SEMIPARAMETRIC INFERENCE FOR INTEGRATED VOLATILITY FUNCTIONALS USING HIGH-FREQUENCY FINANCIAL DATA

Yunxiao Liu

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2017

Approved by:

George Tauchen

Chuanshu Ji

Amarjit Budhiraja

Shankar Bhamidi

Vladas Pipiras

ABSTRACT

YUNXIAO LIU: Essays in High-frequency Financial Econometrics
(Under the direction of George Tauchen and Chuanshu Ji)

With the advent of intraday high-frequency data of financial assets since the late 1990s, the research of financial econometrics has entered into a "big data" era. New theoretical techniques using the theory of continuous time stochastic processes has been extensively developed, and new empirical evidence has been documented. In particular, due to its far-reaching applications in various fields such as risk management and option pricing, the study of volatility, which quantitatively measures the uncertainty of prices of financial assets, has drawn substantial attention from researchers and there has been a large amount of literature devoted to this topic, including both modelling and prediction. In this dissertation, we are firstly concerned with the statistical inference of the so-called *integrated volatility functionals*, which is a general class of quantities that are computed from volatility. Secondly, we also devise a simulation method to recover the probability distribution of prices of financial assets by taking advantage of the information contained in sampled price data.

Accordingly, the dissertation consists of two parts. In the first part, we focus on the estimation of integrated volatility functionals, where the volatility process is assumed to be a long memory Itô semimartingale (LMIS), which is defined as the sum of an Itô semimartingale and a process satisfying certain regularity assumptions that in particular is able to capture the long memory property of financial volatility that has been vastly documented in literature. We provide central limit theorem (CLT) in such context. Furthermore, under the such LMIS assumption, we consider both parametric and nonparametric bootstrap inference methods of integrated volatility functionals, and we show the validity of both bootstrap methods by providing CLTs. Furthermore, with the usual assumption of volatility being Itô semimartingale, we consider an empirical-process form of integrated volatility functionals, and offer functional CLTs when the indexing parameter is of arbitrary finite dimensions. We also consider bootstrap inference in this empirical-process setting.

In the second part, we consider Euler method with estimated spot volatility, from which we are able to regenerate and realize the stochastic dynamics of price of financial asset by taking advantage of the information contained in the observed prices. We provide both theoretical foundation and empirical application of this method.

ACKNOWLEDGEMENTS

It is my highest honor and distinct privilege to develop my Ph.D. career in the greatest Department of Statistics and Operations Research at the University of North Carolina at Chapel Hill, and to work with the world-class econometricians, statisticians and probabilists from both UNC Chapel Hill and Duke University. I would like to express my most sincere gratitude to all who offered generous support over the course of my Ph.D. career in the past three and half years.

Firstly, I would like to extend my deepest gratitude to my dissertation advisor, Professor George Tauchen, for his generous supervision, guidance, patience and encouragement from the very beginning of my Ph.D. study. It was Professor Tauchen who opened to me the door to the appealing field of high-frequency financial econometrics, and who guided me through such a challenging but rewarding journey. His deep understanding of the research problems is always visionary which helps me focus on the most important problems. His critical comments from the weekly seminars have been sharp and insightful throughout, which have proved extremely beneficial to my career, both today and in the future.

Also, I want to thank my coadvisor, Professor Chuanshu Ji, for his understanding, support and guidance. It was Professor Ji's support that enables me to exclusively focus on my research work and make such fast progress. Every conversation with him was quite informative, encouraging and relieving, in particular when there was setback during my second Ph.D. application.

I would also like to express my gratitude to Professor Jia Li for his detailed guidance on my academic research work. Without his generous support and encouragement, it would be impossible for me to go this far on the road of high-frequency financial econometrics. Whenever I got stuck in the long, mysterious proof work, I could always receive illustrating advice from his smartness and sharp insights into the research problems. Whenever I finished a short term task, I could always build up more confidence and enhance my work from his detailed review and comments. Furthermore, I have been inspired by his rigorous attitude to settle down each piece of proof and to finalize each academic paper.

I would like to acknowledge other committee members, Professor Amarjit Budhiraja, Professor Shankar Bhamidi and Professor Vladas Pipiras. The probability courses taught by Professor Buhiraja and Professor Bhamidi offered me rigorous training in probability theory which turned out to be crucially helpful to my research work. I also thank them for their help at the start of my research projects, which saved me a lot of time and energy. I thank Professor Pipiras for his prompt and detailed response to each of my requests on fractional Brownian Motion. His expertise in this area instructively guided me to explore this new topic without any burden.

I appreciate the generous support from all other faculty members from STOR department for their illustrating lectures and discussions on various topics over the years of my Ph.D. study.

I am also grateful to the care and support from my friends and classmates, from both UNC Chapel Hill and Duke University. My special thanks go to Dr. Ruoyu Wu, for his forever generous and illuminating answers to my naive probability questions. My thanks go to senior graduate students, including but not limited to, Dr. Tao Wang, Dr. Yu Zhang, Dr. Qing Feng, Dr. Dongqing Yu, Dr. Minghui Liu, Dr. Qunqun Yu, Dr. Xi Chen, Dr. Hanwei Liu, Dr. Yang Yu, Dr. Meilei Jiang and Dr. Siliang Gong for their generous suggestions on my both academic and industrial careers. My thanks also go to my fellow graduate students, include not limited to Mrs. Yang Yu, Mrs. Yichen Tu, my former roommate Mr. Zheqi Zhang, Mr. Zhengling Qi, Mr. Jianyu Liu and Mrs. Huijun Qian for hundreds of meals, jokes and conservations together. I am grateful to my friends from Duke University. I thank my long-term friend, Mr. Ruibo Ma for staying constantly connected and supportive for the past five years ever from the start of my arrival at the United States. I also thank the group members on financial econometrics from Economics Department at Duke University, including but not limited to, Dr. Ben Zhao, Mrs. Yuan Xue and Ms. Rui Chen, for the valuable discussions and advice. I wish all the best to all of you in the future.

Most importantly, I would like to thank the love and support from my family. None of my graduate life in the U.S. would ever be achieved without the selfless love and support from my parents and grandparents. My thanks also go to Ms. Yan Yang, for her support over the darkest period of time during my graduate life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

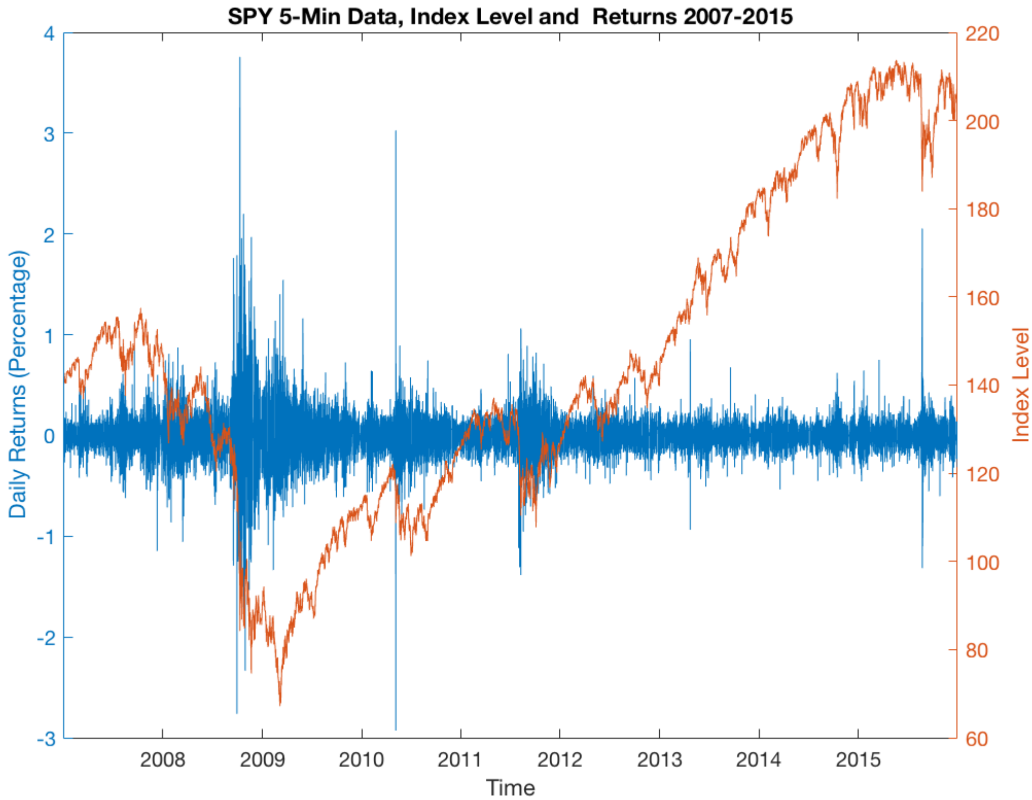
With the advent of intraday high-frequency data of financial assets since the late 1990s, the research of financial econometrics has entered into a “big data” era. New theoretical techniques using the theory of continuous time stochastic processes has been extensively developed, and new empirical evidence has been documented. In particular, due to its far-reaching applications in various fields such as risk management and option pricing, the study of volatility, which quantitatively measures the uncertainty of prices of financial assets, has drawn substantial attention from researchers and there has been a large amount of literature devoted to this topic. In this Introduction, we offer an overview of the dissertation including the basic set-up, the research questions we are to explore, the main results we have obtained and a direction of future work.

We start with the basic statistical setting. For simplicity, we only consider one-dimensional case in the Introduction and the multivariate setting will be discussed in the following chapters. Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions (see e.g., (Jacod and Shiryaev, 2003)), on which are defined a one-dimensional Brownian motion W and a Poisson random measure μ on $\mathbb{R}_+ \times E$ with deterministic intensity $\nu(dt, dz) = dt \otimes \lambda(dz)$. Here E is a Polish space. As is usual the case (e.g. (Aït-Sahalia and Jacod, 2014)), we model the logarithm of the price process X_t of a given stock as an Itô semimartingale in the following Grigelionis form:

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + (\delta 1_{\{\|\delta\|\leq 1\}}) * (\mu - \nu)_t + (\delta 1_{\{\|\delta\|>1\}}) * \nu_t,$$

where σ is an $\mathbb{R}-$valued predictable (or simply progressively measurable) process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and δ is a predictable $\mathbb{R}-$valued function on $\Omega \times \mathbb{R}_+ \times E$. Throughout the dissertation, all stochastic processes, unless otherwise specified, are assumed to be

Figure 1.1: A typical price path and return



(a) The price path and returns of SPY from 01/03/2007-12/31/2015, based on 5-min data.

càdlàg adapted and hence locally bounded. A typical stock price path can be seen from Figure 1.1.

In finance and econometrics, the process σ_t is called the spot (or local) volatility of X_t, and accordingly the spot (local) variance process is defined as $c_t = \sigma_t^2$. Since mathematically the sign of σ_t cannot be identified and c_t is always nonnegative, it is more straightforward to consider c_t in study, which, abusing the terminology, we still call volatility. Moreover, since c_t is latent (not observable), it can only be recovered by using the data of X sampled with high-frequency via certain statistical estimation procedures.

A typical high-frequency sampling setting goes as follows: given a fixed time span $[0, T]$, which typically can be a trading day, the price process X is discretely sampled with equidistant step size Δ_n. Hence for any $i = 1, 2, \ldots, [T/\Delta_n]$, the log-return of X over interval $[(i-1)\Delta_n, i\Delta_n]$ is given by

$$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}.$$

We consider infill asymptotics where the mesh Δ_n of grid of sampling asymptotically tends to 0 as $n \to \infty$.

Recall that in a traditional "low frequency" setting, the daily risk has been measured using daily squared returns, see e.g. (Engle, 1982) and (Bollerslev, 1986), or daily range (difference between maximum and minimum within one day), see e.g. (Alizadeh et al., 2002). With the availability of intraday high-frequency data sampled as above, however, a new measure, which is defined as

$$\int_0^T c_s ds$$

and called *integrated volatility*, becomes prevailing. A consistent and efficient estimator for the intergrated volatility is *realized volatility*, which is defined as the sum of squared intraday returns. Such a method is proposed by (Andersen and Bollerslev, 1999) and popularized by, e.g., (Andersen et al., 2001a) and (Andersen et al., 2003a). More generally, it would be interesting to study the random object of the form

$$S(g) \equiv \int_0^T g(c_s) ds,$$

for some (possibly nonlinear) function g, which is called *integrated volatility functional* and accommodates many quantities that are related to volatility, including integrated volatility as a special case when $g(x) = x$. For other representative examples of g, see Chapter 3.

The first part of this dissertation, which includes Chapter 3, 4 and 5, focuses on the statistical inference for $S(g)$ under various conditions on the variance process c_t and test function g. More precisely, for given g, the estimator for $S(g)$ can be constructed in two steps: firstly, we nonparametrically recover the spot variance c_t over the sampling grid by employing a local average of sum of squared truncated returns (see (Jacod and Protter, 2012), Chapter 9 and 13), that is, for any $0 \leq i \leq N_n \equiv [T/\Delta_n] - k_n$, let

$$\hat{c}_{i\Delta_n} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right)^2 1_{\{||\Delta_{i+j}^n X|| \leq u_n\}},$$

where k_n is a sequence of integers that goes to infinity representing the number of increments employed in a local window and u_n determines the truncation threshold for eliminating jumps in X, see (Mancini, 2001). Next, plugging the estimated spot volatilities into a Riemann approximation framework, the estimator of $S(g)$ is given by

$$S_n(g) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(g(\hat{c}_{i\Delta_n}) - \frac{1}{k_n} g''(\hat{c}_{i\Delta_n}) \hat{c}_{i\Delta_n}^2 \right), \tag{1.1}$$

where a higher order bias term is subtracted off.

Assuming that the volatility process follows an Itô semimartingale, (Jacod and Rosenbaum, 2013a) shows a CLT (see (4.5) in Chapter 4 below) for $S_n(g)$ approximating $S(g)$ with rate $\sqrt{\Delta_n}$ provided test function g and its derivatives satisfy a polynomial growth condition. By a local spatialization argument, (Li et al., 2016a) extends the CLT result to the case of g satisfying a much weaker condition given as Assumption 3.2.1, and (Li and Xiu, 2016) shows an empirical-process-type CLT in a similar setting. However, all these results assume that volatility process is an Itô semimartingale, which is actually not able to capture the long memory property of volatility dynamics that has been widely documented in literature, see for example, (Comte and Renault, 1998). In contrast, in Chapter 3 we derive the same CLT result for $S_n(g)$ approximating $S(g)$ with rate $\sqrt{\Delta_n}$ for a larger class of volatility processes: we assume the volatility process follows a *long-memory Itô semimartingale (LMIS)* which is given by

$$\sigma_t = \sigma_{1,t} + \sigma_{2,t},$$

where $\sigma_{1,t}$ is an Itô semimartingale and $\sigma_{2,t}$ can be a fractional Brownian motion or a Weiner integral with respect to fractional Brownian motion. We state the result below, which can be viewed as one-dimensional version of Theorem 3.4.1.

Theorem 1.0.1. *Under Assumptions 3.1.1-3.4.1, it holds that*

$$\frac{1}{\sqrt{\Delta_n}}(S_n(g) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g)),$$

where $\mathcal{MN}(0, V)$ *is a centered mixed normal distribution with conditional variance*

$$V(g) \equiv 2\int_0^T \left(g'(c_s)\right)^2 c_s^2 ds.$$

Here $\xrightarrow{\mathcal{L}\text{-}s}$ denotes stable convergence in law which will be elaborated in Chapter 2.

In light of Theorem 1.0.1, inference can be done for $S(g)$: for example, one can construct confidence intervals provided that the asymptotic variance $V(g)$ can be consistently estimated, as Corollary 3.7 in (Jacod and Rosenbaum, 2013b). On the other hand, however, a consistent estimator for $S(g)$ is not indispensable to obtain confidence intervals for $S(g)$, as one may turn to bootstrap method.

In Chapter 4, under the assumption that the volatility process σ_t follows LMIS, we propose algorithms for constructing confidence intervals for $S(g)$ via both parametric bootstrap method and nonparametric bootstrap method. Given bootstrap samples of returns $\mathcal{D}_n^* \equiv \{\Delta_i^n X^*, i = 1, \dots, n\}$ generated either parametrically or nonparametrically, the bootstrap estimator for $S(g)$ is given by, not surprisingly, an analogue form of (1.1):

$$S_n(g; \mathcal{D}_n^*) \equiv \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} \left(g(\hat{c}_{n,i}^*) - \frac{1}{k_n} g''(\hat{c}_{n,i}^*)\hat{c}_{n,i}^{*2} \right),$$

where

$$\hat{c}_{n,i}^* = \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X^*\right)^2$$

are bootstrap spot covariance estimators using $\mathcal{D}_n^*$. Here we take $T = 1, \Delta_n = 1/n$ and both $\hat{c}_{n,i}^*$ and $S_n(g; \mathcal{D}_n^*)$ are constructed over non-overlapping blocks $[ik_n/n, (i+1)k_n/n]$ for $i \in \mathcal{I}_n \equiv \{0, \dots, [n/k_n] - 1\}$. Then the bootstrap confidence interval of coverage $1 - \alpha$ is formed as

$$[S_n(g; \mathcal{D}_n) + \widetilde{S}_n(g; \mathcal{D}_n) - q_{1-\alpha/2}(S_n(g; \mathcal{D}_n^*)), S_n(g; \mathcal{D}_n) + \widetilde{S}_n(g; \mathcal{D}_n) - q_{\alpha/2}(S_n(g; \mathcal{D}_n^*))],$$

where $q_{\alpha/2}(S_n(g;\mathcal{D}_n^*))$ and $q_{1-\alpha/2}(S_n(g;\mathcal{D}_n^*))$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of $S_n(g;\mathcal{D}_n^*)$ respectively, computed from a large number of bootstrap repetitions, and

$$\widetilde{S}_n(g;\mathcal{D}_n) = \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} g(\hat{c}_{n,i})$$

is the uncorrected estimator for $S(g)$. Here we use $\mathcal{D}_n \equiv \{\Delta_i^n X, i = 1, \ldots, n\}$ to denote the set of original returns, in contrast to its bootstrap counterpart $\mathcal{D}_n^*$. Theoretically, the asymptotic coverage rate of $1-\alpha$ is guaranteed by Theorem 1.0.2. In the sequel, we use $Z_n \xrightarrow{\mathcal{L}|\mathcal{F}} Z$ to denote $\mathcal{L}(Z_n|\mathcal{F}) \xrightarrow{\mathbb{P}} \mathcal{L}(Z|\mathcal{F})$ for a sequence of random variables $(Z_n)_{n\geq 1}$ and Z, namely, the conditional distribution of Z_n given $\mathcal{F}$ converges to that of Z in probability under Prokhorov metric. Such a mode of convergence in commonly used in the setting of bootstrap, as well as together with stable convergence in law.

Theorem 1.0.2. *Suppose the Assumption 3.1.1-3.4.1, it follows that*

$$\sqrt{n}\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, V(g)),$$

where

$$V(g) \equiv 2\int_0^T \left(g'(c_s)\right)^2 c_s^2 ds.$$

Furthermore, we implement Monte Carlo simulation to study the coverage rates of both the parametric and nonparametric bootstrap confidence intervals in finite sample, the results of which validate our theoretical asymptotic result given in Theorem 1.0.2.

Chapter 5 considers a more general form of test function g. We focus on a functional form of the test function g, namely,

$$g : \mathcal{V} \times \Theta \to \mathbb{R},$$

where $\mathcal{V} \subset \mathbb{R}$ is the range space of spot volatility, and $\Theta \subset \mathbb{R}^{\mathbf{dim}\theta}$ is the space of some indexing parameter θ. So for each fixed value c_t, $g(c_t, \cdot)$ is a function over Θ. For example, when $g(x) = \exp(-ux)$ for $u \in (0, \infty)$, $S(g)$ is the Laplace transform of the volatilty occupation time (Todorov and Tauchen, 2012b), which summarizes the complete spatial information of the volatility process within the time span. See also (Li and Xiu, 2016) Section 3.3 for other econometric applications in this context.

Our goal is to (uniformly) estimate the quantity of the form

$$S(g;\theta) \equiv \int_0^T g(c_s;\theta)ds.$$

Similarly as in (Li and Xiu, 2016), the proposed estimator is

$$S_n(g;\theta;\mathcal{D}_n) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(g(\hat{c}_{i\Delta_n};\theta) - \frac{1}{k_n}\partial_c^2 g\left(\hat{c}_{i\Delta_n};\theta\right)\hat{c}_{i\Delta_n}^2 \right).$$

Under the assumption that the volatility process σ_t is an Itô semimartingale, plus other regularity conditions, we are able to obtain the following functional central limit theorem.

Theorem 1.0.3. *Suppose Assumption 3.1.1 with $\sigma_2 = 0$, and Assumption 3.2.2. Moreover, assume $g : \mathcal{V} \times \Theta \to \mathbb{R}$ satisfies Assumption 3.2.1 with respect to the first variate and is continuously differentiable with respect to $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^{\boldsymbol{dim\theta}}$ is a compact set, with $dim\theta < \infty$. Then the sequence $\Delta_n^{-1/2}\left(S_n(g;\cdot;\mathcal{D}_n) - S(g;\cdot)\right)$ of processes converges $\mathcal{F}-$stably in law under the uniform metric to a process $\xi(\cdot)$ which, conditional on $\mathcal{F}$, is centered Gaussian with covariance function $S_g(\cdot,\cdot)$, where $S_g(\cdot,\cdot)$ is defined as, for any $\theta,\theta' \in \Theta$,*

$$S_g(\theta,\theta') \equiv 2\int_0^T \partial_c g(c_s;\theta)\partial_c g(c_s;\theta')c_s^2 ds.$$

Furthermore, in this functional setting we also develop both parametric and nonparametric bootstrap algorithms to conduct statistical inference as regard to $S(g;\cdot)$. The algorithms are very similar to the ones in Chapter 4 and we provide empirical-process-type asymptotic results to justify both bootstrap algorithms. From an application point of view, such asymptotic results could help constructing (empirical) uniform confidence region for $S(g;\cdot)$.

The second part of the dissertation consists of Chapter 6 alone, where we develop *Euler method with estimated spot volatility.* In the field of financial econometrics, there is always need to simulate the following diffusion process

$$dX_t = b_t dt + \sigma_t dW_t,$$

where W is Brownian motion. Very often X denotes the log-price of financial asset, say stock, and σ is referred to as the volatility process related to X. The most commonly used method to simulate X is the so-called Euler-Maruyama approximation, which is named after Leonhard Euler and Gisiro Maruyama, and is actually a simple generalization of the Euler method for ordinary differential equations to stochastic differential equations. More precisely, to obtain the value of X at terminal time T over a fixed time span $[0, T]$, one uses the recursive equation:

$$X_{\tau_{n+1}} = X_{\tau_n} + b_{\tau_n}(\tau_{n+1} - \tau_n) + \sigma_{\tau_n}(W_{\tau_{n+1}} - W_{\tau_n}),$$

with given discretization grid $0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$. Usually, the equidistant discretization scheme is used, namely, $\tau_{i+1} - \tau_i = \delta$ for some time step $0 < \delta < T$. For a thorough treatment on Euler-Maruyama approximation and its extensions, see (Kloeden and Platen, 1992).

However, to implement such procedure, the values of $(b_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$ have to be prespecified (or simulated) beforehand, which might not replicate the true world as much as possible, even if the specified values for parameters are claimed to be "calibrated to the real world".

Alternatively, instead of specifying particular dynamics for $(\sigma_t)_{t\geq 0}$, we can use estimated spot volatility based on high-frequency data in the Euler method. In other words, we would like to design a data generating mechanism, via Euler method, to regenerate data that mimics the real world more realistically, by taking advantage of the information contained in the observed real data. As seen below, "mimic the real world" is in the sense that the probability distribution of the simulated data generated by our Euler method with estimated spot volatility uniformly approximates (measured by Wasserstein metric) to that of the true data, under certain assumptions. In fact, we have already used this method in constructing parametric bootstrap confidence interval for integrated volatility functionals (see Algorithm 1 in Chapter 4).

Put it more precisely, we assume that the log-price process follows

$$\begin{aligned} dX_t &= \sqrt{c_t} dW_t \\ X_0 &= 0 \end{aligned}$$

where T is the terminal time, c is the variance process and W is one-dimensional Brownian Motion introduced above. In particular, X has neither a drift part nor a jump part. We consider an equally spaced time discretization grid over $[0, T]$ for Euler approximation, i.e., for some $\delta > 0$, let

$$\tau_0 = 0, \quad \tau_i = i\delta.$$

where $i \in \{0, 1, \ldots, \lfloor \frac{T}{\delta} \rfloor\}$. At each discretization time point $i\delta$, the spot volatility estimation is given by

$$\hat{c}_{i\delta} = \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 .$$

Then for fixed n and δ, the global Euler approximation with estimated spot volatility is given by the process:

$$Y_t^{n,\delta} = \sum_{i=0}^{[t/\delta]-1} \sqrt{\hat{c}_{i\delta}} (\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}), \quad 0 \le t \le T,$$

where $\widetilde{W}$ is Brownian motion on the simulation space, which is independent of all the information living in the real world. In particular, for $Y^{n,\delta}$ to be well-defined, the condition $\delta > \Delta_n$ has to be satisfied.

We derive the theoretical results associated with $Y^{n,\delta}$. The very first thing one should notice is that since in simulation only $\widetilde{W}$ is available, $Y^{n,\delta}$ is a "consistent estimator" for the simulated log-price defined by

$$\widetilde{X}_t = \int_0^t \sqrt{c_s} d\widetilde{W}_s, \quad 0 \le t \le T,$$

rather than the true price observed process X_t, which has the same distribution as $\widetilde{X}_t$ under the no leverage assumption, i.e., the volatility process c_t and Brownian motion W_t, both of which are defined from the real world, are independent. To derive the convergence rate of $Y^{n,\delta}$ approximating $\widetilde{X}$, we have

Theorem 1.0.4. *Suppose Assumptions 6.2.1 and 6.2.3. Assume further that $\{c_t : t \geq 0\}$ has sample paths satisfying for any $t > s > 0$,*

$$\mathbb{E}|c_t - c_s|^2 \leq K|t-s|^{2\rho}, \quad 0 < \rho \leq 1,$$

for some constant K. Then it holds for any fixed discretization distance $\delta \in [\Delta_n, T)$ that

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |Y_t^{n,\delta} - \widetilde{X}_t|\right) \leq K\left(\frac{1}{\sqrt{k_n}} + (k_n\Delta_n)^\rho + \delta^\rho + \left(\delta \log\left(\frac{2T}{\delta}\right)\right)^{\frac{1}{2}}\right)$$

for some constant K.

From Theorem 1.0.4 we are able to derive the "optimal" simulation scheme in the sense of fastest convergence rate: to make $Y^{n,\delta}$ converges to $\widetilde{X}$ as fast as possible, one should first take $\delta_n \to 0$ as small as possible, i.e.

$$\delta_n = \Delta_n,$$

which means taking each data sampling point as a discretization point; then we strike balance between statistical error and target error arising from spot volatility estimation, by requiring

$$k_n^{\rho+\frac{1}{2}}\Delta_n^\rho \to \beta \in (0,\infty),$$

or equivalently $k_n \sim \Delta_n^{-\frac{\rho}{\rho+\frac{1}{2}}}$. In this fashion, our "optimal" Euler approximation becomes:

$$Y_t^n = \sum_{i=0}^{\lfloor t/\Delta_n\rfloor-1} \sqrt{\hat{c}_{i\Delta_n}}\Delta_i^n\widetilde{W}, \quad 0 \leq t \leq T,$$

with convergence rate

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\le t\le T}|Y_t^n-\widetilde{X}_t|\right) &\le K\left(\frac{1}{\sqrt{k_n}}+\big(k_n\Delta_n\big)^{\rho}+\Delta_n^{\rho}+\left(\Delta_n\log\big(\frac{2T}{\Delta_n}\big)\right)^{\frac{1}{2}}\right)\\
&\sim \Delta_n^{\frac{1}{2+\frac{1}{\rho}}}.
\end{aligned}$$

We show that such a convergence rate is actually exact (not only an upper bound).

As far as applications are concerned, we are able to evaluate the accuracy of estimation of diffusive beta (see (Reiss et al., 2015)). Simply speaking, it is done by obtaining the sampling distribution of the diffusive beta. More generally, we can use the Euler method with estimated volatility to obtain the sampling distribution for any functional of the path of log-price process.

Future Work: The results obtained above give us a solid foundation for future work.

As for the project of bootstrap inference, we have the following to do

- Consider the over-lapping case, besides the non-overlapping case studied already;
- Prove a similar asymptotic result when the volatility process is of the mixed form considered in Chapter 3.
- Prove a similar asymptotic result when g is a functional, characterized by another indexing parameter θ, as the setting in Chapter 5.

As for the project of Euler method with estimated spot volatility, we can continue the study in both theory and application:

- **Theory:** An interesting direction to generalize our Euler method with estimated volatility is to take into account the so-called leverage effect, which refers to the negative correlation between volatility and returns. Since the Brownian motion $\widetilde{W}$ used in simulation is independent of everything in the real world, to create (negative) correlation between the simulated prices and volatility, we need to use the same $\widetilde{W}$ to regenerate volatility process, which requires to model volatility process as an Itô semimartingale as well and estimate the volatility of volatility (vol. of vol.). As one may imagine, the convergence rate in this situation would be even slower than $\Delta_n^{1/4}$ as both volatility and vol. of vol. are latent.
- **Application:** The daily range of a given price process X, defined as the difference between $\max X_t$ and $\min X_t$ within one day, had been a popular measure to quantify daily risk. Obviously, the daily range depends on the whole price path over a single day, and hence its sample distribution can be realized by the Euler method with estimated volatility. Consequently, we are able to implement empirical study using the Euler method developed here.

CHAPTER 2
Preliminaries

We start with notation that will be used throughout the paper. For a vector B, we use B^j to denote its j-th component. For an integer $d > 0$, $\mathcal{M}_d$ denotes the space of $d \times d$ nonnegative semidefinite matrices. For a matrix A, we use A^{ij} and $A^\intercal$ to denote its (i, j) element and transpose, respectively. For a matrix valued process A_t, the notations A_t^{ij} and $A_t^\intercal$ are interpreted similarly. For a matrix $A \in \mathcal{M}_d$ and a differentiable function g defined on $\mathcal{M}_d$, the first two partial derivatives of g are denoted as $\partial_{jk} g(A) = \partial g(A)/\partial A_{jk}$ and $\partial^2_{jk,lm} g(A) = \partial^2 g(A)/\partial A_{jk} \partial A_{lm}$ respectively. For a set B, $1_B(\cdot)$ denotes the indicator function of set B. The symbol $\equiv$ indicates equality by definition. $||\cdot||$ denotes the Frobenius norm. For any two (possibly random) real-valued sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, we write $a_n = O_p(b_n)$ if a_n/b_n is bounded in probability and write $a_n = o_p(b_n)$ if a_n/b_n converges to 0 in probability. All limits are for $n \to \infty$. We use $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{L}\text{-}s}$ to denote convergence in probability, convergence in law and stable convergence in law, respectively. We use K to denote a generic constant which may vary from line to line.

In this section we give a brief introduce to two important notions that will be used frequently in the rest of dissertation. Section 2.1 introduces Itô semimartingale, which is a basic class of stochastic processes commonly used in econometrics and finance. Section 2.2 discusses the so-called stable convergence in law, which is stronger than the usual convergence in law (or weak convergence).

2.1 Itô semimartingale

We begin with the definition of general semimartingale. For a comprehensive treatment on this topic, together with other notions in stochastic analysis, such as theory of Itô integral, (Poisson) random measures and stochastic integral with respect to random measures, see (Jacod and Shiryaev, 2003), (Jacod and Protter, 2012) and (Aït-Sahalia and Jacod, 2014). We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual condition as given in (Jacod and Shiryaev, 2003) p.2. Throughout the paper, all stochastic processes, unless otherwise specified, are assumed to be càdlàg adapted and hence locally bounded.

Definition 2.1. (a) A semimartingale is a process X of the form $X = X_0 + M + A$ where X_0 is finite-valued and $\mathcal{F}_0-$measurable, M is a local martingale with $M_0 = 0$, and A is a stochastic process of finite variation.

(b) A special semimartingale is a semimartingale X which admits a decomposition $X = X_0 + M + A$ as above, with a process A that is predictable.

Given an $\mathbb{R}^d-$valued process X, the jump measure associated with X is defined as

$$\mu^X = \sum_{s>0:\Delta X_s \neq 0} \epsilon_{(s,\Delta X_s)},$$

where ϵ_a denotes the Dirac measure sitting at a. Then we can rewrite a semimartingale as

$$X_t = X_0 + A'_t + M_t + \sum_{s\leq t} \Delta X_s 1_{\{\|\Delta X_s\|>1\}}$$

where $M_0 = A'_0 = 0$ and A' is of finite variation and M is a local martingale. Then the semimartingale $A' + M$ by construction has jumps of size always smaller than 1. Hence by Lemma 4.24 in (Jacod and Shiryaev, 2003), $A' + M$ is special and we can write

$$A' + M = B + N,$$

where $N_0 = B_0 = 0$ and N is a local martingale and B is a predictable process of finite variation.

Definition 2.2. The characteristic of a $\mathbb{R}^d-$valued semimartingale X is the following triple (B, C, ν):

(i) $B = (B^i)_{1\leq i\leq d}$ is the predictable process of finite variation defined above;

(ii) $C = (C^{ij})_{1\leq i,j\leq d}$ is the quadratic variation of the continuous local martingale part X^c of X, that is, $C^{ij} = \langle X^{i,c}, X^{j,c}\rangle$;

(iii) ν is the predictable compensating measure of the jump measure μ^X of X.

One should note that the characteristic triple does NOT characterize the law of the process except for special cases. An important special case of semimartingale is Levy process, the characteristic triple of which is

$$B_t(\omega) = bt, \quad C_t(\omega) = ct, \quad \nu(\omega, dt, dx) = dt \otimes F(dx),$$

which are not random actually. For a general treatment of Lévy process, see (Bertoin, 1998).

In financial modelling, it is common to use a special class of semimartingales, but which is also a direct extension of Lévy process:

Definition 2.3. A $\mathbb{R}^d-$valued semimartingale X is an Itô semimartingale if its characteristic (B, C, ν) are absolutely continuous with respect to the Lebesgue measure, in the sense that

$$B_t = \int_0^t b_s ds, \quad C_t = \int_0^t c_s ds, \quad \nu(dt, dx) = dtF_t(dx),$$

where $b = (b_t)$ is an $\mathbb{R}^d-$valued process, $c = (c_t)$ is a process with values in $\mathcal{M}_d$, and $F_t = F_t(\omega, dx)$ is for each (ω, t) a measure on $\mathbb{R}^d$.

Now we come to give a fundamental representation theorem for Itô semimartingale, which is usually referred to as the Grigelionis form of Itô semimartingale. The following theorem is Theorem 2.1.2 in (Jacod and Protter, 2012). Let d' be an arbitrary integer with $d' \geq d$, E be a Polish space with a $\sigma-$finite measure λ having no atom, and $q(dt, dx) = dt \otimes \lambda(dx)$.

Theorem 2.1.1. Let X be a $d-$dimensional Itô semimartingale on the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with characteristics (B, C, ν). There is a very good filtered extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$, on which are defined a $d'-$ dimensional Brownian motion W and a

Poisson random measure p on $\mathbb{R}_+ \times E$ with Lévy measure λ, such that

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + (\delta 1_{\{\|\delta\| \leq 1\}}) * (p - q)_t + (\delta 1_{\{\|\delta\| > 1\}}) * p_t, \tag{2.1}$$

and where σ is an $\mathbb{R}^d \otimes \mathbb{R}^{d'}-$valued predictable (or simply progressively measurable) on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, and δ is a predictable $\mathbb{R}^d-$valued function on $\Omega \times \mathbb{R}_+ \times E$.

For a more detailed description of extension of probability space, see (Jacod and Protter, 2012) p.36-37. The point of Theorem 2.1.1 is that any $d-$dimensional Itô semimartingale can be expressed in terms of a Brownian motion and a Poisson random measure, and in fact, (2.1) can be used as the definition for Itô semimartingales, up to extending the space.

Itô semimartingales of form (2.1) have been widely used in modelling prices of financial assets for various reasons. At first, it has been widely known that the prices of financial assets, say stocks, have jumps, which for example occurs when there is significant macroeconomic announcements. Although we consider Poisson random measure with a compensator of product form, $dt \otimes \lambda(dx)$, which is time-homogeneous, the whole jump part in (2.1) is actually time-inhomogeneous since the jump size function δ is random and time-varying. As a consequence the jump part of stock price is driven by a very general class of processes. As for continuous part, the drift part captures the persistence in the process, and also represents the compensation for risk and time, while the continuous martingale part given as a stochastic integral models the small moves.

In fact, as (Back, 1991) points out, special semimartingale appears to be the most general concept of a gains process for which the notion of a local risk premium can be well-defined. On the other hand, (Barndorff-Nielsen and Shephard, 2004a) (Remark 1) and (Barndorff-Nielsen et al., 2006) (Remarks 3) demonstrate the generality of the continuous (local) martingale part $\int_0^t \sigma_s dW_s$ in (2.1). More precisely, by a representation theorem of local martingale as stochastic integral (e.g., (Karatzas and Shreve, 1991) p.170-172), all continuous local martingales with absolutely continuous quadratic variation can be written in the form of $\int_0^t \tilde{\sigma}_s dW_s$ for some process $\tilde{\sigma}_t$. Using the Dambis-Dubins-Schwartz theorem, the difference between the class of continuous local martingale and the class of stochastic integrals with respect to Brownian motion is the local martingales with continuous, but not absolutely continuous quadratic variation. Thus the form of continuous (local) martingale part we consider here is only slightly smaller the class of general continuous local martingale.

2.2 Stable convergence in law

In this subsection we introduce the notion of stable convergence in law, which is stronger than the usual convergence in law or weak convergence. We first review the definition of the latter for illustrative purpose.

Let E be a Polish space, with Borel σ-field $\mathcal{E}$. Let $\{Z_n\}$ be a sequence of E-valued random variables, allowing each of them defined on its own probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$.

Definition 2.4. We say that Z_n converges in law if there is a probability measure μ on $(E, \mathcal{E})$ such that

$$\mathbb{E}\left(f(Z_n)\right) \longrightarrow \int f(x)\mu(dx).$$

for all (Lipschitz) continuous bounded functions f on E.

Usually, one could "realize" the limit as a E-valued random variable Z on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the above convergence reads as

$$\mathbb{E}\left(f(Z_n)\right) \longrightarrow \mathbb{E}\left(f(Z)\right).$$

However, the usual convergence in law defined as above may not be enough in the area of financial econometrics. As one can see, quite often we will be in the following scenario: we need to estimate some (multivariate) parameter θ and we propose a sequence of consistent estimators $\hat{\theta}_n$. We are able to show a central limit theorem with certain convergence rate, say $\sqrt{n}$ and mixed normal limiting distribution, namely,

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma).$$

Very often the limiting variance Σ is random as well, especially in the case of stochastic volatility. In order to do statistical inference from CLT (for example, to construct confidence intervals for θ), one needs to scale the limiting distribution to (multivariate) standard normal distribution. However, this may not be achieved even if Σ can be consistently estimated, as

$$Z_n \xrightarrow{\mathcal{L}} Z, \qquad Y_n \xrightarrow{\mathbb{P}} Y$$

do NOT in general imply

$$(Z_n, Y_n) \xrightarrow{\mathcal{L}} (Z, Y),$$

the only exception being Y is a constant, which is case of the so-called Slutsky Theorem. We hence need a stronger version of convergence in law to make sure the joint convergence $(Z_n, Y_n) \xrightarrow{\mathcal{L}} (Z, Y)$, still holds even if Y is random.

We require E-valued sequence $\{Z_n\}$ of random variables to be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.5. We say that Z_n stably converges in law if there is a probability measure η on the product space $(\Omega \times E, \mathcal{F} \otimes \mathcal{E})$, such that $\eta(A \times E) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$ and

$$\mathbb{E}\left(Y f(Z_n)\right) \longrightarrow \int Y(\omega) f(x) \eta(d\omega, dx)$$

for all bounded (Lipschitz) continuous functions f on E and all bounded random variables Y on $(\Omega, \mathcal{F})$.

As before, we can "realize" the limit Z on an (arbitrary) extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$, then the stable convergence in law above can be written as

$$\mathbb{E}\left(f(Z_n)\right) \longrightarrow \tilde{\mathbb{E}}\left(Y f(Z)\right),$$

provided $\widetilde{\mathbb{P}}(A \cap \{Z \in B\}) = \eta(A \times B)$ for all $A \in \mathcal{F}$ and $B \in \mathcal{E}$. Then we say Z_n converges stably to Z, denoted by

$$Z_n \xrightarrow{\mathcal{L}-s} Z.$$

By definition, it immediately follows that stable convergence in law implies convergence in law. Moreover, we do have the desired property

$$Z_n \xrightarrow{\mathcal{L}-s} Z, \quad Y_n \xrightarrow{\mathbb{P}} Y \implies (Z_n, Y_n) \xrightarrow{\mathcal{L}-s} (Z, Y).$$

In fact, stable convergence in law is very much like convergence in probability: when the limiting variable Z is defined on the same space Ω as all Z_n, it follows that

$$Z_n \xrightarrow{\mathcal{L}-s} Z \iff Z_n \xrightarrow{\mathbb{P}} Z.$$

We end this section with a brief literature retrospection. The notion of stable convergence in law dates back to (Rényi, 1963), and is developed by (Aldous and Eagleson, 1978), (Jacod, 1997) and (Jacod and Protter, 1998). An early use of this concept in econometrics is (Phillips and Ouliaris, 1990). For a brief summary of stable convergence in law used in a high-frequency financial econometrics setting, see (Jacod and Protter, 2012), Section 2.2.1, and a more detailed exposition in general context of stochastic analysis is in (Jacod and Shiryaev, 2003), Chapter VIII 5c. See also (P. and Heyde, 1980) for some different insights on the topic.

CHAPTER 3

Efficient Estimation of Integrated Volatility Functionals with General Volatility Dynamics

3.1 Setting

We start with introducing the formal setup for our analysis. Consider a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Throughout the chapter, all stochastic processes, unless otherwise specified, are assumed to be càdlàg adapted and hence locally bounded. Our basic assumptions of underlying processes are collected in Assumption 3.1.1.

Assumption 3.1.1. *For some constant $r \in [0,1)$, and a sequence of a sequence $(\tau_m)_{m\geq 1}$ of stopping times increasing to ∞, we have*

(i) The process X_t is a $d-$dimensional Itô semimartingale with the form

$$X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad J_t = \sum_{s\leq t} \Delta X_s = \int_0^t \int_{\mathbb{R}} \delta\left(s,z\right) \mu\left(ds,dz\right), \tag{3.1}$$

where the drift b_t is $d-$dimensional; the spot volatility process σ_t is $\mathbb{R}^d \otimes \mathbb{R}^d$ valued; W_t is a $d-$dimensional Brownian motion; μ is a Poisson random measure on $\mathbb{R}_+ \times E$ for an auxiliary Polish space E with the deterministic intensity measure $\nu(dt,dz) = dt \otimes \lambda(dz)$ for some $\sigma-$finite measure λ on E; $\delta : \Omega \times \mathbb{R}_+ \times E \to \mathbb{R}^d$ is a predictable function. Moreover, there are a sequence $(J_m)_{m\geq 1}$ of nonnegative $\lambda-$integrable deterministic functions on E such that $||\delta(\omega,t,z)||^r \wedge 1 \leq J_m(z)$ for all $t \leq \tau_m(\omega)$ and $z \in E$.

(ii) The spot volatility process σ_t is of the form

$$\sigma_t = \sigma_{1,t} + \sigma_{2,t}. \tag{3.2}$$

Moreover, both $\sigma_{1,t}$ *and* $c_{1,t} = \sigma_{1,t}\sigma_{1,t}^\intercal \in \mathcal{M}_d$ *are Itô semimartingales of the following Grigelionis form*

$$\begin{aligned}
\sigma_{1,t} &= \sigma_{1,0} + \int_0^t b_s^{(\sigma_1)} ds + \int_0^t \sigma_s^{(\sigma_1)} dW_s + \int_0^t \int_{\mathbb{R}} \delta^{(\sigma_1)}(s,z) 1_{\{||\delta^{(\sigma_1)}(s,z)||>1\}} \mu(ds,dz) \\
&\quad + \int_0^t \int_{\mathbb{R}} \delta^{(\sigma_1)}(s,z) 1_{\{||\delta^{(\sigma_1)}(s,z)||\leq 1\}} (\mu - \nu)(ds,dz) \\
c_{1,t} &= c_{1,0} + \int_0^t b_s^{(c_1)} ds + \int_0^t \sigma_s^{(c_1)} dW_s + \int_0^t \int_{\mathbb{R}} \delta^{(c_1)}(s,z) 1_{\{||\delta^{(c_1)}(s,z)||>1\}} \mu(ds,dz) \\
&\quad + \int_0^t \int_{\mathbb{R}} \delta^{(c_1)}(s,z) 1_{\{||\delta^{(c_1)}(s,z)||\leq 1\}} (\mu - \nu)(ds,dz)
\end{aligned}$$

where W *and* μ *are the same as in (3.1);* $b^{(\sigma_1)}, b^{(c_1)}, \delta^{(\sigma_1)}, \delta^{(c_1)}$ *are* $d \times d-$*dimensional and* $\sigma^{(\sigma_1)}, \sigma^{(c_1)}$ *are* $d \times d \times d-$*dimensional;* $\delta^{(\sigma_1)}, \delta^{(c_1)}$ *are predictable functions such that for a sequence of nonnegative* λ*-integrable functions* $\tilde{J}_m$ *on* E, $||\delta^{(\sigma_1)}(\omega,t,z)||^2 \wedge 1 \leq \tilde{J}_m(z)$ *and* $||\delta^{(c_1)}(\omega,t,z)||^2 \wedge 1 \leq \tilde{J}_m(z)$ *for all* $t \leq \tau_m(\omega)$ *and* $z \in E$.

On the other hand, $\sigma_{2,t}$ *is a stochastic process satisfying, for some* $\epsilon > 0$,

$$\mathbb{E}\left(||\sigma_{2,t} - \sigma_{2,s}||^2\right) \leq K(t-s)^{1+2\epsilon} \tag{3.3}$$

In finance area, X is usually the logarithm of price of a given stock and σ is the associated volatility process. For a proper introduction to Itô semimartingale, see (Jacod and Protter, 2012), Chapter 2. In particular, up to expanding dimensions, it is no restriction to let all Itô semimartingales be driven by the same Brownian motion and Poisson random measure. We note that Assumption 3.1.1 accommodates a large class of models commonly used in finance and economics, which allows for jumps in both price and volatility processes and for arbitrary dependence structure between components within the model. More importantly, the volatility structure (3.2) considered in Assumption 3.1.1 consists of a general Itô semimartingale plus a component satisfying certain regularity conditions, which covers fractional Brownian motion (and related processes) that may be used to capture the long-memory property of the volatility process. We refer the readers to the seminal works by Comte and Renault (1996,1998), which for the first time introduce the modelling of long-memory property in finance area. In view of such a mixture of Itô semimartingale and long-memory process, in the sequel we refer to model (3.2) as the *long-memory Itô semimartingale* (LMIS) volatility model.

On the technical level, as long as σ_1 is an Itô semimartingale, c_1 is also an Itô semimartingale by Itô's formula. The processes $b^{(c_1)}, \sigma^{(c_1)}$ and $\delta^{(c_1)}$ can be expressed as deterministic functions of σ_1, $b^{(\sigma_1)}$, $\sigma^{(\sigma_1)}$ and $\delta^{(\sigma_1)}$, but we do not need this here. On the other hand, as far as the conditions imposed on the process σ_2 is concerned, since σ_2 may not be a martingale any more (e.g., when σ_2 is fractional Brownian motion), the conditional expectation of $\sigma_{2,t} - \sigma_{2,s}$ given $\mathcal{F}_s$ could be difficult to compute and hence complicates the proof of Theorem 3.4.1 below. This is the reason why more smoothness on the second moment (3.3) is needed, which can be seen as compensation for the loss of martingale property.

Now we state the statistical setting in this chapter. At stage n, we assume that the process X is sampled at times $i\Delta_n$ for some time step Δ_n, for $0 \le i \le n \equiv \lfloor T/\Delta_n \rfloor$, within the *fixed* time interval $[0, T]$. For any process Y, the increments of Y are denoted by

$$\Delta_i^n Y \equiv Y_{i\Delta_n} - Y_{(i-1)\Delta_n}, \quad i = 1, \dots, n. \tag{3.4}$$

Below, we consider an infill asymptotic setting, that is, $\Delta_n \to 0$ as $n \to \infty$.

3.2 Integrated volatility functional

With model (3.1), the spot (co)variance process of X is given by $c = \sigma\sigma^\intercal$, which is also $\mathcal{M}_d$-valued. The (random) object of interest considered in this chapter is the *integrated volatility functional* of the form

$$S(g) \equiv \int_0^T g(c_s)ds, \tag{3.5}$$

for some (possibly nonlinear) test function $g : \mathcal{M}_d \to \mathbb{R}$, which is assumed to satisfy the following assumption. Below, for a compact set $\mathcal{K} \subset \mathcal{M}_d$ and $\epsilon > 0$, we denote the ϵ-enlargement about $\mathcal{K}$ by

$$\mathcal{K}^\epsilon \equiv \{M \in \mathcal{M}_d : \inf_{A \in \mathcal{K}} ||M - A|| < \epsilon\}.$$

Assumption 3.2.1. *There exist a localizing sequence of stopping times* $(\tau_m)_{m\geq 1}$ *and a sequence of convex compact subsets* $\mathcal{K}_m \subseteq \mathcal{M}_d$ *such that* $c_t \in \mathcal{K}_m$ *for* $t \leq \tau_m$ *and* $g \in \mathcal{C}^3(\mathcal{K}_m^\epsilon)$*, the space of three times continuously differentiable functions on* $\mathcal{K}_m^\epsilon$ *for some* $\epsilon > 0$.

Assumption 3.2.1 is easily verified in specific setting, which in particular holds, in one-dimensional case, for $g(c) = \log(c)$ or $\sqrt{c}$, provided that both c_t and $1/c_t$ are locally bounded with $\mathcal{K}_m$ being compact intervals on $(0, \infty)$.

Many quantities of interests in finance and econometrics can be written in the form of (3.5), with Assumption 3.2.1 satisfied. For example, when c is scalar, $g(x) = x$ corresponds to the so-called *integrated volatility* $S(g) = \int_0^T c_t dt$, which has been a popular measure of volatility in high-frequency setting, see (Andersen and Bollerslev, 1999), (Andersen et al., 2001b) and (Andersen et al., 2003b). Moreover, $g(x) = x^2$ corresponds to the *integrated quarticity*, which is the (half of) asymptotic variance when using *realized volatility* to approximate integrated volatility. The more generally defined *power variation* $S(g) = \int_0^T c_t^p dt$ for some $p > 0$ is associated with polynomial test function $g(x) = x^p$, see for example, (Barndorff-Nielsen and Shephard, 2003), (Barndorff-Nielsen and Shephard, 2004b) and (Jacod, 2008). In bivariate case, the beta for the diffusive movement of the stock with respect to the market is given by $\beta_t \equiv c_{12,t}/c_{11,t}$, where the market and the stock are labelled by 1 and 2 respectively, with the test function being $g(A) = A_{12,t}/A_{11,t}$ for $A \in \mathcal{M}_2$, see (Mykland and Zhang, 2009). Moreover, the idiosyncratic spot covariance of the stock can thus be expressed as $c_{22,t} - \beta_t^2 c_{11,t} = c_{22,t} - c_{12,t}^2/c_{11,t}$, with test function $g(A) = A_{22,t} - A_{12,t}^2/A_{11,t}$, see (Mykland and Zhang, 2006). Other examples include: correlation/leverage effect (Kalnina and Xiu, 2016), volatility Laplace transform (Todorov and Tauchen, 2012b), variance betas (Li et al., 2016b), eigenvalues (Aït-Sahalia and Xiu, 2015). Moreover, general forms of $S\left(g\right)$ also serve as integrated moment conditions in specification tests and estimation problems in economic models (Li and Xiu, 2016), following which we will consider a functional version of function g in Chapter 5. We also note that early discussion on the estimation of diffusion process and the sampling frequency of data goes back to (Merton, 1980) and (Zhou, 1996).

In order to give the estimator of $S(g)$ for a given function g, we first nonparametrically recover the spot variance $c_{i\Delta_n}$ by employing a local average of sum of squared truncated returns (see (Jacod and Protter, 2012), Chapter 9 and 13), that is, for any $0 \leq i \leq N_n \equiv [T/\Delta_n] - k_n$, let

$$\hat{c}^{lm}_{i\Delta_n} \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \Delta^n_{i+j} X^l \Delta^n_{i+j} X^m 1_{\{||\Delta^n_{i+j} X|| \leq u_n\}}$$

where $1 \leq l, m \leq d$, k_n is a sequence of integers that goes to infinity representing the number of increments employed in a local window and u_n determines the truncation threshold for eliminating jumps in X, see (Mancini, 2001) and (Mancini, 2009). If X is continuous, then there is no need to truncate in forming $\hat{c}$ by taking $u_n = \infty$. The conditions on tuning parameters k_n and u_n are collected in Assumption 3.2.2. We note that the study of spot covariance estimation dates back to (Foster and Nelson, 1996), which features a continuous setting; one can also see (Aït-Sahalia and Jacod, 2014) on this topic in a more general setting.

Assumption 3.2.2. *$k_n \sim \Delta_n^{-\gamma}$ and $u_n \sim \Delta_n^{\varpi}$ for some constants γ and ϖ satisfying*

$$\frac{r}{2} \vee \frac{1}{3} < \gamma < \frac{1}{2}, \quad \frac{1-\gamma}{2-r} \leq \varpi < \frac{1}{2}.$$

In particular, $k_n \Delta_n \to 0$ and $k_n^2 \Delta_n \to \infty$.

We then define the estimator for $S(g)$ as

$$S_n(g) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \left(g(\hat{c}_{i\Delta_n}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}_{i\Delta_n}) \times (\hat{c}^{jl}_{i\Delta_n} \hat{c}^{km}_{i\Delta_n} + \hat{c}^{jm}_{i\Delta_n} \hat{c}^{kl}_{i\Delta_n}) \right) \quad (3.6)$$

Assuming that the volatility process is an Itô semimartingale, (Jacod and Rosenbaum, 2013a) shows a CLT for $S_n(g)$ approximating $S(g)$ with rate $\sqrt{\Delta_n}$ provided test function g (and its derivative) satisfy a certain growth condition. (Li et al., 2016a) extends the CLT result to the case of g only satisfying Assumption 3.2.1, and (Li and Xiu, 2016) shows an empirical-process-type CLT in a similar setting, while both papers still assuming volatility process is an Itô semimartingale. In contrast, in this chapter we want to derive an associated CLT for $S_n(g)$ approximating $S(g)$ with convergence rate $\sqrt{\Delta_n}$ and the same asymptotic variance as in the aforementioned papers for a larger class of volatility processes given as LMIS (3.2).

3.3 Examples

In this section we provides some concrete examples for the process σ_2 satisfying certain regularity conditions introduced in Assumption 3.1.1. We begin with fractional Brownian motion in Section 3.3.1, and then proceed to Wiener integrals with respect to fractional Brownian motion in Section 3.3.2.

3.3.1 Fractional Brownian motion

Fractional Brownian motion (fBm) $(B_t^H)_{t\geq 0}$ with Hurst index $H \in (0,1)$ is a centered Gaussian process with the covariance function

$$\mathbb{E}(B_t^H B_s^H) = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}),$$

where for simplification we assume $B_0^H = 0$. The process was introduced by (Kolmogorov, 1940), followed by pioneering works including (Hurst, 1951), (Hurst, 1956) and (Mandelbrot, 1983). Fractional Brownian motion has been widely used in hydrology, engineering and finance. When $H = \frac{1}{2}$, the process reduces to the usual standard Brownian motion.

For a more comprehensive description of fractional Brownian motion, see, e.g., (Duncan et al., 2000), (Nualart, 2005), (Nualart, 2006) and (Mishura, 2008). We briefly summarize some important properties of fractional Brownian motion below:

1. Self-similarity: for any $a > 0$, $\{B_{au}^H, u \in \mathbb{R}\} \stackrel{d}{=} a^H \{B_u^H, u \in \mathbb{R}\}$, where $\stackrel{d}{=}$ denotes the equality in any finite-dimensional distributions. This property can be regarded as a "fractal property" in probability.

2. Stationary increments and moment estimates: From definition it follows that the increment of B^H over a finite time interval $[s,t]$ is normally distributed with mean zero and variance
$$\mathbb{E}\left(\left(B_t^H - B_s^H\right)^2\right) = |t-s|^{2H}.$$

Indeed, for any integer $k \geq 1$, we have

$$\mathbb{E}\left(\left(B_t^H - B_s^H\right)^{2k}\right) = \frac{(2k)!}{k!2^k}|t-s|^{2Hk}.$$

Then by Kolmogorov's continuity criterion (e.g. (Revuz and Yor, 1999), Theorem 2.1 in Chapter 1), B^H has a version whose sample paths are $\gamma-$Hölder continuous for any $\gamma < H$.

More generally (see, e.g., Corollary 3.11 in (Duncan et al., 2000)), for any $\alpha > 1$, there is a $C_\alpha < \infty$ such that

$$\mathbb{E}\left|B_t^H - B_s^H\right|^\alpha \leq C_\alpha |t-s|^{\alpha H}. \tag{3.7}$$

3. Long memory property when $H > \frac{1}{2}$: Let $r(n) \equiv \operatorname{Cov}\left(B_1^H, B_{n+1}^H - B_n^H\right)$ be the autocovariance function, then if $H > 1/2$, we have $\sum_{n=1}^\infty r(n) = \infty$, in which case we call that the fractional Brownian motion exhibits long-range dependence.

4. Non-semimartingale: B^H is not a semimartingale when $H \neq 1/2$. This can be proved by studying the $p-$th variation of B^H, see, e.g., Proposition 7.1.1 of (Pipiras and Taqqu, 2016).

5. Prediction formula: The conditional expectation of B_t^H given the past information is given as (3.8). This is first proved by (Gripenberg and Norros, 1996) for the case $H > 1/2$, and extended to $H \in (0, 1/2)$ by (Pipiras and Taqqu, 2001) (Theorem 7.1). To state the result more easily, let $\kappa = H - 1/2$, then for any $0 \leq s \leq t \leq T$ and $\kappa \in (-1/2, 1/2)$, we have

$$\mathbb{E}\left(B_t^\kappa \middle| B_v^\kappa, v \in [0,s]\right) = B_s^\kappa + \int_0^s \Psi^\kappa(s,t,v) dB_v^\kappa, \tag{3.8}$$

where for $v \in (0,t)$,

$$\Psi^\kappa(s,t,v) = \frac{\sin(\pi\kappa)}{\pi} v^{-\kappa}(s-v)^{-\kappa} \int_s^t \frac{z^\kappa (z-s)^\kappa}{z-v} dz.$$

We note that the function $\Psi^{\kappa}(s,t,v)$ is related to the so-called Appell's hypergeometric function, see e.g., Remark 6.4.5 in (Pipiras and Taqqu, 2016).

(Comte and Renault, 1996) and (Comte and Renault, 1998) for the first time introduced fractional Brownian motion to modelling the price and volatility processes of financial assets. In fact, the long memory property of volatility processes has been documented in economics and finance for a long time. As one may expect, it is the long range dependence property described above that makes fractional Brownian motion an ideal stochastic process to capture such features exhibited by volatility. As a consequence, in the remainder of this chapter we will only consider fractional Brownian motion B^H with $H > 1/2$, as well as its continuous version. Then B^H satisfies the conditions imposed on σ_2 in Assumption 3.1.1.

Remark 3.3.1. On the technical level, since B^H is not a semimartingale, and also in light of the prediction formula (3.8), it would be rather hard to verify the estimate

$$|\mathbb{E}\left(B^H_{t+s} - B^H_t \middle| \mathcal{F}_t\right)| \leq Ks$$

for some constant K, which is always true for any Itô semimartingale under certain boundedness conditions (see Lemma 2 in the Proofs). The lack of such an estimate complicates the proof of Theorem 3.4.1 when $\sigma_2 = B^H$. However, the difficulty is overcome by the more smoothness B^H provides as shown in (3.7) when $H > 1/2$.

3.3.2 Wiener integrals w.r.t. fractional Brownian motion

In this example, we focus on the integral over an finite interval $[0,T]$ of the form

$$\int_0^t f(u)dB^H_u, \quad 0 \leq t \leq T, \tag{3.9}$$

where B^H is a fractional Brownian motion with $H > 1/2$. Since fractional Brownian motion is no longer semimartingale, the usual theory of Itô integral cannot be applied to defining stochastic integrals with respect to fractional Brownian motion. Indeed, one needs the theory of Malliavin calculus to define such stochastic integral. However, when $f(u)$ is a deterministic function, (3.9) can be defined in a relatively easier fashion, in which case (3.9) is called fractional Wiener integral.

The integration with respect to general Gaussian processes has been studied for a long time and we refer readers to (Huang and Cambanis, 1978) for an extensive presentation. (Pipiras and Taqqu, 2000) discussed some related questions of Wiener integral of deterministic integrand w.r.t fractional Brownian motion over the real line, and (Pipiras and Taqqu, 2000) discussed that of over a finite interval, which is the case of (3.9) we consider here. The basic idea is to define (3.9) first with f being elementary(step) functions, and then extend the definition to some bigger classes of f in the $L^2(\Omega)$ sense using isometry. Indeed, when $H > 1/2$, (3.9) can be defined for each of the following four increasing classes of integrands:

$$L^2[0,T] \subset L^{1/H}[0,T] \subset |\Lambda|^{\kappa}_T \subset \Lambda^{\kappa}_T,$$

where $\kappa = H - 1/2$ and

$$\begin{aligned} \Lambda^{\kappa}_T &\equiv \left\{ f : [0,T] \to \mathbb{R} \text{ such that } \int_0^T [s^{-\kappa}(I^{\kappa}_{T-}u^{\kappa}f(u))(s)]^2 ds < \infty \right\}, \\ |\Lambda|^{\kappa}_T &\equiv \left\{ f : [0,T] \to \mathbb{R} \text{ such that } \int_0^T \int_0^T |f(u)||f(v)||u-v|^{2\kappa-1} du dv < \infty \right\}. \end{aligned}$$

Here I^{κ}_{T-} is (right-sided) fractional integral operator of order κ defined as

$$(I^{\kappa}_{T-}f)(s) = \frac{1}{\Gamma(\kappa)} \int_s^T f(u)(u-s)^{\kappa-1} du, \quad s \in (0,T), \quad f \in L^1[0,T].$$

For definition of (3.9) for each specific class of integrands, see (Pipiras and Taqqu, 2001) and references therein. For our purpose, it would be sufficient to consider the space $L^{1/H}[0,T]$, as seen from the properties of (3.9) listed below. Moreover, as the theory of fractional Brownian motion is closed related to fractional integrals and fractional derivatives, we refer readers to (Samko and Marichev, 1993) for a comprehensive treatment on this topic.

When $H > 1/2$ (or equivalent $\kappa > 0$), (3.9) has the following properties:

1. Continuity: As explained in (Mishura, 2008), Section 1.11, when $f \in L^{1/H}[0,T]$, the process $\left\{ \int_0^t f(u) dB^H_u \right\}_{0 \le t \le T}$ is continuous in t.

2. Moment estimates: For any $p \geq 1, 0 \leq a < b \leq \infty$, define

$$\|f\|_{L^p(a,b)} \equiv \left(\int_a^b |f(u)|^p du \right)^{1/p}.$$

As proved in Theorem 1.1 in (Mémin et al., 2001), there exists a constant $c(H, \alpha)$ such that for every $\alpha > 0$ and for every a, b with $0 \leq a < b < \infty$, we have

$$\mathbb{E}\left(\left| \int_a^b f(u) dB_u^H \right|^\alpha \right) \leq c(H, \alpha) \|f\|_{L^{1/H}(a,b)}^\alpha. \tag{3.10}$$

3. Prediction formula: Similar to (3.8), for any $0 \leq s < t \leq T$ and $f \in \Lambda_T^\kappa$, we have

$$\mathbb{E}\left(\int_0^t f(v) dB_v^\kappa \middle| B_v^\kappa, v \in [0, s] \right) = \int_0^s f(v) dB_v^\kappa + \int_0^s \Psi_f^\kappa(s, t, v) dB_v^\kappa, \tag{3.11}$$

where for $v \in (0, t)$,

$$\Psi_f^\kappa(s, t, v) = \frac{\sin(\pi\kappa)}{\pi} v^{-\kappa}(s - v)^{-\kappa} \int_s^t \frac{z^\kappa (z - s)^\kappa}{z - v} f(z) dz.$$

(3.11) is proved in Lemma 1 of (Duncan, 2006). One can see (Fink et al., 2013) for derivation of conditional variance, .

By moment estimates (3.10), if f is uniformly bounded from above by some constant $M > 0$ (as in Assumption 3.5.1 in the Proofs), we have

$$\mathbb{E}\left(\left| \int_a^b f(u) dB_u^H \right|^\alpha \right) \leq c(H, \alpha) \|f\|_{L^{1/H}(a,b)}^\alpha \equiv \left(\int_a^b |f(u)|^{1/H} du \right)^{H\alpha} \leq c(H, \alpha) M^\alpha (b - a)^{H\alpha},$$

which is the RHS of (3.7). Therefore, together with the continuity in time, the process $\left\{ \int_0^t f(u) dB_u^H \right\}_{0 \leq t \leq T}$ satisfies the conditions imposed on σ_2.

3.4 Results

In this section we state our main result. In order for the CLT to hold under the setting described in Section 3.1, we need one more assumption on the smoothness of the second moment of σ_2.

Assumption 3.4.1. *Assume* $\epsilon > \frac{\gamma}{2(1-\gamma)}$.

Under Assumption 3.4.1, it holds that $k_n^{1+\epsilon}\Delta_n^{\epsilon} \to 0$, which plays an important role in the proof of Theorem 3.4.1. In light of Assumption 3.2.2, Assumption 3.4.1 implies that $\epsilon > 1/4$, which in the case of fractional Brownian motion corresponds to the Hurst parameter $H > 3/4$. Such a requirement is consistent with the empirical results documented in (Comte and Renault, 1998), (Andersen and Bollerslev, 1997), (Andersen et al., 2001b) and (Bollerslev et al., 2013). In particular, those papers estimate the fractional parameter of the underlying volatility process under both low frequency and high frequency settings, and all of the estimated fractional parameters have a value larger than 0.25.

Now we state our main result.

Theorem 3.4.1. *Under Assumptions 3.1.1-3.4.1, it holds that*

$$\frac{1}{\sqrt{\Delta_n}}(S_n(g) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g)),$$

where $\mathcal{MN}(0, V)$ *is a centered mixed normal distribution with conditional variance*

$$V(g) = \sum_{j,k,l,m=1}^{d} \int_0^T \partial_{jk} g(c_s) \partial_{lm} g(c_s) \left(c_s^{jl} c_s^{km} + c_s^{jm} c_s^{kl} \right) ds.$$

We give some comments as follows on Theorem 3.4.1.

1. Theorem 3.4.1 extends the results in (Jacod and Rosenbaum, 2013a), (Li et al., 2016a) and (Li and Xiu, 2016) by establishing the asymptotic distribution of $S_n(g)$ under the LMIS volatility dynamics. In particular, in the absence of the long-memory component σ_2, Theorem 3.4.1 coincides with those in prior work. Since both the convergence rate and the asymptotic variance remain the same as shown in those papers, the estimator $S_n(g)$ is still *efficient* in the sense of (Jacod and Rosenbaum, 2013a) under the more general LMIS volatility dynamics.

2. The "cost" of including the long-memory component is that we need an additional upper bound for the divergence rate of the local window size k_n, that is, $\gamma < 2\epsilon/(1+2\epsilon)$. This restriction is weaker when ϵ is larger, which corresponds to the case with "longer"

memory. In the extreme case with $\epsilon = 1/2$ (i.e., σ_2 has locally Lipschitz path under the L_2-norm), this restriction is absent, because σ_2 then behaves essentially like a drift term.

3. As already pointed previously, the condition (3.4.1) implicitly imposes a restriction on ϵ, that is, $\epsilon > 1/4$. In other words, σ_2 is Hölder-continuous under the L_2-norm with an index at least 3/4, which shows an apparent discrepancy relative to the 1/2 Hölder continuity of the Itô semimartingale component. This "gap" arises as a compensation for the lack of martingale property in the long-memory component, whereas the martingale property is heavily exploited in previous work based on the Itô semimartingale volatility dynamics. The proofs in the more general LMIS framework thus contains nontrivial additional complications.

4. Last but not least, as far as the application of Theorem 3.4.1 is concerned, one can conduct statistical inference, constructing confidence interval for example, for $S(g)$. More specifically, as shown in (Jacod and Rosenbaum, 2013a), a consistent estimator for the asymptotic variance $V(g)$ is given by

$$\widetilde{S}_n(\bar{h}, \mathcal{D}_n) \equiv \Delta_n \sum_{i=0}^{[T/\Delta_n]-k_n} \bar{h}(\hat{c}_{i\Delta_n}),$$

where $\bar{h}(x) \equiv \sum_{j,k,l,m=1}^{d} \partial_{jk} g(x) \partial_{lm} g(x) \left(x^{jl}x^{km} + x^{jm}x^{kl}\right)$. In particular, $\widetilde{S}_n(\bar{h}, \mathcal{D}_n)$ is robust to the long memory assumption of volatility. Then it follows that

$$\widetilde{S}_n(\bar{h}, \mathcal{D}_n) \xrightarrow{\mathbb{P}} V(g),$$

and hence

$$\frac{S_n(g, \mathcal{D}_n) - S(g)}{\sqrt{\Delta_n \widetilde{S}_n(\bar{h}, \mathcal{D}_n)}} \xrightarrow{\mathcal{L}\text{-}s} \mathcal{N}(0, 1),$$

Confidence intervals for $S(g)$ can be constructed accordingly.

3.5 Proofs

Throughout this section, we use K to denote a generic constant that may change from line to line; we sometimes emphasize the dependence of this constant on some parameter q by writing K_q. Recall that $N_n \equiv [T/\Delta_n] - k_n$, we write $\sum_i$ for $\sum_{i=0}^{N_n}$ for simplicity.

3.5.1 Preliminaries

By a standard localization procedure (see Lemma 4.4.9 in (Jacod and Protter, 2012)), it is enough to show Theorem 3.4.1 under a stronger version of Assumption 3.1.1.

Assumption 3.5.1. *We have Assumption 3.1.1. The process σ takes value in a convex compact set of $\mathcal{M}_d$. Moreover, the processes $b, b^{(\sigma_1)}, b^{(c_1)}$ and $\sigma^{(\sigma_1)}, \sigma^{(c_1)}$ are bounded and there is a bounded $\lambda-$integrable function $J : E \to \mathbb{R}$, such that for all $\omega \in \Omega, t \in [0, T]$ and $z \in E$ we have $||\delta(\omega, t, z)||^r \leq J(z)$ and $||\delta^{(\sigma_1)}(\omega, t, z)||^2 \vee ||\delta^{(\sigma_1)}(\omega, t, z)||^2 \leq J(z)$.*

In the following analysis, it would be much more convenient to consider the continuous part of the process X_t defined by

$$X'_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad 0 \leq t \ \leq T.$$

Accordingly, define for each $i = 0, 1, \ldots, N_n$,

$$\hat{c}'_{i\Delta_n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \Delta^n_{i+j} X' \Delta^n_{i+j} X'^{\intercal}.$$

Then we introduce the following notations that will be used throughout the Proofs, most of which are analogues to those used in (Jacod and Rosenbaum, 2013a):

$$\begin{aligned}
\alpha_{n,i} &\equiv \Delta^n_i X' \Delta^n_i X'^{\intercal} - c_{(i-1)\Delta_n} \Delta_n \\
\tilde{c}_{n,i} &\equiv \hat{c}'_{i\Delta_n} - c_{i\Delta_n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (\alpha_{n,i+j} + (c_{(i+j-1)\Delta_n} - c_{i\Delta_n})\Delta_n)
\end{aligned}$$

With any process Z we associate the variables

$$
\begin{aligned}
\eta_{t,s}(Z) &\equiv \sup_{v\in(0,s]} ||Z_{t+v} - Z_t||^2 \\
\eta_{i,j}^n(Z) &\equiv \sqrt{\mathbb{E}\left(\eta_{(i-1)\Delta_n, j\Delta_n}(Z)\big|\mathcal{F}_{(i-1)\Delta_n}\right)} \\
\eta_i^n(Z) &\equiv \eta_{i,k_n}^n(Z)
\end{aligned}
$$

and we recall Lemma 3.1 of (Jacod and Rosenbaum, 2013b).

Lemma 3.5.1. *For all $t > 0$ and all bounded càdlàg processes Z, we have $\Delta_n \mathbb{E}(\sum_{i=1}^{[t/\Delta_n]} \eta_i^n(Z)) \to 0$, and for all j,k such that $j+k \le k_n$, we have $\mathbb{E}\left(\eta_{i+j,k}^n(Z)|\mathcal{F}_{(i-1)\Delta_n}\right) \le \eta_i^n(Z)$.*

We collect some standard estimates for Itô semimartingale in the following lemma, the proof of which depends heavily on the decomposition of $c_t - c_s$ for $0 \le s < t \le T$,

$$
\begin{aligned}
c_t - c_s &= (\sigma_{1,t} + \sigma_{2,t})(\sigma_{1,t} + \sigma_{2,t})^\intercal - (\sigma_{1,s} + \sigma_{2,s})(\sigma_{1,s} + \sigma_{2,s})^\intercal \\
&= (\sigma_{1,t}\sigma_{1,t}^\intercal - \sigma_{1,s}\sigma_{1,s}^\intercal) + (\sigma_{2,t}\sigma_{2,t}^\intercal - \sigma_{2,s}\sigma_{2,s}^\intercal) + (\sigma_{1,t}\sigma_{2,t}^\intercal + \sigma_{2,t}\sigma_{1,t}^\intercal - \sigma_{1,s}\sigma_{2,s}^\intercal - \sigma_{2,s}\sigma_{1,s}^\intercal) \\
&= c_{1,t} - c_{1,s} + c_{2,t} - c_{2,s} + (\sigma_{1,t} - \sigma_{1,s})\sigma_{2,s}^\intercal + \sigma_{1,t}(\sigma_{2,t}^\intercal - \sigma_{2,s}^\intercal). \qquad (3.12) \\
&\quad + \sigma_{2,s}(\sigma_{1,t}^\intercal - \sigma_{1,s}^\intercal) + (\sigma_{2,t} - \sigma_{2,s})\sigma_{1,t}^\intercal
\end{aligned}
$$

Notice that the third and fourth terms are transposes of the fifth and sixth terms, respectively. Moreover, we have for $i = 1, 2$ (in fact we only need $i = 2$ below, as $c_{1,t}$ is itself an Itô semimartingale by Itô's lemma),

$$
\begin{aligned}
c_{i,t} - c_{i,s} &= \sigma_{i,t}\sigma_{i,t}^\intercal - \sigma_{i,s}\sigma_{i,s}^\intercal \\
&= (\sigma_{i,t} - \sigma_{i,s})\sigma_{i,t}^\intercal + \sigma_{i,s}(\sigma_{i,t}^\intercal - \sigma_{i,s}^\intercal) \qquad (3.13)
\end{aligned}
$$

One will see both decompositions (3.12) and (3.13) will be repeatedly used in the sequel.

Lemma 3.5.2. *Under Assumption 3.5.1, we have*

(1) for any $s,t \geq 0$ and $q \geq 0$,

$$\begin{aligned}
\mathbb{E}\left(\sup_{v\in[0,s]} ||X'_{t+v} - X'_t||^q \big| \mathcal{F}_t\right) &\leq K_q s^{q/2}, \\
||\mathbb{E}\left(X'_{t+s} - X'_t \big| \mathcal{F}_t\right)|| &\leq Ks, \\
\mathbb{E}\left(\sup_{v\in[0,s]} ||\sigma_{1,t+v} - \sigma_{1,t}||^q \big| \mathcal{F}_t\right) &\leq K_q s^{1\wedge q/2}, \\
||\mathbb{E}\left(\sigma_{1,t+s} - \sigma_{1,t} \big| \mathcal{F}_t\right)|| &\leq Ks, \\
\mathbb{E}\left(\sup_{v\in[0,s]} ||c_{1,t+v} - c_{1,t}||^q \big| \mathcal{F}_t\right) &\leq K_q s^{1\wedge q/2}, \\
||\mathbb{E}\left(c_{1,t+s} - c_{1,t} \big| \mathcal{F}_t\right)|| &\leq Ks.
\end{aligned}$$

(2) Let $c_{2,t} = \sigma_{2,t}\sigma_{2,t}^\intercal$, for any $s,t \geq 0$,

$$\begin{aligned}
\mathbb{E}\left(||\sigma_{2,t} - \sigma_{2,s}||\right) &\leq K(t-s)^{1/2+\epsilon} \\
\mathbb{E}\left(||c_{2,t+s} - c_{2,t}||^q\right) &\leq K\mathbb{E}\left(||\sigma_{2,t+s} - \sigma_{2,t}||^q\right), \quad q = 1,2,3,4 \\
||\mathbb{E}\left(c_{t+s} - c_t \big| \mathcal{F}_t\right)|| &\leq K\left(||\mathbb{E}\left(c_{1,t+s} - c_{1,t} \big| \mathcal{F}_t\right)|| + ||\mathbb{E}\left(\sigma_{1,t+s} - \sigma_{1,t} \big| \mathcal{F}_t\right)||\right. \\
&\quad \left. + \mathbb{E}\left(||\sigma_{2,t+s} - \sigma_{2,t}|| \big| \mathcal{F}_t\right)\right) \\
\mathbb{E}\left(||c_{t+s} - c_t||^q\right) &\leq K\left(\mathbb{E}\left(||\sigma_{1,t+s} - \sigma_{1,t}||^q\right) + \mathbb{E}\left(||\sigma_{2,t+s} - \sigma_{2,t}||^q\right)\right), q = 1,2,3,4.
\end{aligned}$$

In particular, as one can see from the proof of the third estimate in (2), it would suffice to consider $\sigma_{1,t} - \sigma_{1,s}$ and $\sigma_{2,t} - \sigma_{2,s}$ when it comes to the difference $c_t - c_s$ in the sequel.

Proof. The estimates in (1) follow from (4.3) in (Jacod and Rosenbaum, 2013a), as X' has no jump part and both $\sigma_{1,t}$ and $c_{1,t}$ are Itô semimartingales.

For part (2), the first estimate is implied by (3.3). For $q = 1,2,3,4$, by (3.13) and the fact that for any matrix A, $||A|| = ||A^\intercal||$ and that σ_2 is bounded, it follows

$$\begin{aligned}
\mathbb{E}\left(||c_{2,t+s} - c_{2,t}||^q\right) &\leq \mathbb{E}\left(||(\sigma_{2,t+s} - \sigma_{2,t})\sigma_{2,t+s}^\intercal + \sigma_{2,t}(\sigma_{2,t+s}^\intercal - \sigma_{2,t}^\intercal)||^q\right) \\
&\leq K\mathbb{E}\left(||(\sigma_{2,t+s} - \sigma_{2,t})\sigma_{2,t+s}^\intercal||^q\right) + K\mathbb{E}\left(||\sigma_{2,t}(\sigma_{2,t+s}^\intercal - \sigma_{2,t}^\intercal)||^q\right) \\
&\leq K\mathbb{E}\left(||(\sigma_{2,t+s} - \sigma_{2,t})||^q\right).
\end{aligned}$$

Hence the second estimate is proved. For the third one, by (3.12)

$$
\begin{aligned}
||\mathbb{E}\left(c_{t+s}-c_t\middle|\mathcal{F}_t\right)|| &= ||\mathbb{E}\left(c_{1,t+s}-c_{1,t}+c_{2,t+s}-c_{2,t}\right.\\
&\quad +(\sigma_{1,t+s}-\sigma_{1,t})\sigma_{2,t}^{\intercal}+\sigma_{1,t+s}(\sigma_{2,t+s}^{\intercal}-\sigma_{2,t}^{\intercal})\\
&\quad \left.+\sigma_{2,t}(\sigma_{1,t+s}^{\intercal}-\sigma_{1,t}^{\intercal})+(\sigma_{2,t+s}-\sigma_{2,t})\sigma_{1,t+s}^{\intercal}\middle|\mathcal{F}_t\right)||\\
&\leq ||\mathbb{E}\left(c_{1,t+s}-c_{1,t}\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left((\sigma_{1,t+s}-\sigma_{1,t})\sigma_{2,t}^{\intercal}\middle|\mathcal{F}_t\right)||\\
&\quad +||\mathbb{E}\left(\sigma_{2,t}(\sigma_{1,t+s}^{\intercal}-\sigma_{1,t}^{\intercal})\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left(c_{2,t+s}-c_{2,t}\middle|\mathcal{F}_t\right)||\\
&\quad +||\mathbb{E}\left(\sigma_{1,t+s}(\sigma_{2,t+s}^{\intercal}-\sigma_{2,t}^{\intercal})\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left((\sigma_{2,t+s}-\sigma_{2,t})\sigma_{1,t+s}^{\intercal}\middle|\mathcal{F}_t\right)||\\
&= ||\mathbb{E}\left(c_{1,t+s}-c_{1,t}\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left((\sigma_{1,t+s}-\sigma_{1,t})\middle|\mathcal{F}_t\right)\sigma_{2,t}^{\intercal}||\\
&\quad +||\sigma_{2,t}\mathbb{E}\left((\sigma_{1,t+s}^{\intercal}-\sigma_{1,t}^{\intercal})\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left(c_{2,t+s}-c_{2,t}\middle|\mathcal{F}_t\right)||\\
&\quad +||\mathbb{E}\left(\sigma_{1,t+s}(\sigma_{2,t+s}^{\intercal}-\sigma_{2,t}^{\intercal})\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left(\sigma_{2,t+s}-\sigma_{2,t}\sigma_{1,t+s}^{\intercal}\middle|\mathcal{F}_t\right)||\\
&\leq K\left(||\mathbb{E}\left(c_{1,t+s}-c_{1,t}\middle|\mathcal{F}_t\right)||+K||\mathbb{E}\left((\sigma_{1,t+s}-\sigma_{1,t})\middle|\mathcal{F}_t\right)||\right)\\
&\quad +K\left(\mathbb{E}\left(||c_{2,t+s}-c_{2,t}||\middle|\mathcal{F}_t\right)+\mathbb{E}\left(||\sigma_{2,t+s}-\sigma_{2,t}||\middle|\mathcal{F}_t\right)\right)\\
&\leq K\left(||\mathbb{E}\left(c_{1,t+s}-c_{1,t}\middle|\mathcal{F}_t\right)||+||\mathbb{E}\left(\sigma_{1,t+s}-\sigma_{1,t}\middle|\mathcal{F}_t\right)||\right.\\
&\quad \left.+\mathbb{E}\left(||\sigma_{2,t+s}-\sigma_{2,t}||\middle|\mathcal{F}_t\right)\right),
\end{aligned}
$$

as desired.

As for the last estimate, by (3.15) and the fact that c_1, σ_1 and σ_2 are all bounded,

$$
\begin{aligned}
\mathbb{E}\left(||c_{t+s}-c_t||^q\right) &\leq K\left(\mathbb{E}\left(||c_{1,t+s}-c_{1,t}||^q\right)+\mathbb{E}\left(||\sigma_{1,t+s}-\sigma_{1,t}||^q\right)+\mathbb{E}\left(||\sigma_{2,t+s}-\sigma_{2,t}||^q\right)\right)\\
&\leq K\left(\mathbb{E}\left(||\sigma_{1,t+s}-\sigma_{1,t}||^q\right)+\mathbb{E}\left(||\sigma_{2,t+s}-\sigma_{2,t}||^q\right)\right).
\end{aligned}
$$

□

3.5.2 Proof of Theorem 3.4.1

In light of a spatial localization argument as in the proof of Theorem 2 in (Li et al., 2016a), we assume that the test function g satisfying Assumption 3.2.1 is indeed compactly supported, and thus both function g and its existent derivatives are bounded from above by some positive constant. This assumption will not be recalled in the sequel.

The next two results are analogues to Lemma 4.1 in (Jacod and Rosenbaum, 2013a), but more involved since the volatility process σ under Assumption 3.1.1 may not be Itô semimartingale any more.

Lemma 3.5.3. *Under Assumption 3.5.1, we have*

$$
\begin{aligned}
&\left|\mathbb{E}\left(\Delta_i^n X'^j \Delta_i^n X'^m | \mathcal{F}_{(i-1)\Delta_n}\right) - c^{jm}_{(i-1)\Delta_n}\Delta_n\right| \\
&\quad \leq K\Delta_n^{3/2}\left(\eta^n_{i,1}(b) + \sqrt{\Delta_n}\right) + K\int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}\left(||\sigma_{2,s} - \sigma_{2,(i-1)\Delta_n}|| \middle| \mathcal{F}_{(i-1)\Delta_n}\right) ds \\
&\mathbb{E}\left|\mathbb{E}\left(\Delta_i^n X'^j \Delta_i^n X'^m | \mathcal{F}_{(i-1)\Delta_n}\right) - c^{jm}_{(i-1)\Delta_n}\Delta_n\right| \\
&\quad \leq K\Delta_n^{3/2}\left(\mathbb{E}\left(\eta^n_{i,1}(b)\right) + \sqrt{\Delta_n} + \Delta_n^{\epsilon}\right)
\end{aligned}
$$

Proof. For the first claim, by using Itô's formula for $f(x, y) = xy$, we have

$$
\begin{aligned}
& \Delta_i^n X'^j \Delta_i^n X'^m - c^{jm}_{(i-1)\Delta_n} \\
= \;& b^j_{(i-1)\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}(X'^m_s - X'^m_{(i-1)\Delta_n})ds + b^m_{(i-1)\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n}(X'^j_s - X'^j_{(i-1)\Delta_n})ds \\
+ \;& \int_{(i-1)\Delta_n}^{i\Delta_n}(X'^m_s - X'^m_{(i-1)\Delta_n})(b^j_s - b^j_{(i-1)\Delta_n})ds + \int_{(i-1)\Delta_n}^{i\Delta_n}(X'^j_s - X'^j_{(i-1)\Delta_n})(b^m_s - b^m_{(i-1)\Delta_n})ds \\
+ \;& \int_{(i-1)\Delta_n}^{i\Delta_n}(c^{jm}_s - c^{jm}_{(i-1)\Delta_n})ds + M_{i\Delta_n} \qquad (3.14)
\end{aligned}
$$

where

$$
M_t = \int_{(i-1)\Delta_n}^{t}(X'^m_s - X'^m_{(i-1)\Delta_n})[\sigma_s]_{j,\cdot}dW_s + \int_{(i-1)\Delta_n}^{t}(X'^j_s - X'^j_{(i-1)\Delta_n})[\sigma_s]_{m,\cdot}dW_s
$$

is martingale vanishing at time $(i-1)\Delta_n$, and $[\sigma_s]_{j,\cdot}$ denotes the $j-$th row of the matrix σ_s. Since b is bounded, by Lemma 3.5.2 and Cauchy-Schwartz inequality, the absolute value of $\mathcal{F}_{(i-1)\Delta_n}$-conditional expectation of the first four terms of RHS of (3.14) is of order

$$
\Delta_n^{3/2}\eta^n_{i,1}(b) + \Delta_n^2.
$$

Now, for the term $\int_{(i-1)\Delta_n}^{i\Delta_n}(c_s^{jm} - c_{(i-1)\Delta_n}^{jm})ds$, note that

$$c^{jm} \quad = \quad c_1^{jm} + c_2^{jm} + \sum_{i=1}^{d}\left(\sigma_1^{ji}\sigma_2^{im} + \sigma_2^{ji}\sigma_1^{im}\right)$$

and hence for any $0 \leq s < t \leq T$, we have a component-wise analogue of (3.12)

$$\begin{aligned} c_t^{jm} - c_s^{jm} \quad = \quad & (c_{1,t}^{jm} - c_{1,s}^{jm}) + (c_{2,t}^{jm} - c_{2,s}^{jm}) + \sum_{i=1}^{d}\Big((\sigma_{1,t}^{ji} - \sigma_{1,s}^{ji})\sigma_{2,s}^{im} + \sigma_{1,t}^{ji}(\sigma_{2,t}^{im} - \sigma_{2,s}^{im}) \\ & + \sigma_{2,s}^{ji}(\sigma_{1,t}^{im} - \sigma_{1,s}^{im}) + (\sigma_{2,t}^{ji} - \sigma_{2,s}^{ji})\sigma_{1,t}^{im}\Big), \end{aligned} \tag{3.15}$$

as well as a component-wise analogue of (3.13)

$$\begin{aligned} c_{2,t}^{jm} - c_{2,s}^{jm} \quad & = \quad \sum_{i=1}^{d}\left(\sigma_{2,t}^{ji}\sigma_{2,t}^{im} - \sigma_{2,s}^{ji}\sigma_{2,s}^{im}\right) \\ & = \quad \sum_{i=1}^{d}\left((\sigma_{2,t}^{ji} - \sigma_{2,s}^{ji})\sigma_{2,t}^{im} + \sigma_{2,s}^{ji}(\sigma_{2,t}^{im} - \sigma_{2,s}^{im})\right) \end{aligned} \tag{3.16}$$

Then in view of (3.16) and Lemma 2 (2), it yields

$$\begin{aligned} & \int_{(i-1)\Delta_n}^{i\Delta_n}\left|\mathbb{E}\left(c_s^{jm} - c_{(i-1)\Delta_n}^{jm}|\mathcal{F}_{(i-1)\Delta_n}\right)\right| ds \\ & \leq K\Delta_n^2 + K\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\left(||\sigma_{2,s} - \sigma_{2,(i-1)\Delta_n}||\big|\mathcal{F}_{(i-1)\Delta_n}\right) ds. \end{aligned}$$

Moreover, by Fubini's Theorem and double expectation theorem, plus Lemma 3.5.2 (2), we have

$$\mathbb{E}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\left(||\sigma_{2,s} - \sigma_{2,(i-1)\Delta_n}||\big|\mathcal{F}_{(i-1)\Delta_n}\right) ds\right) \leq K\Delta_n^{3/2+\epsilon},$$

and hence the second claim follows. □

Lemma 3.5.4. *Under Assumption 3.5.1, we have*

$$\begin{aligned} & \mathbb{E}\Big|\mathbb{E}\left(\Delta_i^n X'^j \Delta_i^n X'^k \Delta_i^n X'^l \Delta_i^n X'^m|\mathcal{F}_{(i-1)\Delta_n}\right) - \\ & \left(c_{(i-1)\Delta_n}^{jk}c_{(i-1)\Delta_n}^{lm} + c_{(i-1)\Delta_n}^{jl}c_{(i-1)\Delta_n}^{km} + c_{(i-1)\Delta_n}^{jm}c_{(i-1)\Delta_n}^{kl}\right)\Delta_n^2\Big| \\ & \leq \quad K\Delta_n^{5/2} \end{aligned}$$

Proof. As in the proof of Lemma 4.1 in (Jacod and Rosenbaum, 2013a), by Itô's formula,

$$\begin{aligned}\prod_{l=1}^{4}\Delta_i^n X'^{j_l} &= \widetilde{M}_{i\Delta_n}+\sum_{l=1}^{4}\int_{(i-1)\Delta_n}^{i\Delta_n} b_s^{j_l}\prod_{1\le m\le 4, m\neq l}(X_s'^{j_m}-X_{(i-1)\Delta_n}'^{j_m})ds\\ &\quad+\frac{1}{2}\sum_{1\le l\neq l'\le 4}\int_{(i-1)\Delta_n}^{i\Delta_n}\left(c_s^{j_l j_{l'}}-c_{(i-1)\Delta_n}^{j_l j_{l'}}\right)\prod_{1\le m\le 4, m\neq l,l'}(X_s'^{j_m}-X_{(i-1)\Delta_n}'^{j_m})ds\\ &\quad+\frac{1}{2}\sum_{1\le l\neq l'\le 4}c_{(i-1)\Delta_n}^{j_l j_{l'}}\int_{(i-1)\Delta_n}^{i\Delta_n}\prod_{1\le m\le 4, m\neq l,l'}(X_s'^{j_m}-X_{(i-1)\Delta_n}'^{j_m})ds.\end{aligned}$$

where $\widetilde{M}_t$ is a martingale vanishing at $(i-1)\Delta_n$. Then by the first estimate in Lemma 3.5.2, Cauchy-Schwartz inequality and the fact that b is bounded, the absolute value of the $\mathcal{F}_{(i-1)\Delta_n}$−conditional expectation of the second term above is of stochastic order $\Delta_n^{5/2}$. In view of the last estimate in Lemma 3.5.2, Cauchy-Schwartz inequality yields that the $\mathcal{F}_{(i-1)\Delta_n}$−conditional expectation of the second term above is of stochastic order $\Delta_n^{5/2}$ as well. For the fourth term, note that

$$\begin{aligned}&c_{(i-1)\Delta_n}^{jk}\int_{(i-1)\Delta_n}^{i\Delta_n}(X_s'^{l}-X_{(i-1)\Delta_n}'^{l})(X_s'^{m}-X_{(i-1)\Delta_n}'^{m})ds\\ &=c_{(i-1)\Delta_n}^{jk}\int_{(i-1)\Delta_n}^{i\Delta_n}\left((X_s'^{l}-X_{(i-1)\Delta_n}'^{l})(X_s'^{m}-X_{(i-1)\Delta_n}'^{m})-c_{(i-1)\Delta_n}^{lm}\right)ds\\ &\quad+c_{(i-1)\Delta_n}^{jk}c_{(i-1)\Delta_n}^{lm}\Delta_n^2\end{aligned}$$

Then the second claim in Lemma 3.5.3, plus the fact that $\eta_{i,1}^n(b)$ is bounded, yields the result for the fourth term. □

Lemma 3.5.3 and 3.5.4 offer us a way to obtain estimates of the statistical error arising from nonparametrically estimating spot covariance, the result of which is summarized in the following lemma.

Lemma 3.5.5. *Under Assumption 3.5.1, we have*

(1) For any $q\ge 0$, $\mathbb{E}\left(||\alpha_{n,i}||^q|\mathcal{F}_{(i-1)\Delta_n}\right)\le k_q\Delta_n^q$

(2) $\mathbb{E}||\mathbb{E}(\alpha_{n,i}|\mathcal{F}_{(i-1)\Delta_n})||\le K\Delta_n^{3/2}\left(\mathbb{E}(\eta_{i,1}^n(b))+\sqrt{\Delta_n}+\Delta_n^{\epsilon}\right)$

(3) $\mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i}^{jk}\alpha_{n,i}^{lm}|\mathcal{F}_{(i-1)\Delta_n}\right)-\left(c_{(i-1)\Delta_n}^{jl}c_{(i-1)\Delta_n}^{km}+c_{(i-1)\Delta_n}^{jm}c_{(i-1)\Delta_n}^{kl}\right)\Delta_n^2\right|\le K\Delta_n^{5/2}$

Proof. The first claim is directly derived from Lemma 3.5.2, or is just (4.10) in (Jacod and Rosenbaum, 2013a). The second claim is exactly the second claim of Lemma 3.5.3, in view of the definition of $\alpha_{n,i}$.

To show (3), Recall that $\alpha_{n,i} = \Delta_i^n X' \Delta_i^n X'^{\intercal} - c_{(i-1)\Delta_n}\Delta_n$, we have

$$\begin{aligned}
& \alpha_{n,i}^{jk}\alpha_{n,i}^{lm} - \left(c^{jl}_{(i-1)\Delta_n}c^{km}_{(i-1)\Delta_n} + c^{jm}_{(i-1)\Delta_n}c^{kl}_{(i-1)\Delta_n}\right)\Delta_n^2 \\
= \quad & \left(\Delta_i^n X'^j \Delta_i^n X'^k - c^{jk}_{(i-1)\Delta_n}\Delta_n\right)\left(\Delta_i^n X'^l \Delta_i^n X'^m - c^{lm}_{(i-1)\Delta_n}\Delta_n\right) \\
& - \left(c^{jl}_{(i-1)\Delta_n}c^{km}_{(i-1)\Delta_n} + c^{jm}_{(i-1)\Delta_n}c^{kl}_{(i-1)\Delta_n}\right)\Delta_n^2 \\
= \quad & \Delta_i^n X'^j \Delta_i^n X'^k \Delta_i^n X'^l \Delta_i^n X'^m \\
& - \left(c^{jk}_{(i-1)\Delta_n}c^{lm}_{(i-1)\Delta_n} + c^{jl}_{(i-1)\Delta_n}c^{km}_{(i-1)\Delta_n} + c^{jm}_{(i-1)\Delta_n}c^{kl}_{(i-1)\Delta_n}\right)\Delta_n^2 \\
& + \Delta_n c^{lm}_{(i-1)\Delta_n}\left(\Delta_n c^{jk}_{(i-1)\Delta_n} - \Delta_i^n X'^j \Delta_i^n X'^k\right) \\
& + \Delta_n c^{jk}_{(i-1)\Delta_n}\left(\Delta_n c^{lm}_{(i-1)\Delta_n} - \Delta_i^n X'^l \Delta_i^n X'^m\right).
\end{aligned}$$

Then by Lemma 3.5.4 and the second claim in Lemma 3.5.3, plus the fact that c and $\eta_i^n(b)$ are both bounded, the result follows. □

With the help of Lemma 3.5.5, we are able to obtain the bounds on spot volatility estimation error

$$\tilde{c}_{n,i} = \hat{c}'_{i\Delta_n} - c_{i\Delta_n} = \frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}(\alpha_{n,i+j} + (c_{(i+j-1)\Delta_n} - c_{i\Delta_n})\Delta_n).$$

Define the "spot covariation" of the continuous martingale parts of X and c_1 and σ_1, respectively, as

$$\left(c_t^{(X,\sigma_1)}\right)^{i,jk} = \sum_{w=1}^{d}\sigma_t^{iw}\left(\sigma_t^{(\sigma_1)}\right)^{jk,w}, \quad \left(c_t^{(X,c_1)}\right)^{i,jk} = \sum_{w=1}^{d}\sigma_t^{iw}\left(\sigma_t^{(c_1)}\right)^{jk,w}.$$

Then both $c^{(X,\sigma_1)}$ and $c^{(X,c_1)}$ are càdlàg adapted.

Lemma 3.5.6. *Under Assumption 3.5.1, we have*

(1) For $q = 2, 3, 4$,

$$\mathbb{E}\left(\|\tilde{c}_{n,i}\|^q\right) \leq K\left(k_n^{-q/2} + k_n\Delta_n + \mathbb{E}\left(\left\|\sigma_{2,i\Delta_n} - \sigma_{2,(i-1)\Delta_n}\right\|^q\right) + \frac{1}{k_n}\sum_{j=1}^{k_n}\mathbb{E}\left(\left\|\sigma_{2,(i+j-1)\Delta_n} - \sigma_{2,i\Delta_n}\right\|^q\right)\right)$$

(2)

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left(\tilde{c}_{n,i}^{jk}\tilde{c}_{n,i}^{lm}\middle|\mathcal{F}_{i\Delta_n}\right) - \frac{1}{k_n}\left(c_{i\Delta_n}^{jl}c_{i\Delta_n}^{km} + c_{i\Delta_n}^{jm}c_{i\Delta_n}^{kl}\right)\right| \\
&\leq K\sqrt{\Delta_n}\left(k_n^{-1/2} + k_n\sqrt{\Delta_n} + k_n^{1/2+\epsilon}\Delta_n^{\epsilon} + \mathbb{E}(\eta_{i+1}^n(b))\right. \\
&\quad \left. + \mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right) + \mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right)\right)
\end{aligned}$$

Proof. For the first claim, we have by definition,

$$\mathbb{E}\left(\|\tilde{c}_{n,i}\|^q\right) \leq K\mathbb{E}\left(\left\|\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q\right) + K\mathbb{E}\left(\left\|\frac{1}{k_n}\sum_{j=1}^{k_n}(c_{(i+j-1)\Delta_n} - c_{i\Delta_n})\right\|^q\right)$$

For the second term on the RHS above, by Jensen's inequality since $q \geq 2$, and Lemma 3.5.2 (2), we have

$$\begin{aligned}
\mathbb{E}\left(\left\|\frac{1}{k_n}\sum_{j=1}^{k_n}(c_{(i+j-1)\Delta_n} - c_{i\Delta_n})\right\|^q\right) &\leq \frac{1}{k_n}\sum_{j=1}^{k_n}\mathbb{E}\left(\left\|c_{(i+j-1)\Delta_n} - c_{i\Delta_n}\right\|^q\right) \\
&\leq K\left(k_n\Delta_n + \frac{1}{k_n}\sum_{j=1}^{k_n}\mathbb{E}\left(\left\|\sigma_{2,(i+j-1)\Delta_n} - \sigma_{2,i\Delta_n}\right\|^q\right)\right)
\end{aligned}$$

We need a bit more work on the first term $\mathbb{E}\left(\left\|\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q\right)$. First note that

$$\mathbb{E}\left(\left\|\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q\right) \leq K(A_{1,n,i} + A_{2,n,i}),$$

where

$$\begin{aligned} A_{1,n,i} &= \mathbb{E}\left(\left\|\sum_{j=1}^{k_n}\left(\alpha_{n,i+j}-\mathbb{E}(\alpha_{n,i+j}|\mathcal{F}_{(i+j-1)\Delta_n})\right)\right\|^q\right) \\ A_{2,n,i} &= \mathbb{E}\left(\left\|\sum_{j=1}^{k_n}\mathbb{E}\left(\alpha_{n,i+j}|\mathcal{F}_{(i+j-1)\Delta_n}\right)\right\|^q\right). \end{aligned}$$

Since $A_{1,n,i}$ is the $q-$th moment of the sum of a martingale difference, by Burkholder-Davis-Gundy inequality, Jensen's inequality (note $q \geq 2$) and double expectation theorem, plus the first estimate in Lemma 3.5.5, we have

$$E(A_{1,n,i}) \leq Kk_n^{q/2}\Delta_n^q.$$

On the other hand, by Jensen's inequality, and the first estimate in Lemma 3.5.3,

$$\begin{aligned} E(A_{2,n,i}) &\leq Kk_n^{q-1}\sum_{j=1}^{k_n}\mathbb{E}\left(\left\|\mathbb{E}\left(\alpha_{n,i+j}|\mathcal{F}_{(i+j-1)\Delta_n}\right)\right\|^q\right) \\ &\leq Kk_n^q(\Delta_n^{3q/2}+\Delta_n^{2q}) \\ &\quad +Kk_n^{q-1}\sum_{j=1}^{k_n}\mathbb{E}\left(\int_{(i+j-1)\Delta_n}^{(i+j)\Delta_n}\mathbb{E}\left(||\sigma_{2,s}-\sigma_{2,(i+j-1)\Delta_n}|||\mathcal{F}_{(i-1)\Delta_n}\right)ds\right)^q \\ &\leq Kk_n^q(\Delta_n^{3q/2}+\Delta_n^{2q}) \\ &\quad +Kk_n^{q-1}\sum_{j=1}^{k_n}\Delta_n^{q-1}\int_{(i+j-1)\Delta_n}^{(i+j)\Delta_n}\mathbb{E}\left(||\sigma_{2,s}-\sigma_{2,(i+j-1)\Delta_n}||^q\right)ds \\ &\leq Kk_n^q(\Delta_n^{3q/2}+\Delta_n^{2q})+Kk_n^q\Delta_n^q\mathbb{E}\left(\left\|\sigma_{2,i\Delta_n}-\sigma_{2,(i-1)\Delta_n}\right\|^q\right) \end{aligned}$$

Therefore

$$\mathbb{E}\left(\left\|\frac{1}{k_n\Delta_n}\sum_{j=1}^{k_n}\alpha_{n,i+j}\right\|^q\right) \leq K\left(k_n^{-q/2}+\Delta_n^{q/2}+\mathbb{E}\left(\left\|\sigma_{2,i\Delta_n}-\sigma_{2,(i-1)\Delta_n}\right\|^q\right)\right)$$

and hence the first claim in Lemma 3.5.6 follows as $k_n^2\Delta_n \to 0$.

Now for the second claim, as in (Jacod and Rosenbaum, 2013a), we let

$$\zeta_{i,j}^n = \alpha_{n,i+j} + (c_{(i+j-1)\Delta_n} - c_{i\Delta_n})\Delta_n,$$

then $\tilde{c}_{n,i} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \zeta_{i,j}^n$, and we can write

$$\tilde{c}_{n,i}^{jk} \tilde{c}_{n,i}^{lm} = \frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \zeta_{i,u}^{n,jk} \zeta_{i,u}^{n,lm} + \frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n-1} \sum_{v=1}^{k_n} \zeta_{i,u}^{n,jk} \zeta_{i,v}^{n,lm} + \frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n-1} \sum_{v=1}^{k_n} \zeta_{i,u}^{n,lm} \zeta_{i,v}^{n,jk} . \tag{3.17}$$

For the first term $\frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \zeta_{i,u}^{n,jk} \zeta_{i,u}^{n,lm}$ in (3.17), we have

$$\frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \zeta_{i,u}^{n,jk} \zeta_{i,u}^{n,lm} - \frac{1}{k_n} \left(c_{(i-1)\Delta_n}^{jl} c_{(i-1)\Delta_n}^{km} + c_{(i-1)\Delta_n}^{jm} c_{(i-1)\Delta_n}^{kl} \right) = R_{1,n,i} + R_{2,n,i} + R_{3,n,i}$$

where

$$\begin{aligned}
R_{1,n,i} &\equiv \frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \left(\zeta_{i,u}^{n,jk} \zeta_{i,u}^{n,lm} - \alpha_{n,i+u}^{jk} \alpha_{n,i+u}^{lm} \right) \\
R_{2,n,i} &\equiv \frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \left(\alpha_{n,i+u}^{jk} \alpha_{n,i+u}^{lm} - \Delta_n^2 \left(c_{(i+u-1)\Delta_n}^{jl} c_{(i+u-1)\Delta_n}^{km} + c_{(i+u-1)\Delta_n}^{jm} c_{(i+u-1)\Delta_n}^{kl} \right) \right) \\
R_{3,n,i} &\equiv \frac{1}{k_n^2} \sum_{u=1}^{k_n} \left(c_{(i+u-1)\Delta_n}^{jl} c_{(i+u-1)\Delta_n}^{km} + c_{(i+u-1)\Delta_n}^{jm} c_{(i+u-1)\Delta_n}^{kl} - c_{i\Delta_n}^{n,jl} c_{i\Delta_n}^{km} - c_{i\Delta_n}^{jm} c_{i\Delta_n}^{kl} \right).
\end{aligned}$$

As derived in the Lemma 4.3 of (Jacod and Rosenbaum, 2013a), it holds that

$$\left| \zeta_{i,u}^{n,jk} \zeta_{i,u}^{n,lm} - \alpha_{n,i+u}^{jk} \alpha_{n,i+u}^{lm} \right| \leq 2\Delta_n \left\| c_{(i+u-1)\Delta_n} - c_{i\Delta_n} \right\| \left\| \alpha_{n,i+u} \right\| + \Delta_n^2 \left\| c_{(i+u-1)\Delta_n} - c_{i\Delta_n} \right\|^2$$

In light of the first estimate in Lemma 3.5.5 and the last estimate in Lemma 3.5.2, Cauchy-Schwartz inequality, plus triangle inequality, yield

$$\mathbb{E} \left(|R_{1,n,i}| \right) \leq K k_n^{-1/2} \sqrt{\Delta_n}.$$

Next, by successive conditioning(tower property), triangle inequality and the third claim in Lemma 3.5.5, we have

$$\begin{aligned}
\mathbb{E} \left(\left| \mathbb{E} \left(R_{2,n,i} | \mathcal{F}_{i\Delta_n} \right) \right| \right) &= \mathbb{E} \left(\left| \mathbb{E} \left(\mathbb{E} \left(R_{2,n,i} | \mathcal{F}_{(i+u-1)\Delta_n} \right) | \mathcal{F}_{i\Delta_n} \right) \right| \right) \\
&\leq \mathbb{E} \left(\left| \mathbb{E} \left(R_{2,n,i} | \mathcal{F}_{(i+u-1)\Delta_n} \right) \right| \right) \\
&\leq K k_n^{-1} \Delta_n^{-1}.
\end{aligned}$$

As for $R_{3,n,i}$, observe that

$$\begin{aligned} & c^{n,jl}_{(i+u-1)\Delta_n} c^{n,km}_{(i+u-1)\Delta_n} - c^{n,jl}_{i\Delta_n} c^{n,km}_{i\Delta_n} \\ = \quad & \left(c^{jl}_{(i+u-1)\Delta_n} - c^{jl}_{i\Delta_n} \right) c^{km}_{(i+u-1)\Delta_n} + c^{jl}_{i\Delta_n} \left(c^{km}_{(i+u-1)\Delta_n} - c^{km}_{i\Delta_n} \right). \end{aligned}$$

By the last estimate in Lemma 3.5.2 and the fact that c_t is bounded, we have

$$\mathbb{E}\left(\left| c^{jl}_{(i+u-1)\Delta_n} c^{km}_{(i+u-1)\Delta_n} - c^{jl}_{i\Delta_n} c^{km}_{i\Delta_n} \right| \right) \leq K \sqrt{k_n \Delta_n}.$$

The same results holds for $c^{jm}_{(i+u-1)\Delta_n} c^{kl}_{(i+u-1)\Delta_n} - c^{jm}_{i\Delta_n} c^{kl}_{i\Delta_n}$. Therefore

$$\mathbb{E}\left(|R_{3,n,i}| \right) \leq K k_n^{-1/2} \sqrt{\Delta_n}.$$

So we have shown that

$$\mathbb{E}\left| \mathbb{E}\left(\frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n} \zeta^{n,jk}_{i,u} \zeta^{n,lm}_{i,u} \middle| \mathcal{F}_{i\Delta_n} \right) - \frac{1}{k_n} \left(c^{jl}_{i\Delta_n} c^{km}_{i\Delta_n} + c^{jm}_{i\Delta_n} c^{kl}_{i\Delta_n} \right) \right| \leq K k_n^{-1/2} \sqrt{\Delta_n}.$$

At last we come to deal with the cross-product term $\frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n-1} \sum_{v=1}^{k_n} \zeta^{n,lm}_{i,u} \zeta^{n,jk}_{i,v}$ in (3.17), and the other term $\frac{1}{k_n^2 \Delta_n^2} \sum_{u=1}^{k_n-1} \sum_{v=1}^{k_n} \zeta^{n,jk}_{i,u} \zeta^{n,lm}_{i,v}$ could be dealt with similarly by interchanging the superscript. Recall that,

$$\zeta^{n,jk}_{i,u} = \alpha^{jk}_{n,i+u} + (c^{jk}_{(i+u-1)\Delta_n} - c^{jk}_{i\Delta_n})\Delta_n,$$

we have the following decomposition by tower property

$$\begin{aligned} \mathbb{E}\left(\zeta^{n,lm}_{i,u} \zeta^{n,jk}_{i,v} \middle| \mathcal{F}_{i\Delta_n} \right) &= \mathbb{E}\left(\mathbb{E}\left(\zeta^{n,lm}_{i,u} \zeta^{n,jk}_{i,v} \middle| \mathcal{F}_{(i+u)\Delta_n} \right) \middle| \mathcal{F}_{i\Delta_n} \right) \\ &= \mathbb{E}\left(\zeta^{n,lm}_{i,u} \mathbb{E}\left(\zeta^{n,jk}_{i,v} \middle| \mathcal{F}_{(i+u)\Delta_n} \right) \middle| \mathcal{F}_{i\Delta_n} \right) \\ &= \mathbb{E}\left(Z_{1,n,i} + Z_{2,n,i} + Z_{3,n,i} + Z_{4,n,i} \middle| \mathcal{F}_{i\Delta_n} \right) \qquad (3.18) \end{aligned}$$

where

$$
\begin{aligned}
Z_{1,n,i} &\equiv \zeta_{i,u}^{n,lm}\left[\mathbb{E}\left(\zeta_{i,v}^{n,jk}\middle|\mathcal{F}_{(i+u)\Delta_n}\right)-\left(c_{(i+u)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\Delta_n\right]\\
Z_{2,n,i} &\equiv \alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk}-c_{(i+u-1)\Delta_n}^{jk}\right)\Delta_n\\
Z_{3,n,i} &\equiv \alpha_{n,i+u}^{lm}\left(c_{(i+u-1)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\Delta_n\\
Z_{4,n,i} &\equiv \left(c_{(i+u-1)\Delta_n}^{lm}-c_{i\Delta_n}^{lm}\right)\left(c_{(i+u)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\Delta_n^2.
\end{aligned}
$$

We proceed from easy to hard. For $Z_{4,n,i}$, by Cauchy-Schwartz inequality and the last estimate in Lemma 3.5.2, we have

$$
\begin{aligned}
\mathbb{E}\left|\mathbb{E}\left(Z_{4,n,i}\middle|\mathcal{F}_{i\Delta_n}\right)\right| &\leq \Delta_n^2\mathbb{E}\left(\left|\left(c_{(i+u-1)\Delta_n}^{lm}-c_{i\Delta_n}^{lm}\right)\left(c_{(i+u)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\right|\right)\\
&\leq Kk_n\Delta_n^3
\end{aligned}
$$

For $Z_{3,n,i}$, by tower property and double expectation theorem, and the fact that c_t is bounded, we have

$$
\begin{aligned}
\mathbb{E}\left|\mathbb{E}\left(Z_{3,n,i}\middle|\mathcal{F}_{i\Delta_n}\right)\right| &= \mathbb{E}\left|\mathbb{E}\left(\mathbb{E}\left(Z_{3,n,i}\middle|\mathcal{F}_{(i+u-1)\Delta_n}\right)\middle|\mathcal{F}_{i\Delta_n}\right)\right|\\
&\leq \mathbb{E}\left|\mathbb{E}\left(Z_{3,n,i}\middle|\mathcal{F}_{(i+u-1)\Delta_n}\right)\right|\\
&= \Delta_n\mathbb{E}\left|\left(c_{(i+u-1)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\mathbb{E}\left(\alpha_{n,i+u}^{lm}\middle|\mathcal{F}_{(i+u-1)\Delta_n}\right)\right|\\
&\leq K\Delta_n\mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i+u}^{lm}\middle|\mathcal{F}_{(i+u-1)\Delta_n}\right)\right|\\
&\leq K\Delta_n^{5/2}\left(\mathbb{E}(\eta_{i+u,1}^n(b))+\sqrt{\Delta_n}+\Delta_n^{\epsilon}\right).
\end{aligned}
$$

On the other hand, by Lemma 3.5.7 and Lemma 3.5.8 respectively, we have

$$
\begin{aligned}
\mathbb{E}\left|\mathbb{E}\left(Z_{1,n,i}\middle|\mathcal{F}_{i\Delta_n}\right)\right| &\leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(b)\right)+k_n\sqrt{\Delta_n}+k_n^{1/2+\epsilon}\Delta_n^{\epsilon}\right).\\
\mathbb{E}\left|\mathbb{E}\left(Z_{2,n,i}\middle|\mathcal{F}_{i\Delta_n}\right)\right| &\leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right)+\mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right)+\Delta_n^{\epsilon}+\sqrt{\Delta_n}\right).
\end{aligned}
$$

Combine all the results derived above, and recall from Lemma 3.5.1 that $\mathbb{E}\left(\eta_{i+j,k}^n(Z)\right)\leq \mathbb{E}\left(\eta_i^n(Z)\right)$ for all bounded càdlàg adapted processes Z and all j,k such that $j+k\leq k_n$, then we have

$$
\begin{aligned}
& \mathbb{E}\left|\mathbb{E}\left(\frac{1}{k_n^2\Delta_n^2}\sum_{u=1}^{k_n-1}\sum_{v=1}^{k_n}\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,jk}\big|\mathcal{F}_{i\Delta_n}\right)\right| \\
\leq \quad & \frac{1}{k_n^2\Delta_n^2}\sum_{u=1}^{k_n-1}\sum_{v=1}^{k_n}\mathbb{E}\left|\mathbb{E}\left(\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,jk}\big|\mathcal{F}_{i\Delta_n}\right)\right| \\
\leq \quad & \frac{1}{k_n^2\Delta_n^2}\sum_{u=1}^{k_n-1}\sum_{v=1}^{k_n}\mathbb{E}\left|\mathbb{E}\left(Z_{1,n,i}+Z_{2,n,i}+Z_{3,n,i}+Z_{4,n,i}\big|\mathcal{F}_{i\Delta_n}\right)\right| \\
\leq \quad & K\sqrt{\Delta_n}\Big(k_n\sqrt{\Delta_n}+\mathbb{E}(\eta_{i+1}^n(b))+k_n^{1/2+\epsilon}\Delta_n^\epsilon+\mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right) \\
& \quad +\mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right)+\sqrt{\Delta_n}+\Delta_n^\epsilon\Big).
\end{aligned}
$$

Thus we prove the second claim in Lemma 3.5.6. □

Lemma 3.5.7. *In the context of Lemma 3.5.6, we have*

$$
\begin{aligned}
& \mathbb{E}\left|\mathbb{E}\left[\zeta_{i,u}^{n,lm}\left(\zeta_{i,v}^{n,jk}-\left(c_{(i+u)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\Delta_n\right)\Big|\mathcal{F}_{i\Delta_n}\right]\right| \\
& \leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(b)\right)+k_n\sqrt{\Delta_n}+k_n^{1/2+\epsilon}\Delta_n^\epsilon\right).
\end{aligned}
$$

Proof. By definition, observe that

$$
\zeta_{i,v}^{n,jk}=\Delta_{i+v}^n X'^j\Delta_{i+v}^n X'^k-c_{i\Delta_n}^{jk}\Delta_n,
$$

simple calculation yields

$$
\begin{aligned}
\zeta_{i,v}^{n,jk}-\left(c_{(i+u)\Delta_n}^{jk}-c_{i\Delta_n}^{jk}\right)\Delta_n \quad = \quad & \Delta_{i+v}^n X'^j\Delta_{i+v}^n X'^k- \\
& c_{(i+v-1)\Delta_n}^{jk}\Delta_n+\left(c_{(i+v-1)\Delta_n}^{jk}-c_{(i+u)\Delta_n}^{jk}\right)\Delta_n.
\end{aligned}
$$

Hence by triangle inequality and tower property, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left(\zeta_{i,v}^{n,jk} - \left(c_{(i+u)\Delta_n}^{jk} - c_{i\Delta_n}^{jk}\right)\Delta_n \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right| \\
&\leq \left|\mathbb{E}\left(\Delta_{i+v}^n X'^j \Delta_{i+v}^n X'^k - c_{(i+v-1)\Delta_n}^{jk}\Delta_n \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right| \\
&\quad + \left|\mathbb{E}\left(\left(c_{(i+v-1)\Delta_n}^{jk} - c_{(i+u)\Delta_n}^{jk}\right)\Delta_n \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right| \\
&\leq \left|\mathbb{E}\left(\mathbb{E}\left(\Delta_{i+v}^n X'^j \Delta_{i+v}^n X'^k - c_{(i+v-1)\Delta_n}^{jk}\Delta_n \middle| \mathcal{F}_{(i+v-1)\Delta_n}\right) \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right| \\
&\quad + \left|\mathbb{E}\left(c_{(i+v-1)\Delta_n}^{jk} - c_{(i+u)\Delta_n}^{jk} \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right|\Delta_n \\
&\leq \mathbb{E}\left(\left|\mathbb{E}\left(\Delta_{i+v}^n X'^j \Delta_{i+v}^n X'^k - c_{(i+v-1)\Delta_n}^{jk}\Delta_n \middle| \mathcal{F}_{(i+v-1)\Delta_n}\right)\right| \middle| \mathcal{F}_{(i+u)\Delta_n}\right) \\
&\quad + \left|\mathbb{E}\left(c_{(i+v-1)\Delta_n}^{jk} - c_{(i+u)\Delta_n}^{jk} \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right|\Delta_n \\
&\leq K\mathbb{E}\Bigg(\Delta_n^{3/2}\eta_{i+v,1}^n(b) + \Delta_n^2 \\
&\quad + \int_{(i+v-1)\Delta_n}^{(i+v)\Delta_n} \mathbb{E}\left(||\sigma_{2,s} - \sigma_{2,(i+v-1)\Delta_n}|| \middle| \mathcal{F}_{(i+v-1)\Delta_n}\right) ds \Bigg| \mathcal{F}_{(i+u)\Delta_n}\Bigg) \\
&\quad + \left(k_n\Delta_n + \mathbb{E}\left(||\sigma_{2,(i+v-1)\Delta_n} - \sigma_{2,(i+u)\Delta_n}|| \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right)\Delta_n
\end{aligned}
$$

where for the last inequality we first use the first claim in Lemma 3.5.3 but replace i by $i+v$, and Lemma 3.5.2.

Therefore, by tower property again, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left[\zeta_{i,u}^{n,lm}\left(\zeta_{i,v}^{n,jk} - \left(c_{(i+u)\Delta_n}^{jk} - c_{i\Delta_n}^{jk}\right)\Delta_n\right) \middle| \mathcal{F}_{i\Delta_n}\right]\right| \\
&= \left|\mathbb{E}\left[\zeta_{i,u}^{n,lm}\mathbb{E}\left(\zeta_{i,v}^{n,jk} - \left(c_{(i+u)\Delta_n}^{jk} - c_{i\Delta_n}^{jk}\right)\Delta_n \middle| \mathcal{F}_{(i+u)\Delta_n}\right) \middle| \mathcal{F}_{i\Delta_n}\right]\right| \\
&\leq \mathbb{E}\left(\left|\zeta_{i,u}^{n,lm}\right| \left|\mathbb{E}\left(\zeta_{i,v}^{n,jk} - \left(c_{(i+u)\Delta_n}^{jk} - c_{i\Delta_n}^{jk}\right)\Delta_n \middle| \mathcal{F}_{(i+u)\Delta_n}\right)\right| \middle| \mathcal{F}_{i\Delta_n}\right) \\
&= W_{1,n,i} + W_{2,n,i} + W_{3,n,i} + W_{4,n,i},
\end{aligned}
$$

where

$$
\begin{aligned}
W_{1,n,i} &\equiv \mathbb{E}\left(\left|\zeta_{i,u}^{n,lm}\right|\mathbb{E}\left(\eta_{i+v,1}^{n}(b)\middle|\mathcal{F}_{(i+u)\Delta_n}\right)\left|\mathcal{F}_{i\Delta_n}\right.\right)\Delta_n^{3/2}\\
W_{2,n,i} &\equiv \mathbb{E}\left(\left|\zeta_{i,u}^{n,lm}\right|\left|\mathcal{F}_{i\Delta_n}\right.\right)\left(\Delta_n^2+k_n\Delta_n^2\right)\\
W_{3,n,i} &\equiv \mathbb{E}\left(\left|\zeta_{i,u}^{n,lm}\right|\times\right.\\
&\quad\left.\mathbb{E}\left(\int_{(i+v-1)\Delta_n}^{(i+v)\Delta_n}\mathbb{E}\left(||\sigma_{2,s}-\sigma_{2,(i+v-1)\Delta_n}||\middle|\mathcal{F}_{(i+v-1)\Delta_n}\right)ds\middle|\mathcal{F}_{(i+u)\Delta_n}\right)\left|\mathcal{F}_{i\Delta_n}\right.\right)\\
W_{4,n,i} &\equiv \mathbb{E}\left(\left|\zeta_{i,u}^{n,lm}\right|\mathbb{E}\left(||\sigma_{2,(i+v-1)\Delta_n}-\sigma_{2,(i+u)\Delta_n}||\middle|\mathcal{F}_{(i+u)\Delta_n}\right)\left|\mathcal{F}_{i\Delta_n}\right.\right)\Delta_n.
\end{aligned}
$$

By the first estimate in Lemma 3.5.5, tower property and and the fact that c_t is bounded, we have

$$
\mathbb{E}\left(\left\|\zeta_{i,u}^{n}\right\|\left|\mathcal{F}_{i\Delta_n}\right.\right)\leq K\Delta_n,\quad \mathbb{E}\left(\left\|\zeta_{i,u}^{n}\right\|^2\left|\mathcal{F}_{i\Delta_n}\right.\right)\leq K\Delta_n^2. \tag{3.19}
$$

Then in view of (3.19), by definition of $\eta_{i+v,1}^{n}(b)$ and Cauchy-Schwartz inequality, we have

$$
\begin{aligned}
\mathbb{E}\left(W_{1,n,i}\right) &\leq \Delta_n^{3/2}\mathbb{E}\left(\eta_{i+1}^{n}(b)\sqrt{\mathbb{E}\left(\left\|\zeta_{i,u}^{n}\right\|^2\left|\mathcal{F}_{i\Delta_n}\right.\right)}\right)\\
&\leq K\Delta_n^{5/2}\mathbb{E}\left(\eta_{i+1}^{n}(b)\right).
\end{aligned}
$$

By (3.19) again, we have

$$
\mathbb{E}\left(W_{2,n,i}\right)\leq Kk_n\Delta_n^3.
$$

Moreover, by using Jensen's inequality multiple times and Cauchy-Schwartz inequality, plus (3.3), it follows

$$
\begin{aligned}
\mathbb{E}\left(W_{3,n,i}\right) &\leq \sqrt{\mathbb{E}\left(\left\|\zeta_{i,u}^{n}\right\|^2\right)}\sqrt{\Delta_n\int_{(i+v-1)\Delta_n}^{(i+v)\Delta_n}\mathbb{E}\left(||\sigma_{2,s}-\sigma_{2,(i+v-1)\Delta_n}||^2\right)ds}\\
&\leq K\Delta_n^{5/2+\epsilon}.
\end{aligned}
$$

At last, by Cauchy-Schwartz inequality, (3.19) and (3.3), it holds that

$$\begin{aligned}
\mathbb{E}\left(W_{4,n,i}\right) &\leq \Delta_n \sqrt{\mathbb{E}\left(\left\|\zeta_{i,u}^n\right\|^2\right)} \sqrt{\mathbb{E}\left(||\sigma_{2,(i+v-1)\Delta_n} - \sigma_{2,(i+u)\Delta_n}||^2\right)} \\
&\leq K\Delta_n^2 (k_n\Delta_n)^{1/2+\epsilon} \\
&= K\Delta_n^{5/2} k_n^{1/2+\epsilon} \Delta_n^{\epsilon}.
\end{aligned}$$

Combine all the results above, we obtain

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left[\zeta_{i,u}^{n,lm}\left(\zeta_{i,v}^{n,jk} - \left(c_{(i+u)\Delta_n}^{jk} - c_{i\Delta_n}^{jk}\right)\Delta_n\right)\Big|\mathcal{F}_{i\Delta_n}\right]\right| \\
&\leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(b)\right) + k_n\sqrt{\Delta_n} + \Delta_n^{\epsilon} + k_n^{1/2+\epsilon}\Delta_n^{\epsilon}\right) \\
&\leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(b)\right) + k_n\sqrt{\Delta_n} + k_n^{1/2+\epsilon}\Delta_n^{\epsilon}\right).
\end{aligned}$$

Hence we prove the lemma. □

Lemma 3.5.8. *In the context of Lemma 3.5.6, we have*

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\Delta_n\middle|\mathcal{F}_{i\Delta_n}\right)\right| \\
&\leq K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right) + \mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right) + \Delta_n^{\epsilon} + \sqrt{\Delta_n}\right)
\end{aligned}$$

Proof. By tower property and double expectation theorem, we have

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\left|\mathcal{F}_{i\Delta_n}\right.\right)\right|\Delta_n \\
&= \mathbb{E}\left|\mathbb{E}\left[\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\left|\mathcal{F}_{(i+u-1)\Delta_n}\right.\right)\left|\mathcal{F}_{i\Delta_n}\right.\right]\right|\Delta_n \\
&\leq \mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\left|\mathcal{F}_{(i+u-1)\Delta_n}\right.\right)\right|\Delta_n
\end{aligned}$$

In view of (3.15), (3.16) and Lemma 3.5.2 (2), we make the following decompositions

$$\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\left|\mathcal{F}_{(i+u-1)\Delta_n}\right.\right) = D_{1,n,i} + D_{2,n,i}$$

where $D_{1,n,i}$ involves the differences $c^{jk}_{1,(i+u)\Delta_n} - c^{jk}_{1,(i+u-1)\Delta_n}$ and $\sigma^{jk}_{1,(i+u)\Delta_n} - \sigma^{jk}_{1,(i+u-1)\Delta_n}$, and $D_{2,n,i}$ involves only difference $\sigma^{jk}_{2,(i+u)\Delta_n} - \sigma^{jk}_{2,(i+u-1)\Delta_n}$ (recall that by (3.16) the difference $c^{jk}_{2,(i+u)\Delta_n} - c^{jk}_{2,(i+u-1)\Delta_n}$ can be expresses via the difference $\sigma^{jk}_{2,(i+u)\Delta_n} - \sigma^{jk}_{2,(i+u-1)\Delta_n}$).

By Cauchy-Schwartz inequality, Lemma 3.5.5 (1), (3.3) and the fact that both σ_1 and σ_2 are bounded, we have

$$\begin{aligned}\mathbb{E}(D_{2,n,i}) &\leq K\sqrt{\mathbb{E}(\|\alpha_{n,i+u})\|^2}\sqrt{\mathbb{E}\left(||\sigma_{2,(i+u)\Delta_n} - \sigma_{2,(i+u-1)\Delta_n}||^2\right)} \\ &\leq K\Delta_n^{3/2+\epsilon}.\end{aligned}$$

As for $D_{1,n,i}$, we are going to show

$$\mathbb{E}|D_{1,n,i}| \leq K\Delta_n^{3/2}\left(\sqrt{\Delta_n} + \mathbb{E}\left(\eta^n_{i+1}(c^{(X,c_1)})\right) + \mathbb{E}\left(\eta^n_{i+1}(c^{(X,\sigma_1)})\right)\right),$$

for which it suffices to show

$$\begin{aligned}&\left|\mathbb{E}\left(\alpha^{lm}_{n,i+u}\left(c^{jk}_{1,(i+u)\Delta_n} - c^{jk}_{1,(i+u-1)\Delta_n}\right)|\mathcal{F}_{(i+u-1)\Delta_n}\right)\right| \\ &\leq K\Delta_n^{3/2}\left(\sqrt{\Delta_n} + \eta^n_{i+1}(c^{(X,c_1)})\right) \qquad (3.20) \\ &\left|\mathbb{E}\Bigg(\alpha^{lm}_{n,i+u}\sum_{w=1}^{d}\left[\left(\sigma^{jw}_{1,(i+u)\Delta_n} - \sigma^{jw}_{1,(i+u-1)\Delta_n}\right)\sigma^{wk}_{2,(i+u-1)\Delta_n}\right.\right. \\ &\qquad \left.\left.+\sigma^{jw}_{2,(i+u-1)\Delta_n}\left(\sigma^{wk}_{1,(i+u)\Delta_n} - \sigma^{wk}_{1,(i+u-1)\Delta_n}\right)\right]|\mathcal{F}_{(i+u-1)\Delta_n}\Bigg)\right| \\ &\leq K\Delta_n^{3/2}\left(\sqrt{\Delta_n} + \eta^n_{i+1}(c^{(X,\sigma_1)})\right) \qquad (3.21)\end{aligned}$$

(3.20) is derived from a word-for-word reproduction of the proof of Lemma 3.2 (c) in (Jacod and Rosenbaum, 2013b), plus the estimates in Lemma 3.5.2. More precisely, for any $i \geq 1$, we have

$$\begin{aligned}\alpha^{lm}_{n,i} &= B_{i\Delta_n} + M_{i\Delta_n}, \\ c^{jk}_{1,i\Delta_n} - c^{jk}_{1,(i-1)\Delta_n} &= B'_{i\Delta_n} + M'_{i\Delta_n},\end{aligned}$$

where B_t and M_t are respectively the drift part and continuous martingale part of

$$\left(X_t'^{l} - X_{(i-1)\Delta_n}'^{l}\right)\left(X_t'^{m} - X_{(i-1)\Delta_n}'^{m}\right) - c_{(i-1)\Delta_n}^{lm}, \quad t \in \{(i-1)\Delta_n, i\Delta_n\}.$$

Likewise, B'_t and M'_t are the drift and (possibly discontinuous) martingale part of $c_{1,t}^{jk} - c_{1,(i-1)\Delta_n}^{jk}$, respectively. In particular, M' may contain the compensated big jumps and the subtraction of the compensator, which is well-defined under Assumption 3.5.1 and is absolutely continuous with respect to Lebesgue measure, is then contained in B'. By (3.14) and Lemma 3.5.2, we have

$$\mathbb{E}\left(B_{i\Delta_n}^2 \middle| \mathcal{F}_{(i-1)\Delta_n}\right) \leq K\Delta_n^3, \quad \mathbb{E}\left(M_{i\Delta_n}^2 \middle| \mathcal{F}_{(i-1)\Delta_n}\right) \leq K\Delta_n^2.$$

Moreover, since $c_{1,t}^{jk} - c_{1,(i-1)\Delta_n}^{jk}$ is itself an one dimensional Itô semimartingale, by Lemma 3.5.2,

$$|B'_{i\Delta_n}| \leq K\Delta_n, \quad \mathbb{E}\left(M_{i\Delta_n}'^2 \middle| \mathcal{F}_{(i-1)\Delta_n}\right) \leq K\Delta_n.$$

Then it follows by (conditional) Cauchy-Schwartz inequality that the absolute values of the $\mathcal{F}_{(i-1)\Delta_n}-$ conditional expectations of $B_{i\Delta_n}B'_{i\Delta_n}$, $B_{i\Delta_n}M'_{i\Delta_n}$ and $M_{i\Delta_n}B'_{i\Delta_n}$ are smaller than $K\Delta_n^2$. As for $M_{i\Delta_n}M'_{i\Delta_n}$, notice that

$$\mathbb{E}\left(M_{i\Delta_n}M'_{i\Delta_n}\middle|\middle|\mathcal{F}_{(i-1)\Delta_n}\right) = \mathbb{E}\left(\langle M, M'\rangle_{i\Delta_n}\middle|\middle|\mathcal{F}_{(i-1)\Delta_n}\right)$$

and

$$\begin{aligned}
\langle M, M'\rangle_{i\Delta_n} &= \left\langle \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^{(c_1)} dW_s\right)^{jk}, \int_{(i-1)\Delta_n}^{i\Delta_n} (X_s'^{m} - X_{(i-1)\Delta_n}'^{m})[\sigma_s]_{l,\cdot} dW_s \right\rangle \\
&\quad + \left\langle \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s^{(c_1)} dW_s\right)^{jk}, \int_{(i-1)\Delta_n}^{i\Delta_n} (X_s'^{l} - X_{(i-1)\Delta_n}'^{l})[\sigma_s]_{m,\cdot} dW_s \right\rangle \\
&= \int_{(i-1)\Delta_n}^{i\Delta_n} \left(c_s^{(X,c_1)}\right)^{l,jk} (X_s'^{m} - X_{(i-1)\Delta_n}'^{m}) ds \\
&\quad + \int_{(i-1)\Delta_n}^{i\Delta_n} \left(c_s^{(X,c_1)}\right)^{m,jk} (X_s'^{l} - X_{(i-1)\Delta_n}'^{l}) ds,
\end{aligned}$$

where the first term can be further written as

$$\begin{aligned}
& \int_{(i-1)\Delta_n}^{i\Delta_n} \left(c_s^{(X,c_1)}\right)^{l,jk} (X'^m_s - X'^m_{(i-1)\Delta_n}) ds \\
= \quad & \left(c_{(i-1)\Delta_n}^{(X,c_1)}\right)^{l,jk} \int_{(i-1)\Delta_n}^{i\Delta_n} (X'^m_s - X'^m_{(i-1)\Delta_n}) ds \\
& + \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\left(c_s^{(X,c_1)}\right)^{l,jk} - \left(c_{(i-1)\Delta_n}^{(X,c_1)}\right)^{l,jk} \right) (X'^m_s - X'^m_{(i-1)\Delta_n}) ds.
\end{aligned}$$

Note X' is continuous Itô semimartingale, by Lemma 2 and (conditional) Cauchy-Schwartz inequality, the $\mathcal{F}_{(i-1)\Delta_n}-$conditional expectation of the above term is smaller than $K\Delta_n^{3/2}\left(\sqrt{\Delta_n} + \eta_{i+1}^n(c^{(X,\sigma_1)})\right)$. It is also true for the term with l and m interchanged.

On the other hand, (3.21) can be proved in a similar fashion, in view of the facts that $\sigma_{2,(i+u-1)\Delta_n}$ is $\mathcal{F}_{(i+u-1)\Delta_n}-$measurable and that σ_2 is bounded.

Therefore, we have shown that

$$\begin{aligned}
& \mathbb{E}\left|\mathbb{E}\left(\alpha_{n,i+u}^{lm}\left(c_{(i+u)\Delta_n}^{jk} - c_{(i+u-1)\Delta_n}^{jk}\right)\left|\mathcal{F}_{i\Delta_n}\right.\right)\right|\Delta_n \\
& \leq K\Delta_n^{3/2}\left(\mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right) + \mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right) + \Delta_n^{\epsilon} + \sqrt{\Delta_n}\right)\Delta_n \\
& = K\Delta_n^{5/2}\left(\mathbb{E}\left(\eta_{i+1}^n(c^{(X,c_1)})\right) + \mathbb{E}\left(\eta_{i+1}^n(c^{(X,\sigma_1)})\right) + \Delta_n^{\epsilon} + \sqrt{\Delta_n}\right),
\end{aligned}$$

as desired. □

Proof of Theorem 3.4.1. With all the preparations above we can move on to the proof of Theorem 3.4.1. As in (Jacod and Rosenbaum, 2013a), we consider the following decomposition

$$\frac{1}{\sqrt{\Delta_n}}(S_n(g) - S(g)) = \sum_{i=1}^{5} V_{i,n},$$

where

$$
\begin{aligned}
V_{1,n} &= \Delta_n^{1/2} \sum_i \Bigg(g(\hat{c}_{i\Delta_n}) - g(\hat{c}'_{i\Delta_n}) \\
&\qquad - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \Big(\partial^2_{jk,lm} g(\hat{c}_{i\Delta_n})(\hat{c}^{jl}_{i\Delta_n} \hat{c}^{km}_{i\Delta_n} + \hat{c}^{jm}_{i\Delta_n} \hat{c}^{kl}_{i\Delta_n}) \\
&\qquad - \partial^2_{jk,lm} g(\hat{c}'_{i\Delta_n})(\hat{c}'^{jl}_{i\Delta_n} \hat{c}'^{km}_{i\Delta_n} + \hat{c}'^{jm}_{i\Delta_n} \hat{c}'^{kl}_{i\Delta_n}) \Big) \Bigg) \\
V_{2,n} &= \Delta_n^{-1/2} \sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} (g(c_{i\Delta_n}) - g(c_s))\, ds - \Delta_n^{-1/2} \int_{(N_n+1)\Delta_n}^{T} g(c_s) ds \\
V_{3,n} &= \Delta_n^{1/2} \sum_i \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \frac{1}{k_n} \sum_{u=1}^{k_n} \Big(c^{lm}_{(i+u-1)\Delta_n} - c^{lm}_{i\Delta_n} \Big) \\
V_{4,n} &= \Delta_n^{1/2} \sum_i \Bigg(g(\hat{c}'_{i\Delta_n}) - g(c_{i\Delta_n}) - \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \tilde{c}^{lm}_{n,i} \\
&\qquad - \frac{1}{2k_n} \partial^2_{jk,lm} g(\hat{c}'_{i\Delta_n}) \Big(\hat{c}'^{jl}_{i\Delta_n} \hat{c}'^{km}_{i\Delta_n} + \hat{c}'^{jm}_{i\Delta_n} \hat{c}'^{kl}_{i\Delta_n} \Big) \Bigg) \\
V_{5,n} &= \Delta_n^{-1/2} k_n^{-1} \sum_i \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \sum_{u=1}^{k_n} \alpha^{lm}_{n,i+u}.
\end{aligned}
$$

In view of this, Theorem 3.4.1 is the consequence of the following claims

$$
\begin{aligned}
V_{i,n} &\xrightarrow{\mathbb{P}} 0, \quad i = 1, 2, 3, 4 \\
V_{5,n} &\xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, V(g)).
\end{aligned}
$$

Now we proceed to show these claims one by one.

Case $i = 1$: Here we extend the proof of (A.18) in (Li and Xiu, 2016) to multidimensional case. Define

$$
h_n(x) = g(x) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(x)(x^{jl} x^{km} + x^{jm} x^{kl}).
$$

Since $g(\cdot)$ is compactly supported and hence uniformly bounded, so is $h(\cdot)$. Then by mean value theorem, we have

$$
\begin{aligned}
\mathbb{E}|V_{1,n}| &\leq K\Delta_n^{1/2}\sum_i \mathbb{E}\left(\left\|\hat{c}_{i\Delta_n} - \hat{c}'_{i\Delta_n}\right\|\right) \\
&\leq K\Delta_n^{1/2}\sum_i a_n\Delta_n^{(2-r)\varpi} \\
&\leq K a_n\Delta_n^{(2-r)\varpi-1/2},
\end{aligned}
$$

for some sequence a_n tending to 0, where for the second inequality we use (4.8) in (Jacod and Rosenbaum, 2013a). By Assumption 3.2.2, $(2-r)\varpi \geq 1/2$, hence $V_{1,n} = o_p$.

Case $i = 2$: The remainder term $\Delta_n^{-1/2}\int_{(N_n+1)\Delta_n}^{T} g(c_s)ds = o_p(1)$ can be proved as in (Jacod and Rosenbaum, 2013a). On the other hand, by Taylor expansion up to the second order

$$
\Delta_n^{-1/2}\sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \left(g(c_s) - g(c_{i\Delta_n})\right) ds = V'_{2,n} + V''_{2,n},
$$

with

$$
\begin{aligned}
V'_{2,n} &= \Delta_n^{-1/2}\sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n})\left(c_s^{lm} - c_{i\Delta_n}^{lm}\right) ds \\
V''_{2,n} &= \Delta_n^{-1/2}\sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\xi_{n,i}(s))\left(c_s^{jk} - c_{i\Delta_n}^{jk}\right)\left(c_s^{lm} - c_{i\Delta_n}^{lm}\right) ds,
\end{aligned}
$$

where $\xi_{n,i}(s)$ lies between c_s and $c_{i\Delta_n}$. Since $\partial^2_{jk,lm} g(\cdot)$ is bounded, Lemma 3.5.2 gives

$$
\begin{aligned}
\mathbb{E}|V''_{2,n}| &\leq K\Delta_n^{-1/2}\sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \mathbb{E}\left(||c_s - c_{i\Delta_n}||^2\right) ds \\
&\leq K\Delta_n^{-1/2}\sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \mathbb{E}\left(||\sigma_{1,s} - \sigma_{1,i\Delta_n}||^2\right) ds \\
&\leq K\Delta_n^{1/2}.
\end{aligned}
$$

As for $V'_{2,n}$, in view of (3.15), (3.16) and Lemma 3.5.2 (2), it suffices to show

$$V'_{2,n,1} \equiv \Delta_n^{-1/2} \sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \left(\sigma_{1,s}^{lm} - \sigma_{1,i\Delta_n}^{lm}\right) ds \quad = \quad o_p(1), \qquad (3.22)$$

$$V'_{2,n,2} \equiv \Delta_n^{-1/2} \sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \left(\sigma_{2,s}^{lm} - \sigma_{2,i\Delta_n}^{lm}\right) ds \quad = \quad o_p(1). \qquad (3.23)$$

(3.22) can be proved using the martingale difference argument used in (Jacod and Protter, 2012), p.153-154. For (3.23), we have by Lemma 3.5.2 (2)

$$\begin{aligned} \mathbb{E}\left|V'_{2,n,2}\right| &\leq K\Delta_n^{-1/2} \sum_i \int_{i\Delta_n}^{(i+1)\Delta_n} \mathbb{E}\left(\left\|\sigma_{2,(i+u-1)\Delta_n} - \sigma_{2,i\Delta_n}\right\|\right) ds \\ &\leq K\Delta_n^{\epsilon}, \end{aligned}$$

which vanishes as $n \to \infty$.

Case $i = 3$: We want to show $V_{3,n} = o_p(1)$. In view of (3.15), (3.16) and Lemma 3.5.2 (2), it suffices to show that

$$V'_{3,n} \quad \equiv \quad \Delta_n^{1/2} \sum_i \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \frac{1}{k_n} \sum_{u=1}^{k_n} \left(\sigma_{1,(i+u-1)\Delta_n}^{lm} - \sigma_{1,i\Delta_n}^{lm}\right) = o_p(1) \qquad (3.24)$$

$$V''_{3,n} \quad \equiv \quad \Delta_n^{1/2} \sum_i \sum_{l,m=1}^{d} \partial_{lm} g(c_{i\Delta_n}) \frac{1}{k_n} \sum_{u=1}^{k_n} \left|\sigma_{2,(i+u-1)\Delta_n}^{lm} - \sigma_{2,i\Delta_n}^{lm}\right| = o_p(1). \qquad (3.25)$$

(3.24) is proved exactly as in (Li and Xiu, 2016), p.7-8 in their Appendix, observing that σ^{lm} is an one dimensional Itô semimartingale. On the other hand, by Lemma 3.5.2 (2), we have

$$\begin{aligned} E|V''_{3,n}| &\leq K\Delta_n^{1/2} \sum_i \frac{1}{k_n} \sum_{u=1}^{k_n} \mathbb{E}\left(\left\|\sigma_{2,(i+u-1)\Delta_n} - \sigma_{2,i\Delta_n}\right\|\right) \\ &\leq K\Delta_n^{-1/2}(k_n\Delta_n)^{1/2+\epsilon} \\ &\leq Kk_n^{1/2+\epsilon}\Delta_n^{\epsilon} \end{aligned}$$

which tends to 0 by Assumption 3.4.1.

Case $i=4$: By Lemma 3.5.6 and Assumption 3.4.1, we have

$$\begin{aligned}
\Delta^{1/2}\sum_i \mathbb{E}\left(\|\tilde{c}_{n,i}\|^3\right) &\rightarrow 0\\
k_n^{-1}\Delta^{1/2}\sum_i \mathbb{E}\left(\|\tilde{c}_{n,i}\|\right) &\rightarrow 0\\
\sum_{j=0}^{k_n-1}\sqrt{k_n^{-1}\left(k_n\Delta_n\right)^{a_4}} &\rightarrow 0
\end{aligned}$$

Then by a multidimensional version of the argument in (Li and Xiu, 2016), we have $V_{4,n} = o_p(1)$.

Case $i=5$: Now we deal with the leading term $V_{5,n}$, which can be rewritten as

$$V_{5,n} = \Delta_n^{-1/2}\sum_{i=0}^{[T/\Delta_n]-1}\sum_{l,m=1}^{d} w_{n,i+1}^{lm}\alpha_{n,i+1}^{lm},$$

where

$$w_{n,i+1}^{lm} = k_n^{-1}\sum_{j=(i-[T/\Delta_n]+k_n)^+}^{i\wedge(k_n-1)} \partial_{lm} g(c_{(i-j)\Delta_n}).$$

Note that in our notations, $w_{n,i+1}^{lm}$ and $\alpha_{n,i+1}$ are $\mathcal{F}_{i\Delta_n}$ and $\mathcal{F}_{(i+1)\Delta_n}$ measurable, respectively. Then by Theorem 2.2.15 in (Jacod and Protter, 2012), it suffices to show

$$\Delta_n^{-1/2}\sum_{i=0}^{[T/\Delta_n]-1} w_{n,i+1}^{lm}\mathbb{E}\left(\alpha_{n,i+1}^{lm}\middle|\mathcal{F}_{i\Delta_n}\right) \xrightarrow{\mathbb{P}} 0, \tag{3.26}$$

$$\begin{aligned}
&\Delta_n^{-1}\sum_{i=0}^{[T/\Delta_n]-1} w_{n,i+1}^{jk}w_{n,i+1}^{lm}\mathbb{E}\left(\alpha_{n,i+1}^{jk}\alpha_{n,i+1}^{lm}\middle|\mathcal{F}_{i\Delta_n}\right)\\
&\xrightarrow{\mathbb{P}} \int_0^T \partial_{jk} g(c_s)\partial_{lm} g(c_s)\left(c_s^{jl}c_s^{km}+c_s^{jm}c_s^{kl}\right)ds,
\end{aligned}\tag{3.27}$$

$$\Delta_n^{-2}\sum_{i=0}^{[T/\Delta_n]-1} \|w_{n,i+1}\|^4\,\mathbb{E}\left(\|\alpha_{n,i+1}\|^4\,\middle|\mathcal{F}_{i\Delta_n}\right) \xrightarrow{\mathbb{P}} 0, \tag{3.28}$$

$$\Delta_n^{-1/2}\sum_{i=0}^{[T/\Delta_n]-1} w_{n,i+1}^{lm}\mathbb{E}\left(\alpha_{n,i+1}^{lm}\Delta_{i+1}^n N\middle|\mathcal{F}_{i\Delta_n}\right) \xrightarrow{\mathbb{P}} 0, \tag{3.29}$$

where $N = W^k$ for some $k = 1, \ldots, d$, or is an arbitrary bounded martingale that is orthogonal to the driving Brownian motion W.

Recall $g(\cdot)$ is compactly supported, $\partial g(\cdot)$ is bounded and so is $w_{n,i}$. Hence by Lemma 3.5.5 (2) and (1), respectively, we have (3.26) and (3.28). Moreover, note c is càdlàg adapted with no fixed time of discontinuity since both σ_t and σ_2 are so, the proof of (4.16) in (Jacod and Rosenbaum, 2013a) gives (3.27).

At last as for 3.29, when $N = W^k$, (3.29) holds because of (3.20) with c_1 replaced by W^k. On the other hand, if N is a bounded martingale orthogonal to W, the "usual argument" gives the result, see (Jacod and Rosenbaum, 2013a) p.21 and the reference therein.

CHAPTER 4

Bootstrap Inference for Integrated Volatility Functionals

In this chapter we introduce alternative ways to do statistical inference for integrated volatility functionals. In the last comment that follows Theorem 3.4.1, it is shown how statistical inference could be done for integrated volatility functionals using the asymptotic result. In this chapter, we propose both parametric and nonparametric bootstrap algorithms to construct confidence intervals for integrated volatility functionals. We justify the two bootstrap methods by both asymptotic results and Monte Carlo simulation.

4.1 Setting

The basic set-up of this chapter is similar to the one in Chapter 3. To make it more convenient for readers to directly start this chapter without going over the previous one, we give a brief introduction to the setting here. In particular, we slightly update the notation to accommodate the bootstrap setting considered in this chapter.

Throughout this chapter, all processes are assumed to be càdlàg adapted. We consider a d-dimensional multivariate Itô semimartingale process X defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{4.1}$$

where b is the d-dimensional drift process, σ is the (co)volatility process taking values in the space $\mathcal{M}_d$, W is a d-dimensional Brownian motion and J denotes the jump process of X. This setting covers most models used in continuous-time economics and finance (see e.g., (Merton, 1992)), allowing for stochastic volatility, jumps and leverage effects.

On the statistical side, we assume that the process X is sampled at times $t_i = i/n$ for $i = 0, 1, \dots, n$, over a fixed time interval $[0, T]$, which may represent a typical trading day.

Without loss of generality, we assume $T = 1$. The increments of X are denoted as

$$\Delta_i^n X \equiv X_{i/n} - X_{(i-1)/n}, \quad i = 1, \ldots, n,$$

and asymptotically the sampling interval tends to 0 as $n \to \infty$.

4.1.1 Integrated volatility functionals

With model (4.1), the spot (co)variance process of X is given by $c = \sigma\sigma^\intercal$, which also takes values in $\mathcal{M}_d$. The goal of this chapter is to provide bootstrap confidence intervals for integrated volatility functionals of the form

$$S(g) \equiv \int_0^1 g(c_s)ds,$$

for some (possibly) nonlinear test function g.

The estimation problem of $S(g)$ with its asymptotic results has been studied in the recent literature, see (Jacod and Rosenbaum, 2013b), (Jacod and Rosenbaum, 2013a) and (Li et al., 2016a). Define

$$\widetilde{S}_n(g; \mathcal{D}_n) = \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} g(\hat{c}_{n,i}) \tag{4.2}$$

where for each $i \in \mathcal{I}_n \equiv \{0, \ldots, [n/k_n] - 1\}$

$$\hat{c}_{n,i} = \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X\right) \left(\Delta_{ik_n+j}^n X\right)^\intercal 1_{\{||\Delta_{ik_n+j}^n X|| \le u_n\}} \tag{4.3}$$

is the local approximation for the spot covariance c_t over the (non-overlapping) interval $[ik_n/n, (i+1)k_n/n]$, and $\mathcal{D}_n \equiv \{\Delta_i^n X, i = 1, \ldots, n\}$ represents the set of returns calculated from the sampled data with the letter $\mathcal{D}$ denoting "data". This is in contrast to the bootstrapped data set $\mathcal{D}^*$. As in Chapter 3, here the tuning parameter k_n is the number of increments employed in a local window and u_n determines the truncation threshold for eliminating jumps in X.

Under mild regularity conditions, $\widetilde{S}_n(g; \mathcal{D}_n)$ is a consistent estimator for $S(g)$; see (Jacod and Protter, 2012). However, for the associated unbiased central limit theorem with rate

$\sqrt{n}$ to hold, a prior de-biasing term has to be added, see (Jacod and Rosenbaum, 2013a). More specifically, define

$$S_n(g;\mathcal{D}_n) = \frac{k_n}{n}\sum_{i=0}^{[n/k_n]-1}\left(g(\hat{c}_{n,i}) - \frac{1}{2k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{n,i})\left(\hat{c}^{jl}_{n,i}\hat{c}^{km}_{n,i}+\hat{c}^{jm}_{n,i}\hat{c}^{kl}_{n,i}\right)\right), \quad (4.4)$$

then under the Assumptions 3.1.1-3.4.1, Theorem 3.4.1 shows that

$$\sqrt{n}\left(S_n(g;\mathcal{D}_n) - S(g)\right) \xrightarrow{\mathcal{L}-s} \mathcal{MN}(0, V(g)) \quad (4.5)$$

with asymptotic variance

$$V(g) = \sum_{j,k,l,m=1}^{d}\int_0^1 \partial_{jk}g(c_s)\partial_{lm}g(c_s)\left(c^{jl}_s c^{km}_s + c^{jm}_s c^{kl}_s\right)ds. \quad (4.6)$$

Some of early works that develop this asymptotic result, under more restricted conditions, though, include (Jacod and Rosenbaum, 2013a) and (Li et al., 2016a).

Based on the asymptotic result (4.5), confidence intervals for $S(g)$ can be formed provided that the asymptotic variance $V(g)$ could be consistently estimated; see (Jacod and Rosenbaum, 2013a). Alternatively, such statistical inference can be done via boostrap methods, which is the goal of the present chapter. In section 4.2, we describe the algorithm to construct parametric bootstrap confidence intervals for $S(g)$, together with theoretical results which justify the algorithm. The nonparametric way, which we call "local IID boostrap", to construct bootstrap confidence intervals using resampling with replacement will be introduced in section 4.3. We emphasize that in this setting we assume the same assumptions that are imposed in Chapter 3, which are Assumption 3.1.1, 3.2.1, 3.2.2 and 3.4.1.

4.2 Parametric Bootstrap

4.2.1 Algorithm

We pick a sequence k_n of width of local window for spot covariance estimation and u_n of truncation threshold for eliminating jumps of X according to Assumption 3.2.2 as given in Chapter 3. Then for a given confidence level $\alpha \in (0, \frac{1}{2})$, the parametric bootstrap confidence interval for $S(g)$ with (asymptotic) coverage probability $1-\alpha$ can be constructed using the following algorithm.

Algorithm 1. (Parametric Bootstrap Confidence Intervals)

Step 1. For each $i \in \mathcal{I}_n$, estimate $\hat{c}_{n,i}$ according to (4.3), and compute $\widetilde{S}_n(g; \mathcal{D}_n)$ and $S_n(g; \mathcal{D}_n)$ according to (4.2) and (4.4) respectively.

Step 2. For each $i \in \mathcal{I}_n$, simulate $\Delta_{ik_n+j}^n X^* \sim \mathcal{N}\left(0, \frac{1}{n}\hat{c}_{n,i}\right)$ for $j = 1, \ldots, k_n$.

Step 3. Compute bootstrap spot covariance estimators using $\Delta_i^n X^*$, namely,

$$\hat{c}_{n,i}^* = \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X^*\right) \left(\Delta_{ik_n+j}^n X^*\right)^\intercal.$$

Step 4. Compute bootstrap estimator for $S(g)$ as

$$S_n(g; \mathcal{D}_n^*) \equiv \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} \left(g(\hat{c}_{n,i}^*) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(\hat{c}_{n,i}^*) \left(\hat{c}_{n,i}^{*jl} \hat{c}_{n,i}^{*km} + \hat{c}_{n,i}^{*jm} \hat{c}_{n,i}^{*kl} \right) \right), \quad (4.7)$$

where $\mathcal{D}_n^* \equiv \{\Delta_i^n X^*, i = 1, \ldots, n\}$ represents the set of returns calculated from parametric bootstrap samples generated from Step 2.

Step 5. Repeat Step 2 - 4 for a large number of times. Set $q_{\alpha/2}(S_n(g; \mathcal{D}_n^*))$ and $q_{1-\alpha/2}(S_n(g; \mathcal{D}_n^*))$ as the $\alpha/2$ and $1-\alpha/2$ quantiles of $S_n(g; \mathcal{D}_n^*)$ respectively. The parametric bootstrap confidence interval of coverage $1-\alpha$ is then formed as

$$[S_n(g; \mathcal{D}_n) + \widetilde{S}_n(g; \mathcal{D}_n) - q_{1-\alpha/2}(S_n(g; \mathcal{D}_n^*)), S_n(g; \mathcal{D}_n) + \widetilde{S}_n(g; \mathcal{D}_n) - q_{\alpha/2}(S_n(g; \mathcal{D}_n^*))]. \quad (4.8)$$

As one can readily see, the confidence interval (4.8) is constructed via parametric bootstrap as bootstrap samples $\Delta_{ik_n+j}^n X^*$ are generated from normal distribution.

4.2.2 Result

Theorem 4.2.1 below justifies the confidence interval (4.8) has asymptotic $1-\alpha$ coverage probability. Intuitively, the bootstrap confidence interval is (asymptotically) valid if

$S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)$ enjoys the same asymptotic result, conditional on the realized original sample, as that of $S_n(g;\mathcal{D}_n) - S(g)$ in (4.5), see e.g. (van der Vaart, 1998).

In the sequel, we use $Z_n \xrightarrow{\mathcal{L}|\mathcal{F}} Z$ to denote $\mathcal{L}(Z_n|\mathcal{F}) \xrightarrow{\mathbb{P}} \mathcal{L}(Z|\mathcal{F})$ for a sequence of random variables $(Z_n)_{n\geq 1}$ and Z, namely, the conditional distribution of Z_n given $\mathcal{F}$ converges to that of Z in probability under Prokhorov metric. Such a mode of convergence in commonly used in the setting of bootstrap, as well as together with stable convergence in law. For the latter situation, see e.g., (Barndorff-Nielsen et al., 2008) Proposition 5 and (Li and Xiu, 2016) Lemma A3.

Theorem 4.2.1. *Suppose the Assumption 3.1.1, 3.2.1, 3.2.2 and 3.4.1 given in Chapter 3 hold, and let $S_n(g;\mathcal{D}_n^*)$ be given by Algorithm 1. It follows that*

$$\sqrt{n}\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, V(g)), \tag{4.9}$$

where

$$V(g) = \sum_{j,k,l,m=1}^{d} \int_0^1 \partial_{jk} g(c_s) \partial_{lm} g(c_s) \left(c_s^{jl} c_s^{km} + c_s^{jm} c_s^{kl}\right) ds. \tag{4.10}$$

Several comments are worth mentioning. Firstly, under the same assumptions as imposed in Theorem 3.4.1, both the convergence rate and the asymptotic variance in Theorem 4.2.1 are exactly the same as that in (4.5) and (4.6), which validates the constructed confidence interval given in (4.8), as argued in (van der Vaart, 1998). Secondly, a closer observation reveals that the left side of (4.9) has only one bias correction that is included in $S_n(g;\mathcal{D}_n^*)$, while there is no biased correction term in $\widetilde{S}_n(g;\mathcal{D}_n)$. This is in line with the original asymptotic result (4.5) where the only bias correction occurs within $S_n(g;\mathcal{D}_n)$.

4.3 The Local IID Bootstrap

In this section we introduce the algorithm to construct local IID bootstrap confidence interval for $S(g)$. The wording "local IID" emerges from the fact that the bootstrap samples in this method are generated by resampling with replacement over each nonoverlapping local window. The theoretical results to justify this procedure is also provided.

4.3.1 Algorithm

Pick a sequence k_n of width of local window for spot covariance estimation and u_n of truncation threshold for eliminating jumps of X according to Assumption 3.2.2 as given in Chapter 3. For a given confidence $\alpha \in (0,1)$, the local IID bootstrap confidence interval for $S(g)$ with (asymptotic) coverage probability $1-\alpha$ can be constructed using the following algorithm.

Algorithm 2. (Local IID Bootstrap Confidence Intervals)

Step 1. For each $i \in \mathcal{I}_n$, estimate $\hat{c}_{n,i}$ according to (4.3), and compute $\widetilde{S}_n(g;\mathcal{D}_n)$ and $S_n(g;\mathcal{D}_n)$ according to (4.2) and (4.4) respectively.

Step 2. For each $i \in \mathcal{I}_n$, compute bootstrap spot covariance estimators

$$\hat{c}^*_{n,i} = \frac{n}{k_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{ik_n+j^*_{i,\ell}} X\right) \left(\Delta^n_{ik_n+j^*_{i,\ell}} X\right)^{\mathsf{T}} 1_{\{\|\Delta^n_{ik_n+j^*_{i,\ell}} X\| \le u_n\}}, \tag{4.11}$$

where for each i and ℓ,

$$j^*_{i,\ell} \sim \text{i.i.d. Uniform}\{1,\ldots,k_n\}.$$

Step 3. Compute bootstrap estimator for $S(g)$ as

$$S_n(g;\mathcal{D}^*_n) \equiv \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} \left(g(\hat{c}^*_{n,i}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}^*_{n,i}) \left(\hat{c}^{*jl}_{n,i} \hat{c}^{*km}_{n,i} + \hat{c}^{*jm}_{n,i} \hat{c}^{*kl}_{n,i} \right) \right). \tag{4.12}$$

where $\mathcal{D}_n \equiv \{\Delta^n_i X^*, i = 1,\ldots,n\}$ represents the set of returns calculated from local IID bootstrap samples generated from Step 2.

Step 4. Repeat Step 2 - 3 for a large number of times. Set $q_{\alpha/2}(S_n(g;\mathcal{D}^*_n))$ and $q_{1-\alpha/2}(S_n(g;\mathcal{D}^*_n))$ as the $\alpha/2$ and $1-\alpha/2$ quantiles of $S_n(g;\mathcal{D}^*_n)$ respectively. The local IID bootstrap confidence interval of coverage $1-\alpha$ is then formed as

$$[S_n(g;\mathcal{D}_n)+\widetilde{S}_n(g;\mathcal{D}_n)-q_{1-\alpha/2}(S_n(g;\mathcal{D}^*_n)), S_n(g;\mathcal{D}_n)+\widetilde{S}_n(g;\mathcal{D}_n)-q_{\alpha/2}(S_n(g;\mathcal{D}^*_n))]. \tag{4.13}$$

Algorithm 2 forms the bootstrap confidence interval in a nonparametric way, featuring that the bootstrap samples $\Delta^n_{ik_n+j} X^*$ are generated from i.i.d. resampling with replacement over each (nonoverlapping) interval $[ik_n/n,(i+1)k_n/n]$, instead of using normal distribution as in the parametric case.

4.3.2 Result

Theorem 4.3.1 below justifies the confidence interval (4.13) has asymptotic $1-\alpha$ coverage probability.

Theorem 4.3.1. *Suppose the Assumption 3.1.1, 3.2.1, 3.2.2 and 3.4.1 given in Chapter 3 hold, and let $S_n(g;\mathcal{D}_n^*)$ be given by Algorithm 2. It follows that*

$$\sqrt{n}\left(S_n(g;\mathcal{D}_n^*)-\widetilde{S}_n(g;\mathcal{D}_n)\right)\xrightarrow{\mathcal{L}|\mathcal{F}}\mathcal{MN}(0,V(g)), \tag{4.14}$$

where

$$V(g)=\sum_{j,k,l,m=1}^{d}\int_0^1\partial_{jk}g(c_s)\partial_{lm}g(c_s)\left(c_s^{jl}c_s^{km}+c_s^{jm}c_s^{kl}\right)ds. \tag{4.15}$$

Theorem 4.3.1 implies that the local IID confidence interval has the same asymptotic efficiency as the parametric method, in spite of the lack of local Gaussianity. On the technical level, this is because the i.i.d. resampling of bootstrap samples $\Delta_{ik_n+j}^n X^*$ with replacement is only implemented locally over each (nonoverlapping) interval, rather than over the whole time span $[0,1]$ as, for example, in (Gonçalves and Meddahi, 2009). Therefore, the constructed confidence interval (4.13) has the desired coverage as argued in (van der Vaart, 1998).

4.4 Monte Carlo Study

4.4.1 The Monte Carlo Set-up

In this section we validate the confidence intervals constructed in (4.8) and (4.13) via Monte Carlo study, where we consider the integrated idiosyncratic variance as mentioned before. We set the time span $T=60$ days which is almost one business quarter, with the time unit being one year or equivalently, 250 trading days. Each day contains 390 1-minute sampled returns, corresponding to 6.5 transaction hours. We also consider 10-minute returns in the study. All continuous-time processes are simulated using Euler scheme with a 5-second mesh. There are 1000 Monte Carlo trials and 500 bootstrap trials.

We consider a bivariate setting, in which the data generating process (DGP) is given as follows:

$$\begin{aligned} dZ_t &= \sqrt{c_{ZZ,t}}dW_t + dJ_{Z,t}, \\ dY_t &= \beta_t\sqrt{c_{ZZ,t}}dW_t + \sqrt{c_{\epsilon,t}}d\widetilde{W}_t + dJ_{Y,t} \end{aligned}$$

where W and $\widetilde{W}$ are two independent Brownian motions, and J_Z and J_Y are two independent compound Poisson processes with intensity equal to 2 jumps per year and jump distribution $\mathcal{N}(0, 0.02^2)$. One can consider Z as the log-price process of the market portfolio and Y as that of some individual asset. The process β, measuring the exposure of Y to Z, follows:

$$\beta_t = 0.5 + 0.1\sin(100t).$$

The market variance processes c_{ZZ} and idiosyncratic variance process c_ϵ satisfy the following factor structure:

$$\sigma_{ZZ,t} = b_t + f_{1,t}, \quad c_{\epsilon,t} = 0.1 + f_{2,t} \tag{4.16}$$

where b_t is a fractional Brownian Motion with Hurst parameter 0.9[1], and the factors f_j, for $j = 1, 2$, are simulated according to

$$d\log(f_{j,t}) = 5\left(\log(0.3^2) - \log(f_{j,t})\right)dt + 5\left(\rho_f dW_t + \sqrt{1-\rho_f^2}dB_{j,t}\right) + dJ^f_{j,t},$$

where the negative correlation $\rho_f = -0.5$ represents the "leverage effect", and $(J^f_j)_{j=1,2}$ are compound Poisson processes, which are mutually independent and independent of other components in the DGP, with intensity equal to 4 jumps per year and jump size distribution being exponential with mean 0.1.

[1]We thank Professor Vladas Pipiras for generously sharing the Matlab code for simulating fractional Brownian motion.

In this setting, the volatility functional under consideration is the so-called idiosyncratic spot covariance of Y, given by

$$g(c_t) = c_{YY,t} - \beta_t^2 c_{ZZ,t} = c_{YY,t} - c_{ZY,t}^2 / c_{ZZ,t},$$

where $c_{ZY,t} = \beta_t^2 c_{ZZ,t} + c_{\epsilon,t}$. We would like to examine the coverage probabilities of the confidence intervals for $S(g; \mathcal{D}_n) = \int_0^T g(s)ds$ as constructed by (4.8) and (4.13).

Tuning parameters in the implementation are set as follows. The truncation threshold for day t is given by $3.5\bar{\sigma}_t \Delta^{0.49}$, where $\bar{\sigma}_t$ is the annualized bipower variation (see (Barndorff-Nielsen and Shephard, 2004c)). We consider two sets of local window $k_n \in \{45, 60, 75\}$ when $\Delta = 1$ minute and $k_n \in \{12, 14, 16\}$ when $\Delta = 10$ minutes, so as to check the robustness of our inference procedure to the choice of tuning parameters.

4.4.2 Results

Monte Carlo coverage probabilities of confidence intervals formed by (4.8) and (4.13) are reported in Table 4.1. The coverage probabilities for both parametric bootstrap confidence intervals and local IID bootstrap confidence intervals are very close to the corresponding nominal confidence levels, and the results are robust to the choices of local window k_n. Astute readers may find that the performance for parametric bootstrap method is slightly better than the local IID method, which is not surprising in this artificial simulation setting as the underlying process dynamics is designed to be normally distributed given the realization of volatility. Other finer observations include, for example, given $\Delta = 10$ minutes, the performance of local IID method of $k_n = 14$ and $k_n = 16$ is better than that of $k_n = 12$. Overall, the simulation results reported here are consistent with our asymptotic theory, and moreover, validate our inference procedure in the finite sample setting.

4.5 Future Work

So far, we have only considered non-overlapping case. It would be interesting to derive the same asymptotic result in the overlapping case from a smoothness point of view. Moreover, althought the bootstrap method could produce confidence intervals for estimat-

Table 4.1: Monte Carlo coverage probabilities for non-overlapping bootstrap confidence intervals.

	$\Delta = 1$ minute			$\Delta = 10$ minutes		
	$k_n = 45$	$k_n = 60$	$k_n = 75$	$k_n = 12$	$k_n = 14$	$k_n = 16$
Panel A. Parametric Bootstrap Method						
$\alpha = 1\%$	0.984	0.986	0.988	0.993	0.992	0.993
$\alpha = 5\%$	0.947	0.952	0.947	0.946	0.959	0.956
$\alpha = 10\%$	0.901	0.888	0.899	0.901	0.911	0.912
Panel B. Local IID Bootstrap Method						
$\alpha = 1\%$	0.978	0.976	0.975	0.973	0.983	0.980
$\alpha = 5\%$	0.926	0.932	0.935	0.916	0.943	0.947
$\alpha = 10\%$	0.873	0.884	0.876	0.860	0.894	0.891

Note: Panel A reports the coverage probabilities for the parametric bootstrap method. Panel B reports the coverage probabilities for the local IID Bootstrap method. The left (resp. right) panel reports results for 1-minute (resp. 10-minute) sampling. Rows and columns respectively correspond to various choices of confidence level α and local window width k_n.

ing integrated volatility functionals without computing the asymptotic variance, it is still required to compute the second order derivatives of the test function g, which could be quite cumbersome in practice. In view of the most recent work in (Li and Xiu, 2017), it would be desired to derive a bootstrap method that has no bias correction term.

4.6 Proofs

In this section, we provide regularity conditions and formal proofs for Theorem 4.2.1 and 4.3.1 in the main text. To do so, we need to complement the notations to be used in the sequel. We use K to denote a positive generic constant, which might vary from line to line. $\mathbb{E}_{\mathcal{F}}(\cdot)$ and $\mathrm{Var}_{\mathcal{F}}(\cdot)$ denote $\mathcal{F}$-conditional expectation and variance, respectively. For two (possibly random) real-valued sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, we write $a_n = O_p(b_n)$ if a_n/b_n is bounded in probability and write $a_n = o_p(b_n)$ if a_n/b_n is converges to 0 in probability.

We point out here that the conditions under which Theorem 4.2.1 and 4.3.1 hold are exactly the same as in Chapter 3. More specifically, we assume the dynamics of X and σ follows Assumption 3.1.1, where, in particular, σ is a LMIS. Furthermore, we assume the test function g satisfies Assumption 3.2.1, and tuning parameters k_n and u_n satisfy

Assumption 3.2.2 noting that $\Delta_n = 1/n$ in current setting. Also, we have Assumption 3.4.1 to compensate for the fact that the long memory part of σ might not be a martingale.

In the following we are going to present formal proofs for Theorem 4.2.1 and 4.3.1. By a standard localization procedure (Lemma 4.4.9 in (Jacod and Protter, 2012)), we can without loss of generality assume all locally bounded process are actually bounded. In light of Lemma 2 in (Li et al., 2016a), we can restrict our attention to the set with probability approaching to 1 on which local covariance estimates $\hat{c}_{n,i}$ are uniformly bounded for $i \in \mathcal{I}_n$. Moreover, by using the spatial localization argument as in the proof of Theorem 2 in (Li et al., 2016a), we can assume that test function g is compactly supported, and hence both function g and its existent derivatives are bounded from above by some positive constant.

4.6.1 Proof of Theorem 4.2.1

Observe that the left hand side of (4.9) can be written as

$$\sqrt{n}\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right) = \sum_{i=0}^{[n/k_n]-1} \frac{k_n}{\sqrt{n}} H_{n,i}$$

where

$$H_{n,i} = g(\hat{c}_{n,i}^*) - g(\hat{c}_{n,i}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}_{n,i}^*)\left(\hat{c}_{n,i}^{*jl}\hat{c}_{n,i}^{*km} + \hat{c}_{n,i}^{*jm}\hat{c}_{n,i}^{*kl}\right).$$

In light of Theorem 2.2.14 in (Jacod and Protter, 2012), which is a set of Lyapunov-type conditions and by subsequence principle, it suffices to show

$$\mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right)\right) \xrightarrow{\mathbb{P}} 0 \tag{4.17}$$

$$\mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right)\right)^2 \xrightarrow{\mathbb{P}} V(g) \tag{4.18}$$

and

$$\sum_{i=0}^{[n/k_n]-1} \mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\frac{k_n}{n}H_{n,i}\right)^4 \xrightarrow{\mathbb{P}} 0, \tag{4.19}$$

which will be proved via the following three steps.

Step 1. We first show (4.17). Simple calculation using normality of $\Delta^n_{ik_n+j}X^*$ gives

$$\mathbb{E}_{\mathcal{F}}(\hat{c}^{*jk}_{n,i}) = \hat{c}^{jk}_{n,i}, \quad \mathbb{E}_{\mathcal{F}}\left(\hat{c}^{*jk}_{n,i} - \hat{c}^{jk}_{n,i}\right)^4 = O_p(k_n^{-2})$$

$$\mathbb{E}_{\mathcal{F}}\left[\left(\hat{c}^{*jk}_{n,i} - \hat{c}^{jk}_{n,i}\right)\left(\hat{c}^{*lm}_{n,i} - \hat{c}^{lm}_{n,i}\right)\right] = \frac{1}{k_n}\left(\hat{c}^{jl}_{n,i}\hat{c}^{km}_{n,i} + \hat{c}^{jm}_{n,i}\hat{c}^{kl}_{n,i}\right).$$

Then by Taylor expansion and the fact that $\hat{c}_{n,i}$ is $\mathcal{F}$-measurable, we have for any $i \in \mathcal{I}_n$,

$$\mathbb{E}_{\mathcal{F}}\left(g(\hat{c}^*_{n,i}) - g(\hat{c}_{n,i})\right) = \frac{1}{2k_n}\sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}_{n,i})\left(\hat{c}^{jl}_{n,i}\hat{c}^{km}_{n,i} + \hat{c}^{jm}_{n,i}\hat{c}^{kl}_{n,i}\right) + G_{n,i},$$

where by Cauchy-Schwartz inequality $G_{n,i}$ is the higher order term satisfying

$$\begin{aligned}\mathbb{E}_{\mathcal{F}}|G_{n,i}| &\leq K\sum_{j,k,l,m,u,v=1}^{d} \mathbb{E}_{\mathcal{F}}\left[\left(\hat{c}^{*jk}_{n,i} - \hat{c}^{jk}_{n,i}\right)\left(\hat{c}^{*lm}_{n,i} - \hat{c}^{lm}_{n,i}\right)\left(\hat{c}^{*uv}_{n,i} - \hat{c}^{uv}_{n,i}\right)\right] \\ &= O_p(k_n^{-2/3}).\end{aligned}$$

By Assumption 3.2.2, $k_n^3/n \to \infty$, it follows that

$$\sqrt{n}\mathbb{E}_{\mathcal{F}}\left(S_n(g;\mathcal{D}^*_n) - \widetilde{S}_n(g;\mathcal{D}_n)\right) = \frac{k_n}{\sqrt{n}}\sum_{i=0}^{[n/k_n]-1} \eta_{n,i} + o_p(1)$$

where

$$\begin{aligned}\eta_{n,i} = &\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}_{n,i})\left(\hat{c}^{jl}_{n,i}\hat{c}^{km}_{n,i} + \hat{c}^{jm}_{n,i}\hat{c}^{kl}_{n,i}\right) \\ &-\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d} \mathbb{E}_{\mathcal{F}}\left[\partial^2_{jk,lm} g(\hat{c}^*_{n,i})\left(\hat{c}^{*jl}_{n,i}\hat{c}^{*km}_{n,i} + \hat{c}^{*jm}_{n,i}\hat{c}^{*kl}_{n,i}\right)\right].\end{aligned}$$

By mean value theorem and Cauchy-Schwartz inequality, one can readily derives that

$$\mathbb{E}|\eta_{n,i}| \leq K k_n^{-3/2},$$

which, plus the fact $k_n^3/n \to \infty$, gives (4.17).

Step 2. Next we show (4.18). In view of step 1, we have uniformly in i,

$$\sqrt{n} H_{n,i} = o_p(1).$$

Since $H_{n,i}$ and $H_{n,j}$ are conditionally independent if $i \neq j$, we have

$$n\mathbb{E}_{\mathcal{F}}\left(\left(S_n(g;\mathcal{D}_n^*) - \widetilde{S}_n(g;\mathcal{D}_n)\right)\right)^2 = \frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{n,i}^2) + o_p(1).$$

Hence it remains to show

$$\frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{n,i}^2) \xrightarrow{\mathbb{P}} V(g) \tag{4.20}$$

To show (4.20), Taylor expansion up to second order yields the following decomposition

$$H_{n,i} = H_{1,n,i} + H_{2,n,i}$$

where

$$H_{1,n,i} = \sum_{j,k=1}^{d} \partial_{jk} g(\hat{c}_{n,i})\left(\hat{c}_{n,i}^{*jk} - \hat{c}_{n,i}^{jk}\right)$$

and, by Cauchy-Schwartz inequality and the fact that $\mathbb{E}_{\mathcal{F}}\left(\hat{c}_{n,i}^{*jk} - \hat{c}_{n,i}^{jk}\right)^4 = O_p(k_n^{-2})$

$$\mathbb{E}_{\mathcal{F}}(H_{2,n,i}^2) = O_p(k_n^{-2}).$$

As for $H_{1,n,i}$, an adaption of Theorem 9.4.1 of (Jacod and Protter, 2012) to the nonoverlapping case yields

$$\begin{aligned}
&\frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{1,n,i}^2)\\
&= \frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\sum_{j,k,l,m=1}^{d}\partial_{jk} g(\hat{c}_{n,i})\partial_{lm} g(\hat{c}_{n,i})\mathbb{E}_{\mathcal{F}}\left[\left(\hat{c}_{n,i}^{*jk} - \hat{c}_{n,i}^{jk}\right)\left(\hat{c}_{n,i}^{*lm} - \hat{c}_{n,i}^{lm}\right)\right]\\
&= \frac{k_n}{n}\sum_{i=0}^{[n/k_n]-1}\sum_{j,k,l,m=1}^{d}\partial_{jk} g(\hat{c}_{n,i})\partial_{lm} g(\hat{c}_{n,i})\left(\hat{c}_{n,i}^{jl}\hat{c}_{n,i}^{km} + \hat{c}_{n,i}^{jm}\hat{c}_{n,i}^{kl}\right)\\
&\xrightarrow{\mathbb{P}} V(g)
\end{aligned}$$

Therefore, by Cauchy-Schwartz inequality

$$\begin{aligned}
\frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{n,i}^2) &= \frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{1,n,i}^2+2H_{1,n,i}H_{2,n,i}+H_{2,n,i}^2) \\
&= \frac{k_n^2}{n}\sum_{i=0}^{[n/k_n]-1}\mathbb{E}_{\mathcal{F}}(H_{1,n,i}^2)+O_p(k_n^{-1/2})+O_p(k_n^{-1}) \\
&\xrightarrow{\mathbb{P}} V(g).
\end{aligned}$$

Thus we prove (4.20), and (4.18) follows.

Step 3. At last we show (4.19). By mean value theorem and simple calculation, it holds that

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}}(H_{n,i}^4) &\leq K\mathbb{E}_{\mathcal{F}}\left(g(\hat{c}_{n,i}^*)-g(\hat{c}_{n,i})\right)+O_p(k_n^{-4}) \\
&= O_p(k_n^{-2}).
\end{aligned}$$

Plugging this into the left hand side of (4.19) gives the result.

Hence we finish the proof for Theorem 4.2.1.

4.6.2 Proof of Theorem 4.3.1

In this section we provide formal proof for Theorem 4.3.1, in which the bootstrap samples are generated using resampling with replacement. One could expect that the proof for Theorem 4.3.1 would be more involved than that of Theorem 4.2.1, due to the lack of (local) Gaussianity.

4.6.2.1 Elimination of jumps and truncation

In the spirit of (Jacod and Rosenbaum, 2013a), we find it useful to replace truncated returns by diffusive returns at the very beginning. To do so, define

$$X'_t = \int_0^t b'_s ds + \int_0^t \sigma_s dW_s,$$

where

$$b'_t = b_t + \int \delta(s,x) 1_{\{||\delta(s,x) \leq 1||\}} \lambda(dx). \tag{4.21}$$

Correspondingly, define

$$\begin{aligned} \hat{c}'_{n,i} &\equiv \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta^n_{ik_n+j} X'\right) \left(\Delta^n_{ik_n+j} X'\right)^\intercal \\ \widetilde{S}'_n(g; \mathcal{D}_n) &\equiv \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} g(\hat{c}'_{n,i}) \end{aligned}$$

In the local IID bootstrap setting, we also need to define

$$\begin{aligned} \hat{c}'^{*}_{n,i} &\equiv \frac{n}{k_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{ik_n+j^*_{i,\ell}} X'\right) \left(\Delta^n_{ik_n+j^*_{i,\ell}} X'\right)^\intercal \\ S'_n(g; \mathcal{D}^*_n) &\equiv \frac{k_n}{n} \sum_{i=0}^{[n/k_n]-1} \left(g(\hat{c}'^{*}_{n,i}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}'^{*}_{n,i}) \left(\hat{c}'^{*,jl}_{n,i} \hat{c}'^{*,km}_{n,i} + \hat{c}'^{*,jm}_{n,i} \hat{c}'^{*,kl}_{n,i} \right) \right) \end{aligned}$$

We will show that under Assumption 3.2.2, it would be enough to consider the diffusive returns $\Delta^n_{ik_n+j} X'$, in place of the truncated returns $\left(\Delta^n_{ik_n+j} X\right) 1_{\{||\Delta^n_{ik_n+j} X|| \leq u_n\}}$, which is mainly due to the following lemma on the convergence in conditional law.

Lemma 4.6.1. *Suppose $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ are two sequences of random variables such that*

$$X_n \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{L}, \quad \mathbb{E}_{\mathcal{F}} |Y_n| \xrightarrow{\mathbb{P}} 0, \tag{4.22}$$

for some distribution $\mathbb{L}$. Then it holds that $X_n + Y_n \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{L}$.

Proof. By definition of convergence in conditional law (e.g., Definition A1 in (Li and Xiu, 2016)), it suffices to show that for any bounded Lipschitz continuous function f,

$$\mathbb{E}_{\mathcal{F}} \left(f(X_n + Y_n) \right) \xrightarrow{\mathbb{P}} \int f(z) \mathbb{L}(dz).$$

Note that

$$\begin{aligned}
&\left|\mathbb{E}_{\mathcal{F}}\left(f(X_n+Y_n)\right)-\int f(z)\mathbb{L}(dz)\right| \\
&\leq \left|\mathbb{E}_{\mathcal{F}}\left(f(X_n+Y_n)\right)-\mathbb{E}_{\mathcal{F}}\left(f(X_n)\right)\right| + \left|\mathbb{E}_{\mathcal{F}}\left(f(X_n)\right)-\int f(z)\mathbb{L}(dz)\right|.
\end{aligned}$$

The first term converges to 0 because of f being Lipschitz and the second condition in (4.22), while the second term vanishes in probability due to the condition $X_n \xrightarrow{\mathcal{L}|\mathcal{F}} \mathbb{L}$ in (4.22). □

The following result officially allows us to replace truncated returns by diffusive returns.

Proposition 4.6.1. *To show* $\sqrt{n}\left(S_n(g;\mathcal{D}_n^*)-\widetilde{S}_n(g;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0,V(g))$, *it suffices to show*

$$\sqrt{n}\left(S_n^{'}(g;\mathcal{D}_n^*)-\widetilde{S}_n'(g;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0,V(g)).$$

Proof. Observe that

$$S_n(g;\mathcal{D}_n^*)-\widetilde{S}_n(g;\mathcal{D}_n) = S_n(g;\mathcal{D}_n^*)-S_n^{'}(g;\mathcal{D}_n^*)+S_n^{'}(g;\mathcal{D}_n^*)-\widetilde{S}_n'(g;\mathcal{D}_n)+\widetilde{S}_n'(g;\mathcal{D}_n)-\widetilde{S}_n(g;\mathcal{D}_n),$$

In the spirit of Lemma 4.6.1, it suffices to show

$$\sqrt{n}\mathbb{E}_{\mathcal{F}}\left|S_n(g;\mathcal{D}_n^*)-S_n^{'}(g;\mathcal{D}_n^*)\right| \xrightarrow{\mathbb{P}} 0, \tag{4.23}$$

$$\sqrt{n}\mathbb{E}\left|\widetilde{S}_n'(g;\mathcal{D}_n)-\widetilde{S}_n(g;\mathcal{D}_n)\right| \xrightarrow{\mathbb{P}} 0. \tag{4.24}$$

To show (4.23), let

$$h(x) = g(x) - \frac{1}{2k_n}\sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm}g(x)(x^{jl}x^{km}+x^{jm}x^{kl}) \tag{4.25}$$

which is in $\mathcal{C}_c^1$, since $g \in \mathcal{C}_c^3$. By mean value theorem,

$$\begin{aligned}
\sqrt{n}\mathbb{E}_{\mathcal{F}}\left|S_n(g;\mathcal{D}_n^*) - S_n^{'}(g;\mathcal{D}_n^*)\right| &\leq \frac{k_n}{\sqrt{n}} \sum_{i=0}^{[n/k_n]-1} \mathbb{E}_{\mathcal{F}}\left|h(\hat{c}_{n,i}^*) - h(\hat{c}_{n,i}^{'*})\right| \\
&\leq \frac{k_n}{\sqrt{n}} \sum_{i=0}^{[n/k_n]-1} \sum_{j,k=1}^{d} \mathbb{E}_{\mathcal{F}}\left|\hat{c}_{n,i}^{*jk} - \hat{c}_{n,i}^{'*jk}\right|
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}}\left|\hat{c}_{n,i}^{*jk} - \hat{c}_{n,i}^{'*jk}\right| &\leq \frac{n}{k_n}\sum_{\ell=1}^{k_n}\mathbb{E}_{\mathcal{F}}\left|\left(\Delta_{ik_n+j_{i,\ell}^*}^n X^j\right)\left(\Delta_{ik_n+j_{i,\ell}^*}^n X^k\right)1_{\{\|\Delta_{ik_n+j_{i,\ell}^*}^n X\|\leq u_n\}}\right. \\
&\qquad \left. - \left(\Delta_{ik_n+j_{i,\ell}^*}^n X^{'j}\right)\left(\Delta_{ik_n+j_{i,\ell}^*}^n X^{'k}\right)\right| \\
&= \frac{n}{k_n}\sum_{\ell=1}^{k_n}\left|\left(\Delta_{ik_n+\ell}^n X^j\right)\left(\Delta_{ik_n+\ell}^n X^k\right)1_{\{\|\Delta_{ik_n+\ell}^n X\|\leq u_n\}}\right. \\
&\qquad \left. - \left(\Delta_{ik_n+\ell}^n X^{'j}\right)\left(\Delta_{ik_n+\ell}^n X^{'k}\right)\right| \\
&\leq K a_n n^{(r-2)\varpi}
\end{aligned}$$

for some sequence of reals a_n converging to 0, and the last inequality is by the argument in page 15 of (Jacod and Rosenbaum, 2013a). Therefore,

$$\sqrt{n}\mathbb{E}_{\mathcal{F}}\left|S_n(g;\mathcal{D}_n^*) - S_n^{'}(g;\mathcal{D}_n^*)\right| \leq K a_n n^{(r-2)\varpi+\frac{1}{2}}$$

which vanishes since $\varpi \geq \frac{1-\gamma}{2-r} > \frac{1}{2(2-r)}$ by Assumption 3.2.2. Hence (4.23) follows.

(4.24) can be proved in the similar (and in fact simpler) way. □

4.6.2.2 Proof for the continuous case

In the spirit of Proposition 4.6.1, we will show

$$\sqrt{n}\left(S_n^{'}(g;\mathcal{D}_n^*) - \widetilde{S}_n^{'}(g;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, V(g)). \tag{4.26}$$

As in the case of parametric bootstrap, we will again use Theorem 2.2.14 in (Jacod and Protter, 2012) plus subsequence principle. To do so, we have to obtain the order of magnitude of returns and spot volatility estimates computed based on bootstrap samples using diffusive returns, which is more involved than the case of parametric bootstrap due to the lack of local Gaussianity. We collect the results in the following lemma, and again we restrict our attention to the set on which spot covariance estimates are uniformly bounded.

Lemma 4.6.2. *We have for any $i \in \mathcal{I}_n$,*

$$\mathbb{E}_{\mathcal{F}}(\hat{c}_{n,i}^{\prime *jk}) = \hat{c}_{n,i}^{\prime jk}, \quad \mathbb{E}_{\mathcal{F}}\left(\hat{c}_{n,i}^{\prime *jk} - \hat{c}_{n,i}^{\prime jk}\right)^4 = O_p(k_n^{-2})$$

$$\mathbb{E}_{\mathcal{F}}\left[\left(\hat{c}_{n,i}^{\prime *jk} - \hat{c}_{n,i}^{\prime jk}\right)\left(\hat{c}_{n,i}^{\prime *lm} - \hat{c}_{n,i}^{\prime lm}\right)\right] = \frac{1}{k_n}\left(\hat{c}_{n,i}^{\prime jl}\hat{c}_{n,i}^{\prime km} + \hat{c}_{n,i}^{\prime jm}\hat{c}_{n,i}^{\prime kl}\right) + O_p(k_n^{-3/2}).$$

Moreover, for any $p \geq 1$,

$$\mathbb{E}_{\mathcal{F}}\left(\hat{c}_{n,i}^{\prime *jk} - \hat{c}_{n,i}^{\prime jk}\right)^p = O_p(1).$$

Proof. The first identity can be obtained by simple calculations. Furthermore, note that

$$\hat{c}_{n,i}^{\prime *jk} - \hat{c}_{n,i}^{\prime jk} = \frac{1}{k_n}\sum_{\ell=1}^{k_n} Z_{n,i,\ell}^{jk}$$

where $Z_{n,i,\ell}^{jk} = \left(\sqrt{n}\Delta_{ik_n+j_{i,\ell}^*}^n X'^j\right)\left(\sqrt{n}\Delta_{ik_n+j_{i,\ell}^*}^n X'^k\right) - \hat{c}_{n,i}^{\prime jk}$. By the fact that $|\hat{c}_{n,i}^{\prime jk}| < K$ and the standard estimates for continuous Itô semimartingale, it follows for any $p \geq 1$,

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}}\left[\left(Z_{n,i,\ell}^{jk}\right)^p\right] &\leq \mathbb{E}_{\mathcal{F}}\left[\left(\sqrt{n}\Delta_{ik_n+j_{i,\ell}^*}^n X'^j\right)^p\left(\sqrt{n}\Delta_{ik_n+j_{i,\ell}^*}^n X'^k\right)^p\right] \\
&= \frac{1}{k_n}\sum_{\ell=1}^{k_n}\left(\sqrt{n}\Delta_{ik_n+\ell}^n X'^j\right)^p\left(\sqrt{n}\Delta_{ik_n+\ell}^n X'^k\right)^p \\
&= O_p(1)
\end{aligned}$$

Then the second equality in Lemma 4.6.2 follows from a direct calculation and the last one can be deduced from Jensen's inequality.

The third claim in Lemma 4.6.2 is a bit more involved. For notational simplicity, let

$$R_{n,i,\ell}^{jklm} = \Delta_{ik_n+\ell}^n X'^j \Delta_{ik_n+\ell}^n X'^k \Delta_{ik_n+\ell}^n X'^l \Delta_{ik_n+\ell}^n X'^m.$$

Direct calculation yields that

$$\mathbb{E}_{\mathcal{F}}\left[\left(\hat{c}_{n,i}^{\prime *jk} - \hat{c}_{n,i}^{\prime jk}\right)\left(\hat{c}_{n,i}^{\prime *lm} - \hat{c}_{n,i}^{\prime lm}\right)\right] = \frac{1}{k_n}\left(\frac{n^2}{k_n}\sum_{\ell=1}^{k_n} R_{n,i,\ell}^{jklm} - \hat{c}_{n,i}^{\prime jk}\hat{c}_{n,i}^{\prime lm}\right)$$

Then it suffices to show

$$\frac{n^2}{k_n}\sum_{\ell=1}^{k_n} R_{n,i,\ell}^{jklm} - \hat{c}_{n,i}^{\prime jk}\hat{c}_{n,i}^{\prime lm} - \hat{c}_{n,i}^{\prime jl}\hat{c}_{n,i}^{\prime km} - \hat{c}_{n,i}^{\prime jm}\hat{c}_{n,i}^{\prime kl} = O_p(k_n^{-1/2}), \tag{4.27}$$

The proof for (4.27) relies on the decomposition

$$\frac{n^2}{k_n}\sum_{\ell=1}^{k_n} R_{n,i,\ell}^{jklm} - \hat{c}_{n,i}^{\prime jk}\hat{c}_{n,i}^{\prime lm} - \hat{c}_{n,i}^{\prime jl}\hat{c}_{n,i}^{\prime km} - \hat{c}_{n,i}^{\prime jm}\hat{c}_{n,i}^{\prime kl} = R_{1,n,i} + R_{2,n,i} + R_{3,n,i}$$

where

$$\begin{aligned}
R_{1,n,i} &= \frac{1}{k_n}\sum_{\ell=1}^{k_n}\left[n^2 R_{n,i,\ell}^{jklm} - \mathbb{E}_{\mathcal{F}}\left(n^2 R_{n,i,\ell}^{jklm}\middle|\mathcal{F}_{(ik_n+\ell-1)/n}\right)\right] \\
R_{2,n,i} &= \frac{1}{k_n}\sum_{\ell=1}^{k_n}\left[\mathbb{E}_{\mathcal{F}}\left(n^2 R_{n,i,\ell}^{jklm}\middle|\mathcal{F}_{(ik_n+\ell-1)/n}\right) - c_{(ik_n+\ell-1)/n}^{jk} c_{(ik_n+\ell-1)/n}^{lm}\right. \\
&\qquad \left. - c_{(ik_n+\ell-1)/n}^{jl} c_{(ik_n+\ell-1)/n}^{km} - c_{(ik_n+\ell-1)/n}^{jm} c_{(ik_n+\ell-1)/n}^{kl}\right] \\
R_{3,n,i} &= \frac{1}{k_n}\sum_{\ell=1}^{k_n}\left(c_{(ik_n+\ell-1)/n}^{jk} c_{(ik_n+\ell-1)/n}^{lm} + c_{(ik_n+\ell-1)/n}^{jl} c_{(ik_n+\ell-1)/n}^{km} + c_{(ik_n+\ell-1)/n}^{jm} c_{(ik_n+\ell-1)/n}^{kl}\right. \\
&\qquad \left. - \hat{c}_{n,i}^{\prime jk}\hat{c}_{n,i}^{\prime lm} - \hat{c}_{n,i}^{\prime jl}\hat{c}_{n,i}^{\prime km} - \hat{c}_{n,i}^{\prime jm}\hat{c}_{n,i}^{\prime kl}\right).
\end{aligned}$$

Note that $R_{1,n,i}$ is the sum of a martingale difference sequence, by the standard estimates for Itô semimartingale, it readily deduces

$$R_{1,n,i} = O_p(k_n^{-1/2}). \tag{4.28}$$

Moreover, by Lemma 3.5.4 in the proof of Theorem 3.4.1 from Chapter 3, we have

$$R_{2,n,i} = O_p(n^{-1/2}). \tag{4.29}$$

Finally, for $R_{3,n,i}$, by estimates of the error of local covariance approximation under LMIS assumption, as given in the first claim of Lemma 3.5.6, we have

$$\mathbb{E}|R_{3,n,i}| \leq n^{-1/2} + k_n^{-1/2} + (k_n/n)^{1/2} \tag{4.30}$$

since $k_n^2/n \to 0$ by Assumption 3.2.2, the leading term is $k_n^{-1/2}$.

Combining (4.28) - (4.30) gives (4.27), and hence proves the lemma. $\square$

Now we come to show (4.26). Write

$$\sqrt{n}\left(S_n^{'}(g;\mathcal{D}_n^*) - \widetilde{S}_n'(g;\mathcal{D}_n)\right) = \sum_{i=0}^{[n/k_n]-1} \frac{k_n}{\sqrt{n}} H'_{n,i}$$

where

$$H'_{n,i} = g(\hat{c}'^*_{n,i}) - g(\hat{c}'_{n,i}) - \frac{1}{2k_n}\sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}'^*_{n,i})\left(\hat{c}'^{*jl}_{n,i}\hat{c}'^{*km}_{n,i} + \hat{c}'^{*jm}_{n,i}\hat{c}'^{*kl}_{n,i}\right).$$

In light of Theorem 2.2.14 in (Jacod and Protter, 2012) and subsequence principle, it suffices to show

$$\mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\left(S_n^{'}(g;\mathcal{D}_n^*) - \widetilde{S}_n'(g;\mathcal{D}_n)\right)\right) \xrightarrow{\mathbb{P}} 0 \tag{4.31}$$

$$\mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\left(S_n^{'}(g;\mathcal{D}_n^*) - \widetilde{S}_n'(g;\mathcal{D}_n)\right)\right)^2 \xrightarrow{\mathbb{P}} V(g) \tag{4.32}$$

$$\sum_{i=0}^{[n/k_n]-1} \mathbb{E}_{\mathcal{F}}\left(\sqrt{n}\frac{k_n}{n}H'_{n,i}\right)^4 \xrightarrow{\mathbb{P}} 0, \tag{4.33}$$

all of which, with the help of Lemma 4.6.2, can be produced word for word as in the parametric case. Hence we finish the prove of (4.26) and hence Theorem 4.3.1.

CHAPTER 5

Empirical-process-type CLTs for Estimating Integrated Volatility Functionals

In this chapter we further generalize the results from Chapter 3 and Chapter 4 on estimating and statistically inferring integrated volatility functionals. We consider a functional form of integrated volatility functionals where the test function has an extra indexing parameter. We extend an empirical-process-type asymptotic result in (Li and Xiu, 2016) to the case of allowing the indexing parameter to be of arbitrary finite dimension. We propose both parametric and nonparametric bootstrap algorithms which provide alternatives for statistically inferring integrated volatility functionals in this setting. As is the case with Chapter 4, we justify the two bootstrap methods by giving empirical-process-type asymptotic results. We emphasize that this project is still on-going with several potential applications to be done in the future.

5.1 Setting

The basic set-up in this chapter is quite similar to that of Chapter 3 and Chapter 4, except for two differences: the first one is that in this chapter we only assume the volatility process is an Itô semimartingale, which in particular does not contain a long-memory part as considered in Assumption 3.1.1; the second difference is that we will consider a functional form of the test function g given by

$$g : \mathcal{V} \times \Theta \to \mathbb{R},$$

where $\mathcal{V} \subset \mathbb{R}$ is the range space of spot volatility, and $\Theta \subset \mathbb{R}^{\mathbf{dim}\theta}$ is the space of indexing parameter θ. Thus for each fixed value c_t, which is already defined in the previous two chapters, $g(c_t, \cdot)$ is a function over Θ. Given $T > 0$, which may be a typical trading day, our goal is to (uniformly) estimate the quantity of the form

$$S(g;\theta) \equiv \int_0^T g(c_s;\theta)ds, \quad \theta \in \Theta.$$

Before we proceed to the estimation and statistical inference of $S(g,\theta)$, we give a couple examples to show the importance of $S(g,\theta)$ defined as such in economics and finance.

- In the degenerate case where θ is fixed, $S(g;\theta)$ reduces to the ordinary integrated volatility functionals $S(g) = \int_0^T g(c_s)ds$ given in (3.5). Then the specific examples include integrated volatility, quarticity, power variation, beta and idiosyncratic variance etc, see Section 3.2 for a more detailed discussion of this case.

- In general, θ varies over some parameter space Θ. In this case, a typical example would be

  $$g(x,\theta) = \exp(-\theta x), \quad \theta \in (0,\infty),$$

  in which case

  $$S(g;\theta) = \int_0^T e^{-\theta c_s}ds, \quad \theta \in (0,\infty)$$

  is the empirical Laplace Transform of c_t, which summarizes the complete spatial information of the volatility process within the time span. We refer readers to (Todorov and Tauchen, 2012b), (Todorov et al., 2012) and (Todorov and Tauchen, 2012a) for more details on realized Laplace transform.

- (Li and Xiu, 2016) considers a variant of GMM setting where moment conditions take the form of temporally integrated functionals of the sample paths of state variables including latent stochastic volatility. More specifically, let X_t be the price of an underlying asset, and Y_t be the price of a derivative contract written on it. Set $Z_t = (t, X_t, r_t, d_t)$ with short interest rate r_t and dividend yield d_t. If (Z_t, c_t) is Markovian under the risk-neutral measure, then

  $$Y_{i\Delta_n} = f(Z_{i\Delta_n}, c_{i\Delta_n}; \theta^*) + a_{i\Delta_n}\chi_i,$$

where f is a $\mathbb{R}-$valued function, θ^* arises from the risk-neutral model for the dynamics of Z_t, and a_t is the stochastic volatility of pricing errors χ_i which in turn satisfies

$$\mathbb{E}(\chi_i|\mathcal{F}) = 0,$$

where $\mathcal{F}$ denotes the whole information set defined over the underlying probability space. Then one can consider the integrated moment condition

$$S(g;\theta^*) = \int_0^T g(Y_t, Z_t, c_t; \theta^*)dt = 0,$$

with

$$g(Y_t, Z_t, c_t; \theta) = \int [Y_t - f(Z_t, c_t; \theta)] \, \mathbb{P}_\chi(d\chi)$$

where $\mathbb{P}_\chi(d\chi)$ denotes the marginal law, independent of $\mathcal{F}$, of pricing errors χ_i. To infer θ^*, we need to estimate $S(g, \theta)$ and study the associated statistical properties.

5.2 An Empirical-process-type Central Limit Theorem

Throughout the rest of this chapter, we assume Assumption 3.1.1 with $\sigma_2 = 0$. That is, the log-price process X is an Itô semimartingale with jumps of finite variation and the volatility process σ is a general Itô semimartingale. In particular, unlike the case of Chapter 3 and 4, we do not impose a long memory part on the volatility process.

We assume that only the process X is sampled at equidistant times $i\Delta_n$ with step size Δ_n at stage n, for $0 \leq i \leq \lfloor T/\Delta_n \rfloor$, within the *fixed* time interval $[0, T]$. Furthermore, the increments of X over $[(i-1)\Delta_n, i\Delta_n]$ are denoted by

$$\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}, \quad i = 1, \ldots, n.$$

Below, we consider an infill asymptotic scheme, that is, $\Delta_n \to 0$ as $n \to \infty$, while T is fixed.

Generalizing the estimator of $S(g,\theta)$ given in (Li and Xiu, 2016) to a multivariate case, the proposed estimator is

$$
\begin{aligned}
S_n(g;\theta;\mathcal{D}_n) \quad \equiv \quad \Delta_n & \sum_{i=0}^{[T/\Delta_n]-k_n} \Bigg(g(\hat{c}_{i\Delta_n};\theta) \\
& - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial^2_{jk,lm} g(\hat{c}_{i\Delta_n};\theta) \times (\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n} + \hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}) \Bigg),
\end{aligned}
$$

where, following the convention in Chapter 4, $\mathcal{D}_n \equiv \{\Delta^n_i X, i = 1,\ldots,n\}$ represents original returns. Note here that the (usual) spot volatility estimator is, for any $0 \le i \le [T/\Delta_n] - k_n$ and $1 \le l,m \le d$,

$$
\hat{c}^{lm}_{i\Delta_n} \equiv \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} \Delta^n_{i+j} X^l \Delta^n_{i+j} X^m 1_{\{||\Delta^n_{i+j}X|| \le u_n\}}
$$

where k_n is a sequence of integers denoting the number of increments employed in a local window and u_n determines the truncation threshold for eliminating jumps in X. Assumption 3.2.2 states the exact conditions imposed on the tuning parameters k_n and u_n.

Based on the proof of (Li and Xiu, 2016), we have the following empirical-process-type central limit theorem:

Theorem 5.2.1. *Suppose Assumption 3.1.1 with $\sigma_2 = 0$, and Assumption 3.2.2. Moreover, assume $g : \mathcal{V} \times \Theta \to \mathbb{R}$ satisfies Assumption 3.2.1 with respect to the first variate and is continuously differentiable with respect to $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^{\boldsymbol{dim\theta}}$ is a compact set. Then if*

$$
dim\theta < 2(1-\gamma)/\gamma, \tag{5.1}
$$

the sequence $\Delta_n^{-1/2}(S_n(g;\cdot;\mathcal{D}_n) - S(g;\cdot))$ of processes converges $\mathcal{F}-$stably in law under the uniform metric to a process $\xi(\cdot)$ which, conditional on $\mathcal{F}$, is centered Gaussian with covariance function $S_g(\cdot,\cdot)$, where $S_g(\cdot,\cdot)$ is defined as, for any $\theta,\theta' \in \Theta$,

$$
S_g(\theta,\theta') \equiv \sum_{j,k,l,m=1}^{d} \int_0^T \partial_{jk} g(c_s;\theta) \partial_{lm} g(c_s;\theta') \left(c^{jl}_s c^{km}_s + c^{jm}_s c^{kl}_s \right) ds.
$$

However, Theorem 5.2.1 is not quite satisfactory due to the restriction (5.1). Put it more precisely, because of Assumption 3.2.2, it follows that $2(1-\gamma)/\gamma < 4$. As a result, it shold hold that $\dim\theta < 4$, which is not only restrictive from a theoretical point of view, but confines the scope of applications.

We point out here that in the proof of (Li and Xiu, 2016), the condition (5.1) arises when showing the stochastic equicontinuity of the empirical-process-type central limit theorem. Nevertheless, by using a method which separates the jump part and the continuous part of the underlying volatility process, we are able to get rid of (5.1), namely,

Theorem 5.2.2. *Suppose Assumption 3.1.1 with $\sigma_2 = 0$, and Assumption 3.2.2. Moreover, assume $g : \mathcal{V} \times \Theta \to \mathbb{R}$ satisfies Assumption 3.2.1 with respect to the first variate and is continuously differentiable with respect to $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^{\boldsymbol{dim\theta}}$ is a compact set, with $dim\theta < \infty$. Then the sequence $\Delta_n^{-1/2}\left(S_n(g;\cdot;\mathcal{D}_n) - S(g;\cdot)\right)$ of processes converges $\mathcal{F}-$stably in law under the uniform metric to a process $\xi(\cdot)$ which, conditional on $\mathcal{F}$, is centered Gaussian with covariance function $S_g(\cdot,\cdot)$, where $S_g(\cdot,\cdot)$ is defined as, for any $\theta, \theta' \in \Theta$,*

$$S_g(\theta,\theta') \equiv \sum_{j,k,l,m=1}^{d} \int_0^T \partial_{jk} g(c_s;\theta)\partial_{lm} g(c_s;\theta')\left(c_s^{jl}c_s^{km} + c_s^{jm}c_s^{kl}\right) ds.$$

Several comments are worth mentioning, besides the condition (5.1) is removed. Firstly, in order to prove the functional central limit theorem stated in the theorem, we need to show the convergence of finite dimensional distributions and the stochastic equicontinuity. Since the convergence of finite dimensional distributions is easily implied by Theorem 3.4.1 with Cramér-Wold device, we only need to show the stochastic equicontinuity, the details of which are presented in the Proofs of this chapter.

Secondly, as far as the *efficiency* is concerned, although it has not been proved that in such functional setting the asymptotic variance $S_g(\cdot,\cdot)$ is the smallest, $S_g(\theta,\theta')$ turns out to be smaller than the asymptotic variance of the functional central limit theorem proved in (Todorov and Tauchen, 2012b) for some specific values of θ and θ'.

Last but not least, from an application point of view, we can use Theorem 5.2.2 to construct confidence band (region) for $\{S_n(g;\theta;\mathcal{D}_n), \theta \in \Theta \subset \mathbb{R}^{dim(\theta)}\}$ by taking advantage of the fact that the "sup" functional is a continuous mapping on the space of continuous functions, together with the continuous mapping theorem.

5.3 Bootstrap Inference

As pointed out above, Theorem 5.2.2 can be used to construct confidence region for $S_n(g;\cdot)$. However, in practice the critical values for confidence region may not be determined from the asymptotic covariance when Θ is not finite, see (Kosorok, 2007). Alternatively, one may turn to bootstrap methods. In the sequel we introduce both parametric and nonparametric bootstrap methods in this functional setting and provide empirical-process-type asymptotic results to (theoretically) justify the two methods.

As is the case with Chapter 4, we only consider non-overlapping case and we defer the study of overlapping case for future research. In view of this change we redefine the aforementioned quantities as follows: for each $i \in \mathcal{I}_n \equiv \{0, \ldots, [T/k_n\Delta_n] - 1\}$, let

$$
\begin{aligned}
\hat{c}_{i\Delta_n} &= \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X\right) \left(\Delta_{ik_n+j}^n X\right)^\intercal 1_{\{||\Delta_{ik_n+j}^n X|| \le u_n\}} \\
\widetilde{S}_n(g;\theta;\mathcal{D}_n) &= k_n\Delta_n \sum_{i=0}^{[T/k_n\Delta_n]-1} g(\hat{c}_{i\Delta_n};\theta) \\
S_n(g;\theta;\mathcal{D}_n) &= k_n\Delta_n \sum_{i=0}^{[T/k_n\Delta_n]-1} \Bigg(g(\hat{c}_{i\Delta_n};\theta) \\
&\quad - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(\hat{c}_{i\Delta_n};\theta) \left(\hat{c}_{i\Delta_n}^{jl} \hat{c}_{i\Delta_n}^{km} + \hat{c}_{i\Delta_n}^{jm} \hat{c}_{i\Delta_n}^{kl} \right) \Bigg).
\end{aligned}
$$

5.3.1 Parametric Bootstrap

We start with the parametric bootstrap method, which regenerates the bootstrap return samples by using normal distribution.

Algorithm 3. (Parametric Bootstrap)

Step 1. For each $i \in \mathcal{I}_n$, estimate $\hat{c}_{i\Delta_n}$, and compute $\widetilde{S}_n(g;\theta;\mathcal{D}_n)$ and $S_n(g;\theta;\mathcal{D}_n)$, respectively

Step 2. For each $i \in \mathcal{I}_n$, simulate $\Delta_{ik_n+j}^n X^* \sim \mathcal{N}(0, \Delta_n \hat{c}_{i\Delta_n})$ for $j = 1, \ldots, k_n$.

Step 3. Compute bootstrap spot covariance estimators using $\Delta_i^n X^*$, namely,

$$\hat{c}_{i\Delta_n}^* = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{ik_n+j}^n X^*\right) \left(\Delta_{ik_n+j}^n X^*\right)^\intercal .$$

Step 4. Compute bootstrap estimator for $S(g;\theta)$ as

$$\begin{aligned} S_n(g;\theta;\mathcal{D}_n^*) &\equiv k_n \Delta_n \sum_{i=0}^{[T/k_n\Delta_n]-1} \Bigg(g(\hat{c}_{i\Delta_n}^*;\theta) \\ &\quad - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 g(\hat{c}_{i\Delta_n}^*;\theta) \left(\hat{c}_{i\Delta_n}^{*jl} \hat{c}_{i\Delta_n}^{*km} + \hat{c}_{i\Delta_n}^{*jm} \hat{c}_{i\Delta_n}^{*kl} \right) \Bigg), \end{aligned}$$

where $\mathcal{D}_n^* \equiv \{\Delta_i^n X^*, i = 1, \ldots, n\}$ represents the set of returns calculated from bootstrap return samples generated from Step 2. Then the empirical distribution of any statistic of $S_n(g;\theta;\mathcal{D}_n)$ could be calculated accordingly.

In theory, such parametric bootstrap method is justified by the following asymptotic result:

Theorem 5.3.1. *Under the same assumptions as in Theorem 5.2.2, and let $S_n(g;\theta;\mathcal{D}_n^*)$ be given by Algorithm 3 (parametric algorithm). It follows that*

$$\Delta_n^{-\frac{1}{2}} \left(S_n(g;\cdot;\mathcal{D}_n^*) - \widetilde{S}_n(g;\cdot;\mathcal{D}_n) \right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, S_g(\cdot,\cdot)),$$

under uniform metric, where for any $\theta, \theta' \in \Theta$,

$$S_g(\theta,\theta') \equiv \sum_{j,k,l,m=1}^{d} \int_0^T \partial_{jk} g(c_s;\theta) \partial_{lm} g(c_s;\theta') \left(c_s^{jl} c_s^{km} + c_s^{jm} c_s^{kl} \right) ds.$$

.

The proof of Theorem 5.3.1 consists of two parts: the convergence of finite dimensional distributions under "$\mathcal{L}|\mathcal{F}$" with Cramér-Wold device, which is true by Theorem 4.2.1, and the stochastic equicontinuity under "$\mathcal{L}|\mathcal{F}$".

5.3.2 The Local IID Bootstrap Bootstrap

Now we proceed to the nonparametric bootstrap method. As in Section 4.3, the name "local IID" comes from the fact the bootstrap return samples are generated by resampling with replacement. The algorithm goes as follows:

Algorithm 4. (Local IID Bootstrap)

Step 1. For each $i \in \mathcal{I}_n$, estimate $\hat{c}_{i\Delta_n}$, and compute $\widetilde{S}_n(g;\theta;\mathcal{D}_n)$ and $S_n(g;\theta;\mathcal{D}_n)$, respectively.

Step 2. For each $i \in \mathcal{I}_n$, compute bootstrap spot covariance estimators

$$\hat{c}^*_{i\Delta_n} = \frac{1}{k_n\Delta_n}\sum_{\ell=1}^{k_n}\left(\Delta^n_{ik_n+j^*_{i,\ell}}X\right)\left(\Delta^n_{ik_n+j^*_{i,\ell}}X\right)^{\intercal} 1_{\{\|\Delta^n_{ik_n+j^*_{i,\ell}}X\|\leq u_n\}},$$

where for each i and ℓ,

$$j^*_{i,\ell} \sim \text{i.i.d. Uniform}\{1,\ldots,k_n\}.$$

Step 3-4 are the same as in Algorithm 3.

The theoretical justification of Algorithm 4 is given by the following theorem.

Theorem 5.3.2. *Under the same assumptions as in Theorem 5.2.2, and let $S_n(g;\theta;\mathcal{D}^*_n)$ be given by Algorithm 4 (local IID algorithm). It follows that*

$$\Delta_n^{-\frac{1}{2}}\left(S_n(g;\cdot;\mathcal{D}^*_n) - \widetilde{S}_n(g;\cdot;\mathcal{D}_n)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{MN}(0, S_g(\cdot,\cdot)),$$

under uniform metric, where for any $\theta,\theta' \in \Theta$,

$$S_g(\theta,\theta') \equiv \sum_{j,k,l,m=1}^{d}\int_0^T \partial_{jk}g(c_s;\theta)\partial_{lm}g(c_s;\theta')\left(c_s^{jl}c_s^{km} + c_s^{jm}c_s^{kl}\right)ds.$$

.

5.4 Future Work

It would be interesting to explore the efficiency of the asymptotic result given in Theorem 5.2.2, namely, whether or not the asymptotic variance $S_g(\cdot,\cdot)$, which is essentially a kernel function, is small enough. In particular, it is worth comparing $S_g(\cdot,\cdot)$ with that of the asymptotic result proved in (Todorov and Tauchen, 2012b), in the special case of estimating empirical Laplace transform of volatility.

On the other hand, we would like to see the finite sample performance of the proposed estimator in such functional setting. We would also like to add more econometric applications, including misspecification test, in this setting. Furthermore, we want to see whether the same asymptotic results would hold for volatility process being LMIS and the overlapping case. Last but not least, it might be possible to estimate $S(g;\theta;\mathcal{D}_n)$ without explicit bias correction in the sense of (Li and Xiu, 2017).

5.5 Proof

The proof in this section follows the proof in (Li and Xiu, 2016). To reduce repetition, we only lay out the steps that are crucial to the theorems under proof, and we refer readers to (Li and Xiu, 2016) for the context.

5.5.1 Proof of Theorem 5.2.2

Following the proof of Lemma A4 in (Li and Xiu, 2016), it suffices to show the uniform convergence of $R_{3,n}(\theta)$ and $R_{4,n}(\theta)$ with respect to θ. Moreover, we inherit the notations from (Li and Xiu, 2016) to facilitate the understanding of the proof here with that of (Li and Xiu, 2016). In particular, we use V_t, instead of c_t, to denote the stochastic covariance process in the proof of Theorem 5.2.2. By polarization argument, we assume $d = 1$.

5.5.1.1 Uniform Convergence for $R_{3,n}(\theta)$ w.r.t. θ

$$R_{3,n}(\theta) \equiv \Delta_n^{1/2} \sum_i \partial_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) k_n^{-1} \sum_{u=1}^{k_n} (V_{(i+u-1)\Delta_n} - V_{i\Delta_n}).$$

We want to show

$$\sup_{\theta\in\Theta} ||R_{3,n}(\theta)|| = o_p(1). \tag{5.2}$$

In order to do so we need to separate the continuous part and the jump part for the volatility process V_t: recall that by Assumption 3.1.1,

$$\begin{aligned} V_t &= V_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,x) 1_{\{||\tilde{\delta}(s,x)\leq 1||\}} (\mu-\nu)(ds,dx) \\ &+ \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,x) 1_{\{||\tilde{\delta}(s,x)>1||\}} (\mu-\nu)(ds,dx) \end{aligned}$$

Define

$$\begin{aligned} V_t^{cts} &\equiv V_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s \\ V_t^{jump} &\equiv \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,x) 1_{\{||\tilde{\delta}(s,x)\leq 1||\}} (\mu-\nu)(ds,dx) \\ &+ \int_0^t \int_{\mathbb{R}} \tilde{\delta}(s,x) 1_{\{||\tilde{\delta}(s,x)>1||\}} (\mu-\nu)(ds,dx) \end{aligned}$$

and correspondingly $R_{3,n}(\theta)$ decomposes into two parts

$$R_{3,n}(\theta) = R_{3,n}^{cts}(\theta) + R_{3,n}^{jump}(\theta),$$

where

$$\begin{aligned} R_{3,n}^{cts}(\theta) &\equiv \Delta_n^{1/2} \sum_i \partial_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) k_n^{-1} \sum_{u=1}^{k_n} (V_{(i+u-1)\Delta_n}^{cts} - V_{i\Delta_n}^{cts}) \\ R_{3,n}^{jump}(\theta) &\equiv \Delta_n^{1/2} \sum_i \partial_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) k_n^{-1} \sum_{u=1}^{k_n} (V_{i\Delta_n}^{jump} - V_{(i+u-1)\Delta_n}^{jump}). \end{aligned}$$

By the martingale difference argument in Li and Xiu (2016), it is known that

$$\sup_{\theta\in\Theta} ||R_{3,n}^{cts}(\theta)|| = o_p(1).$$

Then to show (5.2), it is left to show

$$\sup_{\theta\in\Theta} ||R_{3,n}^{jump}(\theta)|| = o_p(1).$$

which is proved as follows.

First of all, since $\partial_v f(\cdot)$ is uniformly bounded, we have

$$\begin{aligned} ||R_{3,n}^{jump}(\theta)|| &\leq ||\Delta_n^{1/2}\sum_i \partial_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta)k_n^{-1}\sum_{u=1}^{k_n}(V_{i\Delta_n}^{jump} - V_{(i+u-1)\Delta_n}^{jump})|| \\ &\leq \Delta_n^{1/2}\sum_i k_n^{-1}\sum_{u=1}^{k_n}||V_{i\Delta_n}^{jump} - V_{(i+u-1)\Delta_n}^{jump}||. \end{aligned}$$

Since the right hand side does not depend on θ, taking supreme over $\theta \in \Theta$ on the left hand side and then the expectation on both sides gives

$$\begin{aligned} \mathbb{E}\left(\sup_{\theta\in\Theta}||R_{3,n}^{jump}(\theta)||\right) &\leq \Delta_n^{1/2}\sum_i k_n^{-1}\sum_{u=1}^{k_n}\underbrace{\mathbb{E}||(V_{i\Delta_n}^{jump} - V_{(i+u-1)\Delta_n}^{jump})||}_{\leq k_n\Delta_n} \\ &\leq \Delta_n^{1/2}\sum_i k_n\Delta_n \\ &= k_n\Delta_n^{1/2} \to 0. \end{aligned}$$

where we have used the standard estimates of first moments for both compensated small jumps process (Lemma 2.1.5 in Jacod and Protter (2012)), and big jumps process (Lemma 2.1.7 in Jacod and Protter (2012)).

5.5.1.2 Uniform Convergence for $R_{4,n}(\theta)$ w.r.t. θ

Recall

$$R_{4,n}(\theta) = R'_{4,n}(\theta) + R''_{4,n}(\theta).$$

It is already known that

$$\sup_{\theta\in\Theta} ||R'_{4,n}(\theta)|| = o_p(1),$$

see equation (A.37) in Li and Xiu (2016).

Then it remains to show

$$\sup_{\theta\in\Theta} ||R''_{4,n}(\theta)|| = o_p(1), \tag{5.3}$$

where

$$\begin{aligned} R''_{4,n}(\theta) &\equiv \Delta_n^{1/2} \sum_i \frac{1}{2} \partial^2_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) \left(\tilde{v}^2_{n,i} - \mathbb{E}(\tilde{v}^2_{n,i}|\mathcal{F}_{i\Delta_n})\right) \\ \tilde{v}_{n,i} &\equiv \hat{V}'_{i\Delta_n} - V_{i\Delta_n}. \end{aligned}$$

Decompose the local estimation error $\tilde{v}_{n,i}$ into statistical error $S_{n,i}$ and target error $D_{n,i}$ as follows

$$\tilde{v}_{n,i} = \hat{V}'_{i\Delta_n} - \overline{V}_{i\Delta_n} + \overline{V}_{i\Delta_n} - V_{i\Delta_n} \equiv S_{n,i} + D_{n,i}$$

with

$$\overline{V}_{i\Delta_n} \equiv \frac{1}{k_n\Delta_n} \int_{i\Delta_n}^{(i+k_n)\Delta_n} V_s ds.$$

Plug this into $R''_{4,n}(\theta)$ we obtain

$$R''_{4,n}(\theta) = R''_{4,n,S}(\theta) + R''_{4,n,SD}(\theta) + R''_{4,n,D}(\theta)$$

where

$$\begin{aligned} R''_{4,n,S}(\theta) &= \Delta_n^{1/2} \sum_i \frac{1}{2} \partial^2_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) \left(S^2_{n,i} - \mathbb{E}(S^2_{n,i}|\mathcal{F}_{i\Delta_n})\right) \\ R''_{4,n,SD}(\theta) &= \Delta_n^{1/2} \sum_i \frac{1}{2} \partial^2_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) \left(2S_{n,i}D_{n,i} - \mathbb{E}(2S_{n,i}D_{n,i}|\mathcal{F}_{i\Delta_n})\right) \\ R''_{4,n,D}(\theta) &= \Delta_n^{1/2} \sum_i \frac{1}{2} \partial^2_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n};\theta) \left(D^2_{n,i} - \mathbb{E}(D^2_{n,i}|\mathcal{F}_{i\Delta_n})\right) \end{aligned}$$

By the argument used in Li and Xiu (2016), we know that

$$\sup_{\theta\in\Theta} ||R''_{4,n,S}(\theta)|| = o_p(1),$$

mainly due to the fact $||S^2_{n,i}|| \leq k_n^{-1}$.

In view of Cauchy-Schwartz inequality (to take care of $R''_{4,n,SD}(\theta)$), it is only left to show

$$\sup_{\theta\in\Theta} ||R''_{4,n,D}(\theta)|| = o_p(1),$$

the proof of which goes as follows:

Since $\partial^2_{\tilde{Z}} f(\cdot)$ is uniformly bounded, by triangle inequality

$$\begin{aligned}
||R''_{4,n,D}(\theta)|| &= ||\Delta_n^{1/2} \sum_i \frac{1}{2} \partial^2_{\tilde{Z}} f(\tilde{Z}_{i\Delta_n}; \theta) \left(D^2_{n,i} - \mathbb{E}(D^2_{n,i} | \mathcal{F}_{i\Delta_n}) \right) || \\
&\leq \Delta_n^{1/2} \sum_i \left(||D^2_{n,i}|| + \mathbb{E}(D^2_{n,i} | \mathcal{F}_{i\Delta_n}) \right) \\
&\leq \Delta_n^{1/2} \sum_i \left(||D_{n,i}||^2 + \mathbb{E}(D^2_{n,i} | \mathcal{F}_{i\Delta_n}) \right)
\end{aligned}$$

Note that the right hand side does not depend on θ (which is removed together with $\partial^2_v f(\cdot)$), taking supreme w.r.t. θ gives

$$\sup_{\theta\in\Theta} ||R''_{4,n,D}(\theta)|| \leq \Delta_n^{1/2} \sum_i \left(||D_{n,i}||^2 + \mathbb{E}(D^2_{n,i} | \mathcal{F}_{i\Delta_n}) \right)$$

and hence

$$\mathbb{E}\left(\sup_{\theta\in\Theta} ||R''_{4,n,D}(\theta)|| \right) \leq \Delta_n^{1/2} \sum_i \mathbb{E}\left(||D_{n,i}||^2 \right). \tag{5.4}$$

Since

$$
\begin{aligned}
\mathbb{E}(D_{n,i}^2) &= \mathbb{E}(\overline{V}_{i\Delta_n} - V_{i\Delta_n})^2 \\
&= \mathbb{E}\left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} V_s ds - V_{i\Delta_n}\right)^2 \\
&= \mathbb{E}\left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} (V_s - V_{i\Delta_n}) ds\right)^2 \\
&\leq \mathbb{E}\left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} (V_s - V_{i\Delta_n})^2 ds\right) \quad \text{Jensen's inequality} \\
&= \frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} \underbrace{\mathbb{E}(V_s - V_{i\Delta_n})^2}_{\leq k_n\Delta_n} ds \quad \text{Fubini's Thm} \\
&\leq k_n\Delta_n,
\end{aligned}
$$

where we only use the standard estimate for Itö semimartingale for second moment, in which case the magnitudes for both diffusive part and jump part are equal to one.

Substitute into (5.4), we obtain

$$
\mathbb{E}\left(\sup_{\theta\in\Theta} ||R''_{4,n,D}(\theta)||\right) \leq \Delta_n^{1/2} \sum_i k_n\Delta_n = \Delta_n^{1/2}\Delta_n^{-1}k_n\Delta_n = k_n\Delta_n^{1/2} \to 0,
$$

and thus (5.3) is proved.

5.5.2 Proof of Theorem 5.3.1

Now we come to prove Theorem 5.3.1. We start with the following lemma.

Lemma 5.5.1. *For any $p \geq 2$, it holds that*

$$
\mathbb{E}_{\mathcal{F}}\|\hat{c}^*_{i\Delta_n} - \hat{c}_{i\Delta_n}\|^p = O_p(k_n^{-\frac{p}{2}})
$$

Proof. By definition of $\hat{c}^*_{i\Delta_n}$ given in Algorithm 3, we have for some constant $K > 0$,

$$
\begin{aligned}
\mathbb{E}_{\mathcal{F}} \left\| \hat{c}^*_{i\Delta_n} - \hat{c}_{i\Delta_n} \right\|^p &= \mathbb{E}_{\mathcal{F}} \left\| \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta^n_{ik_n+j} X^* \right) \left(\Delta^n_{ik_n+j} X^* \right)^\intercal - \hat{c}_{i\Delta_n} \right\|^p \\
&= k_n^{-p} \mathbb{E}_{\mathcal{F}} \left\| \sum_{j=1}^{k_n} \left(\left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right) \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right)^\intercal - \hat{c}_{i\Delta_n} \right) \right\|^p \\
&\leq K k_n^{-p} \mathbb{E}_{\mathcal{F}} \left\| \sum_{j=1}^{k_n} \left(\left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right) \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right)^\intercal - \hat{c}_{i\Delta_n} \right)^2 \right\|^{p/2} \\
&= K k_n^{-p} k_n^{p/2} \mathbb{E}_{\mathcal{F}} \left\| \frac{1}{k_n} \sum_{j=1}^{k_n} \left(\left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right) \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right)^\intercal - \hat{c}_{i\Delta_n} \right)^2 \right\|^{p/2} \\
&\leq K k_n^{-p/2} \frac{1}{k_n} \sum_{j=1}^{k_n} \mathbb{E}_{\mathcal{F}} \left\| \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right) \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right)^\intercal - \hat{c}_{i\Delta_n} \right\|^p \\
&\leq O_p(k_n^{-p/2}),
\end{aligned}
$$

where the first inequality is by Burkholder-Davis-Gundy inequality and the second last inequality is by Jensen's inequality. The last inequality follows by recognizing that

$$
\mathbb{E}_{\mathcal{F}} \left\| \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right) \left(\frac{\Delta^n_{ik_n+j} X^*}{\sqrt{\Delta_n}} \right)^\intercal - \hat{c}_{i\Delta_n} \right\|^p
$$

is stochastically bounded because of the way the bootstrap returns $\Delta^n_{ik_n+j} X^*$ are generated. Hence we prove the result. □

Now we proceed to prove Theorem 5.3.1. In view of Theorem 5.2.2, is suffices to show the stochastic equicontinuity of $\Delta_n^{-\frac{1}{2}} \left(S_n(g; \cdot; \mathcal{D}^*_n) - \widetilde{S}_n(g; \cdot; \mathcal{D}_n) \right)$. We first have the following decomposition by multivariate Taylor expansion,

$$
\Delta_n^{-\frac{1}{2}} \left(S_n(g; \cdot; \mathcal{D}^*_n) - \widetilde{S}_n(g; \cdot; \mathcal{D}_n) \right) = R_{1,n}(\theta) + R_{2,n}(\theta) + R_{3,n}(\theta)
$$

where

$$
\begin{aligned}
R_{1,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k=1}^{d}\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n}) \\
R_{2,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\frac{1}{2}\Bigg(\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\quad -\frac{1}{k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}^{*}_{i\Delta_n};\theta)\left(\hat{c}^{*jl}_{i\Delta_n}\hat{c}^{*km}_{i\Delta_n}+\hat{c}^{*jm}_{i\Delta_n}\hat{c}^{*kl}_{i\Delta_n}\right)\Bigg), \\
R_{3,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\times \\
&\quad \frac{1}{6}\sum_{j,k,l,m,u,v=1}^{d}\partial^3_{jk,lm,uv}g(\widetilde{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n})(\hat{c}^{*uv}_{i\Delta_n}-\hat{c}^{uv}_{i\Delta_n})
\end{aligned}
$$

where $\widetilde{c}_{i\Delta_n}$ are intermediate values between $\hat{c}^{*}_{i\Delta_n}$ and $\hat{c}_{i\Delta_n}$ as is usual the case with Taylor expansion.

It is easy to see that $R_{3,n}(\theta)$ is stochastic equicontinuous. In fact, recall that we can assume g is compactly supported according to Assumption 3.2.1, then it holds that

$$
\begin{aligned}
\mathbb{E}\left(\sup_{\theta\in\Theta}\|R_{3,n}(\theta)\|\right) &\leq Kk_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\mathbb{E}\|\hat{c}^{*}_{i\Delta_n}-\hat{c}_{i\Delta_n}\|^3 \\
&\leq \Delta_n^{-1/2}k_n^{-3/2},
\end{aligned}
$$

where the last inequality is due to Lemma 5.5.1. Then according to Assumption 3.2.2, $k_n^3\Delta_n\to\infty$, and hence

$$
\sup_{\theta\in\Theta}\|R_{3,n}(\theta)\| = o_p(1).
$$

Therefore, $R_{3,n}(\theta)$ is stochastic equicontinuous.

Next we deal with $R_{1,n}(\theta)$. Note that by definition of $\hat{c}^*_{i\Delta_n}$,

$$
\begin{aligned}
R_{1,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k=1}^{d}\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n}) \\
&= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k=1}^{d}\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)\left(\frac{1}{k_n\Delta_n}\sum_{\ell=1}^{k_n}\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\hat{c}_{i\Delta_n}\right) \\
&= \Delta_n^{-1/2}\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k=1}^{d}\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)\sum_{\ell=1}^{k_n}\left(\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\Delta_n\hat{c}_{i\Delta_n}\right).
\end{aligned}
$$

Then it follows that for any other $\theta'\in\Theta$, we have for any $p>2$,

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{F}}\|R_{1,n}(\theta)-R_{1,n}(\theta')\|^p \\
&=\mathbb{E}_{\mathcal{F}}\Bigg\|\Delta_n^{-1/2}\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k=1}^{d}\left(\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)-\partial_{jk}g(\hat{c}_{i\Delta_n};\theta')\right)\times \\
&\qquad\sum_{\ell=1}^{k_n}\left(\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\Delta_n\hat{c}_{i\Delta_n}\right)\Bigg\|^p \\
&\le K\Delta_n^{-p/2}\mathbb{E}_{\mathcal{F}}\Bigg\|\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{\ell=1}^{k_n}\sum_{j,k=1}^{d}\left(\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)-\partial_{jk}g(\hat{c}_{i\Delta_n};\theta')\right)^2\times \\
&\qquad\left(\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\Delta_n\hat{c}_{i\Delta_n}\right)^2\Bigg\|^{p/2} \\
&\le K\Delta_n^{-p/2}\left(\frac{1}{k_n\Delta_n}k_n\right)^{p/2}\mathbb{E}_{\mathcal{F}}\Bigg\|\left(\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)-\partial_{jk}g(\hat{c}_{i\Delta_n};\theta')\right)^2\times \\
&\qquad\left(\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\Delta_n\hat{c}_{i\Delta_n}\right)^2\Bigg\|^{p/2} \\
&\le K\Delta_n^{-p}\mathbb{E}_{\mathcal{F}}\Bigg\|\left(\partial_{jk}g(\hat{c}_{i\Delta_n};\theta)-\partial_{jk}g(\hat{c}_{i\Delta_n};\theta')\right)\left(\Delta^n_{ik_n+\ell}X^{*j}\Delta^n_{ik_n+\ell}X^{*k}-\Delta_n\hat{c}_{i\Delta_n}\right)\Bigg\|^p
\end{aligned}
$$

where we have in order used Burkholder-Davis-Gundy inequality and Jensen's inequality. Then the continuous differentiability of g with respect to indexing parameter θ and the stochastic boundedness of

$$
\mathbb{E}_{\mathcal{F}}\left\|\frac{\Delta^n_{ik_n+\ell}X^{*j}}{\sqrt{\Delta_n}}\frac{\Delta^n_{ik_n+\ell}X^{*k}}{\sqrt{\Delta_n}}-\hat{c}_{i\Delta_n}\right\|^p
$$

imply

$$\mathbb{E}\|R_{1,n}(\theta) - R_{1,n}(\theta')\|^p \leq K\|\theta - \theta'\|^p.$$

Then by taking $p > \max(\dim(\theta), 2)$, we deduce that $R_{1,n}(\theta)$ is stochastic equicontinuous.

For $R_{2,n}(\theta)$, we observe the following decomposition

$$R_{2,n}(\theta) = R'_{2,n}(\theta) + R''_{2,n}(\theta)$$

where

$$\begin{aligned}
R'_{2,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\frac{1}{2}\Bigg(\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n} - \hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n} - \hat{c}^{lm}_{i\Delta_n}) \\
&\quad -\frac{1}{k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n} + \hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg) \\
R''_{2,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\Bigg(\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n} + \hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right) \\
&\quad -\frac{1}{2k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}^{*}_{i\Delta_n};\theta)\left(\hat{c}^{*jl}_{i\Delta_n}\hat{c}^{*km}_{i\Delta_n} + \hat{c}^{*jm}_{i\Delta_n}\hat{c}^{*kl}_{i\Delta_n}\right)\Bigg).
\end{aligned}$$

By mean value theorem, the fact that $\partial^3_{jk,lm,uv}g(\cdot,\theta)$ is uniformly bounded and that

$$\mathbb{E}_{\mathcal{F}}\|\hat{c}^{*}_{i\Delta_n} - \hat{c}_{i\Delta_n}\| = O_p(k_n^{-\frac{1}{2}}),$$

we deduce that by Assumption 3.2.2

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\|R''_{2,n}(\theta)\|\right) \leq \Delta_n^{-1/2}k_n^{-3/2} \to 0.$$

So we prove that $R''_{2,n}(\theta)$ is stochastic equicontinuous.

As for $R^{'}_{2,n}(\theta)$,

$$
\begin{aligned}
R'_{2,n}(\theta) &= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\frac{1}{2}\Bigg(\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)(\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\quad -\frac{1}{k_n}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg) \\
&= k_n\sqrt{\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\frac{1}{2}\sum_{j,k,l,m=1}^{d}\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)\Bigg((\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\quad -\frac{1}{k_n}\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg)
\end{aligned}
$$

Now for any other $\theta'\in\Theta$, we have for any $p>2$,

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{F}}\|R'_{2,n}(\theta)-R'_{2,n}(\theta')\|^p \\
&\leq Kk_n^p\Delta_n^{p/2}\mathbb{E}_{\mathcal{F}}\Bigg\|\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k,l,m=1}^{d}\left(\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta)-\partial^2_{jk,lm}g(\hat{c}_{i\Delta_n};\theta')\right)\times \\
&\qquad \left((\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n})-\frac{1}{k_n}\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\right)\Bigg\|^p \\
&\leq Kk_n^p\Delta_n^{p/2}\|\theta-\theta'\|^p\mathbb{E}_{\mathcal{F}}\Bigg\|\sum_{i=0}^{[T/k_n\Delta_n]-1}\sum_{j,k,l,m=1}^{d}\Bigg((\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\qquad -\frac{1}{k_n}\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg)\Bigg\|^p \\
&\leq Kk_n^p\Delta_n^{p/2}\|\theta-\theta'\|^p\sum_{j,k,l,m=1}^{d}\mathbb{E}_{\mathcal{F}}\Bigg\|\sum_{i=0}^{[T/k_n\Delta_n]-1}\Bigg((\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\qquad -\frac{1}{k_n}\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg)\Bigg\|^p \\
&\leq Kk_n^p\Delta_n^{p/2}\|\theta-\theta'\|^p\sum_{j,k,l,m=1}^{d}\mathbb{E}_{\mathcal{F}}\Bigg\|\sum_{i=0}^{[T/k_n\Delta_n]-1}\Bigg((\hat{c}^{*jk}_{i\Delta_n}-\hat{c}^{jk}_{i\Delta_n})(\hat{c}^{*lm}_{i\Delta_n}-\hat{c}^{lm}_{i\Delta_n}) \\
&\qquad -\frac{1}{k_n}\left(\hat{c}^{jl}_{i\Delta_n}\hat{c}^{km}_{i\Delta_n}+\hat{c}^{jm}_{i\Delta_n}\hat{c}^{kl}_{i\Delta_n}\right)\Bigg)^2\Bigg\|^{p/2} \\
&\leq Kk_n^p\Delta_n^{p/2}\|\theta-\theta'\|^p(k_n\Delta_n)^{-p/2}\frac{1}{k_n\Delta_n}\sum_{i=0}^{[T/k_n\Delta_n]-1}\mathbb{E}_{\mathcal{F}}\left\|(\hat{c}^{*}_{i\Delta_n}-\hat{c}_{i\Delta_n})^2-\frac{1}{k_n}\right\|^p
\end{aligned}
$$

where we have in order used the continuous differentiability of $\partial^2 g$ with respect to θ and fact that $\partial^3 g$ is uniformly bounded, Jensen's inequality, Burkholder-Davis-Gundy inequality (in view of Lemma 4.6.2) and again, Jensen's inequality. Then by Lemma 5.5.1, we deduce that

$$\mathbb{E}\|R'_{2,n}(\theta) - R'_{2,n}(\theta')\|^p \leq K k_n^{-p/2} \|\theta - \theta'\|^p$$

and hence $R'_{2,n}(\theta)$ is stochastically equicontinuous.

Thus we prove the stochastic equicontinuity of $R_{2,n}$ and that of $\Delta_n^{-\frac{1}{2}} \left(S_n(g; \cdot; \mathcal{D}_n^*) - \widetilde{S}_n(g; \cdot; \mathcal{D}_n) \right)$.

5.5.3 Proof of Theorem 5.3.2

The proof of Theorem 5.3.2 can be produced word for word as in the parametric case. The only difference is that we should use Lemma 4.6.2 instead of the results given in Step 1 of the proof of Theorem 4.2.1.

CHAPTER 6

Euler Method with Estimated Volatility

6.1 Motivation

In the field of financial econometrics, there is always need to simulate the following diffusion process

$$dX_t = b_t dt + \sigma_t dW_t,$$

where W is Brownian motion. Very often X denotes the prices of financial assets, say stocks, and σ is referred to as the volatility process related to X. The most commonly used method to simulate X is the so-called Euler-Maruyama approximation, which is named after Leonhard Euler and Gisiro Maruyama, and is actually a simple generalization of the Euler method for ordinary differential equations to stochastic differential equations. More precisely, to obtain the value of X at terminal time T over a fixed time span $[0.T]$, one uses the following recursive equation:

$$X_{\tau_{n+1}} = X_{\tau_n} + b_{\tau_n}(\tau_{n+1} - \tau_n) + \sigma_{\tau_n}(W_{\tau_{n+1}} - W_{\tau_n}),$$

with given discretization grid $0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$. Usually, the equidistant discretization scheme is used, namely, $\tau_{(i+1)} - \tau_i = \delta$ for some time step $0 < \delta < T$. For a thorough treatment on Euler-Maruyama approximation and its extensions, see (Kloeden and Platen, 1992).

However, to implement such procedure, the values of $(b_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$ have to be prespecified (or simulated) beforehand. For instance, consider a version of Heston model

(Heston, 1993) given as

$$\begin{aligned} dX_t &= (\mu - c_t^{1/2})dt + c_t^{1/2} dW_t \\ dc_t &= \kappa(\alpha - c_t)dt + \gamma c_t^{1/2} dW'_t \end{aligned}$$

where W and W' are two Brownian motion with possible dependence, and parameters μ, α, κ and γ are constants. It is an important example of stochastic volatility models, and to simulate such a system, one has to specify the values for μ, α, κ and γ. If the researcher is only to verify certain stochastic theory via Monte Carol simulation (e.g., finite sample performance of asymptotic theory in a statistical setting), prespecifying the values is totally fine and is actually necessary; however, on the other hand, one has to keep in mind that the values of X generated in this way might not replicate the true world as much as possible, even if the specified values for parameters are claimed to be "calibrated to the real world".

Alternatively, instead of specifying particular dynamics for $(\sigma_t)_{t\geq 0}$, we can use estimated spot volatility based on high-frequency data in the Euler method. In other words, we would like to design a data generating mechanism, via Euler method, to regenerate data that mimics the real world more realistically, by taking advantage of the information contained in the observed real data. As seen below, "mimic the real world" is in the sense that the probability distribution of the surrogate data generated by our Euler method with estimated spot volatility is uniformly approximating to that of the true data, under certain assumptions. In particular, we need the notion of Wasserstein metric to measure the distance between two distributions. With such properties, the data produced in this way can be used to do other empirical studies including evaluating the accuracy of estimation of diffusive beta as discussed in Section 6.5.

We note here that our Euler method with estimated spot volatility has a similar spirit with resampling techniques in statistics, such as bootstrap, which also produces new data by extracting information from the real data. In fact, the application we provide in Section 6.5 demonstrates how we can combine the Euler method with estimated spot volatility with parametric bootstrap.

This chapter is organized as follows. Section 6.2 gives the formal set-up and assumptions. Section 6.3 constructs the Euler approximation with estimated spot volatility. Section 6.4 gives the main theoretical results. Section 6.5 gives an application of the Euler method with estimated spot volatility, which is closely related to bootstrap. Section 6.6 concludes and offers the directions for future research.

6.2 Setting

6.2.1 Product Space

We consider two filtered probability spaces, $(\Omega_1, \mathcal{F}_1, \{\mathcal{F}_t^1\}, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \{\mathcal{F}_t^1\}, \mathbb{P}_2)$. Let $\{W_t, t \geq 0\}$ and $\{\widetilde{W}_t, t \geq 0\}$ be two Brownian Motions defined on $(\Omega_1, \mathcal{F}_1, \{\mathcal{F}_t^1\}, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \{\mathcal{F}_t^1\}, \mathbb{P}_2)$, respectively. In the context of high-frequency econometrics, one may think of the first space as the original probability space on which the stock price processes live, and the second space as the space for simulation (or equivalently, the space for random number generator of MATLAB).

Now we take the product space, i.e., let

$$\Omega = \Omega_1 \times \Omega_2, \qquad \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$$

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_t^1 \otimes \mathcal{F}_t^2, \quad \forall t \geq 0, \quad \mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2.$$

Without loss of generality, we assume $\{\mathcal{F}_t\}$ are complete (otherwise, we can make it complete by standard augmentation procedure). Under this construction, any random object defined on either $(\Omega_1, \mathcal{F}_1, \{\mathcal{F}_t^1\}, \mathbb{P})$ or $(\Omega_2, \mathcal{F}_2, \{\mathcal{F}_t^2\}, \mathbb{P}_2)$ can be naturally extended to the new space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}_1)$, and in particular, $W'(\omega_1, \omega_2) = W(\omega_1)$ and $\widetilde{W}'(\omega_1, \omega_2) = \widetilde{W}(\omega_2)$ are still Brownian Motions with respect to $\{\mathcal{F}_t\}$. For notational simplicity, we continue using W and $\widetilde{W}$ to denote them. At last, by construction of the new space, W and $\widetilde{W}$ are independent from each other on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

6.2.2 Basic models: no jump or leverage effect

For simplicity, throughout we will consider one-dimensional processes, and we conjecture that all the results can be extended to multivariate case by polarization. Let P_t be the raw stock price for a given stock, and define

$$X_t = 100 \times \log P_t, \quad X_0 = x_0,$$

with x_0 being known. Without loss of generality, we will assume in the following $x_0 = 0$. Our basic model is to assume the dynamics of $\{X_t : t \geq 0\}$ follows the following one-dimensional stochastic differential equation (SDE)

$$\begin{aligned} dX_t &= \sqrt{c_t} dW_t \\ X_0 &= 0 \end{aligned} \tag{6.1}$$

where $0 \leq t \leq T$ with T being the terminal time, c is the variance process and W is one-dimensional Brownian Motion introduced above. In particular, X has neither drift part nor jump part.

Note that we assume the drift term in the log-price process X_t to be zero. It is of no restriction since we are considering a high-frequency setting in which the drift term is almost negligible. In reality, there should be a drift in the stock price, but is only visible if we consider a long time span (in years). This is the reason for assuming a log-price dynamics without drift from a economic point of view. Mathematically, (Mykland and Zhang, 2009) gives an argument to remove the drift part in X which involves change of probability measure by using Girsanov theorem.

On the other hand, we can without loss of generality assume there is no jump in X because if otherwise there are jumps in X, we can just truncate the jumps off from the real date and then put them back to the simulate data. Such a step has nothing to do with the Euler method we are going to develop.

We collect all the assumptions imposed on $\{c_t : t \geq 0\}$ as follows.

Assumption 6.2.1. *There is no leverage effect, i.e., the volatility process c_t is independent of the driving Brownian motion W of the stock price dynamics.*

The reason for assuming no leverage effect is two-folded. Firstly, on a technical level, no leverage assumption enables us to use "conditioning on $(c_t)_{t\geq 0}$ argument", see e.g., (Barndorff-Nielsen and Shephard, 2003) and (Barndorff-Nielsen and Shephard, 2004a). More importantly, as will be seen soon, the Euler approximation considered in Section 6.3 can only simulate data without leverage, which makes such an assumption necessary.

Assumption 6.2.2. $\{c_t : t \geq 0\}$ *follows a continuous one-dimensional Itô semimartingale:*

$$c_t = c_0 + \int_o^t b_s^{(c)} ds + \int_0^t \sigma_s^{(c)} dW'_s,$$

where $b^{(c)}$ is locally bounded, $\sigma^{(c)}$ is càdlàg and W' is another Brownian motion living in the original probability space. In the spirit of Assumption 6.2.1, we assume for the moment that $b^{(c)}, \sigma^{(c)}$ and W' are all independent of W, which drives the log-stock price (6.1).

Assumption 6.2.3. $\{c_t : t \geq 0\}$ *is uniformly bounded ($\mathbb{P}^1$-almost surely no matter which realization is) both from infinity and zero, i.e.*

$$\begin{aligned} C &:= \sup_{0\leq t\leq T,\ \mathbb{P}^1-a.s.\omega_1\in\Omega_1} c(t,\omega_1) < \infty, \\ \epsilon_c &:= \mathop{inf}_{0\leq t\leq T,\ \mathbb{P}^1-a.s.\omega_1\in\Omega_1} c(t,\omega_1) > 0. \end{aligned}$$

Moreover, $b^{(c)}$ and $\sigma^{(c)}$ are bounded.

Mathematically, Assumption 6.2.3 can be justified by a standard localization procedure as Lemma 4.4.9 in (Jacod and Protter, 2012). But from financial point of view, in reality the volatility for stocks would just never be infinity or touch zero.

In many situations we may use a weaker condition relative to Assumption 6.2.2, in which we only require the pathwise continuity of $\{c_t : t \geq 0\}$ to certain extent.

Assumption 6.2.4. *The process $\{c_t : t \geq 0\}$ is a one-dimensional adapted stochastic process with sample paths being locally Hölder-ρ continuous for some index $\rho \in (0, 1]$, i.e.,* *for almost surely* $\omega_1 \in \Omega_1$, $\forall s, t \in [0, T]$,

$$|c(t, \omega_1) - c(s, \omega_1)| \leq A(\omega_1)|t - s|^\rho,$$

for some constant $A(\omega_1)$.

Remark 6.2.1. We do not consider $\rho = 0$, in which case the volatility path is only bounded but may not be continuous.

Remark 6.2.2. Recall that the paths for a general Itô semimartingale are locally $\rho'-$Hölder for all $\rho' < \frac{1}{2}$ (e.g., Brownian motion). So Assumption 6.2.2 implies Assumption 6.2.4. In fact, (2.1.43) in (Jacod and Protter, 2012) for estimate for Itô semimartingale (ignoring the jump part) implies that $\forall s, t \in [0, T], t > s$

$$\mathbb{E}\left(|c_t - c_s|^2\right) \leq K\mathbb{E}\left(\left(\int_s^t |b_u^{(c)}|du\right)^2 + \int_s^t |\sigma_u^{(c)}|^2 du\right) \leq K|t - s|,$$

where the second inequality follows from the assumption that $b^{(c)}$ and $\sigma^{(c)}$ are bounded in Assumption 6.2.3.

Lastly, based on all assumptions above, c_t is trivially progressively measurable, and hence the stochastic integral in (6.1) is well-defined.

We end this section with the statistical side of our context. Equidistant observations of X are sampled over a fixed finite time span $[0, T]$ with time step Δ_n at stage n. For each $i = 1, 2, \ldots, \lfloor \frac{T}{\Delta_n} \rfloor$ and any stochastic process Z, define

$$\Delta_i^n Z = Z_{i\Delta_n} - Z_{(i-1)\Delta_n}$$

to be the increment of Z over time increment $[(i-1)\Delta_n, i\Delta_n]$. We will consider infill asymptotics for which $\Delta_n \to 0$ as $n \to \infty$.

6.3 Euler Approximation

In this section we construct our Euler approximation with estimated spot volatility through two steps. In the first we nonparametrically recover the spot volatility from discrete samples of X over a given discretization grid of $[0,T]$ by using a local average of sum of squared returns. Then we plug the spot volatility estimators into the Euler scheme to generate the simulated data.

Consider equally spaced time discretization grid for Euler approximation, i.e., for some $\delta > 0$, let

$$\tau_0 = 0, \quad \tau_i = i\delta,$$

where $i \in \{0, 1, \ldots, \lfloor \frac{T}{\delta} \rfloor\}$. At each discretization time point $i\delta$, the spot volatility estimation is given by

$$\hat{c}_{i\delta} = \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2,$$

where the integer-valued k_n is the length of window we use to estimate spot volatility satisfying

$$k_n \to \infty \quad \text{and} \quad k_n \Delta_n \to 0.$$

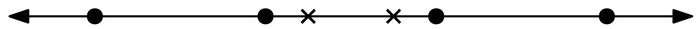
The conditions imposed on the tuning parameter k_n is quite common in nonparametric statistics, which represents a trade-off between bias and variance. In fact, we have

$$\begin{aligned}
\hat{c}_{i\delta} - c_{i\delta} &= \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - c_{i\delta} \\
&= \underbrace{\frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - \frac{1}{k_n \Delta_n} \int_{\lceil \frac{i\delta}{\Delta_n} \rceil \Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + k_n)\Delta_n} c_s ds}_{S_n\text{: statistical error}} \\
&\quad \underbrace{+ \frac{1}{k_n \Delta_n} \int_{\lceil \frac{i\delta}{\Delta_n} \rceil \Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + k_n)\Delta_n} c_s ds - c_{i\delta}}_{D_n\text{: target error}}.
\end{aligned}$$

The two errors S_n and D_n compete with rates $\sqrt{k_n}$ and $\frac{1}{(k_n \Delta_n)^\rho}$, respectively. More precisely, let

$$k_n^{\rho + \frac{1}{2}} \Delta_n^\rho \to \beta,$$

Figure 6.1: Sampling and Discretization Grid when $\delta < \Delta_n$



(a) When $\delta < \Delta_n$, there are two distinct discretization points, $i\delta$ and $(i+1)\delta$ (represented by crosses), between two data points sampled (represented by disks).

then if $\beta = 0$, S_n dominates; if $\beta \in (0, \infty)$, S_n and D_n are of the same magnitude ; if $\beta = \infty$, D_n dominates. The maximum rate is achieved in the second case with the rate being $\Delta_n^{1/4}$, implying that the convergence rate for spot volatility estimation is (much) lower than that of integrated volatility functions (which is $\Delta_n^{1/2}$).

We note that such an estimator for spot volatility estimation dates back to (Foster and Nelson, 1996) under the name rolling volatility estimators; see also (Jacod and Protter, 2012) Chapter 9 and 13 for a detailed treatment of asymptotics.

Then for fixed n and δ, the global Euler approximation with estimated spot volatility is given by the process:

$$Y_t^{n,\delta} = \sum_{i=0}^{[t/\delta]-1} \sqrt{\hat{c}_{i\delta}} (\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}), \quad 0 \le t \le T,$$

where $\widetilde{W}$ is Brownian motion on the simulation space $(\Omega_2, \mathcal{F}_2, \{\mathcal{F}_t^2\}, \mathbb{P}_2)$, which is independent of everything defined on $(\Omega_1, \mathcal{F}_1, \{\mathcal{F}_t^1\}, \mathbb{P}_1)$. In particular, $Y^{\delta,n}$ depends on both discretization grid δ and sampling grid Δ_n.

For $Y^{n,\delta}$ to be well-defined, the condition $\delta > \Delta_n$ is required. One may intuitively understand this from the point of view of non-identifiability: once Δ_n is given, the number of observed data and hence the information provided by the data is given. Since we are trying to mimic the reality from the information provided by the data via simulation, what we can obtain from simulation at most would be no more than how much the data tell us. In other words, the dimension of unknown local volatilities to be recovered (estimated) should not exceed the dimension of known information (sampled data). Mathematically, for given Δ_n (and k_n), there will be more than one discretization points between two data points as soon as δ is small enough. For example, in Figure 6.1, there are two distinct discretization points, $i\delta$ and $(i+1)\delta$ (represented by crosses), between two data points sampled (represented by disks).

Then obviously $\lceil \frac{i\delta}{\Delta_n} \rceil = \lceil \frac{(i+1)\delta}{\Delta_n} \rceil$ and hence

$$\hat{c}_{i\delta} = \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 = \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{(i+1)\delta}{\Delta_n} \rceil + \ell} X \right)^2 = \hat{c}_{(i+1)\delta},$$

since they use the same k_n data points which are right after to $i\delta$. This fact suggests that an arbitrarily small δ (relative to a given Δ_n) may not always improve Y_T.

6.4 Main Results

In this section we present the theoretical results associated with the $Y^{n,\delta}$ constructed above. The very first thing one should notice is that since in simulation only $\widetilde{W}$ is available, $Y^{n,\delta}$ is a "consistent estimator" for the simulated log-price defined by

$$\widetilde{X}_t = \int_0^t \sqrt{c_s} d\widetilde{W}_s, \quad 0 \leq t \leq T,$$

rather than the true price observed process X_t. The appearance of such a $\widetilde{X}$ may look strange at the very beginning, but it actually becomes natural once we recall that our goal is to reproduce the probability distribution of X, of which we think as a random variable taking values in the Polish space $\mathbb{C}([0,T] : \mathbb{R})$. Under the assumption of no leverage, it follows immediately that X and $\widetilde{X}$, which is also a random variable taking values $\mathbb{C}([0,T] : \mathbb{R})$, have the same probability distribution. Hence $(Y^{n,\delta})$ approximating $\widetilde{X}$ is equivalent to $(Y^{n,\delta})$ approximating the distribution of X. Mathematically, such an idea can be realized by using Wasserstein metric, we will discuss this issue at the end of this section.

The second question to ask would be how well the data Y^n generated by the Euler method with estimated volatility approximate $\widetilde{X}$. In light of Section 9.7 in (Kloeden and Platen, 1992), Y^n *weakly converges* (which is actually convergence of moments, and different from the usual synonym for convergence in distribution) to $\widetilde{X}$ as

$$\begin{aligned}\mathbb{E}(\widetilde{X}_t) - \mathbb{E}(Y_t^{\delta,n}) &= \mathbb{E}\left(\mathbb{E}_c\left(\int_0^t \sqrt{c_s}d\widetilde{W}_s\right)\right) - \mathbb{E}\left(\mathbb{E}_c\left(\sum_{i=0}^{\lfloor T/\delta\rfloor-1} \sqrt{\hat{c}_{i\delta}}\Delta_{i+1}\widetilde{W}\right)\right)\\ &= 0 - \mathbb{E}\left(\sum_{i=0}^{\lfloor t/\delta\rfloor-1} \mathbb{E}_c\left(\sqrt{\hat{c}_{i\delta}}\Delta_{i+1}\widetilde{W}\right)\right)\\ &= 0, \qquad 0 \le t \le T,\end{aligned}$$

where we have used the facts that conditioning on c,

$$\int_0^t \sqrt{c_s}d\widetilde{W}_s \sim \mathcal{N}\left(0, \int_0^t c_s ds\right),$$

and that $\hat{c}$, which only depends on the random objects living in the original probability space, is independent of $\widetilde{W}$.

Apparently, convergence of moments alone is not enough for us. More interestingly, we focus our attention on

$$\sup_{0\le t\le T} |Y_t^{n,\delta} - \widetilde{X}_t|, \tag{6.2}$$

which globally measures the distance between $Y^{n,\delta}$ and $\widetilde{X}$. Then a natural question is whether or not the quantity (6.2) tends to 0 as $n \to \infty$ and $\delta_n \to 0$, in which case Y^n is called *strongly convergent* to $\widetilde{X}$ uniformly over $[0, T]$ as in Section 9.6 of (Kloeden and Platen, 1992). If it is, what should the optimal simulation scheme be? Namely, given Δ_n, what are the best choices of δ_n and k_n such that the whole path of $Y^{\delta,n}$ approximates the path of $\widetilde{X}$ as fast as possible? These questions will be answered step by step in Section 6.4.1.

6.4.1 Optimal simulation scheme and rate of convergence

Theorem 6.4.1. *Suppose Assumptions 6.2.1 and 6.2.3, assume further that $\{c_t : t \ge 0\}$ has sample paths satisfying for any $t > s > 0$,*

$$\mathbb{E}|c_t - c_s|^2 \le K|t-s|^{2\rho}, \quad 0 < \rho \le 1, \tag{6.3}$$

for some constant K. Then it holds for any fixed discretization distance $\delta \in [\Delta_n, T)$ that

$$\mathbb{E}\left(\sup_{0\le t\le T}|Y_t^{n,\delta}-\widetilde{X}_t|\right)\le K\left(\frac{1}{\sqrt{k_n}}+(k_n\Delta_n)^\rho+\delta^\rho+\left(\delta\log\left(\frac{2T}{\delta}\right)\right)^{\frac{1}{2}}\right)$$

for some constant K.

Note that (6.3) is weaker than pathwise Holder-ρ continuity, and if $(c_t)_{t\ge 0}$ is an continuous Itô semimartingale, $\rho = \frac{1}{2}$.

Theorem 6.4.1 immediately implies the "consistency" of $Y^{n,\delta}$.

Corollary 6.4.1. *Assume that $k_n \to \infty$ and $k_n\Delta_n \to 0$, then*

$$\begin{aligned}\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}\left(\sup_{0\le t\le T}|Y_t^{n,\delta}-\widetilde{X}_t|\right) &= 0,\\ \lim_{n\to\infty}\limsup_{\delta\to 0}\mathbb{E}\left(\sup_{0\le t\le T}|Y_t^{n,\delta}-\widetilde{X}_t|\right) &= 0.\end{aligned}$$

More importantly, (6.4) provides a concrete structure of the convergence rate (upper bound): the error arising from local volatility estimation (including the statistical error $\frac{1}{\sqrt{k_n}}$ and target error $(k_n\Delta_n)^\rho$), the discretization error δ^ρ from Euler scheme approximation, and the residual error $\left(\delta\log\left(\frac{2T}{\delta}\right)\right)^{\frac{1}{2}}$ are separately additive, from which we are able to derive the "optimal" simulation scheme in the sense of fastest convergence rate: to make $Y^{n,\delta}$ converges to $\widetilde{X}$ as fast as possible, one should first take $\delta_n \to 0$ as small as possible, i.e.

$$\delta_n = \Delta_n,$$

which means taking each data sampling point as a discretization point; then we strike balance between statistical error and target error, by requiring

$$k_n^{\rho+\frac{1}{2}}\Delta_n^\rho \to \beta \in (0,\infty),$$

equivalently $k_n \sim \Delta_n^{-\frac{\rho}{\rho+\frac{1}{2}}}$. In this fashion, our "optimal" Euler approximation becomes:

$$Y_t^n = \sum_{i=0}^{\lfloor t/\Delta_n\rfloor-1}\sqrt{\hat{c}_{i\Delta_n}}\Delta_i^n\widetilde{W}, \quad 0\le t\le T, \tag{6.4}$$

with convergence rate

$$\begin{aligned}\mathbb{E}\left(\sup_{0\le t\le T}|Y_t^n-\widetilde{X}_t|\right) &\le K\left(\frac{1}{\sqrt{k_n}}+\left(k_n\Delta_n\right)^{\rho}+\Delta_n^{\rho}+\left(\Delta_n\log\left(\frac{2T}{\Delta_n}\right)\right)^{\frac{1}{2}}\right)\\ &\sim \Delta_n^{\frac{1}{2+\frac{1}{\rho}}}.\end{aligned}$$

We need to emphasize that there is an end-effect issue in the definition of Y^n: when $i > \lfloor t/\Delta_n\rfloor - k_n$, there will be no enough k_n returns from the right side to give $\hat{c}_{i\Delta_n}$. To overcome this, we assume that we can sample data up to time $T+h$ for some $h>0$. Asymptotically, as $n\to\infty, k_n\Delta_n < h$ and Y^n is well-defined.

Now we know that under the "optimal" simulation scheme, the upper bound of convergence rate for Y^n approximating $\widetilde{X}$ is $\Delta_n^{\frac{1}{2+\frac{1}{\rho}}}$. Then the next natural question is whether this rate is sharp. The answer is, not surprisingly, affirmative. To prove this, we start with the approximation $|Y_T^n-\widetilde{X}_T|$ at terminal time T.

We state the result in the following theorem. We say a sequence of random variables X_n converges to 0 at rate a_n if both $\frac{X_n}{a_n}$ and $\frac{a_n}{X_n}$ are tight.

Theorem 6.4.2. *Under Assumptions 6.2.1,6.2.3 and 6.2.4, assume further that* $\delta_n=\Delta_n$ *and that*

$$\Delta_n\to 0,\quad k_n^{\rho+\frac{1}{2}}\Delta_n^{\rho}\to\beta\in(0,\infty)\quad n\to\infty.$$

Conditioning on $\{c_t : t\ge 0\}$*, the exact convergence rate for* $\sup_{0\le t\le T}|Y_t^n-\widetilde{X}_t|$ *is* $\Delta_n^{\frac{\rho}{2\rho+1}}$.

6.4.2 Special case: constant volatility

Up to now, we have only considered ρ-Hölder continuous for $0<\rho\le 1$. It is also interesting to examine the case when ρ-Hölder continuous for $\rho>1$, i.e., $(c_t)_{t\ge 0}=c$ for some constant c, in which case

$$\widetilde{X}_t=\sqrt{c}\widetilde{W}_t.$$

Then the question is: should we still take a discretization scheme to track the volatility path? Intuitively the answer should be negative as using spot volatility estimates to track a constant volatility path would invoke extra variance. More rigorously, we have the following result.

Proposition 6.4.1. *Assume $c_t \equiv c$, where $c > 0$ is constant. Then for any fixed $\Delta_n < \delta < T$,*

$$\mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{n,\delta} - \widetilde{X}_t|\right) \le K\left(\frac{1}{\sqrt{k_n}} + \left(\delta \log(\frac{2T}{\delta})\right)^{\frac{1}{2}}\right)$$

for some positive constants K_1 and K_2.

Here the approximation error decomposes into a statistical error part $\frac{1}{\sqrt{k_n}}$, and a residual part $\left(\delta\log(\frac{2T}{\delta})\right)^{\frac{1}{2}}$. Similarly as above, to maximize the convergence rate, one should first take δ_n as small as possible, i.e., $\delta_n = \Delta_n$; then take k_n as big as possible, i.e. $k_n = \lfloor \frac{T}{\Delta_n} \rfloor$. In fact, $k_n = \lfloor \frac{T}{\Delta_n} \rfloor$ corresponds to using realized variance to estimate local volatility.

So, the optimal Euler approximation in the case of constant volatility is given by

$$Y_t^{RV} \equiv \sqrt{\frac{RV}{T}}\widetilde{W}_t, \quad 0 \le t \le T$$

where $RV \equiv \sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n X)^2$ is the realized variance. In particular, there exists no residual error from such formulation (otherwise, the residual error would have the dominating rate $(\Delta_n \log(2T/\Delta_n))^{\frac{1}{2}}$).

Theorem 6.4.3. *Assume $c_t \equiv c$, where $c > 0$ is constant, then it holds that*

$$\mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{RV} - \widetilde{X}_t|\right) \le K\sqrt{\Delta_n},$$

for some constant K.

To demonstrate whether $\sqrt{\Delta_n}$ is the exact uniform convergence rate for $\sup_{0\le t\le T} |Y_t^{RV} - \widetilde{X}_t|$, we can use the exact rate for approximating $\widetilde{X}_T$ at terminal time T.

Theorem 6.4.4. *Assume $c_t \equiv c$, where $c > 0$ is constant. Then it holds that*

$$\frac{1}{\sqrt{\Delta_n}}(Y_T^{RV} - \widetilde{X}_T) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{c}{2})\mathcal{N}(0,1),$$

where the two normal random variables on the right hand side are independent.

Theorem 6.4.4, together with uniform integrability and Theorem 6.4.3 imply that

$$K_1\sqrt{\Delta_n} \le \mathbb{E}\left(|Y_T^{RV} - \widetilde{X}_T|\right) \le \mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{RV} - \widetilde{X}_t|\right) \le K_2\sqrt{\Delta_n},$$

and hence we conclude that the exact uniform convergence rate for $\sup_{0\le t\le T} |Y_t^{RV} - \widetilde{X}_t|$ is $\sqrt{\Delta_n}$.

Important remark: In fact there is a generalization of Theorem 6.4.4: even if c_t is not a constant over $[0,T]$, we still have a CLT associated with Y_T^n. More precisely, define the integrated volatility $IV \equiv \int_0^T c_t dt$, then by a standard result on RV approximating IV (e.g., (Aït-Sahalia and Jacod, 2014), p.89-90), we have

$$\frac{1}{\sqrt{\Delta_n}}(RV - IV) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 2\int_0^T c_t^2 dt\right),$$

conditionally on $(c_t)_{0\le\le T}$. Then the delta method gives

$$\frac{1}{\sqrt{\Delta_n}}\left(Y_T^{RV} - \sqrt{\frac{IV}{T}}\widetilde{W}_T\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\int_0^T c_t^2 dt}{2\int_0^T c_t dt}\right),$$

conditionally on $(c_t)_{0\le\le T}$. This implies that under the transitional kernel, Y_T^{RV} is actually a consistent "estimator" for $\sqrt{\frac{IV}{T}}\widetilde{W}_T$ with convergence rate $\sqrt{\Delta_n}$, which has the same distribution as $\widetilde{X}_T$. In light of this, if one is only interested in regenerating the distribution $\mathcal{N}(0, \int_0^T c_t dt)$ of $\widetilde{X}$ at terminal time T, Y_T^{RV} would be better than Y_T^n produced from the Euler method with estimated volatility, as the former has a faster convergence rate $\sqrt{\Delta_n}$. However, Y_T^{RV} can only be used for terminal time T while Y^n uniformly approximates the path of $\widetilde{X}$ over $[0,T]$.

6.4.3 Closing the gap: coupling and Wasserstein metric

In this subsection we revisit an issue brought before. We construct Y^n to approximate the probability distribution of X via an intermediate object $\widetilde{X}$. A mathematical way to formulate such a relation is to use Wasserstein metric.

Definition 6.1. Given metric space (Ω, d), the Wasserstein metric between two probability measures μ and ν on (Ω, d) is defined as

$$d_W(\mu,\nu) = \inf_J \left\{ \mathbb{E}(d(X,Y)) : \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu \right\},$$

where the infimum is taken over all joint distributions J with marginals μ, ν.

Intuitively, one can understand the convergence under Wasserstein metric as the usual weak convergence plus first moment convergence. We note that there exist many metrics on the space of probability measures on a given underlying space, and the choice of which metric to use depends on the specific application at hand. For a brief summary of different probability measure metrics and their relations, see (Gibbs and Su, 2002).

In our context, recall that

$$\begin{aligned}
(X_t)_{0\le t\le T} &\in \mathbb{C}([0,T]:\mathbb{R}),\\
(\widetilde{X}_t)_{0\le t\le T} &\in \mathbb{C}([0,T]:\mathbb{R}),\\
(Y_t^n)_{0\le t\le T} &\in \mathbb{D}([0,T]:\mathbb{R}).
\end{aligned}$$

Take $(\Omega, d) = (\mathbb{D}([0,T]:\mathbb{R}), d_{sk})$ where d_{sk} denotes the Skorohod metric on $\mathbb{D}([0,T]:\mathbb{R})$, it holds that

$$\begin{aligned}
d_W(Y^n, X) &\le \mathbb{E}\left(d_{sk}\left(Y^n, \widetilde{X}\right)\right)\\
&\le \mathbb{E}\left(\sup_{0\le t\le T} |Y_t^n - \widetilde{X}_t|\right) \le \Delta_n^{\frac{1}{2+\frac{1}{\rho}}}.
\end{aligned}$$

Therefore, we emphasize that even if the exact convergence rate for $\sup_{0\le t\le T} |Y_t^n - \widetilde{X}_t|$ is $\Delta_n^{\frac{1}{2+\frac{1}{\rho}}}$, we only know that the distribution of Y^n approximate that of X with rate at most $\Delta_n^{\frac{1}{2+\frac{1}{\rho}}}$, as we are not able to examine over all random objects having the same distributions as Y^n and X.

Finally, on the technical level, one can see that when working together with Y^n, $\widetilde{X}$ is much more tractable than X as the latter completely lives on the space Ω_1. In probability theory, this technique is generally called coupling: to study the convergence of distributions of two sequences of random objects, construct two copies on the same probability space and show L^1 convergence. For example, see (Asmussen, 2003) Chapter VII.

6.4.4 Summary: optimal simulation scheme

At this stage, let's give a brief summary of the "optimal" simulation scheme for the Euler method with estimated spot volatility in different contexts. In practice, we want to choose δ_n and k_n appropriately to maximize the convergence rate of Y^n to $\widetilde{X}$ uniformly over $[0,T]$. At first, we should take $\delta_n = \Delta_n$, namely take discretization points in simulation as many as that of sampling data. Notice that this is the finest discretization frequency we can take, due to an identification issue. One may argue that the smoother the volatility path is, the more sparse the discretization should be, with the concern being the possibility of introducing bias due to fine discretization when c_t is smooth enough. However, we do not need to worry about this since when taking $\delta_n = \Delta_n$, the bias term is always of order Δ_n^ρ uniformly at each discretization point $i\delta_n$. In fact, taking δ_n as such enables us to make the best of data we have: if the volatility path is smooth, then the smooth local estimates would be revealed; if the volatility path is volatile, then this feature would be captured by such finest discretization scheme.

Once δ_n is fixed, the next question would be how to choose the (local) window width k_n. The answer is that the choice of k_n depends on the knowledge of volatility.

(i) $\rho > 1$**: constant volatility** If $\rho > 1$, then c_t has constant paths. Actually one just can think of the Hölder exponent for constant as ∞. By Theorem 6.4.4, we should use RV rather than track the path by using local volatility estimation.

(ii) $\rho < \frac{1}{2}$**: Itô Semimartingale** The common belief of the dynamics of $\{c_t : t \geq 0\}$ is Itô semimartingale, with $\rho = \frac{1}{2}$. However, one should note here that for a general Itô semimartingale, it only satisfies $\frac{1}{2}$-Hölder continuity on average, i.e., for any $t > s > 0$,

$$\mathbb{E}|c_t - c_s| \leq K|t-s|^{\frac{1}{2}},$$

while its paths are ρ'-Hölder for all $\rho' < \frac{1}{2}$ (e.g., Brownian motion) and the continuity constant depends on the path under consideration. With this fact in mind, if we have no other information about the volatility path but Itô semimartingale, we can just take ρ smaller than but arbitrarily close to $\frac{1}{2}$, with choosing window width according to

$$k_n \sim \Delta_n^{-\frac{2}{2+\frac{1}{\rho}}} \to \Delta_n^{-\frac{1}{2}}, \quad \text{as } \rho \uparrow \frac{1}{2}.$$

Then convergence rate for $Y_T - \widetilde{X}_T$ approaches arbitrarily close to $\Delta^{-\frac{1}{4}}$, i.e.:

$$\lim_{\rho \uparrow \frac{1}{2}} \Delta_n^{-\frac{1}{2+\frac{1}{\rho}}} = \Delta_n^{-\frac{1}{4}},$$

but not achieved. We will discuss more about this case in Section 6.4.5.

(iii) $\frac{1}{2} \leq \rho \leq 1$**: more than Itô semartingale** If a researcher knows about the volatility path more than just being an Itô semimartingale, namely, it has paths of ρ-Hölder continuity ($\frac{1}{2} \leq \rho \leq 1$), then s/he should choose k_n according to

$$\sqrt{k_n} \sim \Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}$$

to achieve the maximum possible rate $\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}$.

On the other hand, if the true volatility path is of ρ-Hölder continuity but we do not know this information, we would still treat it as an general Itö semimartingale and choose

$$k_n \sim \Delta^{-\frac{1}{2}}.$$

Obviously the convergence rate would be slower than $\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}$.

Conclusively, the more information we have in hand, the better we design the (global) Euler approximation with estimated volatility.

6.4.5 Extension: more than Hölder continuity

So far, we have been characterizing the path regularity of volatility by ρ-Hölder continuity, as in Assumption 6.2.4. In fact, the theoretical results derived above for Euler method with estimated spot volatility can be extended to any modulus of continuity $g(\cdot)$. A function $g(\cdot)$ is called *a modulus of continuity* for the function $f : [0, T] \to \mathbb{R}$ if $0 \le s < t \le T$ and $t - s < \delta$ imply

$$|f(t) - f(s)| \le g(\delta),$$

for all sufficiently small positive δ, see e.g., (Karatzas and Shreve, 1991). Suppose the sample path of (squared) volatility process $\{c_t : t \ge 0\}$ has modulus of continuity of $g(\cdot)$. Then we could choose k_n according to

$$\sqrt{k_n} g(k_n \Delta_n) \to \beta \in (0, \infty).$$

In the special case of ρ-Hölder continuity with $\rho \in (0, 1]$, we have $g(x) = |x|^\rho$.

The reason to introduce the general modulus of continuity $g(\cdot)$ is that it is well known that almost surely sample paths of a general Itô process is of locally ρ-Hölder continuity for any $\rho < \frac{1}{2}$, but not $\rho = \frac{1}{2}$. Then we may not properly choose k_n according to the mechanism introduced above since ρ could be arbitrarily close to $\frac{1}{2}$. However, the modulus of continuity other than Hölder continuity might be obtainable in such situation.

Example 6.1(Brownian Motion). The first example would be that $\{c_t : t \ge 0\}$ is just a Browinan motion. Then we know the (uniform) modulus of continuity of Brownian path, which is derived by Lévy (1937):

$$\mathbb{P}\left(\limsup_{\delta \to 0^+} \sup_{t,s \in [0,1], |t-s| = \delta} \frac{|W_t - W_s|}{\sqrt{2\delta \log(\frac{1}{\delta})}} = 1 \right) = 1,$$

which implies that there is a finite, non-negative random variable $C(\omega)$ such that for $\mathbb{P}$-almost surely $\omega \in \Omega$,

$$\sup_{t,s \in [0,1], |t-s| = \delta} |W_t - W_s| \le C(\omega) g(\delta)$$

with

$$g(\delta) = \sqrt{\delta \log(\frac{1}{\delta})}$$

for δ sufficiently small. Note that $g(\delta) \to 0$ as $\delta \to 0$ but is faster than any $\delta^{\frac{1}{2}-\epsilon}$ for any $\epsilon > 0$. In this case, we should choose k_n such that

$$\sqrt{k_n}\sqrt{k_n \Delta_n \log(\frac{1}{k_n \Delta_n})} = k_n \Delta_n^{\frac{1}{2}} \sqrt{\log(\frac{1}{k_n \Delta_n})} \to \beta \in (0, \infty).$$

The exact convergence rate for $Y_T^n - \widetilde{X}_T$ is still $\sqrt{k_n}$ which is smaller than but arbitrarily close to $\Delta_n^{-\frac{1}{4}}$. $\diamondsuit$

Example 6.2(A class of Itô processes). For the case of $\{c_t : t \geq 0\}$ being general Itö process, we investigate the case when $\{c_t : t \geq 0\}$ satisfies Assumption 6.2.2, i.e.,

$$c_t = c_0 + \int_o^t b_s^{(c)} ds + \int_0^t \sigma_s^{(c)} dW_s'$$

with $b_t^{(c)}$ and $\sigma_t^{(c)}$ being uniformly bounded from both above and below. Since the drift term is always of higher order, we assume without loss of generality that $b_t^{(c)} \equiv 0$. Then

$$c_t = c_0 + \int_0^t \sigma_s^{(c)} dW_s'$$

is a continuous local martingale with quadratic variation

$$[c]_t = \int_0^t c_s^{(c)} ds, \quad c_s^{(c)} = (\sigma_s^{(c)})^2.$$

Note that $[c]_t$ is a strictly increasing process tending to infinity as $t \to \infty$. Then it is well-known that $\{c_t : t \geq 0\}$ can be time-changed to a Brownian motion run for all time, see, for example, Theorem 9.3 in (Chung and Williams, 1990). More specifically, for any $t \geq 0$,

$$c_t = B_{\int_0^t c_s^{(c)} ds}$$

for some Brownian motion B in $\mathbb{R}$. Then the modulus of continuity of $\{c_t : t \geq 0\}$ boils down to that of the Brownian motion: for $0 < t - s$ sufficiently small

$$\begin{aligned}
|c_t - c_s|^2 &= |B_{\int_0^t c_u^{(c)} du} - B_{\int_0^s c_u^{(c)} du}|^2 \\
&\leq \int_s^t c_u^{(c)} du \log(\frac{1}{\int_0^t c_u^{(c)} du}) \\
&\leq K_1(t-s) \log(\frac{1}{K_2(t-s)}) \\
&= K_1(\log \frac{1}{K_2})(t-s) + K_1(t-s)\log(\frac{1}{t-s})
\end{aligned}$$

where for the second last inequality we used the fact that $c^{(c)}$ is uniformly bounded from both above and below. Since the second term in the last equality dominates, we conclude that

$$|c_t - c_s| \leq \sqrt{(t-s)\log(\frac{1}{t-s})}.$$

Namely in this case the pathwise modulus of continuity of the (squared) volatility process is the same as that for Brownian motion and hence the window width k_n can be chosen accordingly as above. $\diamondsuit$

6.5 Application

Generally speaking, the way to generate Y^n via Euler method with estimated spot volatility can be viewed as parametric bootstrap in the high-frequency setting, as we use the real sampled data to estimate the spot volatility, by which in turn we can regenerate the data. However, different from the usual parametric bootstrap setting where one only needs to estimate a few parameters, the volatility process itself we are to estimate is actually nonparametric: we impose no parametric assumption on it except for certain regularity conditions.

In light of this, as far as the application is concerned, we could use the Euler method with estimated spot volatility the same way as (parametric) bootstrap, but under a high-frequency setting. One such example is to evaluate the accuracy of estimation of regression coefficient.

6.5.1 Estimation accuracy of diffusive beta

We consider the setting proposed in (Reiss et al., 2015). Taking $T = 1$ and $\Delta_n = 1/n$, we have the following bivariate Itô semimartingale

$$
\begin{aligned}
X_t &= X_0 + F_t^X + \int_0^t \sigma_s dW_s \\
Y_t &= Y_0 + F_t^Y + \int_0^t \beta_s \sigma_s dW_s + \int_0^t \sigma'_s dW'_s
\end{aligned}
$$

where F_t^X and F_t^Y are finite variation processes (including both drift part and jump part) and W and W' are two independent Brownian motions. This can be viewed as a continuous version of CAPM model in finance: X plays the role of a systematic risk factor and Y is of an individual asset. Then $\int_0^t \sigma_s dW_s$ is the diffusive part for X, $\int_0^t \sigma'_s dW'_s$ is the so-called idiosyncratic diffusive part for Y, and the process β (continuously) measures the exposure of Y to X.

In such a continuous time regression setting, the coefficient β can be identified as

$$\beta_t = \frac{d\langle X^c, Y^c\rangle_t}{d\langle X^c, X^c\rangle_t},$$

and a natural estimator for it is, after truncating jumps off from returns,

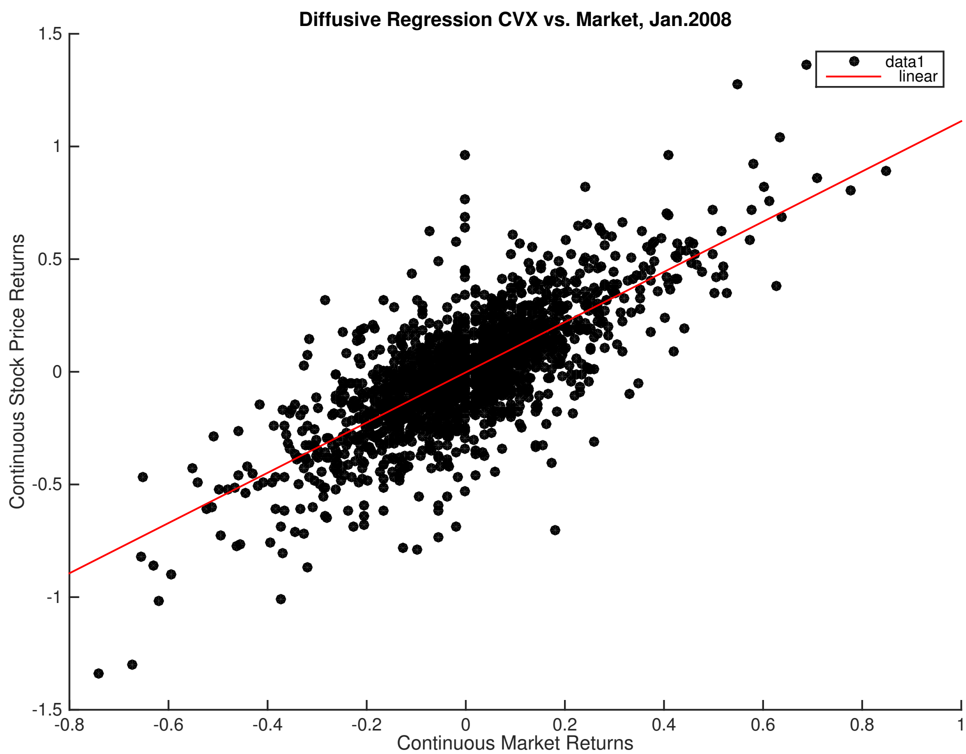
$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (\Delta_i^n Y - \beta \Delta_i^n X)^2, \tag{6.5}$$

which is analogues to the least square estimator in the usual linear regression setting.

Whether the value of β remains constant over certain time interval has been studied in finance for a long time. Economically, the constancy of β is important for justifying the asset pricing model, while statistically, as Figure 6.2 shows, there is a strong linear relationship between continuous returns of market and individual stock revealed from high-frequency data.

(Reiss et al., 2015) develops a statistical procedure to test whether β remains constant over a given time interval under high-frequency setting, and they document that for stocks like IBM, XOM and GLD, a weekly window would be a safe choice for treating market beta

Figure 6.2: Linear Relation between Market and Individual Stock



(a) The 5-min continuous returns of CVX are regressed on those of SPY, for the data sampled over Jan. 2008. The plot reveals a strong linear relation and $\widehat{\beta} = 1.0337$.

as constant in an asset pricing study. In contrast, what we are planning to do here is to use the Euler method with estimated spot volatility to evaluate the accuracy of estimator of β. There are at least two motivations for doing this: on one hand, the first step in the testing procedure proposed by (Reiss et al., 2015) is to estimate β using (6.5); on the other hand, once we accept that β is constant over a given time span (e.g., after the test), we will estimate the value of it for some other empirical use. In either case, we would like to know the accuracy of estimators.

We propose the following algorithm to evaluate the accuracy of $\widehat{\beta}$, which is pathwise dependent:

Step 1: Regress the continuous part of an individual stock on that of market to obtain $\widehat{\beta}$;

Step 2: Estimate c_X over the month, also estimate c_ϵ for $\epsilon = Y - \widehat{\beta}X$;

Step 3: Use Euler method with $\hat{c}_X$ and $\hat{c}_\epsilon$ to simulate one month of market data X^* and residual data ϵ^* respectively. Then form $Y^* = \widehat{\beta}X^* + \epsilon^*$;

Step 4: Re-estimate $\widehat{\beta}^*$ using simulated data X^* and Y^*;

Step 5: Repeat the above for M times, plot the histogram (kernel density estimation) of $\widehat{\beta}^*$.

Using the same dataset as in Figure 6.2, we obtain the sample distribution of $\widehat{\beta}^*$, see Figure 6.3a, from which we may compute the corresponding confidence interval.

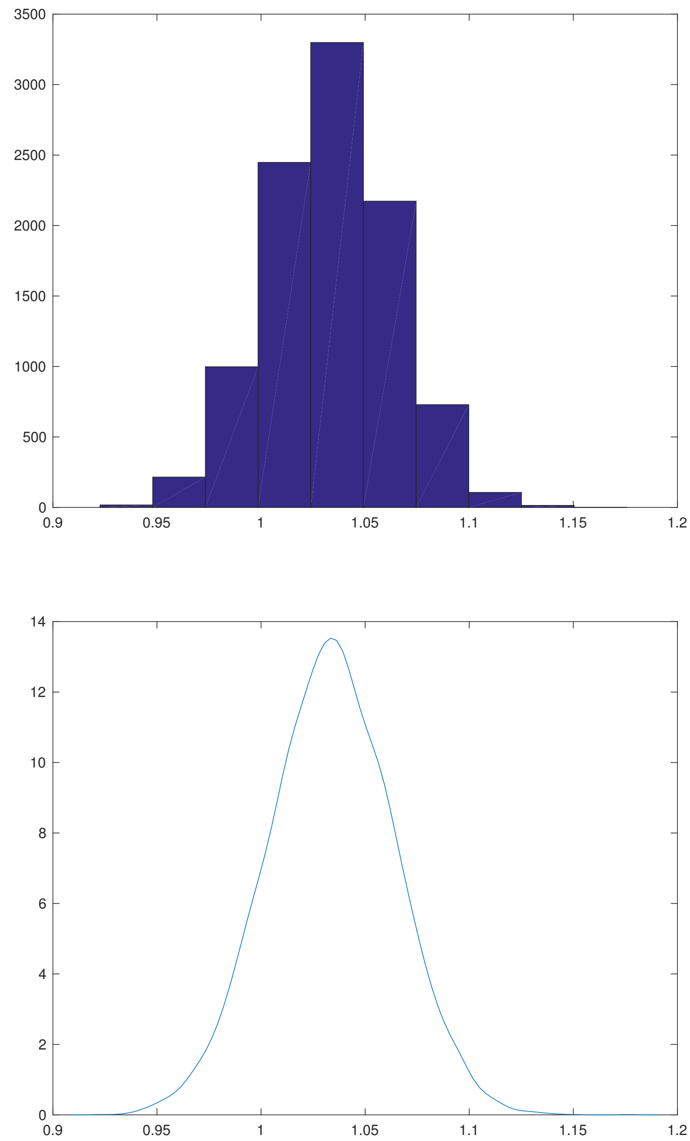
6.5.2 Parametric Bootstrap Inference for Integrated Volatility Functionals

(Liu and Li, 2016) considers (parametric) bootstrap constructing confidence intervals for the so-called integrated volatility functionals of the form

$$\int_0^T g(c_s)ds,$$

for some (nonlinear) function g over a fixed time of interval $[0, T]$. The first few steps in the algorithm to construct such confidence interval can be viewed as using the Euler method we discuss in this paper: nonparametrically estimate the spot volatility first and then generate

Figure 6.3: Sample Distribution of Diffusive Beta



(a) The histogram and kernel density of $\hat{\beta}^*$ generated by the Euler method with estimated with spot volatility. The sample standard deviation of $\widehat{\beta}^* = 0.0299$.

bootstrap samples with these spot volatility estimators plugged in. However, they do not consider the issue of convergence rate of the Euler method of estimated spot volatility but rather use it in their bootstrap inference setting.

6.5.3 Extension: resampling functionals of prices

In the above example, $\widehat{\beta}$ is a statistic that depends upon the whole paths of prices X and Y over $[0, T]$, to which our Euler method with estimated spot volatility can be applied as the convergence of the method is uniform as in Theorem 6.4.2. In principle, our method can be used to regenerate any functionals that depends on the whole sample path of observed prices.

More precisely, consider a general functional $G(\cdot)$, and we would like to obtain the sample distribution of the statistic $G(X)$. An important example could be the daily range, an easy measure of risk, that will be discussed in detail below.

Intuitively speaking, the algorithm to obtain the sample distribution of $G(X)$ will generally go as follows:

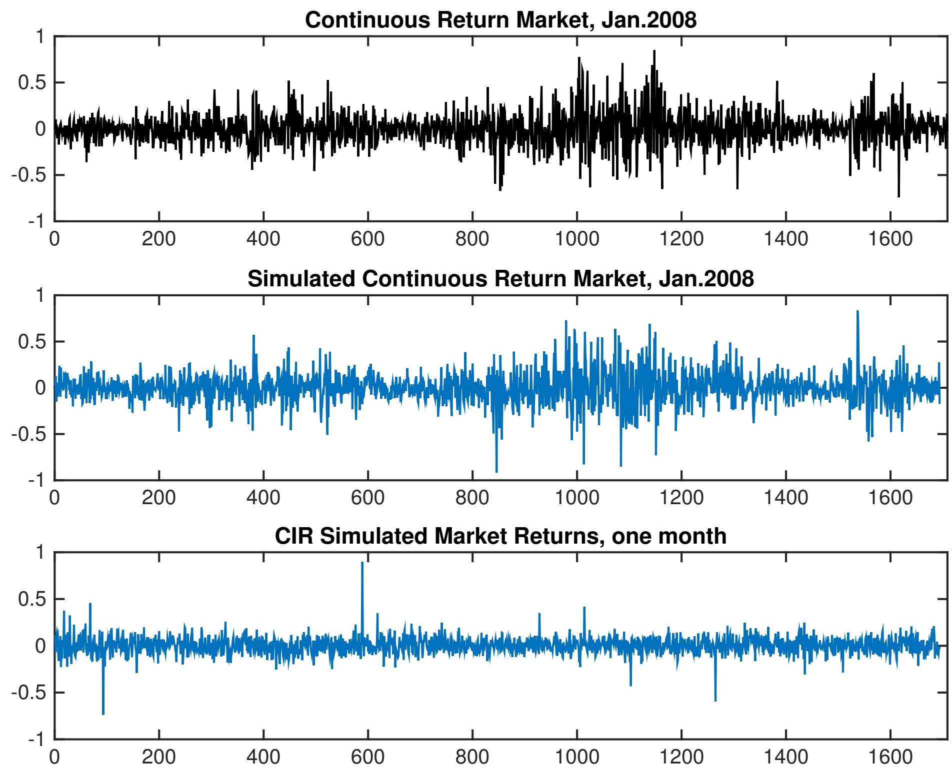
Step 1: Estimate spot volatility using high-frequency observations of price process X;

Step 2: Plug estimated spot volatility into Euler method to regenerate a whole path X^* of price process, and compute $G(X^*)$;

Step 3: Repeat Step 2 M times, and plot $G(X^{*1}), \ldots, G(X^{*M})$.

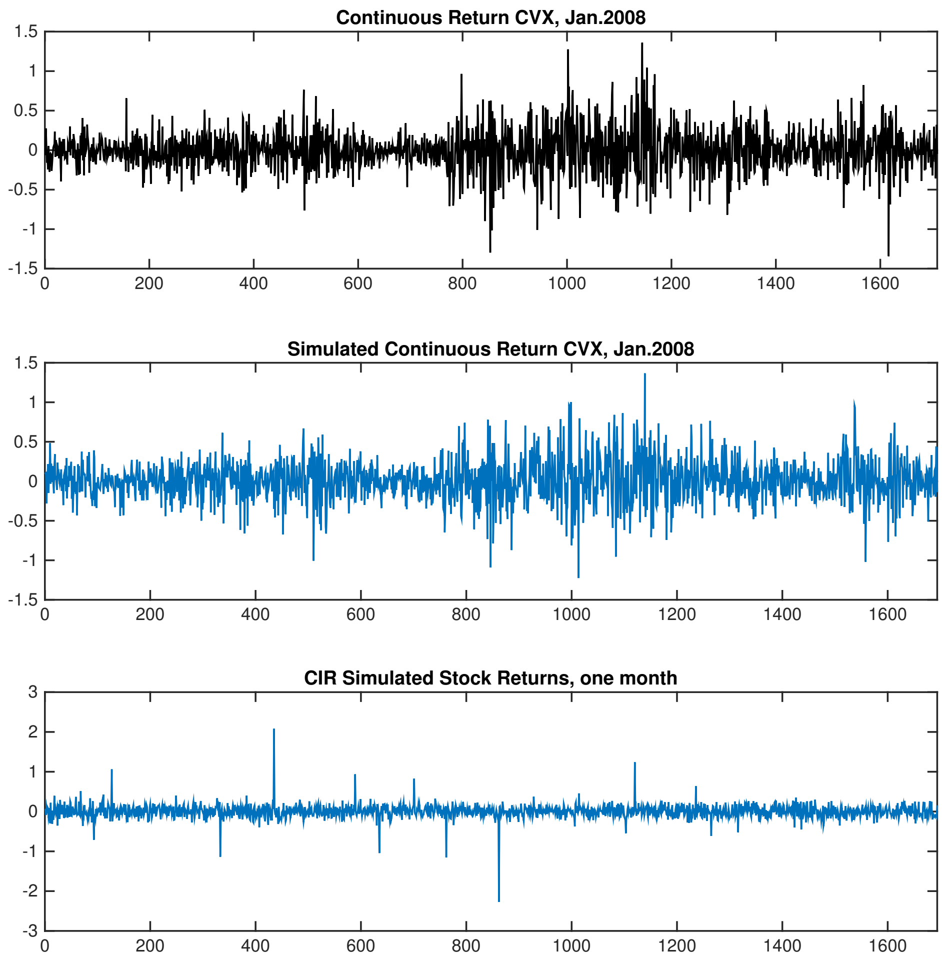
We end this section with a more intuitive illustration of how well the path of price process, produced by our Euler method with estimated spot volatility, is able to replace the true path. In Figure 6.4a and 6.5a, we employ the data used in Figure 6.2 to reproduce a sample path of returns of market and an individual stock, respectively (middle panel). In the upper panels of two figures depicted the true returns, and only for illustrative purpose, the bottom panels show paths generated according to a CIR process specified in the Monte Carlo setting of (Reiss et al., 2015) with parameters prespecified. As one can see, for both market and individual stock, the paths produced by our Euler method are very much like the true ones.

Figure 6.4: Simulated Returns for Market



(a) The top panel shows the 5-min continuous returns for SPY over Jan. 2008; the middle is the continuous returns computed from the resampled prices by using Euler method with estimated volatility; the bottom is simulated according to a CIR process for market prices dynamics with parameters specified in the Monte Carlo study of (Reiss et al., 2015).

Figure 6.5: Simulated Returns for Individual Stock



(a) The top panel shows the 5-min continuous returns for CVX over Jan. 2008; the middle is the continuous returns computed from the resampled prices by using Euler method with estimated volatility; the bottom is simulated according to a CIR process for an individual asset price with parameters specified in the Monte Carlo study of (Reiss et al., 2015).

6.6 Future Work

We can continue the study in both theory and applications:

- **Theory:** An interesting direction to generalize our Euler method with estimated volatility is to take into account the so-called leverage effect, which refers to the negative correlation between volatility and returns. Since the Brownian motion $\widetilde{W}$ used in simulation is independent of everything in the real world, to create (negative) correlation between the simulated prices and volatility, we need to use the same $\widetilde{W}$ to regenerate volatility process, which requires to model volatility process as an Itô semimartingale as well and estimate the volatility of volatility (vol. of vol.). As one may imagine, the convergence rate in this situation would be even slower than $\Delta_n^{1/4}$ as both volatility and vol. of vol. are latent.

- **Applications:** The daily range of a given price process X, defined as the difference between $\max X_t$ and $\min X_t$ within one day, had been a popular measure to quantify daily risk. Obviously, the daily range depends on the whole price path over a single day, and hence its sample distribution can be realized by the Euler method with estimated volatility. Consequently, we are able to implement empirical study using the Euler method developed here.

6.7 Proofs

6.7.1 A Preliminary result

Lemma 6.7.1. Under Assumptions 6.2.1 and 6.2.3 (no Assumption 6.2.2 here). Fix a discretization distance $\delta(\geq \Delta_n)$, for $i = 0, 1, \ldots, \lfloor \frac{T}{\delta} \rfloor - 1$, the statistical error at time $i\delta$ is given by

$$S_{i,n} = \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - \frac{1}{k_n \Delta_n} \int_{\lceil \frac{i\delta}{\Delta_n} \rceil \Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + k_n)\Delta_n} c_s ds,$$

then conditioning on $\{c_t : t \geq 0\}$, it holds that

$$\begin{aligned}
\mathbb{E}_c^1(S_{i,n}) &= 0, \\
\mathbb{E}_c^1(S_{i,n})^2 &= O(\frac{1}{k_n}), \\
\mathbb{E}_c^1(S_{i,n})^3 &= O(\frac{1}{k_n^2}), \\
\mathbb{E}_c^1(S_{i,n})^4 &= O(\frac{1}{k_n^2}),
\end{aligned}$$

and the unconditional moments are of the same order. In particular, the constant on the RHS does not depend on i.

Proof. For notational simplicity, we drop the superscript i in the rest of the proof. We have

$$\begin{aligned}
S_n &= \frac{1}{k_n \Delta_n} \sum_{\ell=1}^{k_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - \frac{1}{k_n \Delta_n} \int_{\lceil \frac{i\delta}{\Delta_n} \rceil \Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + k_n)\Delta_n} c_s ds \\
&= \frac{1}{k_n} \sum_{\ell=1}^{k_n} \underbrace{\left(\frac{1}{\Delta_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds \right)}_{Z_\ell^n} \\
&:= \frac{1}{k_n} \sum_{\ell=1}^{k_n} Z_\ell^n.
\end{aligned}$$

It is important to recognize that conditioning on c, Z_ℓ^n are independent as ℓ varies, and we will first estimate various moments of Z_ℓ^n.

For each $\ell \in \{1, 2, \ldots, k_n\}$, the first moment of Z_ℓ^n are given by

$$\begin{aligned}
\mathbb{E}_c^1(Z_\ell^n) &= \mathbb{E}_c^1 \left(\frac{1}{\Delta_n} \left(\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X \right)^2 - \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds \right) \\
&= \frac{1}{\Delta_n} \mathbb{E}_c^1 \left((\Delta^n_{\lceil \frac{i\delta}{\Delta_n} \rceil + \ell} X)^2 \right) - \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds \\
&= \frac{1}{\Delta_n} \mathbb{E}_c^1 \left(\left(\int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} \sqrt{c_s} dW_s \right)^2 \right) - \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds \\
&= \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds - \frac{1}{\Delta_n} \int_{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell - 1)\Delta_n}^{(\lceil \frac{i\delta}{\Delta_n} \rceil + \ell)\Delta_n} c_s ds \\
&= 0,
\end{aligned}$$

and the second moment

$$\begin{aligned}
\mathbb{E}_c^1\left((Z_\ell^n)^2\right) &= \mathbb{E}_c^1\left(\frac{1}{\Delta_n}\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^2 - \frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&= \frac{1}{\Delta_n^2}\mathbb{E}_c^1\left((\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X)^4\right) + \left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&\quad -2\frac{1}{\Delta_n}\underbrace{\mathbb{E}_c^1\left((\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X)^2\right)}_{\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds}\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds \\
&= \frac{1}{\Delta_n^2}\mathbb{E}_c^1\left((\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X)^4\right) - \left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&= \frac{1}{\Delta_n^2}3\left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 - \left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&= 2\left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&\sim O(1),
\end{aligned}$$

where the last integral being of order 1 since $\{c_t : t \geq 0\}$ is uniformly bounded from below and above. Moreover

$$\begin{aligned}
\mathbb{E}_c^1\left((Z_\ell^n)^3\right) &= \mathbb{E}\left(\frac{1}{\Delta_n}\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^2 - \frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^3 \\
&= \frac{1}{\Delta_n^3}\mathbb{E}\left(\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^6 - 3\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^4\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right. \\
&\quad \left.+3\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^2\left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 - \left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^3\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\Delta_n^3}\left(15\left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^3\right. \\
&\qquad -9\left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds \\
&\qquad +3\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds \left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^2 \\
&\qquad \left. -\left(\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^3\right) \\
&= 8\left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^3 \\
&\sim O(1)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}_c^1\left((Z_\ell^n)^4\right) &= \mathbb{E}_c^1\left(\frac{1}{\Delta_n}\left(\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X\right)^2 - \frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^4 \\
&= \left(\frac{1}{\Delta_n}\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds\right)^4 \mathbb{E}_c^1\left(\left(\underbrace{\frac{\Delta^n_{\lceil\frac{i\delta}{\Delta_n}\rceil+\ell}X}{\sqrt{\int_{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell-1)\Delta_n}^{(\lceil\frac{i\delta}{\Delta_n}\rceil+\ell)\Delta_n} c_s ds}}}_{\sim\mathcal{N}(0,1)}\right)^2 - 1\right)^4 \\
&\sim O(1)
\end{aligned}
$$

In fact, all the moments of Z_ℓ^n are of order 1, due to the fact that $\{c_t : t \geq 0\}$ is uniformly bounded from below and above.

With all the moments computed above, we can calculate the associated moments of S_n. Then

$$
\begin{aligned}
\mathbb{E}_c^1(S_n) &= 0 \\
\mathbb{E}_c^1(S_n^2) &= \mathbb{E}_c^1\left(\frac{1}{k_n}\sum_{\ell=1}^{k_n} Z_\ell^n\right)^2 \\
&= \frac{1}{k_n^2}\sum_{\ell=1}^{k_n}\mathbb{E}_c^1\left((Z_\ell^n)^2\right) + 0(\textit{product terms vanishes}) \\
&\sim \frac{1}{k_n^2}k_n O(1) = O(\frac{1}{k_n}) \\
\mathbb{E}_c^1(S_n^3) &= \mathbb{E}_c^1\left(\frac{1}{k_n}\sum_{\ell=1}^{k_n} Z_\ell^n\right)^3 \\
&= \frac{1}{k_n^3}\mathbb{E}_c\left(\sum_{\ell=1}^{k_n}(Z_\ell^n)^3 + 3\sum_{i\neq j}(Z_i^n)^2 Z_j^n + 6\sum_{i\neq j\neq k} Z_i^n Z_j^n Z_k^n\right) \\
&= \frac{1}{k_n^3}\sum_{\ell=1}^{k_n}\mathbb{E}_c(Z_\ell^n)^3 \sim O(\frac{1}{k_n^2}).
\end{aligned}
$$

At last, we have

$$
\begin{aligned}
\mathbb{E}_c^1(S_n^4) &= \mathbb{E}_c^1\left(\frac{1}{k_n}\sum_{\ell=1}^{k_n} Z_\ell^n\right)^4 \\
&= \frac{1}{k_n^4}\mathbb{E}_c\left(\sum_{\ell=1}^{k_n}(Z_\ell^n)^4 + 6\sum_{i\neq j}(Z_i^n)^2(Z_j^n)^2\right) \\
&\sim \frac{1}{k_n^4}k_n + \frac{1}{k_n^4}k_n(k_n-1) \sim O(\frac{1}{k_n^2}).
\end{aligned}
$$

Since all the estimates of convergence order of moments of S_n do not depend on c_t (because $\{c_t : t \geq 0\}$ is uniformly bounded from above and below), by law of iterated expectation, the unconditional moments have the same convergence order as those of conditional ones. Moreover, the convergence order does not depend on i, namely, does not depend on which discretization time point $i\delta$ to consider, which can be easily seen from the proof. $\square$

6.7.2 Proof of Theorem 6.4.1

Fix finite horizon T, and at stage n, fix discretization frequency δ, sampling frequency Δ_n and length of window k_n for spot volatility estimation, consider Euler approximation defined in Section 6.3:

$$Y_t^{n,\delta} = \sum_{i=0}^{[t/\delta]-1} \sqrt{\hat{c}_{i\delta}}(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}), \quad 0 \le t \le T,$$

which is now viewed as a stochastic process with time index t. By convention, if $t < \delta$, the sum above is just $Y_0^{n,\delta} = 0$. We are going to derive the global/pathwise approximation of $Y^{n,\delta}$ to $(\widetilde{X}_t)_{0\le t\le T}$. First note that for any $t \in [0,T]$,

$$\begin{aligned}
Y_t^{n,\delta} - \widetilde{X}_t &= \sum_{i=0}^{[t/\delta]-1} \sqrt{\hat{c}_{i\delta}}(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}) - \int_0^t \sqrt{c_s} d\widetilde{W}_s \\
&= \underbrace{\sum_{i=0}^{[t/\delta]-1} (\sqrt{\hat{c}_{i\delta}} - \sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta})}_{\text{local estimation problem}} \qquad (6.6) \\
&\quad + \underbrace{\sum_{i=0}^{[t/\delta]-1} \sqrt{c_{i\delta}}(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}) - \int_0^{[\frac{t}{\delta}]\delta} \sqrt{c_s} d\widetilde{W}_s}_{\text{discretization part}} \\
&\quad + \underbrace{\int_0^{[\frac{t}{\delta}]\delta} \sqrt{c_s} d\widetilde{W}_s - \int_0^t \sqrt{c_s} d\widetilde{W}_s}_{\text{residual part}} \qquad (6.7)
\end{aligned}$$

Then by triangle inequality it follows that

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{n,\delta} - \widetilde{X}_t|\right) &\le \mathbb{E}\left(\sup_{0\le t\le T} \left|\sum_{i=0}^{[t/\delta]-1} (\sqrt{\hat{c}_{i\delta}} - \sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta})\right|\right) \\
&\quad + \mathbb{E}\left(\sup_{0\le t\le T} \left|\sum_{i=0}^{[t/\delta]-1} \sqrt{c_{i\delta}}(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta}) - \int_0^{[\frac{t}{\delta}]\delta} \sqrt{c_s} d\widetilde{W}_s\right|\right) \\
&\quad + \mathbb{E}\left(\sup_{0\le t\le T} \left|\int_{[\frac{t}{\delta}]\delta}^t \sqrt{c_s} d\widetilde{W}_s\right|\right).
\end{aligned}$$

The rest of the proof is to derive the above upper bound step by step.

Step 1: Local estimation part. We first deal with the "local estimation" part. By Hölder inequality,

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{0\le t\le T}\left|\sum_{i=0}^{[t/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta})\right|\right)\\
&\le \left\{\mathbb{E}\left(\sup_{0\le t\le T}\left|\sum_{i=0}^{[t/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta})\right|\right)^2\right\}^{\frac{1}{2}}
\end{aligned}
$$

The natural next step is to apply Doob's inequality to convert the supreme value of the stochastic process over time interval $[0,T]$ into the value of the process at terminal time T, which is easier to handle. However, to validate the use of Doob's inequality, we have to show that the stochastic process

$$
Z_t^{n,\delta} := \sum_{i=0}^{[t/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta}), \quad 0\le t\le T,
$$

is a martingale associated with certain filtration.

Lemma 6.7.2. Define continuous-time filtration $(\mathcal{G}_t)_{0\le t\le T}$ as

$$
\mathcal{G}_t = \mathcal{F}^1 \otimes \mathcal{F}^2_{[\frac{t}{\delta}]}, \quad 0\le t\le T.
$$

Then the stochastic process $(Z_t^{n,\delta})_{0\le t\le T}$ is a martingale associated with $(\mathcal{G}_t)_{0\le t\le T}$.

The proof of the lemma is straightforward by the definition of martingale. In light of this fact, by Doob's inequality we have

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{0\le t\le T}\left|\sum_{i=0}^{[t/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta})\right|\right)^2\\
&=\mathbb{E}\left(\sup_{0\le t\le T}\left|\sum_{i=0}^{[t/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta})\right|^2\right)\\
&\le 4\mathbb{E}\left(\sum_{i=0}^{[T/\delta]-1}(\sqrt{\hat{c}_{i\delta}}-\sqrt{c_{i\delta}})(\widetilde{W}_{(i+1)\delta}-\widetilde{W}_{i\delta})\right)^2 \quad \text{evaluated at terminal time } T\\
&\le K\left(\frac{1}{k_n}+(k_n\Delta_n)^{2\rho}\right)
\end{aligned}
$$

where the last inequality follows from Proposition 6.7.1 plus Assumption 6.2.3 and 6.2.4, replacing Δ_n by δ.

Step 2: Discretization part. Next we try to deal with the discretization part, which is essentially a global approximation of discretization of a stochastic integral to the stochastic integral itself. Define

$$
\tilde{c}_s=\sum_{i=0}^{[T/\delta]-1} c_{i\delta}\cdot 1_{[i\delta,(i+1)\delta)}(s), \quad s\in[0,T],
$$

which is an adapted càdlàg process with paths being piecewise constant. Then we have

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,T]}\left|\sum_{i=0}^{[t/\delta]-1}\sqrt{c_{i\delta}}\Delta_{i+1}^{n}\widetilde{W}-\int_{0}^{[\frac{t}{\delta}]\delta}\sqrt{c_s}d\widetilde{W}_s\right|\right)\\
&=\mathbb{E}\left(\sup_{t\in[0,T]}\left|\int_{0}^{[\frac{t}{\delta}]\delta}(\sqrt{\tilde{c}_s}-\sqrt{c_s})d\widetilde{W}_s\right|\right)\\
&=\mathbb{E}\left(\sup_{t\in[0,[\frac{T}{\delta}]\delta]}\left|\int_{0}^{t}(\sqrt{\tilde{c}_s}-\sqrt{c_s})d\widetilde{W}_s\right|\right)\\
&\leq K\mathbb{E}\left(\int_{0}^{[\frac{T}{\delta}]\delta}(\sqrt{\tilde{c}_s}-\sqrt{c_s})^2ds\right)^{\frac{1}{2}} \qquad \text{by Burkholder-Davis-Gundy}\\
&\leq K\mathbb{E}^1\left\{\int_{0}^{[\frac{T}{\delta}]\delta}(\tilde{c}_s-c_s)^2ds\right\}^{\frac{1}{2}} \qquad c_t \text{ is bounded away from 0}\\
&=K\left\{\mathbb{E}^1\left(\int_{0}^{[\frac{T}{\delta}]\delta}(\tilde{c}_s-c_s)^2ds\right)\right\}^{\frac{1}{2}} \qquad \text{Jensen's inequality}\\
&=K\left\{\int_{0}^{[\frac{T}{\delta}]\delta}\mathbb{E}^1(\tilde{c}_s-c_s)^2ds\right\}^{\frac{1}{2}} \qquad \text{by Fubini's theorem}\\
&=K\left\{\sum_{i=0}^{[T/\delta]-1}\int_{i\delta}^{(i+1)\delta}\mathbb{E}^1(\tilde{c}_s-c_s)^2ds\right\}^{\frac{1}{2}}\\
&=K\left\{\sum_{i=0}^{[T/\delta]-1}\int_{i\delta}^{(i+1)\delta}\mathbb{E}^1(c_{i\delta}-c_s)^2ds\right\}^{\frac{1}{2}} \qquad \text{by definition of } \tilde{c}_s\\
&=K\left\{\sum_{i=0}^{[T/\delta]-1}\int_{i\delta}^{(i+1)\delta}(i\delta-s)^{2\rho}ds\right\}^{\frac{1}{2}} \qquad \text{by definition of } \tilde{c}_s\\
&\leq K\left(\sum_{i=0}^{[T/\delta]-1}\int_{i\delta}^{(i+1)\delta}\delta^{2\rho}ds\right)^{\frac{1}{2}}\\
&\leq K\sqrt{T}\delta^{\rho}.
\end{aligned}
$$

Hence by using Burkholder-Davis-Gundy inequality, we show that the convergence rate (upper bound) for the uniform "discretization part" is the same as that for only considering the terminal time.

Step 3: Residual part. At last we deal with the residual part. It is not as easy as it sounds like since we are essentially trying to control the moments of modulus of continuity of Itö process. Fortunately, (Fischer and Nappo, 2010) proves those results for us. More specifically, by Theorem 1 in (Fischer and Nappo, 2010) and the assumption that $(c_t)_{t\geq 0}$ is uniformly bounded from both below and above, we have

$$\mathbb{E}\left(\sup_{0\leq t\leq T}\left|\int_{[\frac{t}{\delta}]\delta}^{t}\sqrt{c_s}d\widetilde{W}_s\right|\right)\leq K\left(\delta\log\left(\frac{2T}{\delta}\right)\right)^{\frac{1}{2}},$$

as desired.

6.7.3 Proof of Theorem 6.4.2

We already know that under the optimal simulation scheme, we have upper bound

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|Y_t^n-\widetilde{X}_t|\right)\leq K\Delta_n^{\frac{1}{2+\frac{1}{\rho}}}$$

for some constant K.

Since $\mathbb{E}\left(|Y_T^n-\widetilde{X}_T|\right)\leq\mathbb{E}\left(\sup_{0\leq t\leq T}|Y_t^n-\widetilde{X}_t|\right)$, it suffices to show that the exact convergence rate for $|Y_T^n-\widetilde{X}_T|$ is of order $\Delta_n^{\frac{1}{2+\frac{1}{\rho}}}$ as well.

Similar to the proof of Theorem 6.4.1, we decompose the difference between Y_T^n and $\widetilde{X}_T$ as follows:

$$\begin{aligned}
Y_T^n - \widetilde{X}_T &= \sum_{i=0}^{[T/\Delta_n]-1} \sqrt{\hat{c}_{i\Delta_n}}\Delta_{i+1}\widetilde{W} - \int_0^T \sqrt{c_s}d\widetilde{W}_s \\
&= \sum_{i=0}^{[T/\Delta_n]-1} \sqrt{\hat{c}_{i\Delta_n}}\Delta_{i+1}\widetilde{W} - \sum_{i=0}^{[T/\Delta_n]-1} \sqrt{c_{i\delta}}\Delta_{i+1}\widetilde{W} \\
&\quad + \sum_{i=0}^{[T/\Delta_n]-1} \sqrt{c_{i\Delta_n}}\Delta_{i+1}\widetilde{W} - \int_0^T \sqrt{c_s}d\widetilde{W}_s \\
&= \underbrace{\sum_{i=0}^{[T/\Delta_n]-1} (\sqrt{\hat{c}_{i\Delta_n}} - \sqrt{c_{i\Delta_n}})\Delta_{i+1}\widetilde{W}}_{\text{local estimation problem}} \\
&\quad + \underbrace{\sum_{i=0}^{[T/\Delta_n]-1} \sqrt{c_{i\Delta_n}}\Delta_{i+1}\widetilde{W} - \int_0^{[\frac{T}{\Delta_n}]\Delta_n} \sqrt{c_s}d\widetilde{W}_s}_{\text{discretization part}} \\
&\quad + \underbrace{\int_0^{[\frac{T}{\Delta_n}]\Delta_n} \sqrt{c_s}d\widetilde{W}_s - \int_0^T \sqrt{c_s}d\widetilde{W}_s}_{\text{residual part}}
\end{aligned}$$

We first deal with the "residual part". Note that

$$\begin{aligned}
\mathbb{E}\left(\int_0^{[\frac{T}{\Delta_n}]\Delta_n} \sqrt{c_s}d\widetilde{W}_s - \int_0^T \sqrt{c_s}d\widetilde{W}_s\right)^2 &= \mathbb{E}\left(\int_{[\frac{T}{\Delta_n}]\Delta_n}^T \sqrt{c_s}d\widetilde{W}_s\right)^2 \\
&= \mathbb{E}\left(\int_{[\frac{T}{\Delta_n}]\Delta_n}^T c_s ds\right) \quad \text{by Itö isometry} \\
&\leq K\Delta_n \qquad c_t \text{ is bounded from above}
\end{aligned}$$

It follows from Jensen's inequality that

$$\mathbb{E}\left|\int_0^{[\frac{T}{\Delta_n}]\Delta_n} \sqrt{c_s}d\widetilde{W}_s - \int_0^T \sqrt{c_s}d\widetilde{W}_s\right| \leq K\Delta_n^{\frac{1}{2}}.$$

As for the discretization part, by the Step 2 in proof of Theorem 6.4.1, we have

$$\mathbb{E}\left|\sum_{i=0}^{[T/\Delta_n]-1}\sqrt{c_{i\Delta_n}}\Delta_{i+1}^n\widetilde{W}-\int_0^{[\frac{T}{\Delta_n}]\Delta_n}\sqrt{c_s}d\widetilde{W}_s\right|\leq K\sqrt{T}\Delta_n^\rho.$$

The local estimation part is the leading term, and is more involved to deal with.

Lemma 6.7.3. Under Assumptions 6.2.1, 6.2.3 and 6.2.4, assume that as $n\to\infty$

$$\Delta_n\to 0,\quad k_n^{\rho+\frac{1}{2}}\Delta_n^\rho\to\beta\in(0,\infty)$$

Conditioning on $\{c_t : t\geq 0\}$, the following functional convergence in law holds

$$\frac{\sum_{i=0}^{[\cdot/\Delta_n]-1}\frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})\Delta_{i+1}^n\widetilde{W}}{\alpha_{\cdot,n}}\Longrightarrow Y\quad\text{on}\quad[0,\mathrm{T}]$$

where Y is a continuous centered Gaussian process with independent increments having

$$\mathbb{E}(Y_t^2)=1,\quad\forall\ t\in[0,T]$$

and

$$\alpha_{t,n}^2=\sum_{i=0}^{[t/\Delta_n]-1}\frac{\Delta_n}{4c_{i\Delta_n}}\left(\underbrace{\frac{2}{k_n^2}\sum_{\ell=1}^{k_n}\left(\frac{1}{\Delta_n}\int_{(i+\ell-1)\Delta_n}^{(i+\ell)\Delta_n}c_s ds\right)^2}_{\text{due to statistical error in local estimation}}+\underbrace{\left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n}c_s ds-c_{i\Delta_n}\right)^2}_{\text{due to target error in local estimation}}\right).$$

Remark 6.7.1. $\{\sqrt{t}Y_t\}:t\in[0,T]\}$ is standard Brownian motion on $[0,T]$.

Proof. Fix a sample point $\omega\in\Omega$. Throughout this proof, we will be conditioning on the volatility path $\{c_t(\omega):t\geq 0\}$. To emphasize this, we use, for example, $\mathbb{E}_c(\cdot)$ with subscript c when taking expectation conditioning on $\{c_t(\omega):t\geq 0\}$.

By Assumption 6.2.4, there exists $\rho\in(0,1]$ such that $\forall s,t\in[0,T]$,

$$|c(t,\omega)-c(s,\omega)|\leq A(\omega)|t-s|^\rho,$$

for some constant $A(\omega)$. Here by requiring

$$k_n^{\rho+\frac{1}{2}}\Delta_n^\rho \to \beta \in (0,\infty),$$

we are actually taking

$$k_n \sim \Delta_n^{-\frac{\rho}{\rho+\frac{1}{2}}} = \Delta_n^{-\frac{2\rho}{2\rho+1}},$$

or equivalently

$$k_n^{-\frac{1}{2}} \sim (k_n\Delta_n)^\rho \sim \Delta_n^{-\frac{2\rho}{2\rho+1}},$$

which means we strike the balance between the convergence rate of statistical error of local estimation and that of target error. This immediately implies that

$$k_n\Delta_n \sim \Delta_n^{1-\frac{\rho}{\rho+\frac{1}{2}}} \to 0.$$

We fit our setting into Theorem 2.2.13 in (Jacod and Protter, 2012): for each $n \geq 1$ and $i = 0, 1, \ldots, [\frac{T}{\Delta_n}] - 1$, define

$$\begin{aligned}
\eta_i^n &= \frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})\Delta_{i+1}^n\widetilde{W} \\
\mathcal{G}_i^n &= \mathcal{F}^1_{(i+k_n)\Delta_n} \otimes \mathcal{F}^2_{(i+1)\Delta_n}
\end{aligned}$$

with stopping rule $N_n(t)$ being deterministic $\lfloor \frac{t}{\Delta_n} \rfloor$ (note here the index i starts from 0 while in (Jacod and Protter, 2012) i starts from 1). Then immediately $\eta_i^n \in \mathcal{G}_i^n$ (in fact, we can think of $\mathcal{G}_i^n$ as the filtration generated by η_i^n).

Note that for any $i = 0, 1, \ldots, [\frac{T}{\Delta_n}] - k_n$,

$$\begin{aligned}
\hat{c}_{i\delta_n} = \hat{c}_{i\Delta_n} &= \frac{1}{k_n\Delta_n}\sum_{j=\lceil \frac{i\delta_n}{\Delta_n}\rceil+1}^{\lceil \frac{i\delta_n}{\Delta_n}\rceil+k_n}\left(\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s\right)^2 \\
&= \frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\left(\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s\right)^2 \quad \delta_n = \Delta_n.
\end{aligned}$$

Then as usual we decompose local estimation error into statistical error and target error:

$$
\begin{aligned}
\hat{c}_{i\Delta_n} - c_{i\Delta_n} &= \underbrace{\frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\left(\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s\right)^2 - \frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\int_{(j-1)\Delta_n}^{j\Delta_n}c_s ds}_{S_{i,n}} \\
&\quad \underbrace{+\frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\int_{(j-1)\Delta_n}^{j\Delta_n}c_s ds - c_{i\Delta_n}}_{D_{i,n}} \\
&:= S_{i,n} + D_{i,n},
\end{aligned}
$$

where $S_{i,n}$ can be written as

$$
\begin{aligned}
S_{i,n} &= \frac{1}{k_n}\sum_{j=i+1}^{i+k_n}\left(\frac{1}{\Delta_n}\left(\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s\right)^2 - \frac{1}{\Delta_n}\int_{(j-1)\Delta_n}^{j\Delta_n}c_s ds\right) \\
&:= \frac{1}{k_n}\sum_{j=i+1}^{i+k_n} Z_{i,j,n}
\end{aligned}
$$

with $Z_{i,j,n}$ being centered and independent as j varies (recall the moments estimates for $S_{i,n}$ are given by Lemma 6.7.1). Then for any $t \in [0,T]$,

$$
\begin{aligned}
\alpha^2_{t,n} &= \mathbb{E}_c\left(\sum_{i=0}^{[t/\Delta_n]-1}\eta_i^n\right)^2 = \sum_{i=0}^{[t/\Delta_n]-1}\mathbb{E}_c\left(\eta_i^n\right)^2 \\
&= \sum_{i=0}^{[t/\Delta_n]-1}\frac{\Delta_n}{4c_{i\Delta_n}}\mathbb{E}_c\left(\hat{c}_{i\Delta_n} - c_{i\Delta_n}\right)^2 = \sum_{i=0}^{[t/\Delta_n]-1}\frac{\Delta_n}{4c_{i\Delta_n}}\left(\mathbb{E}_c(S_{i,n})^2 + D^2_{i,n}\right) \\
&= \sum_{i=0}^{[t/\Delta_n]-1}\frac{\Delta_n}{4c_{i\Delta_n}}\left(\frac{2}{k_n^2}\sum_{\ell=1}^{k_n}\left(\frac{1}{\Delta_n}\int_{(i+\ell-1)\Delta_n}^{(i+\ell)\Delta_n}c_s ds\right)^2 + D^2_{i,n}\right)
\end{aligned}
$$

By undergraduate analysis, it follows that

$$
\begin{aligned}
k_n\sum_{i=0}^{[t/\Delta_n]-1}\frac{\Delta_n}{4c_{i\Delta_n}}\frac{2}{k_n^2}\sum_{\ell=1}^{k_n}\left(\frac{1}{\Delta_n}\int_{(i+\ell-1)\Delta_n}^{(i+\ell)\Delta_n}c_s ds\right)^2 &\to \frac{1}{2}\int_0^t c_s ds < \infty \\
k_n D^2_{i,n} = (\sqrt{k_n}D_{i,n})^2 &< K(\beta)\frac{1}{4}\int_0^t \frac{1}{c_s}ds < \infty
\end{aligned}
$$

Therefore

$$k_n \alpha_{t,n}^2 \sim O(1)$$

or in other words, the exact convergence rate of $\alpha_{t,n}^2$ is k_n.

Now note

$$\begin{aligned}
\mathbb{E}_c(\eta_i^n | \mathcal{G}_{i-1}^n) &= \mathbb{E}_c\left(\frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})\Delta_{i+1}^n \widetilde{W} | \mathcal{G}_{i-1}^n\right) \\
&= \mathbb{E}_c\left(\frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n}) | \mathcal{G}_{i-1}^n\right)\mathbb{E}_c(\Delta_{i+1}^n \widetilde{W}) = 0
\end{aligned}$$

Then it suffices to show

$$\frac{1}{\alpha_{t,n}^2} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^2 | \mathcal{G}_{i-1}^n) \quad \xrightarrow{\mathbb{P}} \quad 1, \quad \forall\, t > 0 \tag{6.8}$$

$$\frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^4 | \mathcal{G}_{i-1}^n) \quad \xrightarrow{\mathbb{P}} \quad 0, \quad \forall\, t > 0 \tag{6.9}$$

To prove (6.8), first note

$$\begin{aligned}
& \mathbb{E}_c \left\{ \frac{1}{\alpha_{t,n}^2} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^2 | \mathcal{G}_{i-1}^n) - 1 \right\}^2 \\
= \quad & \frac{1}{\alpha_{t,n}^4} \mathbb{E}_c \left\{ \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^2 | \mathcal{G}_{i-1}^n) - \alpha_{t,n}^2 \right\}^2 \\
= \quad & \frac{1}{\alpha_{t,n}^4} \mathbb{E}_c \left\{ \sum_{i=0}^{[t/\Delta_n]-1} \frac{\Delta_n}{4c_{i\Delta_n}} \mathbb{E}_c\left((\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 | \mathcal{G}_{i-1}^n\right) - \sum_{i=0}^{[t/\Delta_n]-1} \frac{\Delta_n}{4c_{i\Delta_n}} \mathbb{E}_c(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 \right\}^2 \\
= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \mathbb{E}_c \left\{ \sum_{i=0}^{[t/\Delta_n]-1} \frac{1}{4c_{i\Delta_n}} \left(\mathbb{E}_c\left((\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 | \mathcal{G}_{i-1}^n\right) - \mathbb{E}_c(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2\right) \right\}^2 \\
:= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \mathbb{E}_c \left\{ \sum_{i=0}^{[t/\Delta_n]-1} H_{i,n} \right\}^2
\end{aligned}$$

where $H_{i,n} := \frac{1}{4c_{i\Delta_n}}\left(\mathbb{E}_c\left((\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2|\mathcal{G}_{i-1}^n\right) - \mathbb{E}_c(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2\right)$, and in particular the randomness of $H_{i,n}$ is up to time $(i+k_n-1)\Delta_n$. Then we immediately have

$$\mathbb{E}_c(H_{i,n}) = 0$$

Since

$$\begin{aligned}
& \mathbb{E}_c(|\eta_i^n|^2|\mathcal{G}_{i-1}^n) = \mathbb{E}_c(\frac{1}{4c_{i\Delta_n}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2(\Delta_{i+1}^n\widetilde{W})^2|\mathcal{G}_{i-1}^n) \\
= \quad & \frac{\Delta_n}{4c_{i\Delta_n}}\mathbb{E}_c((\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2|\mathcal{G}_{i-1}^n) \\
= \quad & \frac{\Delta_n}{4c_{i\Delta_n}}\left(\mathbb{E}_c(S_{i,n}^2|\mathcal{G}_{i-1}^n) + 2D_{i,n}\mathbb{E}_c(S_{i,n}|\mathcal{G}_{i-1}^n) + D_{i,n}^2\right) \\
= \quad & \frac{\Delta_n}{4c_{i\Delta_n}}\left(\frac{1}{k_n^2}\left(\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right)^2 + \frac{2}{k_n^2}\left(\frac{1}{\Delta_n}\int_{(i+k_n-1)\Delta_n}^{(i+k_n)\Delta_n} c_s ds\right)^2 + \frac{2D_{i,n}}{k_n}\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n} + D_{i,n}^2\right),
\end{aligned}$$

it follows that

$$\begin{aligned}
& \mathbb{E}_c(H_{i,n}^2) \\
= \quad & \frac{1}{16c_{i\Delta_n}^2}\mathbb{E}_c\left\{\frac{1}{k_n^2}\left(\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right)^2 + \frac{2}{k_n^2}\left(\frac{1}{\Delta_n}\int_{(i+k_n-1)\Delta_n}^{(i+k_n)\Delta_n} c_s ds\right)^2 + \frac{2D_{i,n}}{k_n}\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right. \\
& \quad \left. + D_{i,n}^2 - \mathbb{E}_c(S_{i,n}^2) - D_{i,n}^2\right\}^2 \\
= \quad & \frac{1}{16c_{i\Delta_n}^2}\mathbb{E}_c\left\{\frac{1}{k_n^2}\left(\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right)^2 + \frac{2}{k_n^2}\left(\frac{1}{\Delta_n}\int_{(i+k_n-1)\Delta_n}^{(i+k_n)\Delta_n} c_s ds\right)^2\right. \\
& \left. + \frac{2D_{i,n}}{k_n}\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n} - \mathbb{E}_c(S_{i,n}^2)\right\}^2 \\
\leq \quad & \frac{1}{16c_{i\Delta_n}^2}\left\{\frac{1}{k_n^4}\mathbb{E}_c\left(\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right)^4 + \frac{4}{k_n^4}\left(\frac{1}{\Delta_n}\int_{(i+k_n-1)\Delta_n}^{(i+k_n)\Delta_n} c_s ds\right)^4 + \right. \\
& \quad \left. \frac{4D_{i,n}^2}{k_n^2}\mathbb{E}_c\left(\sum_{j=i+1}^{i+k_n-1} Z_{i,j,n}\right)^2 + (\mathbb{E}_c(S_{i,n}^2))^2\right\} \\
\sim \quad & O(\frac{1}{k_n^2})
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& \mathbb{E}_c \left\{ \frac{1}{\alpha_{t,n}^2} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^2 | \mathcal{G}_{i-1}^n) - 1 \right\}^2 \\
= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \mathbb{E}_c \left\{ \sum_{i=0}^{[t/\Delta_n]-1} H_{i,n} \right\}^2 = \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(H_{i,n}^2) + \frac{2}{\alpha_{t,n}^4} \mathbb{E}_c \left(\sum_{0 \le i < m}^{[t/\Delta_n]-1} H_{i,n} H_{m,n} \right) \\
= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(H_{i,n}^2) + \frac{2\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-2} \mathbb{E}_c \left\{ H_{i,n} \sum_{m=i+1}^{[t/\Delta_n]-1} H_{m,n} \right\} \\
= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(H_{i,n}^2) + \frac{2\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-2} \mathbb{E}_c \left\{ H_{i,n} \sum_{m=i+1}^{\min\{[t/\Delta_n]-1, i+k_n-2\}} H_{m,n} \right\} \\
= \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(H_{i,n}^2) + \frac{2\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-2} \sum_{m=i+1}^{\min\{[t/\Delta_n]-1, i+k_n-2\}} \mathbb{E}_c \left\{ H_{i,n} H_{m,n} \right\} \\
\le \quad & \frac{\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(H_{i,n}^2) + \frac{2\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-2} \sum_{m=i+1}^{\min\{[t/\Delta_n]-1, i+k_n-2\}} \sqrt{\mathbb{E}_c(H_{i,n}^2)} \sqrt{\mathbb{E}_c(H_{m,n}^2)} \\
\sim \quad & \Delta_n + k_n \Delta_n \to 0,
\end{aligned}
$$

where for the third equality is due to the fact that $H_{i,n}$ and $H_{m,n}$ are independent if $|m-i| > k_n - 1$ (finite range dependence) and for the inequality we use the Cauchy-Schwartz inequality. Thus we have shown (6.8). Next to show (6.9)

$$
\begin{aligned}
\mathbb{E}_c \left| \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^4 | \mathcal{G}_{i-1}^n) \right| &\leq \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c \left| \mathbb{E}_c(|\eta_i^n|^4 | \mathcal{G}_{i-1}^n) \right| \\
&= \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \mathbb{E}_c(|\eta_i^n|^4) \\
&= \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \frac{1}{16 c_{i\Delta_n}^2} \mathbb{E}_c \left((\hat{c}_{i\Delta_n} - c_{i\Delta_n}) \Delta_{i+1}^n \widetilde{W} \right)^4 \\
&= \frac{1}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \frac{1}{16 c_{i\Delta_n}^2} \mathbb{E}_c (\hat{c}_{i\Delta_n} - c_{i\Delta_n})^4 \mathbb{E}_c (\Delta_{i+1}^n \widetilde{W})^4 \\
&= \frac{3\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} \frac{1}{16 c_{i\Delta_n}^2} \mathbb{E}_c (\hat{c}_{i\Delta_n} - c_{i\Delta_n})^4 \\
&= \frac{3\Delta_n^2}{\alpha_{t,n}^4} \sum_{i=0}^{[t/\Delta_n]-1} O(\frac{1}{k_n^2}) \\
&\sim \Delta_n \to 0.
\end{aligned}
$$

Therefore (6.9) holds and the theorem follows. □

Proposition 6.7.1. Under Assumptions 6.2.1, 6.2.3 and 6.2.4, assume that as $n \to \infty$

$$
\Delta_n \to 0, \quad k_n^{\rho+\frac{1}{2}} \Delta_n^{\rho} \to \beta \in (0, \infty).
$$

Then conditioning on $\{c_t : t \geq 0\}$, it holds that

$$
\frac{\sum_{i=0}^{[T/\Delta_n]-1} (\sqrt{\hat{c}_{i\Delta_n}} - \sqrt{c_{i\Delta_n}}) \Delta_{i+1}^n \widetilde{W}}{\alpha_{T,n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)
$$

where

$$
\alpha_{T,n}^2 = \sum_{i=0}^{[T/\Delta_n]-1} \frac{\Delta_n}{4 c_{i\Delta_n}} \left(\underbrace{\frac{2}{k_n^2} \sum_{\ell=1}^{k_n} \left(\frac{1}{\Delta_n} \int_{(i+\ell-1)\Delta_n}^{(i+\ell)\Delta_n} c_s ds \right)^2}_{\text{due to statistical error in local estimation}} + \underbrace{\left(\frac{1}{k_n \Delta_n} \int_{i\Delta_n}^{(i+k_n)\Delta_n} c_s ds - c_{i\Delta_n} \right)^2}_{\text{due to target error in local estimation}} \right).
$$

Proof. Fix a sample point $\omega \in \Omega$. As in the proof of Lemma 6.7.3, we will be conditioning on the volatility path $\{c_t(\omega) : t \geq 0\}$ throughout this proof.

By Taylor expansion, for each $i = 0, 1, \ldots, [T/\Delta_n] - 1$,

$$\sqrt{\hat{c}_{i\Delta_n}} = \sqrt{c_{i\Delta_n}} + \frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n}) - \frac{1}{8}\xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2,$$

where $\xi_{i,n} = \theta_{i,n}\hat{c}_{i\Delta_n} + (1-\theta_{i,n})c$ with $\theta_{i,n} \in [0,1]$. Then

$$\begin{aligned}
& \sum_{i=0}^{[T/\Delta_n]-1} (\sqrt{\hat{c}_{i\Delta_n}} - \sqrt{c_{i\Delta_n}})\Delta_{i+1}^n \widetilde{W} \\
= \quad & \sum_{i=0}^{[T/\Delta_n]-1} \left(\frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n}) - \frac{1}{8}\xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 \right) \Delta_{i+1}^n \widetilde{W} \\
= \quad & \underbrace{\sum_{i=0}^{[T/\Delta_n]-1} \frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})\Delta_{i+1}^n \widetilde{W}}_{I} - \frac{1}{8} \underbrace{\sum_{i=0}^{[T/\Delta_n]-1} \xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 \Delta_{i+1}^n \widetilde{W}}_{II}.
\end{aligned}$$

Conditioning on $\{c_t : t \geq 0\}$, by Lemma 6.7.3 and continuous mapping theorem (projection to terminal time T), we have

$$\frac{\sum_{i=0}^{[T/\Delta_n]-1} \frac{1}{2\sqrt{c_{i\Delta_n}}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})\Delta_{i+1}^n \widetilde{W}}{\alpha_{T,n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

By Slutsky's Theorem, what is left to show now is that the part II converges to 0 in mean square (in probability) after being scaled by $\alpha_{T,n}$.

$$\begin{aligned}
& \mathbb{E}_c \left(\frac{1}{\alpha_{T,n}} \sum_{i=0}^{[T/\Delta_n]-1} \xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 \Delta_{i+1}^n \widetilde{W} \right)^2 \\
& = \frac{1}{\alpha_{T,n}^2} \sum_{i=0}^{[T/\Delta_n]-1} \mathbb{E}_c \left(\xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2 \right)^2 \mathbb{E}_c(\Delta_{i+1}^n \widetilde{W})^2 \\
& = \frac{\Delta_n}{\alpha_{T,n}^2} \sum_{i=0}^{[T/\Delta_n]-1} \mathbb{E}_c \left(\xi_{i,n}^{-3}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^4 \right).
\end{aligned}$$

Notice that $\forall i$

$$
\begin{aligned}
&\mathbb{E}_c\left(\xi_{i,n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4\right)\\
&=\mathbb{E}_c\left(\xi_{i,n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4 1_{\{\hat{c}_{i\Delta_n}>c_{i\Delta_n}\}}\right)+\mathbb{E}_c\left(\xi_{i,n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4 1_{\{\hat{c}_{i\Delta_n}<c_{i\Delta_n}\}}\right)\\
&\le \mathbb{E}_c\left(c^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4 1_{\{\hat{c}_{i\Delta_n}>c_{i\Delta_n}\}}\right)+\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4 1_{\{\hat{c}_{i\Delta_n}<c_{i\Delta_n}\}}\right)\\
&\le \underbrace{\mathbb{E}_c\left((\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4\right)}_{\sim k_n^{-2}}+\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4\right),
\end{aligned}
$$

Recall that the exact convergence rate for $\alpha_{T,n}^2$ is k_n, so we have

$$
\frac{\Delta_n}{\alpha_{T,n}^2}\sum_{i=0}^{[T/\Delta_n]-1}\mathbb{E}_c\left((\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4\right)\sim O(\frac{1}{k_n})\to 0.
$$

On the other hand, note

$$
\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-3}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^4\right) \quad\le\quad \sqrt{\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-6}\right)}\sqrt{\mathbb{E}_c(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^8}
$$

We first consider $\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-6}\right)$, note that

$$
\begin{aligned}
\hat{c}_{i\Delta_n} &= \frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\left(\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s\right)^2\\
&= \frac{1}{k_n\Delta_n}\sum_{j=i+1}^{i+k_n}\underbrace{\int_{(j-1)\Delta_n}^{j\Delta_n}c_s ds}_{\ge\epsilon_c\Delta_n}\left(\underbrace{\frac{\int_{(j-1)\Delta_n}^{j\Delta_n}\sqrt{c_s}dW_s}{\sqrt{\int_{(j-1)\Delta_n}^{j\Delta_n}c_s ds}}}_{\sim\mathcal{N}(0,1)}\right)^2\\
&\ge \frac{1}{k_n}\chi_{k_n,i}^2
\end{aligned}
$$

and hence

$$
\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-6}\right)\le k_n^6\mathbb{E}_c\left(\frac{1}{\chi_{k_n}^2}\right)^6 \overset{k_n\ge 12}{\sim} k_n^6\cdot k_n^{-6}\sim O(1).
$$

Now we come to focus on $\sqrt{\mathbb{E}_c\left(\hat{c}_{i\Delta_n} - c_{i\Delta_n}\right)^8)}$.

$$
\begin{aligned}
\mathbb{E}_c(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^8 &= \mathbb{E}_c(S_{i,n} + D_{i,n})^8 \\
&\leq \mathbb{E}_c(S_{i,n}^8) + D_{i,n}^8
\end{aligned}
$$

where as for $D_{i,n}^8$,

$$
\begin{aligned}
D_{i,n}^8 &= \left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} c_s ds - c_{i\Delta_n}\right)^8 \\
&\leq \left(\frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} \left(c_s - c_{i\Delta_n}\right) ds\right)^8 \\
&\leq \frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} \left(c_s - c_{i\Delta_n}\right)^8 ds \\
&\leq \frac{1}{k_n\Delta_n}\int_{i\Delta_n}^{(i+k_n)\Delta_n} \left((K_n+1)\Delta_n\right)^{8\rho} ds \\
&\sim (k_n\Delta_n)^{8\rho} \\
&\sim (k_n^{-\frac{1}{2}})^8 \sim k_n^{-4}.
\end{aligned}
$$

For $\mathbb{E}_c(S_{i,n}^8)$, recall

$$
\mathbb{E}_c(S_{i,n}^8) = \mathbb{E}_c\left(\frac{1}{k_n}\sum_{j=i+1}^{i+k_n} Z_{i,j,n}\right)^8
$$

with $Z_{i,j,n}$ being centered and independent as j varies. In particular, any moments of $Z_{i,j,n}$ is of order 1 (behaves like constant), independent of i and j. Therefore,

$$
\begin{aligned}
\mathbb{E}_c(S_{i,n}^8) \quad &= \mathbb{E}_c\left(\frac{1}{k_n}\sum_{j=i+1}^{i+k_n} Z_{i,j,n}\right)^8 \\
&= \frac{1}{k_n^8}\mathbb{E}_c\Bigg(\sum_{j=i+1}^{i+k_n} Z_{i,j,n}^8 \\
&\qquad + \sum_{j\neq m} Z_{i,j,n}^6 Z_{i,m,n}^2 + \sum_{j\neq m} Z_{i,j,n}^5 Z_{i,m,n}^3 + \sum_{j\neq m} Z_{i,j,n}^4 Z_{i,m,n}^4 \\
&\qquad + \sum_{j\neq m\neq k} Z_{i,j,n}^4 Z_{i,m,n}^2 Z_{i,k,n}^2 + \sum_{j\neq m\neq k} Z_{i,j,n}^3 Z_{i,m,n}^3 Z_{i,k,n}^2 \\
&\qquad + \sum_{j\neq m\neq k\neq \ell} Z_{i,j,n}^2 Z_{i,m,n}^2 Z_{i,k,n}^2 Z_{i,\ell,n}^2\Bigg) \\
&\sim \frac{1}{k_n^8} k_n^4 \sim k_n^{-4},
\end{aligned}
$$

as the four additive terms are of order $k_n, k_n(k_n-1), k_n(k_n-1)(k_n-2)$ and $k_n(k_n-1)(k_n-2)(k_n-3)$, respectively. Taking all above together, we have

$$
\begin{aligned}
\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-3}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^4\right) \quad &\leq \quad \sqrt{\mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-6}\right)}\sqrt{\mathbb{E}_c(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^8} \\
&\sim \quad O(1)\sqrt{O(k_n^{-4})} \\
&\sim \quad O(k_n^{-2})
\end{aligned}
$$

which implies

$$
\frac{\Delta_n}{\alpha_{T,n}^2}\sum_{i=0}^{[T/\Delta_n]-1} \mathbb{E}_c\left(\hat{c}_{i\Delta_n}^{-3}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^4\right) \sim O(\frac{1}{k_n}) \longrightarrow 0.
$$

Consequently, we prove

$$
\mathbb{E}_c\left(\frac{1}{\alpha_{T,n}}\sum_{i=0}^{[T/\Delta_n]-1} \xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n} - c_{i\Delta_n})^2\Delta_{i+1}^n\widetilde{W}\right)^2 \longrightarrow 0
$$

and hence

$$\frac{1}{\alpha_{T,n}}\sum_{i=0}^{[T/\Delta_n]-1}\xi_{i,n}^{-\frac{3}{2}}(\hat{c}_{i\Delta_n}-c_{i\Delta_n})^2\Delta_{i+1}^n\widetilde{W}\xrightarrow{\mathbb{Q}(\omega,\cdot)}0,$$

where $\mathbb{Q}(\omega,\cdot)$ is a transition probability kernel, given the volatility path ω being conditioned on. Thus by Slutsky's theorem, the result follows. □

Continuation of Proof of Theorem 6.4.2: Now we come to decomposition 6.8. At this stage we can blow up our original target $Y_T^n-\widetilde{X}_T$ by

$$\sqrt{k_n}\sim\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}},$$

which gives

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}(Y_T^n-\widetilde{X}_T)\\
=\quad&\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\sum_{i=0}^{[T/\Delta_n]-1}(\sqrt{\hat{c}_{i\Delta_n}}-\sqrt{c_{i\Delta_n}})\Delta_{i+1}^n\widetilde{W}\\
&+\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\left(\sum_{i=0}^{[T/\Delta_n]-1}\sqrt{c_{i\Delta_n}}\Delta_{i+1}^n\widetilde{W}-\int_0^{[\frac{T}{\Delta_n}]\Delta_n}\sqrt{c_s}d\widetilde{W}_s\right)\\
&+\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\left(\int_0^{[\frac{T}{\Delta_n}]\Delta_n}\sqrt{c_s}d\widetilde{W}_s-\int_0^T\sqrt{c_s}d\widetilde{W}_s\right)\\
=\quad&\underbrace{\left(\alpha_{T,n}\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\right)}_{\text{as constant}}\underbrace{\frac{1}{\alpha_{T,n}}\sum_{i=0}^{[T/\Delta_n]-1}(\sqrt{\hat{c}_{i\Delta_n}}-\sqrt{c_{i\Delta_n}})\Delta_{i+1}^n\widetilde{W}}_{\to\mathcal{N}(0,1),\text{ by Proposition 6.7.1}}\\
&+\underbrace{\left(\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\Delta_n^{\rho}\right)}_{\to 0}\underbrace{\Delta_n^{-\rho}\left(\sum_{i=0}^{[T/\Delta_n]-1}\sqrt{c_{i\Delta_n}}\Delta_{i+1}^n\widetilde{W}-\int_0^{[\frac{T}{\Delta_n}]\Delta_n}\sqrt{c_s}d\widetilde{W}_s\right)}_{\text{tight}}\\
&+\underbrace{\left(\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}\Delta_n^{\frac{1}{2}}\right)}_{\to 0}\underbrace{\Delta_n^{-\frac{1}{2}}\left(\int_0^{[\frac{T}{\Delta_n}]\Delta_n}\sqrt{c_s}d\widetilde{W}_s-\int_0^T\sqrt{c_s}d\widetilde{W}_s\right)}_{\text{tight}}
\end{aligned}$$

where we have used the facts that the exact convergence rate for $\alpha_{T,n}$ is $\sqrt{k_n} \sim \Delta_n^{-\frac{1}{2+\frac{1}{\rho}}}$, and that for $\rho > 0$,

$$\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}} \Delta_n^{\rho} = \Delta_n^{(-\frac{1}{2\rho+1}+1)\rho} = \Delta_n^{\frac{2\rho}{2\rho+1}\rho} \to 0,$$

$$\Delta_n^{-\frac{1}{2+\frac{1}{\rho}}} \Delta_n^{\frac{1}{2}} = \Delta_n^{\frac{1}{2}-\frac{1}{2+\frac{1}{\rho}}} \to 0.$$

6.7.4 Proof of Proposition 6.4.1

When c_t is constant, namely the pathwise continuity is of $\rho > 1$, we have for $0 \le t \le T$,

$$\widetilde{X}_t = \int_0^t \sqrt{c} d\widetilde{W}_s = \sqrt{c}\widetilde{W}_t.$$

In light of (6.7), there exist neither the target error induced from spot volatility estimation nor the discretization error, and as a result the convergence rate (upper bound) becomes

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{n,\delta} - \widetilde{X}_t|\right) &\le \mathbb{E}\left(\sup_{0\le t\le T}\left|\sum_{i=0}^{[t/\delta]-1} (\sqrt{\hat{c}_{i\delta}} - \sqrt{c})(\widetilde{W}_{(i+1)\delta} - \widetilde{W}_{i\delta})\right|\right) \\
&+ \mathbb{E}\left(\sup_{0\le t\le T} \sqrt{c}|\widetilde{W}_t - \widetilde{W}_{\delta[\frac{t}{\delta}]}|\right) \\
&\le K\left(\frac{1}{\sqrt{k_n}} + \left(\delta \log(\frac{2T}{\delta})\right)^{\frac{1}{2}}\right).
\end{aligned}$$

6.7.5 Proof of Theorem 6.4.3

First note that

$$\begin{aligned}
\mathbb{E}\left(\sup_{0\le t\le T} |Y_t^{RV} - \widetilde{X}_t|\right) &= \mathbb{E}\left(\sup_{0\le t\le T}\left|\sqrt{\frac{RV}{T}}\widetilde{W}_t - \sqrt{c}\widetilde{W}_t\right|\right) \\
&= \mathbb{E}\left(\sup_{0\le t\le T}\left|\sqrt{\frac{RV}{T}} - \sqrt{c}\right| |\widetilde{W}_t|\right) \\
&= \mathbb{E}\left(\left|\sqrt{\frac{RV}{T}} - \sqrt{c}\right| \sup_{0\le t\le T} |\widetilde{W}_t|\right) \\
&= \mathbb{E}\left(\left|\sqrt{\frac{RV}{T}} - \sqrt{c}\right|\right) \mathbb{E}\left(\sup_{0\le t\le T} |\widetilde{W}_t|\right),
\end{aligned}$$

where the last equality is due to independence. Now, on one hand, we have

$$\left(\mathbb{E}\left(\sup_{0\le t\le T}|\widetilde{W}_t|\right)\right)^2 \le \mathbb{E}\left(\sup_{0\le t\le T}|\widetilde{W}_t|\right)^2 = \mathbb{E}\left(\sup_{0\le t\le T}|\widetilde{W}_t|^2\right) \le 4\mathbb{E}(\widetilde{W}_T^2) < \infty,$$

where the two inequalities are due to Jensen's inequality and Doob's inequality, respectively.

On the other hand, we have

$$\mathbb{E}(RV) = \mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]}(\Delta_i^n X)^2\right) = \sum_{i=1}^{[T/\Delta_n]}\mathbb{E}\left((\Delta_i^n X)^2\right) = \sum_{i=1}^{[T/\Delta_n]}c\Delta_n = c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n$$

and

$$\begin{aligned}
\mathbb{E}(RV)^2 &= Var(RV) + (E(RV))^2 = Var\left(\sum_{i=1}^{[T/\Delta_n]}(\Delta_i^n X)^2\right) + (E(RV))^2 \\
&= \sum_{i=1}^{[T/\Delta_n]} Var\left((\Delta_i^n X)^2\right) + \left(c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n\right)^2 \\
&= \sum_{i=1}^{[T/\Delta_n]} c^2\Delta_n^2 Var(\chi_1^2) + \left(c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n\right)^2 \\
&= 2c^2\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n^2 + \left(c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n\right)^2.
\end{aligned}$$

Therefore, it holds that

$$\begin{aligned}
\left\{\mathbb{E}\left(\left|\sqrt{\frac{RV}{T}}-\sqrt{c}\right|\right)\right\}^2 &\le \mathbb{E}\left(\left|\sqrt{\frac{RV}{T}}-\sqrt{c}\right|^2\right) \\
&= \frac{1}{T}\mathbb{E}\left(\sqrt{RV}-\sqrt{Tc}\right) \\
&\le K\mathbb{E}(RV-Tc)^2 \\
&= K\left(E(RV)^2 - 2TcE(RV) + (Tc)^2\right) \\
&= K\left(2c^2\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n^2 + \left(c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n\right)^2 - 2Tc^2\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n + (Tc)^2\right) \\
&= K\left(2c^2\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n^2 + \left(c\lfloor\frac{T}{\Delta_n}\rfloor\Delta_n - Tc\right)^2\right) \\
&\le K\left(2c^2T\Delta_n + \Delta_n^2\right),
\end{aligned}$$

which implies

$$\mathbb{E}\left(\left|\sqrt{\frac{RV}{T}}-\sqrt{c}\right|\right) \leq K(\Delta_n+\Delta_n^2)^{\frac{1}{2}} \leq K(\sqrt{\Delta_n}+\Delta_n),$$

as desired.

6.7.6 Proof of Theorem 6.4.4

Note $\widetilde{W}_0 = 0$ and that

$$\begin{aligned} Y_T^{RV} - \widetilde{X}_T &= \sqrt{\frac{RV}{T}}(\widetilde{W}_T-\widetilde{W}_0) - \sqrt{c}(\widetilde{W}_T-\widetilde{W}_0) \\ &= \left(\sqrt{\frac{RV}{T}}-\sqrt{c}\right)(\widetilde{W}_T-\widetilde{W}_0) \\ &= \frac{1}{\sqrt{T}}(\sqrt{RV}-\sqrt{Tc})(\widetilde{W}_T-\widetilde{W}_0). \end{aligned}$$

By a classical result from integrated volatility estimation (see, e.g.,(Aït-Sahalia and Jacod, 2014)), we have the following central limit theorem

$$\frac{1}{\sqrt{\Delta_n}}(RV-Tc) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2Tc^2).$$

By Delta's method ($f(x)=\sqrt{x}, f'(x)=\frac{1}{2\sqrt{x}}$), it follows that

$$\begin{aligned} \frac{1}{\sqrt{\Delta_n}}(\sqrt{RV}-\sqrt{Tc}) &\xrightarrow{\mathcal{L}} \mathcal{N}(0,2Tc^2)\frac{1}{2\sqrt{Tc}} \\ &\sim \mathcal{N}(0,\frac{c}{2}). \end{aligned}$$

Notice that $\forall n$, $\frac{1}{\sqrt{\Delta_n}}(\sqrt{RV}-\sqrt{Tc})$ is independent of $\widetilde{W}_T-\widetilde{W}_0$, then

$$\left(\frac{1}{\sqrt{\Delta_n}}(\sqrt{RV}-\sqrt{Tc}), \widetilde{W}_T-\widetilde{W}_0\right) \xrightarrow{\mathcal{L}} \left(\mathcal{N}(0,\frac{c}{2}), \mathcal{N}(0,T)\right).$$

By continuous mapping theorem, it follows

$$\frac{1}{\sqrt{\Delta_n}}(\sqrt{RV}-\sqrt{Tc})(\widetilde{W}_T-\widetilde{W}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\frac{c}{2})\mathcal{N}(0,T).$$

Therefore,

$$
\begin{aligned}
\frac{1}{\sqrt{\Delta_n}}\left(Y_T^{RV} - \widetilde{X}_T\right) &= \frac{1}{\sqrt{\Delta_n}}\frac{1}{\sqrt{T}}(\sqrt{RV} - \sqrt{Tc})(\widetilde{W}_T - \widetilde{W}_0) \\
&\xrightarrow{\mathcal{L}} \frac{1}{\sqrt{T}}\mathcal{N}(0, \frac{c}{2})\mathcal{N}(0, T) \\
&\sim \mathcal{N}(0, \frac{c}{2})\mathcal{N}(0, 1)
\end{aligned}
$$

BIBLIOGRAPHY

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. and Xiu, D. (2015). Principal component analysis of high frequency data. Technical report.

Aldous, D. and Eagleson, G. (1978). On mixing and stability of limit theorems. *Annals Probab.*, 6:325–331.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Rangebased estimation of stochastic volatility models. *The Journal of Finance*, 57:1047–1091.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2001a). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96:42–55.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003a). Modelling and forecasting realized volatility. *Econometrica*, 71:579–625.

Andersen, T. G. and Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The journal of Finance*, 52(3):975–1005.

Andersen, T. G. and Bollerslev, T. (1999). "answering the skeptics: Yes, standard volatility models do provide accurate forecast. *International Economic Review*, 39:885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003b). Modeling and forecasting realized volatility. *Econometrica*, 71(2):pp. 579–625.

Asmussen, S. (2003). *Applied Probability and Queues*. Springer-Verlag, New York.

Back, K. (1991). Asset pricing for general processes*. *Journal of Mathematical Economics*, 20:371–395.

Barndorff-Nielsen, O. and Shephard, N. (2003). Realized Power Variation and Stochastic Volatility Models. *Bernoulli*, 9:243–265.

Barndorff-Nielsen, O. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., and Shephard, N. (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory*, 22:677–719.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2004c). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2:1–37.

Bertoin, J. (1998). *Lévy Process*. Cambridge University Press.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31:307–327.

Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2):409–424.

Chung, K. and Williams, R. (1990). *Introduction to Stochastic Integration*. Birkhäuser, Boston, 2nd edition.

Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics*, 73:101–149.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323.

Duncan, T. (2006). Prediction for some processes related to a fractional brownian motion. *Statistics & Probability Letters*, 76:128–134.

Duncan, T. E., Hu, Y., and Pasik-Duncan, B. (2000). Stochastic calculus for fractional brownian motion i. theory. *SIAM J. Control Optim.*, 38:582612.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom. *Econometrica*, 50:987–1008.

Fink, H., Klüppelberg, C., and Zähle, M. (2013). Conditional distributions of processes related to fractional brownian motion. *Journal of Applied Probability*, 50:166–183.

Fischer, M. and Nappo, G. (2010). On the moments of the modulus of continuity of itö processes. *Stochastic Analysis and Applications*, 28:103–122.

Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, 64:139–174.

Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, 70:419–435.

Gonçalves, S. and Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica*, 77(1):283–306.

Gripenberg, G. and Norros, I. (1996). On the prediction of fractional brownian motion. *Journal of Applied Probability*, 33:400–410.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Review of Financial Studies*, 6:327–343.

Huang, S. and Cambanis, S. (1978). Stochastic and multiple wiener integrals for gaussian processes. *The Annals of Probability*, 6:585614.

Hurst, H. (1951). Long-term storage capacity in reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:400410.

Hurst, H. (1956). Methods of using long-term storage in reservoirs. *Proc. Inst. Civil Engineers Part I, Chapter 5*, pages 519–590.

Jacod, J. (1997). On continuous gaussian martingales and stable convergence in law. *Sminaire de Probabilits XXXI*, page 169212.

Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic processes and their applications*, 118(4):517–559.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the euler method for stochastic differential equations. *Annals of Probab*, 26:297307.

Jacod, J. and Protter, P. (2012). *Discretization of Processes.* Springer.

Jacod, J. and Rosenbaum, M. (2013a). Quarticity and Other Functionals of Volatility: Efficient Estimation. *Annals of Statistics*, 41:1462–1484.

Jacod, J. and Rosenbaum, M. (2013b). Quarticity and Other Functionals of Volatility: Efficient Estimation. *Annals of Statistics*, 118:1462–1484.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York, second edition.

Kalnina, I. and Xiu, D. (2016). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association.*

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, Berlin, second edition.

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Springer.

Kolmogorov, A. N. (1940). Wienersche spiralen und einige andere interessante kurven im hilbertschen raum. c. r. *(Doklady) Acad. URSS (N.S.)*, 26:115118.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference.* Springer Science & Business Media.

Li, J., Todorov, V., and Tauchen, G. (2016a). Adaptive estimation of continuous-time regression models using high-frequency data. Technical report, Duke Univeristy.

Li, J., Todorov, V., and Tauchen, G. (2016b). Inference theory for volatility functional dependencies. *Journal of Econometrics*, 193(1):17–34.

Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, 84:1613–1633.

Li, J. and Xiu, D. (2017). Efficient estimation of integrated volatility functionals via multiscale jacknife.

Liu, Y. and Li, J. (2016). Bootstrap inference for integrated volatility functionals. Technical report, UNC Chapel Hill and Duke University.

Mancini, C. (2001). Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion. *Giornale dell'Istituto Italiano degli Attuari*, LXIV:19–47.

Mancini, C. (2009). Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps. *Scandinavian Journal of Statistics*, 36:270–296.

Mandelbrot, B. (1983). *The Fractal Geometry of Nature.* W. H. Freeman and Company, San Francisco, CA.

Mémin, J., Mishura, Y., and Valkeila, E. (2001). Inequalities for the moments of wiener integrals with respect to a fractional brownian motion. *Statistics & Probability Letters*, 51:197–206.

Merton, R. (1992). *Continuous Time Finance*. Basil Blackwell, New York, N.Y.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8(4):323–361.

Mishura, Y. S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Springer, Berlin Heidelberg.

Mykland, P. and Zhang, L. (2006). ANOVA for Diffusions and Ito Processes. *Annals of Statistics*, 34:1931–1963.

Mykland, P. and Zhang, L. (2009). Inference for Continuous Semimartingales Observed at High Frequency. *Econometrica*, 77:1403–1445.

Nualart, D. (2005). *The Malliavin Calculus and Related Topics*. Springer-Verlag, New York, USA, 2nd edition.

Nualart, D. (2006). Fractional brownian motions: Stochastic calculus and applications. *Proceedings of the International Congress of Mathematicians, Madrid, Spain.*

P., H. and Heyde, C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Phillips, P. and Ouliaris, S. (1990). Asymptotic properties of residual based test for cointegration. *Econometrica*, 58:165–193.

Pipiras, V. and Taqqu, M. S. (2000). Integration questions related to fractional brownian motion. *Probability Theory and Related Fields*, 118:251–291.

Pipiras, V. and Taqqu, M. S. (2001). Integration questions related to fractional brownian motion. *Bernoulli*, 7(6):873–897.

Pipiras, V. and Taqqu, M. S. (2016). *Long-Range Dependence and Self-Similarity.* Cambridge University Press.

Reiss, M., Todorov, V., and Tauchen, G. (2015). Nonparametric test for a constant beta between ito semimartingales based on high-frequency data. *Stochastic Processes and their Applications*, 125:2955–2988.

Rényi, A. (1963). On stable sequences of events. *Sankya, Series A*, 25:293–302.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, Germany, third edition.

Samko, S.G., A. K. and Marichev, O. (1993). *Fractional Integrals and Derivatives*. Gordon and Breach Science Publishers, New York.

Todorov, V. and Tauchen, G. (2012a). Inverse realized laplace transforms for nonparametric volatility density estimation in jump-diffusions. *Journal of the American Statistical Association*, 107(498):622–635.

Todorov, V. and Tauchen, G. (2012b). The Realized Laplace Transform of Volatility. *Economtrica*, 80:1105–1127.

Todorov, V., Tauchen, G., et al. (2012). Realized laplace transforms for pure-jump semimartingales. *The Annals of Statistics*, 40(2):1233–1262.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.