{i=1}^n \begin{pmatrix} \ell_{\alpha\alpha'}(x_i, y_i) & \ell_{\alpha\beta'}(x_i, y_i) \\ \ell_{\beta\alpha'}(x_i, y_i) & \ell_{\beta\beta'}(x_i, y_i) \end{pmatrix}$$

where

$$\begin{aligned}
\ell_{\alpha\alpha'}(x_i, y_i) &= \frac{\mu_{\alpha\alpha'}(x_i)}{\sigma^2(x_i)}\big(y_i - x_i - \Delta\mu(x_i)\big) - \frac{\Delta\mu_\alpha\mu'_\alpha(x_i)}{\sigma^2(x_i)} \\
\ell_{\alpha\beta'}(x_i, y_i) &= -\frac{2\mu_\alpha\sigma'_\beta(x_i)}{\sigma^3(x_i)}\big(y_i - x_i - \Delta\mu(x_i)\big) \\
\ell_{\beta\beta'}(x_i, y_i) &= \frac{1}{\Delta\sigma^4(x_i)}\Big[\big(\sigma\sigma_{\beta\beta'}(x_i) - 3\sigma_\beta\sigma'_\beta(x_i)\big)\big[\big(y_i - x_i - \Delta\mu(x_i)\big)^2 - \Delta\sigma^2(x_i)\big] \\
&\qquad\qquad - 2\Delta\sigma^2\sigma_\beta\sigma'_\beta(x_i)\Big].
\end{aligned}$$

Thus,it's easily derived from Lemma 9 and 10, that

$$\begin{aligned}
\sum_{i=1}^n \ell_\alpha(x_i, y_i) &= \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t + O_p(\sqrt{\Delta T}(\kappa_{\mu'_\alpha} - \kappa_{\mu_\alpha}\kappa_{\sigma^{\cdot}}\kappa_{\sigma^{-1}})(T)) \\
\sum_{i=1}^n \ell_\beta(x_i, y_i) &= \sqrt{\frac{2}{\Delta}}\int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t + O_p(\Delta^{-1/4-\zeta}F_{\sigma_\beta\sigma^{-3}}(T))
\end{aligned}$$

and this proves the first part of the proposition. Note that, for example,

$$\int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t = O_p\big(\sqrt{T}\kappa_{\mu_\alpha}\kappa_{\sigma^{-1}}(T)\big)$$

and

$$\sqrt{\Delta T}(\kappa_{\mu'_\alpha} - \kappa_{\mu_\alpha}\kappa_{\sigma^{\cdot}}\kappa_{\sigma^{-1}})(T) \to 0$$

from Assumption 15. From Lemma 9, 10 and 1, we have

$$
\begin{aligned}
\sum_{i=1}^{n} \ell_{\alpha\alpha'}(x_i, y_i) &= -\int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'}} - \kappa_{\mu_{\alpha\alpha'}} \kappa_{\sigma^{\cdot}} \kappa_{\sigma^{-1}})(T)) \\
\sum_{i=1}^{n} \ell_{\alpha\beta'}(x_i, y_i) &= -2\int_0^T \frac{\mu_\alpha \sigma'_\beta}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_\alpha} \kappa'_{\sigma_\beta} \kappa_{\sigma^{-1}} - 2\kappa_{\mu_\alpha} \kappa'_{\sigma_\beta} \kappa_{\sigma^{\cdot}} \kappa^2_{\sigma^{-1}} + \kappa_{\mu_\alpha} \kappa'_{\sigma^{\cdot}_\beta} \kappa_{\sigma^{-1}})(T)) \\
\sum_{i=1}^{n} \ell_{\beta\beta'}(x_i, y_i) &= -\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t)dt + O_p\left(\sqrt{\frac{T}{\Delta}}(\kappa_\sigma \kappa_{\sigma_{\beta\beta'}} - 3\kappa_{\sigma_\beta} \kappa'_{\sigma_\beta})\kappa^2_{\sigma^{-1}}(T)\right).
\end{aligned}
$$

Note that for $\ell_{\alpha\alpha}$ term, the second term will be of smaller order from Assumption 6 when $T \to \infty$ and $\Delta \to 0$, but when T is fixed, both the first term and the second term will be the leading term in the asymptotics. It's also easy to extend the vector case by applying these lemmas elementwise. As for the diagonality, it's easy to check that $H_0(\theta_0)$ will be block diagonal from

$$
\begin{aligned}
&\frac{\sqrt{\Delta}}{T}\kappa^2_\sigma \kappa^{-1}_{\mu_\alpha}(T) \int_0^T \frac{\mu_\alpha \sigma'_\beta}{\sigma^2}(X_t)dW_t \; \kappa^{-1\prime}_{\sigma_\beta}(T) \\
&\qquad = \frac{\sqrt{\Delta}}{T}\kappa^2_\sigma \kappa^{-1}_{\mu_\alpha}(T) \; O_p\big(\sqrt{T}\kappa_{\mu_\alpha} \kappa'_{\sigma_\beta} \kappa^2_{\sigma^{-1}}(T)\big) \; \kappa^{-1\prime}_{\sigma_\beta}(T) \to_p 0
\end{aligned}
$$

as $T \to \infty$ and $\Delta \to 0$. (The inverse operator is elementwise, for the notational convenience.)

Part 2: Milstein ML Case

Here I also denote $x_i = X_{(i-1)\Delta}$ and $y_i = X_{i\Delta}$. It's straightforward from the functional form of the score and Hessian functions, using Lemma 1-11 and 13. The basic procedure is same as the Euler case, but I will not go in detail for each case here. For example, for the score function with respect to the drift term parameter,

$$\frac{\partial \ell(x_i, y_i)}{\partial \alpha} = \left(\frac{e^+(x_i, y_i) - e^-(x_i, y_i)}{e^+(x_i, y_i) + e^-(x_i, y_i)}\right) \frac{\sqrt{\Delta}\mu_\alpha}{\sigma^\cdot G}(x_i, y_i) + \frac{\mu_\alpha}{\sigma\sigma^\cdot}(x_i, y_i) + \frac{\Delta^2 \sigma\sigma^\cdot \mu_\alpha}{G^2}(x_i, y_i)$$

where $G(x_i, y_i) = \left(\Delta\, \sigma(x_i)(\sigma(x_i) + \Delta\, \sigma\sigma^{\cdot 2}(x_i) + 2\, \sigma^\cdot(x_i)(y_i - x_i - \Delta\, \mu(x_i)))\right)^{1/2}$, suppressing all the arguments for the functions. Note that for the term containing $\frac{e^+(x,y)-e^-(x,y)}{e^+(x,y)+e^-(x,y)}$, it's same as finding the limiting distribution without $\frac{e^+(x,y)-e^-(x,y)}{e^+(x,y)+e^-(x,y)}$ from Lemma 11, and for the terms with $B(x, y)$, they can be taken care of by Lemma 13, and as a result, we get the following terms.

$$\begin{aligned} \sum_{i=1}^{n} \frac{\partial \ell(x_i, y_i)}{\partial \alpha} =& \sum_{i=1}^{n} \frac{\mu_\alpha}{\sigma^2}(x_i)(y_i - x_i - \Delta\, \mu(x_i)) \\ &- \frac{3}{2} \sum_{i=1}^{n} \frac{\mu_\alpha \sigma^\cdot}{\sigma^3}(x_i) \left[(y_i - x_i - \Delta\, \mu(x_i))^2 - \Delta\sigma^2(x_i)\right] \\ &+ O_p(\Delta\sqrt{T}\kappa_{\sigma^\cdot}^2 \kappa_{\mu_\alpha} \kappa_{\sigma^{-1}}^2(T)) \end{aligned}$$

So the rest of the step is to find the limiting distribution of each terms, and we get

$$\sum_{i=1}^{n} \frac{\partial \ell(x_i, y_i)}{\partial \alpha} = \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t + O_p(\sqrt{\Delta T}\kappa_{\mu_\alpha}\kappa_{\sigma^\cdot}\kappa_{\sigma^{-1}}(T))$$

using Lemma 8 and 10.

Part 3: Exact ML Case

For the score terms w.r.t. α, what we want to show is

$$\sum_{i=1}^{n} \ell_\alpha(x_i, y_i, \Delta) \approx \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t. \tag{A.1}$$

Since the function ℓ_α is not only a function of x, but also a function of y and Δ, we can consider the following Taylor expansion

$$\begin{aligned}\sum_{i=1}^{n}\ell_{\alpha}(x_i,y_i,\Delta)=&\sum_{i=1}^{n}\ell_{\alpha}(x_i,x_i,\Delta)+\sum_{i=1}^{n}\ell_{\alpha y}(x_i,x_i,\Delta)(y_i-x_i)\\&+\frac{1}{2}\sum_{i=1}^{n}\ell_{\alpha yy}(x_i,x_i,\Delta)(y_i-x_i)^2+O_p(R_{1(T,\Delta)})\end{aligned}$$

for some order $R_{1(T,\Delta)}$. (Hereafter, I will denote the order of the remainder term as $R_{k(T,\Delta)}$.) Note that $R_{1(T,\Delta)}$ will be of smaller order from Assumption 11. Denoting $\mathcal{W}_i = W_{i\Delta} - W_{(i-1)\Delta}$ for the simplicity, we can replace $(y_i - x_i)$ with

$$y_i - x_i = \Delta\mu(x_i) + \sigma(x_i)\mathcal{W}_i + R_i$$

where R_i is a remainder term, and we have

$$\begin{aligned}\sum_{i=1}^{n}\ell_{\alpha}(x_i,y_i,\Delta)=&\sum_{i=1}^{n}\ell_{\alpha y}(x_i,x_i,\Delta)\sigma(x_i)\mathcal{W}_i\\&+\sum_{i=1}^{n}\ell_{\alpha}(x_i,x_i,\Delta)+\Delta\sum_{i=1}^{n}\ell_{\alpha y}(x_i,x_i,\Delta)\mu(x_i)\\&+\frac{1}{2}\sum_{i=1}^{n}\ell_{\alpha yy}(x_i,x_i,\Delta)\sigma^2(x_i)\mathcal{W}_i^2+O_p(R_{2(T,\Delta)})\\=&A_T+B_T+O_p(R_{2(T,\Delta)}).\end{aligned}$$

Note that R_2 becomes of smaller order by Assumption 8. To make (A.1) hold, we should have

$$A_T \approx \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t \tag{A.2}$$

and B_T should be of smaller order than A_T. To show (A.2), we can do the Taylor expansion w.r.t. Δ again, then

$$A_T=\sum_{i=1}^{n}\ell_{\alpha y}(x_i,x_i,0)\sigma(x)\mathcal{W}_i+\Delta\sum_{i=1}^{n}\ell_{\alpha y\Delta}(x_i,x_i,0)\sigma(x_i)\mathcal{W}_i+O_p(R_3)$$

and from the following condition in Assumption 12,

$$\ell_{\alpha y}(x_i, x_i, 0) = \frac{\mu_\alpha}{\sigma^2}(x_i),$$

we can check that (A.2) holds. Note that R_3 is of smaller order by Assumption 10. Similarly, applying Taylor expansions w.r.t. Δ for each term in B_T, and using the following condition,

$$\ell_{\alpha\Delta}(x_i, x_i, 0) + \frac{1}{2}\ell_{\alpha yy}(x_i, x_i, 0)\sigma^2(x_i) = \frac{\mu\mu_\alpha}{\sigma^2}(x_i)$$

with $\ell_\alpha(x_i, x_i, 0) = 0$, we can be sure that B_T is of smaller order than A_T. Thus, with the following conditions,

$$\begin{aligned}
\ell_{\alpha y}(x_i, x_i, 0) &= \frac{\mu_\alpha}{\sigma^2}(x_i) \\
\ell_{\alpha\Delta}(x_i, x_i, 0) + \frac{1}{2}\ell_{\alpha yy}(x_i, x_i, 0)\sigma^2(x_i) &= \frac{\mu\mu_\alpha}{\sigma^2}(x_i) \\
\ell_\alpha(x_i, x_i, 0) &= 0
\end{aligned}$$

we can show that (A.1) holds. For the scores w.r.t. β, we want show

$$\sum_{i=1}^{n} \ell_\beta(x_i, y_i, \Delta) \approx \sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t.$$

and following the similar steps, it can be shown under the following conditions as in Assumption 12,

$$\begin{aligned}
\ell_\beta(x_i, x_i, 0) &= -\frac{\sigma_\beta}{\sigma}(x_i) \\
\Delta\ell_{\beta yy}(x_i, x_i, \Delta) &\to \frac{2\sigma_\beta}{\sigma^3}(x_i) \quad \text{as } \Delta \to 0.
\end{aligned}$$

For the Hessian terms w.r.t. α, we want to show

$$\sum_{i=1}^{n} \ell_{\alpha\alpha}(x_i, y_i, \Delta) = -\int_0^T \frac{\mu_\alpha^2}{\sigma^2}(X_t)dt\,\big(1 + o_p(1)\big),$$

and similarly, the conditions to make this leading term is

$$\ell_{\alpha\alpha}(x_i, x_i, 0) = 0$$

$$\ell_{\alpha\alpha y}(x_i, x_i, 0) = 0$$

$$\ell_{\alpha\alpha\Delta}(x_i, x_i, 0) + \frac{1}{2}\ell_{\alpha\alpha yy}(x_i, x_i, 0)\sigma^2(x_i) = -\frac{\mu_\alpha^2}{\sigma^2}(x_i)$$

as in Assumption 12. For the Hessian w.r.t. β, we want show

$$\sum_{i=1}^n \ell_{\beta\beta}(x_i, y_i, \Delta) \approx -\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta^2}{\sigma^2}(X_t)dt,$$

and the conditions are

$$\ell_{\beta\beta}(x_i, x_i, \Delta) + \frac{\Delta}{2}\ell_{\beta\beta yy}(x_i, x_i, \Delta)\sigma^2(x_i) \to -\frac{2\sigma_\beta^2}{\sigma^2}(x_i) \quad \text{as } \Delta \to 0.$$

For the off-diagonal blocks of the Hessian, we should have

$$\ell_{\alpha\beta}(x_i, x_i, 0) = 0$$

$$\sqrt{\Delta}\ell_{\alpha\beta y}(x_i, x_i, \Delta) \to 0 \quad \text{as } \Delta \to 0$$

$$\sqrt{\Delta}\ell_{\alpha\beta\Delta}(x_i, x_i, \Delta) + \frac{\sqrt{\Delta}}{2}\ell_{\alpha\beta yy}(x_i, x_i, \Delta)\sigma^2(x_i) \to 0 \quad \text{as } \Delta \to 0$$

to make them asymptotically negligible.

2. Proof of Proposition 2

Part 1: Euler ML Case

We need to show

$$\sup_{\theta\in\mathcal{N}} \left| v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1\prime} \right| \to_p 0$$

where $\mathcal{N} = \{\theta : |v'(\theta - \theta_0)| \leq 1\}$. Here, I let w as defined in AD2, and $v = T^{-\varepsilon}w$ for some $\varepsilon > 0$, so that it satisfies $vw^{-1} \to 0$. To prove this, note that we have

$$\begin{aligned}\sup_{\theta\in\mathcal{N}}\left|v^{-1}\big(\mathcal{H}(\theta)-\mathcal{H}(\theta_0)\big)v^{-1'}\right| &= \sup_{\theta\in\mathcal{N}}\left|v^{-1}\mathcal{J}(\tilde{\theta})\big((\theta-\theta_0)\otimes v^{-1'}\big)\right| \\ &\leq v^{-1}\sup_{\theta\in\mathcal{N}}\big|\mathcal{J}(\theta)\big|\big(\bar{v}\otimes v^{-1'}\big)\end{aligned}$$

where $\tilde{\theta}$ is a value in the line connecting θ and θ_0, K is the number of the parameters, i.e., the length of a vector θ, $\mathcal{J}(\theta) = \frac{\partial}{\partial\theta}\mathrm{vec}(\mathcal{H}(\theta))'$, and $\bar{v} = \mathrm{diag}\big(v^{-1}\big)$. To show that this converges to 0, I will first show that the order difference between $\mathcal{H}(\theta_0)$ and $\mathcal{J}(\theta_0)$ is small enough compared to $\bar{v}$, and next, that $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ has the same asymptotic order as $\mathcal{J}(\theta_0)$. And after that, the rest is just an application of these results, to show

$$v^{-1}\sup_{\theta\in\mathcal{N}}\big|\mathcal{J}(\theta)\big|\big(\bar{v}\otimes v^{-1'}\big) \to_p 0.$$

as $T \to \infty$ and $\Delta \to 0$. Hereafter, I will denote $x_i = X_{(i-1)\Delta}$ and $y_i = X_{i\Delta}$ for the simplicity.

Step 1. To check the difference of the order between $\mathcal{H}(\theta_0)$ and $\mathcal{J}(\theta_0)$, let's first consider the order of $\mathcal{H}(\theta_0)$. Denoting (j,l) element of $\mathcal{H}(\theta_0)$ as $\bar{h}_{jl} = \sum_{i=1}^{n} h_{i,jl}$, note that $h_{i,jl}$ has the following form,

$$h_{i,jl}(\theta) = \Delta^{s_1}\big(y_i - x_i - \Delta\mu(x_i,\theta)\big)^r a(x_i,\theta) + \Delta^{s_2}b(x_i,\theta)$$

with $r = 1, 2$. On the other hand, we have

$$\begin{aligned}\frac{\partial}{\partial\theta}h_{i,jl}(\theta) &= \Delta^{s_1}\big(y_i - x_i - \Delta\mu(x_i,\theta)\big)^r a_\theta(x_i,\theta) + \Delta^{s_2}b_\theta(x_i,\theta) \\ &\quad + r\Delta^{s_1+1}\big(y_i - x_i - \Delta\mu(x_i,\theta)\big)^{r-1}a\mu_\theta(x_i,\theta).\end{aligned}$$

Note that the derivative has a same form as $h_{i,jl}(\theta)$ but only with derivatives of each function, so it's easy to check that

$$v^{-1}\mathcal{J}(\theta_0)\big(\bar{v}\otimes v^{-1'}\big) = T^{3\varepsilon}w^{-1}\mathcal{J}(\theta_0)\big(\bar{w}\otimes w^{-1'}\big)$$

and $\bar{w} = \mathrm{diag}\big(w^{-1}\big)$. Since there exists $a > 0$ such that $T^a\bar{w} \to_p 0$ and we can choose $\varepsilon < a/3$, we have

$$T^{3\varepsilon}w^{-1}\mathcal{J}(\theta_0)\big(\bar{w}\otimes w^{-1'}\big) \leq T^{3\varepsilon-a}w^{-1}\mathcal{J}(\theta_0)\big(\iota_k\otimes w^{-1'}\big) \to_p 0$$

for large enough T from Assumption 6, where ι_k is k by 1 one vector and k is the number of rows in $\bar{w}$.

Step 2. We will next show that $\mathcal{J}(\theta_0)$ and $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ have the same asymptotic order, i.e., if we have

$$\eta\mathcal{J}(\theta_0)\big(\bar{\eta}\otimes\eta'\big) = O_p(1),$$

with an appropriate matrix and a vector η and $\bar{\eta}$, then we also have

$$\sup_{\theta\in\mathcal{N}}\Big|\eta\mathcal{J}(\theta)\big(\bar{\eta}\otimes\eta'\big)\Big| = O_p(1).$$

For this, denoting $\bar{h}_k^{jl}(\theta)$ as (j,lk) element of $\mathcal{J}(\theta)$ and η_k^{jl} as its corresponding convergence rate, it's enough to show that,

$$\sup_{\theta\in\mathcal{N}}\Big|\eta_k^{jl}\bar{h}_k^{jl}(\theta)\Big| = O_p(1)$$

when we have $\eta_k^{jl}\bar{h}_k^{jl}(\theta_0) = O_p(1)$, for each k, j and l. We will suppress all the superscripts hereafter for the simplicity, that is, $\bar{h} = \bar{h}_k^{jl}$ and $\eta = \eta_k^{jl}$. Note that $\bar{h}$ also has the following form as previously denoted,

$$\bar{h}(\theta) = \sum_{i=1}^{n}\Delta^{s_1}\big(y_i - x_i - \Delta\mu(x_i,\theta)\big)^r a(x_i,\theta) + \Delta^{s_2}b(x_i,\theta).$$

Denoting $\eta(\theta)$ as the order of $\bar{h}(\theta)$, i.e.,

$$\eta(\theta)\bar{h}(\theta) = O_p(1)$$

note that $\eta(\theta)$ has a form $\Delta^{s1}T^{s2}\kappa_g(T,\theta)$, where κ_g is a product of some asymptotic order functions which appear in Assumption 13. Explicitly denoting $\eta = \eta(\theta_0)$ as a function $\eta(\theta)$ evaluated at θ_0, we have

$$\begin{aligned}\sup_{\theta\in\mathcal{N}}\left|\eta(\theta_0)\bar{h}(\theta)\right| =& \sup_{\theta\in\mathcal{N}}\left|\eta(\theta_0)\eta(\theta)^{-1}\eta(\theta)\bar{h}(\theta)\right| \\ =& \sup_{\theta\in\mathcal{N}}\left|\eta(\theta_0)\eta(\theta)^{-1}\right| O_p(1).\end{aligned}$$

If we only consider the case of one function for the simplicity, we have

$$\eta(\theta) = T\kappa_f(T,\theta)$$

and we have

$$\sup_{\theta\in\mathcal{N}}\left|\eta(\theta_0)\eta(\theta)^{-1}\right| = \sup_{\theta\in\mathcal{N}}\left|\frac{\kappa_f(T,\theta_0)}{\kappa_f(T,\theta)}\right| \to_p 1$$

by Assumption 7. Generalization for multiple product is also not difficult. So now I showed that $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ has the same order as $\mathcal{J}(\theta_0)$, and the rest steps are same as already described in the beginning.

Part 2: Milstein ML Case

We need to show

$$\sup_{\theta\in\mathcal{N}}\left|v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1\prime}\right| \to_p 0$$

where $\mathcal{N} = \{\theta : |v'(\theta-\theta_0)| \leq 1\}$. Here, I let w as defined in AD2, and $v = T^{-\varepsilon}w$ for some $\varepsilon > 0$, so that it satisfies $vw^{-1} \to 0$. To prove this, note that we have

$$\begin{aligned}\sup_{\theta\in\mathcal{N}} \left| v^{-1}\big(\mathcal{H}(\theta) - \mathcal{H}(\theta_0)\big)v^{-1'} \right| &= \sup_{\theta\in\mathcal{N}} \left| v^{-1}\mathcal{J}(\tilde{\theta})\big((\theta-\theta_0)\otimes v^{-1'}\big) \right| \\ &\le v^{-1}\sup_{\theta\in\mathcal{N}} \big|\mathcal{J}(\theta)\big|\big(\bar{v}\otimes v^{-1'}\big)\end{aligned}$$

where $\tilde{\theta}$ is a value in the line connecting θ and θ_0, K is the number of the parameters, i.e., the length of a vector θ, $\mathcal{J}(\theta) = \frac{\partial}{\partial\theta}\mathrm{vec}(\mathcal{H}(\theta))'$, and $\bar{v} = \mathrm{diag}\big(v^{-1}\big)$. To show that this converges to 0, I will first show that the order difference between $\mathcal{H}(\theta_0)$ and $\mathcal{J}(\theta_0)$ is small enough compared to $\bar{v}$, and next, that $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ has the same asymptotic order as $\mathcal{J}(\theta_0)$. And after that, the rest is just an application of these results, to show

$$v^{-1}\sup_{\theta\in\mathcal{N}} \big|\mathcal{J}(\theta)\big|\big(\bar{v}\otimes v^{-1'}\big) \to_p 0.$$

as $T \to \infty$ and $\Delta \to 0$. Hereafter, I will denote $x_i = X_{(i-1)\Delta}$ and $y_i = X_{i\Delta}$ for the simplicity.

Step 1. To check the difference of the order between $\mathcal{H}(\theta_0)$ and $\mathcal{J}(\theta_0)$, let's first consider the order of $\mathcal{H}(\theta_0)$. Denoting (j,l) element of $\mathcal{H}(\theta_0)$ as $\bar{h}_{jl} = \sum_{i=1}^n h_{i,jl}$, note that $h_{i,jl}$ has the following form,

$$\begin{aligned}h_{i,jl}(\theta) = \sum_{r=1}^{\bar{r}} \Big[& A(x_i,y_i)^{p_r}B(x_i,y_i)^{q_r}\Delta^{s_r}\big(y_i - x_i - \Delta\mu(x_i,\theta)\big)^{w_r} d(x_i,\theta)\times \\ & \big[a(x_i,\theta) + \Delta b(x_i,\theta) + \big(y_i - x_i - \Delta\mu(x_i,\theta)\big)c(x_i,\theta)\big]^{u_r}\Big]\end{aligned}$$

for some $\bar{r} < \infty$, where

$$A(x_i,y_i) = \frac{e^+(x_i,y_i) - e^-(x_i,y_i)}{e^+(x_i,y_i) + e^-(x_i,y_i)} \quad \text{and} \quad B(x_i,y_i) = \frac{1}{e^+(x_i,y_i) + e^-(x_i,y_i)},$$

p_r, q_r, w_r being non-negative (possibly zero) integers and s_r, u_r being possibly negative real numbers. As for this functional form, one can check this by looking at the actual derivatives of log-likelihood function in Appendix II. For this function $h_{i,jl}$, here I will only focus on the terms with $q_r = 0$ since it can be shown from the proof of Lemma 11 that those terms with positive q_r will be of smaller order than the other terms. Now with this functional form, and assuming that we have the biggest order term for $r = r^*$, we can write $h_{i,jl}(\theta)$ as, ignoring all the smaller order terms,

$$\begin{aligned} h_{i,jl}(\theta) &= A(x_i, y_i)^{p_*} \Delta^{s_*} \big(y_i - x_i - \Delta\mu(x_i, \theta)\big)^{w_*} d(x_i, \theta) \\ &\qquad \big[a(x_i, \theta) + \Delta b(x_i, \theta) + \big(y_i - x_i - \Delta\mu(x_i, \theta)\big) c(x_i, \theta)\big]^{u_*} \big(1 + o_p(1)\big) \\ &\equiv \Delta^{s_*} h(x_i, y_i, \Delta, \theta)\big(1 + o_p(1)\big), \end{aligned}$$

where the subscript $_*$ denotes the corresponding values for $r = r^*$. By Lemma 12, we have

$$v(\theta_0)^{-1} \Delta^{s_*} \sum_{i=1}^{n} h(x_i, y_i, \Delta, \theta_0) = O_p(1)$$

where $v(\theta_0)^{-1}$ is the order of

$$\Delta^{s_*} \sum_{i=1}^{n} h^*(x_i, \theta_0) = \Delta^{s_*} \sum_{i=1}^{n} v(x_i, \theta) a^{u_*}(x_i, \theta).$$

Note that from Assumption 13, $v(x_i, \theta) a^{u_*}(x_i, \theta)$ will have the following form of the product of several functions, but here, I will only consider when there are 2 functions only, such that,

$$v(x_i, \theta) a^{u}(x_i, \theta) = f(x_i, \theta)^p g(x_i, \theta)^q$$

for some asymptotically homogeneous functions f and g and real numbers p and q. Also, in this case, if we think about $\mathcal{J}(\theta)$, a derivative of $\mathcal{H}(\theta)$ w.r.t. the parameters, the biggest order term will be

$$\begin{aligned}\Delta^{s_r}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}h^*(x_i,\theta_0) =&\Delta^{s_r}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}v(x_i,\theta_0)a^u(x_i,\theta_0)\\ =&\Delta^{s_r}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}f(x_i,\theta_0)^p g(x_i,\theta_0)^q\\ =&\Delta^{s_r}\sum_{i=1}^{n}\big(pf(x_i,\theta_0)^{p-1}f^{\cdot}(x_i,\theta_0)+qg(x_i,\theta_0)^{q-1}g'(x_i,\theta_0)\big)\end{aligned}$$

and from Assumption 15, it's obvious that

$$\sum_{k=1}^{K} v^{-1}\mathcal{J}(\theta)\big(\bar{v}\otimes v^{-1'}\big)\to_p 0.$$

A generalization for the cases that consists of multiple product of functions is also straightforward.

Step 2. We will next show that $\mathcal{J}(\theta_0)$ and $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ have the same asymptotic order, i.e., if we have

$$\eta\mathcal{J}(\theta_0)\big(\bar{\eta}\otimes\eta'\big)=O_p(1),$$

with an appropriate matrix and a vector η and $\bar{\eta}$, then we also have

$$\sup_{\theta\in\mathcal{N}}\Big|\eta\mathcal{J}(\theta)\big(\bar{\eta}\otimes\eta'\big)\Big|=O_p(1).$$

For this, denoting $\bar{h}_k^{jl}(\theta)$ as (j,lk) element of $\mathcal{J}(\theta)$ and η_k^{jl} as its corresponding convergence rate, it's enough to show that,

$$\sup_{\theta\in\mathcal{N}}\Big|\eta_k^{jl}\bar{h}_k^{jl}(\theta)\Big|=O_p(1)$$

when we have $\eta_k^{jl}\bar{h}_k^{jl}(\theta_0)=O_p(1)$, for each k and (j,l). We will suppress all the superscripts hereafter for the simplicity, that is, $\bar{h}=\bar{h}_k^{jl}$ and $\eta=\eta_k^{jl}$. Note that h_i

($\bar{h} = \sum_{i=1}^{n} h_i$) also has the following form as previously denoted,

$$h_i(\theta) = \sum_{r=1}^{\bar{r}} \Big[A_i^{p_r} B_i^{q_r} \Delta^{s_r} \big(y_i - x_i - \Delta\mu(x_i, \theta)\big)^{w_r} v(x_i, \theta) \times$$
$$\big[a(x_i, \theta) + \Delta b(x_i, \theta) + \big(y_i - x_i - \Delta\mu(x_i, \theta)\big) c(x_i, \theta)\big]^{u_r} \Big].$$

We will only focus on the terms with $q_r = 0$ with the same reason as before. Again, we can express $\bar{h}(\theta)$ as

$$\bar{h}(\theta) = \Delta^{s_r} \sum_{i=1}^{n} h(x_i, y_i, \Delta, \theta),$$

and by Lemma 12,

$$\eta(\theta) \Delta^{s_r} \sum_{i=1}^{n} h(x_i, y_i, \Delta, \theta) = O_p(1)$$

where $\eta(\theta)$ is the order of

$$\Delta^{s_*} \sum_{i=1}^{n} h^*(x_i, \theta) = \Delta^{s_*} \sum_{i=1}^{n} v(x_i, \theta) a^{u_*}(x_i, \theta).$$

So, since η is the order of $\Delta^{s_*} \sum h^*(x_i, \theta_0)$, explicitly denoting $\eta = \eta(\theta_0)$ as a function $\eta(\theta)$ evaluated at θ_0, we have

$$\begin{aligned} \sup_{\theta \in \mathcal{N}} \left| \eta(\theta_0) \bar{h}(\theta) \right| &= \sup_{\theta \in \mathcal{N}} \left| \eta(\theta_0) \eta(\theta)^{-1} \eta(\theta) \bar{h}(\theta) \right| \\ &= \sup_{\theta \in \mathcal{N}} \left| \eta(\theta_0) \eta(\theta)^{-1} \right| O_p(1). \end{aligned}$$

Note that from Assumption 13, $v(x, \theta) a^{u_*}(x, \theta)$ also consists of the product of the functions, so if I only consider the case of one function,

$$v(x, \theta) a^{u_*}(x, \theta) = f(x, \theta)$$

and in this case, we have

$$\eta(\theta) = T\kappa_f(T,\theta)$$

and we have

$$\sup_{\theta\in\mathcal{N}}\left|\eta(\theta_0)\eta(\tilde{\theta})^{-1}\right| = \sup_{\theta\in\mathcal{N}}\left|\frac{\kappa_f(T,\theta_0)}{\kappa_f(T,\theta)}\right| \to_p 1$$

by Assumption 6. So now I showed that $\sup_{\theta\in\mathcal{N}}|\mathcal{J}(\theta)|$ has the same order as $\mathcal{J}(\theta_0)$, and the rest steps are same as already described in the beginning.

3. Useful Lemmas

Lemma 1. *Let f be a twice differentiable function, and let f and its derivatives satisfy Assumption 2. Then,*

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})\Delta = \int_0^T f(X_t)dt + O_p\big(\Delta T(\kappa_\mu\kappa_{f^\cdot} + \kappa_\sigma^2\kappa_{f^{\cdot\cdot}})(T)\big) + O_p\big(\Delta\sqrt{T}\kappa_\sigma\kappa_{f^\cdot}(T)\big).$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} f(X_{(i-1)\Delta})\Delta &= \int_0^T f(X_t)dt - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big(f(X_t) - f(X_{(i-1)\Delta})\big)dt \\
&= \int_0^T f(X_t)dt - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t}\left(\mu f^\cdot + \frac{\sigma^2 f^{\cdot\cdot}}{2}\right)(X_s)dsdt \\
&\qquad - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t}\sigma f^\cdot(X_s)dW_sdt \\
&= \int_0^T f(X_t)dt + A_T + B_T
\end{aligned}$$

by Itô's lemma.

$$\begin{aligned} A_T &= \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} (s-(i-1)\Delta) \left(\mu f^{\cdot} + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right) (X_s) ds \\ &\leq \Delta \int_0^T \left| \mu f^{\cdot} + \frac{\sigma^2 f^{\cdot\cdot}}{2} \right| (X_t) dt \qquad \text{(A.3)} \\ &= O_p\big(\Delta T \kappa_\mu \kappa_{f^{\cdot}}(T)\big) + O_p\big(\Delta T \kappa_\sigma^2 \kappa_{f^{\cdot\cdot}}(T)\big). \end{aligned}$$

Also,

$$B_T = \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} (s-(i-1)\Delta) \sigma f^{\cdot}(X_s) dW_s$$

and this is a martingale whose quadratic variation is bounded by

$$\Delta^2 \int_0^T \sigma^2 f^{\cdot 2}(X_t) dt = O_p\big(\Delta^2 T \kappa_\sigma^2 \kappa_{f^{\cdot}}^2(T)\big).$$

So the remainder terms are of order

$$A_T + B_T = O_p\big(\Delta T \kappa_\mu \kappa_{f^{\cdot}}(T)\big) + O_p\big(\Delta T \kappa_\sigma^2 \kappa_{f^{\cdot\cdot}}(T)\big) + O_p\big(\Delta \sqrt{T} \kappa_\sigma \kappa_{f^{\cdot}}(T)\big).$$

□

Lemma 2. *For g and f satisfying Assumption 2,*

(a) if the following repeated integrations only consist of the time (dt) integtation,

$$\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^{s} f(X_r) dr \cdots dt = O_p(\Delta^{k-1} T \kappa_g \kappa_f(T))$$

where k is the number of the repeated integrations,

(b) and otherwise,

$$\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^{s} f(X_r) dr \cdots dW_t = O_p(\Delta^{(2k_1+k_2-1)/2} \sqrt{T} \kappa_g \kappa_f(T))$$

where k_1 is the number of integrals w.r.t. the time, and k_2 is the number of integral w.r.t. the Brownian motion. Here, though I could not write appropriately, in the *expression for the repeated integration, the integral can be with respect to either time (dt) or the Brownian motion (dW_t) with any combinations of the two, which has at least one dW_t term.*

Proof. (a) Applying the same technique in the proof of Lemma 1, we can show that

$$\begin{aligned}
&\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \cdots \int_{(i-1)\Delta}^{s} f(X_r)dr \cdots dt \\
&\quad = \sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (r-(i-1)\Delta)^{k-1} f(X_r)dr \\
&\quad \leq \Delta^{k-1} \sum_{i=1}^{n} |g(X_{(i-1)\Delta})| \int_{(i-1)\Delta}^{i\Delta} |f(X_r)|dr \\
&\quad \leq \Delta^{k-1} T \sup_{0\leq t\leq T} \left|g(X_t)\right| \sup_{0\leq t\leq T} \left|f(X_t)\right| \\
&\quad = O_p(\Delta^{k-1} T \kappa_g \kappa_f(T)).
\end{aligned}$$

(b) First, note that we can make the most outer integration w.r.t. the Brownian motion by change of the integration, to have

$$\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \frac{(i\Delta-u)^k}{k!} \int_{(i-1)\Delta}^{u} \cdots \int_{(i-1)\Delta}^{s} f(X_r)dr \cdots dW_v dW_u$$

where k is the number of dt integrations at the most outer side. This is a martingale with a quadratic variation bounded by

$$\Delta^{2k} \sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{u} \cdots \int_{(i-1)\Delta}^{s} f(X_r)dr \cdots dW_v \right)^2 du$$

and since this is a positive process, its order is the same as the order of its expectation. We have

$$\begin{aligned}
&\Delta^{2k} \mathbb{E} \left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{u} \cdots \int_{(i-1)\Delta}^{s} f(X_r)dr \cdots dW_v \right)^2 du \right) \\
&\quad = \Delta^{2k} \mathbb{E} \left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \mathbb{E}_{(i-1)\Delta} \left(\int_{(i-1)\Delta}^{u} \cdots \int_{(i-1)\Delta}^{s} f(X_r)dr \cdots dW_v \right)^2 du \right).
\end{aligned}$$

For the condition expectation part, we can change the order of the integrals to reduce the number of integrations. When the most inner integral is w.r.t. the Brownian motion, this is bounded by

$$\Delta^{2k_1+k_2-1}\sum_{i=1}^{n} g^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}_{(i-1)\Delta}f^2(X_r)dr = O_p(\Delta^{2k_1+k_2-1}T\kappa_g^2\kappa_f^2(T))$$

Note that this is also a positive process so the order is same as the expectation

$$\begin{aligned}
\Delta^{2k_1+k_2-1}\mathbb{E}&\left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}_{(i-1)\Delta}f^2(X_r)dr\right)\\
&= \Delta^{2k_1+k_2-1}\mathbb{E}\left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}f^2(X_r)dr\right)\\
&\leq \Delta^{2k_1+k_2-1}T\,\mathbb{E}\left(\sup_{0\leq t\leq T}\left|g^2(X_t)\right|\sup_{0\leq t\leq T}\left|f^2(X_t)\right|\right)\\
&= O_p(\Delta^{2k_1+k_2-1}T\kappa_g^2\kappa_f^2(T)).
\end{aligned}$$

Also,

$$\Delta^{2k_1+k_2-3}\sum_{i=1}^{n} g^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}_{(i-1)\Delta}\left(\int_{(i-1)\Delta}^{t}f(X_s)ds\right)^2dt$$

has the same order as

$$\begin{aligned}
\Delta^{2k_1+k_2-3}\mathbb{E}&\left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\left(\int_{(i-1)\Delta}^{t}f(X_s)ds\right)^2dt\right)\\
&\leq \Delta^{2k_1+k_2-3}\mathbb{E}\left(\sum_{i=1}^{n}\sup_{0\leq t\leq T}\left|g^2(X_t)\right|\int_{(i-1)\Delta}^{i\Delta}\left(\int_{(i-1)\Delta}^{t}\sup_{0\leq t\leq T}\left|f(X_t)\right|ds\right)^2dt\right)\\
&\leq \Delta^{2k_1+k_2-1}T\,\mathbb{E}\left(\sup_{0\leq t\leq T}\left|g^2(X_t)\right|\sup_{0\leq t\leq T}\left|f(X_t)\right|^2\right)\\
&= O_p(\Delta^{2k_1+k_2-1}T\kappa_g^2\kappa_f^2(T))
\end{aligned}$$

when the most inner integral is w.r.t. the time. $\square$

Lemma 3. *Let A_{ij} be one of the followings,*

$$\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f_j(X_s)dsdt$$
$$\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f_j(X_s)dsdW_t$$
$$\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f_j(X_s)dW_sdt$$
$$\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f_j(X_s)dW_sdW_t$$

for f_j's satisfying Assumption 2, and B_{ij} be the same integral without the function $f_j(X_s)$. Let $V_{i\Delta} = (W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^k B_{ij}$.

(a) If $\mathbb{E}V_\Delta = 0$, we have

$$\sum_{i=1}^n g(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^k A_{ij} = O_p(\kappa_{1\Delta}\sqrt{T}\kappa_g\kappa_{f_1}\cdots\kappa_{f_k}(T))$$

where $\kappa_{1\Delta} = \left(\mathbb{E}V_\Delta^2/\Delta\right)^{1/2}$

(b) Otherwise,

$$\sum_{i=1}^n g(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^k A_{ij} = O_p(\kappa_{2\Delta}T\kappa_g\kappa_{f_1}\cdots\kappa_{f_k}(T))$$

where $\kappa_{2\Delta} = \mathbb{E}V_\Delta/\Delta$.

($\kappa_{1\Delta} = c_1\Delta^{2k_1+k_2+\frac{3}{2}k_3+\frac{3}{2}k_4+\frac{1}{2}k_w-\frac{1}{2}}$ and $\kappa_{2\Delta} = c_2\Delta^{2k_1+k_2+\frac{3}{2}k_3+\frac{3}{2}k_4+\frac{1}{2}k_w-1}$.)

Proof. Replacing each $f_j(X_t)$ with $f_j(X_{(i-1)\Delta}) + (f_j(X_t) - f_j(X_{(i-1)\Delta}))$ and arranging them, we have

$$\sum_{i=1}^n g(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^k A_{ij}$$
$$= \sum_{i=1}^n g(X_{(i-1)\Delta}) \prod_{j=1}^k f_j(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^k B_{ij} + R$$

where R represents a remainder term. When $\mathbb{E}V_\Delta = 0$, the order of the leading term can be obtained from the expectation of the square, that is,

$$\mathbb{E}\left(\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^{k} B_{ij}\right)^2$$
$$=\mathbb{E}\left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j^2(X_{(i-1)\Delta}) \mathbb{E}_{(i-1)\Delta}\left((W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^{k} B_{ij}\right)^2\right)$$

from the independent increment property of the Brownian motion, so

$$\mathbb{E}\left(\mathbb{E}_{(i-1)\Delta} V_{i\Delta}^2 \sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j^2(X_{(i-1)\Delta})\right) = O_p\big(\mathbb{E}V_\Delta^2 \Delta^{-1} T \kappa_g^2 \kappa_{f_1}^2 \cdots \kappa_{f_k}^2(T)\big).$$

When $\mathbb{E}V_\Delta \neq 0$,

$$\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^{k} B_{ij}$$
$$= \mathbb{E}V_\Delta \sum_{i=1}^{n} g(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j(X_{(i-1)\Delta})$$
$$+ \sum_{i=1}^{n} g(X_{(i-1)\Delta}) \prod_{j=1}^{k} f_j(X_{(i-1)\Delta}) \left((W_{i\Delta} - W_{(i-1)\Delta})^p \prod_{j=1}^{k} B_{ij} - \mathbb{E}V_\Delta\right)$$

so the first term is of order $O_p(\mathbb{E}V_\Delta \Delta^{-1} T \kappa_g \kappa_{f_1} \cdots \kappa_{f_k}(T))$ and it's also easy to check the order of the second term is smaller than the first term by taking expectation of the square with the same steps as in the previous case. The order of the remainder term R can be obtained using Lemma 2 and the Schwartz inequality, but here, I will show a simple case as an example. For

$$\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} f_1(X_s) ds dW_t \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} f_2(X_s) ds dW_t$$

we can rewrite it as

$$
\begin{aligned}
&\sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_1(X_{(i-1)\Delta}) f_2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} dsdW_t \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} dsdW_t \\
&+\sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_1(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} dsdW_t \times \\
&\qquad\qquad \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} \left(f_2(X_s) - f_2(X_{(i-1)\Delta})\right) dsdW_t \\
&+\sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} dsdW_t \times \\
&\qquad\qquad \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} \left(f_1(X_s) - f_1(X_{(i-1)\Delta})\right) dsdW_t \\
&+\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} \left(f_1(X_s) - f_1(X_{(i-1)\Delta})\right) dsdW_t \times \\
&\qquad\qquad \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} \left(f_2(X_s) - f_2(X_{(i-1)\Delta})\right) dsdW_t
\end{aligned}
$$

For the first term,

$$
\sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_1(X_{(i-1)\Delta}) f_2(X_{(i-1)\Delta}) \left(\int_{(i-1)\Delta}^{i\Delta} tdW_t \int_{(i-1)\Delta}^{i\Delta} tdW_t - \frac{\Delta^3}{3} \right)
$$

the order can be obtained from the expectation of the square,

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) f_1^2(X_{(i-1)\Delta}) f_2^2(X_{(i-1)\Delta}) \mathbb{E}_{(i-1)\Delta} \left(\int_{(i-1)\Delta}^{i\Delta} tdW_t \int_{(i-1)\Delta}^{i\Delta} tdW_t - \frac{\Delta^3}{3} \right)^2 \right) \\
&= \mathbb{E}\left(\frac{2\Delta^6}{9} \sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) f_1^2(X_{(i-1)\Delta}) f_2^2(X_{(i-1)\Delta}) \right) \\
&= O_p(\Delta^5 T \kappa_g^2 \kappa_{f_1}^2 \kappa_{f_2}^2(T))
\end{aligned}
$$

while

$$
\frac{\Delta^3}{3} \sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_1(X_{(i-1)\Delta}) f_2(X_{(i-1)\Delta}) = O_p(\Delta^2 T \kappa_g \kappa_{f_1} \kappa_{f_2}(T))
$$

so this will become the leading term. This is guaranteed from the order of the remainder terms,

$$
\begin{aligned}
&\sum_{i=1}^{n} g(X_{(i-1)\Delta}) f_1(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} ds dW_t \times \\
&\qquad\qquad \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(f_2(X_s) - f_2(X_{(i-1)\Delta})\big) ds dW_t \\
&\leq \sqrt{\sum_{i=1}^{n} g^2(X_{(i-1)\Delta}) f_1^2(X_{(i-1)\Delta}) \left(\int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} ds dW_t\right)^2} \times \\
&\quad \sqrt{\sum_{i=1}^{n} \left(\int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(f_2(X_s) - f_2(X_{(i-1)\Delta})\big) ds dW_t\right)^2} \\
&= O_p(\Delta \sqrt{T} \kappa_g \kappa_{f_1}(T)) O_p(\Delta^{3/2} \sqrt{T} \kappa_\sigma \kappa_{f_2^{\cdot}}(T)) \\
&= O_p(\Delta^{5/2} T \kappa_g \kappa_{f_1} \kappa_\sigma \kappa_{f_2^{\cdot}}(T))
\end{aligned}
$$

Note that

$$
f_2(X_s) - f_2(X_{(i-1)\Delta}) = \int_{(i-1)\Delta}^{s} \left(\mu f^{\cdot} + \frac{f^{\cdot\cdot}}{2}\right)(X_t) dt + \int_{(i-1)\Delta}^{s} \sigma f^{\cdot}(X_t) dW_t
$$

and all terms can be taken care of by Lemma 2 and Schwartz inequality, and the same thing can be done to show

$$
\begin{aligned}
&\sum_{i=1}^{n} g(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(f_1(X_s) - f_1(X_{(i-1)\Delta})\big) ds dW_t \times \\
&\qquad\qquad \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(f_2(X_s) - f_2(X_{(i-1)\Delta})\big) ds dW_t \\
&= O_p(\Delta^3 T \kappa_\sigma^2 \kappa_g \kappa_{f_1^{\cdot}} \kappa_{f_2^{\cdot}}(T))
\end{aligned}
$$

□

Lemma 4. *If $Y_t = O_p(g(t))$ as $t \to \infty$ for some positive function g, then we have*

$$
\int_0^T Y_t dt = O_p\big(G(T)\big)
$$

as $T \to \infty$, where $G(x) = \int^x g(s)ds$.

Proof. The condition means that for any $\varepsilon > 0$, there exists M such that

$$\mathbf{P}\left\{\left|\frac{Y_t}{g(t)}\right| > M\right\} < \varepsilon \tag{A.4}$$

holds for all $t \geq t_0$. Note that

$$\left|\int_0^T Y_t dt\right| = \left|\int_0^{t_0} Y_t dt + \int_{t_0}^T Y_t dt\right| \leq \left|\int_0^{t_0} |Y_t| dt\right| + \left|\int_{t_0}^T |Y_t| dt\right| = A_T + B_T.$$

Since we can always find $\tilde{M}$ such that

$$\mathbf{P}\left\{\left|\int_0^{t_0} Y_t dt\right| > \tilde{M}\left|\int_0^{t_0} g(t)dt\right|\right\} < \varepsilon,$$

we have

$$A_T \leq \tilde{M}\int_0^{t_0} g(t)dt$$

with probability $1-\varepsilon$. Also, we have

$$B_T \leq M\int_{t_0}^T g(t)dt$$

with probability $1-\varepsilon$ from (A.4) and from these, we can find $\bar{M}$ that makes

$$\mathbf{P}\left\{\left|\int_0^T Y_t dt\right| > \bar{M}\left|\int_0^T g(t)dt\right|\right\} < \varepsilon,$$

which completes the proof. □

Lemma 5. *Let f satisfy Assumption 2.*

(a) If $k > 0$ is an odd number,

$$\sum_{i=1}^n f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k = O_p(\Delta^{(k-1)/2}\sqrt{T}\kappa_f(T))$$

(b) If $k > 0$ is an even number,

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k = O_p(\Delta^{(k-2)/2} T \kappa_f(T)).$$

Proof. (a)

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k\right)^2 \\
&\quad = \mathbb{E}\left(\sum_{i=1}^{n} f^2(X_{(i-1)\Delta})\mathbb{E}_{(i-1)\Delta}(W_{i\Delta} - W_{(i-1)\Delta})^{2k}\right)
\end{aligned}$$

since $(W_{i\Delta} - W_{(i-1)\Delta})^k$ and $(W_{j\Delta} - W_{(j-1)\Delta})^k$ are independent for $i \neq j$.

$$\mathbb{E}_{(i-1)\Delta}(W_{i\Delta} - W_{(i-1)\Delta})^{2k} = (2k-1)!!\Delta^k$$

so

$$\mathbb{E}\left((2k-1)!!\Delta^k \sum_{i=1}^{n} f^2(X_{(i-1)\Delta})\right) = O_p(\Delta^{k-1} T \kappa_f^2(T))$$

from Lemma 1. So

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k = O_p(\Delta^{(k-1)/2} \sqrt{T} \kappa_f(T))$$

(b) We rewrite the expression as

$$\begin{aligned}
\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k &= \Delta^{k/2}(k-1)!! \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \\
&\quad + \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big((W_{i\Delta} - W_{(i-1)\Delta})^k - \Delta^{k/2}(k-1)!!\big)
\end{aligned}$$

We can show

$$\Delta^{k/2}(k-1)!! \sum_{i=1}^{n} f(X_{(i-1)\Delta}) = O_p\big(\Delta^{(k-2)/2} T \kappa_f(T)\big)$$

from Lemma 1 and note that

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n} f(X_{(i-1)\Delta})\big((W_{i\Delta}-W_{(i-1)\Delta})^k-\Delta^{k/2}(k-1)!!\big)\right)^2 \\
&\quad=\mathbb{E}\left(\sum_{i=1}^{n} f^2(X_{(i-1)\Delta})\mathbb{E}_{(i-1)\Delta}\big[(W_{i\Delta}-W_{(i-1)\Delta})^k-\Delta^{k/2}(k-1)!!\big]^2\right)
\end{aligned}$$

Note that

$$\mathbb{E}_{(i-1)\Delta}\big[(W_{i\Delta}-W_{(i-1)\Delta})^k-\Delta^{k/2}(k-1)!!\big]^2=\big((2k-1)!!-((k-1)!!)^2\big)\Delta^k$$

so

$$\mathbb{E}\left(\big((2k-1)!!-((k-1)!!)^2\big)\Delta^k\sum_{i=1}^{n} f^2(X_{(i-1)\Delta})\right)=O_p(\Delta^{k-1}T\kappa_f^2(T))$$

from Lemma 1. So the second term is of order $O_p(\Delta^{(k-1)/2}\sqrt{T}\kappa_f(T))$ which is smaller than the first term. □

Lemma 6. *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 $t\in[(j-1)\Delta, j\Delta)$, $j=1,\ldots,n+1$. Then

$$V^{\Delta}\to_p V$$

for a standard Brownian motion V independent of W, and

$$V_T^{\Delta}-V_T=O_p\big((\Delta T)^{1/4}\big).$$

Proof. Clearly, V^{Δ} is a continuous martingale with quadratic variation given by

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

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

We have

$$\begin{aligned}[V^\Delta]_t - t =& \frac{2}{\Delta} \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left[(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta) \right] ds \\ &+ \frac{2}{\Delta} \int_{(j-1)\Delta}^{t} \left[(W_s - W_{(j-1)\Delta})^2 - (s - (j-1)\Delta) \right] ds + O(\Delta) \end{aligned} \tag{A.5}$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \ldots, n+1$, uniformly in $t \in [0, T]$. Therefore, ignoring $O(\Delta)$ term in (A.17) that is unimportant, it follows that

$$\begin{aligned}\mathbb{E}\left([V^\Delta]_t - t\right)^2 =& \left(\frac{2}{\Delta}\right)^2 \sum_{i=1}^{j-1} \mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta} \left[(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta) \right] ds \right)^2 \\ &+ \left(\frac{2}{\Delta}\right) \mathbb{E}\left(\int_{(j-1)\Delta}^{t} \left[(W_s - W_{(j-1)\Delta})^2 - (s - (j-1)\Delta) \right] ds \right)^2 \end{aligned} \tag{A.6}$$

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

$$\begin{aligned}&\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta} \left[(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta) \right] ds \right)^2 \\ &\quad \le \Delta \int_{(i-1)\Delta}^{i\Delta} \mathbb{E}\left[(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta) \right]^2 ds = \frac{2\Delta^4}{3} \end{aligned} \tag{A.7}$$

for $i = 1, \ldots, n$. Moreover, we may deduce from (A.18) and (A.19) that

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

under our assumption. Consequently, it follows that

$$\sup_{0 \le t \le T} \mathbb{E}\left([V^\Delta]_t - t\right)^2 \to 0$$

in our asymptotic framework. This implies that

$$V^\Delta \to_p V,$$

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

$$\begin{aligned}
&\mathbb{E}[V^\Delta, W]_t^2 &\text{(A.8)}\\
&= \frac{2}{\Delta}\left[\sum_{i=1}^{j-1}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}(W_s - W_{(i-1)\Delta})ds\right)^2 + \mathbb{E}\left(\int_{(j-1)\Delta}^{t}(W_s - W_{(j-1)\Delta})ds\right)^2\right]
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}(W_s - W_{(i-1)\Delta})ds\right)^2 &\le \Delta\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}(W_s - W_{(i-1)\Delta})^2 ds\\
&= \Delta\int_{(i-1)\Delta}^{i\Delta}(s-(i-1)\Delta)\,ds = \frac{\Delta^3}{2} \qquad \text{(A.9)}
\end{aligned}$$

for $i = 1, \ldots, n$. Therefore, it can be deduced from (A.20) and (A.21) that

$$\begin{aligned}
\mathbb{E}[V^\Delta, W]_t^2 &\le \frac{2}{\Delta}\sum_{i=1}^{n}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}(W_s - W_{(i-1)\Delta})ds\right)^2\\
\frac{2}{\Delta}n\frac{\Delta^3}{2} &= \Delta T \to 0,
\end{aligned}$$

and that

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

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

For the second statement, note that V_t^Δ can be represented as a time changed

Brownian motion $V_{[V^\Delta]_t}$ from the DDS representation. Thus we have

$$\frac{V_{[V^\Delta]_t} - V_t}{\sqrt{\left|[V^\Delta]_t - t\right|}}\sqrt{\left|[V^\Delta]_t - t\right|} = O_p(1)\sqrt{\left|O_p(\sqrt{\Delta T})\right|} = O_p\big((\Delta T)^{1/4}\big).$$

$\square$

Lemma 7. *Let f be a two times differentiable function and let f and its derivatives satisfy Assumption 2. Then*

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) = \int_0^T f(X_t)dW_t + O_p\big(\sqrt{\Delta T}\kappa_\sigma\kappa_{f^\cdot}(T)\big)$$

and

$$\begin{aligned}
&\sqrt{\frac{2}{\Delta}}\sum_{i=1}^{n} f(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{s} dW_u dW_s \\
&\qquad = \int_0^T f(X_t)dV_t + O_p\big(\Delta^{1/4}T^{1/4}\kappa_f(T)\big) \\
&\qquad\qquad + O_p\big(\Delta^{1/4}T^{3/4}\kappa_\sigma^2\kappa_{f^\cdot}^2(T)\big) \\
&\qquad\qquad + O_p(\Delta^{1/4}T^{5/4}(\kappa_\mu\kappa_{f^\cdot} + \kappa_\sigma^2\kappa_{f^{\cdot\cdot}} + \sqrt{\kappa_\mu\kappa_{f^\cdot}\kappa_{f^{\cdot\cdot}}}\kappa_\sigma)(T)),
\end{aligned}$$

where V is as defined in Lemma 16.

Proof. For the first statement,

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

The last line is due to Lemma 3.

For the second statement,

$$
\begin{aligned}
&\sqrt{\frac{2}{\Delta}} \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} dW_u dW_s \\
&\quad = \int_0^T f(X_t)dV_t + \int_0^T f(X_t)d(V^\Delta - V)_t \\
&\qquad - \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \big(f(X_t) - f(X_{(i-1)\Delta})\big)dV_t^\Delta \\
&\quad = \int_0^T f(X_t)dV_t + A_T + B_T.
\end{aligned}
$$

We will show the order of A_T in Part 1, and in Part 2, the order of B_T.

Part 1. For A_T, note that

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

from integration by parts exploiting the notation for the quadratic covariation term. The orders of the first two terms can be easily obtained from the order of $f(X_T)$,

$V_T^\Delta - V_T$, and from Lemma 4. For the first term,

$$f(X_T)(V_T^\Delta - V_T) = O_p(\kappa_f(T))O_p\big((\Delta T)^{1/4}\big) = O_p\big((\Delta T)^{1/4}\kappa_f(T)\big)$$

and for the second term,

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

We have

$$\begin{aligned}C_T &\le \sqrt{\int_0^T (V_t^\Delta - V_t)^2 dt \int_0^T \left(\mu f^{\cdot} + \frac{\sigma^2 f^{\cdot\cdot}}{2}\right)^2 (X_t)dt} \\ &= O_p(\Delta^{1/4}T^{3/4})\big[O_p(\sqrt{T}\kappa_\mu\kappa_{f^{\cdot}}(T)) + O_p(\sqrt{T}\kappa_\sigma^2\kappa_{f^{\cdot\cdot}}(T)) \\ &\qquad + O_p(\sqrt{T\kappa_\mu\kappa_{f^{\cdot}}\kappa_{f^{\cdot\cdot}}}\kappa_\sigma(T))\big] \\ &= O_p(\Delta^{1/4}T^{5/4}\kappa_\mu\kappa_{f^{\cdot}}(T)) + O_p(\Delta^{1/4}T^{5/4}\kappa_\sigma^2\kappa_{f^{\cdot\cdot}}(T)) \\ &\quad + O_p(\Delta^{1/4}T^{5/4}\sqrt{\kappa_\mu\kappa_{f^{\cdot}}\kappa_{f^{\cdot\cdot}}}\kappa_\sigma(T))\end{aligned}$$

from Lemma 4, and D_T is a martingale with a quadratic variation

$$\begin{aligned}\int_0^T (V_t^\Delta - V_t)^2\sigma^2 f^{\cdot 2}(X_t)dt &\le \sqrt{\int_0^T (V_t^\Delta - V_t)^4 dt \int_0^T \sigma^4 f^{\cdot 4}(X_t)dt} \\ &= O_p(\Delta^{1/2-2\zeta}T^{1-2\zeta})O_p(\sqrt{T}\kappa_\sigma^4\kappa_{f^{\cdot}}^4(T))\end{aligned}$$

from Lemma 4 also, so

$$D_T = O_p\big(\Delta^{1/4}T^{3/4}\kappa_\sigma^2\kappa_{f^{\cdot}}^2(T)\big).$$

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 same as the quadratic covariation of

$$\int_0^t\sigma f^{\cdot}(X_s)dW_s$$

and

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

as in the definition of V_t^{Δ}. So we have

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

To obtain its order, note that

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}f(X_s)\int_{(i-1)\Delta}^{s}dW_u ds\\
&\quad=\sum_{i=1}^{n}f(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{s}dW_u ds\\
&\qquad+\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big(f(X_s)-f(X_{(i-1)\Delta})\big)\int_{(i-1)\Delta}^{s}dW_u ds\\
&\quad=A_{1T}+A_{2T}.
\end{aligned}$$

We have

$$A_{1T}=\sum_{i=1}^{n}f(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta-u)dW_u$$

and this is a martingale with a quadratic variation bounded by

$$\Delta^2 \sum_{i=1}^{n} f^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} du = \Delta^3 \sum_{i=1}^{n} f^2(X_{(i-1)\Delta}) = O_p(\Delta^2 T \kappa_f^2(T))$$

from Lemma 1. For A_{2T},

$$\begin{aligned} A_{2T} =& \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \left(\mu f^{\cdot} + \frac{\sigma^2 f^{\cdot\cdot}}{2}\right)(X_u) du \int_{(i-1)\Delta}^{s} dW_u ds \\ &+ \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \sigma f^{\cdot}(X_u) dW_u \int_{(i-1)\Delta}^{s} dW_u ds \\ =& A_{21T} + A_{22T}. \end{aligned}$$

and

$$\begin{aligned} A_{21T} \leq & \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \left|\mu f^{\cdot} + \frac{\sigma^2 f^{\cdot\cdot}}{2}\right| (X_u) du ds \\ =& O_p(\Delta T \kappa_\mu \kappa_{f^{\cdot}}(T)) + O_p(\Delta T \kappa_\sigma^2 \kappa_{f^{\cdot\cdot}}(T)) \end{aligned}$$

from Lemma 3. For A_{22T},

$$A_{22T} \leq \sqrt{\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} \sigma f^{\cdot}(X_u) dW_u\right)^2 ds \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} dW_u\right)^2 ds}.$$

Note that

$$\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} \sigma f^{\cdot}(X_u) dW_u\right)^2 ds = O_p(\Delta T^2 \kappa_\sigma^2 \kappa_{f^{\cdot}}^2(T))$$

since

$$\int_{(i-1)\Delta}^{s} \sigma f^{\cdot}(X_u) dW_u = O_p(\sqrt{\Delta T} \kappa_\sigma \kappa_{f^{\cdot}}(T)),$$

so

$$A_{22T} = O_p\big(\sqrt{\Delta} T \kappa_\sigma \kappa_{f^{\cdot}}(T)\big) O_p\big(\sqrt{\Delta T}\big) = O_p(\Delta T^{3/2} \kappa_\sigma \kappa_{f^{\cdot}}(T)),$$

and the order of quadratic covariation becomes

$$
\begin{aligned}
[f(X),&(V^{\Delta} - V)]_T = O_p\big(\sqrt{\Delta}\sqrt{T}\kappa_{\sigma}\kappa_{f^{\cdot}}(T)\big) + O_p\big(\sqrt{\Delta}T\kappa_{\mu}\kappa_{\sigma}\kappa_{f^{\cdot\cdot}}(T)\big) \\
&+ O_p\big(\sqrt{\Delta}T\kappa_{\mu}\kappa_{\sigma^{\cdot}}\kappa_{f^{\cdot}}(T)\big) + O_p\big(\sqrt{\Delta}T\kappa_{\sigma}^2\kappa_{\sigma^{\cdot}}\kappa_{f^{\cdot\cdot}}(T)\big) + O_p\big(\sqrt{\Delta}T\kappa_{\sigma}^3\kappa_{f^{\cdot\cdot\cdot}}(T)\big) \\
&+ O_p\big(\sqrt{\Delta}T\kappa_{\sigma}^2\kappa_{\sigma^{\cdot\cdot}}\kappa_{f^{\cdot}}(T)\big) + O_p\big(\sqrt{\Delta}T\kappa_{\sigma}^2\kappa_{f^{\cdot\cdot}}(T)\big) + O_p\big(\sqrt{\Delta}T\kappa_{\sigma}\kappa_{\sigma^{\cdot}}\kappa_{f^{\cdot}}(T)\big) \\
&+ O_p\big(\sqrt{\Delta}\sqrt{T\kappa_{\sigma}^2\kappa_{f^{\cdot\cdot}}}(T)\big) + O_p\big(\sqrt{\Delta}\sqrt{T\kappa_{\sigma}\kappa_{\sigma^{\cdot}}\kappa_{f^{\cdot}}}(T)\big).
\end{aligned}
$$

Since these are all of smaller order than the previous, we have

$$
\begin{aligned}
A_T = &O_p\big(\Delta^{1/4}T^{1/4}\kappa_f(T)\big) + O_p(\Delta^{1/4}T^{5/4}\kappa_{\mu}\kappa_{f^{\cdot}}(T)) + O_p(\Delta^{1/4}T^{5/4}\kappa_{\sigma}^2\kappa_{f^{\cdot\cdot}}(T)) \\
&+ O_p(\Delta^{1/4}T^{5/4}\sqrt{\kappa_{\mu}\kappa_{f^{\cdot}}\kappa_{f^{\cdot\cdot}}}\kappa_{\sigma}(T)) + O_p\big(\Delta^{1/4}T^{3/4}\kappa_{\sigma}^2\kappa_{f^{\cdot}}^2(T)\big)
\end{aligned}
$$

as a result.

Part 2. For B_T,

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

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

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

is a martingale with a quadratic variation

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

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

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big([V^\Delta]_{i\Delta}-[V^\Delta]_s\big)^2ds\right) \\
&=\mathbb{E}\left(\frac{4}{\Delta^2}\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left(\int_s^{i\Delta}(W_u-W_s)^2du\right)^2ds\right) \qquad \text{(A.10)}\\
&=\mathbb{E}\left(\frac{4}{\Delta^2}\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}_{(i-1)\Delta}\left(\int_s^{i\Delta}(W_u-W_s)^2du\right)^2ds\right).
\end{aligned}
$$

and since

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

we have

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

and

$$B_{11T} = O_p(\Delta\sqrt{T}).$$

For B_{12T},

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

so

$$B_{1T} = O_p(\Delta\sqrt{T})O_p(\Delta\sqrt{T}\kappa_f^2(T)) = O_p(\Delta\sqrt{T}\kappa_f(T)).$$

For B_{2T}, note that

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

is a martingale with a quadratic variation

$$\begin{aligned}&\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left(\int_{(i-1)\Delta}^{t} f(X_s)dW_s\right)^2 d[V^{\Delta}]_t \\ &=\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f^2(X_s)dsd[V^{\Delta}]_t \\ &\quad + 2\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f(X_s)\int_{(i-1)\Delta}^{s} f(X_u)dW_u dW_s d[V^{\Delta}]_t \\ &=B_{21T} + 2B_{22T}.\end{aligned}$$

For B_{21T},

$$\begin{aligned} B_{21T} &= \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left([V^\Delta]_{i\Delta} - [V^\Delta]_s\right) f^2(X_s) ds \\ &\le \sqrt{\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left([V^\Delta]_{i\Delta} - [V^\Delta]_s\right)^2 ds \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} f^4(X_s) ds}. \end{aligned}$$

so

$$B_{21T} = O_p(\Delta\sqrt{T}) O_p(\sqrt{T}\kappa_f^2(T)) = O_p(\Delta T \kappa_f^2(T)).$$

For B_{22T},

$$B_{22T} = \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left([V^\Delta]_{i\Delta} - [V^\Delta]_s\right) f(X_s) \int_{(i-1)\Delta}^{s} f(X_u) dW_u dW_s$$

and this is a martingale with a quadratic variation

$$\begin{aligned} &\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left([V^\Delta]_{i\Delta} - [V^\Delta]_s\right)^2 f^2(X_s) \left(\int_{(i-1)\Delta}^{s} f(X_u) dW_u\right)^2 ds \\ &\le \sqrt{\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left([V^\Delta]_{i\Delta} - [V^\Delta]_s\right)^4 ds \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} f^4(X_s) \left(\int_{(i-1)\Delta}^{s} f(X_u) dW_u\right)^4 ds}. \end{aligned}$$

Note that

$$\int_{(i-1)\Delta}^{s} f(X_u) dW_u = O_p(\sqrt{\Delta T}\kappa_f(T))$$

since it's a martingale with a quadratic variation

$$\int_{(i-1)\Delta}^{s} f^2(X_u) du = O_p(\Delta T \kappa_f^2(T)),$$

and since

$$f(X_s) = O_p(\kappa_f(T)),$$

we have

$$\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta} f^4(X_s)\left(\int_{(i-1)\Delta}^{s} f(X_u)dW_u\right)^4 ds = O_p(\Delta^2 T^3 \kappa_f^8(T))$$

and

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

since

$$\begin{aligned}\mathbb{E}_{(i-1)\Delta}\left(\int_s^{i\Delta}(W_u - W_s)^2 du\right)^4 &\leq (i\Delta - s)^2 \mathbb{E}_{(i-1)\Delta}\left(\int_s^{i\Delta}(W_u - W_s)^4 du\right)^2 \\ &\leq (i\Delta - s)^3 \int_s^{i\Delta} \mathbb{E}_{(i-1)\Delta}(W_u - W_s)^8 du \\ &= 21(i\Delta - s)^8\end{aligned}$$

with the same way as in (A.22). So

$$B_{22T} = O_p(\Delta\sqrt{T}\kappa_f(T))$$

and we can check that B_T has a smaller order than A_T. $\square$

Lemma 8. *For a positive integer k and for a four times differentiable function f, let f and its derivatives satisfy Assumption 2. Then*

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^k = O_p\big(\Delta^{(k-1)/2}\sqrt{T}\kappa_\sigma^k \kappa_f(T)\big)$$

when k is odd and

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^k = O_p\big(\Delta^{(k-2)/2} T\kappa_\sigma^k \kappa_f(T)\big)$$

when k is even.

Proof. Note that, by Ito's Lemma,

$$
\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} =& \int_{(i-1)\Delta}^{i\Delta} \mu(X_t)dt + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t)dW_t \\
=& \Delta\mu(X_{(i-1)\Delta}) + \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
& + \int_{(i-1)\Delta}^{i\Delta} (\mu(X_t) - \mu(X_{(i-1)\Delta}))dt \\
& + \int_{(i-1)\Delta}^{i\Delta} (\sigma(X_t) - \sigma(X_{(i-1)\Delta}))dW_t
\end{aligned}
$$

and

$$
\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}) =& \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
& + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} (\mu\mu^{\cdot} + \frac{\sigma^2\mu^{\cdot\cdot}}{2})(X_s)dsdt \\
& + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\mu^{\cdot}(X_s)dW_s dt \\
& + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} (\mu\sigma^{\cdot} + \frac{\sigma^2\sigma^{\cdot\cdot}}{2})(X_s)dsdW_t \\
& + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\sigma^{\cdot}(X_s)dW_s dW_t.
\end{aligned}
$$

so we have

$$
\begin{aligned}
& \sum_{i=1}^{n} f(X_{(i-1)\Delta})(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^k \\
&= \sum_{i=1}^{n} f(X_{(i-1)\Delta})(\sigma^k(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k + R_i)
\end{aligned}
$$

where R_i are the cross products of each terms and

$$
\sum_{i=1}^{n} f(X_{(i-1)\Delta})\sigma^k(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k = O_p\big(\Delta^{(k-1)/2}\sqrt{T}\kappa_\sigma^k\kappa_f(T)\big)
$$

when k is odd and

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})\sigma^k(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^k = O_p\big(\Delta^{(k-2)/2}T\kappa_\sigma^k\kappa_f(T)\big)$$

when k is even from Lemma 5. And also,

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})R_i$$

can be dealt with Lemma 3 and can be shown to be of smaller order. □

Lemma 9.

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))$$
$$= \int_0^T \sigma f(X_t)dW_t + O_p\big(\sqrt{\Delta T}\kappa_\sigma\kappa_{\sigma'}\kappa_f(T)\big) + O_p\big(\sqrt{\Delta T}\kappa_\sigma^2\kappa_{f'}(T)\big)$$

Proof. The proof is same as the proof of Lemma 8 only with a difference that $k = 1$ and

$$\sum_{i=1}^{n} f\sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) = \int_0^T \sigma f(X_t)dW_t + O_p\big(\sqrt{\Delta T}\kappa_\sigma\kappa_{\sigma'}\kappa_f(T)\big)$$
$$+ O_p\big(\sqrt{\Delta T}\kappa_\sigma^2\kappa_{f'}(T)\big)$$

from Lemma 15 □

Lemma 10. *For a four times differentiable function f, let f and its derivatives satisfy Assumption 2. Then*

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^2 - \Delta\sigma^2(X_{(i-1)\Delta})\big]$$
$$= \sqrt{2\Delta}\int_0^T \sigma^2 f(X_t)dV_t + O_p(\Delta^{1/4}F_f(T))$$

for any $\zeta > 0$, where $F_f(T)$ is a function of T as defined as in (A.11) and (A.12) according to f, and V is a standard Brownian motion independent of W.

Proof. Denoting $V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta} = \int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} dW_s dW_t$, we can write as

$$\begin{aligned}
&X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}) \\
&\qquad = \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + \sigma\sigma^{\cdot}(X_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta}) + R_i,
\end{aligned}$$

where R_i is a remainer term, from the first equation of the proof of Lemma 8. Replacing this into the following, we have

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^2 - \Delta\sigma^2(X_{(i-1)\Delta})\big] \\
&\quad = \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[(\sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + \sigma\sigma^{\cdot}(X_{(i-1)\Delta})(V_{i\Delta} - V_{(i-1)\Delta}) + R_i)^2 \\
&\qquad\qquad - \Delta\sigma^2(X_{(i-1)\Delta})\big] \\
&\quad = \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[\sigma^2(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^2 \\
&\qquad\qquad + 2\sigma^2\sigma^{\cdot}(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta}) \\
&\qquad\qquad + \sigma^2\sigma^{\cdot 2}(X_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta})^2 + R_i^* \\
&\qquad\qquad - \Delta\sigma^2(X_{(i-1)\Delta})\big],
\end{aligned}$$

where R_i^* denotes the terms multiplied by R_i. From Lemma 15,

$$\begin{aligned}
\sum_{i=1}^{n} f\sigma^2(X_{(i-1)\Delta})[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta] &= \sqrt{2\Delta}\int_0^T f\sigma^2(X_t)dV_t \\
&\quad + O_p(\Delta^{3/4}T^{1/4}\kappa_f\kappa_\sigma^2(T)) \qquad \text{(A.11)} \\
&\quad + O_p(\Delta^{3/4}T^{3/4}\kappa_\sigma^4(\kappa_\sigma\kappa_{f^{\cdot}} + 2\kappa_f\kappa_{\sigma^{\cdot}})^2(T)) \\
&\quad + O_p(\Delta^{3/4}T^{5/4}F(T))
\end{aligned}$$

where

$$F(T) = \kappa_\sigma \Big(\kappa_{f^{\cdot\cdot}} \kappa_\sigma^3 + 2\left(2\kappa_{f^\cdot}\kappa_{\sigma^\cdot} + \kappa_f \kappa_{\sigma^{\cdot\cdot}}\right)\kappa_\sigma^2 + 2\kappa_f \kappa_{\sigma^\cdot}^2 \kappa_\sigma + \kappa_\mu \left(\kappa_\sigma \kappa_{f^\cdot} + 2\kappa_f \kappa_{\sigma^\cdot}\right) \tag{A.12}$$
$$+ \left(\kappa_\mu \kappa_\sigma \left(\kappa_\sigma \kappa_{f^\cdot} + 2\kappa_f \kappa_{\sigma^\cdot}\right)\left(\kappa_{f^{\cdot\cdot}}\kappa_\sigma^2 + 4\kappa_{f^\cdot}\kappa_{\sigma^\cdot}\kappa_\sigma + 2\kappa_f \kappa_{\sigma^{\cdot\cdot}}\kappa_\sigma + 2\kappa_f \kappa_{\sigma^\cdot}^2\right)\right)^{1/2}\Big)(T)$$

and

$$\sum_{i=1}^n f\sigma^2\sigma^\cdot(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})(V_{i\Delta}^\Delta - V_{(i-1)\Delta}^\Delta) = O_p(\Delta T \kappa_f \kappa_\sigma^2 \kappa_{\sigma^\cdot}(T))$$
$$\sum_{i=1}^n f\sigma^2\sigma^{\cdot 2}(X_{(i-1)\Delta})(V_{i\Delta}^\Delta - V_{(i-1)\Delta}^\Delta)^2 = O_p(\Delta T \kappa_f \kappa_\sigma^2 \kappa_{\sigma^\cdot}^2(T))$$

by the same steps in the proof of Lemma 3 from the independent increments of the Brownian motion and $\mathbb{E}\big((W_{i\Delta} - W_{(i-1)\Delta})(V_{i\Delta}^\Delta - V_{(i-1)\Delta}^\Delta)\big) = 0$, and the remainder term $\sum_{i=1}^n f(X_{(i-1)\Delta})R_i^*$ can be also shown to be of smaller order by Lemma 3. □

Lemma 11. *Let $\{Z_i\}$ be a sequence of random variables. Denoting $x_i = X_{(i-1)\Delta}$ and $y_i = X_{i\Delta}$, we have*

$$\sum_{i=1}^n \left(\frac{e^+(x_i,y_i) - e^-(x_i,y_i)}{e^+(x_i,y_i) + e^-(x_i,y_i)}\right)^p Z_i = \sum_{i=1}^n Z_i + O_p\left(\sqrt{\frac{T\nu}{\Delta \exp(\Delta^{-1})}}\right)$$

for a finite integer $p > 0$, as $\Delta \to 0$ and $T \to \infty$, where ν is a sequence satisfying $\sum_{i=1}^n Z_i^2 = O_p(\nu)$.

Proof. By expanding the terms and arranging them, we have

$$\sum_{i=1}^n \left(\frac{e^+(x_i,y_i) - e^-(x_i,y_i)}{e^+(x_i,y_i) + e^-(x_i,y_i)}\right)^p Z_i = \sum_{i=1}^n \left(1 - \frac{2e^-(x_i,y_i)}{e^+(x_i,y_i) + e^-(x_i,y_i)}\right)^p Z_i$$
$$= \sum_{i=1}^n Z_i + \sum_{i=1}^n \left(\sum_{j=1}^p C_{p,j}\left(\frac{2e^-(x_i,y_i)}{e^+(x_i,y_i) + e^-(x_i,y_i)}\right)^j\right) Z_i$$

where $C_{p,j} = \binom{c_p^j}{p}$ with $c_p^j = \min(j-1, p-j)$. For each term in the remainder,

$$\begin{aligned}
\sum_{i=1}^{n} C_{p,j} \left(\frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)+e^-(x_i,y_i)} \right)^j Z_i &\le \sqrt{\sum_{i=1}^{n} C_{p,j}^2 \left(\frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)+e^-(x_i,y_i)} \right)^{2j} \sum_{i=1}^{n} Z_i^2} \\
&= \sqrt{O_p(n\exp(-\Delta^{-1}))O_p(\nu)} \\
&= O_p\left(\sqrt{\frac{T\nu}{\Delta\exp(\Delta^{-1})}} \right)
\end{aligned}$$

since

$$\left(\frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)+e^-(x_i,y_i)} \right)^{2j} = O_p\big(\exp(-\Delta^{-1})\big). \tag{A.13}$$

for each i. To show (A.13), we will first obtain the order of $X_{i\Delta} - X_{(i-1)\Delta}$ in Part 1, and will prove (A.13) by finding out the order of $e^-(x_i,y_i)$ and $e^-(x_i,y_i)$ in Part 2. Hereafter, we will explicitly denote the arguments for $e^+(x_i,y_i)$ and $e^-(x_i,y_i)$ such as

$$\begin{aligned}
&e^+(X_{(i-1)\Delta}, X_{i\Delta}) \qquad \text{(A.14)} \\
&e^-(X_{(i-1)\Delta}, X_{i\Delta}).
\end{aligned}$$

Part 1. By Itô's lemma,

$$\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} &= \Delta\mu(X_{(i-1)\Delta}) + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \left(\mu\mu^{\cdot} + \frac{\sigma^2\mu^{\cdot\cdot}}{2} \right)(X_s)ds dt \\
&\qquad + \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\mu^{\cdot}(X_s)dW_s dt + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t)dW_t \\
&= \Delta\mu(X_{(i-1)\Delta}) + A_T + B_T + C_T.
\end{aligned}$$

For A_T, note that

$$\int_{(i-1)\Delta}^{i\Delta} f(X_s)ds \le \Delta \sup_{s\in[(i-1)\Delta, i\Delta]} \big|f(X_s)\big| \le \Delta \sup_{s\in[0,T]} \big|f(X_s)\big|$$

and

$$\sup_{t\in[0,T]} \left|f(X_t)\right| = O_p(\kappa_f(T)).$$

Thus we have

$$\begin{aligned} A_T &= \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \left(\mu\mu^{\cdot} + \frac{\sigma^2 \mu^{\cdot\cdot}}{2} \right) (X_s) ds \\ &\le \Delta \int_{(i-1)\Delta}^{i\Delta} \left| \mu\mu^{\cdot} + \frac{\sigma^2 \mu^{\cdot\cdot}}{2} \right| (X_s) ds \\ &= O_p(\Delta^2 \kappa_\mu \kappa_{\mu^{\cdot}}(T)) + O_p(\Delta^2 \kappa_\sigma^2 \kappa_{\mu^{\cdot\cdot}}(T)). \end{aligned}$$

Also,

$$B_T = \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s) \sigma \mu^{\cdot}(X_s) dW_s$$

is a martingale whose quadratic variation is

$$[B]_T = \int_{(i-1)\Delta}^{i\Delta} (i\Delta - s)^2 \sigma^2 \mu^{\cdot 2}(X_s) ds = O_p\big(\Delta^3 \kappa_\sigma^2 \kappa_{\mu^{\cdot}}^2(T)\big),$$

so we have

$$B_T = O_p\big(\Delta^{3/2} \kappa_\sigma \kappa_{\mu^{\cdot}}(T)\big).$$

Also, C_T is a martingale whose quadratic variation is

$$[C]_T = \int_{(i-1)\Delta}^{i\Delta} \sigma^2(X_t) dt = O_p(\Delta \kappa_\sigma^2(T)),$$

so we have

$$C_T = O_p\big(\Delta^{1/2} \kappa_\sigma(T)\big).$$

Combining these results, we have

$$\begin{aligned} X_{i\Delta} - X_{(i-1)\Delta} =& \Delta\mu(X_{(i-1)\Delta}) + O_p(\Delta^2\kappa_\mu\kappa_{\mu^\cdot}(T)) + O_p(\Delta^2\kappa_\sigma^2\kappa_{\mu^{\cdot\cdot}}(T)) \\ & + O_p\big(\Delta^{3/2}\kappa_\sigma\kappa_{\mu^\cdot}(T)\big) + O_p\big(\Delta^{1/2}\kappa_\sigma(T)\big) \\ =& O_p\big(\Delta^{1/2}\kappa_\sigma(T)\big). \end{aligned}$$

Note that we have

$$f(X_{(i-1)\Delta}) \leq \sup_{t\in[0,T]} \big|f(X_t)\big| = O_p(\kappa_f(T)) \tag{A.15}$$

for each i.

Part 2. From the order results in Part 1 and (A.15), we know that

$$\frac{1}{\Delta\sigma^{\cdot 2}(X_{(i-1)\Delta})}$$

is the biggest order term in (A.14), so we have

$$e^+(X_{(i-1)\Delta}, X_{i\Delta}) = O_p\left(\exp(\Delta^{-1}\kappa^2_{\sigma^{\cdot -1}}(T))\right)$$

and

$$e^-(X_{(i-1)\Delta}, X_{i\Delta}) = O_p\left(\exp(-\Delta^{-1}\kappa^2_{\sigma^{\cdot -1}}(T))\right).$$

From $e^-(x_i, y_i)/\exp(-\Delta^{-1}) \to_p 0$, we have

$$e^-(x_i, y_i) = O_p(\exp(-\Delta^{-1})),$$

and with this order, we can show by CMT,

$$\begin{aligned}\frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)+e^-(x_i,y_i)} &= \frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)+o_p(1)}\\ &= \frac{2e^-(x_i,y_i)}{e^+(x_i,y_i)}\big(1+o_p(1)\big)\\ &= O_p(\exp(-\Delta^{-1})).\end{aligned}$$

□

Lemma 12. *Let f be a function of a form of,*

$$\begin{aligned}f(X_{(i-1)\Delta}, X_{i\Delta}, \Delta) = \big(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta})\big)^v d(X_{(i-1)\Delta})\\ \big[\sigma(X_{(i-1)\Delta}) + \Delta b(X_{(i-1)\Delta}) + \big(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta})\big)c(X_{(i-1)\Delta})\big]^u\end{aligned}$$

with $v = 0, 1, 2$ and a real number $u < 2$, for $b(x)$, $c(x)$ and $d(x)$ being four times differentiable functions. These functions and their derivatives satisfy Assumption 2. Let

$$f^*(X_{(i-1)\Delta}) = d(X_{(i-1)\Delta})\sigma(X_{(i-1)\Delta})^u\big(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta})\big)^v$$

If we have a decreasing sequence μ_T satisfying

$$\mu_T\Delta^{3/2}\sum_{i=1}^{n} f^*(X_{(i-1)\Delta}) \to_p A$$

for some A as $T \to \infty$ and $\Delta \to 0$, we also have

$$\mu_T\Delta^{3/2}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}, \Delta) \to_p A$$

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

Proof. Note that

$$f(X_{(i-1)\Delta}, X_{i\Delta}, \Delta) = (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta})\sigma(X_{(i-1)\Delta})^u + R$$

where R is the remainder term which is

$$
\begin{aligned}
R \leq\ & (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta}) \\
& \quad u\,\sigma(X_{(i-1)\Delta})^{u-1}\left(\Delta b(X_{(i-1)\Delta}) + (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))c(X_{(i-1)\Delta})\right) \\
& + |(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))|^v |d(X_{(i-1)\Delta})| \\
& \quad A_{sup}\left(\Delta b(X_{(i-1)\Delta}) + (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))c(X_{(i-1)\Delta})\right)^2
\end{aligned}
$$

where

$$
\begin{aligned}
A_{sup} = & \sup_{t\in[0,T]}\left|\frac{u(u-1)}{2}\left(\sigma(X_t) + \Delta b(X_t) + (X_{t+\Delta} - X_t - \Delta\mu(X_t))c(X_t)\right)^{u-2}\right| \\
& + \sup_{t\in[0,T]}\left|\frac{u(u-1)}{2}\sigma^{u-2}(X_t)\right|
\end{aligned}
$$

since a power function is monotonic. Thus,

$$
\begin{aligned}
\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}, \Delta) \ & = \sum_{i=1}^{n}(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta})\sigma(X_{(i-1)\Delta})^u \\
& + \Sigma R
\end{aligned}
$$

where ΣR is the sum of the remainder terms, such that

$$
\begin{aligned}
\Sigma R \leq\ & \sum_{i=1}^{n}(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta}) \\
& \quad u\,\sigma(X_{(i-1)\Delta})^{u-1}\left(\Delta b(X_{(i-1)\Delta}) + (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))c(X_{(i-1)\Delta})\right) \\
& + A_{sup}\sum_{i=1}^{n} |(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))|^v |d(X_{(i-1)\Delta})| \\
& \quad \left(\Delta b(X_{(i-1)\Delta}) + (X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))c(X_{(i-1)\Delta})\right)^2 \\
& = A_T + B_T
\end{aligned}
$$

It is easy to check

$$A_T = \begin{cases} O_p(\Delta^{v/2}\sqrt{T}\kappa_\sigma^{v+1}\kappa_d\kappa_{\sigma^{-1}}^{1-u}\kappa_c(T)) & \text{if } u < 1 \\ O_p(\Delta^{v/2}\sqrt{T}\kappa_\sigma^{v+u}\kappa_d\kappa_c(T)) & \text{if } u \geq 1 \end{cases}$$

when $v+1$ is odd and

$$A_T = \begin{cases} O_p(\Delta^{(v-1)/2}T\kappa_\sigma^{v+1}\kappa_d\kappa_{\sigma^{-1}}^{1-u}\kappa_c(T)) & \text{if } u < 1 \\ O_p(\Delta^{(v-1)/2}T\kappa_\sigma^{v+u}\kappa_d\kappa_c(T)) & \text{if } u \geq 1 \end{cases}$$

when $v+1$ is even from Lemma 8, and for B_T,

$$B_T = \begin{cases} O_p(\Delta^{v/2}T\kappa_\sigma^{v+2}\kappa_{\sigma^{-1}}^{2-u}\kappa_d\kappa_c^2(T)) & \text{if } u < 2 \\ O_p(\Delta^{v/2}T\kappa_\sigma^{v+u}\kappa_d\kappa_c^2(T)) & \text{if } u \geq 2 \end{cases}$$

following the same steps in the proof of Lemma 11.

On the other hand,

$$\sum_{i=1}^{n}(X_{i\Delta}-X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta})\sigma(X_{(i-1)\Delta})^u$$
$$= \begin{cases} O_p(\Delta^{(v-1)/2}\sqrt{T}\kappa_\sigma^{v}\kappa_d\kappa_{\sigma^{-1}}^{-u}(T)) & \text{if } u < 0 \\ O_p(\Delta^{(v-1)/2}\sqrt{T}\kappa_\sigma^{v+u}\kappa_d(T)) & \text{if } u \geq 0 \end{cases}$$

when v is odd and

$$\sum_{i=1}^{n}(X_{i\Delta}-X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^v d(X_{(i-1)\Delta})\sigma(X_{(i-1)\Delta})^u$$
$$= \begin{cases} O_p(\Delta^{(v-2)/2}T\kappa_\sigma^{v}\kappa_d\kappa_{\sigma^{-1}}^{-u}(T)) & \text{if } u < 0 \\ O_p(\Delta^{(v-2)/2}T\kappa_\sigma^{v+u}\kappa_d(T)) & \text{if } u \geq 0 \end{cases}$$

when v is even. Under our condition, this becomes the leading term, which completes the proof. □

Lemma 13. *Let g be a power function $g(x) = x^p$ and f be a four times differentiable* *function. Also let f and its derivatives satisfy Assumption 2. Denoting*

$$D(X_{(i-1)\Delta}, X_{i\Delta}) = \Delta\sigma\sigma^{\cdot 2}(X_{(i-1)\Delta}) + 2\sigma^{\cdot}(X_{(i-1)\Delta})\big(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta})\big)$$

we have

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta}) g\big(\sigma(X_{(i-1)\Delta}) + D(X_{(i-1)\Delta}, X_{i\Delta})\big) \\
&\qquad = \sum_{i=1}^{n} f(X_{(i-1)\Delta})\Big[g(\sigma(X_{(i-1)\Delta})) + g'(\sigma(X_{(i-1)\Delta}))D(X_{(i-1)\Delta}, X_{i\Delta}) + \cdots \\
&\qquad\qquad + g^{(k)}(\sigma(X_{(i-1)\Delta}))D(X_{(i-1)\Delta}, X_{i\Delta})^{k}\Big] + O_p(R),
\end{aligned}$$

where $R = \Delta^{k/2}(\sqrt{T}\kappa_{\sigma}^{p}\kappa_f\kappa_{\sigma^{\cdot}}(T) + T\kappa_f\kappa_{\sigma}^{p}(T))$ when k is an even number, and $R = \Delta^{(k-1)/2}T\kappa_{\sigma}^{p}\kappa_f\kappa_{\sigma^{\cdot}}(T)$ when k is odd if $p \geq 0$. If $p < 0$, $R = \Delta^{k/2}(\sqrt{T}\kappa_{\sigma^{-1}}^{-p}\kappa_f\kappa_{\sigma^{\cdot}}(T) + T\kappa_f\kappa_{\sigma^{-1}}^{-p}(T))$ when k is even, and $R = \Delta^{(k-1)/2}T\kappa_{\sigma^{-1}}^{-p}\kappa_f\kappa_{\sigma^{\cdot}}(T)$ when k is odd.

Proof. Let's denote $\mathrm{D}_i = \mathrm{D}(X_{(i-1)\Delta}, X_{i\Delta})$ for the simplicity hereafter.

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta}) g\big(\sigma(X_{(i-1)\Delta}) + \mathrm{D}_i\big) \\
&= \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[g(\sigma(X_{(i-1)\Delta})) + g'(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i + \cdots + \frac{1}{k!}g^{(k)}(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i^{k}\big] \\
&\quad + R
\end{aligned}$$

where R is a remainder term which is

$$\begin{aligned}
R &\leq \sum_{i=1}^{n} \frac{1}{(k+1)!} f(X_{(i-1)\Delta}) g^{(k+1)}(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i^{k+1} + \sum_{i=1}^{n} \frac{G_{sup}}{(k+2)!}|f(X_{(i-1)\Delta})||\mathrm{D}_i|^{k+2} \\
&= A_T + B_T
\end{aligned}$$

where

$$G_{sup} = \sup_{t\in[0,T]} \Big| g^{(k+2)} \big[\sigma(X_t) + \Delta\sigma\sigma^{\cdot 2}(X_t) + 2\sigma^{\cdot}(X_t)(X_{t+\Delta} - X_t - \Delta\mu(X_t)) \big] \Big|$$
$$+ \sup_{t\in[0,T]} \Big| g^{(k+2)} \big[\sigma(X_t) \big] \Big|$$

since a power function is monotone. With the same steps as in the proof of Lemma 12, we can show

$$A_T = \begin{cases} O_p(\Delta^{k/2}\sqrt{T}\kappa_\sigma^p \kappa_f \kappa_{\sigma^\cdot}(T)) & \text{if } p > 0 \\ O_p(\Delta^{k/2}\sqrt{T}\kappa_{\sigma^{-1}}^{-p} \kappa_f \kappa_{\sigma^\cdot}(T)) & \text{if } p \leq 0 \end{cases}$$

when $k+1$ is odd and

$$A_T = \begin{cases} O_p(\Delta^{(k-1)/2} T\kappa_\sigma^p \kappa_f \kappa_{\sigma^\cdot}(T)) & \text{if } p > 0 \\ O_p(\Delta^{(k-1)/2} T\kappa_{\sigma^{-1}}^{-p} \kappa_f \kappa_{\sigma^\cdot}(T)) & \text{if } p \leq 0 \end{cases}$$

when $k+1$ is even. Also

$$B_T = \begin{cases} O_p(\Delta^{k/2} T\kappa_f \kappa_\sigma^p(T)) & \text{if } p > 0 \\ O_p(\Delta^{k/2} T\kappa_f \kappa_{\sigma^{-1}}^{-p}(T)) & \text{if } p \leq 0 \end{cases}$$

So as a result,

$$\sum_{i=1}^n f(X_{(i-1)\Delta}) g\big(\sigma(X_{(i-1)\Delta}) + \mathrm{D}_i\big)$$
$$= \sum_{i=1}^n f(X_{(i-1)\Delta}) \big[g(\sigma(X_{(i-1)\Delta})) + g'(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i + \cdots + \frac{1}{k!} g^{(k)}(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i^k \big]$$
$$+ \begin{cases} O_p(\Delta^{k/2}\sqrt{T}\kappa_\sigma^p \kappa_f \kappa_{\sigma^\cdot}(T)) + O_p(\Delta^{k/2} T\kappa_f \kappa_\sigma^p(T)) & \text{if } p > 0 \\ O_p(\Delta^{k/2}\sqrt{T}\kappa_{\sigma^{-1}}^{-p} \kappa_f \kappa_{\sigma^\cdot}(T)) + O_p(\Delta^{k/2} T\kappa_f \kappa_{\sigma^{-1}}^{-p}(T)) & \text{if } p \leq 0 \end{cases}$$

when k is even and

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta}) g\big(\sigma(X_{(i-1)\Delta}) + \mathrm{D}_i\big) \\
&\quad = \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \big[g(\sigma(X_{(i-1)\Delta})) + g'(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i + \cdots + \frac{1}{k!} g^{(k)}(\sigma(X_{(i-1)\Delta}))\mathrm{D}_i^k\big] \\
&\qquad + \begin{cases} O_p(\Delta^{(k-1)/2} T \kappa_\sigma^p \kappa_f \kappa_{\sigma'}(T)) & \text{if } p > 0 \\ O_p(\Delta^{(k-1)/2} T \kappa_{\sigma^{-1}}^{-p} \kappa_f \kappa_{\sigma'}(T)) & \text{if } p \le 0 \end{cases}
\end{aligned}$$

when k is odd. $\square$

B. Asymptotics of the Log-Likelihood Derivatives

1. Euler ML Estimator Asymptotics

For the scores of the Euler approximated log-likelihood function, we have

$$\mathcal{S}_\alpha(\theta) = \sum_{i=1}^{n} \ell_\alpha(x_i, y_i) = \int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t + O_p\big(\sqrt{\Delta T}(\kappa_{\mu_\alpha'} + \kappa_{\mu_\alpha}\kappa_{\sigma'}\kappa_{\sigma^{-1}})(T)\big)$$

$$\mathcal{S}_\beta(\theta) = \sum_{i=1}^{n} \ell_\beta(x_i, y_i) = \sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t + O_p\big(\Delta^{-1/4} F_{\sigma_\beta \sigma^{-3}}(T)\big)$$

and for the Hessians, we have

$$
\begin{aligned}
\mathcal{H}_{\alpha\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'}(x_i, y_i) &= -\int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'}} - \kappa_{\mu_{\alpha\alpha'}}\kappa_{\sigma^{\cdot}}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\alpha\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'}(x_i, y_i) &= -2\int_0^T \frac{\mu_\alpha \sigma'_\beta}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_\alpha}\kappa'_{\sigma_\beta}\kappa_{\sigma^{-1}} - \kappa_{\mu_\alpha}\kappa'_{\sigma_\beta}\kappa_{\sigma^{\cdot}}\kappa^2_{\sigma^{-1}} + \kappa_{\mu_\alpha}\kappa'_{\sigma^{\cdot}_\beta}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\beta\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'}(x_i, y_i) &= -2\int_0^T \frac{\sigma_\beta \mu'_\alpha}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma_\beta}\kappa'_{\mu^{\cdot}_\alpha}\kappa_{\sigma^{-1}} - \kappa_{\sigma_\beta}\kappa'_{\mu_\alpha}\kappa_{\sigma^{\cdot}}\kappa^2_{\sigma^{-1}} + \kappa_{\sigma^{\cdot}_\beta}\kappa'_{\mu_\alpha}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\beta\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\beta'}(x_i, y_i) &= -\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t)dt \\
&\quad + O_p\left(\sqrt{\frac{T}{\Delta}}(\kappa_\sigma \kappa_{\sigma_{\beta\beta'}} - 3\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta})\kappa^2_{\sigma^{-1}}(T)\right)
\end{aligned}
$$

2. Milstein ML Estimator Asymptotics

For the scores of the Milstein approximated log-likelihood function, we have

$$
\begin{aligned}
\mathcal{S}_\alpha(\theta) &= \sum_{i=1}^{n} \ell_\alpha(x_i, y_i) = \int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_\alpha} + \kappa_{\mu_\alpha}\kappa_{\sigma^{\cdot}}\kappa_{\sigma^{-1}})(T)\big) \\
\mathcal{S}_\beta(\theta) &= \sum_{i=1}^{n} \ell_\beta(x_i, y_i) = \sqrt{\frac{2}{\Delta}}\int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t + O_p\big(\Delta^{-1/4}F_{\sigma_\beta \sigma^{-3}}(T)\big)
\end{aligned}
$$

and for the Hessians, we have

$$
\begin{aligned}
\mathcal{H}_{\alpha\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'}(x_i, y_i) &= -\int_0^T \frac{\mu_\alpha \mu'_\alpha}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'}} - \kappa_{\mu_{\alpha\alpha'}}\kappa_{\sigma^{\cdot}}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\alpha\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'}(x_i, y_i) &= -2\int_0^T \frac{\mu_\alpha \sigma'_\beta}{\sigma^2}(X_t)dW_t + 3\int_0^T \frac{\mu_\alpha \sigma'_\beta \sigma^{\cdot}}{\sigma^2}(X_t)dt \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_\alpha}\kappa'_{\sigma_\beta}\kappa_{\sigma^{-1}} - \kappa_{\mu_\alpha}\kappa'_{\sigma_\beta}\kappa_{\sigma^{\cdot}}\kappa^2_{\sigma^{-1}} + \kappa_{\mu_\alpha}\kappa'_{\sigma^{\cdot}_\beta}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\beta\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'}(x_i, y_i) &= -2\int_0^T \frac{\sigma_\beta \mu'_\alpha}{\sigma^2}(X_t)dW_t + 3\int_0^T \frac{\sigma_\beta \mu'_\alpha \sigma^{\cdot}}{\sigma^2}(X_t)dt \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma_\beta}\kappa'_{\mu^{\cdot}_\alpha}\kappa_{\sigma^{-1}} - \kappa_{\sigma_\beta}\kappa'_{\mu_\alpha}\kappa_{\sigma^{\cdot}}\kappa^2_{\sigma^{-1}} + \kappa_{\sigma^{\cdot}_\beta}\kappa'_{\mu_\alpha}\kappa_{\sigma^{-1}})(T)) \\
\mathcal{H}_{\beta\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\beta'}(x_i, y_i) &= -\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta \sigma'_\beta}{\sigma^2}(X_t)dt \\
&\quad + O_p\left(\sqrt{\frac{T}{\Delta}}(\kappa_\sigma \kappa_{\sigma_{\beta\beta'}} - 3\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta})\kappa^2_{\sigma^{-1}}(T)\right)
\end{aligned}
$$

C. Proofs and Useful Lemmas for Chapter III

1. Proof of Proposition 3 and 5

Part 1: Euler ML Case

Denote $x = X_{(i-1)\Delta}$ and $y = X_{i\Delta}$. Note that we have the scores of the likelihood $\mathcal{L}$ as $S(\theta_0) = \sum_{i=1}^n \big(\ell_\alpha(x,y), \ell_\beta(x,y)\big)'$, where

$$
\begin{aligned}
\ell_\alpha(x,y) &= \frac{\mu_\alpha(x)}{\sigma^2(x)}\big(y - x - \Delta\mu(x)\big) \\
\ell_\beta(x,y) &= \frac{\sigma_\beta(x)}{\Delta\sigma^3(x)}\big[\big(y - x - \Delta\mu(x)\big)^2 - \Delta\sigma^2(x)\big]
\end{aligned}
$$

and for the Hessians, we have

$$H(\theta_0) = \sum_{i=1}^{n} \begin{pmatrix} \ell_{\alpha\alpha}(x,y) & \ell_{\alpha\beta}(x,y) \\ \ell_{\alpha\beta}(x,y) & \ell_{\beta\beta}(x,y) \end{pmatrix}$$

where

$$\begin{aligned}
\ell_{\alpha\alpha}(x,y) &= \frac{\mu_{\alpha\alpha}(x)}{\sigma^2(x)}\big(y - x - \Delta\mu(x)\big) - \frac{\Delta\mu_\alpha^2(x)}{\sigma^2(x)} \\
\ell_{\alpha\beta}(x,y) &= -\frac{2\mu_\alpha\sigma_\beta(x)}{\sigma^3(x)}\big(y - x - \Delta\mu(x)\big) \\
\ell_{\beta\beta}(x,y) &= \frac{1}{\Delta\sigma^4(x)}\Big[\big(\sigma\sigma_{\beta\beta}(x) - 3\sigma_\beta^3(x)\big)\big[\big(y - x - \Delta\mu(x)\big)^2 - \Delta\sigma^2(x)\big] - 2\Delta\sigma^2\sigma_\beta^2(x)\Big].
\end{aligned}$$

Also we have

$$J(\theta_0) = \sum_{i=1}^{n} \begin{pmatrix} \ell_{\alpha\alpha\alpha}(x,y) & \ell_{\alpha\alpha\beta}(x,y) \\ \ell_{\alpha\alpha\beta}(x,y) & \ell_{\alpha\beta\beta}(x,y) \\ \ell_{\alpha\alpha\beta}(x,y) & \ell_{\alpha\beta\beta}(x,y) \\ \ell_{\alpha\beta\beta}(x,y) & \ell_{\beta\beta\beta}(x,y) \end{pmatrix},$$

where

$$\begin{aligned}
\ell_{\alpha\alpha\alpha}(x,y) &= -\frac{\mu_{\alpha\alpha\alpha}(x)}{\sigma^2(x)}\big(y - x - \Delta\mu(x)\big) - \frac{3\Delta\mu_\alpha\mu_{\alpha\alpha}(x)}{\sigma^2(x)} \\
\ell_{\alpha\alpha\beta}(x,y) &= -\frac{2\mu_{\alpha\alpha}\sigma_\beta(x)}{\sigma^3(x)}\big(y - x - \Delta\mu(x)\big) + \frac{2\Delta\mu_\alpha^2\sigma_\beta(x)}{\sigma^3(x)} \\
\ell_{\alpha\beta\beta}(x,y) &= \frac{2\mu_\alpha\big(3\sigma_\beta^2(x) - \sigma\sigma_{\beta\beta}\big)}{\sigma^4(x)}\big(y - x - \Delta\mu(x)\big) \\
\ell_{\beta\beta\beta}(x,y) &= \frac{1}{\Delta\sigma^5(x)}\big(\sigma^2\sigma_{\beta\beta\beta}(x) - 9\sigma\sigma_\beta\sigma_{\beta\beta}(x) + 12\sigma_\beta^3\big)\big[\big(y - x - \Delta\mu(x)\big)^2 - \Delta\sigma^2(x)\big] \\
&\quad + \frac{1}{\sigma^3(x)}\big(10\sigma_\beta^3(x) - 6\sigma\sigma_\beta\sigma_{\beta\beta}(x)\big)
\end{aligned}$$

and

$$
\begin{aligned}
K_{\alpha\alpha\alpha\alpha}(\theta_0) &= \sum_{i=1}^{n} \ell_{\alpha\alpha\alpha\alpha}(x,y) \\
&= \sum_{i=1}^{n} \left[\frac{\mu_{\alpha\alpha\alpha\alpha}(x)}{\sigma^2(x)} \big(y - x - \Delta\mu(x)\big) - \frac{\Delta(3\mu_{\alpha\alpha}^2 + 4\mu_\alpha\mu_{\alpha\alpha\alpha})(x)}{\sigma^2(x)} \right].
\end{aligned}
$$

From this, it's easily derived from Lemma 21 and 22, that

$$
\begin{aligned}
\sum_{i=1}^{n} \ell_\alpha(x,y) &\approx \left[\int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \left(\mu_\alpha^\cdot - \frac{\mu_\alpha \sigma^\cdot}{\sigma} \right)(X_t)dV_t \right] \\
\sum_{i=1}^{n} \ell_\beta(x,y) &\approx \left[\sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t \right. \\
&\qquad \left. + \frac{2}{(3\Delta)^{1/4}} \int_0^T \left(\sigma_\beta^\cdot - \frac{\sigma_\beta \sigma^\cdot}{\sigma} \right)(X_t) \sqrt{\left| \sqrt{\frac{2}{3}} V_t + \frac{1}{\sqrt{3}} Z_t \right|} dU_t \right]
\end{aligned}
$$

which proves the first part of the proposition, and also from Lemma 21, 22 and 2, we have

$$
\begin{aligned}
\sum_{i=1}^{n} \ell_{\alpha\alpha}(x,y) &\approx \left[-\int_0^T \frac{\mu_\alpha^2}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha}}{\sigma}(X_t)dW_t \right. \\
&\qquad \left. - \sqrt{\frac{\Delta}{2}} \int_0^T \left(\mu_{\alpha\alpha}^\cdot - \frac{\mu_{\alpha\alpha}\sigma^\cdot}{\sigma} \right)(X_t)dV_t \right] \\
\sum_{i=1}^{n} \ell_{\alpha\beta}(x,y) &\approx \left[-2 \int_0^T \frac{\mu_\alpha \sigma_\beta}{\sigma^2}(X_t)dW_t \right. \\
&\qquad \left. + \sqrt{2\Delta} \int_0^T \left(\frac{2\mu_\alpha \sigma_\beta \sigma^\cdot}{\sigma^2} - \frac{\mu_\alpha^\cdot \sigma_\beta}{\sigma} - \frac{\mu_\alpha \sigma_\beta^\cdot}{\sigma} \right)(X_t)dV_t \right] \\
\sum_{i=1}^{n} \ell_{\beta\beta}(x,y) &\approx \left[-\frac{2}{\Delta} \int_0^T \frac{\sigma_\beta^2}{\sigma^2}(X_t)dt + \sqrt{\frac{2}{\Delta}} \int_0^T \left(\frac{\sigma_{\beta\beta}}{\sigma} - \frac{3\sigma_\beta^2}{\sigma^2} \right)(X_t)dV_t \right].
\end{aligned}
$$

I omit the results for J and K here. Note that for $\ell_{\alpha\alpha}$ term, the second term will be of smaller order from Assumption 5 when $T \to \infty$ and $\Delta \to 0$, but when T is fixed, both the first term and the second term will be the leading term in the asymptotics. It's also easy to extend the vector case by applying these lemmas elementwise. As for the diagonality, since

$$w = \mathrm{diag}\Big(\sqrt{T}\kappa_{\mu_\alpha}\kappa_\sigma^{-1}(\nu(T)), \sqrt{T/\Delta}\kappa_{\sigma_\beta}\kappa_\sigma^{-1}(\nu(T))\Big)$$

it's easy to check that $H_0(\theta_0)$ will be block diagonal from

$$\begin{aligned}\frac{\sqrt{\Delta}}{T}\kappa_{\mu_\alpha}^{-1}\kappa_{\sigma_\beta}^{-1}\kappa_\sigma^2(\nu(T))\int_0^T \frac{\mu_\alpha\sigma_\beta}{\sigma^2}(X_t)dW_t \\ = \frac{\sqrt{\Delta}}{T}\kappa_{\mu_\alpha}^{-1}\kappa_{\sigma_\beta}^{-1}\kappa_\sigma^2(\nu(T))O_p\big(\sqrt{T}\kappa_{\mu_\alpha}\kappa_{\sigma_\beta}\kappa_\sigma^{-2}(\nu(T))\big) \to_p 0\end{aligned}$$

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

Part 2: Milstein ML Case

It's straightforward from the functional form of the score and Hessian functions, using Lemma 1-12, 14, 21 and 22. The basic procedure is same as the Euler case, but I'll not go in detail for each case here. For example, for the score function with respect to the drift term parameter,

$$\frac{\partial \ell_i}{\partial \alpha} = \left(\frac{e^+ - e^-}{e^+ + e^-}\right)\frac{\sqrt{\Delta}\mu_\alpha}{\sigma' B} + \frac{\mu_\alpha}{\sigma\sigma^\cdot} + \frac{\Delta^2\sigma\sigma^\cdot\mu_\alpha}{B^2}$$

where $B = \big(\Delta\,\sigma(\sigma + \Delta\,\sigma\sigma^{\cdot 2} + 2\,\sigma^\cdot(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\,\mu))\big)^{1/2}$, suppressing all the arguments for the functions. Note that for the term containing $\frac{e^+ - e^-}{e^+ + e^-}$, it's same as finding the limiting distribution without $\frac{e^+ - e^-}{e^+ + e^-}$ from Lemma 12, and for the terms with B, they can be taken care of by Lemma 14, and as a result, we get the following terms.

$$\begin{aligned}\sum_{i=1}^n \frac{\partial \ell_i}{\partial \alpha} =& \sum_{i=1}^n \frac{\mu_\alpha}{\sigma^2}(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\,\mu) \\ &- \frac{3}{2}\sum_{i=1}^n \frac{\mu_\alpha\sigma'}{\sigma^3}\left[(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\,\mu)^2 - \Delta\sigma^2\right] + O_p(\Delta\kappa_T)\end{aligned}$$

So the rest of the step is to find the asymptotic expansions of each terms, and we get

$$\sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \alpha} \approx \left[\int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \left(\dot{\mu_\alpha} + \frac{2\mu_\alpha \dot{\sigma}}{\sigma} \right) (X_t)dV_t \right]$$

using Lemma 21 and 22.

2. Proof of Theorem 2 and 3

We begin this proof from (3). Following the notations in Appendix, note that

$$\mathcal{S}(\theta) = \begin{pmatrix} \mathcal{S}_\alpha(\theta) \\ \mathcal{S}_\beta(\theta) \end{pmatrix} \quad \text{and} \quad \mathcal{H}(\theta) = \begin{pmatrix} \mathcal{H}_{\alpha\alpha'}(\theta) & \mathcal{H}_{\alpha\beta'}(\theta) \\ \mathcal{H}'_{\alpha\beta'}(\theta) & \mathcal{H}_{\beta\beta'}(\theta) \end{pmatrix},$$

and for $\mathcal{J}(\theta)$, jth $k \times k$ block of this $k^2 \times k$ matrix is

$$\mathcal{J}_j(\theta) = \begin{pmatrix} \mathcal{J}_{\alpha\alpha'\alpha_j}(\theta) & \mathcal{J}_{\alpha\beta'\alpha_j}(\theta) \\ \mathcal{J}'_{\alpha\beta'\alpha_j}(\theta) & \mathcal{J}_{\beta\beta'\alpha_j}(\theta) \end{pmatrix},$$

for $1 \leq j \leq k_1$ where α_j is the jth element of α, and

$$\mathcal{J}_j(\theta) = \begin{pmatrix} \mathcal{J}_{\alpha\alpha'\beta_j}(\theta) & \mathcal{J}_{\alpha\beta'\beta_j}(\theta) \\ \mathcal{J}'_{\alpha\beta'\beta_j}(\theta) & \mathcal{J}_{\beta\beta'\beta_j}(\theta) \end{pmatrix},$$

for $k_1 + 1 \leq j \leq k$ where β_j is the $(j - k_1)$th element of β. Note that, for example, $\mathcal{J}_{\alpha\alpha'\alpha_j}(\theta)$ is the jth $k \times k$ block of $\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta)$, i.e.,

$$\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta) = \begin{pmatrix} \mathcal{J}_{\alpha\alpha'\alpha_1}(\theta) \\ \vdots \\ \mathcal{J}_{\alpha\alpha'\alpha_{k_1}}(\theta) \end{pmatrix}.$$

Now applying the following block matrix inversion formula to $\mathcal{H}(\theta)^{-1}$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} \big(A - BD^{-1}C\big)^{-1} & -A^{-1}B\big(D - CA^{-1}B\big)^{-1} \\ -\big(D - CA^{-1}B\big)^{-1}CA^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{pmatrix}$$

and arranging the terms based on the Δ orders, we can find the first order term of D_T becomes

$$\begin{aligned} -\frac{1}{2}\mathcal{H}_{\alpha\alpha'}(\theta_0)^{-1}\big(I_k \otimes (\hat{\alpha} - \alpha_0)'\big)\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta_0)(\hat{\alpha} - \alpha_0) \\ \approx -\frac{1}{2}H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} \end{aligned}$$

for the α part, and

$$\begin{aligned} -\frac{1}{2}\mathcal{H}_{\beta\beta'}(\theta_0)^{-1}\Big(&\big(I_k \otimes (\hat{\alpha} - \alpha_0)'\big)\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta_0)(\hat{\alpha} - \alpha_0) \\ &+ \big(I_k \otimes (\hat{\beta} - \beta_0)'\big)\mathcal{J}_{\beta\beta'\otimes\beta}(\theta_0)(\hat{\beta} - \beta_0)\Big) \\ &+ \frac{1}{2}\mathcal{H}_{\beta\beta'}(\theta_0)^{-1}\mathcal{H}_{\alpha\beta'}'(\theta_0)\mathcal{H}_{\alpha\alpha'}(\theta_0)^{-1}\big(I_k \otimes (\hat{\alpha} - \alpha_0)'\big)\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta_0)(\hat{\alpha} - \alpha_0) \\ \approx -\frac{\Delta}{2}H_{\beta\beta,1}^{-1}\Big(&\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} + \big(I_k \otimes S_{\beta,1}' H_{\beta\beta,1}^{-1}\big)J_{\beta\beta\beta,1}H_{\beta\beta,1}^{-1}S_{\beta,1}\Big) \\ &+ \frac{\Delta}{2}H_{\beta\beta,1}^{-1}H_{\alpha\beta,1}' H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} \end{aligned} \tag{A.16}$$

for the β part.

For C_T, denoting

$$\mathcal{H}(\theta_0) \approx \begin{pmatrix} H_{\alpha\alpha,1} + \sqrt{\Delta}H_{\alpha\alpha,2} & H_{\alpha\beta,1} + \sqrt{\Delta}H_{\alpha\beta,2} \\ H_{\alpha\beta,1}' + \sqrt{\Delta}H_{\alpha\beta,2}' & \frac{1}{\Delta}H_{\beta\beta,1} + \frac{1}{\sqrt{\Delta}}H_{\beta\beta,2} \end{pmatrix}$$

and

$$\mathcal{S}(\theta_0) \approx \begin{pmatrix} S_{\alpha,1} + \sqrt{\Delta}S_{\alpha,2} \\ \frac{1}{\sqrt{\Delta}}S_{\beta,1} + \frac{1}{\Delta^{1/4}}S_{\beta,2} \end{pmatrix}.$$

Applying the block matrix inversion formula again, we have

$$
\begin{aligned}
&-\mathcal{H}(\theta_0)^{-1}\mathcal{S}(\theta_0)\\
&\quad\approx \begin{pmatrix} -H_{\alpha\alpha,1}^{-1}S_{\alpha,1} - \sqrt{\Delta}H_{\alpha\alpha,1}^{-1}\Big(H_{\alpha\alpha,2}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} + S_{\alpha,2} - H_{\alpha\beta,1}H_{\beta\beta,1}^{-1}S_{\beta,1}\Big) \\ -\sqrt{\Delta}H_{\beta\beta,1}^{-1}S_{\beta,1} - \Delta^{3/4}H_{\beta\beta,1}^{-1}S_{\beta,2} \end{pmatrix}
\end{aligned}
$$

eliminating all the higher order terms which are smaller than the second term. Now arranging the terms again, we have

$$
\begin{aligned}
\hat{\alpha} - \alpha_0 &\approx -H_{\alpha\alpha,1}^{-1}S_{\alpha,1} - \frac{1}{2}H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1}\\
&\quad - \sqrt{\Delta}H_{\alpha\alpha,1}^{-1}\Big(H_{\alpha\alpha,2}H_{\alpha\alpha,1}^{-1}S_{\alpha,1} + S_{\alpha,2} - H_{\alpha\beta,1}H_{\beta\beta,1}^{-1}S_{\beta,1}\Big)\\
&= A_{1T} + A_{2T} + \sqrt{\Delta}A_{3T}\\
\hat{\beta} - \beta_0 &\approx -\sqrt{\Delta}H_{\beta\beta,1}^{-1}S_{\beta\beta,1} - \Delta^{3/4}H_{\beta\beta,1}^{-1}S_{\beta,2}\\
&= \sqrt{\Delta}B_{1T} + \Delta^{3/4}B_{2T}
\end{aligned}
$$

since (A.16) is of smaller order than $\Delta^{3/4}B_{2T}$.

The proof of Theorem 3 is the same as the one of Theorem 2, but ignoring the Δ order terms. Beginning rom (4), since we are ignoring Δ order terms, it's easy to see that all the higher order terms of A_T are coming from $-H_{\alpha\alpha,1}^{-1}S_{\alpha,1}$, and also the higher terms of B_T come from

$$
-\frac{1}{2}H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)J_{\alpha\alpha\alpha,1}H_{\alpha\alpha,1}^{-1}S_{\alpha,1}.
$$

For the leading term of C_T, we can check that it is

$$
-\frac{1}{6}H_{\alpha\alpha,1}^{-1}\big(I_k \otimes S_{\alpha,1}' H_{\alpha\alpha,1}^{-1}\big)K_{\alpha\alpha\alpha\alpha,1}\big(H_{\alpha\alpha,1}^{-1}S_{\alpha,1} \otimes H_{\alpha\alpha,1}^{-1}S_{\alpha,1}\big).
$$

3. Useful Lemmas

Lemma 14. *Define*

$$Z_t^{\Delta} = \frac{6}{\Delta^{3/2}} \left(\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3} \right) \int_{(i-1)\Delta}^{s} dW_u dW_s \right.$$
$$\left. + \int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3} \right) \int_{(j-1)\Delta}^{s} dW_u dW_s \right)$$

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

$$Z^{\Delta} \to_p Z$$

for a standard Brownian motion Z *which is independent of* W *and* V. *Also,*

$$Z_t^{\Delta} - Z_t = O_p\big((\Delta T)^{1/4}\big).$$

Proof. Clearly, Z^{Δ} is a continuous martingale with quadratic variation given by

$$[Z^{\Delta}]_t = \frac{36}{\Delta^3} \left[\sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3} \right)^2 (W_s - W_{(i-1)\Delta})^2 ds \right.$$
$$\left. + \int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3} \right)^2 (W_s - W_{(j-1)\Delta})^2 ds \right]$$

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

$$\begin{aligned}[Z^{\Delta}]_t - t = & \frac{36}{\Delta^3} \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3} \right)^2 \big[(W_s - W_{(i-1)\Delta})^2 - (s - (i-1)\Delta) \big] ds \\ & + \frac{36}{\Delta^3} \int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3} \right)^2 \big[(W_s - W_{(j-1)\Delta})^2 - (s - (j-1)\Delta) \big] ds \\ & + O(\Delta) \qquad \text{(A.17)}\end{aligned}$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \ldots, n+1$, uniformly in $t \in [0, T]$. Therefore, ignoring

$O(\Delta)$ term in (A.17) that is unimportant, it follows that

$$
\begin{aligned}
&\mathbb{E}\left([Z^{\Delta}]_t - t\right)^2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.18)}\\
=&\frac{36^2}{\Delta^6}\sum_{i=1}^{j-1}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)^2\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]ds\right)^2\\
&+\frac{36^2}{\Delta^6}\mathbb{E}\left(\int_{(j-1)\Delta}^{t}\left(i\Delta - s - \frac{\Delta}{3}\right)^2\left[(W_s - W_{(j-1)\Delta})^2 - (s-(j-1)\Delta)\right]ds\right)^2
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)^2\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]ds\right)^2\\
&\quad\leq \Delta\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)^4\mathbb{E}\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]^2 ds = \frac{11\Delta^8}{2835}
\end{aligned}
\tag{A.19}
$$

for $i = 1, \ldots, n$. Moreover, we may deduce from (A.18) and (A.19) that

$$
\begin{aligned}
&\mathbb{E}\left([Z^{\Delta}]_t - t\right)^2\\
&\quad\leq \frac{36^2}{\Delta^6}\sum_{i=1}^{n}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}(s-(i-1)\Delta)^2\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]ds\right)^2\\
&\quad= \frac{36^2}{\Delta^6}n\frac{11\Delta^8}{2835} = \frac{176}{35}\Delta T \to 0
\end{aligned}
$$

under our assumption. Consequently, it follows that

$$
\sup_{0\leq t\leq T}\mathbb{E}\left([Z^{\Delta}]_t - t\right)^2 \to 0
$$

in our asymptotic framework. This implies that

$$
Z^{\Delta} \to_p Z,
$$

where Z is the standard Brownian motion.

We now prove that Z is independent of V. For this, we note that

$$[Z^\Delta, V^\Delta]_t = \frac{6\sqrt{2}}{\Delta^2}\left[\sum_{i=1}^{j-1}\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds \right. \\ \left. + \int_{(j-1)\Delta}^{t}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(j-1)\Delta})^2 ds\right]$$

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

$$\mathbb{E}\left([Z^\Delta, V^\Delta]_t\right)^2 = \frac{72}{\Delta^4}\left[\sum_{i=1}^{j-1}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds\right)^2 \right. \\ \left. + \mathbb{E}\left(\int_{(j-1)\Delta}^{t}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(j-1)\Delta})^2 ds\right)^2\right] \quad \text{(A.20)}$$

for $t \in [(j-1)\Delta, j\Delta)$, $j = 1, \ldots, n+1$, due to the independent increment property of Brownian motion and to that

$$\mathbb{E}\left[\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds\right] = 0.$$

Moreover, we have by Cauchy-Schwarz

$$\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds\right)^2 \\ \leq \Delta\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)^2\mathbb{E}(W_s - W_{(i-1)\Delta})^4 ds = \frac{2\Delta^6}{45} \quad \text{(A.21)}$$

for $i = 1, \ldots, n$. Therefore, it can be deduced from (A.20) and (A.21) that

$$\mathbb{E}\left([Z^\Delta, V^\Delta]_t\right)^2 = \frac{72}{\Delta^4}\left[\sum_{i=1}^{j-1}\mathbb{E}\left(\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds\right)^2 \right. \\ \left. + \mathbb{E}\left(\int_{(j-1)\Delta}^{t}\left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})^2 ds\right)^2\right] \\ \leq \frac{72}{\Delta^4}n\frac{2\Delta^6}{45} = \frac{16}{5}\Delta T \to 0,$$

and that

$$\sup_{0\le t\le T} \mathbb{E}\left([Z^\Delta, V^\Delta]_t\right)^2 \to 0$$

in our asymptotic framework.

To prove that Z is independent of W, note that

$$\begin{aligned}
[Z^\Delta, W]_t &= \frac{6}{\Delta^{3/2}} \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})ds \\
&\quad + \frac{6}{\Delta^{3/2}} \int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})ds
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\big([Z^\Delta, W]_t\big)^2 &= \frac{36}{\Delta^3} \sum_{i=1}^{j-1} \mathbb{E}\Big[\int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})ds\Big]^2 \\
&\quad + \frac{36}{\Delta^3}\mathbb{E}\Big[\int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3}\right)(W_s - W_{(i-1)\Delta})ds\Big]^2 \\
&\le \frac{36}{\Delta^2} \sum_{i=1}^{j-1} \int_{(i-1)\Delta}^{i\Delta} \left(i\Delta - s - \frac{\Delta}{3}\right)^2 \mathbb{E}(W_s - W_{(i-1)\Delta})^2 ds \\
&\quad + \frac{36}{\Delta^2} \int_{(j-1)\Delta}^{t} \left(i\Delta - s - \frac{\Delta}{3}\right)^2 \mathbb{E}(W_s - W_{(i-1)\Delta})^2 ds \\
&\le \Delta^2 n = \Delta T \to 0
\end{aligned}$$

For the last statement, note that Z_t^Δ can be represented as a time changed Brownian motion $Z_{[Z^\Delta]_t}$ from the DDS representation. Thus we have

$$\frac{Z_{[Z^\Delta]_t} - Z_t}{\sqrt{\big|[Z^\Delta]_t - t\big|}} \sqrt{\big|[Z^\Delta]_t - t\big|} = O_p(1)\sqrt{\big|O_p(\sqrt{\Delta T})\big|} = O_p\big((\Delta T)^{1/4}\big).$$

□

Lemma 15.

$$[V^\Delta]_t - t = \frac{2\sqrt{2}}{3}\sqrt{\Delta} V_t + \frac{2}{3}\sqrt{\Delta} Z_t + O_p(\Delta^{3/4} T^{1/4})$$

Proof. From

$$
\begin{aligned}
[V^{\Delta}]_t - t &= \frac{2}{\Delta}\sum_{i=1}^{j-1}\int_{(i-1)\Delta}^{i\Delta}\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]ds \\
&\quad + \frac{2}{\Delta}\int_{(j-1)\Delta}^{t}\left[(W_s - W_{(i-1)\Delta})^2 - (s-(i-1)\Delta)\right]ds \\
&= \frac{4}{\Delta}\sum_{i=1}^{j-1}\int_{(i-1)\Delta}^{i\Delta}(i\Delta - s)\int_{(i-1)\Delta}^{s} dW_r dW_s + \frac{4}{\Delta}\int_{(j-1)\Delta}^{t}(i\Delta - s)\int_{(i-1)\Delta}^{s} dW_r dW_s \\
&= \frac{4}{3}\sum_{i=1}^{j-1}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{s} dW_r dW_s + \frac{4}{\Delta}\sum_{i=1}^{j-1}\int_{(i-1)\Delta}^{i\Delta}\left(i\Delta - s - \frac{\Delta}{3}\right)\int_{(i-1)\Delta}^{s} dW_r dW_s \\
&\quad + \frac{4}{3}\int_{(j-1)\Delta}^{t}\int_{(i-1)\Delta}^{s} dW_r dW_s + \frac{4}{\Delta}\int_{(j-1)\Delta}^{t}\left(i\Delta - s - \frac{\Delta}{3}\right)\int_{(i-1)\Delta}^{s} dW_r dW_s \\
&= \frac{2\sqrt{2}}{3}\sqrt{\Delta}V_t^{\Delta} + \frac{2}{3}\sqrt{\Delta}Z_t^{\Delta}
\end{aligned}
$$

it easily follows from Lemma 10 and Lemma 14. □

Lemma 16. *Define*

$$U_t^{\Delta} = \int_0^t \frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}} dW_s.$$

Then

$$U^{\Delta} \to_p U,$$

where U is a standard BM independent of V, Z and W.

Proof. Note that U^{Δ} is a continuous martingale with quadratic variation given by

$$[U^{\Delta}]_t = \int_0^t \left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 ds.$$

We have

$$\mathbb{E}\big([U^{\Delta}]_t - t\big)^2 = \mathbb{E}\left(\int_0^t \left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right] ds\right)^2$$
$$= \int_0^t \int_0^t \mathbb{E}\left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right]\left[\left(\frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}\right)^2 - 1\right] dsdr.$$

Note here that

$$\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 \sim \chi_1^2$$

for any s and Δ so

$$\mathbb{E}\left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right]\left[\left(\frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}\right)^2 - 1\right]$$

is a covariance between two χ_1^2 random variables. Thus,

$$\left|\mathbb{E}\left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right]\left[\left(\frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}\right)^2 - 1\right]\right| \leq 2$$

for any s, r and Δ, and for any $s \neq r$, (shown in Part A)

$$\left|\mathbb{E}\left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right]\left[\left(\frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}\right)^2 - 1\right]\right| = O(\Delta T)$$

So we have

$$\int_0^t \int_0^t \mathbb{E}\left[\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}}\right)^2 - 1\right]\left[\left(\frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}\right)^2 - 1\right] dsdr$$
$$\leq \int_0^t \int_0^t \left[2 \cdot 1_{\{t=s\}} + O(\Delta T)\right] dsdr$$
$$= O(\Delta T^3) \to 0$$

if $\Delta T^3 \to 0$. Thus we have

$$\sup_{0 \le t \le T} \mathbb{E}\big([U^\Delta]_t - t\big)^2 \to 0$$

and this proves

$$U^\Delta \to_p U,$$

where U is a standard Brownian motion.

Part A. Let's denote

$$A_\Delta = \left(\frac{V_s^\Delta - V_s}{\sqrt{|[V^\Delta]_s - s|}}\right)^2 - 1 \quad \text{and} \quad B_\Delta = \left(\frac{V_r^\Delta - V_r}{\sqrt{|[V^\Delta]_r - r|}}\right)^2 - 1.$$

Then

$$\mathbb{E} A_\Delta B_\Delta \le 2 \int_{-\infty}^{\infty} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx$$

where F_s^Δ and f_s^Δ are distribution and density functions of $[V^\Delta]_s$, and similarly for F_r^Δ and f_r^Δ with $[V^\Delta]_r$. The above inequality is because it's a probability that A_Δ and B_Δ will be dependent. To deal with this integral, let's divide it by

$$\begin{aligned}
&\int_{-\infty}^{\infty} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx \\
&\quad = \int_{-\infty}^{c_1} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx + \int_{c_1}^{c_2} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx \\
&\qquad + \int_{c_2}^{\infty} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx \\
&\quad = A + B + C
\end{aligned}$$

where $r < c_1 < s < c_2$. Then

$$\begin{aligned}
A &\le \int_{-\infty}^{c_1} 1\cdot f_r^\Delta(x)dx = F_r^\Delta(c_1) \\
B &\le \int_{c_1}^{c_2} (1 - F_s^\Delta(c_1)) f_r^\Delta(x)dx = (1 - F_s^\Delta(c_1)) \int_{c_1}^{c_2} f_r^\Delta(x)dx \le 1 - F_s^\Delta(c_1) \\
C &\le \int_{c_2}^{\infty} 1\cdot f_r^\Delta(x)dx = 1 - F_r^\Delta(c_2).
\end{aligned}$$

For B and C,

$$\begin{aligned}
B \le 1 - F_s^\Delta(c_1) = \mathbf{P}\{[V^\Delta]_s \ge c_1\} &\le \mathbf{P}\{\big|[V^\Delta]_s - s\big| \ge c_1 - s\} \\
&\le \frac{\mathbb{E}\big([V^\Delta]_s - s\big)^2}{(c_1 - s)^2} = O(\Delta T) \\
C \le 1 - F_r^\Delta(c_2) = \mathbf{P}\{[V^\Delta]_r \ge c_2\} &\le \mathbf{P}\{\big|[V^\Delta]_r - r\big| \ge c_2 - r\} \\
&\le \frac{\mathbb{E}\big([V^\Delta]_r - r\big)^2}{(c_2 - r)^2} = O(\Delta T)
\end{aligned}$$

since $c_1 > s$ and $c_2 > r$. For A,

$$\begin{aligned}
A \le F_r^\Delta(c_1) = \mathbf{P}\{[V^\Delta]_r \le c_1\} &\le \mathbf{P}\{\big|[V^\Delta]_r - r\big| \ge r - c_1\} \\
&\le \frac{\mathbb{E}\big([V^\Delta]_r - r\big)^2}{(r - c_1)^2} = O(\Delta T)
\end{aligned}$$

since $r > c_1$. So we have

$$\mathbb{E} A_\Delta B_\Delta = O(\Delta T)$$

Part B.

Independency 1. (U independent of W)

$$[U^\Delta, W]_t = \int_0^t \frac{V_s^\Delta - V_s}{\sqrt{|V_s^\Delta - s|}}\, ds$$

We have

$$\begin{aligned}
\mathbb{E}[U^{\Delta}, W]_t^2 &= \int_0^t \int_0^t \mathbb{E}\left(\frac{V_s^{\Delta} - V_s}{\sqrt{|V_s^{\Delta} - s|}} \frac{V_r^{\Delta} - V_r}{\sqrt{|V_r^{\Delta} - r|}} \right) dsdr \\
&\leq \int_0^t \int_0^t \left[1_{\{s=r\}} + O(\Delta T) \right] dsdr \\
&= O(\Delta T^3) \to 0
\end{aligned}$$

if $\Delta T^3 \to 0$. For the second line, let's denote

$$A_{\Delta} = \frac{V_s^{\Delta} - V_s}{\sqrt{|[V^{\Delta}]_s - s|}} \quad \text{and} \quad B_{\Delta} = \frac{V_r^{\Delta} - V_r}{\sqrt{|[V^{\Delta}]_r - r|}}.$$

Then the rest steps are the same as in Part A.

Independency 2. (U independent of V)

$$[U^{\Delta}, V]_t = \int_0^t \frac{V_s^{\Delta} - V_s}{\sqrt{|[V]_s^{\Delta} - s|}} \, d[W, V]_s = 0$$

since

$$\frac{V_s^{\Delta} - V_s}{\sqrt{|[V]_s^{\Delta} - s|}} = O_p(1).$$

Note that

$$\begin{aligned}
\int_0^t \frac{V_s^{\Delta} - V_s}{\sqrt{|[V]_s^{\Delta} - s|}} \, d[W, V]_s &\leq \sqrt{\int_0^t \frac{(V_s^{\Delta} - V_s)^2}{|[V]_s^{\Delta} - s|} ds \int_0^t [W, V]^2 ds} \\
&= O_p(\sqrt{t}) \cdot 0 = 0
\end{aligned}$$

and

$$\begin{aligned}\int_0^t \frac{V_s^\Delta - V_s}{\sqrt{|[V]_s^\Delta - s|}}\, d[W,V]_s &\geq -\int_0^t \frac{|V_s^\Delta - V_s|}{\sqrt{|[V]_s^\Delta - s|}}\, d[W,V]_s \\ &\geq -\sqrt{\int_0^t \frac{(V_s^\Delta - V_s)^2}{|[V]_s^\Delta - s|} ds \int_0^t [W,V]^2 ds} \\ &= O_p(\sqrt{t}) \cdot 0 = 0\end{aligned}$$

Independency 3. (U independent of Z) Same as above replacing V with Z. $\square$

Lemma 17.

$$\frac{1}{\Delta^{1/4}} \int_0^T f(X_t)(V_t^\Delta - V_t) dW_t \approx \sqrt{\frac{2}{3}} \int_0^T f(X_t) \sqrt{\left|\sqrt{2} V_t + Z_t\right|} dU_t$$

where U is a standard BM independent of V and W.

Proof. With an equi-spaced partition $(t_0, t_1, \cdots, t_n)$ with $t_0 = 0$ and $t_n = T$, let $\delta = t_i - t_{i-1}$. Denoting

$$M_t^\Delta = \int_0^t \frac{V_s^\Delta - V_s}{\sqrt{|[V^\Delta]_s - s|}} dW_s$$

we can rewrite

$$\begin{aligned}\frac{1}{\Delta^{1/4}} \int_0^T f(X_t)(V_t^\Delta - V_t) dW_t &= \int_0^T f(X_t) \frac{\sqrt{|[V^\Delta]_t - t|}}{\Delta^{1/4}} \frac{V_t^\Delta - V_t}{\sqrt{|[V^\Delta]_t - t|}} dW_t \\ &= \operatorname*{plim}_{\delta \to 0} \sum_{i=1}^n f(X_{t_{i-1}}) \frac{\sqrt{|[V^\Delta]_{t_{i-1}} - t_{i-1}|}}{\Delta^{1/4}} (M_{t_i}^\Delta - M_{t_{i-1}}^\Delta)\end{aligned}$$

from the definition of the Itô integral since

$$dM_t^\Delta = \frac{V_t^\Delta - V_t}{\sqrt{|[V^\Delta]_t - t|}} dW_t.$$

From Lemma 16,

$$M_{t_i}^\Delta - M_{t_{i-1}}^\Delta = \int_{t_{i-1}}^{t_i} \frac{V_s^\Delta - V_s}{\sqrt{|[V^\Delta]_s - s|}} dW_s \to_p U_{t_i} - U_{t_{i-1}}$$

and from Lemma 15,

$$\frac{\sqrt{|[V^{\Delta}]_{t_{i-1}} - t_{i-1}|}}{\Delta^{1/4}} \to_p \frac{\sqrt{2}}{3^{1/4}} \sqrt{\left| \sqrt{\frac{2}{3}} V_{t_{i-1}} + \frac{1}{\sqrt{3}} Z_{t_{i-1}} \right|},$$

and the both convergences are uniform in i, so we can exchange the limits

$$\begin{aligned}
\lim_{\Delta \to 0} \operatorname*{plim}_{\delta \to 0} \sum_{i=1}^{n} & f(X_{t_{i-1}}) \frac{\sqrt{|[V^{\Delta}]_{t_{i-1}} - t_{i-1}|}}{\Delta^{1/4}} (M_{t_i}^{\Delta} - M_{t_{i-1}}^{\Delta}) \\
&= \operatorname*{plim}_{\delta \to 0} \lim_{\Delta \to 0} \sum_{i=1}^{n} f(X_{t_{i-1}}) \frac{\sqrt{|[V^{\Delta}]_{t_{i-1}} - t_{i-1}|}}{\Delta^{1/4}} (M_{t_i}^{\Delta} - M_{t_{i-1}}^{\Delta}) \\
&= \operatorname*{plim}_{\delta \to 0} \frac{\sqrt{2}}{3^{1/4}} \sum_{i=1}^{n} f(X_{t_{i-1}}) \sqrt{\left| \sqrt{\frac{2}{3}} V_{t_{i-1}} + \frac{1}{\sqrt{3}} Z_{t_{i-1}} \right|} (U_{t_i} - U_{t_{i-1}}) \\
&= \frac{\sqrt{2}}{3^{1/4}} \int_0^T f(X_t) \sqrt{\left| \sqrt{\frac{2}{3}} V_t + \frac{1}{\sqrt{3}} Z_t \right|} dU_t
\end{aligned}$$

□

Lemma 18.

$$\int_0^T f(X_t)(V_t^{\Delta} - V_t) dt = O_p(\Delta^{3/8} T^{13/8} \kappa_f(\nu(T)))$$

Proof. **Part A.**

$$\mathbb{E}\left(\int_0^T f(X_t)(V_t^{\Delta} - V_t) dt \right)^2 = \int_0^T \int_0^T \mathbb{E}\left[f(X_t) f(X_s)(V_t^{\Delta} - V_t)(V_s^{\Delta} - V_s) \right] dt ds$$

Note that

$$
\begin{aligned}
&\mathbb{E}\left[f(X_t)f(X_s)(V_t^{\Delta}-V_t)(V_s^{\Delta}-V_s)\right]\\
&\quad=\mathbb{E}\Big(\mathbb{E}\big[\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)(V_s^{\Delta}-V_s)\big|f(X_t)f(X_s)\big]f(X_t)f(X_s)\Big)\\
&\quad\le\sqrt{\mathbb{E}\Big(\big(\mathbb{E}\big[\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)(V_s^{\Delta}-V_s)\big|f(X_t)f(X_s)\big]\big)^2\Big)\mathbb{E}\big(f^2(X_t)f^2(X_s)\big)}\\
&\quad\le\sqrt{\mathbb{E}\left(\sup_{0\le t,s\le T}\big|f^2(X_t)f^2(X_s)\big|\right)}\times\\
&\qquad\sqrt{\mathbb{E}\Big(\big(\mathbb{E}\big[\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)(V_s^{\Delta}-V_s)\big|f(X_t)f(X_s)\big]\big)^2\Big)}
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\int_0^T\int_0^T\mathbb{E}\left[f(X_t)f(X_s)(V_t^{\Delta}-V_t)(V_s^{\Delta}-V_s)\right]dtds\\
&\qquad\le\sqrt{\mathbb{E}\left(\sup_{0\le t,s\le T}\big|f^2(X_t)f^2(X_s)\big|\right)}\times\\
&\qquad\quad\int_0^T\int_0^T\sqrt{\mathbb{E}\Big(\big(\mathbb{E}\big[\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)(V_s^{\Delta}-V_s)\big|f(X_t)f(X_s)\big]\big)^2\Big)}dtds\\
&\qquad=O_p(\Delta^{3/4}T^{13/4}\kappa_f^2(T^r))
\end{aligned}
$$

Note that $\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)=O_p(\sqrt{\Delta}T)$ (shown in Part B) and $V_s^{\Delta}-V_s=O_p\big((\Delta T)^{1/4}\big)$, thus

$$
\sqrt{\mathbb{E}\Big(\big(\mathbb{E}\big[\mathbb{E}\big(V_t^{\Delta}-V_t|V_s^{\Delta}-V_s\big)(V_s^{\Delta}-V_s)\big|f(X_t)f(X_s)\big]\big)^2\Big)}=O_p(\Delta^{3/4}T^{5/4})
$$

and

$$
\mathbb{E}\left(\sup_{0\le t,s\le T}\big|f^2(X_t)f^2(X_s)\big|\right)=O_p\big(\kappa_f^2(\nu(T))\big)
$$

Thus

$$
\int_0^T f(X_t)(V_t^{\Delta}-V_t)dt=O_p(\Delta^{3/8}T^{13/8}\kappa_f(\nu(T))).
$$

Part B. Since

$$\begin{aligned}\mathbb{E}\Big(\mathbb{E}\big(V_t^\Delta - V_t | V_s^\Delta - V_s\big)^2\Big) &\leq \mathbb{E}\Big(\mathbb{E}\big((V_t^\Delta - V_t)^2 | V_s^\Delta - V_s\big)\Big) \\ &= \mathbb{E}\big((V_t^\Delta - V_t)^2\big) \\ &\leq 2T\end{aligned}$$

we have

$$\mathbb{E}\Big(\mathbb{E}\big(V_t^\Delta - V_t | V_s^\Delta - V_s\big)^2\Big) \leq 2T \int_{-\infty}^{\infty} (1 - F_s^\Delta(x)) f_r^\Delta(x) dx = O_p(\Delta T^2)$$

from the same steps as in the proof of Lemma 16 since $\mathbb{E}\big(V_t^\Delta - V_t | V_s^\Delta - V_s\big) = 0$ when $V_t^\Delta - V_t$ and $V_s^\Delta - V_s$ are independent. Thus

$$\mathbb{E}\big(V_t^\Delta - V_t | V_s^\Delta - V_s\big) = O_p(\sqrt{\Delta} T).$$

$\square$

Lemma 19.

$$\begin{aligned}&\sqrt{\frac{2}{\Delta}} \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} dW_u dW_s \\ &\qquad \approx \left(\int_0^T f(X_t) dV_t + \Delta^{1/4} \sqrt{\frac{2}{3}} \int_0^T \sigma f'(X_t) \sqrt{\left|\sqrt{2} V_t + Z_t\right|} dU_t \right)\end{aligned}$$

where V is a standard Brownian motion independent of W, and U is a standard Brownian motion independent of W and V.

Proof. **Part 1.**

$$\begin{aligned}
&\sqrt{\frac{2}{\Delta}}\sum_{i=1}^{n} f(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{s} dW_u dW_s \\
&\quad= \int_0^T f(X_t)dV_t + \int_0^T f(X_t)d(V^\Delta - V)_t \\
&\qquad - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big(f(X_t) - f(X_{(i-1)\Delta})\big)dV_t^\Delta \\
&\quad= \int_0^T f(X_t)dV_t + A_T + B_T.
\end{aligned}$$

For A_T, note that

$$\begin{aligned}
A_T &= f(X_T)(V_T^\Delta - V_T) - \int_0^T (V_t^\Delta - V_t)df(X_t) - [f(X),(V^\Delta - V)]_T \\
&= A_{1T} - A_{2T} - A_{3T}
\end{aligned}$$

from integration by parts exploiting the notation for the quadratic covariation term. Under suitable conditions, we can show (in Part 2)

$$B_T = O_p(\sqrt{\Delta T}\kappa_\sigma\kappa_{f'}(\nu(T)))$$

and

$$A_{3T} = O_p\big(\sqrt{\Delta T}\kappa_\sigma\kappa_{f'}(\nu(T))\big).$$

For this, note that

$$f(X_t) = f(X_0) + \int_0^t \left(\mu f' + \frac{\sigma^2 f''}{2}\right)(X_s)ds + \int_0^t \sigma f'(X_s)dW_s$$

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

$$\int_0^t \sigma f'(X_s)dW_s$$

and

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

as in the definition of V_t^{Δ}. So we have

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

To obtain its order, note that

$$\begin{aligned}
&\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} f(X_s) \int_{(i-1)\Delta}^{s} dW_u ds \\
=& \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} dW_u ds \\
&+ \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(f(X_s) - f(X_{(i-1)\Delta}) \right) \int_{(i-1)\Delta}^{s} dW_u ds \\
=& A_{1T} + A_{2T}.
\end{aligned}$$

We have

$$A_{1T} = \sum_{i=1}^{n} f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} (i\Delta - u) dW_u$$

and this is a martingale with a quadratic variation bounded by

$$\Delta^2 \sum_{i=1}^{n} f^2(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} du = \Delta^3 \sum_{i=1}^{n} f^2(X_{(i-1)\Delta}) = O_p(\Delta^2 T \kappa_f^2(\nu(T)))$$

from Lemma 1. For A_{2T},

$$\begin{aligned}
A_{2T} =& \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \left(\mu f' + \frac{\sigma^2 f''}{2} \right) (X_u) du \int_{(i-1)\Delta}^{s} dW_u ds \\
&+ \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \sigma f'(X_u) dW_u \int_{(i-1)\Delta}^{s} dW_u ds \\
=& A_{21T} + A_{22T}.
\end{aligned}$$

and

$$
\begin{aligned}
A_{21T} \leq & \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{s} \left| \mu f' + \frac{\sigma^2 f''}{2} \right| (X_u) du ds \\
= & O_p(\Delta T \kappa_\mu \kappa_{f'}(\nu(T))) + O_p(\Delta T \kappa_\sigma^2 \kappa_{f''}(\nu(T)))
\end{aligned}
$$

from Lemma 2. For A_{22T},

$$
A_{22T} \leq \sqrt{\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} \sigma f'(X_u) dW_u \right)^2 ds \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} dW_u \right)^2 ds}.
$$

Note that

$$
\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left(\int_{(i-1)\Delta}^{s} \sigma f'(X_u) dW_u \right)^2 ds = O_p(\Delta T^2 \kappa_\sigma^2 \kappa_{f'}^2(\nu(T)))
$$

since

$$
\int_{(i-1)\Delta}^{s} \sigma f'(X_u) dW_u = O_p(\sqrt{\Delta T} \kappa_\sigma \kappa_{f'}(\nu(T))),
$$

so

$$
A_{22T} = O_p\big(\sqrt{\Delta} T \kappa_\sigma \kappa_{f'}(\nu(T))\big) O_p\big(\sqrt{\Delta T}\big) = O_p(\Delta T^{3/2} \kappa_\sigma \kappa_{f'}(\nu(T))),
$$

and the order of quadratic covariation becomes

$$
\begin{aligned}
[f(X), (V^\Delta - V)]_T = & O_p\big(\sqrt{\Delta}\sqrt{T} \kappa_\sigma \kappa_{f'}(\nu(T))\big) + O_p\big(\sqrt{\Delta} T \kappa_\mu \kappa_\sigma \kappa_{f''}(\nu(T))\big) \\
& + O_p\big(\sqrt{\Delta} T \kappa_\mu \kappa_{\sigma'} \kappa_{f'}(\nu(T))\big) + O_p\big(\sqrt{\Delta} T \kappa_\sigma^2 \kappa_{\sigma'} \kappa_{f''}(\nu(T))\big) + O_p\big(\sqrt{\Delta} T \kappa_\sigma^3 \kappa_{f'''}(\nu(T))\big) \\
& + O_p\big(\sqrt{\Delta} T \kappa_\sigma^2 \kappa_{\sigma''} \kappa_{f'}(\nu(T))\big) + O_p\big(\sqrt{\Delta} T \kappa_\sigma^2 \kappa_{f''}(\nu(T))\big) + O_p\big(\sqrt{\Delta} T \kappa_\sigma \kappa_{\sigma'} \kappa_{f'}(\nu(T))\big) \\
& + O_p\big(\sqrt{\Delta} \sqrt{T \kappa_\sigma^2 \kappa_{f''}}(\nu(T))\big) + O_p\big(\sqrt{\Delta} \sqrt{T \kappa_\sigma \kappa_{\sigma'} \kappa_{f'}}(\nu(T))\big).
\end{aligned}
$$

For A_{1T},

$$f(X_T)(V_T^\Delta - V_T) \approx \frac{2\Delta^{1/4}\bar{V}}{3^{1/4}}\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}f(X_T)$$
$$= O_p\big((\Delta T)^{1/4}\kappa_f(\nu(T))\big)$$

where $\bar{V} \sim \mathbf{N}(0,1)$, and for A_{2T},

$$\int_0^T (V_t^\Delta - V_t)df(X_t) = \int_0^T (V_t^\Delta - V_t)\left(\mu f' + \frac{\sigma^2 f''}{2}\right)(X_t)dt$$
$$+ \int_0^T (V_t^\Delta - V_t)\sigma f'(X_t)dWt$$
$$= A_{21T} + A_{22T}.$$

For A_{21T},

$$A_{21T} = O_p\big(\Delta^{3/8}T^{13/8}(\kappa_\mu\kappa_{f'} + \kappa_\sigma^2\kappa_{f''})(\nu(T))\big)$$

from Lemma 18. For A_{22T},

$$A_{22T} = O_p(\Delta^{1/4}T^{3/4}\kappa_\sigma\kappa_{f'}(\nu(T)))$$

From Lemma 17. Note that A_{22T} cannot be of smaller order than B_T or A_{1T} no matter how, so we need to find the exact asymptotic distribution of A_{22T}.

$$A_{22T} \approx \left(\frac{\sqrt{2}\Delta^{1/4}}{3^{1/4}}\int_0^T \sigma f'(X_t)\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}dU_t\right)$$

Part 2. For B_T,

$$
\begin{aligned}
B_T = & \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \left(\mu f' + \frac{\sigma^2 f''}{2} \right) (X_s) ds dV_t^{\Delta} \\
& + \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma f'(X_s) dW_s dV_t^{\Delta} \\
= & B_{1T} + B_{2T}
\end{aligned}
$$

from Itô's lemma. For B_{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

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

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

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left([V^{\Delta}]_{i\Delta}-[V^{\Delta}]_{s}\right)^{2}ds\right)\\
&\quad=\mathbb{E}\left(\frac{4}{\Delta^{2}}\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left(\int_{s}^{i\Delta}(W_u-W_s)^{2}du\right)^{2}ds\right)\\
&\quad=\mathbb{E}\left(\frac{4}{\Delta^{2}}\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\mathbb{E}_{(i-1)\Delta}\left(\int_{s}^{i\Delta}(W_u-W_s)^{2}du\right)^{2}ds\right).
\end{aligned} \tag{A.22}$$

and since

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

we have

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

and

$$B_{11T}=O_p(\Delta\sqrt{T}).$$

For B_{12T},

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

so

$$B_{1T}=O_p(\Delta\sqrt{T})O_p(\Delta\sqrt{T}\kappa_f^{2}(\nu(T)))=O_p(\Delta\sqrt{T}\kappa_f(\nu(T))).$$

For B_{2T}, note that

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

is a martingale with a quadratic variation

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left(\int_{(i-1)\Delta}^{t} f(X_s)dW_s\right)^2 d[V^{\Delta}]_t \\
&=\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f^2(X_s)dsd[V^{\Delta}]_t \\
&\quad + 2\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t} f(X_s)\int_{(i-1)\Delta}^{s} f(X_u)dW_u dW_s d[V^{\Delta}]_t \\
&=B_{21T} + 2B_{22T}.
\end{aligned}$$

For B_{21T},

$$\begin{aligned}
B_{21T} =&\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big([V^{\Delta}]_{i\Delta} - [V^{\Delta}]_s\big) f^2(X_s)ds \\
\leq&\sqrt{\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big([V^{\Delta}]_{i\Delta} - [V^{\Delta}]_s\big)^2 ds \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta} f^4(X_s)ds.}
\end{aligned}$$

so

$$B_{21T} = O_p(\Delta\sqrt{T})O_p(\sqrt{T}\kappa_f^2(\nu(T))) = O_p(\Delta T\kappa_f^2(\nu(T))).$$

For B_{22T},

$$B_{22T} = \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\big([V^{\Delta}]_{i\Delta} - [V^{\Delta}]_s\big) f(X_s)\int_{(i-1)\Delta}^{s} f(X_u)dW_u dW_s$$

and this is a martingale with a quadratic variation

$$\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left([V^{\Delta}]_{i\Delta}-[V^{\Delta}]_{s}\right)^{2}f^{2}(X_s)\left(\int_{(i-1)\Delta}^{s}f(X_u)dW_u\right)^{2}ds$$
$$\leq\sqrt{\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left([V^{\Delta}]_{i\Delta}-[V^{\Delta}]_{s}\right)^{4}ds\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}f^{4}(X_s)\left(\int_{(i-1)\Delta}^{s}f(X_u)dW_u\right)^{4}ds}.$$

Note that

$$\int_{(i-1)\Delta}^{s}f(X_u)dW_u=O_p(\sqrt{\Delta T}\kappa_f(\nu(T)))$$

since it's a martingale with a quadratic variation

$$\int_{(i-1)\Delta}^{s}f^{2}(X_u)du=O_p(\Delta T\kappa_f^{2}(\nu(T))),$$

and since

$$f(X_s)=O_p(\kappa_f(\nu(T))),$$

we have

$$\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}f^{4}(X_s)\left(\int_{(i-1)\Delta}^{s}f(X_u)dW_u\right)^{4}ds=O_p(\Delta^{2}T^{3}\kappa_f^{8}(\nu(T)))$$

and

$$\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\left([V^{\Delta}]_{i\Delta}-[V^{\Delta}]_{s}\right)^{4}ds=O_p(\Delta^{4}T)$$

since

$$\begin{aligned}\mathbb{E}_{(i-1)\Delta}\left(\int_{s}^{i\Delta}(W_u-W_s)^{2}du\right)^{4}&\leq(i\Delta-s)^{2}\mathbb{E}_{(i-1)\Delta}\left(\int_{s}^{i\Delta}(W_u-W_s)^{4}du\right)^{2}\\&\leq(i\Delta-s)^{3}\int_{s}^{i\Delta}\mathbb{E}_{(i-1)\Delta}(W_u-W_s)^{8}du\\&=21(i\Delta-s)^{8}\end{aligned}$$

with the same way as in (A.22). So

$$B_{22T} = O_p(\Delta\sqrt{T}\kappa_f(\nu(T)))$$

and we can check that B_T has a smaller order than A_T. $\square$

Lemma 20. *For a three times differentiable function f with asymptotically homogeneous derivatives,*

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) = \int_0^T f(X_t)dW_t - \sqrt{\frac{\Delta}{2}}\int_0^T \sigma f'(X_t)dV_t$$
$$+ O_p(\Delta^{3/4}T^{3/4}\kappa_\sigma(\kappa_\sigma\kappa_{f''} + \kappa_{\sigma'}\kappa_{f'})(T^\gamma))$$

Proof. By Itô's lemma,

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
&\qquad = \int_0^T f(X_t)dW_t - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}(f(X_t) - f(X_{(i-1)\Delta}))dW_t \\
&\qquad = \int_0^T f(X_t)dW_t - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t}\sigma f'(X_s)dW_s dW_t \\
&\qquad\qquad - \sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta}\int_{(i-1)\Delta}^{t}(\mu f' + \frac{\sigma^2 f''}{2})(X_s)ds dW_t \\
&\qquad = \int_0^T f(X_t)dW_t - A_T - B_T
\end{aligned}$$

Note that

$$B_T = O_p(\Delta T\kappa_\mu\kappa_{f'}(\nu(T))) + O_p(\Delta T\kappa_\sigma^2\kappa_{f''}(\nu(T)))$$

from Lemma 2 and

$$
\begin{aligned}
A_T &= \sum_{i=1}^{n} \sigma f'(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} dW_s dW_t \\
&\quad + \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(\sigma f'(X_s) - \sigma f'(X_{(i-1)\Delta})\big) dW_s dW_t \\
&= A_{1T} + A_{2T}
\end{aligned}
$$

We have

$$
\begin{aligned}
A_{2T} &= \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \int_{(i-1)\Delta}^{s} \Big[\mu(\sigma f'' + \sigma' f') \\
&\qquad\qquad + \frac{\sigma^2(\sigma f''' + 2\sigma' f'' + \sigma'' f')}{2}\Big](X_u) du dW_s dW_t \\
&\quad + \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \int_{(i-1)\Delta}^{s} \sigma(\sigma f'' + \sigma' f')(X_u) dW_u dW_s dW_t \\
&= O_p(\Delta^{3/2}\sqrt{T}(\kappa_\mu(\kappa_\sigma \kappa_{f''} + \kappa_{\sigma'}\kappa_{f'}) + \kappa_\sigma^2(\kappa_\sigma \kappa_{f'''} + \kappa_{\sigma'}\kappa_{f''} + \kappa_{\sigma''}\kappa_{f'}))(\nu(T))) \\
&\quad + O_p(\Delta\sqrt{T}\kappa_\sigma(\kappa_\sigma \kappa_{f''} + \kappa_{\sigma'}\kappa_{f'})(\nu(T)))
\end{aligned}
$$

from Lemma 2 and

$$
A_{1T} = \sqrt{\frac{\Delta}{2}} \int_0^T \sigma f'(X_t) dV_t + O_p(\Delta^{3/4} T^{3/4} \kappa_\sigma(\kappa_\sigma \kappa_{f''} + \kappa_{\sigma'}\kappa_{f'})(\nu(T)))
$$

from Lemma 19. Thus,

$$
\begin{aligned}
\sum_{i=1}^{n} f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) &= \int_0^T f(X_t) dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \sigma f'(X_t) dV_t \\
&\quad + O_p(\Delta^{3/4} T^{3/4} \kappa_\sigma(\kappa_\sigma \kappa_{f''} + \kappa_{\sigma'}\kappa_{f'})(\nu(T)))
\end{aligned}
$$

$\square$

Lemma 21.

$$\sum_{i=1}^{n} f(X_{(i-1)\Delta})(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))$$
$$= \int_0^T \sigma f(X_t)dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \sigma^2 f^{\cdot}(X_t)dV_t$$
$$+ O_p(\Delta^{3/4}T^{3/4}\kappa_\sigma(\kappa_\sigma^2\kappa_{f''} + \kappa_\sigma\kappa_{\sigma'}\kappa_{f'} + \kappa_\sigma\kappa_{\sigma''}\kappa_f + \kappa_{\sigma'}^2\kappa_f)(\nu(T)))$$

Proof. Note that

$$\begin{aligned}
X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}) =& \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) \\
&+ \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} (\mu\mu^{\cdot} + \frac{\sigma^2\mu^{\cdot\cdot}}{2})(X_s)dsdt \\
&+ \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\mu^{\cdot}(X_s)dW_sdt \\
&+ \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} (\mu\sigma^{\cdot} + \frac{\sigma^2\sigma^{\cdot\cdot}}{2})(X_s)dsdW_t \\
&+ \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\sigma^{\cdot}(X_s)dW_sdW_t.
\end{aligned}$$

We have

$$\sum_{i=1}^{n} \sigma f(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) = \int_0^T \sigma f(X_t)dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \sigma(\sigma^{\cdot} f + \sigma f^{\cdot})(X_t)dV_t$$
$$+ O_p(\Delta^{3/4}T^{3/4}\kappa_\sigma(\kappa_\sigma^2\kappa_{f''} + \kappa_\sigma\kappa_{\sigma'}\kappa_{f'} + \kappa_\sigma\kappa_{\sigma''}\kappa_f + \kappa_{\sigma'}^2\kappa_f)(\nu(T)))$$

from Lemma 20, and

$$\begin{aligned}
\sum f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \sigma\sigma^{\cdot}(X_s) dW_s dW_t \\
&= \sum \sigma\sigma^{\cdot} f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} dW_s dW_t \\
&\quad + \sum f(X_{(i-1)\Delta}) \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} \big(\sigma\sigma^{\cdot}(X_s) - \sigma\sigma^{\cdot}(X_s)\big) dW_s dW_t \\
&= \sqrt{\frac{\Delta}{2}} \int_0^T \sigma\sigma^{\cdot} f(X_t) dV_t + O_p(\Delta\sqrt{T}\kappa_f(\sigma\sigma^{\cdot 2} + \sigma^2\sigma^{\cdot\cdot})(\nu(T)))
\end{aligned}$$

from Lemma 11 and Lemma 2. The rest of the terms can be shown to be of smaller order, thus combining these results, we have the stated result. □

Lemma 22. *For a four times differentiable asymptotically homogeneous function f,*

$$\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta}) \big[(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^2 - \Delta\sigma^2(X_{(i-1)\Delta})\big] \\
&\approx \left[\sqrt{2\Delta} \int_0^T f\sigma^2(X_t) dV_t + \Delta^{3/4} \frac{2}{\sqrt{3}} \int_0^T \sigma(\sigma^2 f' + 2\sigma\sigma' f)(X_t) \sqrt{\big|\sqrt{2}V_t + Z_t\big|} dU_t\right]
\end{aligned}$$

Proof. Denoting $V_{i\Delta}^{\Delta} - V_{(i-1)\Delta}^{\Delta} = \int_{(i-1)\Delta}^{i\Delta} \int_{(i-1)\Delta}^{t} dW_s dW_t$, we can write as

$$\begin{aligned}
&X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}) \\
&\quad = \sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + \sigma\sigma'(X_{(i-1)\Delta})(V_{i\Delta}^{\Delta} - V_{(i-1)\Delta}^{\Delta}) + R_i,
\end{aligned}$$

where R_i is a remainer term, from the first equation of the proof of Lemma 3. Replacing this into the following, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[(X_{i\Delta} - X_{(i-1)\Delta} - \Delta\mu(X_{(i-1)\Delta}))^2 - \Delta\sigma^2(X_{(i-1)\Delta})\big] \\
&\quad = \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[(\sigma(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta}) + \sigma\sigma'(X_{(i-1)\Delta})(V_{i\Delta} - V_{(i-1)\Delta}) + R_i)^2 \\
&\qquad\qquad\qquad - \Delta\sigma^2(X_{(i-1)\Delta})\big] \\
&\quad = \sum_{i=1}^{n} f(X_{(i-1)\Delta})\big[\sigma^2(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})^2 \\
&\qquad\qquad\qquad + 2\sigma^2\sigma'(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta}) \\
&\qquad\qquad\qquad + \sigma^2\sigma'^2(X_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta})^2 + R_i^* \\
&\qquad\qquad\qquad - \Delta\sigma^2(X_{(i-1)\Delta})\big],
\end{aligned}
$$

where R_i^* denotes the terms multiplied by R_i. From Lemma 19,

$$
\begin{aligned}
&\sum_{i=1}^{n} f\sigma^2(X_{(i-1)\Delta})[(W_{i\Delta} - W_{(i-1)\Delta})^2 - \Delta] \qquad \text{(A.23)} \\
&\approx \left[\sqrt{2\Delta}\int_0^T f\sigma^2(X_t)dV_t + \frac{2\Delta^{3/4}}{3^{1/4}}\int_0^T \sigma(\sigma^2 f' + 2\sigma\sigma' f)(X_t)\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}dU_t\right]
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{i=1}^{n} f\sigma^2\sigma'(X_{(i-1)\Delta})(W_{i\Delta} - W_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta}) &= O_p(\Delta T\kappa_f\kappa_\sigma^2\kappa_{\sigma'}(\nu(T))) \\
\sum_{i=1}^{n} f\sigma^2\sigma'^2(X_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta})^2 &= O_p(\Delta T\kappa_f\kappa_\sigma^2\kappa_{\sigma'}^2(\nu(T)))
\end{aligned}
$$

by the same steps in the proof of Lemma 3 from the independent increments of the Brownian motion and $\mathbb{E}\big((W_{i\Delta} - W_{(i-1)\Delta})(V^{\Delta}_{i\Delta} - V^{\Delta}_{(i-1)\Delta})\big) = 0$, and the remainder term $\sum_{i=1}^{n} f(X_{(i-1)\Delta})R_i^*$ can be also shown to be of smaller order by Lemma 3. $\square$

D. Asymptotic Expansions of the Log-Likelihood Derivatives

1. Euler ML Estimator Asymptotics

For the scores of the Euler approximated log-likelihood function, we have

$$\begin{aligned}
\mathcal{S}_\alpha(\theta) = \sum_{i=1}^{n} \ell_\alpha(x,y) &\approx \left[\int_0^T \frac{\mu_\alpha}{\sigma}(X_t)dW_t - \sqrt{\frac{\Delta}{2}}\int_0^T \left(\mu_\alpha^{\cdot} - \frac{\mu_\alpha\sigma^{\cdot}}{\sigma}\right)(X_t)dV_t\right] \\
&= \left[S_{\alpha,1} + \sqrt{\Delta}S_{\alpha,2}\right] \\
\mathcal{S}_\beta(\theta) = \sum_{i=1}^{n} \ell_\beta(x,y) &\approx \Bigg[\sqrt{\frac{2}{\Delta}}\int_0^T \frac{\sigma_\beta}{\sigma}(X_t)dV_t \\
&\qquad + \frac{2}{(3\Delta)^{1/4}}\int_0^T \left(\sigma_\beta^{\cdot} - \frac{\sigma_\beta\sigma^{\cdot}}{\sigma}\right)(X_t)\sqrt{\left|\sqrt{\frac{2}{3}}V_t + \frac{1}{\sqrt{3}}Z_t\right|}dU_t\Bigg] \\
&= \left[\frac{1}{\sqrt{\Delta}}S_{\beta,1} + \frac{1}{\Delta^{1/4}}S_{\beta,2}\right]
\end{aligned}$$

and for the Hessians, we have

$$\begin{aligned}
\mathcal{H}_{\alpha\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'}(x,y) &\approx \Bigg[-\int_0^T \frac{\mu_\alpha\mu_\alpha'}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t)dW_t \\
&\qquad - \sqrt{\frac{\Delta}{2}}\int_0^T \left(\mu_{\alpha\alpha'}^{\cdot} - \frac{\mu_{\alpha\alpha'}\sigma^{\cdot}}{\sigma}\right)(X_t)dV_t\Bigg] \\
&= \left[H_{\alpha\alpha,1} + \sqrt{\Delta}H_{\alpha\alpha,2}\right] \\
\mathcal{H}_{\alpha\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'}(x,y) &\approx \Bigg[-2\int_0^T \frac{\mu_\alpha\sigma_\beta'}{\sigma^2}(X_t)dW_t \\
&\qquad + \sqrt{2\Delta}\int_0^T \left(\frac{2\mu_\alpha\sigma_\beta'\sigma^{\cdot}}{\sigma^2} - \frac{\mu_\alpha^{\cdot}\sigma_\beta'}{\sigma} - \frac{\mu_\alpha\sigma_\beta^{\cdot\prime}}{\sigma}\right)(X_t)dV_t\Bigg] \\
&= \left[H_{\alpha\beta,1} + \sqrt{\Delta}H_{\alpha\beta,2}\right] \\
\mathcal{H}_{\beta\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'}(x,y) &\approx \left[H_{\alpha\beta,1}' + \sqrt{\Delta}H_{\alpha\beta,2}'\right]
\end{aligned}$$

$$
\begin{aligned}
\mathcal{H}_{\beta\beta'}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\beta'}(x,y) \\
&\approx \left[-\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta\sigma_\beta'}{\sigma^2}(X_t)dt + \sqrt{\frac{2}{\Delta}}\int_0^T \left(\frac{\sigma_{\beta\beta'}}{\sigma} - \frac{3\sigma_\beta\sigma_\beta'}{\sigma^2}\right)(X_t)dV_t \right] \\
&= \left[\frac{1}{\Delta}H_{\beta\beta,1} + \frac{1}{\sqrt{\Delta}}H_{\beta\beta,2}\right].
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\alpha}(x,y) &= -3\int_0^T \frac{\mu_\alpha \diamond \mu_{\alpha\alpha'}}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'\otimes\alpha}}{\sigma}(X_t)dW_t \\
&\quad + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big) \\
&= J_{\alpha\alpha\alpha,1} + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big)
\end{aligned}
$$

where

$$
\mu_\alpha \diamond \mu_{\alpha\alpha'} = \big(\mu_\alpha \otimes \mu_{\alpha\alpha'} + \mu_{\alpha\alpha'} \otimes \mu_\alpha + \mu_\alpha' \otimes \mathrm{vec}(\mu_{\alpha\alpha'})\big)/3.
$$

Also,

$$
\begin{aligned}
&\mathcal{J}_{\alpha\alpha'\otimes\beta}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\beta}(x,y) = 2\int_0^T \left(\frac{\mu_\alpha\mu_\alpha' \otimes \sigma_\beta}{\sigma^3}\right)(X_t)dt - 2\int_0^T \frac{\mu_{\alpha\alpha'} \otimes \sigma_\beta}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\mu^{\cdot}_{\alpha\alpha'}} \otimes \kappa_{\sigma_\beta} + \kappa_{\sigma^{\cdot}}\kappa_{\mu_{\alpha\alpha'}} \otimes \kappa_{\sigma_\beta} + \kappa_\sigma\kappa_{\mu_{\alpha\alpha'}} \otimes \kappa_{\sigma^{\cdot}_\beta})/\kappa_\sigma^2(\nu(T))) \\
&= J_{\alpha\alpha\beta,1} + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\mu^{\cdot}_{\alpha\alpha'}} \otimes \kappa_{\sigma_\beta} + \kappa_{\sigma^{\cdot}}\kappa_{\mu_{\alpha\alpha'}} \otimes \kappa_{\sigma_\beta} + \kappa_\sigma\kappa_{\mu_{\alpha\alpha'}} \otimes \kappa_{\sigma^{\cdot}_\beta})/\kappa_\sigma^2(\nu(T))) \\
&\mathcal{J}_{\beta\alpha'\otimes\alpha}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'\otimes\alpha}(x,y) = 2\int_0^T \left(\frac{\sigma_\beta \otimes \mu_\alpha\mu_\alpha'}{\sigma^3}\right)(X_t)dt - 2\int_0^T \frac{\sigma_\beta \otimes \mu_{\alpha\alpha'}}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\sigma_\beta} \otimes \kappa_{\mu^{\cdot}_{\alpha\alpha'}} + \kappa_{\sigma^{\cdot}}\kappa_{\sigma_\beta} \otimes \kappa_{\mu_{\alpha\alpha'}} + \kappa_\sigma\kappa_{\sigma^{\cdot}_\beta} \otimes \kappa_{\mu_{\alpha\alpha'}})/\kappa_\sigma^2(\nu(T))) \\
&= J^{\triangledown}_{\alpha\alpha\beta,1} + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\sigma_\beta} \otimes \kappa_{\mu^{\cdot}_{\alpha\alpha'}} + \kappa_{\sigma^{\cdot}}\kappa_{\sigma_\beta} \otimes \kappa_{\mu_{\alpha\alpha'}} + \kappa_\sigma\kappa_{\sigma^{\cdot}_\beta} \otimes \kappa_{\mu_{\alpha\alpha'}})/\kappa_\sigma^2(\nu(T)))
\end{aligned}
$$

$$\mathcal{J}_{\alpha\beta'\otimes\alpha}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'\otimes\alpha}(x,y) = 2\int_0^T \left(\frac{\sigma'_\beta \otimes \mathrm{vec}(\mu_\alpha \mu'_\alpha)}{\sigma^3}\right)(X_t)dt$$

$$-2\int_0^T \frac{\sigma'_\beta \otimes \mathrm{vec}(\mu_{\alpha\alpha'})}{\sigma^2}(X_t)dW_t$$

$$+O_p(\sqrt{\Delta T}(\kappa_\sigma \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu^\cdot_{\alpha\alpha'}}) + \kappa_{\sigma^\cdot}\kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}) + \kappa_\sigma \kappa'_{\sigma^\cdot_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}))/\kappa^2_\sigma(\nu(T)))$$

$$= J^\Delta_{\alpha\alpha\beta,1}$$

$$+O_p(\sqrt{\Delta T}(\kappa_\sigma \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu^\cdot_{\alpha\alpha'}}) + \kappa_{\sigma^\cdot}\kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}) + \kappa_\sigma \kappa'_{\sigma^\cdot_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}))/\kappa^2_\sigma(\nu(T)))$$

$$\mathcal{J}_{\beta\beta'\otimes\alpha}(\theta) = \sum_{i=1}^{n} \ell_{\beta\beta'\otimes\alpha}(x,y) = \int_0^T \left(\frac{6\sigma_\beta\sigma'_\beta \otimes \mu_\alpha}{\sigma^3} - \frac{2\sigma_{\beta\beta'} \otimes \mu_\alpha}{\sigma^2}\right)(X_t)dW_t$$

$$+O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} \otimes \kappa_{\mu_\alpha} + \kappa_\sigma(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} \otimes \kappa_{\mu^\cdot_\alpha} + \kappa_{\sigma^\cdot}\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu_\alpha} + \kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta} \otimes \kappa_{\mu_\alpha})$$

$$+\kappa^2_\sigma(\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu^\cdot_\alpha} + \kappa_{\sigma^\cdot_{\beta\beta'}} \otimes \kappa_{\mu_\alpha}))(\nu(T)))$$

$$= J_{\alpha\beta\beta,1}$$

$$+O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} \otimes \kappa_{\mu_\alpha} + \kappa_\sigma(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} \otimes \kappa_{\mu^\cdot_\alpha} + \kappa_{\sigma^\cdot}\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu_\alpha} + \kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta} \otimes \kappa_{\mu_\alpha})$$

$$+\kappa^2_\sigma(\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu^\cdot_\alpha} + \kappa_{\sigma^\cdot_{\beta\beta'}} \otimes \kappa_{\mu_\alpha}))(\nu(T)))$$

$$\mathcal{J}_{\alpha\beta'\otimes\beta}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'\otimes\beta}(x,y) = \int_0^T \left(\frac{6\mu_\alpha \otimes \sigma_\beta\sigma'_\beta}{\sigma^3} - \frac{2\mu_\alpha \otimes \sigma_{\beta\beta'}}{\sigma^2}\right)(X_t)dW_t$$

$$+O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_\sigma(\kappa_{\mu^\cdot_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_{\sigma^\cdot}\kappa_{\mu_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})$$

$$+\kappa^2_\sigma(\kappa_{\mu^\cdot_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma^\cdot_{\beta\beta'}}))(\nu(T)))$$

$$= J^\Delta_{\alpha\beta\beta,1}$$

$$+O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_\sigma(\kappa_{\mu^\cdot_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_{\sigma^\cdot}\kappa_{\mu_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})$$

$$+\kappa^2_\sigma(\kappa_{\mu^\cdot_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma^\cdot_{\beta\beta'}}))(\nu(T)))$$

$$
\begin{aligned}
\mathcal{J}_{\beta\alpha'\otimes\beta}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\alpha'\otimes\beta}(x,y) = \int_0^T \left(\frac{6\mu'_\alpha \otimes \mathrm{vec}(\sigma_\beta\sigma'_\beta)}{\sigma^3} - \frac{2\mu'_\alpha \otimes \mathrm{vec}(\sigma_{\beta\beta'})}{\sigma^2} \right)(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_\sigma(\kappa'_{\mu^\cdot_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) \\
&\qquad + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})) + \kappa^2_\sigma(\kappa'_{\mu^\cdot_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma^\cdot_{\beta\beta'}})))(\nu(T))) \\
&= J^{\nabla}_{\alpha\beta\beta,1} \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_\sigma(\kappa'_{\mu^\cdot_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) \\
&\qquad + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})) + \kappa^2_\sigma(\kappa'_{\mu^\cdot_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma^\cdot_{\beta\beta'}})))(\nu(T)))
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{J}_{\beta\beta'\otimes\beta}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\beta'\otimes\beta}(x,y) = \frac{1}{\Delta}\int_0^T \left(\frac{10\sigma_\beta\sigma'_\beta \otimes \sigma_\beta}{\sigma^3} - \frac{6\sigma_\beta \diamond \sigma_{\beta\beta'}}{\sigma^2} \right)(X_t)dt \\
&\quad + O_p(\sqrt{T/\Delta}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}\otimes \kappa_{\sigma_\beta} + \kappa^2_\sigma\kappa_{\sigma_{\beta\beta'\otimes\beta}} + \kappa_\sigma\kappa_{\sigma_\beta}\diamond\kappa_{\sigma_{\beta\beta'}})/\kappa^3_\sigma(\nu(T))) \\
&= \frac{1}{\Delta}J_{\beta\beta\beta,1} + O_p(\sqrt{T/\Delta}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}\otimes \kappa_{\sigma_\beta} + \kappa^2_\sigma\kappa_{\sigma_{\beta\beta'\otimes\beta}} + \kappa_\sigma\kappa_{\sigma_\beta}\diamond\kappa_{\sigma_{\beta\beta'}})/\kappa^3_\sigma(\nu(T))).
\end{aligned}
$$

where

$$
\sigma_\beta \diamond \sigma_{\beta\beta'} = \big(\sigma_\beta \otimes \sigma_{\beta\beta'} + \sigma_{\beta\beta'} \otimes \sigma_\beta + \sigma'_\beta \otimes \mathrm{vec}(\sigma_{\beta\beta'})\big)/3.
$$

Lastly,

$$
\begin{aligned}
\mathcal{K}_{\alpha\alpha'\otimes\alpha\alpha'}(\theta) &= \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\alpha\alpha'}(x,y) = -\int_0^T \frac{3\mu_{\alpha\alpha'} \otimes \mu_{\alpha\alpha'} + 4\mu_\alpha \dot{\diamond} \mu_{\alpha\alpha'\otimes\alpha}}{\sigma^2}(X_t)dt \\
&\quad + \int_0^T \frac{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}{\sigma}(X_t)dW_t + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^\cdot_{\alpha\alpha'\otimes\alpha\alpha'}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}\kappa_{\sigma^\cdot}/\kappa_\sigma)(\nu(T))\big) \\
&= K_{\alpha\alpha\alpha\alpha,1} + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^\cdot_{\alpha\alpha'\otimes\alpha\alpha'}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}\kappa_{\sigma^\cdot}/\kappa_\sigma)(\nu(T))\big).
\end{aligned}
$$

where

$$
\mu_\alpha \dot{\diamond} \mu_{\alpha\alpha'\otimes\alpha} = \big(\mu'_\alpha \otimes \mu_{\alpha\alpha'\otimes\alpha} + \mu_{\alpha\alpha'\otimes\alpha} \otimes \mu'_\alpha + \mu_\alpha \otimes \mu'_{\alpha\alpha'\otimes\alpha} + \mu'_{\alpha\alpha'\otimes\alpha} \otimes \mu_\alpha\big)/4.
$$

2. Milstein ML Estimator Asymptotics

For the scores of the Milstein approximated log-likelihood function, we have

$$\begin{aligned}
\mathcal{S}_\alpha(\theta) = \sum_{i=1}^{n} \ell_\alpha(x,y) &\approx \left[\int_0^T \frac{\mu_\alpha}{\sigma}(X_t) dW_t - \sqrt{\frac{\Delta}{2}} \int_0^T \left(\mu_\alpha^{\cdot} + \frac{2\mu_\alpha \sigma^{\cdot}}{\sigma} \right) (X_t) dV_t \right] \\
&= \left[S_{\alpha,1} + \sqrt{\Delta} S_{\alpha,2} \right] \\
\mathcal{S}_\beta(\theta) = \sum_{i=1}^{n} \ell_\beta(x,y) &\approx \Bigg[\sqrt{\frac{2}{\Delta}} \int_0^T \frac{\sigma_\beta}{\sigma}(X_t) dV_t \\
&\quad + \frac{2}{(3\Delta)^{1/4}} \int_0^T \left(\sigma_\beta^{\cdot} - \frac{\sigma_\beta \sigma^{\cdot}}{\sigma} \right) (X_t) \sqrt{\left| \sqrt{\frac{2}{3}} V_t + \frac{1}{\sqrt{3}} Z_t \right|} dU_t \Bigg] \\
&= \left[\frac{1}{\sqrt{\Delta}} S_{\beta,1} + \frac{1}{\Delta^{1/4}} S_{\beta,2} \right]
\end{aligned}$$

and for the Hessians, we have

$$\begin{aligned}
\mathcal{H}_{\alpha\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'}(x,y) &\approx \Bigg[- \int_0^T \frac{\mu_\alpha \mu_\alpha'}{\sigma^2}(X_t) dt + \int_0^T \frac{\mu_{\alpha\alpha'}}{\sigma}(X_t) dW_t \\
&\quad - \sqrt{\frac{\Delta}{2}} \int_0^T \left(\mu_{\alpha\alpha'}^{\cdot} + \frac{2\mu_{\alpha\alpha'} \sigma^{\cdot}}{\sigma} \right) (X_t) dV_t \Bigg] \\
&= \left[H_{\alpha\alpha,1} + \sqrt{\Delta} H_{\alpha\alpha,2} \right] \\
\mathcal{H}_{\alpha\beta'}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\beta'}(x,y) &\approx \Bigg[- 2 \int_0^T \frac{\mu_\alpha \sigma_\beta'}{\sigma^2}(X_t) dW_t + 3 \int_0^T \frac{\mu_\alpha \sigma_\beta' \sigma^{\cdot}}{\sigma^2}(X_t) dt \\
&\quad + \sqrt{\frac{\Delta}{2}} \int_0^T \left(\frac{2\mu_\alpha^{\cdot} \sigma_\beta'}{\sigma} - \frac{\mu_\alpha \sigma_\beta^{\cdot\prime}}{\sigma} - \frac{\mu_\alpha \sigma_\beta' \sigma^{\cdot}}{\sigma^2} \right) (X_t) dV_t \Bigg] \\
&= \left[H_{\alpha\beta,1} + \sqrt{\Delta} H_{\alpha\beta,2} \right] \\
\mathcal{H}_{\beta\alpha'}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'}(x,y) &\approx \left[H_{\alpha\beta,1}' + \sqrt{\Delta} H_{\alpha\beta,2}' \right]
\end{aligned}$$

$$\begin{aligned}
\mathcal{H}_{\beta\beta'}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\beta'}(x,y) \\
&\approx \left[-\frac{2}{\Delta}\int_0^T \frac{\sigma_\beta \sigma_\beta'}{\sigma^2}(X_t)dt + \sqrt{\frac{2}{\Delta}}\int_0^T \left(\frac{\sigma_{\beta\beta'}}{\sigma} - \frac{3\sigma_\beta\sigma_\beta'}{\sigma^2}\right)(X_t)dV_t \right] \\
&= \left[\frac{1}{\Delta} H_{\beta\beta,1} + \frac{1}{\sqrt{\Delta}} H_{\beta\beta,2} \right].
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathcal{J}_{\alpha\alpha'\otimes\alpha}(\theta) = \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\alpha}(x,y) &= -3\int_0^T \frac{\mu_\alpha \diamond \mu_{\alpha\alpha'}}{\sigma^2}(X_t)dt + \int_0^T \frac{\mu_{\alpha\alpha'\otimes\alpha}}{\sigma}(X_t)dW_t \\
&\quad + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big) \\
&= J_{\alpha\alpha\alpha,1} + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big),
\end{aligned}$$

where

$$\mu_\alpha \diamond \mu_{\alpha\alpha'} = \big(\mu_\alpha \otimes \mu_{\alpha\alpha'} + \mu_{\alpha\alpha'} \otimes \mu_\alpha + \mu_\alpha' \otimes \mathrm{vec}(\mu_{\alpha\alpha'})\big)/3.$$

Also,

$$\begin{aligned}
\mathcal{J}_{\alpha\alpha'\otimes\beta}(\theta) &= \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\beta}(x,y) = \int_0^T \left(\frac{2\mu_\alpha\mu_\alpha' \otimes \sigma_\beta}{\sigma^3} + \frac{3\mu_{\alpha\alpha'} \otimes \sigma_\beta\sigma^{\cdot}}{\sigma^2} \right)(X_t)dt \\
&\quad - 2\int_0^T \frac{\mu_{\alpha\alpha'} \otimes \sigma_\beta}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\mu^{\cdot}_{\alpha\alpha'}}\otimes \kappa_{\sigma_\beta} + \kappa_{\sigma^{\cdot}}\kappa_{\mu_{\alpha\alpha'}}\otimes \kappa_{\sigma_\beta} + \kappa_\sigma\kappa_{\mu_{\alpha\alpha'}}\otimes \kappa_{\sigma^{\cdot}_\beta})/\kappa_\sigma^2(\nu(T))) \\
&= J_{\alpha\alpha\beta,1} + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\mu^{\cdot}_{\alpha\alpha'}}\otimes \kappa_{\sigma_\beta} + \kappa_{\sigma^{\cdot}}\kappa_{\mu_{\alpha\alpha'}}\otimes \kappa_{\sigma_\beta} + \kappa_\sigma\kappa_{\mu_{\alpha\alpha'}}\otimes \kappa_{\sigma^{\cdot}_\beta})/\kappa_\sigma^2(\nu(T))) \\
\mathcal{J}_{\beta\alpha'\otimes\alpha}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\alpha'\otimes\alpha}(x,y) = \int_0^T \left(\frac{2\sigma_\beta \otimes \mu_\alpha\mu_\alpha'}{\sigma^3} + \frac{3\sigma_\beta \otimes \mu_{\alpha\alpha'}\sigma^{\cdot}}{\sigma^2} \right)(X_t)dt \\
&\quad - 2\int_0^T \frac{\sigma_\beta \otimes \mu_{\alpha\alpha'}}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\sigma_\beta}\otimes \kappa_{\mu^{\cdot}_{\alpha\alpha'}} + \kappa_{\sigma^{\cdot}}\kappa_{\sigma_\beta}\otimes \kappa_{\mu_{\alpha\alpha'}} + \kappa_\sigma\kappa_{\sigma^{\cdot}_\beta}\otimes \kappa_{\mu_{\alpha\alpha'}})/\kappa_\sigma^2(\nu(T))) \\
&= J^{\nabla}_{\alpha\alpha\beta,1} + O_p(\sqrt{\Delta T}(\kappa_\sigma\kappa_{\sigma_\beta}\otimes \kappa_{\mu^{\cdot}_{\alpha\alpha'}} + \kappa_{\sigma^{\cdot}}\kappa_{\sigma_\beta}\otimes \kappa_{\mu_{\alpha\alpha'}} + \kappa_\sigma\kappa_{\sigma^{\cdot}_\beta}\otimes \kappa_{\mu_{\alpha\alpha'}})/\kappa_\sigma^2(\nu(T)))
\end{aligned}$$

$$
\begin{aligned}
\mathcal{J}_{\alpha\beta'\otimes\alpha}(\theta) &= \sum_{i=1}^{n} \ell_{\alpha\beta'\otimes\alpha}(x,y) = \int_0^T \left(\frac{2\sigma'_\beta \otimes \mathrm{vec}(\mu_\alpha \mu'_\alpha)}{\sigma^3} + \frac{3\sigma'_\beta \otimes \mathrm{vec}(\mu_{\alpha\alpha'})\sigma^{\cdot}}{\sigma^2} \right)(X_t)dt \\
&\quad - 2\int_0^T \frac{\sigma'_\beta \otimes \mathrm{vec}(\mu_{\alpha\alpha'})}{\sigma^2}(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu^{\cdot}_{\alpha\alpha'}}) + \kappa_{\sigma^{\cdot}} \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}) + \kappa_\sigma \kappa'_{\sigma^{\cdot}_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}))/\kappa^2_\sigma(\nu(T))) \\
&= J^{\Delta}_{\alpha\alpha\beta,1} \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_\sigma \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu^{\cdot}_{\alpha\alpha'}}) + \kappa_{\sigma^{\cdot}} \kappa'_{\sigma_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}) + \kappa_\sigma \kappa'_{\sigma^{\cdot}_\beta} \otimes \mathrm{vec}(\kappa_{\mu_{\alpha\alpha'}}))/\kappa^2_\sigma(\nu(T)))
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{J}_{\beta\beta'\otimes\alpha}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\beta'\otimes\alpha}(x,y) \\
&= \int_0^T \left(\frac{3\sigma_\beta \sigma^{\cdot\prime}_\beta \otimes \mu_\alpha}{\sigma^2} + \frac{3\sigma^{\cdot}_\beta \sigma'_\beta \otimes \mu_\alpha}{\sigma^2} + \frac{3\sigma_{\beta\beta'} \otimes \mu_\alpha \sigma^{\cdot}}{\sigma^2} - \frac{15\sigma_\beta \sigma'_\beta \otimes \mu_\alpha \sigma^{\cdot}}{\sigma^3} \right)(X_t)dt \\
&\quad + \int_0^T \left(\frac{6\sigma_\beta \sigma'_\beta \otimes \mu_\alpha}{\sigma^3} - \frac{2\sigma_{\beta\beta'} \otimes \mu_\alpha}{\sigma^2} \right)(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^{\cdot}} \kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} \otimes \kappa_{\mu_\alpha} + \kappa_\sigma(\kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} \otimes \kappa_{\mu^{\cdot}_\alpha} + \kappa_{\sigma^{\cdot}} \kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu_\alpha} + \kappa_{\sigma_\beta} \kappa'_{\sigma^{\cdot}_\beta} \otimes \kappa_{\mu_\alpha}) \\
&\qquad + \kappa^2_\sigma(\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu^{\cdot}_\alpha} + \kappa_{\sigma^{\cdot}_{\beta\beta'}} \otimes \kappa_{\mu_\alpha}))(\nu(T))) \\
&= J_{\alpha\beta\beta,1} \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^{\cdot}} \kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} \otimes \kappa_{\mu_\alpha} + \kappa_\sigma(\kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} \otimes \kappa_{\mu^{\cdot}_\alpha} + \kappa_{\sigma^{\cdot}} \kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu_\alpha} + \kappa_{\sigma_\beta} \kappa'_{\sigma^{\cdot}_\beta} \otimes \kappa_{\mu_\alpha}) \\
&\qquad + \kappa^2_\sigma(\kappa_{\sigma_{\beta\beta'}} \otimes \kappa_{\mu^{\cdot}_\alpha} + \kappa_{\sigma^{\cdot}_{\beta\beta'}} \otimes \kappa_{\mu_\alpha}))(\nu(T)))
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{J}_{\alpha\beta'\otimes\beta}(\theta) &= \sum_{i=1}^{n} \ell_{\alpha\beta'\otimes\beta}(x,y) \\
&= \int_0^T \left(\frac{3\mu_\alpha \otimes \sigma_\beta \sigma^{\cdot\prime}_\beta}{\sigma^2} + \frac{3\mu_\alpha \otimes \sigma^{\cdot}_\beta \sigma'_\beta}{\sigma^2} + \frac{3\mu_\alpha \otimes \sigma_{\beta\beta'} \sigma^{\cdot}}{\sigma^2} - \frac{15\mu_\alpha \otimes \sigma_\beta \sigma'_\beta \sigma^{\cdot}}{\sigma^3} \right)(X_t)dt \\
&\quad + \int_0^T \left(\frac{6\mu_\alpha \otimes \sigma_\beta \sigma'_\beta}{\sigma^3} - \frac{2\mu_\alpha \otimes \sigma_{\beta\beta'}}{\sigma^2} \right)(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^{\cdot}} \kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} + \kappa_\sigma(\kappa_{\mu^{\cdot}_\alpha} \otimes \kappa_{\sigma_\beta} \kappa'_{\sigma_\beta} + \kappa_{\sigma^{\cdot}} \kappa_{\mu_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma_\beta} \kappa'_{\sigma^{\cdot}_\beta}) \\
&\qquad + \kappa^2_\sigma(\kappa_{\mu^{\cdot}_\alpha} \otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha} \otimes \kappa_{\sigma^{\cdot}_{\beta\beta'}}))(\nu(T)))
\end{aligned}
$$

$$
\begin{aligned}
&= J^{\Delta}_{\alpha\beta\beta,1} \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa_{\mu_\alpha}\otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_\sigma(\kappa_{\mu_\alpha^\cdot}\otimes \kappa_{\sigma_\beta}\kappa'_{\sigma_\beta} + \kappa_{\sigma^\cdot}\kappa_{\mu_\alpha}\otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha}\otimes \kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta}) \\
&\qquad + \kappa^2_\sigma(\kappa_{\mu_\alpha^\cdot}\otimes \kappa_{\sigma_{\beta\beta'}} + \kappa_{\mu_\alpha}\otimes \kappa_{\sigma^\cdot_{\beta\beta'}}))(\nu(T))) \\
&\mathcal{J}_{\beta\alpha'\otimes\beta}(\theta) = \sum_{i=1}^{n} \ell_{\beta\alpha'\otimes\beta}(x,y) \\
&= \int_0^T \Bigg(\frac{3\mu'_\alpha \otimes \mathrm{vec}(\sigma_\beta\sigma^{\cdot\prime}_\beta)}{\sigma^2} + \frac{3\mu'_\alpha \otimes \mathrm{vec}(\sigma^\cdot_\beta\sigma'_\beta)}{\sigma^2} + \frac{3\mu'_\alpha \otimes \mathrm{vec}(\sigma_{\beta\beta'})\sigma^\cdot}{\sigma^2} \\
&\qquad - \frac{15\mu'_\alpha \otimes \mathrm{vec}(\sigma_\beta\sigma'_\beta)\sigma^\cdot}{\sigma^3}\Bigg)(X_t)dt \\
&\quad + \int_0^T \left(\frac{6\mu'_\alpha \otimes \mathrm{vec}(\sigma_\beta\sigma'_\beta)}{\sigma^3} - \frac{2\mu'_\alpha \otimes \mathrm{vec}(\sigma_{\beta\beta'})}{\sigma^2}\right)(X_t)dW_t \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_\sigma(\kappa'_{\mu_\alpha^\cdot}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) \\
&\qquad + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})) + \kappa^2_\sigma(\kappa'_{\mu_\alpha^\cdot}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma^\cdot_{\beta\beta'}})))(\nu(T))) \\
&= J^{\nabla}_{\alpha\beta\beta,1} \\
&\quad + O_p(\sqrt{\Delta T}(\kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_\sigma(\kappa'_{\mu_\alpha^\cdot}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}) + \kappa_{\sigma^\cdot}\kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) \\
&\qquad + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma_\beta}\kappa'_{\sigma^\cdot_\beta})) + \kappa^2_\sigma(\kappa'_{\mu_\alpha^\cdot}\otimes \mathrm{vec}(\kappa_{\sigma_{\beta\beta'}}) + \kappa'_{\mu_\alpha}\otimes \mathrm{vec}(\kappa_{\sigma^\cdot_{\beta\beta'}})))(\nu(T)))
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{J}_{\beta\beta'\otimes\beta}(\theta) &= \sum_{i=1}^{n} \ell_{\beta\beta'\otimes\beta}(x,y) = \frac{1}{\Delta}\int_0^T \left(\frac{10\sigma_\beta\sigma'_\beta \otimes \sigma_\beta}{\sigma^3} - \frac{6\sigma_\beta \diamond \sigma_{\beta\beta'}}{\sigma^2}\right)(X_t)dt \\
&\quad + O_p(\sqrt{T/\Delta}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}\otimes \kappa_{\sigma_\beta} + \kappa^2_\sigma\kappa_{\sigma_{\beta\beta'\otimes\beta}} + \kappa_\sigma\kappa_{\sigma_\beta}\diamond\kappa_{\sigma_{\beta\beta'}})/\kappa^3_\sigma(\nu(T))) \\
&= \frac{1}{\Delta}J_{\beta\beta\beta,1} + O_p(\sqrt{T/\Delta}(\kappa_{\sigma_\beta}\kappa'_{\sigma_\beta}\otimes \kappa_{\sigma_\beta} + \kappa^2_\sigma\kappa_{\sigma_{\beta\beta'\otimes\beta}} + \kappa_\sigma\kappa_{\sigma_\beta}\diamond\kappa_{\sigma_{\beta\beta'}})/\kappa^3_\sigma(\nu(T))).
\end{aligned}
$$

where

$$
\sigma_\beta \diamond \sigma_{\beta\beta'} = \big(\sigma_\beta \otimes \sigma_{\beta\beta'} + \sigma_{\beta\beta'} \otimes \sigma_\beta + \sigma'_\beta \otimes \mathrm{vec}(\sigma_{\beta\beta'})\big)/3.
$$

Lastly,

$$\begin{aligned}\mathcal{K}_{\alpha\alpha'\otimes\alpha\alpha'}(\theta) &= \sum_{i=1}^{n} \ell_{\alpha\alpha'\otimes\alpha\alpha'}(x,y) = -\int_0^T \frac{3\mu_{\alpha\alpha'}\otimes\mu_{\alpha\alpha'} + 4\mu_\alpha\dot{\diamond}\mu_{\alpha\alpha'\otimes\alpha}}{\sigma^2}(X_t)dt \\ &+ \int_0^T \frac{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}{\sigma}(X_t)dW_t + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha\alpha'}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big) \\ &= K_{\alpha\alpha\alpha\alpha,1} + O_p\big(\sqrt{\Delta T}(\kappa_{\mu^{\cdot}_{\alpha\alpha'\otimes\alpha\alpha'}} + \kappa_{\mu_{\alpha\alpha'\otimes\alpha\alpha'}}\kappa_{\sigma^{\cdot}}/\kappa_\sigma)(\nu(T))\big).\end{aligned}$$

where

$$\mu_\alpha\dot{\diamond}\mu_{\alpha\alpha'\otimes\alpha} = \big(\mu'_\alpha\otimes\mu_{\alpha\alpha'\otimes\alpha} + \mu_{\alpha\alpha'\otimes\alpha}\otimes\mu'_\alpha + \mu_\alpha\otimes\mu'_{\alpha\alpha'\otimes\alpha} + \mu'_{\alpha\alpha'\otimes\alpha}\otimes\mu_\alpha\big)/4.$$

VITA

Name: Minsoo Jeong

Address: Department of Economics, Texas A&M University

4228 TAMU, College Station, TX 77843

Email Address: jongms@tamu.edu

Education: B.A., Economics, Seoul National University, 2001

M.A., Economics, Seoul National University, 2003

The typist for this dissertation was Minsoo Jeong.